



**POLITECNICO**  
**MILANO 1863**

Master of Science in Management Engineering

***A novel model to assess the impact of surgeons’  
fatigue on surgery outcome***

Supervisor:

**Professor Paolo Trucco**

Co-supervisor:

**Professor Rossella Onofrio**

Candidates:

**Giada Marani 900469**

**Margherita Munari 905558**

Academic year 2019/2020

*To nonna, for the phone calls, the laughs and for her unconditional love*

*To mum and dad, for being my role models and staying by my side since day one*

*To my family, for overcoming the distance between us with their constant presence*

*To my second family, Mara, Matti, Mario, Alu and Eli, for being my home away from home*

*To my flatmates, for the everlasting memories in Mac Mahon 104*

*To Maggie, for your attitude towards life and for being the sister I never had*

*To the old and new friends, thank you.*

*Giada*

*To my family, for allowing me to undertake this journey with my head held high and never making me feel second to no one*

*To nonna Carla, for gifting me with unconditional happiness and love at my every result*

*To my travel companions, for paving the path towards this goal with lightheartedness and memories*

*To the four Italians that made Compiegne an unforgettable experience*

*To the five Veronesi with whom I shared the ups and downs of the last years*

*And lastly to Giada without whom reaching this goal wouldn't have been possible.*

*Margherita*

*We would like to thank our supervisor Professor Paolo Trucco and co-supervisor Professor Rossella Onofrio for giving us the opportunity to develop this thesis. Furthermore, we would like to thank Dr. Antonio Galfano, Dr. Luciano Cristofori and Dr. Isabella Frigerio for their time, availability and helpfulness.*

*Giada & Margherita*

## Abstract

In the healthcare context, committing an error could mean endangering human lives. The number of errors made by doctors is so high, it makes them one of the most frequent causes of death. The intent of this thesis was to study fatigue in its broadest manifestations and to develop an instrument that would lead reactive philosophy, now present in hospitals, to shift towards proactivity. This means having an a priori approach that makes it possible to predict errors and reduce the probability of their occurrence. From a literary point of view, the gap in the definition of fatigue, in its measurement and application, especially in the healthcare context, is evident. A model has been developed with a dual offline and real-time application. It was then applied to a real case of laparoscopic prostatectomy at the Niguarda Hospital in Milan. The objective of this thesis is to allow the a priori (offline) identification of surgical procedures at risk. For these tasks, the measurement in real-time of the level of fatigue of surgeons takes place, in order to establish whether they are authorized to operate or not.

Keywords: fatigue, healthcare, surgery, Human Reliability Analysis, human error, proactive, performance

Nel contesto sanitario commettere un errore può voler significare mettere in pericolo vite umane. Gli errori commessi dai medici sono pari ad un numero così elevato che li porta ad essere una delle cause più frequenti di morte. L'intento della tesi è stato quello di studiare la fatica nelle sue più ampie sfaccettature e sviluppare uno strumento che portasse la filosofia reattiva, oggi presente negli ospedali, a diventare proattiva. Questo significa avere un approccio preventivo che consenta di prevedere gli errori e ridurre la loro probabilità di occorrenza. Da un punto di vista letterario, il gap nella definizione di fatica, nella sua misurazione e nella sua applicazione, soprattutto in contesto sanitario, è evidente. È stato sviluppato un modello con una

duplice implementazione di tipo offline e real-time. È stato poi applicato in una operazione chirurgica reale di prostatectomia laparoscopica all'ospedale Niguarda di Milano. L'obiettivo di questa tesi è rendere possibile l'identificazione a priori (offline) di procedure chirurgiche a rischio e per questi task misurare in real-time il livello di affaticamento dei chirurghi per stabilire se a loro sia concesso di operare o meno.

Keywords: fatica, sanità, chirurgia, Human Reliability Analysis, errore umano, proattivo, prestazione

# Table of contents

<i>Abstract</i>	3
<i>Table of contents</i>	5
<i>Index of Figures</i>	8
<i>Index of Tables</i>	10
<i>Acronyms</i>	11
<b>1. Introduction</b>	<b>13</b>
1.1 Problem statement	16
<b>2. State of art on Human Reliability Analysis</b>	<b>22</b>
2.1 Overview and evolution of Human Reliability Analysis	23
2.2 Performance Shaping Factors	26
2.3 Human Reliability Analysis in healthcare	27
<b>3. Literature review</b>	<b>30</b>
<b>3.1 Literature search methodology</b>	<b>31</b>
3.1.1 Prisma Flow Diagram	31
3.1.2 Background context	32
3.1.3 Database search strategy	32
<b>3.2 Content analysis</b>	<b>33</b>
3.2.1 General overview	35
3.2.2 Fatigue	36
3.2.3 Human performance	41
3.2.4 Night vs day	44
3.2.5 Prefrontal Cortex	45
3.2.6 Proactive vs reactive	48
3.2.7 Naturalistic Decision Making	49

3.2.8 Complex vs simple tasks	54
3.2.9 Wearable devices	55
3.2.10 Standardization	57
<b>3.3 Literature gap</b>	<b>60</b>
<b>3.4 Aim of the work</b>	<b>62</b>
<b>4. Model development</b>	<b>66</b>
<b>4.1 Model overview</b>	<b>67</b>
<b>4.2 Functional Job Analysis</b>	<b>69</b>
4.2.1 Task analysis	70
<b>4.3 Horizontal axis: task complexity</b>	<b>73</b>
4.3.1 Definition of task complexity	74
4.3.2 The study of complexity – Functional Job Analysis	74
4.3.3 Questionnaire	77
4.3.3.1 Final score and ranges	81
<b>4.4 Vertical axis: fatigue impact</b>	<b>82</b>
4.4.1 Definition of fatigue impact	82
4.4.2 Convergent and divergent thinking	83
4.4.2.1 Convergent thinking	83
4.4.2.2 Divergent thinking	83
4.4.2.3 Joint role of convergent and divergent thinking	84
4.4.3 Questionnaire	86
4.4.3.1 Explanation and examples in medical context	88
4.4.3.2 Final score and ranges	92
<b>4.5 Integrating Human Error Probability in the model</b>	<b>93</b>
<b>4.6 Off-line mapping analysis</b>	<b>94</b>
<b>4.7 Real time analysis – the fatigue level evaluation framework</b>	<b>96</b>
4.6.1 Epworth Sleepiness Scale	98

4.6.2 Modified Fatigue Impact Scale	100
4.6.4 Pittsburgh Sleep Quality Index	102
4.6.5 Psychomotor Vigilance Test	103
4.8 The model flow chart	108
<b>5. Model validation and calibration</b>	<b>113</b>
5.1 Overview of the process	114
5.2 Validation of the model	115
5.2.1 Fatigue impact scale validation	115
5.2.2 Task complexity scale validation	115
5.3 Calibration of the model in the robotic surgical context	117
5.3.1 Bocciardi-Approach Robot-Assisted Radical Prostatectomy procedure	118
5.3.2 Ex-ante feature calibration	121
5.3.2.1 Fatigue impact scale	122
5.3.2.2 Task complexity scale	123
5.3.3 Scenarios hypotheses	123
5.3.3.1 Measure of reliability of the task complexity scale	126
5.3.3.2 Measure of reliability of the Fatigue impact scale	126
5.4 Real-time feature assessment	127
<b>6. Discussion</b>	<b>132</b>
<b>7. Conclusion</b>	<b>137</b>
<b>Bibliography</b>	<b>140</b>
<b>Appendix 1: Epworth Sleepiness Scale</b>	<b>150</b>
<b>Appendix 2: Modified Fatigue Impact Scale</b>	<b>151</b>
<b>Appendix 3: Pittsburgh Sleep Quality Index</b>	<b>152</b>
<b>Appendix 4: PSQI Scoring Instructions</b>	<b>154</b>

## Index of Figures

Figure 1: Prisma Flow Diagram of the literature review process _____	31
Figure 2: Papers published on fatigue, performance and HRA topics _____	35
Figure 3: Papers published on fatigue, performance and HRA topics in healthcare _____	35
Figure 4: Most recurrent journals included in the study _____	36
Figure 5: Papers published about fatigue in healthcare context _____	37
Figure 6: Fatigue causes in healthcare _____	38
Figure 7: Mental fatigue measurement methods _____	40
Figure 8: Most used metrics for fatigue evaluation overall _____	40
Figure 9: Most used metrics for fatigue evaluation in healthcare _____	40
Figure 10: Papers published about performance in healthcare context _____	41
Figure 11: Terms used for performance overall _____	42
Figure 12: Terms used for performance in healthcare _____	42
Figure 13: Most used metrics for performance overall _____	43
Figure 14: Most used metrics for performance in healthcare _____	43
Figure 15: Prefrontal Cortex _____	45
Figure 16: Gambling Task results (Killgore, Balkin, & Wesensten, 2006) _____	47
Figure 17: Recognition-Primed Decision Model (Klein, 1993) _____	52
Figure 18: How mental fatigue is measured through wearables _____	56
Figure 19: Location of the wearable device on the body _____	56
Figure 20: Fatigue properties _____	60
Figure 21: Fatigue properties and emerging needs _____	63
Figure 22: Fatigue Impact - Task Complexity model _____	67
Figure 23: Horizontal axis of the FI-TC model: task complexity _____	73
Figure 24: Functional Job Analysis scale dimensions: Things, Data, People (Fine, 2014) _____	76
Figure 25: Questionnaire for the assessment of task complexity (part I) _____	79
Figure 26: Questionnaire for the assessment of task complexity (part II) _____	80
Figure 27: Vertical axis of the FI-TC model: fatigue impact _____	82
Figure 28: Consequences of differing combinations of divergent and convergent thinking (Cropley, 2006) _____	85
Figure 29: Questionnaire for the assessment of task fatigue impact level _____	87
Figure 30: HEP in the FI-TC model _____	93



Figure 31: The definition of the FI-TC model "zones": green, orange and red	95
Figure 32: Fatigue level evaluation framework	96
Figure 33: Fatigue Severity Scale (Flachenecker et al., 2002)	101
Figure 34: Psychomotor Vigilance Test execution phases	104
Figure 35: PVT performance under TSD conditions	107
Figure 36: The model flow chart diagram	110
Figure 37: The process in the evaluation of the FI-TC model	114
Figure 38: Differences in incision techniques in prostatectomy	118
Figure 39: The three components of DaVinci robot: the console, the robotic arm cart and the cart view.	119
Figure 40: Scenario hypothesis in the robotic application 1 – coherence and overlapping of results	124
Figure 41: Scenario hypothesis in the robotic application 2 – coherence and overlapping of Fatigue Impact results	124
Figure 42: Scenario hypothesis in the robotic application 3 – coherence and overlapping of Task Complexity results	125
Figure 43: Scenario hypothesis in the robotic application 4 - dispersion of results	125
Figure 44: Questionnaire for subjective fatigue assessment tools evaluation	129
Figure 45: The development process of FI-TC model creation	133
Figure 46: Comparative results of IF perception in two different surgical context (Onofrio & Trucco, 2018).	135

## Index of Tables

Table 1: Number practicing medical doctors per hundred thousand inhabitants (Eurostat) _____	16
Table 2: Number of surgeons per hundred thousand inhabitants (Eurostat)_____	17
Table 3: PSFs and corresponding HRA methods (Pan & Wu, 2018) _____	27
Table 4: Database search keywords and results _____	33
Table 5: Functional Job Analysis Rating scales and brief definitions (Fine, 2014)____	75
Table 6: Summary of the steps for the calculation of the final score task complexity evaluation _____	81
Table 7: Effect sizes of outcome metrics for acute TSD and chronic PSD (by Basner & Dinges, 2011) _____	105
Table 8: Performance score on PVT _____	108
Table 9: Task complexity scale weighted by expert surgeons_____	116
Table 10: Fatigue impact scale dimentions _____	122
Table 11: Results of the questionnaire on subjective tool quality _____	130

## Acronyms

HCEG	Healthcare Executive Group
EWTD	European Working Time Directive
BMA	British Medical Association
FRMS	Fatigue Risk Management System
HRA	Human Reliability Analysis
BAC	Blood Alcohol Concentration
HEP	Human Error Probability
THERP	Technique for Human Error Prediction
ASEP	Accident Sequence Evaluation Program
HCR	Human Cognitive Reliability
HEART	Human Error Assessment and Reduction Technique
OATS	Operator Action Trees System
JHEDI	Justification of Human Error Data Identification
SLIM	Success Likelihood Index Method
MTO	Men-Technology Organization
CREAM	Cognitive Reliability and Error Analysis Method
SPAR-H	Standardized Plant Analysis Risk-Human Reliability Analysis
ATHEANA	A Technique for Human Event Analysis
CES	Cognitive Environmental Simulation
CAHR	Connectionism Assessment of Human Reliability
COSIMO	Cognitive Simulation Model
NARA	Nuclear Action Reliability Assessment
PSF	Performance Shaping Factor
AE	Adverse Event
OCHRA	Observational Clinical Human Reliability Analysis
EDIT	Error type, Direct treat and Indirect Treat
CAT	Competency Assessment Tool
OSCE	Objective Structured Clinical Examination

PVT	Psychomotor Vigilance Test
EEG	Electroencephalogram
EOG	Electrooculography
ECG	Electrocardiogram
KSS	Karolinska Sleepiness Scale
PFC	Prefrontal Cortex
SD	Sleep Deprivation
IGT	Iowa Gambling Task
VM	Ventromedial
RPD	Recognition-Primed Decision
FJA	Functional Job Analysis
NTS	Non-Technical Skills
HTA	Hierarchical Task Analysis
SHERPA	Systematic Human Error Reduction and Prediction Approach
ESS	Epworth Sleepiness Scale
ASP	Average Sleep Propensity
SSP	Situational Sleep Propensity
FIS	Fatigue Impact Scale
MFIS	Modified Fatigue Impact Scale
FSS	Functional System Scores
VAS	Visual Analog Scale
MS	Multiple Sclerosis
PSQI	Pittsburgh Sleep Quality Index
RT	Reaction Time
TSD	Total Sleep Deprivation
PSD	Partial Sleep Deprivation
EJE	Expert Judgement Elicitation
RARP	Robot-Assisted Radical Prostatectomy
BA-RARP	Bocciardi-Approach Robot-Assisted Prostatectomy
MIS	Minimally invasive surgery

# 1. Introduction

---

Chapter number one consists of an introduction of the thesis and the problem statement. The problem statement aims to outline the reasons which gave rise to the need of dealing with the topics treated in the present work.

---

The healthcare context can be defined as a high risk and complex system. Struggle between life and death is usual in surgeons' daily life where jobs are performed under stressing conditions. Numerous are the responsibilities on their shoulders and an error committed may be the cause of serious consequences regarding patients' health. The work shifts imposed on surgeons and on-call doctors do not facilitate the doctors' services, which are extremely delicate. Being a doctor today involves numerous sacrifices as night shift, little free time and unsocial hours which leads to tiredness, stress, drowsiness etc. Tiredness impairs medical members' alertness and the ability to safely operate.

The topic of fatigue has been studied for decades since it has been recognized as the cause for the majority of errors in high risk contexts as aviation, nuclear, petroleum, transportation and healthcare. From the study of literature, it has emerged that fatigue is qualitatively difficult to be defined and quantitatively difficult to be measured. Few papers propose a definition of fatigue, which is also referred to as a synonym of sleep deprivation or lack of sleep. It has many different ways of manifesting and can be caused by numerous factors which means also that fatigue impacts differently on individuals.

With the aim of reducing risks and increasing safety, high risk contexts have already implemented tools and programs for managing fatigue, considering also all the other influencing factors of operators' performance. One of the most used method is called Human Reliability Analysis. It was born in the industrial sector, but it is now developed in other numerous fields. The principal characteristic of HRA is its systematic approach, through which human contribution is seen within the technical and organizational context. The aspect that makes it interesting is its proactive approach in defining ex-ante Human Error Probabilities related to specific contexts.

The HRA application in healthcare is of extreme importance in the identification of the most critical factors that could influence surgeons' performance. Since 1980, numerous studies tried to adapt and translate the HRA technique adopted in high risk

contexts to healthcare and they succeeded in constructing modified versions. What was exposed by the literature study, is the presence of numerous validated versions of HRA methods to be applied to healthcare but with some limitations in the definition of the so-called Influencing Factors, more precisely in the consideration of fatigue. Fatigue is included in the analysis as one of the Influencing Factors, but it is not developed in all of its dimensions. It is underestimated and not always considered.

Since the enormous impact that fatigue has been proven to have on operators, this thesis concerns the development of a model that consider both the Influencing Factors of HRA in healthcare, and the analysis of fatigue independently, considering it at 360°. The model proposed is structured to be proactive: providing in advance suggestions about the criticality of tasks performed by surgeons and if necessary, assessing the fatigue level of surgeons in real-time.

In order to validate the model, it was applied to a real surgical operation. For its high complexity and robustness, the Bocciardi-Approach Robot-Assisted Radical Prostatectomy (BA-RARP) was chosen. The BA-RARP procedure is a high technological operation which requires an extremely high level of mental load and human-machines interactions.

The difficult situation of the past few months caused by Coronavirus, impaired the final phase of validation and calibration of the model. A large amount of data and feedback from doctors working at Niguarda Hospital in Milan would have been needed to draw useful and justified conclusions. Since the number of patients in intensive care has increased dramatically, hospitals were totally reorganized, and doctors had no spare time to respond to questionnaires.

## 1.1 Problem statement

In the healthcare system while costs rise, and workplaces are limited, patients increase. Costs are seen as the first challenge and issue for 2020, as announced by the HealthCare Executive Group annual forum. There is the need for investments but the lack of capital. Italy is one of the European countries with the lowest capital invested in healthcare (Ferrucci, 2018). The consequences are observed in the lack of beds, nurses and doctors which are forced to work ridiculous hours out of their normal shiftwork in order to satisfy demand represented by the increasing amount of admitted patients.

The table below represents the number practicing medical doctors per hundred thousand inhabitants – the first eight European countries are shown.

	2014	2015	2016	2017
<b>Austria</b>	504.61	509.12	512.96	518.28
<b>Norway</b>	442.92	440.35	451.22	465.85
<b>Lithuania</b>	430.74	433.92	446.69	455.63
<b>Switzerland</b>	412.58	419.71	425.06	429.78
<b>Germany</b>	410.82	413.93	418.65	424.88
<b>Bulgaria</b>	398.69	404.54	413.76	424.49
<b>Italy</b>	388.04	383.83	395.27	398.95

Table 1: Number practicing medical doctors per hundred thousand inhabitants (Eurostat)

Italy is in 7<sup>th</sup> position behind Austria, Norway, Lithuania, Switzerland, Germany and Bulgaria. Italy in 2017 had 398.95 practicing medical doctors per hundred thousand inhabitants. More than the number itself, it is interesting to see how in these four years this ratio incremented only by 3% – the lowest increment together with Austria.

Even more evident is the data regarding the number of surgeons per hundred thousand inhabitants shown below.



	2014	2015	2016	2017
<b>Greece</b>	136.18	134.87	139.53	131.56
<b>Germany</b>	112.69	114.91	116.38	118.21
<b>Bulgaria</b>	94.44	103.83	101.32	108.54
<b>Lithuania</b>	103.09	103.65	106.86	107.38
<b>Cyprus</b>	89.15	95.44	102.99	106.69
<b>Austria</b>	100.05	103.24	101.55	103.15
<b>Italy</b>	94.53	94.21	98.68	99.88

Table 2: Number of surgeons per hundred thousand inhabitants (Eurostat)

Italy in fact, has yet to overcome the 100 surgeons/hundred thousand inhabitants' threshold and is again behind six European states.

Thirty-two percent of the total American healthcare and social assistance workforce – 16 million individuals – reported they do not have enough sleep (Caruso, 2014). Especially first aid and emergency medicine doctors, internists, anesthesiologists, specialists in general surgery who work up to 48 hours in a row, are subject to this lack of rest. Sleep loss results essentially in the degradation of diverse aspects of functioning, including cognitive processes, vigilance, physical coordination, judgment and decision making, communication, outlook, and numerous other parameters. It can even lead to injuries, obesity and a wide range of chronic diseases (Caruso, 2014). Patients' safety is undermined, and the theme of clinical risk is a topical issue. This is proven by a recent study (Makary & Daniel, 2016) that estimated medical errors as the third biggest cause of death in hospitalizations in the United States. In Italy there are about 80-90 deaths in a day with 320 thousand damaged people a year due to medical errors ("Sanità, gli errori dei medici provocano 90 morti al giorno," 2006). Given the fact that fifty percent of the errors are potentially preventable (Contino, 2018), there is a moral duty to look for a solution that reduces the impact of human error in the treatment path. One of the possible solutions concerns to safeguard

medical workforce putting limitations on their work shift calendar avoiding extended wakefulness and sleep loss which can impair medical members' alertness and ability to safely operate.

Several changes have been done in the European and US laws underlying a sense of confusion in what is good and what is not for the health of doctors, the safety of patients and the demand satisfaction. In the United States, the Council in charge of setting rules for training doctors scrapped the 16-hour cap on shifts worked by doctors who have just graduated from medical school. So, in 2017 they decided to extend working shift allowing interns to work again up to 28-hours shifts. Dr. Paul E. Sax, a clinical director of the Division of Infectious Diseases at Brigham and Women's Hospital, and a professor at Harvard Medical School, responded to the maneuver of US Council with a declaration on WBUR Radio describing his detrimental experience as an intern when the typical shift was 28 hours long. He and his colleagues were on call every third or fourth night as interns, they were tired, the work-life balance was messed up and they tended to look at patients as components of an assembly line work (Dr. Paul E. Sax, 2017). Something different happened in Europe where in August 2009 the full implementation of a European Working Time Directive (EWTD) settled out the legal parameters for minimum rest-breaks, holidays, night-time work and a maximum workweek of 48 h (Fitzgerald & Caesar, 2012). The Italian Parliament excluded hospital physicians from the application of the Directive that resulted in extended and excessive work time for the category. It was after five years that the Italian Parliament declared the willingness to defend the right of medical workforce issuing the law 161 which followed the directive of 2009 (Gnerre et al., 2017).

Associations and journals have never stopped monitoring hospitals and interviewing doctors to ensure the success of the directive. In the last decade several newspapers and doctors themselves have denounced the critical condition of medical workforce highlighting the violation of the law. The British Medical Association (BMA) survey of junior doctors showed that more than half still worked more than 56 h per week, to fill rota gaps or because of perceived pressure to work additional hours (The Lancet,

2010). Furthermore, there are doctors that directly admit the riskiness that sleep deprivation and fatigue can have on patients' safety, condemning the Local Health Authority for the grueling working conditions (Lupino, 2016). In 2018, the term "auto-resignation" was coined by the vice-president of FNOMCeO referring to the process through which doctors are quitting their job due to the causes listed above (Cavalcanti, 2018). It is fair to say that the introduction of laws, with the objective to safeguard doctors' health, and consequently patients' safety, limiting shift working hours, have not been effective in the solution of the problem.

The healthcare sector given its unique characteristics has always been deemed a stand-alone context. This is true to some extent, but from many other points of view the techniques and methods adopted in other industries – such as the manufacturing context – can and must be modified to be adopted in the medical field. In the past years the focus has always been on the differences between the healthcare context and all the other sectors. The main difference is represented by the enormously vast and diverse set of activities of which healthcare is made up of: surgery, primary care, emergency medicine just to mention a few. Another important difference is its very high unreliability even when compared to many other hazardous industries. This is because work in sectors such as nuclear power tends to be routine and departures from normality are avoided. In healthcare on the other hand, unpredicted events are part of the daily life. Lastly, more than any other industry, the healthcare context relies on human-human interactions and human-machine ones are only secondary or specific to certain areas. Having said this, a change of point of view should be implemented and the possible similarities and parallelism to leverage must be highlighted. The general structure of a healthcare system is similar to that of aviation, nuclear power and petroleum industries: complex, hazardous and very large. The healthcare context, as other sectors, is defined as a high risk and complex system, in which human error is the most contributing factor in the 80% of incidents (Hollnagel, 1993). High risk condition means a condition with a potential to threaten human life, properties, health, or the environment (Oliver, 2013). The work of anesthetists given

its high vigilance level and technology surveillance, can be compared to that of commercial pilots. Emergency medicine can be likened to military or fire-fighting rapid response teams (Lyons, Adams, Woloshynowych, & Vincent, 2004).

Sectors such as aviation, nuclear plant, petroleum industry and transportation have a very extensive background literature that investigates on Human Reliability Analysis and Risk Analysis methods both in researches and real application. On the contrary, healthcare literature lacks that knowledge. The first publication about safety concerns in industrial sectors in regards to working hours of transportation operators dates back to 1930s (Gander et al., 2011). Since that moment up to the end of 20<sup>th</sup> century, numerous researches have been done with the scope to understand the causes of human errors and a way to decrease them. It was found out there were several stimuli in the working environment for human error and it is believed that fatigue is one of them. A study by NASA revealed that 70% of all aviation accidents were related to fatigue (Givi & Jaber, 2014). In transportation, the National Highway Traffic Safety Administration estimates that 100 000 crashes are caused by drowsy drivers, which results in more than 1500 fatalities and 71 000 injuries each year in the U.S. Regarding the naval context, a 1996 U.S. Coast Guard analysis of 279 incidents showed that fatigue contributed to 16% of critical vessel casualties and 33% of personal injuries (Ji, Lan, & Looney, 2006). This data has been the fuse in developing fatigue management models with the objective to reduce errors. As previously said, there are already applications of these models in certain sectors except healthcare. In aviation, for example, in 2011 the table of contents of the implementation guide for operators was published. The reason of its publication was the introduction of Fatigue Risk Management System, which aims to ensure that flight and cabin crew members are sufficiently alert, so they can operate to a satisfactory level of performance. Traditionally, crewmember fatigue had been managed through prescribed limits on maximum flight and duty hours, based on a historical understanding of fatigue through simple work and rest period relationships. New knowledge related to the effects of sleep and circadian rhythms provides an additional dimension to the

management of fatigue risks. An FRMS provides a means of adding this safety dimension, allowing operators to work both safer and more efficiently (Adriana Patricia Muñoz Zapata, 2011). This is only one of the examples of the successful realization and application of fatigue management systems.

Given the extremely positive progress obtained in only fourteen years now, the hope is that industrial knowledge of complex socio-technical systems can be imported to contribute to patient safety. The benefits of transferring and applying risk analysis methods to healthcare services are fully recognized in the patient safety literature (Cagliano, Grimaldi, & Rafele, 2011; Lyons, 2009; Lyons et al., 2004; Verbano & Turra, 2010; Charles Vincent, Taylor-Adams, & Stanhope, 1998). The models have been modified in order for them to be applied in a context that has different variables and characteristics with respect to the one for which they were created. The first techniques imported have been Root Cause Analysis and Failure Mode and Effect Analysis. Because of some limitations, alternative solutions, like Human Reliability Analysis, were proposed by specialists (Lyons, 2009; Tang, Hanna, Joice, & Cuschieri, 2004; Verbano & Turra, 2010). The main problem is that in the models considered the most adaptable to the healthcare context, fatigue is not well developed. Up to now the variable has only had a marginal impact on the predicting process whilst it should be having a contribution of primary importance.

## 2. State of art on Human Reliability Analysis

---

In this chapter the literature on Human Reliability Analysis is presented. An in depth understanding of HRA in the industrial sector together with its evolution and subsequent application in the healthcare field is introduced.

---

## 2.1 Overview and evolution of Human Reliability Analysis

Human Reliability Analysis (HRA) techniques “aim to determine the impact of human error on a system. The techniques draw on systems engineering and cognitive and behavioral science methods to understand and evaluate the human contribution to system reliability and safety” (Sujan, Embrey, & Huang, 2018). Therefore, the novelty of HRA techniques is their anticipatory method of analysis that changes completely the original approach used by healthcare, from day one tied to ex-post approaches using incident reporting techniques. The study of human reliability is defined as a hybrid scientific specialized subfield related to psychology, ergonomics, reliability analysis, engineering, and system analysis.

To determine the impact within HRA means to identify which kind of errors humans could make, model the system into which they operate and quantify the probability of human errors. The nature of Human Error Probability is quantitative and is the result of operators’ activities calculation. It is determined and influenced by Performance Shaping Factors which can be defined as any aspects of human, environmental, organizational or task characteristics that affect human performance, positively or negatively. The definition of PSFs is one of the most important steps in the implementation of HRA because, providing numerical basis, they influence the final nominal result of HEP levels. Considering that HRA comprises the study of human behavior under a cognitive aspect, the second key step is the interpretation of human behavior and the simulation of it which includes studies of physical and psychological factors. The multidimensionality of human behavior makes it difficult to be defined. Several studies over the years have tried to obtain a final quantitative value that indicates the probability of error to predict and prevent unsafe behaviors.

Human Reliability Analysis was born around the years 50’ and 60’. Subsequently to disasters that happened from 1950 to date, HRA spread all over the world, undergoing transformations that lead to the diversification of techniques creating a wide umbrella of HRA methods. The first implementation of HRA was in the aeronautical

field, answering the need of examinee tasks, processes, systems and organizational structures, with the aim to identify weaknesses of the system and give a numerical value to human reliability. The spread of HRA followed the tragedy of Three Mile Island, the most serious nuclear accident in the United States. In 1983 the *U.S. Nuclear Regulatory Commission* used the *Technique for Human Error Rate Prediction* (Swain & Guttman, 1983) as principal method for determining the reliability of the operators. At the same time the focus on human error and the collection of data on performances increased, leading to the diffusion of first-generation methods: THERP, ASEP, HCR, HEART, PHRA, OATS, JHEDI, SLIM (Pasquale, Iannone, Miranda, & Riemma, 2013). Because of their structure they are also called “behavioral techniques”, which is also the reasons why they are so criticized. The analysis of the event is purely descriptive and based on formal external behavioral characteristics of the errors’ analysis, without considering the reasons and cognitive mechanisms that induce the errors. Basically, they see men as a mechanical component. In this way, the aspects of dynamic interaction with the working environment, both physical and social, are lost. Many are the criticisms for the lack of and superficial consideration of some PSFs which are considered as minor factors. HEPs, with the aim of identifying the characteristics of the operators’ tasks, are defined a major factor. THERP and HCR are the most used methods of the first-generation models overcoming the criticisms thanks to their ease of use and highly quantitative aspects.

The separation of techniques from the first to the second-generation is driven by the evolution of the concept of error with the development of cognitive human models, which enable the creation of cognitive methodologies on human reliability. The errors are no more seen as something caused by human or by technical failures, but as a consequence of the interaction of more components in relation to each other and with the environment. Components have technological, human and organizational natures. The focus on quantitative valuations of the first-generation shift to a qualitative assessment of human error, analyzing through developed cognitive models the multi interactions between operators and production processes at



different levels. Considering the dependence of operators from personal factors (such as level of knowledge, stress, etc.) the actual situation, and from the different kind of man-machine interfaces, humans in this case have to be seen in the integrated system of men-technology-organizations (Pasquale et al., 2013). Another difference that characterizes second-generation models is the deeper consideration of PSFs. Previously they were used for describing the impact of the environment on the operators, whereas now they become indicators for the cognitive impact on operators. In first-generation HRA PSFs were not adequately treated and none of the methods tried to explain the effect of PSFs on performance. The most used second-generation techniques are: CREAM, SPAR-H, ATHEANA, CES, CAHR, MERMOS.

There exists a third-generation or next-generation of HRA techniques, born in the last years, raised by the development of pre-existing methodologies. Working on weaknesses and limitations of the second-generation methodologies, experts improved existing models and did new researches. Those HRA methods based on the Cognitive Simulation Models are considered a part of this classification (Pan, Lin, & He, 2017). One of the methods classified in this category is Nuclear Action Reliability Assessment, which is in fact an advanced version of the HEART for the nuclear fields.

Independently from the generation of the methodology, HRA analysis is developed in two phases:

- qualitative phase, an informal analysis of the tasks with the aim to identify and model the possible human errors and the contribution to the errors
- quantitative phase, in which the probability of occurrence of human errors effects is estimated.

## 2.2 Performance Shaping Factors

The increase in complexity of the systems, due to the higher number of machines and people involved, makes the human cognition an input which is no more negligible. Failures become more complex and the analysis develops a broader search of causes. Performance Shaping Factors introduced by HRA methodologies are adopted to describe the variety of human-machine interaction.

A PSF can be defined as an environmental, personal, or task-oriented factor that influences human performance in complex systems. Improving or worsening the performance, these factors impact the probability of human error. Examples of psychological and physiological PSFs are repetitive work, displeasure, task speed, confusion, distraction, emergency situations, high temperature and fatigue (Griffith & Mahadevan, 2015). They are also defined as contributing factors and are divided in two classes: internal and external. The external factors are the circumstances in which the task is performed. The internal factors indicate the psychological and physiological conditions of the operators. The environment in which the operators perform is made up by these two typologies of factors, also called context. HRA's objective is to analyze the context finding causes of human errors and measure the errors according to these causes. Depending on the HRA techniques the name of these contributing factors changes. The firsts that introduced the PSFs were Swain and Guttman developing the THERP methodology. It followed an evolution of the term even if today, PSFs is the most used and common name for contributing and influencing factors on operators' performance. The table below shows some of the different appellation for different methods that consider different factors. There are significant differences between PSFs of various HRAs and there is no uniform standard for their selection (Pan & Wu, 2018).

Appellation PSFs	Corresponding HRA method	Considered factors
PSFs	THERP	(1) External PSFs: working environment factors, task manuals, equipment use list, etc. (2) Internal PSFs: experience, qualifications, skills, physical condition, personality, knowledge, team recognition and ability, etc. (3) Stress level: occupation stress, etc.
	HCR	Operator experience, stress level and quality of operator–plant interface
	SLIM SPAR-H	A series of important PSFs that are determined with experts' judgment Available time, stress, complexity, training, procedures, work efficiency, appropriateness of duties and working procedure
PIFs	ASEP	Same as THERP
	HDT ADS-IDAC	4–8 factors in general, such as procedures, human–machine interface and leader commanding quality Cognitive modes and tendencies, emotional arousal, strains and feelings, perception and appraisal, intrinsic characteristics, memorized information, organizational factors, team-related factors, conditioning events, environmental factors and physical factors
CPCs	CREAM	Adequacy of organization, working conditions, adequacy of MMI/operational support, availability of procedures/plans, number of simultaneous goals, available time, time of day (circadian rhythm), adequacy of training/expertise, crew collaboration quality
EPCs	HEART	38 EPCs, such as unfamiliarity with the situation (the situation may be important; however, it rarely occurs), lack of available time for error checking and correction, low SNR, susceptible to a strong message or characteristics, etc.
EFCs	ATHEANA	(1) PSFs: human–machine interface, stress level, etc. (2) Organizational factors/common power-plant conditions: power plant configuration, instrument control availability, etc.
PAFs	CREATE	PSFs that are involved in activating and utilizing knowledge, and in influencing human problem-solving behavior

Note: ADS-IDAC = accident dynamic simulator with the information decision and action in crew context; ASEP = accident sequence evaluation program; ATHEANA = a technique for human event analysis; CPC = common performance condition; CREAM = cognitive reliability and error analysis method; CREATE = cognitive reliability assessment technique; EFC = error forcing context; EPC = error producing condition; HCR = human cognitive reliability; HDT = holistic decision tree; HEART = human error assessment and reduction technique; HRA = human reliability analysis; MMI = man–machine interactions; PAF = performance affecting factor; PIF = performance influence factor; PSF = performance shaping factor; SLIM = success likelihood index method; SNR = signal-to-noise ratio; SPAR-H = standardized plant analysis risk human reliability analysis; THERP = technique for human error rate prediction.

Table 3: PSFs and corresponding HRA methods (Pan & Wu, 2018)

Every PSF is calculated on the base of some rules according to the HRA methodology chosen for the analysis. Sometimes the direct measure or the extrapolation is not possible for the evaluation of the influencing factor because it is too complicated or too abstractive that an expert judgement is needed. Experiments, audits, field experiences are needed as direct valuation method to measure the state of PSFs.

### 2.3 Human Reliability Analysis in healthcare

The introduction of HRA techniques in healthcare was fostered by the evolution of the mindset of human error and the development of technology in surgery. Even though healthcare has been rapidly evolving from a medical point of view trying to understand patients' needs, on the other hand it has neglected to deepen all the

aspects on patient safety, consequently related to human errors. It is only in the last years that the sector felt the need to identify initiatives with the aim of increasing patient safety and decreasing errors. The need comes from the analysis of incident reporting, used to collect and analyze data about past patient harms, which records that one out of ten patients admitted to hospital could be victim of an Adverse Event. Of these AE, more than 50% involve surgical procedures (C. Vincent, Neale, & Woloshynowych, 2001). In this field the development of technology brought a revolution with its climax in Minimally Invasive Robotic Surgery where the technology changed completely the way of operating. With this new way of doing practices, surgery has become more complex and riskier than before. The level of technological complexity and the interaction of a plurality of professional figures, with a new concept of human error, seen as cause for organizational-managerial insufficiencies of the system, requires the implementation of a systemic approach to manage the risk. This typology of approach, unlike the others, considers the organizational context into which the incidents are verified. For this reason, the application in surgery of security knowledge, that usually was adopted in industrial sector, has recently gained interest.

The field that has seen the broadest application of HRA is surgery. The first real application in the field was made by Joice, Hanna, & Cuschieri, 1998 of the *Department of Surgery and Surgical Skills Unit of University of Dundee*. They studied errors made during an endoscopic surgery and classified them with a method called subsequently *Observational Clinical Human Reliability Analysis (OCHRA)*, which derives from HRA but is based on direct observation of surgical procedures. The same method was applied to further studies which saw in OCHRA a potentially useful method. The objectives were to evaluate surgeons' competencies, describe the learning curve and the frequency and the typology of errors committed during the execution of particular tasks or laparoscopic procedures (Tang et al., 2004).

As previously mentioned, HRA techniques were born in the industrial field and saw their development in aviation, nuclear plants, petroleum industry and transportation. The applications in healthcare derived from the modification of original HRA methodologies and their adaptation to the specific context. Until today, the techniques that have been modified and applied in this field are:

- *Observational Clinical Human Reliability Analysis* – OCHRA
- *Human Error Assessment and Reduction Technique* – HEART modified
- *Error type, Direct threat, and Indirect threat* – EDIT
- *Competency Assessment Tool Score* – CAT
- *Objective Structured Clinical Examination* – OSCE

The principal objective of these procedures is to record the nature and the incidence of technical errors, through the observation and the categorization of skill-based errors. Another important goal is the identification of frequency and patterns of technical errors to vary surgeons experience that do the procedures. Furthermore, PSFs are fundamental for the evaluation of HEP and the formulation of HRA. Because the costs associated with adverse risks are very large, the application of HRA, with the identification of errors and possible remedies, is also implemented for obtaining the reduction of expenditures.

### 3. Literature review

---

Chapter number three is structured to present the phase of literature research and literature content explanation. The literature search methodology is explained, and all gathered literature is presented and discussed in detail. Conclusions are drawn from the content analysis and explicated in the literature gap which acts as a bridge between literature review and the model development. Finally, from the gaps found in literature the aims of the presented study are outlined and explored.

---

### 3.1 Literature search methodology

The literature review was structured in two initial horizontal steps combined with four vertical phases. Horizontally the database search of documents was joined with the initially supplied articles relevant for a background study. The four vertical dimensions follow the phases of the Prisma Flow Chart model presented below.

#### 3.1.1 Prisma Flow Diagram

To outline the overall process followed in the study of literature the Prisma Flow Diagram is presented below.

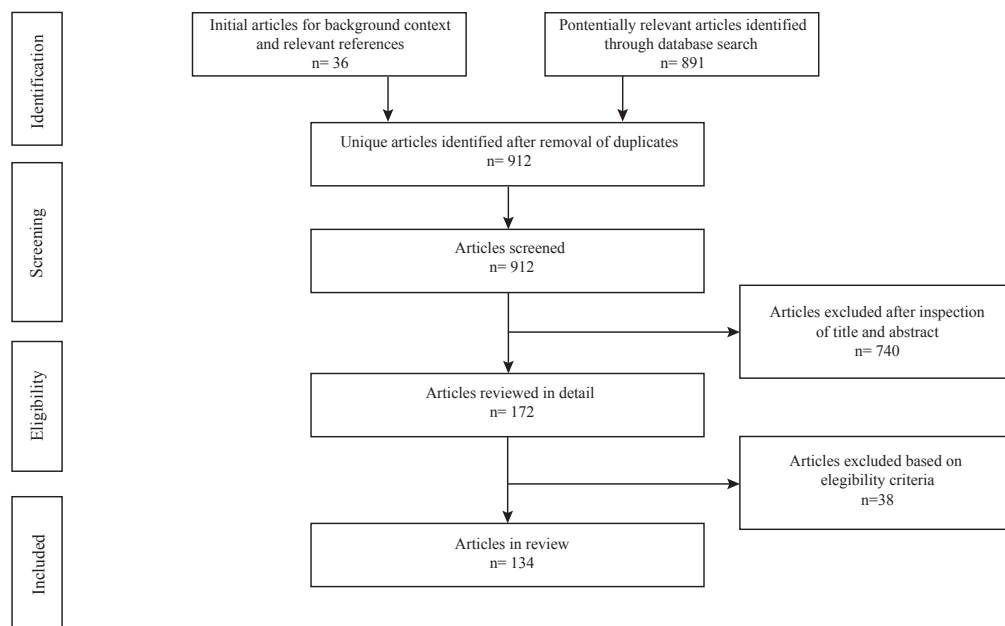


Figure 1: Prisma Flow Diagram of the literature review process

The systematic review started with two different steps: the identification of a basket of initial documents and their relative references and the database research. After the download of all the documents the duplicates were removed, and the remaining were screened. Through the analysis of the title and the abstract the documents

coherent with the scope of research were selected and the ones out of scope were deleted. After the screening, 172 documents were read and 38 of them discarded due to eligibility criteria. The final set of documents that were deemed relevant for the research were 134.

### 3.1.2 Background context

To gain a general knowledge of the context, a first reading was made on a selected number of papers which provided an overview of the most common words used in the field of fatigue, performance and wearable devices.

References of this initial basket of documents were reviewed and those coherent with the scope of research were screened. This first step was essential for making an effective choice of the keywords that were used for the database search.

A strong relationship between the words fatigue and sleep deprivation emerged as they are used in most cases as synonymous; sleep deprivation and shift work are often mentioned together, showing they are strictly interconnected with fatigue and performance mostly in the healthcare context; furthermore BAC (Blood Alcohol Concentration) appears to be a commonly used practice to evaluate human capabilities and performance in sleep deprived subjects.

### 3.1.3 Database search strategy

A systematic search was then performed on Scopus. PubMed and Web of Science were discarded after their limited quantitative results during the first researches. Restrictions based on date of publication were applied only to searches involving wearable devices, where the most recent from 2010 coupled with the most cited were chosen. The search was limited to publications written in English for all queries.

The keywords, their combination and number of results are presented in Table 4.



Database	Query used	Results
Scopus	"Fatigue assessment" AND "human performance"	1
	Fatigue AND "human reliability analysis"	10
	Fatigue AND "human performance" AND healthcare	4
	Fatigue AND "human performance" AND surgery	10
	"Human performance" AND healthcare	101
	"Human reliability analysis" AND healthcare	22
	"Sleep deprivation" AND "human performance"	72
	Bac AND "sleep deprivation"	21
	Circadian AND performance AND healthcare	146
	"Sleep deprivation" AND "human error"	31
	"Sleep deprivation" AND surgeon AND performance	65
	"Shift work" AND healthcare AND performance	46
	"Wearable technology" OR "wearable devices" OR "wearable sensors" AND fatigue	340

Table 4: Database search keywords and results

The studies were screened by reviewing the title and the abstract. 711 were excluded after this brief review based on evident mismatches between the context and aim of the present work, and the results gained from the database.

### 3.2 Content analysis

The content analysis paragraph presents all the topics and related aspects that came out from the literature review analysis. The topics described in the sub-chapters below, have been considered important for the development of the thesis. Given the breadth of the topics covered, a summary of what will be presented is proposed:

- *General overview* shows the numbers of papers published inherent with the thesis topic and their relative journals
- *Fatigue* describes fatigue properties, the causes of fatigue (sleep deprivation, shift work and circadian and homeostatic rhythms), and the metrics utilized for its measurement
- *Human performance* presents human performance definitions and the metrics used for its measurement. The connection between fatigue and performance is highlighted
- *Night vs day* explains how differently fatigue impacts on the individual during different parts of the day. The differences between acute and chronic sleep loss and the importance of the circadian and homeostatic rhythms in performing well are evidenced
- *Prefrontal Cortex* describes what the Prefrontal Cortex is, where it is positioned and how it can be used to develop knowledge about fatigue context. It presents the deep dependences between Prefrontal Cortex, sleep deprivation and performance.
- *Naturalistic decision making* describes the way through which human beings make decisions and how much it is related to the PFC and influenced by sleep deprivation
- *Proactive vs reactive* presents the differences between reactive and proactive approaches and the advantages in applying the latter to implement error risk mitigation in healthcare
- *Complex vs simple task* describes how differently the fatigue impact is on individuals whilst performing complex and simple tasks
- *Wearable devices* offer an overview on wearable technologies potentiality in health status measurement followed by a description on how the devices are able

to measure mental fatigue, pointing out the part of the body in which they are applied

- o *Standardization* evidences how much healthcare lacks standardization and why this sector should develop in that direction. This chapter is more about hospital organization rather than fatigue or performance and aims to highlight problems and lacks in current healthcare.

### 3.2.1 General overview

A general overview of the gathered information and data retrieved from the literature review is of fundamental importance to have a bird eye view of all analyzed content.

The topic of fatigue and performance related to Human Reliability Analysis, as said before, is acquiring importance and relevance year by year. In Figures 2 and 3 below the number of papers per years can be seen.

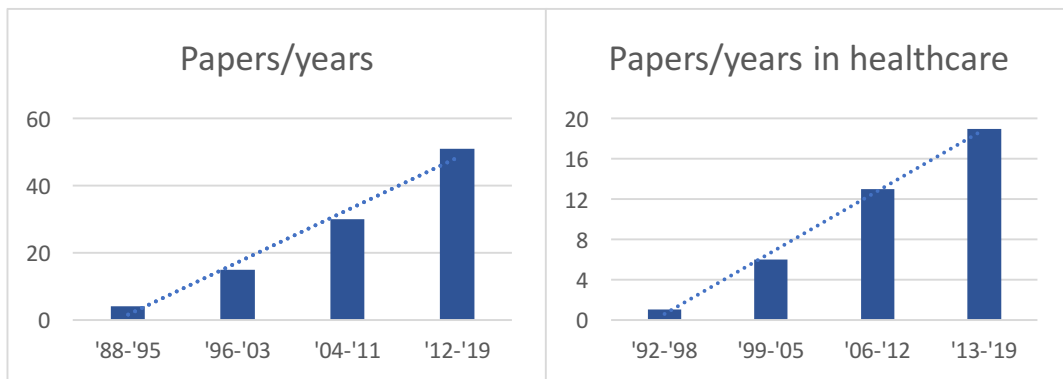


Figure 2: Papers published on fatigue, performance and HRA topics

Figure 3: Papers published on fatigue, performance and HRA topics in healthcare

Figure 2 groups all analyzed papers and it shows a steeply increasing trend. The numbers of papers from '88-'95 to '12-'19 has literally increased by ten times. Figure 3 on the other hand, focuses on the healthcare context and the steadily increasing trend line shows how even in this subgroup an increase of ten times in the number of papers in under 30 years has occurred. By looking closely to the y-axis, papers in the

healthcare context are just one third of total papers which shows how, specifically in the medical area, not enough attention has been brought to the topic.

The 134 papers included in the review come from more than 70 different journals. Some journals were more recurrent than others as Figure 4 shows.

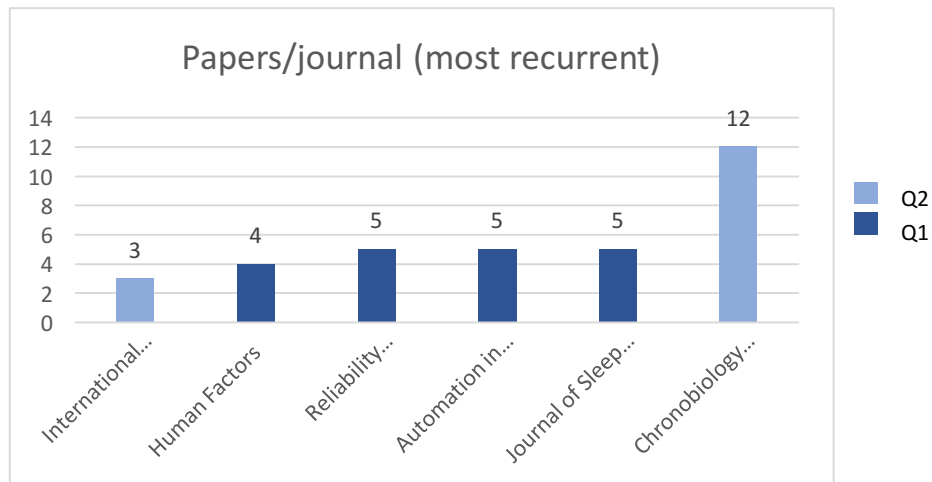


Figure 4: Most recurrent journals included in the study

With the use of *Scimago Journal and Country Rank* the validity of the journal was evaluated. Only those papers featured in Q1 and Q2 journals were deemed to be of sufficient reliability and solidity and were included in the review. *Chronobiology International* (Q2) was the most recurrent journal followed by *Journal of Sleep Research* (Q1), *Automation in Construction* (Q1) and *Reliability Engineering and System Safety* (Q1).

### 3.2.2 Fatigue

83 articles out of the 134 included in this review treated the topic of fatigue – without considering the 31 that referred specifically to wearable devices which will be dealt with in the following chapters. Of these 83 articles only 33 were specific to the healthcare context. Figure 5 highlights again how the topic of fatigue in healthcare is becoming of increasing importance and relevance in the last decade. Whilst only 9

papers treated the argument in the first ten years of the 20<sup>th</sup> century this number virtually tripled in the following ten.

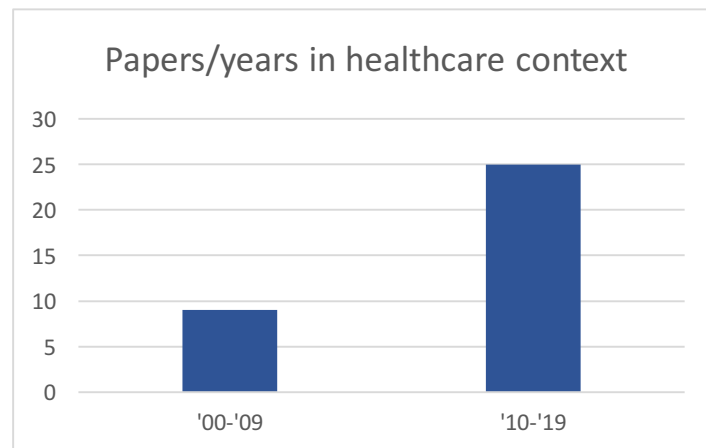


Figure 5: Papers published about fatigue in healthcare context

Fatigue is multi-dimensional and is mainly defined by listing its symptoms such as tiredness, lack of energy, sleepiness, drowsiness, lack of motivation, memory lapses, reduced concentration and reduced alertness (Dean, Scott, & Rogers, 2006; Jaber, Givi, & Neumann, 2013; Shen et al., 2006; Winwood, Winefield, Dawson, & Lushington, 2005). The Oxford English Dictionary defines fatigue as the lassitude or weariness resulting from either bodily or mental exertion. A similar definition is given by Gander, Millar, Webster, & Merry, 2008 “Fatigue is the inability to function at one’s optimum level because physical and mental exertion (of all waking activities, not only work) exceeds current capacity.” From these two first definitions it results clear how fatigue is divided in two typologies: mental – or cognitive – and physical – or physiological – fatigue. Physical fatigue can manifest as a reduction in strength or muscle power (Griffith & Mahadevan, 2011) up to physical exhaustion (Peters & Peters, 2007). On the other hand cognitive fatigue is seen as an overwhelming feeling of tiredness (Griffith & Mahadevan, 2011), altered concentration, slowed reaction time, and reduced problem-solving abilities (Dean et al., 2006). Both types of fatigue may manifest as a temporary state called acute fatigue. An ongoing acute fatigue in

healthy individuals may though evolve into a chronic condition which may represent a long-term illness. (Barker & Nussbaum, 2011)

Another predominant method for defining this extremely intricated and multi-factorial term which fatigue is, is by focusing on its causes. Through the in-depth screening of the articles three elements emerged as its main consequences (Figure 6): sleep deprivation, shift work and circadian and homeostatic rhythms.

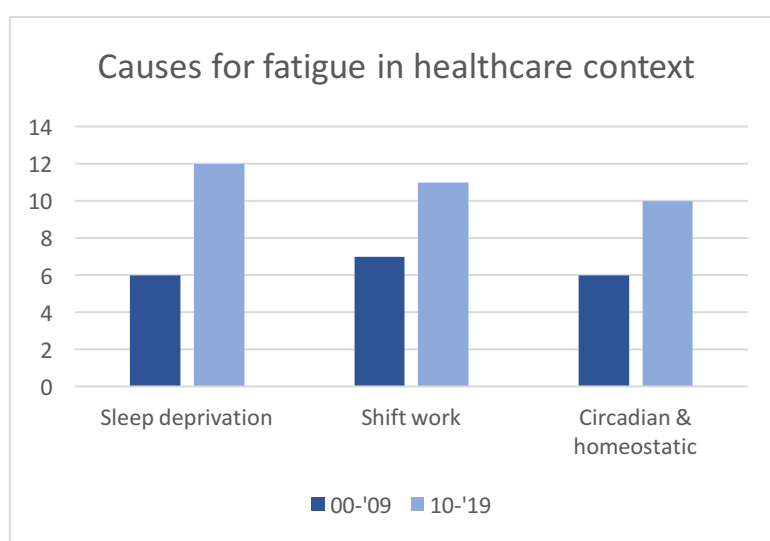


Figure 6: Fatigue causes in healthcare

It can be seen from Figure 6 how sleep deprivation emerges as the main cause of human fatigue closely followed by shift work and circadian & homeostatic rhythms. It is interesting to see how these three elements are closely intertwined, literally one causing the other. In simple terms: shift work is a common cause of sleep deprivation in individuals which itself is caused by the misalignment of circadian and homeostatic cycles given by prolonged working hours and compromised day-night cycles caused by shift work.

When speaking of *sleep deprivation* there are many controversial opinions ranging from one extreme to the other. Fatigue for many is better classified as sleep

deprivation (Patterson, Lochtefeld, Larson, & Christensen-Salem, 2019; Posada-Quintero, Bolkhovsky, Qin, & Chon, 2018; Rose & Giray, 2013), using these terms as exact synonyms. For others fatigue is seen as the direct effect of a lack of sleep or extended wakefulness (Cohen et al., 2010; Dean et al., 2006; Gaba & Howard, 2002; Goel, Basner, Rao, & Dinges, 2013; Gunzelmann, Gross, Gluck, & Dinges, 2009; Mehta et al., 2016; Parry, Oeppen, Amin, & Brennan, 2018; Pasquale et al., 2013; Sturm et al., 2011). Lastly in one paper fatigue and sleepiness are referred to as distinct and independent phenomena (Shen et al., 2006).

Dall’Ora, Ball, Recio-Saucedo, & Griffiths, 2016 set out a scoping review of 35 papers to analyze in-depth the characteristics of *shift work* that have effect of employees’ performance and well-being. This review resulted in finding six characteristics which impacted on employees, them being: shift length, weekly work hours, compressed working week, overtime, night work/rotating or fixed shifts and rest opportunities. Even though conflicting results are reported, it emerges overall that long shifts ( $\geq 12$  hours) and more than 40 weekly hours of shift work may have negative effects on performance. Overtime and night shift working are associated with increased errors and the circadian misalignment impacts on the ability to perform during these night shifts. This misalignment may be counteracted through fixed shifts as permanent night work may reduce the disruption of the sleep-wake schedule.

The two-process model refers to the *homeostatic and circadian* processes which are combined to determine sleep-wake cycles. The homeostatic process represents the drive for sleep: it increases exponentially during wakefulness and decreases at a similar rate whilst sleeping. It is based on thresholds: when the homeostat reaches a certain threshold sleep is triggered, when it decreases below a certain threshold, wakefulness is set off. The circadian process refers to daily oscillatory between these thresholds and promotes more wakefulness than sleep. It is diverse from individual to individual and is revealed by characteristics that define the so called “morning people”, which tend to be more fresh in the early hours of the day, or more “night people”, which do not feel the drive to sleep until late at night (Goel et al., 2013).

Mental fatigue in order to be evaluated must be measured through specific metrics. Metrics are divided in subjective and objective – or instrument-based ones, as depicted in Figure 7. Subjective assessment methods include questionnaires and self-reported scales whilst objective ones are task performance testing such as the Psychomotor Vigilance Test; electrophysiological state monitoring such as EEG, EOG and ECG; physical activity monitoring such as eye-tracking, skin response and body temperature. (J. Li et al., 2020)

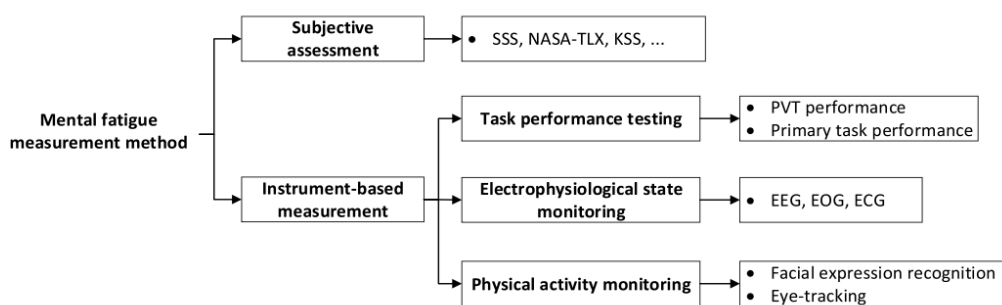


Figure 7: Mental fatigue measurement methods

Figure 8 and 9 reveal an interesting aspect. Whilst the overall most used metric in the included papers was seen to be an objective one: PVT and reaction time evaluation, when focusing only on the healthcare sector the Kalinska Sleepiness Scale – a subjective measure – takes first place as most used.

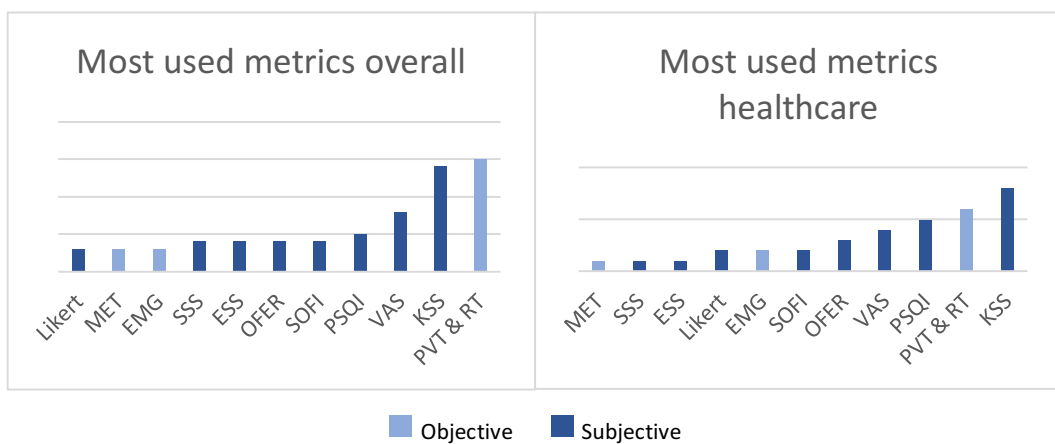


Figure 8: Most used metrics for fatigue evaluation overall

Figure 9: Most used metrics for fatigue evaluation in healthcare



The PVT is used in several different contexts: aviation, transport, nuclear power plants and healthcare. It provides an objective measure of fatigue and sleepiness by focusing on performance in the form of reaction time and alertness. Visual or auditory stimuli are presented to individuals at random intervals and they must react to them. Different versions of the test are created for specific situations. (Gore B.F., Casner S.M., Wickens C.D., Pallesen S., 2018). The KSS on the other hand evaluates subjective sleepiness of the individual through a p-point scale which goes from 1 being “very alert” to 9 being “very sleepy, fighting sleep” (Ganesan et al., 2019). Being it a subjective scale people undertaking it might tend to underestimate their level of sleepiness and results may be biased.

### 3.2.3 Human performance

84 out of the 134 included articles referred specifically to performance. 34 of these were specific to the medical context. Even here we can see in Figure 10 how the theme of human performance in the healthcare sector is becoming more and more treated.

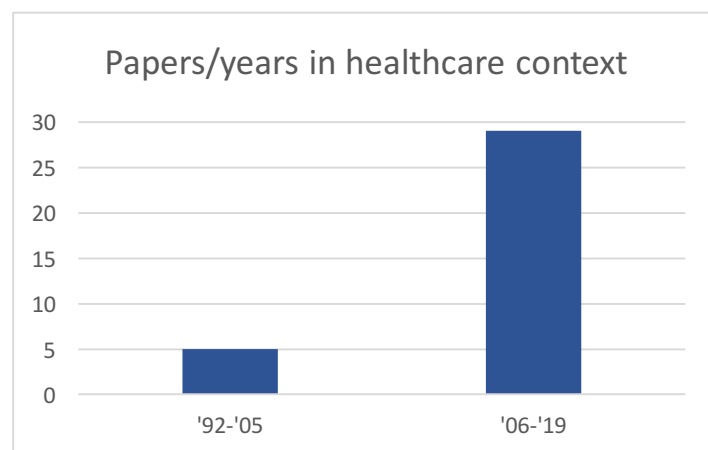


Figure 10: Papers published about performance in healthcare context

Between 1992 and 2005 only 5 of the included documents spoke explicitly about human performance whilst between 2006 and 2019 an outstanding number of 29 papers treated the argument with regards to the healthcare sector.

Even though the number of papers speaking of the argument was satisfyingly high, it was found that papers actually defining explicitly human performance was extremely low. Only 7 articles gave a definition, the most complete of which takes the form of a function of technical and non-technical skills, randomness and cultural and organizational context – in which stress and fatigue are inserted (Casali et al., 2019). Though this function does bring together most aspects that influence human performance it does not specifically show in-depth how they impact on it.

Performance, specifically in the medical context, can be seen as reduced mortality, reduction of error rate and the standardization of effective processes and techniques that bring to successful results and positive outcomes. Figure 11 and 12 compare the most used terms in papers regarding performance both overall and in the healthcare context.

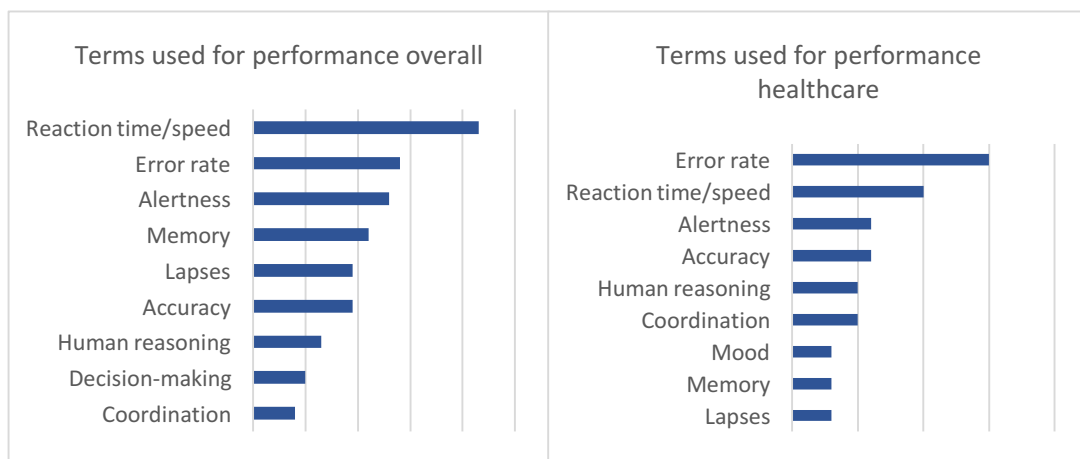


Figure 11: Terms used for performance overall

Figure 12: Terms used for performance in healthcare

Whilst speed and reaction time are overall the most used followed by error rate, alertness, memory and lapses, in the medical sector error rate takes first place. Reaction time and speed, alertness and accuracy follow. It is known that there is a percentage of variability and randomness in every operation – being it even the simplest. This combined with human error causes patient outcomes which differ from that of a perfectly successful operation. If the error rate can be decreased to a rate close to zero, the probability of having satisfying outcomes increases exponentially. Measuring human performance with specifically designed metrics is of fundamental importance to be able to evaluate the effect that fatigue has on it. As it can be seen below in Figures 13 and 14, PVT and reaction time appear again here as most used measure.

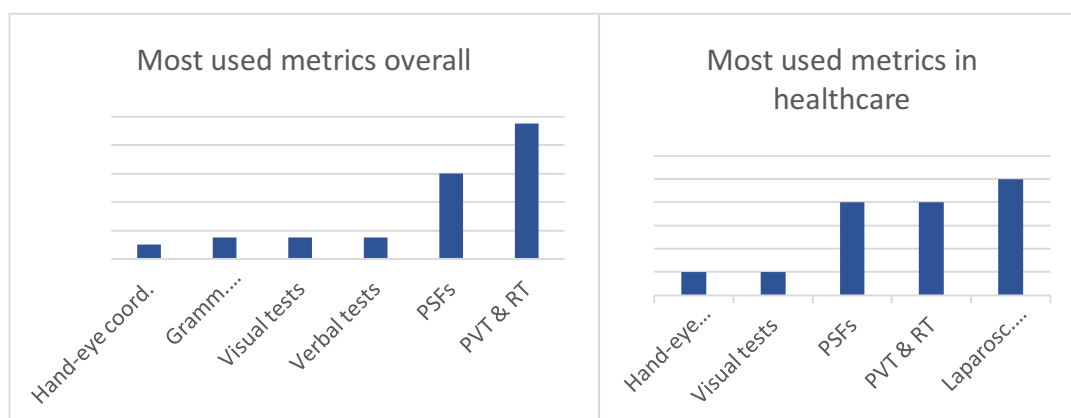


Figure 13: Most used metrics for performance overall

Figure 14: Most used metrics for performance in healthcare

PVT and reaction time can be intended both as measures of fatigue and performance as they aim to measure changes in performance under the effect of different influencing factors – such as fatigue. PSFs take second place overall as HRA is becoming of common use in many sectors, followed by verbal and visual tests, grammatical reasoning and hand-eye coordination. In Figure 14 we can see the most used metrics in the healthcare sector. In first place there is a totally new metric, unused in any other context: laparoscopic tasks. Laparoscopic simulation has been used extensively to assess surgical skills and performance and may be of different

types: ectopic pregnancy procedural model (Amirian, 2014), Fundamentals of Laparoscopic Surgery peg transfer task (Wilson et al., 2011), tissue attachment under traction, placement of a silicon wire or laparoscopic square knot (Huizinga et al., 2018), cup drop, rope passing or pegboard exchange (Mohtashami, Thiele, Karreman, & Thiel, 2014). The measured outcome also changes from study to study: time in simulation, dexterity – computed as instrument path length and instrument angular path – and error – estimated as blood loss (Wilson et al., 2011), time to completion, instrument movement and application of force through the Force-Motion-Surgical-Trainer (Huizinga et al., 2018), depth perception, bimanual dexterity and efficiency (Mohtashami et al., 2014).

### 3.2.4 Night vs day

Fatigue does not impact on performance with the same magnitude during different parts of the day. As a matter of fact, many factors such as the circadian and homeostatic rhythm, backward rotating shift work and poor adaptation to night shifts, enhance fatigue's detrimental effects on performance during night hours. From surgery to anesthetic it has been thoroughly proven that at nighttime risks to incur in accidents caused by fatigue or sleep-deprived individuals increase exponentially.

Whilst anesthesia in Australia is deemed to be one of the safest in the world, this trend of poor human performance in night shifts may take this primacy away from it. The risk of needlestick injuries and accidental dural puncture during epidural insertion were found to be higher at night. Not to mention the peak rates of mortality and intensive care unit readmission of patients discharged between 0300 and 0400. It is for these reasons that an eight-week audit of all cases was put in to place at Perth Royal Hospital (Australia) in order to evaluate which procedures done during night hours could not have been put off to the morning as the hospital has a precise policy that states that only Category 1 and Category 2 procedures (based on severity) which cannot wait until the morning should be done between 2230 and 0800. It was found

that an outstanding 28% of cases done during this hour period were in breach of this policy putting at an unnecessary risk the lives of these patients (O'Loughlin, Smithies, & Corcoran, 2010).

As mentioned before the two-process model does affect an individual's performance level differently during day and night. It has also been found that acute and chronic sleep loss have distinct homeostatic mechanisms. The study in question found that there are at least two well defined and separate components influenced differently by short-term sleep loss (acute) and long-term sleep loss (chronic). What this means in practical terms is that even though an individual has fully recovered from a night of acute sleep loss, he can develop a chronic sleep loss debt. This can be seen in people working long hours for sustained periods which believe that by sleeping in the weekends they have fully retrieved their lost sleep. Overall individuals impaired by chronic sleep loss were found to have a deteriorated performance during the biological night with respect to those induced by acute sleep loss alone (Cohen et al., 2010).

### 3.2.5 Prefrontal Cortex

The Treccani dictionary defines the Prefrontal Cortex as the portion of the front lobe that corresponds to the upper frontal and middle convolutions and orbital convolutions. It is an area which receives afferents from almost all cortical areas, from the thalamus and the numerous subcortical centers. It's subdivided in three structures: dorsolateral, mesial and orbital. The dorsolateral portion is appointed to



Figure 15: Prefrontal Cortex

the organization and planning of several complex behaviors and superior level cognitions such as voluntary actions which are logically ordered, motor programming, verbal fluidity and learning and use of concepts and strategies. The mesial portion plays a role in the motivation both cognitive and emotional. Lastly, the orbital portion

regulates and inhibits the elaboration of the stimuli which interfere with the task in progress and controls instinctual thrust.

The PFC is not responsible for easy, straightforward and mechanical behaviors or instinctive responses such as reacting to an unexpected noise or movement. Its role becomes significant in “top-down” processing which exists when behavior necessitates the direction of inner states or intents. PFC though is critical in circumstances where the aligning between sensory inputs, understanding and actions are speedily changing (Miller & Cohen, 2001).

Harrison & Horne, 2000 in their study summarized papers which focused on tasks that required higher than average levels of technical conscious mental activities. Their study and arguments were brought forward all with a common thread line: the cortical region which has the most central role and functions the most (meaning it has the highest metabolic rate of all cortical regions) during wakefulness is the PFC. Moreover, viewing slow sleep wave typology of sleep as an almost certain exclusive manner of recovery of the cerebral cortex together with a more intense EEG activity in the PFC when blood flow in this region is low, strongly indicates that the PFC may suffer as a consequence of SD. The two authors thus propose that a detrimental performance due to SD is mainly consequence of a reduced activity of the PFC. In other words, they state that a sleep deprived subject’s performance would be impacted negatively while carrying out any task that would require the activation of the PFC. These tasks are of different nature and typology and the areas of main concerns which are impacted are:

- being able to evaluate a situation of complex nature while maintaining the lucidity to avoid distractions;
- capability to develop and evolve strategies along time and different scenarios;
- mood and empathy;
- flexibility to incur in lateral thinking and ability to develop innovative solutions;
- having an ex-ante attitude by anticipating consequences and possible risks;

- having high language and communication skills such as word fluency;
- temporal memory;
- showing insight and understanding into own performance;
- maintaining an interested attitude towards outcome.

These tasks coupled with novelty of situations and unpredictability of developments increase the possible impact of SD on performance.

Killgore, Balkin, & Wesensten, 2006 hypothesized that 49 hours of sleep deprivation would impact negatively on the quality of decision-making and would bring about an increased risk-taking behavior in the Iowa Gambling Task. Thirty-four healthy participants completed the IGT both at rest (baseline) and after the two nights of sleep loss. The IGT consists of a card game on a computer screen with four apparently identical decks of cards. The objective of the game is to win as much money as possible by selecting one card at a time. At every selection, participants win or lose a predetermined amount of money showed on the screen. Two out of the four decks are considered “disadvantageous” or “bad decks” (A’ and B’) as the amount won is generally high but also is the loss. The other two are deemed “advantageous” or “good decks” (C’ and D’) because playing them consistently leads to a net gain through smaller wins and lower penalties. Overall the IGT monitors the ability of participants to modify and adapt decision making strategies according to implicit learning of penalties and reward occurrences.

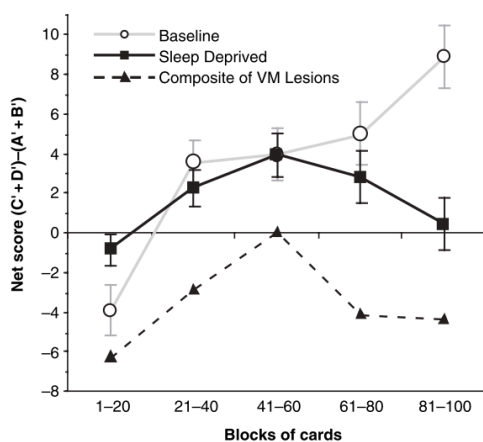


Figure 16: Gambling Task results (Killgore, Balkin, & Wesensten, 2006)

In Figure 16 on the left the results from the study in question are presented. At well-rested baseline (O) participants showed a gradual and constant linear shift from disadvantageous decks towards advantageous decks. In contrast, after the two nights of sleep loss the same subjects had a similar trend in the first half of the

game whilst in the second half there was a complete shift of strategy moving back to the disadvantageous decks. This shift in behavior shows an evident similarity with the pattern of IGT performances for patients with lesions to the ventromedial prefrontal cortex (dashed line). Overall when compared to the rested baseline performance, sleep deprived individuals in this study showed to be less able to understand and weigh appropriately the immediate benefits of lower short-term rewards against longer-term higher penalties. This cognitive ability appears to depend strongly on the integrity of the PFC which is consistent with the above-mentioned findings of Harrison & Horne, 2000. The similarity in strategies between sleep deprived individuals and patients with ventromedial lesions of the PFC are due to the fact that sleep loss is associated with a decreased blood flow to the ventral regions of the prefrontal cortex. This suggests a very strong finding: sleep loss may unfavorably impact on the correct operating of these same regions of the PFC. The VM PFC is an important convergence zone where external stimuli are connected to internal emotional body states – these links are called somatic markers. These somatic markers potentially can affect or bias decision making according to how “good” or “bad” a course of action or decision makes the individual feel. There is a possibility that prolonged wakefulness has a particularly important effect on those executive functions that depend strongly on emotions together with other cognitive procedures.

### 3.2.6 Proactive vs reactive

Independently from fatigue typology and how fatigue impacts on performance there are two methods that can be implemented for managing risks associated to fatigue and fatigue itself: reactive or proactive approach.

A reactive approach focuses on reacting and responding to events after they occur. It is an ex-post technique. A proactive approach is based on foreseeing issues and possibly eliminating problems before they occur, implementing an ex-ante solution.



The difference between the two approaches is the perspective each one provides in assessing actions and events. While the reactive approach has the objective to mitigate and reduce the effects of fatigue, solving the problems that may arise, the proactive approach predicts and prevents the occurrence of the problems.

If the adoption of proactive approaches is common in sectors like nuclear installations, aerospace and military operations that use HRA for safety management of safety-critical systems, in healthcare it is not yet common. The healthcare context is still characterized by a reactive approach culture and the necessity of a transition to a much more proactive approach is underlined by the fact that 50% of human errors in healthcare systems are deemed to be potentially preventable. Through the prediction of human error it is possible to implement error risk mitigation decisions leading to a significant decrease in patients' death and the utilization of healthcare system resources (Vaughn-Cooke, Nembhard, & Ulbrecht, 2010). Even if healthcare, more than other sectors, relies on people rather than automation to deliver its services, safety management is a relatively recent field. Furthermore, the differences in culture and drivers make the transaction of HRA techniques from general sectors to healthcare very difficult. The singular characteristics of the sector is the frequent interaction across organizational boundaries and the demand of a significant tolerance for increased levels of uncertainty. The transition must be done with caution and has to be adapted appropriately.

### 3.2.7 Naturalistic Decision Making

Naturalistic Decision Making has been at the center of recent studies due to its more realistic nature. Although other models for decision making do exist – such as the Rationalistic and Normative Descriptive ones – they have been studied in laboratories with nonprofessional participants, providing in this way distorted and misleading results (Harrison & Horne, 2000b).

Within the group of Naturalistic Decision Making models the Recognition-Primed Decision Model of Rapid Decision Making stood out (Klein, 1993). This model avoids some limitations of analytical strategies and shows how experienced people make decisions in operational settings without evaluating and comparing a range of options. Klein, 1993 observed and obtained information by urban fire ground commanders in emergency situations. The main points which emerged were:

- fire ground commanders felt their decision-making process did not fit into a decision-tree structure;
- they were not consciously considering and evaluating different alternatives and possibilities;
- they were not assessing probabilities of occurrence;
- their decisions were not based on optimal choices as it would have taken too long, choices had to be prompt, efficient, effective and adaptable to changing situations;
- they reacted to situations basing their actions on circumstances they had previously dealt with;
- plans of actions were defined promptly, put into action, overseen and modified based on the development of the situation.

It is incorrect to state that the fire ground commanders were not aware of different paths of actions. Instead in emergency cases where time is the major constraint, their priority was not to evaluate advantages and disadvantages but to acknowledge and categorize a specific situation and behave consequently. In contexts where rapid decision-making is necessary, recognitional strategy appears to be the most efficient. Based on previous experience a practicable and feasible option is chosen.

Klein, 1993 represent the RPD model with three cases (Figure 17):

- A. Simple match → it's the simplest case where the set of circumstances is identified, and the consequent reaction are put into effect;

- B. Developing a course of action → slightly more complex case where the decision maker carries out conscious considerations of the reaction through a mental simulation to uncover possible issues before they occur;
- C. Complex RPD strategy → the most complex and intricate case where the conscious considerations reveal problematics which require modifications, or the chosen reaction is deemed to be inadequate and rejected in favor of the next viable one.

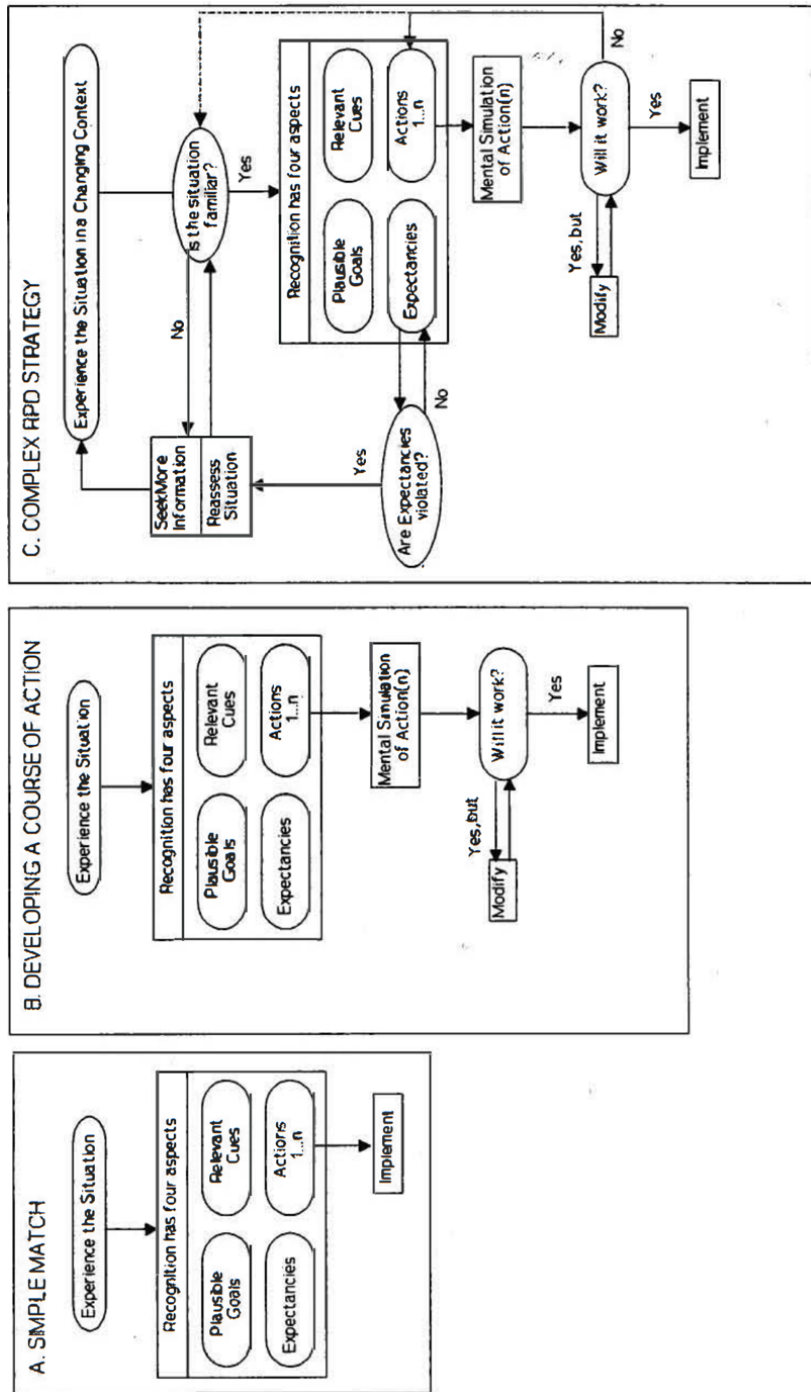


Figure 17: Recognition-Primed Decision Model (Klein, 1993)

The recognition of a situation is divided into four main aspects according to Klein, 1993 a) plausible goals, that refer to the comprehension of the typologies of objectives which can be rationally achieved b) relevant cues, to determine which signals are fundamental in the specific context c) expectancies, developing expectations that can aid as an evaluation of the correctness and precision of the situation assessment d) actions 1...n, recognizing the typical actions to take.

Harrison & Horne, 2000 stated that emergency personnel are those to implement more often the RPD model as they must establish the situation rapidly and take consequent effective actions. Moreover, they observed that several sections of the RPD model rely heavily and require activation of the PFC implicitly implying the risk of being heavily impacted by SD. The initial phases i.e. the soaking in of information and assessment of the critical situation, are deemed to remain resilient even when the individual is subject to SD. The four stages of recognition of the situation on the other hand could be impacted:

- Plausible goals → during SD lateral thinking and flexibility are strongly impacted which may result in poor planning and a negatively affected ex ante approach;
- Relevant cues → there is a risk for relevant signals to be overlooked as the ability to filter relevant from distracting information is impacted by SD;
- Expectancies → the evaluation of the scope and expectations of the situation may be quite difficult in complex, rapidly evolving and dynamic circumstances;
- Actions → SD impacts on the ability to develop and evolve strategies. Most of all the ability to change an already undertaken course of action – even though it is not appropriate anymore – is affected.

Overall it is clear how the majority of skills required to take out a decision using the RPD model require activation of the PFC and therefore are particularly vulnerable to SD.

### 3.2.8 Complex vs simple tasks

Not only is human performance heterogeneous when moving from day to night but also when evaluated through complex or simple tasks. Simple tasks can be the PVT, the Job Performance Scale or sequential key pressing tests whilst with complex tasks laparoscopic simulations, hand-eye coordination or simulation of medical crisis are intended. What has been found in different studies is that performance is more impaired, and effects of fatigue are more evident when sustaining tasks which are less cognitively demanding.

Pilcher, Band, Odle-Dusseau, & Muth, 2007 proposed a study to evaluate the effect of acute sleep deprivation on sustained operations. As this effect had been examined in several previous studies with results that weren't able to fully explain it, a range of tasks which required different information processing were chosen. An actiwatch was implemented to monitor wrist accelerations, objectively measuring sleep and wakefulness. The range of tasks included both cognitive and vigilance tasks. The most complex, stress-inducing and demanding of the cognitive tasks was the Wombat test. The Wombat is a computerized task which was specifically created to evaluate circumstances awareness, endurance to stress, and awareness and attentiveness management capabilities. Training of about 2 hours is required before-hand, time in which participants become familiar with the task. The other two cognitive tasks were shorter but still required training and were the Cardinal Direction Task and the Automated Performance Test System. The latter measures cognitive and motor functions and includes grammatical reasoning and math processing whilst the former simulates a real-life work environment where participants are controlling unmanned aerial vehicles. The three different vigilance tests on the other hand were a simulated chemical plant task, the PVT and a computerized task to evaluate long-term vigilance. Training for these tasks was much shorter. Results from this studied showed how acute sleep deprivation impacted only slightly on the cognitively challenging tasks with respect to the vigilance tasks where performance was significantly impaired. Moreover, the performance on the Wombat even improved during the night of sleep

deprivation which suggests participants continued learning and were more engaged given the game-like features of this test. The decrease in performance in the vigilance tests supports the previously stated hypothesis of Harrison & Horne, 2000 as they required activation of the PFC. The arousal theory is also generally supported as those tasks in which arousal could be deemed low (vigilance tasks) have the most impaired performance as opposed to those which require sustained attention, activation and a high arousal level (cognitive tasks) that result in being less affected.

### 3.2.9 Wearable devices

A very relevant and recent trend involves wearable devices and technologies. Although having been around for some years now, new functionalities and uses are being discovered day by day. Wearable devices are no more Bluetooth headphones, now there are smartwatches that can monitor your position, heartbeat, can receive and send messages and this is just the tip of a whole iceberg of functionalities. Wearable technologies are starting to be applied to different contexts. An outstanding potentiality has been seen in fatigue monitoring and HRA.

The 31 included documents regarding wearable technologies all focused on fatigue monitoring, tackling the theme from different points of view. 24 wearables measure mental fatigue, 16 of which mention specifically the term drowsiness and 3 eye fatigue. Physical fatigue was monitored in 9 papers. After having defined the measured aspect – in this case mental fatigue – the study of these technologies was divided based on how the measurement was taken and the position of the device on the body (Figures 18 and 19).

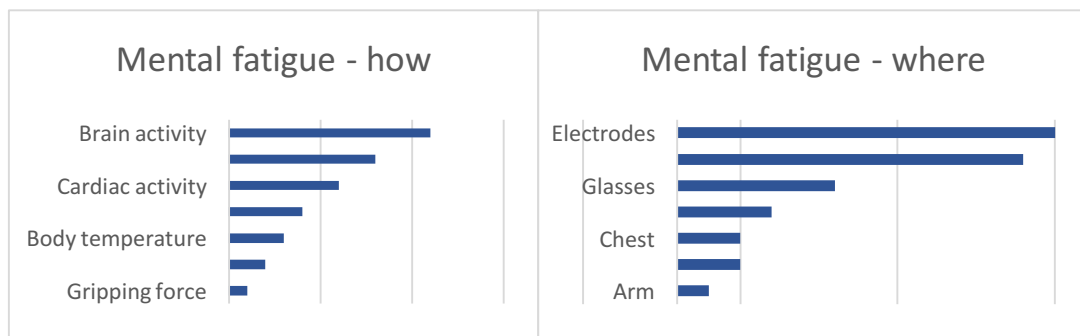


Figure 18: How mental fatigue is measured through wearables

Figure 19: Location of the wearable device on the body

Figure 18 shows how mental fatigue is measured. Eye movement is the second most used metric to evaluate mental fatigue. Although it is an extremely useful and precise method it requires the individual being monitored to wear glasses which may interfere with the delicate tasks which the role of a surgeons requires to be done. The most used measure is brain activity through Electroencephalography. Brain activity is measured through electrodes generally placed on the scalp. Construction workers were found to have inadequate rest and high alcohol intake this impacting on their job performance. This is why many pre-service fatigue screening approaches are put into place, something which is definitely missing in the healthcare sector. A novel and up-to-date screening method which measures mental fatigue was proposed and validated by H. Li et al., 2019 with a wearable EEG headpiece and four assessing indicators.

Moving on to more intricate wearables we have those which put together different measurement techniques like the single EEG electrode brain implemented by Kartsch, Benatti, Schiavone, Rossi, & Benini, 2018. It is capable to detect 5 different levels of drowsiness with an average accuracy of 92,5%. This work is innovative as it aimed and succeeded in putting together different measurement techniques: behavioral and physiological. Whilst as mentioned above the physiological information comes from the EEG signals, behavioral data are extrapolated from an Inertial Measurement Unit. These two technologies are put together in a minimally invasive wearable device which is applicable to the scalp and does not interfere with any operation.



### 3.2.10 Standardization

Given its high-risk characteristics and high variability features, the medical and healthcare sector has been inward-looking and inflexible on many aspects. It is wrong though to assume that this field is totally apart and isolated from other contexts. Manufacturing for example provides foods of thought and operating standards that could be adapted to healthcare. Lessons can be learnt from manufacturing on how to diminish costs and minimize variations through a strong focus on standardization of processes.

Modic, Lasalvia, & Merges, 2019 distinguished between two types of variations in healthcare:

- “Bad” variations are those that are undesirable. They are inconsistencies in care delivery and result in increased costs, a failure in outcome improvement and higher human error probabilities with difficult end result estimation. These variations are referred to as “bad” when patients with the same illness receive an inconsistent treatment based on the location in which they are treated or from whom they receive the treatment;
- “Good” variations on the other hand are those customized and individualized treatments reserved to patients based on their specific conditions.

The real challenge is to distinguish between the above-mentioned two variations and implement standardization and individualization techniques to reduce “bad” inconsistencies whilst monitoring and maintaining the “good” ones to continue providing a high-quality service.

Modic et al. again went on in highlighting the importance of a continuous improvement strategy where each worker has a clearly defined and understood role and processes have been refined and optimized. Standardization in fact can only be effective if a continuously bettering of standards, techniques and processes is put into place and if evidence-based guidelines are personalized and adapted by leaders and experts. The main goal must be implementing norms and procedures in order to make

sure that “never events” – those events which should never happen – are not treated just as inevitable but are ensured to never happen.

A very recent paper by Casali et al., 2019 focuses on the importance of non-technical skills. These skills include circumstance awareness, decision-making up to all the Performance-Shaping Factors considered in HRA techniques. Whilst standardization through development of technical skills shows tangible improvements, the same cannot be stated when referring to NTS. In fact, NTS are characterized by a very high variability, which means that the bettering of these skills would have significant positive multiplicative effects on outcome. Technological advances have made it possible for certain surgical procedures to gain a satisfactory level of standardization (ex. quality of closure of bronchi, veins and arteries in lobectomy procedures) but this homogenization is still not as widely spread as it could and should be.

Again, the medical field is presented with a second main challenge: standardizing NTS. Evidence-based practices improvement is very limited in those skills which don't refer specifically to technically-acquired knowledge. A higher level of standardization means less variability and less adverse outcomes which are essential in the healthcare context.

After having recognized two real challenges in the medical field it is important to highlight that there are ways to reduce their impact. Sujan et al., 2018 based their paper on the poor levels of accuracy in probability estimations, the uniqueness of patients and absence of strict regulative schemes typical of the healthcare industry. The Safer Clinical Systems consisted in a five-year patient safety improvement program structured in five steps. The first two steps which were the definition of pathway, goals, risks and threats were sustained by tools such as the Safety Culture Index, process mapping, Hierarchical Task Analysis and SHERPA. The last three steps included risks being better understood, the development of options to reduce them and planning and putting into action of interventions. These were supported by rapid

improvement cycles and statistical approaches together with the re-application of the methods used in steps 1 and 2. The SCS brought about three main findings:

- Clinical processes, procedures and tasks often are marked by poor reliability. Although this was already well known, more evidence was found to support this thesis and the above-mentioned problem of poor consistency comes about again. This poor reliability was noticed to be linked to processes and tasks which had never been consciously defined, designed or specified.
- The methodical and structured application of HRA could enhance the awareness of the previously stated problem and support the healthcare sector in the standardization process.
- Hierarchical Task Analysis was an essential tool in aiding experts and clinical teams in gaining a more in-depth knowledge and comprehension of the tasks they had to perform, clearly defining boundaries and explicating task characteristics.

### 3.3 Literature gap

The extremely detailed literature review brought forward in this thesis, gave rise to many topics which require a deepening of understanding and gaps which feel the need to be filled. One was the main gap which enhanced the drive to create the Fatigue Impact – Task Complexity model.

#### *G1 – Poor coherence and focus on fatigue’s properties*

To the best of the author’s knowledge, extant literature was seen to consider and integrate fatigue in existing methodologies in a very superficial manner. Fatigue has a vast list of properties and due to this, there is not a univocal definition for it. Out of the 84 papers regarding healthcare that treated the topic of fatigue, 25 definitions were found, 4 papers did not mention the term fatigue, 5 papers assumed it as a synonym of sleep deprivation whilst the remaining 50 were not able to provide a definition. This inability or lack of interest to define the term, suggests the poor importance it has been given.

Below fatigue has been analyzed and its main properties have been identified. This was done to highlight through consequent argumentations the large gaps in the present literature.

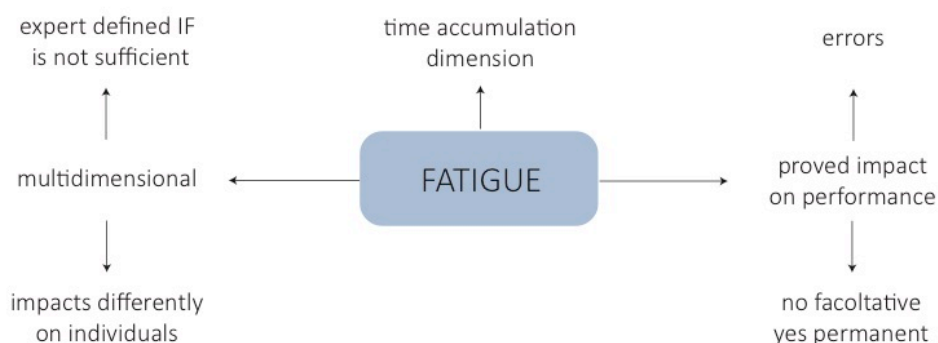


Figure 20: Fatigue properties

Starting from the left side of the graph, fatigue's first property emerges: multidimensionality. As deeply explained in chapter 2.2.4 fatigue has many different ways of manifesting and can be caused by numerous factors. This means not only that it impacts differently on individuals but that an expert defined IF does not sufficiently reflect its importance. Up to now in fact, fatigue has been seen as an optional Performance Shaping Factor to consider or not based on experts' perception. There is a very strong incoherence in acknowledging that fatigue has a subjective impact and defining it through a constant multiplier.

Moreover, many papers and studies have demonstrated an impaired performance in fatigued individuals. Alertness, good decision making, high level communication and reactivity and promptness to respond to critical situations are no optional when operating in the medical field. Impaired performance results in an increase of errors. Reminding ourselves the context we are analyzing, errors in many cases could mean irreversible damages sometimes causing death. Fatigue's proved impact on performance, calls for a permanent integration of it in HRA (or fatigue evaluating) techniques. Linking back also to the use of IFs as a measure of impact of fatigue, optionality must not be accepted anymore. Fatigue is an evanescent threat to monitor and take into account constantly.

Lastly, fatigue's time accumulation dimension adds to its complexity. This longitudinal time characteristic must be considered as fatigue alters as time passes and a prolonged sleep deprivation has cumulative effects. Recovery of fatigue does not correspond to sleeping in to regain the hours lost the night before. This is because several nights of sleep deprivation – better seen as prolonged wakefulness – have cumulative impacts on performance and are to be recouped with several nights of total sleep.

Overall, fatigue has not yet been considered and integrated in a satisfying manner. When dealing with the topic of fatigue the tendency is to consider only partially its

properties. Fatigue's very severe consequences have been evaluated but not taken enough into consideration when applying HRA and error predicting techniques.

### 3.4 Aim of the work

The Fatigue Impact – Task Complexity model presented in this thesis aims to fulfill the main gap emerged from the literature review whilst integrating HRA techniques and Functional Job Analysis. Healthcare's reluctance to learn from other sectors can be seen in the poor use of HRA. Figure 3 in chapter 3.2.1 highlights the lack of adoption of HRA methods in this industry as only in very recent years it has become a topic worth treating. Task description techniques are an example of very effective HRA technique extremely useful for error analysis. They have been successfully implemented in many areas of surgical training, but an even more extended application is required to counteract also the poor standardization common in healthcare. In fact, poor standardization is another issue related to healthcare's stand-alone nature. Standardizing operating procedures benefits to performance management, the reduction of errors and higher reliability and replicability. HRA and Functional Job Analysis are the methods able to bridge the gap between healthcare and all the other industries it can learn from, whilst highlighting the critical issues of the medical sector.

Returning to G1 regarding fatigue, the previously identified characteristics of fatigue were supplemented with the consequent needs that emerged.

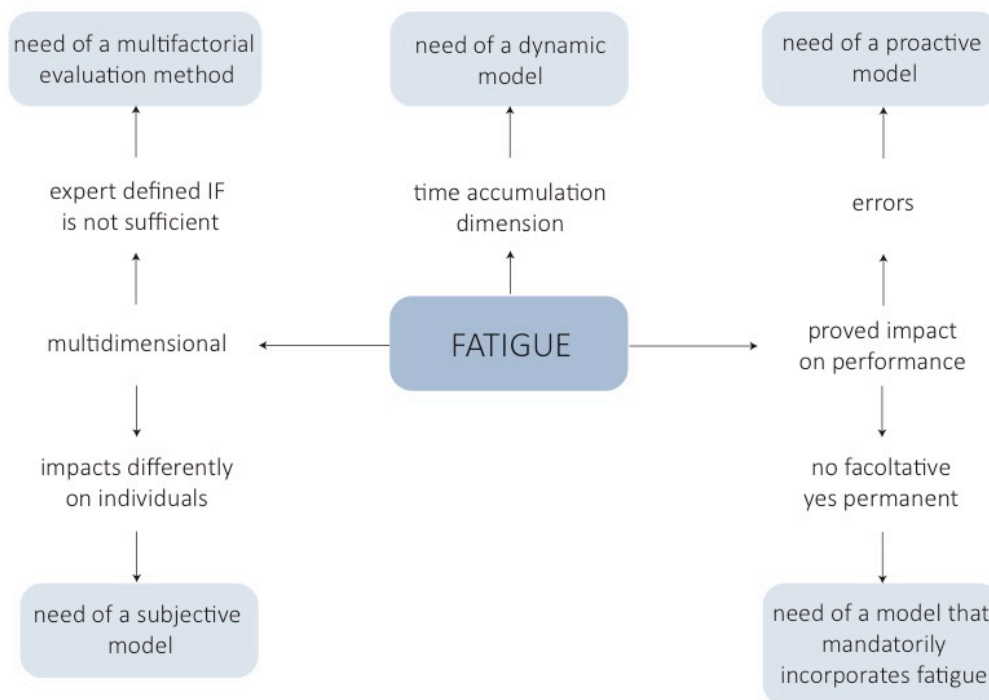


Figure 21: Fatigue properties and emerging needs

The deficiencies found are reported in the following points together with the solutions adopted:

- Need of a proactive model: fatigue’s proved negative impact on performance and error making, highlights the necessity of a shift from reactivity to proactivity within the healthcare context. Industries which have advanced safety cultures are described as informed at all levels, highly adaptable to change, distinguished by a high level of trust by all and worried – intended as far from complacency (Hudson, 2003). The medical culture though, lacks systematic risk management which makes it a completely reactive sector. Measures are taken ex-post and they involve repairing damages – which in most cases are unreparable.
- Need of a model that mandatorily incorporates fatigue: stressing the fact that fatigue’s impact on performance is certain for all individuals, its inclusion in in HRA methods must be mandatory. The possibility to arbitrarily select or not its

inclusion could be very dangerous due to misleading Human Error Probabilities results.

- Need of a multifactorial evaluation method: fatigue's multidimensionality calls for the need of a more precise way of evaluating fatigue.
- Need of a subjective model: given its different ranges of impact on individuals, the model must be subjective.
- Need of a dynamic model: due to its time accumulation dimension the model must comprise a dynamic feature to take into account this evolving nature.

The solutions to these deficiencies take shape in different aspects of the Fatigue Impact – Task complexity model. The FI-TC model presents itself as a tool to support ex-ante detection of critical tasks within healthcare coupled with a real-time fatigue level evaluation framework. The point was to enable the integration of a measure of fatigue impact on surgeons' technical performance, with a measure of complexity of the task. The main gap which stemmed the need of a more complete integration and consideration of fatigue in Human Reliability Analysis was merged with many other shortcomings: a) the lack of a measure of perception/influence of fatigue coupled with the increasing knowledge about fatigue's effect on the Prefrontal Cortex gave rise to the y-axis of the model: *fatigue impact* b) the importance of task complexity as a task characteristic that influences and predicts human performance and behaviors gave rise to the x-axis of the model: *task complexity* c) the poor use of task analysis and standardization in healthcare suggested the use of *Functional Job Analysis* as the technique to divide jobs into task to be evaluated d) the necessity of the implementation of HRA techniques in healthcare to become a customary procedure suggested the addition of an HRA component to the model: *HEP* e) G1 gave rise to the *fatigue level evaluation framework*

Overall the aims of the FI-TC model can be summed up as:

- Integrating fatigue having a 360° view of its properties
- Creating a model which can be easy replicable in different contexts



- Moving towards a higher standardization of tasks
- Shifting from reactivity to proactivity by reducing possible errors and not repairing through the ex-ante and real-time analysis

## 4. Model development

---

Chapter number four is structured to present the model and its inherent features. The model is introduced visually with a brief overview. The chapter unrolls by describing Functional Job Analysis which stands at the base of its configuration and each feature of the model in detail: x-axis, y-axis and HEP. The previously identified aims find shape in the off-line mapping analysis and real time analysis through the fatigue level evaluation framework.

---

## 4.1 Model overview

The model presented in this thesis – the axes of which will be analyzed in-depth in the chapters below – differs from previously seen models for many reasons. The main reason is its deeply adaptable nature to a wide variety of contexts. The aim is the reduction of risks through a replicable framework that pushes the medical field in particular towards standardization of processes and proactive acting.

The Fatigue Impact – Task Complexity model is a support tool in the ex-ante detection of critical tasks that will need a real-time evaluation of fatigue level of operators. In the construction of the model the focus was to relate tasks' fatigue impact to the level of complexity of that task not disregarding Human Error Probability. The model will be described with terms and notions from the healthcare context as it is proposed as a solution for the gaps previously presented. Applicability of the model to a variety of context out of the healthcare one is not excluded. The objective, in fact, is to create a flexible framework with a deeply adaptable nature. In order to be implemented in real-time, the FI-TC model requires a fatigue level evaluation framework – for the real-time analysis – which is later on described. In Figure 20 the FI-TC model is presented on the left showing its integration with the fatigue level evaluation framework.

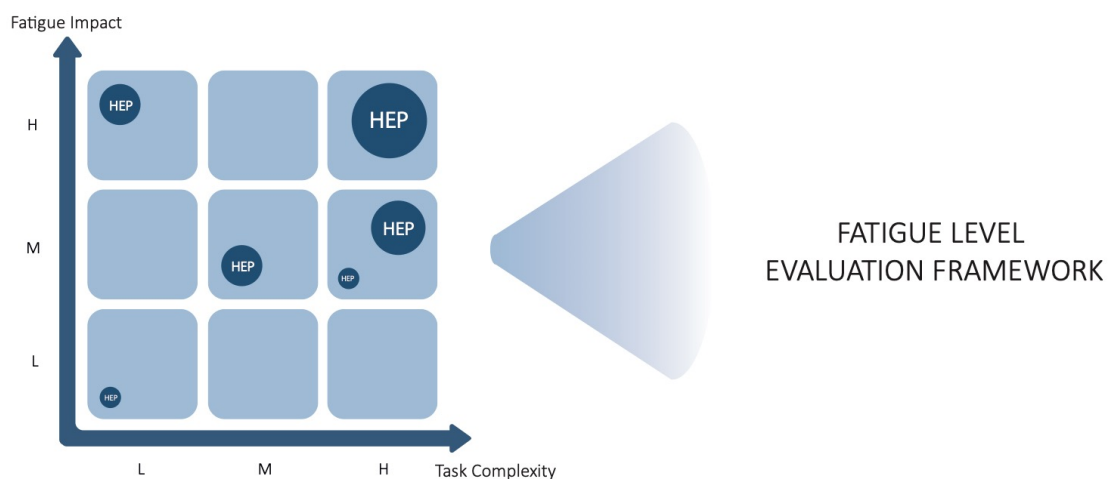


Figure 22: Fatigue Impact - Task Complexity model

The FI-TC model is suitable in moments where uncertainty arises, and numerous variables make the situation difficult to be evaluated. In healthcare the FI-TC model is provided with the aim to help human resource management in those tasks deemed critical and evaluating whether surgeons are in a condition to operate or not, decreasing the risks of error making. The model considers the complexity of the task to be performed, the impact of fatigue on performance, and the likelihood of human error. The latter is obtained through the adoption of HEP analysis in HRA methods. It has been included in the model to widen the evaluation method. Therefore, fatigue and task complexity are not the only variables considered as a cause of impaired performance, but the Influencing Factors used in the HEP calculation are included. As presented in chapter two, the IFs are all the environmental, personal, or task-oriented factors that influence human performance in complex systems. The model has been structured to consider as many variables as possible in order to obtain the most reliable output.

The FI-TC model is structured in two axes and nine areas. The horizontal axis is defined by task complexity whilst the vertical axis represents fatigue impact. Both are tripartite axes which generate nine areas, each of which is characterized by a certain level of task complexity and a level of fatigue impact. The nine areas are then divided in three groups. Each of them has a different level of criticality. Within the nine regions, circle-shaped Human Error Probabilities are included. The bigger the bubble, the higher the probability level of human error.

The aim is to define in advance which of the tasks that constitute a surgical operation, need a real time evaluation of the operator's fatigue level. The model does this process in two phases. The first one takes place offline without including values related to the health status of the operator. In the first phase, each task that constitutes the surgical operation under analysis, will be positioned into the model. The task could be located in a low risk area where no other analysis is needed or in a critical area where it is necessary to look at the HEP value of the task. If the HEP value of the task is higher than the nominal one accepted by the model, real-time

verification is required. The second phase occurs in real-time through the use of the fatigue level evaluation framework. So, the application in real-time of the model is guided by the model itself in an offline mode which previously studies the task and their criticality.

The fatigue level evaluation framework is proposed as the tool through which the model finds its applicability. It is composed of four different methods that aim at measuring the level of fatigue of the individual. These methods are the Epworth Sleepiness Scale, the Modified Fatigue Impact Scale, the Pittsburgh Sleep Quality Index and the Psychomotor Vigilance Test. The first three cited are subjective questionnaires whilst the latter is an objective measure of vigilance level. Different scales were considered even if they aim at measuring the same variable. The intent was to present all the questionnaires that from the analysis of the the literature proved to be the most used and reliable.

## 4.2 Functional Job Analysis

Functional Job Analysis was implemented to define the number and typologies of tasks composing a job. The complexity of the tasks in question was also rated according to the FJA method. FJA is a theory developed by Fine & Wiley, 1971, derived from researches that started sixty years earlier which worked on job descriptions to establish transferability among jobs. Initially, activities were identified through titles. So, different jobs had different job titles. However, Fine discovered that job titles were unstable, unrealistic and ultimately imprecise and non-descriptive. What was stable were tasks, the fundamental unit of job design, job management and job performance (Moore, 1999). FJA is a methodology for describing work, through a process of job analysis by studying tasks. It can be defined as:

- a *conceptual system*, which defines the elements of work activity, e.g., people, environment, data;

- an *observational method*, where observing and recording people working takes place;
- a *method of analysis*, which creates connection between all work output and macro-goals of the organization with competences, know-how and approaches of the employees.

FJA is used in the field of human resource management and organizational strategic planning. It is adopted by European and Asian laws in many industries meanwhile in the United States and Canada it is at the basis of their national occupational classification systems. FJA provides a job description that is accurate, objective, reliable and steady. It responds to the difficulty of delivering a descriptive job information, without incurring in the introduction of subjective elements in the descriptive materials. The task statement is composed by verbs, nouns and adjectives. Through an analysis made by observation without the use of the FJA scale there is the possibility that analysts or researches refer to a) the same thing with different appellatives b) the same elements with different nouns and c) different actions with same verb. The analysis in this way, is unreliable. Functional Job Analysis proposes itself as a tool for defining tasks and measuring their complexity, with higher objectivity. The implementation of FJA sees the creation of task statements and adopts the 10-component scale to evaluate and rate the descriptive material of jobs. The final result is a conceptual framework accessible to anyone for the evaluation and examination of the level of complexity that can be correlated to a specific task (Fine, 2014). Further explanation for each step of the methodology is given in the paragraphs below.

#### 4.2.1 Task analysis

Task statements are the capstones of the descriptive and objective job information method called Functional Job Analysis. Task is the tiniest comprehensive piece of work activity in which a job can be split into. Therefore, if jobs are performed for reaching

a scope, every time that a task is accomplished, part of the objective is achieved. The task is included in the workplace and at its completion it becomes an input to the realization of the final scope. It is defined as an "...action or action sequence grouped through time designed to contribute a specified end result to the accomplishment of an objective for which functional levels and orientation(s) can be reliably assigned" (Wiley & Fine, 1970).

Task statement is not only a list of actions but contains supplementary information used to describe the situation in which the operator performs the task. In addition to actions, it specifies any person involved in the task and with which the worker interacts with, such as co-workers or patients in the case of the healthcare field. Other information concerns which adopted materials, involved tools and used work aids (in the case of healthcare for example patient records, equipment, anatomy references).

FJA recognizes *Things*, *Data* and *People* as the objects of workers behaviours. This information is captured through observation of the phases of the job. The observation of the process should occur several times in different parts of the day, with the focus on the actions performed rather than on who is performing it and when. When the observation takes place, it is better for someone competent to be doing the tasks observed in order to acquire data that is not biased. In taking out this the list of five questions proposed by Fine should be used to acquire all the information that should be included in the task statement. These questions are:

- Question 1. Who is doing the task?

The subject should be analysed according to his or her actions and no reference should be given about his or her identity and occupation. The statement has to be adaptable to any person, independently of their profession, without incurring in rewrite.

- Question 2. What are the actions performed?

This is the list of distinct actions performed by the worker. It is important to represent actions with verbs that are as much specific as possible avoiding misinterpretation of the task statement by the reader. Instead of writing actions such as “operates”, which actions operating requires such as “insert”, “open”, “cut”, etc must be inserted. Verbs representing elementary actions should be used. This is particularly important when task statements are used in the process of assigning jobs.

- Question 3. What are the tools, materials or work aids used?

The task statement should include materials and tools that are used by the worker in performing the task. Moreover, every person with which the worker interacts while doing the task, should be mentioned and included.

- Question 4. What are the instructions to follow?

It is important to write how the worker receives the instruction of execution: if he must follow verbal or written instructions, if there are general routine procedures set by the hospital or if the worker is free to choose how to act. This part includes the level of discretion that should be used by a worker. It can happen that the actions described are so specific that the source of instructions is implied.

- Question 5. Which are the results to accomplish?

One of the most important aspect is the definition of the outcome, the reason behind the implementation of the series of actions. Again, the higher the specificity in the definition of the objective, the clearer the task statement will be to the reader, and the more explicit it will be to the one in charge of performing the task.

The task statement is used as a job description and is supposed to be an effective tool in communicating work assignments, so there is no room for misinterpretation



regarding what the operator should do. Its structure is composed by a first part called action part and by a small final phrase made by few words called outcome part. The action part is made by actions that can be mental, physical, interpersonal or a combination of them. Example of mental actions are “decides”, “chooses” and “compares”, example of physical actions are “grabs”, “inserts” and “touches” while interpersonal actions could be “responds”, “dialogues” and “informs”. The outcome section is located at the end of the statement, composed by only one verb and preceded by “in order to”, or just “to”. Below an example of task statement where the action part and the outcome one are divided by “//” and verbs are highlighted by italics, is given.

*Ex. Receives patient, asks for the medical receipt, inserts data of the patient, in order to // maintain updated the medical record.*

### 4.3 Horizontal axis: task complexity

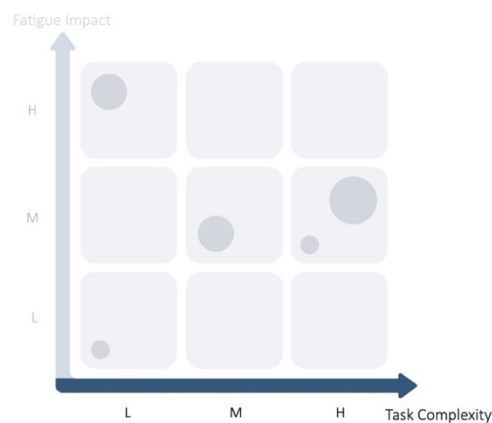


Figure 23: Horizontal axis of the FI-TC model: task complexity

The x-axis of the FI-TC model proposed describes the level of task complexity. The axis is divided in three parts: from left to right there is “low”, “medium” and “high”. These three ranges are defined through the use of a task complexity evaluation scale which will be presented in the paragraphs below. The critical analysis of the literature together with further researches, regarding the importance of standardization and precise definition of tasks within the medical field, suggested the implementation of Functional Job Analysis.

#### 4.3.1 Definition of task complexity

Task complexity is an invariant property of a task characterized by the number of components, inputs, products, and the relationships among all three involved in performing a task (Hysong, Amspoker, & Petersen, 2016). In this thesis, task complexity is operationalized by individual task ratings of 10 dimensions recommended by the Functional Job Analysis (FJA) methodology.

#### 4.3.2 The study of complexity – Functional Job Analysis

Functional Job Analysis was presented in the chapter above as a descriptive tool in job representation. When a task statement is written by an analyst, the perceptions about task complexity and task difficulty are indirectly integrated in the statement, in the moment in which he chooses words for job description instead of others. Fine, in the original version of Functional Job Analysis, integrated in his methodology a 10 dimensions scale used for language control. The original objective of this scale therefore was of aiding analysts in the selection of the correct terms – in regard to the complexity of the task in question – when writing the task statement. Below this use of the scale will be presented, explaining subsequently what differed between the implementation chosen in this thesis and the original FJA one.

The task complexity scale used in FJA analysis is composed of 10 items. The dimensions of the scale are: *Things, Data, People, Worker Instructions, Reasoning, Mathematics, Language, Worker Technology, Worker Interaction* and *Human-Error Consequence*. The first three dimensions are grouped together because *Things, Data* and *People* embed behavioural terms, which are used to describe the relationship between the worker and each one of the three dimensions. The relationship between the worker and *Things* is physical, mental when the worker deals with *Data* and interpersonal when the worker interacts with *People*. These are defined functional scales indicating behavioural levels that can actually be defined as skill levels. The *Worker Instruction* scale indicates the mix of levels of autonomy that the job necessitates together with the level of prescription imposed. The *Reasoning, Math,* and *Language* scales delineate the levels of educational development required to perform the work. *Worker Technology, Worker Interaction* and *Human-Error Consequence* indicate the level of difficulty in carrying out the job (Fine, 2014). The table below presents the scale with the range of scores and the definition of each dimensions.

Scale	Range of scores*			Definition
	Low	Med	High	
Things	1–2	3	4	Physical interaction with and response to tangibles—touched, felt, observed, and related to in space; images visualized spatially
Data	1–2	3–4	5–6	Information, ideas, facts, statistics, specification of output, knowledge of conditions, techniques; mental operations
People	1–2	3–4	5–8	Live interaction among people, and between people and animals
Worker Instructions	1–2	3–4	5–8	Amount of autonomy afforded worker, based on the degree to which inputs, outputs, tools, and procedures required to accomplish task are specified
Reasoning	1–2	3–4	5–6	Knowledge, ability to deal with theory versus practice, abstract versus concrete and many versus few variables
Mathematics	1–2	3	4–5	Knowledge and ability to deal with mathematical problems and operations from counting and simple addition to higher mathematics
Language	1–2	3–4	5–6	Knowledge and ability to speak, read, or write language materials from simple verbal instructions to complex sources or written information and ideas
Worker Technology	1–2	3–4	5–8	Means and methods employed in completing a task or work assignment (tools, machines, equipment or work procedures, processes or any other aids to assist in the handling, processing or evaluation of things or data
Worker Interaction	1–2	3–4	5–8	Degree to which, when working with others (through direct or indirect contact), workers assist each other, coordinate their efforts and adapt their style and behavior to accommodate atypical or unusual circumstances and conditions; this effort leads to achieving employer goals to given standards
Human-Error Consequence	1–2	3–4	5–8	Degree of responsibility imposed upon the performer with respect to possible mental or physical harm to persons (including performer, recipients, respondents, co-workers, or the public), resulting from errors in performance of the task being scaled

*1 is the lowest level of complexity associated with each scale, representing the lowest behavioral benchmark for the scale in question. Higher numbers mean higher degrees of complexity on each given scale. Each scale has a different maximum, because the scales are benchmarked to their natural behavioral limits. For example, complexity with respect to data was benchmarked on six naturally occurring levels: (1) comparing, (2) copying, (3) computing/compiling, (4) analyzing, (5) innovating, and (6) synthesizing. Each scale is benchmarked in a similar manner, yielding to different natural ranges. See Fine and Cronshaw<sup>2</sup> for benchmark levels associated with each FJA scale. Fine and Getkate<sup>19</sup> defined low, medium, and high ranges for the Things, Data, and People scales. Ranges on the remaining scales were grouped into low, medium, and high, accordingly*

Table 5: Functional Job Analysis Rating scales and brief definitions (Fine, 2014)

Each scale dimension is benchmarked according to different levels of complexity. The lowest level of complexity is represented by the numerical value 1, which represents the lowest behavioural benchmark for the dimension taken into consideration. As it is shown, the range of scores is not equal for each dimension. For example, the highest value for *Things* is 4, while the highest for *People* is 8 and for *Language* is 6. The differences in the maximum values are due to the benchmark system referring to the natural behavioural limits of each dimension. An example of the functional scale dimensions of natural behavioural limits is presented below.

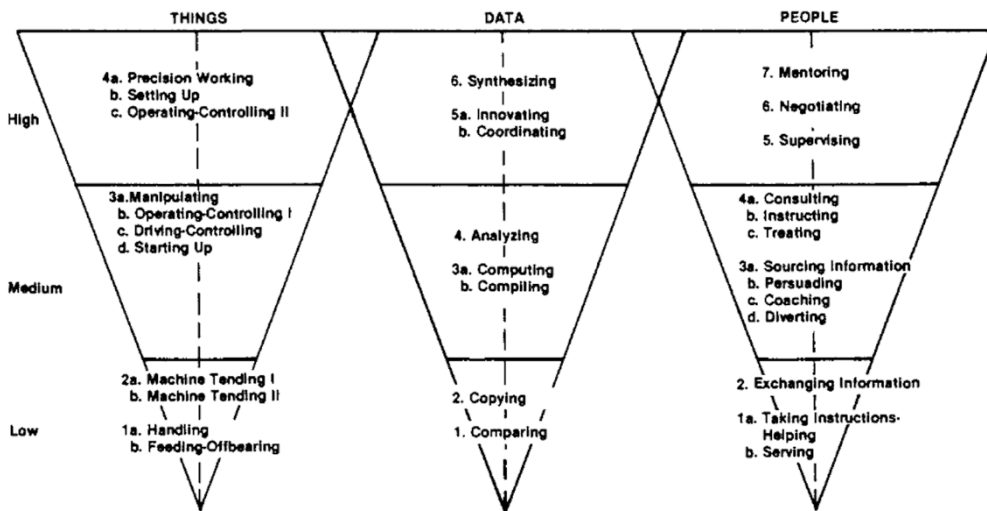


Figure 24: Functional Job Analysis scale dimensions: Things, Data, People (Fine, 2014)

As it is shown, the triangles are divided in three sections named “low”, “medium” and “high” with reference to the degree of complexity. What these triangles symbolize is the choice of terms for each item of the scale according to the complexity of the task. In other words, if the item *People* is deemed to have a high complexity then in the task statement words such as “supervising”, “negotiating” and “mentoring” will be used. Again, if *Data* as well is seen to have a high complexity then “innovating”, “coordinating” and “synthesizing” will be the chosen terms in the drafting of the task statement. Each hierarchy is independent from the others and must not be read across only because they appear to be on the same complexity level. A worker

assigned to high functional level of *Data*, is not necessarily involved in high level of *Things* and *People*. It occurs normally that for lower level of the triangles on the sides, a worker is assigned to higher functions of *Data*. Vice versa though, high levels of complexity of *People* and *Things* involves directly a high or even higher level in the *Data* dimension. For example, when a worker is in the condition of “consulting”, the fourth level of *People*, he is likely to be in the process of “analysing” and possibly “innovating” and “coordinating”. Another intrinsic characteristic of the model regards the inclusiveness of the lower numbered features under the highest one. So, a worker that interacts with *Things* with “precision working”, is also “manipulating”, “machine tending” and “handling” (Fine, 2014).

#### 4.3.3 Questionnaire

In paragraph 4.3.2 the original use of the task complexity scale in FJA was outlined. For the purposes of this thesis, the task complexity scale was extrapolated from its “language evaluation” function and used for the mere aim to assess task complexity. What this means, is that the scale is not used in the task statement writing but only subsequently, when experts are asked to evaluate task complexity through a purposely designed questionnaire.

The questionnaire to assess the level of complexity of a task is presented below. The scale proposed by S. Fine in the Functional Job Analysis was chosen as the structure of the questionnaire. After a critical literature reading, FJA methodology was considered one of the best methods for the complexity level evaluation of the tasks. Even if FJA was born in industrial and organizational fields, numerous applications can be seen in the healthcare context, most of the time with the object of reallocating work between doctors and system redesign.

The questionnaire is structured in ten sentences, each of which are composed by the name of the scale dimension and the explanation of it. Just below of each sentence, the number of blocks is drawn representing the complexity level in accordance with

the benchmarks defined by FJA. For example, the first dimension *Things*, according to FJA theory, has a range of scores from 1 to 4 so, four blocks are drawn in representation of different levels. The surgeons are asked to tick as many blocks as they think is the level of complexity concerning that dimension. The decision to hide the number of each block, was taken in order to avoid the surgeons being confused by different lengths. The formulas and the process of results evaluation are presented in the next chapter.

DATE: \_\_\_\_\_

NAME: \_\_\_\_\_ SURNAME: \_\_\_\_\_

ROLE: \_\_\_\_\_

OPERATION: \_\_\_\_\_

TASK: \_\_\_\_\_

TASK CODE: \_\_\_\_\_

---

TICK THE BOXES ACCORDING TO THE LEVEL OF COMPLEXITY OF THE TASK ACCORDING TO:

THINGS - Physical interaction with and response to tangibles—touched, felt, observed, and related to in space; images visualized spatially  
 LOW     HIGH

DATA - Information, ideas, facts, statistics, specification of output, knowledge of conditions, techniques; mental operations  
 LOW       HIGH

PEOPLE - Live interaction among people, and between people and animals  
 LOW         HIGH

WORKER INSTRUCTIONS - Amount of autonomy afforded worker, based on the degree to which inputs, outputs, tools, and procedures required to accomplish task are specified  
 LOW         HIGH

REASONING - Knowledge, ability to deal with theory versus practice, abstract versus concrete and many versus few variables  
 LOW       HIGH

MATHEMATICS - Knowledge and ability to deal with mathematical problems and operations from counting and simple addition to higher mathematics  
 LOW      HIGH

Figure 25: Questionnaire for the assessment of task complexity (part I)

LANGUAGE - Knowledge and ability to speak, read, or write language materials from simple verbal instructions to complex sources or written information and ideas

LOW       HIGH

WORKER TECHNOLOGY - Means and methods employed in completing a task or work assignment (tools, machines, equipment or work procedures, processes or any other aids to assist in the handling, processing or evaluation of things or data

LOW         HIGH

WORKER INTERACTION - Degree to which, when working with others (through direct or indirect contact), workers assist each other, coordinate their efforts and adapt their style and behavior to accommodate atypical or unusual circumstances and conditions; this effort leads to achieving employer goals to given standards

LOW         HIGH

HUMAN-ERROR CONSEQUENCE - Degree of responsibility imposed upon the performer with respect to possible mental or physical harm to persons (including performer, recipients, respondents, co-workers, or the public), resulting from errors in performance of the task being scaled

LOW         HIGH

Figure 26: Questionnaire for the assessment of task complexity (part II)



#### 4.3.3.1 Final score and ranges

The final score is obtained in two steps. The first step comprises the calculation of the average of the scores collected for each dimension. It is required to sum up all the attendees answer which are in the form of ticked boxes. The total number of boxes ticked by the attendees must be divided with the number of people that have participated. The average is used to identify the level of complexity which can be equal to “Low” (L), “Medium” (M) or “High” (H) according to the ranges defined in Table 5 previously described.

Scale	Range of scores	Scores	Average ( $\mu$ )	Level
Things	1-4	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Data	1-6	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
People	1-8	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Worker Instruction	1-8	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Reasoning	1-6	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Mathematics	1-5	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Language	1-6	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Worker Technology	1-8	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Worker Interaction	1-8	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H
Human-Error Consequences	1-8	$x_1 \dots x_n$	$\frac{\sum_{i=0}^n x_i}{n}$	L or M or H

Table 6: Summary of the steps for the calculation of the final score task complexity evaluation

Table 6 resumes all the relevant data required for the final score evaluation. The final score level is easily identifiable looking at the last column of the table and by selecting the level type that has been mentioned the most.

## 4.4 Vertical axis: fatigue impact

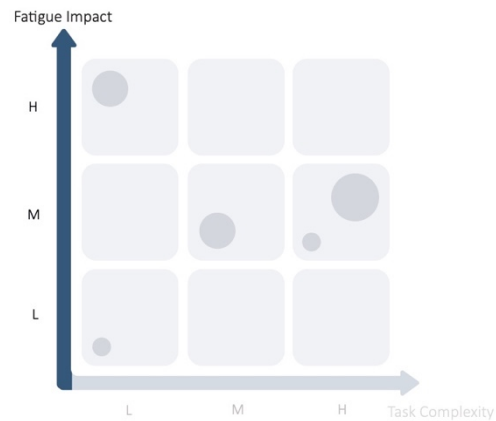


Figure 27: Vertical axis of the FI-TC model: fatigue impact

The y-axis of the FI-TC model proposed is defined by fatigue impact. Fatigue impact is evaluated based on task characteristics and other findings inferred from a critical analysis of the literature. It appears as a tripartite axis: the lower part of the segment representing a “low” of fatigue impact, followed by “medium” and “high” moving upwards further away from zero.

### 4.4.1 Definition of fatigue impact

The impact of fatigue has been referred to in extant literature based on individuals’ characteristics through the application of evaluation scales such as the Fatigue Impact Scale and the Modified Fatigue Impact Scale. These impact assessment techniques are composed of a list of statements which describe the effects of fatigue on the individual.

In this thesis the concept of fatigue impact was initially isolated from the individual subjective matter and transferred to the topic of tasks’ characteristics. Fatigue impact has in fact been defined in the present work as “the influence that a measured level

of fatigue has on the individual. It is based on the skills used in the duties carried out. Such skills are identified through the evaluation of task characteristics”.

#### 4.4.2 Convergent and divergent thinking

The questionnaire which will be presented below is mainly based on the difference between convergent and divergent types of tasks. For this reason, paragraphs on these concepts were deemed of great importance. Many authors have addressed the argument, not only focusing on the description of each but also on the joint role they have together in the creation of novelty and creativity.

##### 4.4.2.1 Convergent thinking

Convergent thinking is based on the concept that one solution is correct and logical criteria is used to reach this solution. Generally the problem to be solved through convergent thinking is well-defined. Cropley, 2006 states that convergent thinking leverages techniques and knowledge already acquired and is mainly about recognizing familiar circumstances. There are two main aspects to convergent thinking: firstly, manipulation and adaptation of preexisting information must take place followed by an increase of this knowledge as a result of the application of it. This type of thinking generally gives rise to *orthodoxy*.

##### 4.4.2.2 Divergent thinking

Divergent thinking, contrarily, allows many correct solutions to exist. Solutions are not mutually exclusive and great flexibility of thoughts is required. As opposed to convergent thinking, the problem to be solved is rather vague with unclear boundaries and a hazy selection criteria. Cropley, 2006 affirms that even divergent thinking involves previous knowledge but is based on alternative ways to combine readily available information. It requires identifying the know-how and expertise to

leverage, understanding possible connections among apparently unlinked information and transforming it into a new solution. Different people solving the same problem through divergent thinking will almost certainly follow two diverse paths, both valuable. This type of thinking generally gives rise to *variability*.

#### *4.4.2.3 Joint role of convergent and divergent thinking*

Cropley, 2006 focused not only in defining extensively convergent and divergent thinking but gave great importance to the joint role of the two. Creativity creation is based on two main phases: *generating* novelty and *exploring* it once it has been generated. The former involves the actual production of novelty but with the integration of the latter it may lead to quasi-creativity in most cases. The transformation of basic novelty into creativity – functional, productive and successful novelty – requires both processes to be tightly interconnected. Generation of novelty requires divergent thinking whilst exploration is made possible through convergent thinking. It is not enough to merely “have the idea”, this must be followed by a feasibility, workability and scalability analysis.

When novelty is introduced there are many risks in which individuals could incur, even with a correct implementation of exploration.

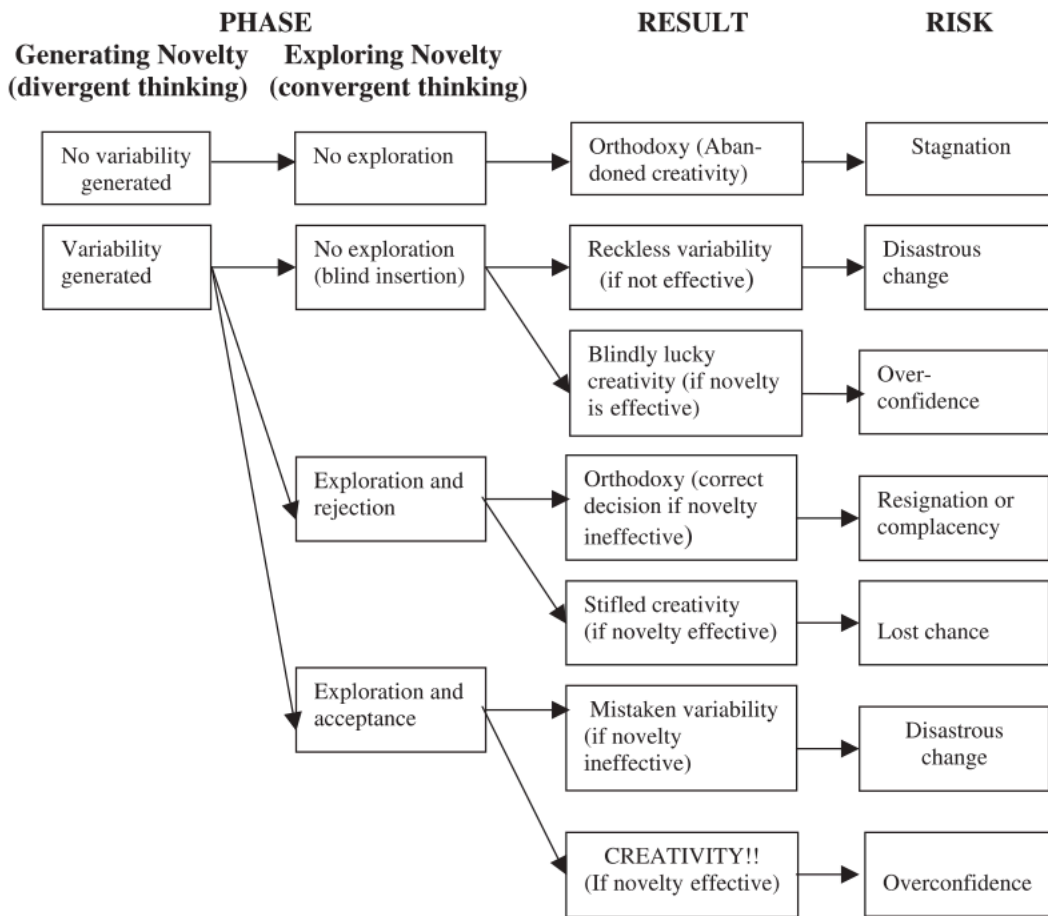


Figure 28: Consequences of differing combinations of divergent and convergent thinking (Cropley, 2006)

The first row depicts the problem of orthodoxy where no variability is generated (no divergent thinking) together with no exploration (no convergent thinking). The risk is to incur in stagnation remaining in the current situation with no willingness to change. When variability is generated a whole new set of possibilities opens up. Variability with no exploration may generate a) recklessness if variability is not effective, possibly resulting in disastrous change b) blind luck creativity if novelty is effective, possibly causing over-confidence as it lacks logic and proofing. Convergent thinking can be defined “correct” or “incorrect”. When convergent thinking correctly rejects an ineffective novelty, a disastrous change is avoided. There is a risk though for resignation or complacency which may lead to decreased stimuli to generate new variability. Convergent thinking can also be incorrect a) in rejecting effective

knowledge which results in a lost chance and b) in accepting an ineffective novelty again causing a disastrous change. Last but not least a correct acceptance of an effective novelty results in creativity.

Different models reflect the link between these two typologies of thinking. The *summation model* states that convergent and divergent thinking add to each other and make up for each other's deficiencies. The *threshold model* has a more dynamic point of view: below a threshold level of convergent thinking, divergent thinking is considered effectively impossible. If convergent thinking approaches this threshold from below, the likelihood of divergent thinking occurring raises accordingly. The *channel model* argues that convergent thinking creates the channel through which how much and what kind of knowledge and information is necessary for divergent thinking is established. Similarly, the *capacity model* believes that convergent thinking is necessary in order to define the amount of information to be applied in divergent thinking.

The strong interdependence between the two styles of thinking makes it impossible to refer to one without mentioning the other. In generating creativity both divergent and convergent thinking are indispensable, but it may occur that with very tight time-constraints and high-risk situations, divergent behaviors stand out more.

#### 4.4.3 Questionnaire

The questionnaire to assess the extent to which an evaluated level of fatigue may impact on a specific task is presented below. The overall structure is divided in two sections. The first section consists of a mutually exclusive, multiple choice question with two options. The second part takes a further step, going into major detail and allowing the participant to choose which of the 10 sentences presented best reflect the task in question.

DATE: \_\_\_\_\_

NAME: \_\_\_\_\_ SURNAME: \_\_\_\_\_

ROLE: \_\_\_\_\_

OPERATION: \_\_\_\_\_

TASK: \_\_\_\_\_

TASK CODE: \_\_\_\_\_

---

DO YOU CONSIDER THIS TASK

CONVERGENT – SOLUTION CAN BE FOUND BY LOGICAL DEDUCTION

DIVERGENT – REQUIRES CREATIVE THINKING AND A “NON-LINEAR” APPROACH TO THE PROBLEM

---

TICK THE BOXES THAT IN YOUR OPINION BEST DESCRIBE THE ABOVE-MENTIONED TASK

RAPIDLY CHANGING INFORMATION

MANY VARIABLES TO TAKE INTO CONSIDERATION

MAY REQUIRE SEVERAL CHANGES OF PLAN OF ACTION

REQUIRES LATERAL THINKING

REQUIRES BEING INNOVATIVE

REQUIRES TAKING RISKS

REQUIRES HIGH LEVEL OF COMMUNICATION

CALL FOR TEMPORAL MEMORY (MEMORY FOR WHEN EVENTS OCCUR)

LENGTHY WITH RESPECT TO THE WHOLE OPERATION

DONE IN THE SECOND HALF OF THE OPERATION

Figure 29: Questionnaire for the assessment of task fatigue impact level

#### 4.4.3.1 Explanation and examples in medical context

The questionnaire is mainly based on two concepts which influence the magnitude of the impact of fatigue on an individual: PFC and time. Points a and b and 1 through to 8 refer specifically to those tasks which cause an activation of the Prefrontal Cortex. They are based on the paper Harrison & Horne, 2000 – thoroughly analyzed above – which defines and evaluates typical divergent tasks heavily impacted by sleep deprivation. The last two points – 9 and 10 – concern the length of the task and its temporal collocation within the whole operation. Below each item of the questionnaire is analyzed and examples in the neurosurgical field are mentioned.

The operator is asked to choose whether the task is reflected more by the term “convergent” or “divergent”. This choice is mutually exclusive and sets a very generic idea of the pathway followed by the tasks – which will be considered in the final score.

- a. *convergent – solution can be found by logical deduction*

A task can be defined convergent – implicitly meaning that it uses convergent thinking – if a problem can be solved through the use of logical deductions and relying on previously stored knowledge. It implies the use of rules and thoughts through a process which has been already validated.

- b. *divergent – requires creative thinking and a “non-linear” approach to the problem*

A task is divergent – uses divergent thinking – when it requires being unconventional. It is based on the retrieving of a broad range of existing knowledge, the combination of disparate information and the production of multiple answers.

After the first double-choice question, the participant is confronted with ten sentences. All the sentences which reflect the task must be selected. All points are coupled with an example in the medical context given by a neurosurgeon.



- 1. *rapidly changing information*

Sleep deprivation impairs individuals' visual and auditory abilities. So, whilst performance in complex tasks may remain overall equal, there is a lack of focused attention. Focused attention can be defined as the type of attention which enables people to promptly detect relevant stimuli. It differs from sustained attention – which is not largely impacted from sleep deprivation – as the latter is the ability to focus on a task for a prolonged amount of time without being distracted.

Example in medical context:

Intraoperative deliquoration of cerebrospinal fluid is not predictable and involves rapidly changing circumstances and information.

- 2. *many variables to take into consideration*

When there are many variables to take into consideration, in most cases some are of shallow importance. It is important to be able to rapidly detect these relevant variables without being distracted by the complex array of those of lower importance. Selective attention comes into play in these circumstances as it is the ability to select and attend to a specific stimulus avoiding distractions.

Example in medical context:

Operation done with the aid of a neuronavigator. A neuronavigator is a computerized location system that is used during brain surgery. It's a system that consists of a station equipped with a monitor, connected to two infrared cameras and small refractive spheres, that aid the recompositing of the image. The neuronavigator provides real-time information by overlapping on the monitor both the position indicator of the surgical instrument with the images of computed tomography or magnetic resonance imaging of the patient acquired beforehand. For these reasons the use of the neuronavigator requires taking many variables into consideration.

- 3. *may require several changes of plan of action*

As seen in the paragraph “Prefrontal Cortex”, Killgore, Balkin, & Wesensten, 2006 found that sleep deprived individuals were more likely to persevere in their initially chosen strategy rather than change to a more appropriate one. SD impacts highly on this ability to evaluate the suitability of a plan of action and eventually on the necessity to modify it. Sleep deprived individuals tend to be very reluctant in altering initial decision.

Example in medical context:

Recognizing the biological nature of the tumor mass may require several changes of plan of action as the heterogeneity of tumor masses and the uniqueness of some cases may be revealed only in the middle of the operation.

- 4. *requires lateral thinking*
- 5. *requires being innovative*

Points 4 and 5 are two ways of expressing a similar concept. It was chosen to insert both these sentences as slightly different ways of explicating a concept could reflect better what different individuals experience. Lateral thinking and being innovative are at the base of the definition of divergent impact and are seen to be strongly impacted by fatigue.

Example in medical context:

Individuation on a non-predictable anatomical variable.

- 6. *requires taking risks*

Referring again to the paper by Killgore, Balkin, & Wesensten, 2006, it was seen that sleep deprived participants had a greater willingness to take risks. There is a very evident decrease in concern of negative consequences, even more when faced with high rewards. This could be very dangerous in delicate situations and fields such as healthcare where human lives are potentially at risk.

Example in medical context:

Clipping of a cerebral aneurysm which consists in the exclusion of the aneurysm from the cerebral circulation through the application of one or more “clips” on the neck of the aneurysm. It is a very risky task within an operation.

- *7. requires high level of communication*

SD impacts on individuals also in the form of language processing alterations. An obvious qualitative difference in speech and verbal communication skills has been seen. Pilcher, Jennings, Phillips, & McCubbin, 2016 found that even one simulated night shift under sleep deprivation settings is detrimental to auditory attention and language comprehension.

Example in medical context:

Processes done during an awake brain surgery also called craniotomy. This type of operation is done to treat some neurological conditions including also brain tumors and epileptic seizures. Awake craniotomy is done to reduce the probability of damaging critical brain areas that control speech and other important skills. It requires high level of communication between the surgical team as they must all be aligned, from the neurosurgeon to the anesthesiologist and speech-language pathologist.

- *8. calls for temporal memory (memory for when events occur)*

Although it's a rather recent matter of study, temporal memory has been seen to be affected by sleep deprivation. Harrison & Horne, 2000a found that SD did not significantly impair the recognition of faces. It did though affect negatively temporal memory (recency) aspects of tests despite the test being designed in order to gain optimum performances (novelty, stimulating, short etc.).

Example in medical context:

Timing of the clipping of a blood vessel.

- 9. *lengthy with respect to the whole operation*

Example in medical context:

Preparation of the patient in a particular position.

- 10. *done in the second half of the operation*

Participants are generally seen to feel increasingly tired with time-on-task (Roy, Bonnet, Charbonnier, & Campagne, 2013). This is why the length of the task and its collocation within the whole operation is important.

Example in medical context:

Verifying results.

#### 4.4.3.2 Final score and ranges

The final score was calculated using a simple formula. The number of ticked boxes from point 1 to 10 were summed. If the individual had chosen a) convergent, -1 was added at the end of the formula. If, on the other hand, he had chosen b) divergent, then +1 was added.

$$y = \sum_{i=1}^{10} x_i \pm 1$$

*where + 1 if divergent and – 1 if convergent*

The total range of scores goes from 0 (the only possible negative result of -1 is rounded to 0) to 11. The three ranges “low”, “medium” and “high” can be seen below.

$$if \begin{cases} 0 \leq y \leq 3 \text{ then range} = \text{low} \\ 4 \leq y \leq 8 \text{ then range} = \text{medium} \\ 9 \leq y \leq 11 \text{ then range} = \text{high} \end{cases}$$

## 4.5 Integrating Human Error Probability in the model

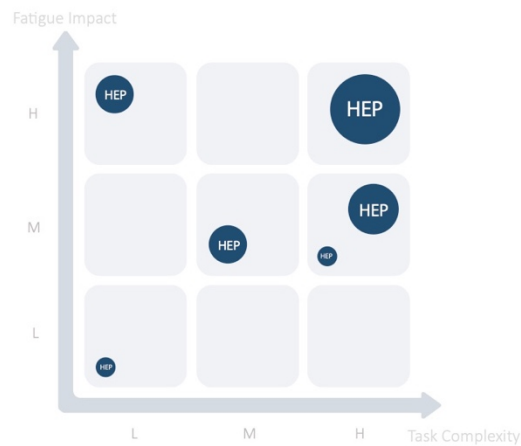


Figure 30: HEP in the FI-TC model

Within the nine regions, circle-shaped Human Error Probabilities are included to add depth to the model. The bigger the bubble, the higher the probability level of human error. Task complexity on the x-axis is measured relative only to the a priori characteristics of the tasks, all the other attributes and features are within the HEP. HEP is measured with any HRA technique of choice, integrating all IF different from fatigue – that will be assessed subsequently.

Human Error Probability can be defined as the quantification of the unreliability of humans in performing tasks. In first generation HRA techniques, it is actually calculated in the same way as the assessment of failure rate of products. Second generation techniques, on the other hand, explain human behavior through cognitive models and HEP is based on cognitive analysis (Pan & Wu, 2018).

Although the calculation of the HEP differs slightly according to the implemented technique, the majority of quantitative HRA techniques follow a common process as stated by Onofrio & Trucco, 2018. Firstly, the conditional or nominal HEP is assessed in a preliminary step. Secondly, the relevant Influencing Factors or Performance

Shaping Factors are taken into consideration to modify the basic HEP according to the following formula:

$$HEP = NominalHEP * \left( \sum_{i=1}^n PFS_i * W_i \right)$$

Where:

$W_i$  = weight of the i-th PFS for the specific task

$PSF_i * W_i$  = weighting effect for the i-th PSF on the nominal HEP

The presence of the HEP within the model considers the importance of having an HRA aspect to raise awareness of the high unreliability of the healthcare sector. HEP are not directly linked to *fatigue impact* and *task complexity* which gives an interesting cue of analysis which will be brought forward in the following paragraphs. The possibility to evaluate HEPs through the use of any preferred HRA technique, keeps highlighting the highly versatile use that can be made of the presented model. It gives great possibilities of adaptability to different fields and circumstances.

#### 4.6 Off-line mapping analysis

The first step coincides with the first aim of the FI-TC model. It regards the individuation of the critical areas on which the subsequent step – the fatigue evaluation framework – must be applied. Mapping the areas of the model induces an off-line analysis where the matrix becomes an ex-ante evaluation tool. The vertical dimension of fatigue impact is necessary to map the fatigue trait induced by the task net of the entry level of the operator. This is why criteria on how to discriminate between tasks is needed and will be provided below. Criteria will discriminate between those tasks named “green zone” and those tasks for which the real-time evaluation of the operator’s fatigue is needed.

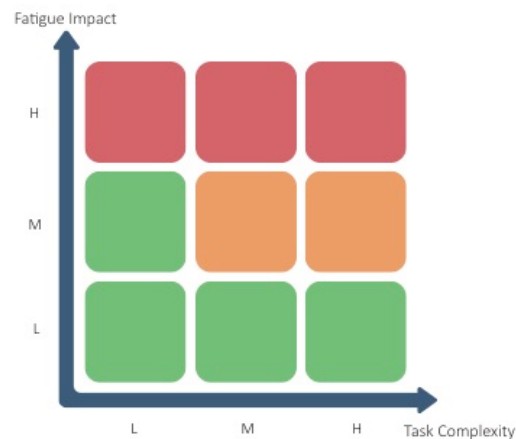


Figure 31: The definition of the FI-TC model "zones": green, orange and red

- *Green zone*: this zone comprises the three boxes where fatigue impact is low and the box where fatigue impact is medium and task complexity low. It was named “green zone” as according to the information analyzed in the literature study, for any of these tasks, the real-time fatigue level evaluation is not needed. The activation of the real-time evaluation depends firstly on the fatigue impact level of the task and secondarily on its complexity.
- *Orange zone*: this zone is made up of two boxes of the model. The boxes where fatigue impact is medium and task complexity is both medium and high. Even though fatigue impact is at the same level of a box in the “green zone”, the influence of task complexity is considered and real-time fatigue evaluation for certain levels of HEP is assumed necessary. Real-time analysis will be activated if:

$$HEP > \alpha$$

The  $\alpha$  level of HEP cannot be defined a priori. This is because different surgical realities may and will have diverse calibrations of the parameter. It is important to understand though that Human Error Probability is the final calibration parameter that decides if real-time analysis must be done.

- *Red zone*: this zone comprises all three boxes where fatigue impact level is high for any level of task complexity. The activation of the real-time analysis is based again on a HEP level arranged beforehand.

$$HEP > \beta$$

In the same way as for the  $\alpha$  parameter, also the  $\beta$  value cannot be defined a priori but needs to be contextualized. The only indication that can be given is that  $\beta < \alpha$  as it must reflect the higher criticality of the “red zone”.

#### 4.7 Real time analysis – the fatigue level evaluation framework

The fatigue level evaluation framework (Figure 30) was created as an extension of the FI-TC model. Its aim is to provide an assessment of the individual’s fatigue level when a real time evaluation is required.

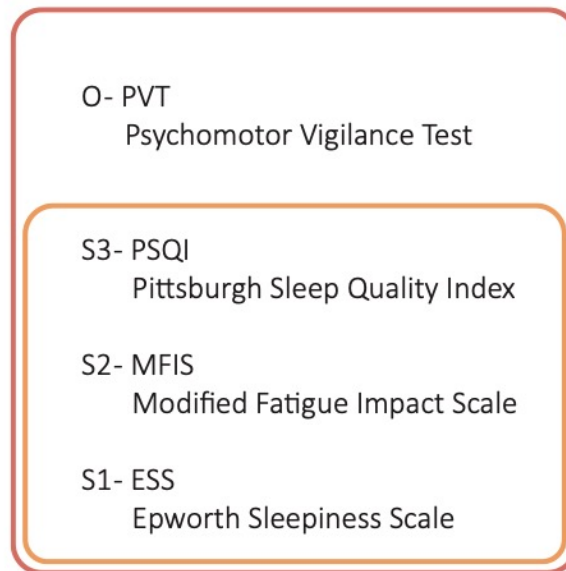


Figure 32: Fatigue level evaluation framework

As shown in the figure, it is composed of four different tests. The Epworth Sleepiness Scale (S1), the Modified Fatigue Impact Scale (S2) and the Pittsburgh Sleep Quality Index (S3) are all subjective questionnaires whilst the Psychomotor Vigilance Test (O) is an objective test. The capital letter as an identifying code distinguishes between these typologies. Each of the methods used to evaluate fatigue will be deeply presented in the chapters below.



Both the test and the questionnaires aim at the same purpose – evaluating the fatigue level of an operator – even though it is done through different questions and techniques. After the literature content analysis, three subjective methods were selected from all existing questionnaires. The primary reason for them being considered is their high level of validation and consistency. The real time application of the fatigue evaluation tools is not always required. As described in the previous paragraph, it depends on the color of the regions to which tasks belong, and their HEP value. For all of the tasks that are positioned in the “green zone” the framework should not be applied. If the tasks are placed in the “orange zone” with the level of HEP higher than the nominal value set, the implementation of a subjective tool for the fatigue level evaluation is required. As mentioned, three are the subjective tools proposed as valid by literature studies but just one is required for the evaluation. If the task is placed in the “red zone” with the level of HEP higher than the nominal value set, a deeper analysis is required because of the riskiness of the situation. The implementation of the framework foresees the execution of both objective and subjective fatigue evaluation methods in this zone I in order to be precise in the tiredness evaluation of the operator. To deny a surgeon the ability to operate both of the tests should result negative. The rating scales of each test will be presented in the next paragraph.

In deciding timing of when surgeons are to complete the test and questionnaires of the framework, studies on circadian rhythms resulted useful (Valdez, 2019a).(Valdez, 2019b)(Valdez, 2019b) As it has been seen many times, fatigue has many effects on individuals. One of the main effects is an impaired attention level. Attention in simple words refers to that cognitive process which allows individuals to effectively interact with the external environment. REF in his study on circadian rhythms in attention defined attention according to Posner and Rafal’s model. According to this model attention is made up of four components: a) tonic alertness, which is the ability to react to an event in the environment, seen as a general activation of the organism b) phasic alertness, which on the other hand refers to the response to an event after a

warning signal takes place c) selective attention, which refers to the ability to select and attend to a specific stimulus avoiding distractions d) sustained attention, that is the ability to focus on a task for a prolonged amount of time without being distracted. Evaluating circadian rhythms in attention it was found that components of attention are at a low level in the morning (7am to 10am) because of circadian rhythms reaching their lowest level. From 10am to 2pm attention levels increase but then decrease again after lunch (2pm to 4pm) in the so-called post-lunch dip. Attention improves from 4pm to 10pm to again decrease between 10pm and 4am and reach its lowest level at dawn from 4am to 7am. These time ranges apply to an individual that sleeps approximately 8 hours per night from 11pm to 7am. Attention so was seen to change on average every 4 hours. According to this the results obtained from the measurement of the fatigue level on an individual have a duration of approximately 4 hours and necessitate to be revised after this time has passed.

#### 4.6.1 Epworth Sleepiness Scale

The Epworth Sleepiness Scale was developed by Dr. Murray Johns in 1990 and lately modified in 1997 (the final version can be seen in Appendix 1). It consists in a self-administered, subjective questionnaire to evaluate a person's Average Sleep Propensity in daily life. The higher the score in the ESS, the higher the individual's ASP. The questionnaire takes 2-3 minutes to complete and respondents are required to rate on a 4-point scale (0-3) their chance of dozing off in eight described situations in recent times (Johns, 1997).

More in detail, the ESS evaluates the chances of dozing off in activities that differ extensively based on their somnificity. These eight ESS scores give estimates of different Situational Sleep Propensities for the individual. Putting these eight SSP scores together the overall ASP is assessed. It is interesting to deepen the knowledge of the meaning of these three quite unused terms:

- *Somnificity* is a term introduced by Dr. Johns in 2002 to describe how different postures, circumstances and activities ease or delay sleep emergence in individuals. It was discovered that it is not dependent on characteristics of people or their lifestyles. For example, lying down in bed with dimmed lights and ceasing voluntary movements such as talking coupled with closing eyelids enhances sleep in less than 15 minutes;
- *Situational Sleep Propensity* measures the propensity of an individual to fall asleep repeating the same activity at the same time of the day. It is highly dependent on circumstances, so the method used to calculate it is very relevant. For example, SSP of a same activity can be measured with the Multiple Sleep Latency Test or the Maintenance of Wakefulness Test and will result in different outputs;
- *Average Sleep Propensity* is a hypothetical construct defined by Dr. Johns which reflects a general level of sleepiness of the individual under different circumstances. This ASP is the output of the ESS which sums eight different SSPs.

Overall, people with an atypically elevated level of ASP are seen to be more likely to doze off when performing those activities characterized by a low somnificity than “average” people do. The Epworth Sleepiness Scale is based on this belief.

The ESS final score ranges from 0 to 24 and is the result of the sum of the eight scores. General ESS scores, as stated by Johns, 1997 can be interpreted in such way:

- 0 - 5 lower normal daytime sleepiness;
- 6 – 10 higher normal daytime sleepiness;
- 11 – 12 mild excessive daytime sleepiness;
- 13 – 15 moderate excessive daytime sleepiness;
- 16 – 24 severe excessive daytime sleepiness.

For the purpose of the model presented in this thesis the above five ranges were re-grouped into the following three:

- 0 – 5 low, comprising only the lower normal daytime sleepiness, accepted in the “red zone”;
- 6 – 12 medium, comprising the higher normal and mild excessive daytime sleepiness, accepted in the “orange zone” and rejected in the “red zone”;
- 13 – 24 high, comprising the moderate excessive and severe excessive daytime sleepiness, which denies the operator to perform regardless of the zone.

The ESS has been validated extensively and can be considered a well-grounded subjective evaluation measure. It does though have various limitations. Some of these limitations include: being a subjective measure in may be influenced by bias and inaccuracy as other reports; it is unable to make accurate predictions; it can't distinguish which factors cause a particular level of ASP and lastly it is not suitable for people with a serious cognitive impairment.

#### 4.6.2 Modified Fatigue Impact Scale

The Modified Fatigue Impact Scale (Appendix 2) is a modified version of the 40-item Fatigue Impact Scale. The FIS was originally developed to evaluate fatigue on quality of life of patients affected by chronic disease such as Multiple Sclerosis. The MFIS became such when the Multiple Sclerosis Quality of Life Inventory was created and went to cover phase 2. The step from the 40 items of the FIS to the 21 of the MFIS was made by eliminating content-redundant items (Larson, 2013).

The MFIS consists of 21 statements that describe effects of fatigue. Individuals have to rate the statements from 0 (never) to 4 (almost always) basing their answers on the past 4 weeks and the questionnaire takes 5-10 minutes to complete. Items of the MFIS can be aggregated in three subscales:

- Physical subscale comprises items 4, 6, 7, 10, 13, 14, 17, 20 and 21. This scale ranges from 0 to 36.

- Cognitive subscale comprises items 1, 2, 3, 5, 11, 12, 15, 16, 18 and 19. This scale ranges from 0 to 40.
- Psychosocial subscale comprises items 8 and 9 and ranges from 0 to 8.

Overall the MFIS final score goes from 0 to 84. To the best of the authors' knowledge, no articles have been published to date regarding specific cutoffs for ranges except for one. Flachenecker et al., 2002 compared four different fatigue rating scales: the FSS, MFSS, MFIS and the VAS on patients affected from MS. Patients were classified beforehand based on their answers to a structured questionnaire (Multiple Sclerosis Council, 1998). Individuals who responded that fatigue a) represented one of their three most debilitating symptoms b) occurred on a daily basis or on most days c) restricted and limited daily activities at home and work, were classified as MS-related fatigue group (MS-F). Those who fulfilled none of the above criteria were categorized as MS-related non-fatigue group (MS-NF). Lastly, those who fulfilled partly the criteria were classified as "borderline".

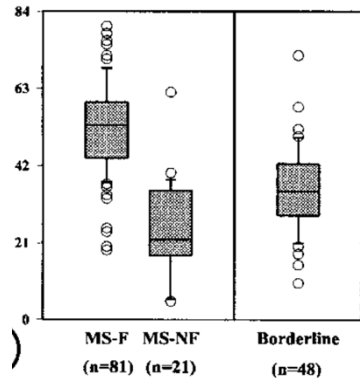


Figure 33: Fatigue Severity Scale (Flachenecker et al., 2002)

On the left in Figure 31 the results of MFIS scale in MS-F (left), MS-NF (middle) and borderline (right) patients is presented. MFIS saw very discriminative results with no overlap of the 10<sup>th</sup> and 90<sup>th</sup> percentiles for MS-F and MS-NF groups. Without considering the outliers, approximately MS-F ranged from 38 to 69, MS-NF from 4 to 38 and borderline from 21 to 45. Considering that those in the MS-F group responded positively to three very strong statements, that range of responses was considered high. Borderline values were considered to define the cut-off between low and medium ranges. Overall the three defined ranges are:

- 0 – 21 low, accepted in the "red zone";
- 22 – 37 medium, accepted in the "orange zone" and rejected in the "red zone";
- 38 – 84 high, operator is denied performance regardless of the zone.

MFIS is a highly used and validated subjective measure of fatigue. It gives a basic understanding of an individual's level of fatigue in order to allow him to or to not pursue specific tasks which are impacted by fatigue.

#### 4.6.4 Pittsburgh Sleep Quality Index

Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R., & Kupfer presented the Pittsburgh Sleep Quality Index (Appendix 3) in 1998, a sleep questionnaire developed with several goals: a) provide a solid, applicable and standardized validated measurement of sleep quality technique b) distinguish clearly between "good" quality sleepers and "poor" quality sleepers c) make available for use an easy to use and understandable index for researchers and clinicians to interpret d) provide a synthetic and useful evaluation of different sleep disturbances. The PSQI determines sleep quality focusing on the previous month. This time frame was chosen as it is intermediate between post sleep inventories and survey type questionnaires. The latter usually assesses difficulties in sleep in the previous year or more, the former focuses only on the previous night's sleep.

The PSQI consists of 19 self-rated questions and five to be answered by a roommate or bed partner. These last questions are not considered in the final score but are for clinical information only. The quality of sleep is well determined by the question assessment on seven different factors through which they are grouped and valued as seven component scores. From the questions it is possible to evaluate the operator's sleep duration, how the operator has slept in terms of sleep quality and sleep efficiency, how long it takes him to fall asleep (sleep latency), if the operator was on any medications while sleeping, daytime dysfunction and sleep disturbances. The PSQI score rules (Appendix 4) are then used for the transposition of the scores in a scale from 0 to 3 in order to add them all up easily. The final global score has a range of 0-21, where the lowest scores refer to good sleep quality while higher it goes worst are the sleep quality conditions. A global PSQI score greater than 5 distinguishes good

and poor sleepers. Above this threshold (>5) the operator's sleep is defined as poor quality (Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R., & Kupfer, 1998). For this reason, the chosen ranges are:

- < 5 is accepted in the “red zone”
- ≤ 6 is accepted in the “orange zone”.

The scale was tested for eighteen months before being published. It raised consensus with the opinions of patients that find the index easy to use. It demonstrated then high level of internal consistency looking at the 19 individual questions and their relative seven components. Another strength of the index is the stability over time of the global and component scores and individual question answers. The unique limitation is represented by the difficulty in understanding the sense of what is written but over the time PSQI has been translated in 56 different languages.

#### 4.6.5 Psychomotor Vigilance Test

The Psychomotor Vigilance Test is an objective measure of reduced alertness caused by fatigue and sleep deprivation. Through a critical analysis of the literature and a statistical study, PVT resulted the most used metric in fatigue evaluation while it is at the second place if only the healthcare context is taken into consideration (Figures 8 and 9). Because of its high usage with respect to the others, PVT has been selected for the framework structure.

The PVT is a simple reaction time task at random intervals, alternated by seconds of pause that vary between 2 or 10 s in the 10 min PVT – the most commonly used. The aim is to present visual or auditory stimuli to individuals which they must react to (Gore B.F., Casner S.M., Wickens C.D., Pallesen S., 2018). The subject in analysis is required to press a button as soon as he/she hears a sound or sees the light. The final number of lapses of attention on each test session measure the level of instability in sustained attention caused by sleep deprivation. Lapses of attention are

quantitatively represented by the response time. The response time (RT) is the time taken by a system or functional unit to react to a given input. There is more than one type of PVT because they may differ in terms of operative time and tools (laptop, tablet or smartphone) with which the test is performed.

The PVT implemented in the framework (Basner, Mollicone, & Dinges, 2011) was chosen because of its speed of execution, flexibility, ease of use and repeatability. It is performed on tablets though it may be transferred on smartphones, it takes 3 minutes to complete and the inter-trial intervals range from 1 to 4 s. The point of the test is the generation of a stimulus as a rolling millisecond counter that appears inside a red box, written in yellow, on a black screen. Individuals are required to react as quickly as possible finger-tapping the tablet or smartphone touch screen, immediately below the stimulus using the index of the dominant hand. If the action produces a right answer the rolling millisecond counter stops and displays the response time for 1 s. An example is shown in the Figure 32 below.

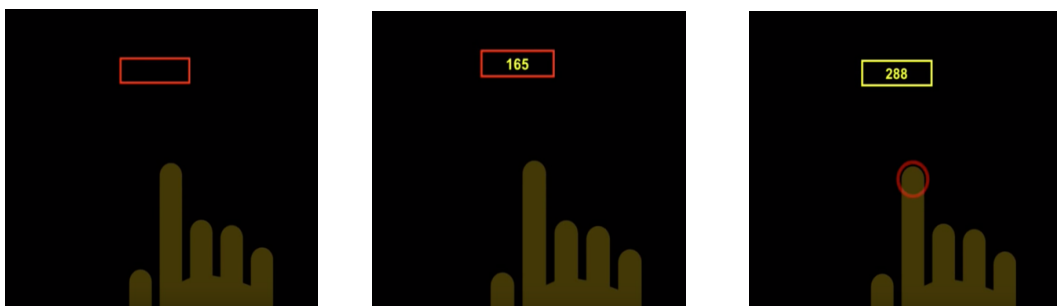


Figure 34: Psychomotor Vigilance Test execution phases

Although the test was shortened by 70%, the average number of stimuli decreased only by 32,4% (from 93,6 to 62,3). This specific type of PVT is based on a threshold level of RT of 355 ms. If reaction time is greater than 355 ms, the response is considered failed and the run is considered a lapse of attention (Grant, Honn, Layton, Riedy, & Van Dongen, 2017). If the response time is below 100 ms the run is



considered a false start, as in the case of the absence of stimulus. In this case the run is not valid and not taken into account. Answers are not considered also when pressing the wrong button or failing to release it for 3s or longer.

The study was structured by gathering data from a total sleep deprivation study (Basner & Rubinstein, 2011) and from a partial sleep deprivation protocol (Basner et al., 2011). The TSD study involved 33h of wakefulness. Test runs 1-7 (9am to 9pm) were averaged and reflected non-sleep deprived individuals. Test runs 8-17 (11pm to 5pm) were averaged and reflected the sleep deprived state, based on the fact that PVT performance begins to decline after 16h of wakefulness (Van Dongen, Maislin, Mullington, & Dinges, 2003) . The PSD study involved two initial baseline nights (B1 and B2) of 10h time in bed (10pm to 8pm), followed by 5 nights (R1 to R5) of partial sleep deprivation of 4h TIB (4am to 8am).

The chosen outcome measures were based on a previous paper by Basner & Dinges, 2011 where outcome metrics for a 10-min PVT were ranked based on their effect sizes for acute total and chronic partial sleep deprivation (Table 7).

Rank	Acute Total Sleep Deprivation		Chronic Partial Sleep Deprivation	
	Outcome Metric	Effect Size (95% CI)	Outcome Metric	Effect Size (95% CI)
1	Number of Lapses and False Starts	1.94 (1.53; 2.69)	Mean 1/RT	1.21 (0.94; 1.59)
2	Mean 1/RT	1.93 (1.55; 2.65)	Fastest 10% RT	1.13 (0.87; 1.51)
3	Number of Lapses	1.86 (1.45; 2.59)	Slowest 10% 1/RT	1.11 (0.78; 1.57)
4	Slowest 10% 1/RT	1.86 (1.54; 2.58)	Median RT	0.96 (0.74; 1.29)
5	Performance Score	1.68 (1.34; 2.28)	Number of Lapses	0.91 (0.71; 1.18)
6	Lapse Probability	1.59 (1.27; 2.16)	Number of Lapses and False Starts	0.90 (0.69; 1.17)
7	Fastest 10% RT	1.54 (1.22; 2.08)	Performance Score	0.88 (0.69; 1.13)
8	Number of False Starts	1.29 (0.98; 1.78)	Lapse Probability	0.88 (0.69; 1.13)
9	Mean RT	0.64 (0.53; 1.23)	Mean RT	0.43 (0.35; 0.91)
10	Median RT	0.38 (0.33; 1.83)	Number of False Starts	0.39 (0.14; 0.76)

RT, response time; CI, confidence interval; Effect sizes of Mean 1/RT, Slowest 10% 1/RT, and the Performance Score were multiplied by -1 to facilitate comparisons. Nonparametric 95% bootstrap confidence intervals are shown in parenthesis.

Table 7: Effect sizes of outcome metrics for acute TSD and chronic PSD (by Basner & Dinges, 2011)

Out of these 10 measures Basner et al., 2011 chose: a) fastest 10% of RT b) mean 1/RT (or reciprocal response time) c) slowest 10% 1/RT d) number of lapses e) newly

developed performance score. The latter is calculated as 100% minus the number of lapses and false starts (100% optimal performance, 0% worst possible performance).

For the purpose of defining the ranges of the PVT presented in the fatigue level evaluation framework only the TSD was considered. This decision was made in order to give clear and strict cut-offs, not considering the fuzzy boundaries of PSD. In Figure 33 below the averaged results of 31 subjects' performances during a 33h period of sleep deprivation are depicted (on the left). On the right for each of the five outcome variables, results were centered around alert performance (bouts 1-7 from 9am to 9pm).

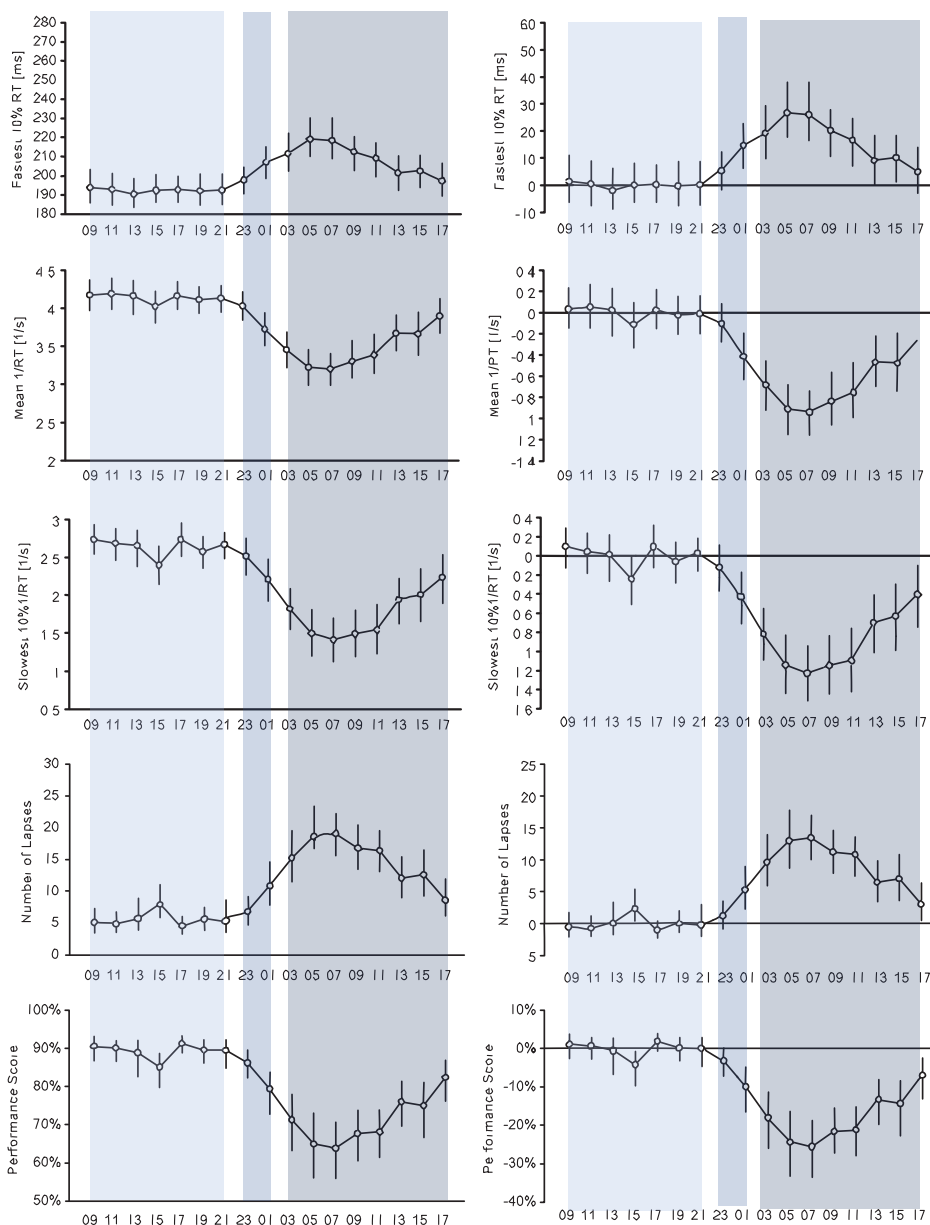


Figure 35: PVT performance under TSD conditions

Results from Table 8 were combined with the need of a simple technique with a high easiness of use. Performance score with an effect size of 1.68 for acute total sleep deprivation and 0.88 for chronic partial sleep deprivation seemed to be the right compromise. By grouping the results obtained at the start of the period that reflects

sleep deprived individuals' performances 11pm – 1am), boundaries are more clearly notable (Table 8 below).

	9am – 9 pm	11pm – 1am	3am – 5pm
<b>Performance score</b>	89,4%	82%	71,5%

Table 8: Performance score on PVT

Performance score results have been averaged according to three time slots:

- 9am – 9pm, non-sleep deprived
- 11pm – 1am, start of sleep deprivation
- 3am – 5pm, fully sleep deprived

In order to be able to perform tasks placed in the “red zone” with a high fatigue impact, the threshold has been fixed at a minimum performance score of 85,7% - the average between non-sleep deprived and start of sleep deprivation results. In other words, a maximum of 14 lapses and false starts in the 3-min PVT is required to be considered fully capable to complete successfully certain tasks.

The PVT is a very commonly used metric and its facility makes it very adaptable to different context. This specific PVT and its duration of just 3 minutes makes it perfect to be inserted in the framework together with other metrics, not extending the time of assessment above the limits. Its objectivity is the main reason it has been appointed to the assessment of fatigue level in those individuals that need to perform tasks highly impacted by fatigue.

#### 4.8 The model flow chart

The flow chart represents the stages that have to be followed to correctly implement the FI-TC model. It has been inserted in the thesis to summarize all the concepts previously described and help the reader in the understanding of the temporal

sequence of the steps to follow. It is made up of rectangles, arrows and colored squares. The former represents actions to be performed. The arrows indicate the direction in which the diagram should be read and therefore, they indicate order of precedence of actions. The colored squares instead, stand for positive or negative results, depending on whether the square is respectively green or red. After the action called "check of region" it is possible to see arrows of different colors exiting the box. The colors of the arrows - green, orange, red - stand for the colors of the model zones which refers to the criticality level of the task.

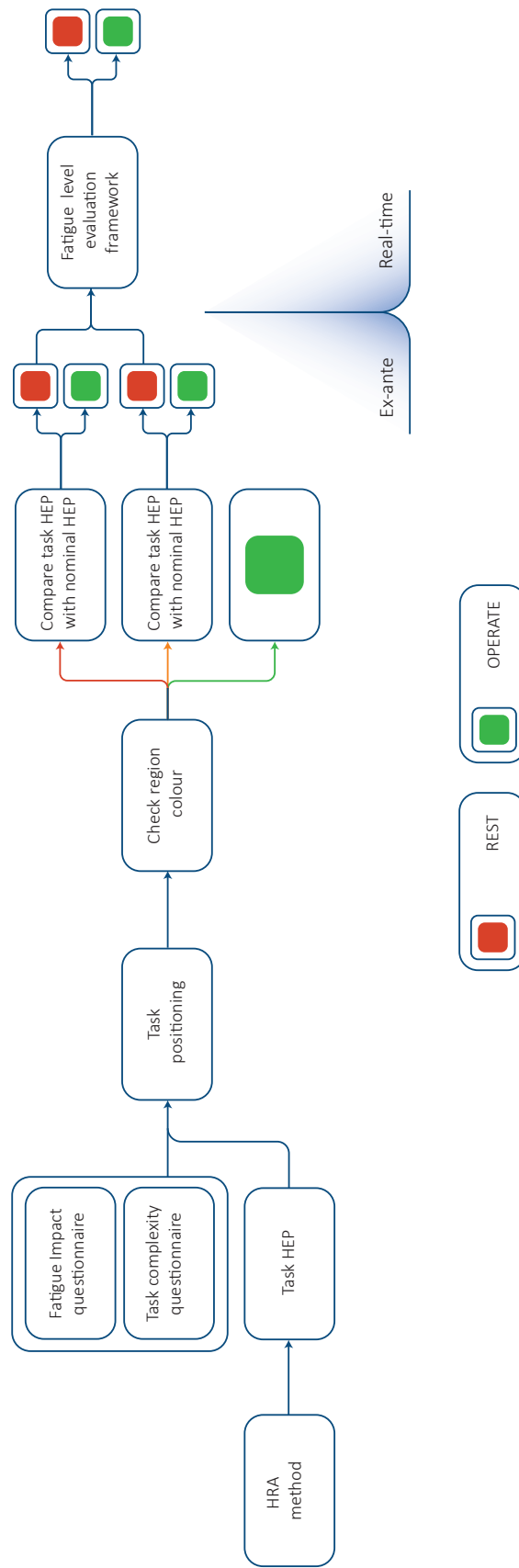


Figure 36: The model flow chart diagram

Let's assume a hospital wants to decrease the probability of errors occurring in a certain surgery specialty and operation. The chief decides to adopt the FI-TC model in order to identify in advance whether the operation in question is composed of critical tasks. If it does, the model will be also used for real-time assessment of surgeons' fatigue. The actions that must be accomplish are listed below.

- Distribution of the fatigue impact and task complexity questionnaires surgeons. The number of surgeons considered changes according to how many surgeons are involved in the surgery team in question
- Utilization of an HRA method of choice to calculate the HEP on each task that forms the operation
- Positioning of each task on the model according to the results obtained by the first and the second step. The questionnaires' answers indicate which level of fatigue impact and task complexity characterize the task and its positioning on the axes whilst the HEP indicates the dimension of the bubble
- Check of the color of the zones where the tasks have been positioned. If the zone is green, follow the green arrow, and so on
- If the followed arrow is green the task into question is not critical, so it does not require other controls. The surgeon is free to operate
- If the followed arrow is orange or red, it means that another check is required. The HEP calculated at the second step is here compared with the nominal HEP value. If the nominal HEP is not exceeded, go to the green square, contrarily go to the red square
- If the square is green the surgeon is free to operate. If the square is red move on to the next step.

The process that has been described until now is made up by all those actions that form the ex-ante model. The latter occurs totally offline, without including values related to the health status of surgeons. It consists in an advanced critical task detection tool. The last two steps include:

- Utilization of the fatigue level evaluation framework
- According to the results the surgeon can operate (green), or he will be advised to rest (red).



## 5. Model validation and calibration

---

The research method of the work presented in this chapter consists in the validation and calibration of the Fatigue Impact -Task Complexity model. Given the current state of the situation, due to the health emergency caused by Covid-19, surgeons and healthcare operators are all unavailable. The calibration and validation procedure of the model via expert judgement will be explained but will be missing the section of full data collection and analysis.

The model validation is developed through a theoretical explanation of how the process would have been taken out and what was actually obtained. The calibration was explained through the application study in robotic surgery. Lastly, the subjective fatigue evaluating metrics are analyzed through a real-time feature assessment.

---

## 5.1 Overview of the process

The process adopted in evaluating the FI-TC model is comprised of different sections and diverse aspects are under analysis (Figure 35).

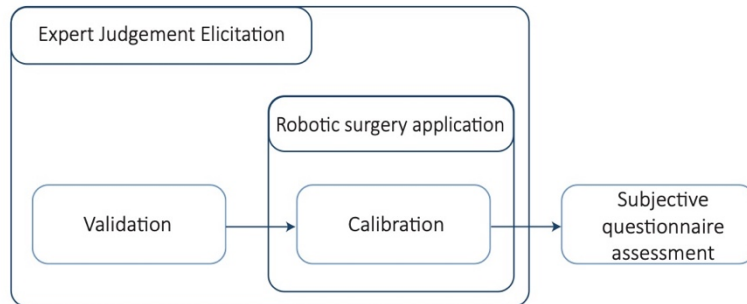


Figure 37: The process in the evaluation of the FI-TC model

The process includes validation and calibration taken out through Expert Judgement Elicitation, and the subjective questionnaire assessment. EJE was implemented in a two-step procedure. First of all, it was used for the validation of the ex-ante feature of the FI-TC. The validation step was expanded with feedbacks from surgeons outside of the robotic environment, departing from a context-centric validation. This was done in order to broaden reasoning to further considerations. Secondly, through the specific case-study regarding the robotic context, EJE was adopted to calibrate the ex-ante feature. The second step, contrarily to the first, is tightly task-centric and not surgeon-centric.

Expert Judgement Elicitation stands at the basis of the validation and calibration method. EJE refers to the systematic approach of gathering and organizing subjective judgements of experts on a specific subject. It aims to make explicit and accessible the intangible and unprinted knowledge of experts. It is based on their personal experience and insights they may give on the current limitations and weaknesses of the published knowledge (Knol, Sluijs, & Slottje, 2008).

## 5.2 Validation of the model

Validation of the items refer to them being easily readable, usable and understandable. Even though the quantity of data gathered was not satisfying nor enough to be generalized, there was an effective possibility of receiving negative feedbacks. Negative feedbacks would have sprung the need of modifying terminology, labels or propose new scales.

### 5.2.1 Fatigue impact scale validation

The first scale to be validated was the fatigue impact one with its 2 generic and 10 specific elements. Initial intentions were to involve the whole urology surgical team of Niguarda hospital in Milan. Having feedbacks from different people from within the same context would have been of extreme usefulness to draw context-specific conclusions. As explained above, this was not possible. What was achieved though was to have the responses of two surgeons: a neurosurgeon of Borgo Trento Hospital in Verona and a urologist of Niguarda Hospital in Milan. Both doctors agreed on the readability, usability and understandability of the fatigue impact scale items.

### 5.2.2 Task complexity scale validation

As stated in the paragraph above, intentions were to expand the validation process. Again, with the task complexity scale this was not feasible. To the two doctors mentioned above, a third participated to the validation of the scale in question, a general surgeon of Perderzoli Hospital in Peschiera. All three doctors were perfectly aligned and confirmed the legibility, clearness and comprehensibility of the items. Having had the possibility to submit the task complexity questionnaires to two surgeons belonging to an open context rather than a robotic one, opened a world which had already been intuited.

A very relevant aspect emerged from the discussion with experts during the validation in regard to the relevance of some items of the scale with respect to others. In other words, the usability of the scale could be increased with the introduction of weights. Weights could be juxtaposed to each item of the task complexity questionnaire. This would give experts the “power” to order items based on their impact, could possibly induce to the elimination of useless items and would make the scale more context specific. The three previously mentioned surgeons were asked to weight the items of the scale and eliminate those items they deemed purposeless. Even though three results are not enough to extract a pattern, it is very interesting to see how some items totally differ whilst others are aligned. The results from the surgeons are presented in the table below.

Item	(Robotic) Urologist	(Open) General surgeon	(Open) Neurosurgeon
Things	5	15	55
Data	/	9	10
People	20	10	2
Worker instructions	/	19	7
Reasoning	25	18	20
Mathematics	/	/	/
Language	/	5	1
Worker technology	5	9	2
Worker interaction	20	15	3
Human-error consequences	25	/	/
	<b>100</b>	<b>100</b>	<b>100</b>

Table 9: Task complexity scale weighted by expert surgeons

Even though overall validity was deemed positive by all three doctors, when coming to analyzing the single items of the scale, they believe some items should be deleted. Both surgeons coming from the open context eliminated the *Mathematics* component as it is not relevant to their role and the *Human-Error Consequences* as every action and consequences are 100% responsibility of the surgeon and evaluating them in a scale, they believe is redundant. On the left of the table the results given by the urologist from the robotic context differ largely. The *Mathematics* component is again seen as not required, followed by *Data*, *People* and *Language*. Looking more precisely at the actual values given to the items the only one that can be comparable is that given to *Reasoning*, which received 25, 18 and 20 respectively from the urologist, the general surgeon and the neurosurgeon. What these results suggest are not only the necessity to use the model in a specific context (robotic or open) but also in a specific surgical specialty.

### 5.3 Calibration of the model in the robotic surgical context

Calibrating the model via expert judgement is necessary to give a measure of reliability of the scales. Though both scales have a very strong literature background, their applicability to the surgical context and to any other context in which they may be used, needs a reliability evaluation.

The calibration process of the model was structured as an application in robotic surgery at the urology department of the Niguarda Hospital, Milan. The specific operation that was targeted was the Bocciardi-Approach Robot-Assisted Radical Prostatectomy which is detailly explained in paragraph 5.3.1. What the targeted operation implies is the extremely technological context in which the simulation is developed. It's a unique circumstance where human variables are as important and vital as the technological ones.

### 5.3.1 Bocciardi-Approach Robot-Assisted Radical Prostatectomy procedure

The Bocciardi-Approach Robot-Assisted Radical Prostatectomy (BA-RARP) is a surgical operation for clinically localized prostate cancer. It is a highly technological and Minimally Invasive Surgery. What this means is that the surgical operation is performed through cuts of few centimeters instead of several centimeters as in open surgery. The technique utilized in a MIS allows to reach the area to be treated with instruments such as cameras, lights and scalpels, which pass through holes of a few centimeters, minimizing the trauma of the operation. These differences are shown in the figure below.

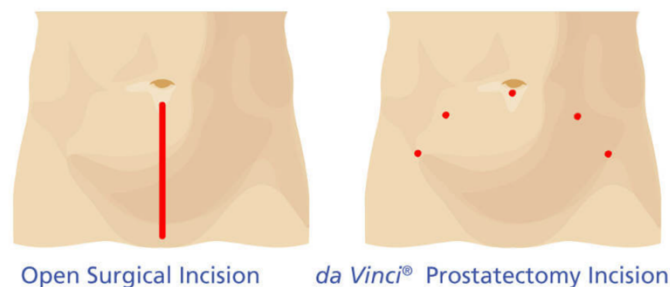


Figure 38: Differences in incision techniques in prostatectomy

The robot-assisted technique has lower mental and physical impact on patients and a significant reduction in complications of the surgical wound, resulting in a reduction in postoperative hospitalization and reduced costs. Traditionally the radical prostatectomy leads to urinary incontinence and impotence consequences whilst with the robot assistance it's possible to use retro-bladder access and lower greatly the bad consequences of the operation. The technology used is called *DaVinci* System. It is famous for its great versatility of movements thanks to the *EndoWrist* technology that it incorporates. The latter provides up to 7 degrees of freedom in forearm and wrist movements. At the same time, it reduces the tremor of surgeon's hands or involuntary movements that could affect the performance. *DaVinci* is also recognized for its magnified 3D high-definition system which overcomes the limitations of traditional laparoscopy surgery such as the flat, two-dimensional vision. Other improvements have been made with respect to unnatural surgeon positions or

dissociation between instrument control and vision proper of traditional surgery. A highly precision, control, quality of vision is obtained with the *DaVinci* utilization. The robotic setting is presented in the figure below.

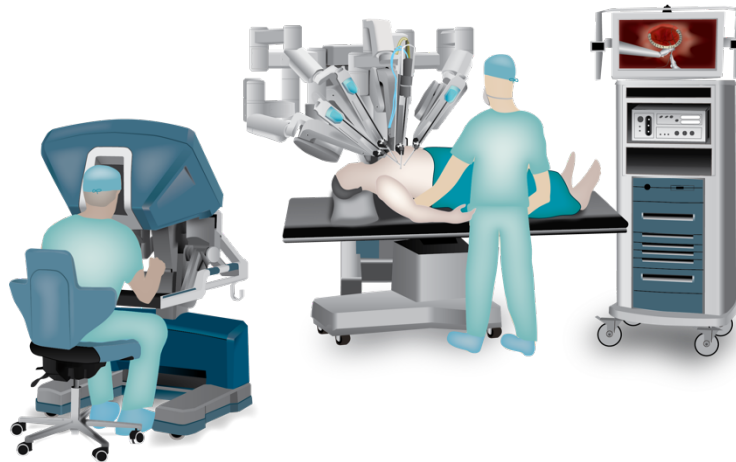


Figure 39: The three components of DaVinci robot: the console, the robotic arm cart and the cart view.

Starting from the left side we have the console, the robotic arm cart and the cart view. The surgeon sits at the console, resting his head on the viewer which allows him to see inside the patient's body in 3D. The images come from a couple of mini cameras inserted into the patient's body. The system activates and is ready to receive orders only when the surgeon places his head on the console correctly. The robot does not operate autonomously but translates and transmits the movements of the surgeon's hands (sitting at the console) to the instruments that move inside the patient's body. The surgical procedure of BA-RARP is composed of 5 steps: anesthesia, room and robot preparation, patient preparation, surgery execution and the end phase.

- The first step concerns all those tests needed by the anesthetist to validate the response condition to anesthesia of the patient. The process varies according to hospitals, health condition of patients and their ages. The anesthesia is general with total pain control. Therefore, it requires blood tests, urine tests, electrocardiogram and all the other tests on the check list

- The preparation of the room and robot in a highly technologic operation like this, is crucial. It consists of instrument cart preparation and *DaVinci* connection. The console, the robotic arm cart and the cart view should be connected together, and the robot's arms should be dressed. Optical instruments have to be prepared
- The third step concerns the patient preparation. This phase includes the pre-anesthetization of the patient which has been positioned on the surgical bed, in supine position. The abdominal part is the only part not covered by sterile towels with the 6 strategic point drawn for the robotic doors.
- The surgery execution of BA-RARP consists of the complete removal of prostate. In some more severe situations, there is the possibility that even the surrounding lymph nodes are removed. The procedure is entirely performed through the utilization of the robot which characterizes its high dependence on technological instruments. The process starts with the patient's skin and subcutaneous tissue incision, the positioning of the robot arms and their fastening. The subsequent phases are summarized as:
  - Peritoneum engraving and isolation of seminal vesicles;
  - Suspension of peritoneum;
  - Isolation of posterior surface of prostate and lateral pedicles;
  - Isolation of bladder neck;
  - Isolation of anterior surface of prostate and apex;
  - Removal of prostate;
  - Anastomosis;
  - Drain positioning.



- The final phase of the laparoscopic prostatectomy is the removal of the *DaVinci* components that for all the course of the surgery where inside the patient's body. It is performed on average after one hour and a half after the start of the surgery.

### 5.3.2 Ex-ante feature calibration

The ex-ante feature calibration was intended again to involve the whole surgical team mentioned above. They were to be asked to evaluate the same three tasks on the fatigue impact and task complexity scales. The tasks in question are the three critical tasks of the BA-RARP procedure identified thanks to expert judgement and medical evidence (Onofrio & Trucco, 2019):

- 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

Clearly it was not possible to obtain the desired feedback and comment on the overlapping or separation of data on the plane. At the same time for both fatigue impact and task complexity a very limited amount of data was collected which does not allow conclusions to be drawn but does concede reflections to be made.

### 5.3.2.1 Fatigue impact scale

In calibrating the fatigue impact scale, data was gathered from the urologist on the three critical tasks of prostatectomy (Table 10).

	Task 1	Task 2	Task 3
Convergent		-1	-1
Divergent	+1		
Rapidly changing information			
Many variables to take into consideration			
May require several changes of plan of action	+1		+1
Requires lateral thinking		+1	+1
Requires being innovative		+1	+1
Requires taking risks	+1		
Requires high level of communication	+1		
Call for temporal memory			
Lengthy with respect to the whole operation	+1		+1
Done in the second half of the operation		+1	+1
	<b>5</b>	<b>2</b>	<b>4</b>
	<b>MEDIUM</b>	<b>LOW</b>	<b>MEDIUM</b>

Table 10: Fatigue impact scale dimensions

Task 1 was defined as divergent by the doctor, which added 1 to the total of 4 selected elements. With an overall score of, 5 the isolation of lateral peduncles and of posterior prostate surface resulted as the task with the highest fatigue impact placing it in the medium area. Task 2 gained a much lower score positioning itself in the low section with a total of 2, making it the task with the smallest impact of fatigue. Lastly the

anastomosis task with a score of 4 went to locate itself again in the medium section just below Task 1.

From these results, Task 2 being low, lies within the “green zone” which does not require the subsequent real-time analysis. Task 1 and 3 though may be both belonging to the “green zone” or “orange zone” according to the results coming from the task complexity scale.

#### *5.3.2.2 Task complexity scale*

The same approach had with the fatigue impact scale was not able to be applied to the task complexity one due to the lack of responses. Submitting also the task complexity questionnaire would have given precise indication on where the tasks would have located on the FI-TC model. The extension of the questionnaire to the other members of the surgical teams would have given the possibility to study overlapping or dispersion of the data, making measures of reliability of the scale feasible. For this reason, scenarios were hypothesized in the following paragraph together with possible measures and remedies for reliability.

#### *5.3.3 Scenarios hypotheses*

Scenarios hypotheses were linked to the application study within the robotic context. This was done in order to understand the extent to which the model is context, task or surgeon specific. The results and observations presented below confirm the context-centric feature of the FI-TC model which enhances the necessity of taking out targeted studies on specific tasks in predetermined contexts. After the scenarios are presented measures of reliability and remedies for both scales are discussed.

There are four different types of results that could emerge from the gathered data based on how the answers are spread on the Cartesian plane of the model. The four possible outcomes are discussed below:

- Scenario 1 – coherence and overlapping of results  
If most of the data results grouped in a single area of the Cartesian plane it is said to converge to the same results as it can be seen in Figure 38.

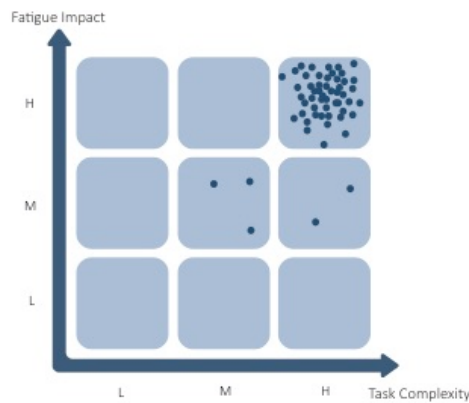


Figure 40: Scenario hypothesis in the robotic application 1 – coherence and overlapping of results

- Scenario 2 – coherence and overlapping of fatigue impact results  
If most of the data is to follow a horizontal development on the plane, then it is said to converge to the same fatigue impact level.

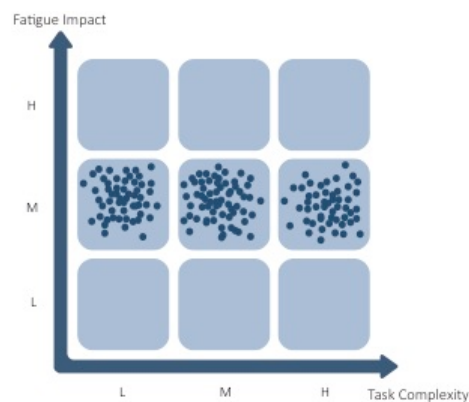


Figure 41: Scenario hypothesis in the robotic application 2 – coherence and overlapping of Fatigue Impact results

- Scenario 3 – coherence and overlapping of task complexity results  
If most of the data is to follow a vertical development on the plane, then it is said to converge to the same task complexity level.

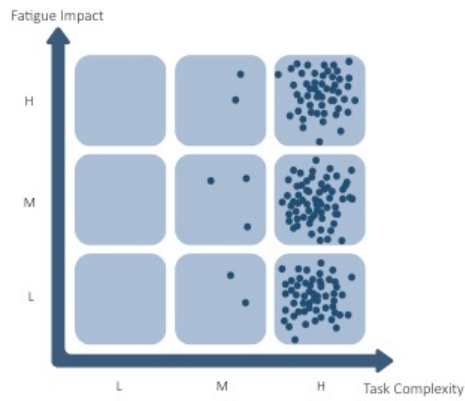


Figure 42: Scenario hypothesis in the robotic application 3 – coherence and overlapping of Task Complexity results

- Scenario 4 – dispersion of results  
If data results spread throughout all the plane, not following any logic then it is said to diverge.

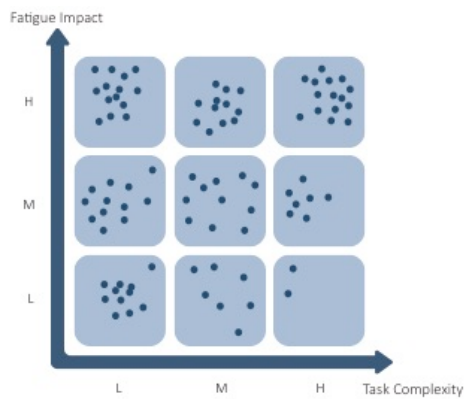


Figure 43: Scenario hypothesis in the robotic application 4 - dispersion of results

#### *5.3.3.1 Measure of reliability of the task complexity scale*

Once the overlapping or dispersion of results has been identified a further analysis is needed to evaluate the reliability of the interested scale.

If Scenarios 2 or 4 were to occur, the task complexity scale would result in having a very weak, almost null reliability and would need total reframing. Total reframing could consist in a) change in terms used in the questionnaire or b) structural change of the questionnaire, if the feedback from the rest of the surgical team on scale validity were to be negative. These two solutions would go to improve readability and understandability of the questionnaire. If validity responses though were to be positive a third solution in this case would be a complete change of questionnaire, not using the Functional Job Analysis one anymore. These three options do not cover all possibilities but aim to give suggestions.

If Scenarios 1 or 3 were to occur, the task complexity scale would result in having a very strong reliability. Gathered data merges towards a same area of the graph with regards to task complexity meaning the scale used reflects the analyzed situation. Stronger or weaker reliability depends on how dispersed data is on the plane. If data is highly concentrated in a restricted area, task complexity scale will be strongly reliable. If data is characterized by quite a few dispersed points, then the scale will be less reliable. Increasing reliability of the scale could be done through the adjustment of the weights given to its items in the validation section.

#### *5.3.3.2 Measure of reliability of the Fatigue impact scale*

The same concepts applied to the task complexity scale were applied to the fatigue impact one.

If Scenarios 3 or 4 were to occur, the fatigue impact scale would be said to have a very weak or null reliability. This could mean that, as with the TC scale, a total reframing could be needed. Reframing could be a) a change in terminology used in the

questionnaire b) structural change of the questionnaire – maybe eliminating the first two questions of convergent/divergent – if validity responses from the rest of the components of the surgical team were to be negative. Contrarily, option c) complete change of questionnaire, could be the solution. The need of a total change of questionnaire could be due to Prefrontal Cortex not reflecting fatigue impact correctly in the healthcare sector.

If Scenarios 1 or 2 were to occur, the fatigue impact scale would result in having a quite strong reliability. As for the TC scale, this reliability would depend on how spread out the data is on the plane. Data highly concentrated in a small area means high reliability and confirms the appropriateness of the scale. A lower reliability due to more spread-out data means that there are areas for improvement. A possible improvement could include adding items to the FI questionnaire to the 10 already existing. This could be done both by splitting existing items in more detailed and specific ones or adding new bullet points to cover aspects the current items aren't able to convey.

#### 5.4 Real-time feature assessment

The fatigue level evaluation framework is the real-time tool used to evaluate the fatigue level of operators. The fatigue-evaluation topic is well developed by an extensive literature, in which it was possible to find numerous articles that developed, analyzed and validated tools for fatigue evaluation. Essentially, two are the categories into which these tools can be classified: subjective tools and objective tools. The subjective tool structure consists mostly of self-administered questionnaires. Different approaches are used in the questions' structuring and in their evaluation to obtain the final score. The number of questions varies, and grades are assigned in different ways. In developing the FI-TC model and the subsequent fatigue level evaluation framework, three were the subjective questionnaires considered: the Epworth Sleepiness Scale, the Modified Fatigue Impact Scale and the Pittsburgh Sleep

Quality Index. The choice made was based on the literature content analysis and on the reliability of the questionnaire. The three questionnaires in question have proved to be the most used and reliable. For this reason, they have all been proposed as subjective tools to possibly be chosen for the fatigue level evaluation.

It has to be considered that the focus of the thesis is on healthcare and surgeons' operativity. In view of the lack of time that surgeons have and the impossibility to perform all three subjective questionnaires, a phase of questionnaires' assessment is necessary. Through experts' feedbacks, the three pre-selected questionnaires are evaluated. The aim is to find out which of the three questionnaire is the most appropriate and easy to use. In order to gather the opinions of experts on the qualities of the subjective tools and collect the judgements, a new questionnaire is needed (Figure 42). The new questionnaire includes also the three subjective metrics which are not shown in the figure to avoid redundancy. The questionnaire is formed by six numbered questions and one question, on the bottom of the sheet, that aims to extract one overall preference.



Could we please request you take 10 minutes to complete the three fatigue evaluating questionnaires.

After having completed all three questionnaires, answer the questions below. In each question you must choose only one out of the three subjective questionnaires:

[ESS](#) for the Epworth Sleepiness Scale,

[MFIS](#) for the Modified Fatigue Impact Scale,

[PSQI](#) for the Pittsburgh Sleep Quality Index.

**SURGICAL SPECIALTY:** \_\_\_\_\_

\_\_\_\_\_

1. It has the most fully understandable questions

\_\_\_\_\_

2. It is the most accurate

\_\_\_\_\_

3. It has the shortest completion time

\_\_\_\_\_

4. It kept you interested the most throughout its completion

\_\_\_\_\_

5. It is the easiest to fill in

\_\_\_\_\_

6. It is the most complete

\_\_\_\_\_

Overall if you had to pick only one of the subjective questionnaires based on the 6 points above you would choose: \_\_\_\_\_

Why?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Figure 44: Questionnaire for subjective fatigue assessment tools evaluation

The six questions were formulated on the basis of the characteristics that define the qualities of a good questionnaire (Crawford, 1997). A good questionnaire should:

- Foster participants to answer with precise, complete and impartial information
- Stimulate the attention of the participants throughout the questionnaire and be brief
- Make it easy for those who fill in the questionnaire to give useful answers and necessary information
- Meet research objectives.

The first bullet point gave rise to numbered questions 1,2 and 3 of the questionnaire. The second bullet point gave rise to the fourth, and so on. Therefore, each of the six questions defines one or more of the characteristics that a well-structured questionnaire should have. The last question, instead, asks the respondent which of the three questionnaires he would choose if he had to pick only one of them.

When all the questionnaires have been completed by the sample of surgeons, the answers should be analyzed in order to define the final result: which of the three subjective questionnaire is the most appropriate. The table below summarizes all the answers and shows the results of the test. The left column indicates the characteristics that each question outlines and the right one indicates the most mentioned test. The last row shows which of the questionnaires is the most quoted by surgeons.

Characteristic	Result
Most understandable	--
Most accurate	--
Shortest completion time	--
Most interesting	--
Easiest	--
The most complete	--
If you had to pick only one	--

Table 11: Results of the questionnaire on subjective tool quality

The results give the possibility to select a different questionnaire based on a specific feature. For example, if the feature sought in a questionnaire is completeness or speed of completion, row three and six are to be considered. It would be also interesting to understand if the context-specific and task-specific characteristics of the model are present also in the selection of the subjective questionnaires dividing results by context or task.

## 6. Discussion

---

This chapter aims to discuss the main aspects and points touched by the Fatigue Impact – Task Complexity model. The construction process will be reviewed and topics regarding HRA, standardization and the application study in the BA-RARP procedure will be covered. Lastly, the chapter will end with the discussion of the limitations of the study.

---

The idea of the development of the Fatigue Impact – Task Complexity model, stemmed from an in-depth analysis of fatigue. Taking a step backwards and re-analyzing this process could be very beneficial to understanding the features, reasoning and limitations which stand at the base of it.

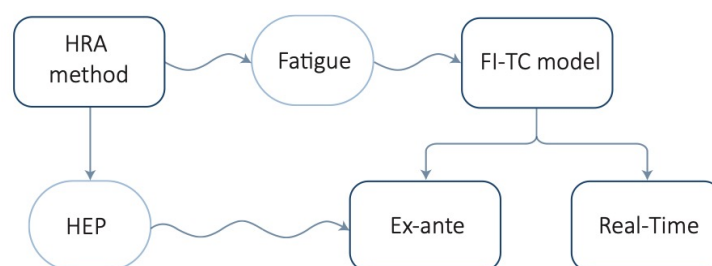


Figure 45: The development process of FI-TC model creation

The graph above shows the process behind the construction the model and can be outlined in three main points:

- Step 1 – extraction of fatigue from HRA method
 

Onofrio & Trucco, 2018 identified 21 Influencing Factors in surgery of which fatigue was included as number 15. As with all Human Reliability Analysis, the IFs which are believed to impact on the situation are chosen, leaving out those which are regarded as not relevant to HEP calculation. The modified HEART methodology (Onofrio & Trucco, 2019) applied to the three critical tasks of BA-RARP procedure also used in the present study, gave surgeons the possibility to select the IFs. IF number 15 – fatigue – was not selected for any of the three tasks meaning it wasn't deemed to impact on possible error making. This case in particular gave rise to the idea of creating the FI-TC model. By extraction of fatigue from the HRA method it is intended its deletion from the 21 existing IFs – which would then become 20.
- Step 2 – inclusion of fatigue in FI-TC model
 

This was believed to be the most appropriate way to deal with fatigue for the need to find a method that includes it mandatorily. After being extracted from the HRA

methods and deleted as IF, fatigue becomes an input of the FI-TC model. Fatigue finds its place in both features of the model: a) within the ex-ante feature where its impact on tasks is identified and b) in the real-time feature where fatigue level of operators is measured.

- Step 3 – inclusion of HEP in ex-ante feature of the model

Lastly, HEP which is the output of the HRA method used, becomes also the input of the FI-TC model in the ex-ante feature. It figures as a bubble with dimensions according to its value. HEP has the role of activating the real-time analysis in those areas named “orange zone” and “red zone”.

These three steps confirm the importance of Human Reliability Analysis in healthcare by highlighting its inclusion from the start to the end of the process. Fatigue is extracted from the HRA method, HEP is the output of the HRA method. HRA was of vital importance in the development of the FI-TC which cannot function without it. For this reason, the selection of which HRA technique to use was left as a free choice. In this way, each context will choose the technique it believes to adapt better to its characteristics based on the features it values the most. No obligations are imposed which could make people relentless in adopting the model. This freedom also gives the possibility to use already calculated HEPs from other studies without having to redo it fully.

Another interesting aspect to analyze recalls the modified HEART technique in robotic surgery (Onofrio & Trucco, 2019) mentioned above. As already stated, fatigue was not chosen as relevant IF for none of the critical tasks evaluated. When submitting the fatigue impact questionnaire to the urologist though, Tasks 1 and 2 resulted in having a medium fatigue impact. Regardless the activation or not of the real-time analysis, a medium fatigue impact level implies that fatigue does have a relevant effect on taking out those tasks, in contradiction with the IF selection. This is believed to be because when presented with a list of factors which could possibly affect performance, fatigue tends to be underestimated and overlooked as its effects aren't tangible and evident.

Referring again back to the study by Onofrio & Trucco, 2018 on identifying the 21 IFs relevant to surgery, another point is worth discussing. In this study IF 1 noise and ambient talk and IF 4 verbal interruption were seen to have different results when in an open surgery setting compared to a Mini-Invasive Surgery setting.

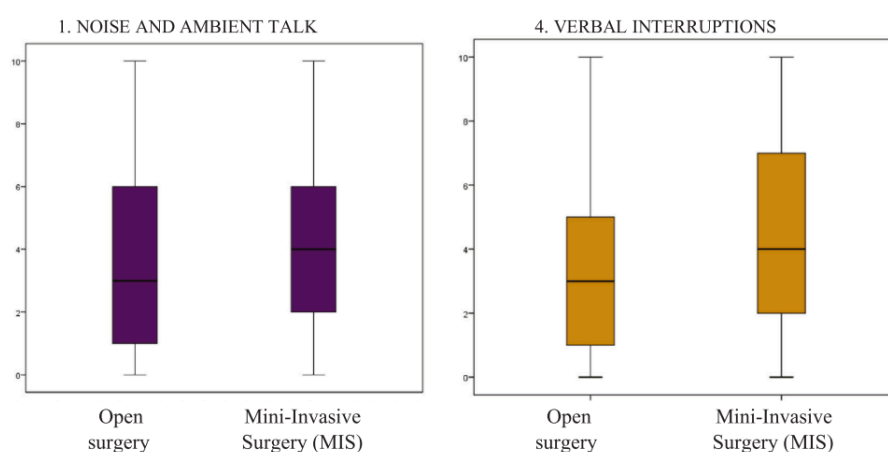


Figure 46: Comparative results of IF perception in two different surgical context (Onofrio & Trucco, 2018).

The results show how surgeons operating in MIS have the impression to be more exposed and sensitive to various sources of distraction. This was interpreted and conceptualized as possibly due to the higher cognitive workload and team coordination needed in MIS compared to open surgery setting.

The distinction between open and MIS setting was maintained in the application study of the FI-TC model. If cognitive workload is deemed higher in MIS, then also the impact of fatigue will be as these two factors are highly correlated. This choice set the base of the model being strictly context and task-centric. The BA-RARP procedure was chosen mainly due to its surgical setting being one of the most complex nowadays. Its high reliance on technology, complemented with an extremely elevated dependence on human resources and their interaction makes it one of the most avant-garde operations in existence. This gives the application study a very strong rationale and

provides an incentive for the FI-TC model to be replicated in other contexts for other operations.

As briefly mentioned above, standardization and task-oriented measures are the key for the model functioning at its best. Standardizing tasks translates in lower HEPs thanks to the identification of best practices and the provision of benchmarks according to which corrective actions can be applied at the right time. Also, an easier and more effective way of educating and training is made possible through a provision of prescribed detailed steps (Armstrong et al., 2012). Not only does high standardization mean a lower probability of human errors but it induces new ways of lowering task complexity. Splitting a task into smaller tasks – microtasks – implies a shift to the left on the task complexity axis. This shift is seen as a structural transition made possible through the resizing of the task. The simplification of a task could create a domino-effect action by causing a lowering of its fatigue impact level and a reduction of its Human Error Probability bringing to a transition towards the safest zone – the “green zone” – of the FI-TC model.

Overall though, this thesis presents some limitations. Firstly, the study is intended only for specific applications. This could be seen as a positive aspect as it implies precision and highly calibrated measures, but it also requires initiation in every context that varies even slightly. Secondly, the data gathering process was not able to be completed given the current health emergency circumstances. Data is missing and only assumptions and hypotheses could be made. Lastly, the Fatigue Impact – Task Complexity model can appear as time consuming. Initiation is needed, an HRA method has to be taken out, questionnaires for fatigue impact and task complexity require completion, the subjective fatigue evaluating metric has to be chosen and the real time analysis needs to be implemented.



## 7. Conclusion

---

The final chapter reviews the gaps fulfilled by the developed model. It moves on suggesting how further studies could be structured and which direction they could take. Lastly, final comments about the dissertation are made.

---

The Fatigue Impact -Task Complexity model goes to fulfill the gap of poor integration of fatigue in existing Human Reliability Analysis techniques in healthcare. It does so by taking into consideration all aspects and characteristics of fatigue which stem from its multifactorial nature. When speaking of poor integration, it cannot be seen as a per se statement but must be related to a very crucial point: the gravity of possible consequences of fatigued individuals in the context in question. Sleep deprived individuals in a high-risk context such as healthcare – specifically surgeons in the surgery context – put at serious risk human lives, reason for which fatigue requires a 360° consideration. Not only must fatigue be considered in all its features but it must have a central role in the organization of human resources within the hospital. This is exactly what occurs in the FI-TC:

- Fatigue has a central role as it is considered both in the ex-ante analysis through the y-axis of fatigue impact and in the real-time analysis of the fatigue level evaluation framework
- Fatigue has been fully analyzed and all features find their place in different aspects of the model such as its multifactorial nature in the fatigue level evaluation framework, its subjectivity in the selection of a range of subjective questionnaires and its proved strong impact on performance in the mandatory inclusion and proactive nature of the model.

In the previous chapter the limitations that characterize the present work were outlined. These limitations gave rise to two directions for possible future developments and studies. This model exists as an application to very specific contexts. In this thesis the BA-RARP procedure has been described and used as a simulation study but a broadening of the model's boundaries towards other procedures is not only worthwhile but necessary. Slowly but steadily, its application could be extended to other roles, an example of which could be the role of anesthetists at first. Secondly, an extension to nurses could be done being them at the center of debates on over working, long shift hours and sleep deprivation. Enlarging the calibration of the method to more procedures and more hospitals –

ideally to all of Italy – could spring the ignition of the second direction future studies could take. This direction involves the setting-up of a database containing the data necessary to promptly initiate the model in any context. In other words, the widening of the calibration to the whole state would bring to a very large amount of gathered data. HEP values and fatigue impact and task complexity levels would be defined and all tasks of all procedures and operations could be mapped. This would improve performances, effectiveness, efficiency and make the model less time consuming and immediately usable. Lastly, another possible direction was identified which did not stem from the limitations but from the content analysis of the literature review. Wearable devices in fatigue monitoring are becoming of increasing importance in today's world. A future of the model could include the real-time fatigue level measurement through a wearable device worn by surgeons throughout the day. This would go to reduce the duration of the real-time analysis and provide users with constantly updated levels of fatigue.

These proposed future steps to improve the model's performance go to enhance its applicability and usability. The Fatigue Impact – Task Complexity model as it is, lays the foundations for a more controlled workplace within hospitals. The intangibility and alterable nature of fatigue coupled with the tendency of it to be underestimated, triggers the need of careful surveillance. More monitored human resources mean less errors which brings to a safer environment in which patients feel shielded. High quality hospital services are safe, effective and people-centered. Upgrading the quality of services provided in hospitals has the power to improve overall trust that patients have in health institutions. ("Hospitals Management and Quality," 2018)

## Bibliography

- Adriana Patricia Muñoz Zapata, L. C. C. (2011). FRMS Implementation Guide for Operators July 2011. *Phys. Rev. E*, (July), 24. Retrieved from [http://ridum.umanizales.edu.co:8080/jspui/bitstream/6789/377/4/Muñoz\\_Zapata\\_Adriana\\_Patricia\\_Artículo\\_2011.pdf](http://ridum.umanizales.edu.co:8080/jspui/bitstream/6789/377/4/Muñoz_Zapata_Adriana_Patricia_Artículo_2011.pdf)
- Amirian, I. (2014). The impact of sleep deprivation on surgeons' performance during night shifts. *Danish Medical Journal*, 61(9), 1–14.
- Armstrong, T., Yu, D., Frischknecht, A., Minter, R., Andreatta, P., & Kasten, S. (2012). Standardization of surgical procedures for identifying best practices and training. *Work*, 41(SUPPL.1), 4673–4679. <https://doi.org/10.3233/WOR-2012-0108-4673>
- Barker, L. M., & Nussbaum, M. A. (2011). Fatigue, performance and the work environment: A survey of registered nurses. *Journal of Advanced Nursing*, 67(6), 1370–1382. <https://doi.org/10.1111/j.1365-2648.2010.05597.x>
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of PVT to Sleep Loss. *Sleep*, 34(5), 581–591.
- Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, 69(11–12), 949–959. <https://doi.org/10.1016/j.actaastro.2011.07.015>
- Basner, M., & Rubinstein, J. (2011). Fitness for duty: A 3-minute version of the psychomotor vigilance test predicts fatigue-related declines in luggage-screening performance. *Journal of Occupational and Environmental Medicine*, 53(10), 1146–1154. <https://doi.org/10.1097/JOM.0b013e31822b8356>
- Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R., & Kupfer, D. J. (1998). The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Psychiatry Research*, Vol. 28, pp. 193–213.
- Cagliano, A. C., Grimaldi, S., & Rafele, C. (2011). A systemic methodology for risk management in healthcare sector. *Safety Science*, 49(5), 695–708. <https://doi.org/10.1016/j.ssci.2011.01.006>

- Caruso, C. C. (2014). Negative impacts of shiftwork and long work hours. *Rehabilitation Nursing, 39*(1), 16–25. <https://doi.org/10.1002/rnj.107>
- Casali, G., Cullen, W., & Lock, G. (2019). The rise of human factors: Optimising performance of individuals and teams to improve patients' outcomes. *Journal of Thoracic Disease, 11*(Suppl 7), S998–S1008. <https://doi.org/10.21037/jtd.2019.03.50>
- Cohen, D. A., Wang, W., Wyatt, J. K., Kronauer, R. E., Dijk, D.-J., Czeisler, C. A., & Klerman, E. B. (2010). Uncovering Residual Effects of Chronic Sleep Loss on Human Performance. *Sci Transl Med., 46*(4), 564–574. <https://doi.org/10.1016/j.cortex.2009.08.003>. Predictive
- Contino, G. (2018). Schermata 2015-10-12 alle 14.
- Crawford, I. (1997). Chapter 4: Questionnaire Design. In *Marketing Research and Information Systems*. Retrieved from [http://www.fao.org/docrep/w3241e/w3241e05.htm#choose the method\(s\) of reaching target respondents](http://www.fao.org/docrep/w3241e/w3241e05.htm#choose%20the%20method(s)%20of%20reaching%20target%20respondents)
- Cropley, A. J. (2006). Intervention in Creativity With Children Aged 10 and 11 Years : Impact of a Play Program on Verbal and Graphic – Figural Creativity. *Creativity Research Journal, 0419*(January), 37–41. <https://doi.org/10.1207/s15326934crj1803>
- Dall'Ora, C., Ball, J., Recio-Saucedo, A., & Griffiths, P. (2016). Characteristics of shift work and their impact on employee performance and wellbeing: A literature review. *International Journal of Nursing Studies, 57*, 12–27. <https://doi.org/10.1016/j.ijnurstu.2016.01.007>
- Dean, G. E., Scott, L. D., & Rogers, A. E. (2006). Infants at Risk: When Nurse Fatigue Jeopardizes Quality Care. *Advances in Neonatal Care, 6*(3), 120–126. <https://doi.org/10.1016/j.adnc.2006.02.001>
- Ferrucci, M. (2018). Ocse : in Italia la spesa sanitaria è tra le più basse in Europa. *IlManifesto, 1*. Retrieved from <https://ilmanifesto.it/ocse-in-italia-spesa-sanitaria-e-tra-le-piu-basse-in-europa/>
- Fine, S. A. (2014). Benchmark Tasks for Job Analysis. *Benchmark Tasks for Job Analysis*. <https://doi.org/10.4324/9781315806204>

- Fine, S. A., & Wiley, W. W. (1971). An Introduction to Functional Job Analysis; A Scaling of Selected Tasks from The Social Welfare Field. *Methodes for Manpower Analysis, No. 4*, (3), 94.
- Fitzgerald, J. E., & Caesar, B. C. (2012). The European Working Time Directive: a practical review for surgical trainees. *International Journal of Surgery (London, England)*, *10*(8), 399–403. <https://doi.org/10.1016/j.ijssu.2012.08.007>
- Flachenecker, P., Kümpfel, T., Kallmann, B., Gottschalk, M., Grauer, O., Rieckmann, P., ... Toyka, K. V. (2002). Fatigue in multiple sclerosis: A comparison of different rating scales and correlation to clinical parameters. *Multiple Sclerosis*, *8*(6), 523–526. <https://doi.org/10.1191/1352458502ms839oa>
- Gaba, D. M., & Howard, S. K. (2002). Fatigue among clinicians and the safety of patients. *New England Journal of Medicine*, *347*(16), 1249–1255. <https://doi.org/10.1056/NEJMsa020846>
- Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis and Prevention*, *43*(2), 573–590. <https://doi.org/10.1016/j.aap.2009.11.007>
- Gander, P., Millar, M., Webster, C., & Merry, A. (2008). Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiology International*, *25*(6), 1077–1091. <https://doi.org/10.1080/07420520802551428>
- Ganesan, S., Magee, M., Stone, J. E., Mulhall, M. D., Collins, A., Howard, M. E., ... Sletten, T. L. (2019). The Impact of Shift Work on Sleep, Alertness and Performance in Healthcare Workers. *Scientific Reports*, *9*(1), 1–13. <https://doi.org/10.1038/s41598-019-40914-x>
- Givi, Z. S., & Jaber, M. Y. (2014). Human error due to learning and fatigue. *IIE Annual Conference and Expo 2014*, 1933–1942.
- Gnerre, P., Montemurro, D., Rossi, A. P., Troise, C., Palermo, C., Amati, D., ... Ragazzo, F. (2017). Lack of application of the European Work Time Directive: Effects on workload, work satisfaction and burnout among Italian physicians. *Italian Journal of Medicine*, *11*(2), 159–163. <https://doi.org/10.4081/itjm.2017.714>

- Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science*, 119, 155–190. <https://doi.org/10.1016/B978-0-12-396971-2.00007-5>
- Gore B.F., Casner S.M., Wickens C.D., Pallesen S., B. B. (2018). Workload and Fatigue. In *Space Safety and Human Performance*.
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., & Van Dongen, H. P. A. (2017). 3-Minute Smartphone-Based and Tablet-Based Psychomotor Vigilance Tests for the Assessment of Reduced Alertness Due To Sleep Deprivation. *Behavior Research Methods*, 49(3), 1020–1029. <https://doi.org/10.3758/s13428-016-0763-8>
- Griffith, C. D., & Mahadevan, S. (2011). Inclusion of fatigue effects in human reliability analysis. *Reliability Engineering and System Safety*, 96(11), 1437–1447. <https://doi.org/10.1016/j.res.2011.06.005>
- Griffith, C. D., & Mahadevan, S. (2015). Human reliability under sleep deprivation: Derivation of performance shaping factor multipliers from empirical data. *Reliability Engineering and System Safety*, 144, 23–34. <https://doi.org/10.1016/j.res.2015.05.004>
- Gunzelmann, G., Gross, J. B., Gluck, K. A., & Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science*, 33(5), 880–910. <https://doi.org/10.1111/j.1551-6709.2009.01032.x>
- Harrison, Y., & Horne, J. A. (2000a). Sleep loss and temporal memory. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 53(1), 271–279. <https://doi.org/10.1080/713755870>
- Harrison, Y., & Horne, J. A. (2000b). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, 6(3), 236–249. <https://doi.org/10.1037/1076-898X.6.3.236>
- Hollnagel, E. (1993). *Human reliability analysis: context and control*.
- Hospitals Management and Quality. (2018). Retrieved from Who website: <https://www.who.int/hospitals/management-and-quality/en/>

- Hudson, P. (2003). Applying the lessons of high risk industries to health care. *Quality and Safety in Health Care*, 12(SUPPL. 1), 7–12. [https://doi.org/10.1136/qhc.12.suppl\\_1.i7](https://doi.org/10.1136/qhc.12.suppl_1.i7)
- Huizinga, C. R. H., de Kam, M. L., Stockmann, H. B. A. C., van Gerven, J. M. A., Cohen, A. F., & van der Bogt, K. E. A. (2018). Evaluating Fitness to Perform in Surgical Residents after Night Shifts and Alcohol Intoxication: The development of a “Fit-to-Perform” test. *Journal of Surgical Education*, 75(4), 968–977. <https://doi.org/10.1016/j.jsurg.2018.01.010>
- Hysong, S. J., Amspoker, A. B., & Petersen, L. A. (2016). A Novel Method for Assessing Task Complexity in Outpatient Clinical-Performance Measures. *Journal of General Internal Medicine*, 31, 28–35. <https://doi.org/10.1007/s11606-015-3568-z>
- Jaber, M. Y., Givi, Z. S., & Neumann, W. P. (2013). Incorporating human fatigue and recovery into the learning-forgetting process. *Applied Mathematical Modelling*, 37(12–13), 7287–7299. <https://doi.org/10.1016/j.apm.2013.02.028>
- Ji, Q., Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 36(5), 862–875. <https://doi.org/10.1109/TSMCA.2005.855922>
- Johns, M. (1997). Epworth Sleepiness Scale official website. Retrieved from <https://epworthsleepinessscale.com/about-the-ess/>
- Joice, P., Hanna, G. B., & Cuschieri, A. (1998). Errors enacted during endoscopic surgery - A human reliability analysis. *Applied Ergonomics*, 29(6), 409–414. [https://doi.org/10.1016/S0003-6870\(98\)00016-7](https://doi.org/10.1016/S0003-6870(98)00016-7)
- Kartsch, V. J., Benatti, S., Schiavone, P. D., Rossi, D., & Benini, L. (2018). A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems. *Information Fusion*, 43(January 2017), 66–76. <https://doi.org/10.1016/j.inffus.2017.11.005>
- Killgore, W. D. S., Balkin, T. J., & Wesensten, N. J. (2006). Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research*, 15(1), 7–13. <https://doi.org/10.1111/j.1365-2869.2006.00487.x>



- Klein, G. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In *Decision Making in Action: Models and Methods*. <https://doi.org/10.4324/9780429440861-6>
- Knol, A., Sluijs, J. P. Van Der, & Slottje, P. (2008). Expert Elicitation : Methodological suggestions for its use in environmental health impact assessments. *RIVM Letter Report 630004001/2008*, 56. Retrieved from [http://www.nusap.net/downloads/reports/Expert\\_Elicitation.pdf](http://www.nusap.net/downloads/reports/Expert_Elicitation.pdf)
- Larson, R. D. (2013). Psychometric properties of the modified fatigue impact scale. *International Journal of MS Care*, 15(1), 15–20. <https://doi.org/10.7224/1537-2073.2012-019>
- Li, H., Wang, D., Chen, J., Luo, X., Li, J., & Xing, X. (2019). Pre-service fatigue screening for construction workers through wearable EEG-based signal spectral analysis. *Automation in Construction*, 106(December 2018), 102851. <https://doi.org/10.1016/j.autcon.2019.102851>
- Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction*, 109(October 2019), 103000. <https://doi.org/10.1016/j.autcon.2019.103000>
- Lupino, F. (2016). Viterbo, turni massacranti: 18 chirurghi denunciano la Asl: «A rischio le operazioni sui pazienti». *IlMattino*.
- Lyons, M. (2009). Towards a framework to select techniques for error prediction: Supporting novice users in the healthcare sector. *Applied Ergonomics*, 40(3), 379–395. <https://doi.org/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 and Safety in Medicine*, 16(4), 223–237.
- Makary, M., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *Bmj.Com*, 2016–2018. Retrieved from <https://www.bmj.com/content/353/bmj.i2139.abstract%0Ahttp://www.fiercehealthcare.com/story/hospital-medical-errors-third-leading-cause-death-dispute-to-err-is-human-report/2013-09-20>

- Mehta, R. K., Peres, S. C., Steege, L. M., Potvin, J. R., Wahl, M., Stanley, L. M., & Nesthus, T. E. (2016). Fatigue monitoring and management across different industries. *Proceedings of the Human Factors and Ergonomics Society*, 992–995. <https://doi.org/10.1177/1541931213601230>
- Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24(1), 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Modic, M. T., Lasalvia, L., & Merges, R. (2019). *Standardization and personalization: lessons from other industries*. (6), 16.
- Mohtashami, F., Thiele, A., Karreman, E., & Thiel, J. (2014). Comparing technical dexterity of Sleep-Deprived versus intoxicated surgeons. *Journal of the Society of Laparoendoscopic Surgeons*, 18(4), 1–7. <https://doi.org/10.4293/JLS.2014.00142>
- Moore, F. I. (1999). Functional job analysis. *World Health Organization*, 9, 43–46. <https://doi.org/10.4324/9781410602497>
- O’Loughlin, E., Smithies, W. J., & Corcoran, T. B. (2010). Out-of-hours surgery - A snapshot in time. *Anaesthesia and Intensive Care*, 38(6), 1059–1063. <https://doi.org/10.1177/0310057x1003800616>
- Oliver, J. (2013). Human reliability, error, and human factors in engineering maintenance. In *Journal of Chemical Information and Modeling* (Vol. 53). <https://doi.org/10.1017/CBO9781107415324.004>
- Onofrio, R., & Trucco, P. (2018). Human Reliability Analysis (HRA) in surgery: Identification and assessment of Influencing Factors. *Safety Science*, 110(August), 110–123. <https://doi.org/10.1016/j.ssci.2018.08.004>
- Onofrio, R., & Trucco, P. (2019). *A Methodology for Dynamic Human Reliability Analysis in Robotic Surgery*.
- Pan, X., Lin, Y., & He, C. (2017). A Review of Cognitive Models in Human Reliability Analysis. *Quality and Reliability Engineering International*, 33(7), 1299–1316. <https://doi.org/10.1002/qre.2111>
- Pan, X., & Wu, Z. (2018). Performance shaping factors in the human error probability modification of human reliability analysis. *International Journal of Occupational Safety and Ergonomics*, 0(0), 1–13. <https://doi.org/10.1080/10803548.2018.1498655>

- Parry, D. A., Oeppen, R. S., Amin, M. S. A., & Brennan, P. A. (2018). Sleep: its importance and the effects of deprivation on surgeons and other healthcare professionals. *British Journal of Oral and Maxillofacial Surgery*, *56*(8), 663–666. <https://doi.org/10.1016/j.bjoms.2018.08.001>
- Pasquale, V. Di, Iannone, R., Miranda, S., & Riemma, S. (2013). An Overview of Human Reliability Analysis Techniques in Manufacturing Operations. *Intech*, 221–240. <https://doi.org/http://dx.doi.org/10.5772/57353>
- Patterson, R. E., Lochtefeld, D., Larson, K. G., & Christensen-Salem, A. (2019). Computational Modeling of the Effects of Sleep Deprivation on the Vigilance Decrement. *Human Factors*. <https://doi.org/10.1177/0018720819829949>
- Peters, G. A., & Peters, B. J. (2007). Medical error and patient safety: Human factors in medicine. *Medical Error and Patient Safety: Human Factors in Medicine*, pp. 1–344. <https://doi.org/10.1108/ijhcqa.2007.06220gae.003>
- Pilcher, J. J., Band, D., Odle-Dusseau, H. N., & Muth, E. R. (2007). Human performance under sustained operations and 'acute sleep deprivation conditions: Toward a model of controlled attention. *Aviation Space and Environmental Medicine*, *78*(5 II).
- Pilcher, J. J., Jennings, K. S., Phillips, G. E., & McCubbin, J. A. (2016). Auditory attention and comprehension during a simulated night shift: Effects of task characteristics. *Human Factors*, *58*(7), 1031–1043. <https://doi.org/10.1177/0018720816654377>
- Posada-Quintero, H. F., Bolkhovsky, J. B., Qin, M., & Chon, K. H. (2018). Human Performance Deterioration Due to Prolonged Wakefulness Can Be Accurately Detected Using Time-Varying Spectral Analysis of Electrodermal Activity. *Human Factors*, *60*(7), 1035–1047. <https://doi.org/10.1177/0018720818781196>
- Rose, M., & Giray, N. (2013). Universal fatigue management strategies. *Sleep Medicine Clinics*, *8*(2), 255–263. <https://doi.org/10.1016/j.jsmc.2013.02.004>
- Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6607–6610. <https://doi.org/10.1109/EMBC.2013.6611070>
- Sanità, gli errori dei medici provocano 90 morti al giorno. (2006). *La Repubblica*.

- Shen, J., Botly, L. C. P., Chung, S. A., Gibbs, A. L., Sabanadzovic, S., & Shapiro, C. M. (2006). Fatigue and shift work. *Journal of Sleep Research*, *15*(1), 1–5. <https://doi.org/10.1111/j.1365-2869.2006.00493.x>
- Sturm, L., Dawson, D., Vaughan, R., Hewett, P., Hill, A. G., Graham, J. C., & Maddern, G. J. (2011). Effects of fatigue on surgeon performance and surgical outcomes: A systematic review. *ANZ Journal of Surgery*, *81*(7–8), 502–509. <https://doi.org/10.1111/j.1445-2197.2010.05642.x>
- Sujan, M. A., Embrey, D., & Huang, H. (2018). On the application of Human Reliability Analysis in healthcare: Opportunities and challenges. *Reliability Engineering and System Safety*, (June), 106189. <https://doi.org/10.1016/j.ress.2018.06.017>
- Swain, D., & Guttman, H. E. (1983). *Handbook of reliability analysis with emphasis on nuclear plant applications*. (August).
- Tang, B., Hanna, G. B., Joice, P., & Cuschieri, A. (2004). Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Archives of Surgery*, *139*(11), 1215–1220. <https://doi.org/10.1001/archsurg.139.11.1215>
- The Lancet. (2010). Doctors' training and the European Working Time Directive. *The Lancet*, *375*(9732), 2121. [https://doi.org/10.1016/S0140-6736\(10\)60977-4](https://doi.org/10.1016/S0140-6736(10)60977-4)
- Valdez, P. (2019a). Circadian rhythms in attention. *Yale Journal of Biology and Medicine*, *92*(1), 81–92.
- Valdez, P. (2019b). Homeostatic and circadian regulation of cognitive performance. *Biological Rhythm Research*, *50*(1), 85–93. <https://doi.org/10.1080/09291016.2018.1491271>
- Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness. *Sleep (Rochester)*, *26*, 117–126. <https://doi.org/10.1093/sleep/26.2.117>
- Vaughn-Cooke, M., Nembhard, H. B., & Ulbrecht, J. (2010). Reformulating human reliability in healthcare systems. *IIE Annual Conference and Expo 2010 Proceedings*.
- Verbano, C., & Turra, F. (2010). A human factors and reliability approach to clinical risk management: Evidence from Italian cases. *Safety Science*, *48*(5), 625–639. <https://doi.org/10.1016/j.ssci.2010.01.014>

- Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: Preliminary retrospective record review. *British Medical Journal*, 322(7285), 517–519. <https://doi.org/10.1136/bmj.322.7285.517>
- Vincent, Charles, Taylor-Adams, S., & Stanhope, N. (1998). Framework for analysing risk and safety in clinical medicine. *Bmj*, 316, 1154–1157. <https://doi.org/10.5694/j.1326-5377.2002.tb04537.x>
- Wiley, & Fine. (1970). A systems approach to new careers. *Differences*.
- Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., & Masters, R. S. W. (2011). Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX). *World Journal of Surgery*, 35(9), 1961–1969. <https://doi.org/10.1007/s00268-011-1141-4>
- Winwood, P. C., Winefield, A. H., Dawson, D., & Lushington, K. (2005). Development and validation of a scale to measure work-related fatigue and recovery: The Occupational Fatigue Exhaustion/Recovery scale (OFER). *Journal of Occupational and Environmental Medicine*, 47(6), 594–606. <https://doi.org/10.1097/01.jom.0000161740.71049.c4>

## Appendix 1: Epworth Sleepiness Scale

### Epworth Sleepiness Scale

Name: \_\_\_\_\_ Today's date: \_\_\_\_\_

Your age (Yrs): \_\_\_\_\_ Your sex (Male = M, Female = F): \_\_\_\_\_

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired?

This refers to your usual way of life in recent times.

Even if you haven't done some of these things recently try to work out how they would have affected you.

Use the following scale to choose the **most appropriate number** for each situation:

- 0 = would **never** doze
- 1 = **slight chance** of dozing
- 2 = **moderate chance** of dozing
- 3 = **high chance** of dozing

*It is important that you answer each question as best you can.*

Situation	Chance of Dozing (0-3)
Sitting and reading _____	_____
Watching TV _____	_____
Sitting, inactive in a public place (e.g. a theatre or a meeting) _____	_____
As a passenger in a car for an hour without a break _____	_____
Lying down to rest in the afternoon when circumstances permit _____	_____
Sitting and talking to someone _____	_____
Sitting quietly after a lunch without alcohol _____	_____
In a car, while stopped for a few minutes in the traffic _____	_____

**THANK YOU FOR YOUR COOPERATION**

© M.W. Johns 1990-97

## Appendix 2: Modified Fatigue Impact Scale

### Modified Fatigue Impact Scale (MFIS)

Fatigue is a feeling of physical tiredness and lack of energy that many people experience from time to time. But people who have medical conditions like MS experience stronger feelings of fatigue more often and with greater impact than others.

Following is a list of statements that describe the effects of fatigue. Please read each statement carefully, the circle the one number that best indicates how often fatigue has affected you in this way during the past 4 weeks. (If you need help in marking your responses, tell the interviewer the number of the best response.) Please answer every question. If you are not sure which answer to select choose the one answer that comes closest to describing you. Ask the interviewer to explain any words or phrases that you do not understand.

#### Because of my fatigue during the past 4 weeks

	Never	Rarely	Sometimes	Often	Almost Always
1. I have been less alert.	0	1	2	3	4
2. I have had difficulty paying attention for long periods of time.	0	1	2	3	4
3. I have been unable to think clearly.	0	1	2	3	4
4. I have been clumsy and uncoordinated.	0	1	2	3	4
5. I have been forgetful.	0	1	2	3	4
6. I have had to pace myself in my physical activities.	0	1	2	3	4
7. I have been less motivated to do anything that requires physical effort.	0	1	2	3	4
8. I have been less motivated to participate in social activities.	0	1	2	3	4
9. I have been limited in my ability to do things away from home.	0	1	2	3	4
10. I have trouble maintaining physical effort for long periods.	0	1	2	3	4
11. I have had difficulty making decisions.	0	1	2	3	4
12. I have been less motivated to do anything that requires thinking	0	1	2	3	4
13. My muscles have felt weak	0	1	2	3	4
14. I have been physically uncomfortable.	0	1	2	3	4
15. I have had trouble finishing tasks that require thinking.	0	1	2	3	4
16. I have had difficulty organizing my thoughts when doing things at home or at work.	0	1	2	3	4
17. I have been less able to complete tasks that require physical effort.	0	1	2	3	4
18. My thinking has been slowed down.	0	1	2	3	4
19. I have had trouble concentrating.	0	1	2	3	4
20. I have limited my physical activities.	0	1	2	3	4
21. I have needed to rest more often or for longer periods.	0	1	2	3	4

## Appendix 3: Pittsburgh Sleep Quality Index

### PITTSBURGH SLEEP QUALITY INDEX (PSQI)

**INSTRUCTIONS:** The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions.

1. During the past month, when have you usually gone to bed at night?  
USUAL BED TIME \_\_\_\_\_
2. During the past month, how long (in minutes) has it usually take you to fall asleep each night?  
NUMBER OF MINUTES \_\_\_\_\_
3. During the past month, when have you usually gotten up in the morning?  
USUAL GETTING UP TIME \_\_\_\_\_
4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spend in bed.)  
HOURS OF SLEEP PER NIGHT \_\_\_\_\_

**INSTRUCTIONS:** For each of the remaining questions, check the one best response. Please answer all questions.

5. During the past month, how often have you had trouble sleeping because you...
 

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
(a) ...cannot get to sleep within 30 minutes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) ...wake up in the middle of the night or early morning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) ...have to get up to use the bathroom	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(d) ...cannot breathe comfortably	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(e) ...cough or snore loudly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(f) ...feel too cold	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(g) ...feel too hot	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(h) ...had bad dreams	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(i) ...have pain	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(j) Other reason(s), please describe				
How often during the past month have you had trouble sleeping because of this?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



	Very good	Fairly good	Fairly bad	very bad
6. During the past month, how would you rate your sleep quality overall?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
7. During the past month, how often have you taken medicine (prescribed or "over the counter") to help you sleep?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	No problem at all	Only a very slight problem	Somewhat of a problem	A very big problem
9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	No bed partner or roommate	Partner/ roommate in other room	Partner in same room, but not same bed	Partner in same bed
10. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If you have a roommate or bed partner, ask him/her how often in the past month you have had...

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
(a) ...loud snoring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) ...long pauses between breaths while asleep	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) ...legs twitching or jerking while you sleep	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(d) ...episodes of disorientation or confusion during sleep	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(e) Other restlessness while you sleep; please describe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Appendix 4: PSQI Scoring Instructions

---

### SCORING INSTRUCTIONS FOR THE PITTSBURGH SLEEP QUALITY INDEX:

The Pittsburgh Sleep Quality Index (PSQI) contains 19 self-rated questions and 5 questions rated by the bed partner or roommate (if one is available). Only self-rated questions are included in the scoring. The 19 self-rated items are combined to form seven "component" scores, each of which has a range of 0-3 points. In all cases, a score of "0" indicates no difficulty, while a score of "3" indicates severe difficulty. The seven component scores are then added to yield one "global" score, with a range of 0-21 points, "0" indicating no difficulty and "21" indicating severe difficulties in all areas.

Scoring proceeds as follows:

---

#### Component 1: Subjective sleep quality

Examine question #6, and assign scores as follows:

Response	Component 1 score
"Very good"	0
"Fairly good"	1
"Fairly bad"	2
"Very bad"	3

*Component 1 score:* \_\_\_\_\_

---

#### Component 2: Sleep latency

1. Examine question #2, and assign scores as follows:

Response	Score
≤15 minutes	0
16-30 minutes	1
31-60 minutes	2
> 60 minutes	3

*Question #2 score:* \_\_\_\_\_

2. Examine question #5a, and assign scores as follows:

Response	Score
Not during the past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

*Question #5a score:* \_\_\_\_\_

3. Add #2 score and #5a score

*Sum of #2 and #5a:* \_\_\_\_\_

4. Assign component 2 score as follows:

Sum of #2 and #5a	Component 2 score
0	0
1-2	1
3-4	2
5-6	3

*Component 2 score:* \_\_\_\_\_

**Component 3: Sleep duration**

Examine question #4, and assign scores as follows:

<b>Response</b>	<b>Component 3 score</b>
> 7 hours	0
6-7 hours	1
5-6 hours	2
< 5 hours	3

**Component 3 score:** \_\_\_\_\_

---

**Component 4: Habitual sleep efficiency**

1. Write the number of hours slept (question #4) here: \_\_\_\_\_

2. Calculate the number of hours spent in bed:

Getting up time (question #3): \_\_\_\_\_

Bedtime (question #1): \_\_\_\_\_

\_\_\_\_\_

*Number of hours spent in bed:* \_\_\_\_\_

3. Calculate habitual sleep efficiency as follows:

(Number of hours slept/Number of hours spent in bed) X 100 = Habitual sleep efficiency (%)

( \_\_\_\_\_ / \_\_\_\_\_ ) X 100 = %

4. Assign component 4 score as follows:

<b>Habitual sleep efficiency %</b>	<b>Component 4 score</b>
> 85%	0
75-84%	1
65-74%	2
< 65%	3

**Component 4 score:** \_\_\_\_\_

---

**Component 5: Step disturbances**

1. Examine questions #5b-5j, and assign scores for each question as follows:

<b>Response</b>	<b>Score</b>
Not during the past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3
<i>5b score:</i>	_____
<i>5c score:</i>	_____
<i>5d score:</i>	_____
<i>5e score:</i>	_____
<i>5f score:</i>	_____
<i>5g score:</i>	_____
<i>5h score:</i>	_____
<i>5i score:</i>	_____
<i>5j score:</i>	_____

2. Add the scores for questions #5b-5j:

*Sum of #5b-5j:* \_\_\_\_\_

3. Assign component 5 score as follows:

<b>Sum of #5b-5j</b>	<b>Component 5 score</b>
0	0
1-9	1
10-18-4	2
19-27	3

**Component 5 score:** \_\_\_\_\_

---

**Component 6: Use of sleeping medication**

Examine question #7 and assign scores as follows:

<b>Response</b>	<b>Component 6 score</b>
Not during the past month	0
Less than once a week	1
Once or twice a week	2
Three or more times a week	3

**Component 6 score:** \_\_\_\_\_

---

**Component 7: Daytime dysfunction**

1. Examine question #8, and assign scores as follows:

<b>Response</b>	<b>Score</b>
Never	0
Once or twice	1
Once or twice each week	2
Three or more times each week	3

*Question#8 score:* \_\_\_\_\_

2. Examine question #9, and assign scores as follows:

<b>Response</b>	<b>Score</b>
No problem at all	0
Only a very slight problem	1
Somewhat of a problem	2
A very big problem	3

*Question #9 score:* \_\_\_\_\_

3. Add the scores for question #8 and #9:

*Sum of #8 and #9:* \_\_\_\_\_

4. Assign component 7 score as follows:

<b>Sum of #8 and #9</b>	<b>Component 7 score</b>
0	0
1-2	1
3-4	2
5-6	3

***Component 7 score:*** \_\_\_\_\_

---

**Global PSQI Score**

Add the seven component scores together:

***Global PSQI Score:*** \_\_\_\_\_