DISS. ETH NO: 25467

Development and application of a Human Reliability Analysis method for radiotherapy applications

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH Zurich

(Dr. sc. ETH Zurich)

presented by

DHRUV PANDYA

MSc, EPFL

born on 05.03.1988

Citizen of India

accepted on the recommendation of

Prof. Dr. Ali Mosleh

Prof. Dr. Antony J. Lomax, Prof. Dr. Giovanni Sansavini, Dr. Luca Podofillini

2018

Acknowledgement

This PhD thesis summarizes my research work performed in Risk and Human Reliability group at Paul Scherrer Institut from February 2014 to January 2018. First and foremost, I would like to thank my professor Prof. Giovanni Sansavini and supervisor Dr. Vinh Dang for providing their valuable guidance throughout the research work and routinely monitoring my progress. This helped me in bettering my research approach to solve the set research objectives.

This research would not have been possible without the help of two people: Dr. Luca Podofillini and Dr. Frank Emert. They both have played an instrumental and central role in this research work and I would like to specially thank both of them for their priceless contributions to my work. Dr. Luca Podofillini, my direct supervisor at PSI, helped me in understanding the details of the field of Human Reliability. He, with his enormous experience in the field, has guided, reviewed and challenged my research work very closely. He has supported me to always think about novel approaches, critically evaluate and improve my solutions, and perform work in systematic manner. He always showed a lot of patience towards my mistakes and consequently motivated me by giving constructive criticism. I cannot be grateful enough that he was my direct supervisor at PSI. Dr. Frank Emert, from Center for Proton Therapy (CPT), helped me in understanding the details of the Proton therapy workflow. He was always available in helping and clearing my doubts about the process. He helped in organizing talk-throughs with personnel in the CPT and played a very important role during the expert elicitation process. He has also helped me in reviewing my papers and thesis from the CPT experts' perspective. I would like to thank him for all his help.

I would like to say special thanks to my office mates Dr. Lusine Mkrtchyan, Dr. Slavka Prvakova, and Mr. Salvatore Francesco Greco. They have been the best office mates one could ever have. They supported and motivated me during the PhD life and also shared a good social life at and outside PSI which helped in de-stressing through-out the thesis.

I would also like to thank my other colleagues at PSI, Dr. Edin Alijagic, Dr. Matteo Spada, Ms. Xiaojin Zhang and Dr. Calvin Whealton for all the memorable social interactions I had with them.

I would like to thank my closest friends Saswati Das, Suraj Roy and Mitesh Gulati, who have been my emotional and mental support structure for more than a decade. They have given me countless advices on my problems and have always shown their faith in me, which have provided

me the strength to overcome my hurdles and consequently developing an open mind and growing my personality.

Finally, I would like to express my deepest gratitude to my family, specially my parents, who have silently made sacrifices so that I can pursue better quality education and have always motivated me to do and achieve better in life. I dedicate my dissertation work to them

Funding Acknowledgement

This work is funded by a PSI inter-departmental funding initiative (CROSS), PSI's Center for Proton Therapy, PSI's Energy Divisions, and the Future Resilient Systems program. I would also like to thank the personnel working at PSI's CPT for their cooperation and support. I would like to acknowledge the support from the Future Resilient Systems program at the Singapore-ETH Centre, established between the Swiss Federal Institute of Technology in Zurich (ETH Zurich) and Singapore's National Research Foundation (FI 370074011).

ABSTRACT

Radiotherapy treatment is a complex process that involves communication between multiple expertise and continuous interaction with human-machine interface systems, to ensure safe and efficient patient handling. These attributes present a risk of failure with the consequence that patient safety is compromised, and incidents or accidents may occur. A common approach to prevent such risks is by reducing undesired occurrences. This is achieved by conducting retrospective analysis of accidents and incidents aimed to identify and classify the contributing factors, and, then, recommend prevention/mitigation strategies in form of directives and guidelines for patient safety for the clinics. These form the basis for the quality assurance program in each clinic.

Literature research of outcomes of retrospective analyses of incidents and accidents from global databases, literature, and reports, indicate humans to be dominant contributors in 82-97% incidents. Recent safety guidelines point to the need of proactive risk assessment, building on and advancing beyond retrospective investigations. For this purpose, Failure Mode and Effects Analysis and Probabilistic Safety Assessment studies are conducted and have produced useful results; yet, when adopting these techniques, the systematic inclusion of possible human failures in the safety assessment is challenged by the lack of methods directly applicable to the specific radiotherapy domain. Indeed, as shown by literature research, Human Reliability Analysis (HRA) methods have been evaluated for their applicability to radiotherapy or healthcare in general to tackle and model human failures. Their application to these domains revealed that the existing methods do not address several human tasks specific to healthcare nor do they address the specificities of the radiotherapy context. The need to tailor HRA methods to specific domains is further supported by recent developments of HRA methods addressing domain-specific tasks and error producing conditions, e.g. railways, nuclear etc.

Therefore, this thesis develops the first HRA method for radiotherapy domain and applies the method to study failure sequences in the radiotherapy workflow of a specific therapy center, the Center for Proton Therapy at the Paul Scherrer Institute of Switzerland.

First, the thesis identifies and characterizes the taxonomies of factors influencing the radiotherapy personnel performance (performance influencing factors (PIF)) and tasks representative (Generic Task Types (GTTs)) of the radiotherapy domain that formed the building blocks of the HRA method. A total of six GTT and nine PIFs with definitions are developed for the method. A generic methodology is proposed to systematically and traceably identify set of PIFs affecting a GTT. It includes direct use of a cognitive framework to progressively map GTTs to failure modes, failure causes, failure mechanisms and PIFs. This provides a strong theoretical basis to the method. Then, the methodology is applied to the radiotherapy domain and develops GTT-PIF structures for the method. A total of eighteen GTT-PIF structures are developed for radiotherapy based on the proposed methodology. Further, these structures are validated against existing literature.

Building on the developed qualitative assessment, the thesis addresses the quantification approach for the developed HRA method. To this aim, the Decision Tree (DT) methodology is chosen as the quantification methodology to compute the influence of the identified PIFs on the failure probabilities of the GTT-failure mode. Eighteen DTs are developed (one for each GTT-failure mode- PIF structure), in which (a) each branch point is the PIF and (b) each DT path represents the Human Error Probabilities (HEPs) due to the influence of a PIF or of a combination of PIFs. Once developed, the HEPs are estimated for paths of the DTs via a structured elicitation of judgment from domain experts. The experts assess the importance of specific human factors on the failure probability by means of a qualitative scale. Expert inputs are converted into statements about the order of magnitude of the probability values; these statements are then combined via an expert aggregation method, developed specifically for HRA. To build confidence on the developed methodology, the thesis validates the elicitation results against relevant applicable HEPs from existing HRA methods.

Finally, the thesis combines the results of the two building blocks, i.e. the identified GTT-PIF structures and DT for the quantification of the HEPs and investigates ten failure sequences for the 4D radiotherapy treatment workflow at Paul Scherrer Institut, to systematically assess and quantify the associated failure probabilities. The analysis transferred into safety-enhancing proposals related to the implementation of checks and to the improvement of their effectiveness.

SOMMARIO

Il trattamento radioterapico è considerato un processo di natura complessa che generalmente richiede un confronto tra diverse aree di competenze e una continua interazione con interfacce uomo-macchina, al fine di garantire un sicuro ed efficiente trattamento dei pazienti. A tali elementi è possibile associare un determinato rischio di fallimento le cui conseguenze comporterebbero una compromissione della sicurezza del paziente, nonché l'occorrenza di eventi incidentali di varia entità. Un approccio comunemente impiegato allo scopo di ridurre tali rischi consiste nella mitigazione della probabilità di occorrenza di questi ultimi. Tale obiettivo viene raggiunto mediante lo svolgimento di analisi retrospettive di eventi incidentali, il cui scopo è quello di identificare e classificare i fattori che contribuiscono al rischio e, successivamente, suggerire strategie di prevenzione e mitigazione sotto forma di direttive o linee guida per la sicurezza del paziente. Tali strategie costituiscono le basi del programma di garanzia della qualità in ciascuna clinica o struttura ospedaliera specializzata.

I risultati forniti dalle analisi retrospettive di eventi incidentali provenienti da diverse fonti (database globali, letteratura, report etc.) mostrano come il fattore umano contribuisca in modo dominante nel 82-97% dei casi. I più recenti orientamenti in materia di sicurezza sottolineano la necessità di usare tecniche di analisi dei rischi di tipo proattivo, sviluppate a partire da, e che vadano oltre le indagini di tipo retrospettivo. A tal proposito, sono stati intrapresi numerosi studi di sicurezza di tipo probabilistico (Probabilistic Safety Assessment, PSA) e analisi FMEA (Failure Mode and Effects Analysis), alcuni dei quali con promettenti risultati. Ciononostante, nell'applicare tali tecniche, l'inclusione sistematica di tutti i possibili errori umani negli studi di sicurezza è minata dalla mancanza di metodi che siano direttamente applicabili allo specifico contesto radioterapico.

In tal senso, come riportato in letteratura, negli anni recenti sono stati effettuati numerosi studi di valutazione dell'applicabilità di diversi metodi di analisi dell'affidabilità umana (Human Reliability Analysis, HRA) al settore della radioterapia, o più in generale nel settore sanitario, in merito alla loro bontà nel caratterizzare e modellare gli errori umani. Dalle applicazioni HRA in tali contesti si evince come i metodi esistenti non siano in grado di rappresentare adeguatamente

un cospicuo numero di azioni umane tipiche del settore sanitario, né di caratterizzare opportunamente alcune specificità del contesto radioterapico. La necessità di adeguare i metodi HRA agli specifici ambiti di applicazione trova ulteriore riscontro nei recenti orientamenti della comunità scientifica verso lo sviluppo di nuovi metodi che siano in grado di caratterizzare in modo ottimale le performance degli operatori, nonché le condizioni che promuovono l'errore umano, nel relativo dominio di applicazione del metodo (ad es. nucleare, trasporti ferroviari etc.).

Pertanto, il presente lavoro di tesi si propone di sviluppare il primo metodo HRA specifico per il settore della radioterapia e, conseguentemente, di applicare tale metodo allo scopo di investigare le possibili sequenze di errori nei processi radioterapici riguardanti uno specifico centro di trattamento, il Center of Proton Therapy operante in Svizzera presso il Paul Scherrer Institute.

In primo luogo, la tesi identifica e caratterizza la tassonomia degli elementi costitutivi del metodo HRA, nella fattispecie i fattori che influenzano la performance degli operatori (Performance Influencing Factors, PIFs) in radioterapia e le task caratteristiche di tale settore (Generic Task Types, GTTs). In totale, sei GTTs e nove PIFs sono stati sviluppati per il metodo proposto, ciascuno con le rispettive definizioni. La tesi propone inoltre una metodologia generica che consenta di identificare, in modo sistematico e tracciabile, i gruppi di PIFs che influenzano ciascun GTT. Tale metodologia prevede l'uso diretto di modelli cognitivi allo scopo di accoppiare in modo progressivo i GTT alle relative modalità di errore, cause di errore, meccanismi di errore e PIFs. L'approccio adottato consente pertanto di rafforzare le basi teoriche del metodo HRA sviluppato nel lavoro di tesi. La metodologia proposta è successivamente applicata al settore della radioterapia allo scopo di sviluppare il set di GTT-PIF alla base del metodo HRA. In totale, diciotto set GTT-PIF sono stati individuati relativamente al contesto radioterapico e successivamente validati tramite una revisione della letteratura disponibile.

A partire dagli aspetti qualitativi definiti nella prima fase del lavoro, la tesi ha in seguito affrontato lo sviluppo della struttura quantitativa del metodo HRA. A tal scopo, alberi di decisione (Decision Tree, DT) sono stati adottati come modello quantitativo per computare l'influenza dei PIFs sulle probabilità di errore umano delle coppie GTT-modalità di errore. In totale sono stati sviluppati diciotto DTs (uno per ciascuna struttura GTT-modalità di errore-PIF), all'interno dei quali (a) ciascun nodo costituisce un PIF e (b) ciascun cammino lungo l'albero

rappresenta la probabilità di errore umano (Human Error Probability, HEP) relativa all'influenza di un PIF o ad una combinazione di più PIFs. I valori di HEPs associati a ciascun cammino all'interno del DT sono stati estrapolati facendo ricorso al giudizio di esperti del settore. Nella fattispecie, gli esperti stimano l'importanza di specifici fattori umani sulla probabilità di fallimento della performance utilizzando una scala qualitativa. Ciascun giudizio fornito dagli esperti viene dapprima convertito in modo tale da fornire una stima quantitativa dell'ordine di grandezza della probabilità di errore. Successivamente, tali stime vengono combinate mediante un metodo di aggregazione sviluppato appositamente per applicazioni HRA. Allo scopo di consolidare la struttura quantitativa del metodo HRA proposto, i risultati ottenuti dal processo di aggregazione sono stati confrontati, laddove possibile, con gli omologhi valori di HEP provenienti dai metodi HRA esistenti.

Infine, la tesi combina i risultati dei due blocchi di lavoro (ovvero, le strutture GTT-PIF identificate nella fase qualitativa e i modelli DT per la quantificazione delle HEPs) allo scopo di investigare dieci potenziali sequenze di fallimento nei processi di lavoro relativi al trattamento radioterapico di tipo 4D, in fase di sviluppo presso il Paul Scherrer Institute, e quantificarne le relative probabilità di errore umano. Le analisi effettuate hanno fornito importanti raccomandazioni atte a migliorare la sicurezza dei processi lavorativi, relativamente all'introduzione di opportune azioni di controllo e al miglioramento dell'efficacia di quest'ultime.

List of Abbreviations

Abbreviation	Full form
3D/4D	3rd Dimension or 4th Dimension
A	Action
AAPM	American Association for Physicist in Medicine
ATHEANA	A Technique for Human Error Analysis
CARA	Controller Action Reliability Assessment
CIRS	Critical Incident Reporting System
CPT	Center for Proton Therapy
CT	Computed Tomography
CTV	Clinical Treatment Volume
DIEW	Distraction/interruptions and excessive workload
DM	Decision-making
D/N	Detecting and noticing
DT	Decision Tree
DVH	Dose Volume Histogram
E	Extreme
ED	Environmental distractions
FMEA	Failure Mode and Effects Analysis
GTT	Generic Task Type
GTV	Gross Treatment Volume
GUI	Graphic User Interface
H	High
HEART	Human Error Assessment and Reduction Technique
HEP	Human Error Probability
HFACS	Human Factors Analysis and Classification System
HFE	Human Factors and Ergonomics
HRA	Human Reliability Analysis

ICRP	International Commission on Radiological Protection
ICU-V	Information content unclear- Verbal
IMPT	Intensity Modulated Therapy
IU	Information unclear
L	Low
LTE	Lack of experience and training
LVDTENE	Low vigilance due to expecting no error
M	Moderate
MCF	Macrocognitive Function
MD	Medical Directive
MRI	Magnetic Resonance Imaging
MU	Monitor Units
NARA	Nuclear Action Reliability Assessment
PACS	Picture Archiving System
PATBASE	Patient Database
PIF	Performance Influencing Factors
PPV	Patient Positioning Verification software
PSA	Probabilistic Safety Assessment
PSI	Paul Scherrer Institut
PTV	Planning Treatment Volume
QA	Quality Assurance
RARA	Railways Action Reliability Assessment
ROSIS	Radiation Oncology Safety Information System
RMA	Risk Matrix Approach
SAFRON	Safety in Radiation Oncology
SF	Steering File
SPAR-H	Simplified Plant Analysis Risk Human Reliability Assessment
SM/U	Sense-making and understanding
SRU	Software or resource unavailable

TC	Tumor complexity
T-C	Team-coordination
TG100	Task Group 100
THERP	A Technique for Human Error-rate prediction
TP	Time pressure
TPS	Therapy Planning System
VOI	Volumes Of Interest

Contents

Table of Figures

Table of Tables

Chapter 1: Introduction

This chapter gives an overview into the PhD research conducted to develop a Human Reliability Analysis method for the radiotherapy field, which is explained in detail in the upcoming chapters. It presents the motivation and the rationale behind the research project. It then presents the key research aims and objectives for different parts of the thesis, and the key contributions addressed in this thesis. The last section of Chapter 1 presents the outline of the thesis and the overview of the approach taken to achieve each objective in the respective chapters.

1.1: Background to the problem: Motivation

Radiotherapy is the process of treating tumors using radiation, which can be delivered either by placing the radiative source inside the body (termed as Brachytherapy) or by guiding the beams from outside (termed as external beam radiotherapy) [1]. The goal of external beam radiotherapy (EBRT) is to safely and effectively deliver a dose distribution to the tumor while sparing healthy tissue around it [2]. Modern radiotherapy methods allow the treatment of cancer tumors even in the proximity of critical organs for example brain, spinal cord etc. To achieve this, a high resolution medical imaging [3] is combined with state-of-the-art radiation delivery techniques [4] based on sophisticated treatment planning systems [5]. Particle radiotherapy (PRT) exploits the potential of protons and carbon ions to deliver the dose in a favorable spatial distribution due to their precisely defined, energy dependent, finite range (Bragg peak) [6], [7]. This sophisticated level of spatial precision in photon and especially PRT requires an exact knowledge of the target position of tumor in space and time throughout the entire treatment [8, 9].

Figure 1 shows a generic radiotherapy process, highlighting the diversity of technologies and personnel expertise involved in each treatment stage [10]. The exact number of stages, their interfaces, and the specific allocation of responsibilities can be different depending on the radiation type and treatment center. For example, at the Paul Scherrer Institute’s Center for Proton Therapy, the treatment process involves nine phases and five types of expertise: radiation oncologists, radiation therapist, medical physicists, planners/dosimetrists and medical assistants.

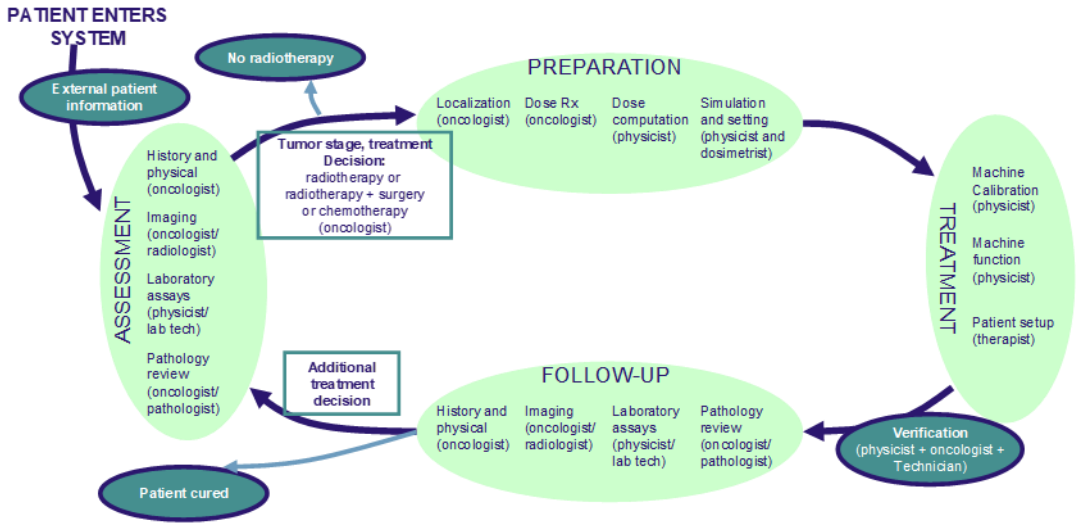


Figure 1: Radiotherapy process map: main treatment stages, technologies and personnel involved [10]

All the above makes the radiotherapy treatment ***a complex process, which requires interaction of expertise from various backgrounds to successfully deliver the treatment***. Besides technological aspects, safety and reliability of the correct dose strongly depends on human performance [1, 11, 12].

Radiotherapy, as any other domain of healthcare prioritizes safety of the patients [11, 12, 13]. Following Medical Exposure Directives, for example General MED 97/43/EURATOM by the European Council [14], most of the facilities maintain incident databases and perform retrospective investigations of past events, to gather information of the event causes [15]. Moreover, collaborative initiatives exist, at the national (e.g. Switzerland with ROSIS [16]) and international levels with the International Atomic Energy Agency's SAFRON [17]. The aim of such public databases is to promote incident learning across different facilities, such to prevent similar occurrences in other facilities.

In line with the central role of the personnel in the treatment process, the investigation of past incidents, near-misses and accidents consistently indicate that ***human errors are the dominant contributor to clinical incidents*** [18]. Of special importance is, that they occur in various phases of the treatment workflow. A study by the World Health Organization of reported radiotherapy incidents in the last 30 years shows that the largest number of injurious events were reported during the ***Preparation*** stage, (see **Figure 1**) (54% of injurious events, which is 22% of all incidents). For near-miss events, failures in the ***information transfer*** (across different treatment stages in **Figure 1**) were dominant (38% of the near-misses). Generally, human errors are found as a dominant contributor in 82 to 97% of the surveyed incidents [1, 19, 20, 21, 22, 23].

The vulnerabilities identified by retrospective investigations have been incorporated into guidelines and recommendations for assuring patient safety [e.g. 11, 12, 13, 24, 25]. These guidelines and recommendations highlight the relevance of the safety efforts in the field, which also manifests in the introduction of advanced technologies intended to improve accuracy, safety, and efficiency of the treatment. These guidelines form the basis for assuring safety at the facility (i.e. treatment clinic), which implement these as risk, safety and quality assurance, control, and management measures. However, concerns have been raised regarding the ***disproportionate focus of guidelines on equipment*** and on the need to ***better prioritize limited resources for enhancing safety and reliability of dose delivery*** [26, 27].

Taxonomies based on retrospective analysis have also been developed to ***characterize and address the root causes of occurred events (including human, technical and organizational causes), and to propose effective safety-enhancing measures***. For example:

(1) Reference [28] proposes to classify errors into prescription, preparation and treatment domains of the process based on their type. It identified three types of error categories: Prescription (dose or volume errors), Occurrence (Systematic or sporadic) and Source (process or infrastructure).
(2) Reference [29] uses an adapted version of the Human Factors Analysis and Classification System (HFACS) from air traffic to analyze incidents. The analysis based on HFACS was used to provide knowledge on types of cognitive failures and more importantly on the recurring causal factors that affect the performance. The analysis of human errors has been done based on skill-based, rule-based and knowledge-based error.
(3) Reference [30] proposes a practice-based taxonomy to study the causal factors affecting the treatment. This taxonomy is focused to where in the process the error occurred and where it was stopped. To classify human errors it uses skill, rule and knowledge based classification.
(4) The most recent and detailed taxonomy is developed by reference [31]. It has a detailed hierarchical taxonomy for root causes which is intended to be a part of a risk management / incident learning system.

All these analyses are based on retrospective investigations of occurred events and ***do not include prospective and proactive assessment of risk (both technical and human) to reduce it***. The radiotherapy community recognizes the need to complement the retrospective analyses and evaluate the effectiveness of the safety measures and quality assurance procedures proactively and prospectively at individual clinics [21, 26, 27].

For risk analyses, the American Association of Physicists in Medicine (AAPM) recommends the use of a widely established technique: ***Failure Mode and Effects Analysis (FMEA)***. FMEA supports a comprehensive and systematic analysis of hazards and safety barriers and implements a fairly detailed methodology to understand human failures. Many radiotherapy center/clinics have applied it and converted the results in safety-enhancing measures [32, 33, 34]. Another regulatory organization, the Ibero-American Forum of Radiological and Nuclear Regulatory Agencies, proposes the use of the Risk Matrix Approach (RMA) [35], a semi-quantitative approach to estimate the probability and severity of harmful events including the assessment of barriers [15].

Furthermore, the International Commission on Radiological Protection recommends the use of quantitative methods for a comprehensive assessment of risks in the process [21]. In particular, it suggests the use of ***Probabilistic Safety Assessment (PSA) techniques***, which enable system perspective, orientation towards scenarios combining the interactions of equipment, personnel, and safety systems, and measures. These features make PSA a complementary tool to FMEA to produce an overall, quantitative assessment of the facility risk and quantification of the relevant contributors to risk [36]. The use of PSA techniques to assess the safety of radiotherapy has been investigated and produced useful results [e.g. 18, 37, 38, 39]. However, ***the analysis of human errors and estimation of human failure probabilities are a major impediment to PSA***. This is due to the lack of established failure probability values for the human tasks involved in radiotherapy (and in the healthcare domain in general). Indeed, conscious use of screening can support practical results and conclusions [18], but detailed techniques would allow more realistic risk profiles as well as a systematic assessment of the performance influencing factors.

Recently, the Task Group 100 by the AAPM's developed a risk management tool based on the combination of FMEA and PSA technique: Fault Trees (FTs) [40]. The tool aims at studying causal scenarios using FTs for the highest risk priority number events obtained from FMEA performed at a specific clinic. This is the latest approach to study human failures by combining the strengths of the two types of techniques, i.e. systems perspective, scenario combination from PSA and relative prioritization coming from FMEA.

On the other hand, there have been recent efforts on studying systematically the impact of factors influencing human performance and human errors quantitatively in radiotherapy (and healthcare in general). This showcases the growing importance of quantification of human failure probabilities for the radiotherapy field. A review about the applicability of existing Human

Reliability Analysis[(1)](HRA) methods to healthcare in general was performed by [41, 42]. This included few studies in which HRA methods have been applied in radiotherapy or in healthcare safety assessments. For example, reference [43] applied A Technique for Human Error Analysis (ATHEANA) HRA [44] method to Cobalt therapy to gain qualitative insights. However, the conclusion of the review was that ***the available methods for HRA do not address a number of human tasks specific to healthcare safety assessments nor do they address the specificities of the radiotherapy context***, e.g. the factors that influence the reliability of tasks in radiotherapy and in the healthcare domain. As another example, reference [45] includes the application of existing HRA methods, i.e. Human Error Assessment and Reduction Technique (HEART) [46, 47], to a radiotherapy task. This study acknowledged that the HEART methodology is quick to deploy but may not be suitably applicable to all healthcare-related tasks without further development or modification [45]. Reference [48] uses the Bayesian Belief Network (BBN) approach to study the relative importance of tasks and factors affecting each task, and thereby studying preventive measures in the radiotherapy process (external and internal radiotherapy). Furthermore, the authors acknowledged that the approach allows understanding critical tasks and factors in the radiotherapy process but the application will be facility-specific, i.e. the BBN and its quantification are dependent upon the procedure of the facility and will need to be modified accordingly when applied to other facilities. A recent NUREG study also concludes that a detailed radiotherapy-specific HRA would allow better characterization of the risk and of its main contributors, and the identification and evaluation of specific improvements to the process [48].

Besides its relevance for the radiotherapy domain, the present work contributes to the recent trend by the HRA community to ***develop domain-specific HRA methods, which acknowledge the specificities of the domains of applications***. A strong interest has been witnessed in establishing domain-specific HRA methods beyond the nuclear domain in recent years. This is intended to promote HRA methods that characterize tasks and error-producing conditions specific to the domain under assessment, to possibly prioritize risk contributors and, additionally, to also promote the use of human error probabilities tailored to specific domains. These include:

- US National Aeronautics and Space Administration's guidance on application of four HRA methods on space mission analysis [50];

(1) HRA is the field of study that is based on sets of methods and tools to identify and analyze (qualitatively and quantitatively) possible human failures.

- the development of Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) based HRA method for oil and gas field [51, 52];
- and, lastly, sector specific variants of HEART, i.e. the Nuclear Action Reliability Analysis (NARA) for Nuclear, the Railway Action Reliability Analysis (RARA) for railways, and the Controller Action Reliability Analysis (CARA) for the air traffic control domain [53, 55, 54].

The research directions in the field of human failures in radiotherapy and the findings reviewed in this chapter point to the need of systematically identify and characterize human failures and study in detail the conditions that influence the performance of the humans involved in the radiotherapy process. Therefore, all of these together form the motivation of this Ph.D. thesis for the development of a HRA method specific for radiotherapy to help better understand, identify and estimate human failures qualitatively and quantitatively. Consequently, apply the method to enable process improvements, which can prevent, mitigate and protect against human failures in proactive terms.

1.2: Aims and objectives

Given the above background, the main aims of this dissertation are first, to develop a new human reliability analysis method specific for external beam radiotherapy (referred to as radiotherapy in the rest of the thesis) and, second, to apply it to the Center for Proton Therapy (CPT) at the Paul Scherrer Institute (PSI) for an assessment of risks in specific failure sequences and, possibly, for suggesting safety-enhancing measures. The developed HRA method must meet four main requirements:

- it should address a set of safety-relevant tasks representative of the radiotherapy treatment process and the relevant failure modes;
- it should identify the factors that influence the reliability of the performance of these tasks in different situations and guide their evaluation;
- it should support the estimation of task failure probabilities, accounting for the influence of the identified performance influencing conditions;
- the method development process should be traceable: in particular, in the identification of the representative tasks, of the associated influencing factors and in the use of expert judgement (Traceability allows method review and increases acceptability).

To achieve the overall aim, smaller objectives in the development of the HRA method have been formulated and are presented below.

1. Identification of factors that influence the performance of the personnel working in radiotherapy. (**Chapter 2: HRA method- qualitative aspects**)
   1.1 High-level factor definitions of the Performance Influencing Factors (PIFs) including the list of specific situations in which the PIF manifests to affect performance, termed "negative conditions".
2. A systematic and traceable procedure for developing the taxonomy of the representative set of tasks carried out by the facility personnel as part of the patient handling process in a radiotherapy facility for the radiotherapy HRA method. (**Chapter 2: HRA method- qualitative aspects**)
   2.1 Identification of the radiotherapy work-process specific tasks and their failure modes.
   2.2 Concept development for the categorization of the identified tasks into a representative task set called Generic Task Types (GTTs).
   2.3 Identification of sets of PIFs affecting the failure modes for the GTTs previously established in the radiotherapy industry using the cognitive framework.
3. Development of a quantification framework to systematically incorporate the quantitative impact of PIFs on the task failure probabilities (**Chapter 3: HRA method- quantitative elements**)
   3.1 Developing the overall concept of the quantification approach for the method.
   3.2 Developing the method's error representation framework (a set of cause-based Decision Trees) that facilitates the systematic incorporation of the impact of the PIF and their assessment.
4. Quantification of human error probabilities (HEPs) of tasks given the negative performance conditions in traceable and systematic way using expert judgment (**Chapter 3: HRA method- quantitative elements**)
   4.1 Developing and performing expert elicitation sessions with radiotherapy experts.
   4.2 A systematic and traceable way of computation of HEPs and their validation with existing HRA methods.
5. HRA method description with assumptions, limitations and application guidance (**Chapter 4: HRA method for radiotherapy: overview**)
   5.1 Combine results from previous objectives and form the method

5.2 List assumptions and limitations for method use; and provide guidance in method use.

6. Application of the developed method to potential failure sequences (**Chapter 5: Identification of failure sequences as case studies for method application** and **Chapter 6: Application of HRA method to failure sequences**)
   6.1 Identification of potential failure sequences.
   6.2 Generating HEPs for the failure sequences using the method.

Figure 2 links the above objectives and the relevant chapters of the thesis.

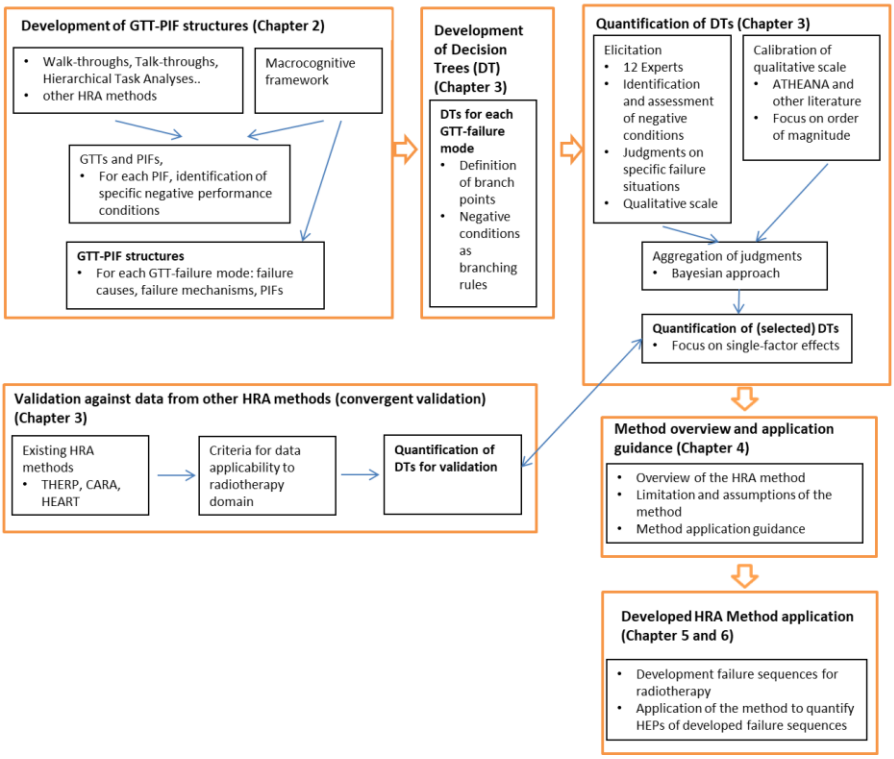


Figure 2: HRA method development and application overview

1.3: Key contributions of this research

The key contributions addressed and the original methodologies developed in this thesis are discussed in the following sections: **Section 1.3.1: A new HRA method for radiotherapy applications** to **Section 1.3.5: Applicability to other domains**.

1.3.1: A new HRA method for radiotherapy applications

To the author's knowledge, till date, no HRA method has been developed specifically for radiotherapy applications. This thesis has taken the first steps and developed the first HRA method tailored to the radiotherapy domain. It has identified and characterized tasks specific to radiotherapy and factors influencing the performance of personnel in radiotherapy. Further, it has quantified task failure probabilities accounting for the impact of the identified performance influencing factors. The developed method is generic for the domain and opens its application to other radiotherapy centers.

1.3.2: Traceability of method development

This thesis proposes a systematic and traceable methodology to develop the set of GTTs for an HRA method. These properties are needed to ease the review of the process and of its results, and, in case, to identify and implement changes. The proposed methodology entails a step-by-step approach to develop GTT- Performance Influencing Factors (PIFs) structures, which includes the explicit use of an underlying macrocognitive model [56]. This approach strengthens the link between HRA methods and cognitive models, as recently advocated to enhance the empirical basis of HRA methods [56, 57, 58, 59]. The methodology follows the progressive identification of failure modes, causes, mechanisms and performance influencing factors to develop structures of task-PIFs. The products of the proposed methodology are the GTTs and the GTT-PIF structures.

Remarkably, the thesis also validates, at least partially, the derived GTT-PIF structures against the radiotherapy-specific FMEA by the Task Group 100 from the AAPM [40].

1.3.3: Traceable use of expert judgment to quantify HEPs

Expert judgment becomes necessary to quantify HEPs for HRA methods for domains where data is insufficient. These HEPs are obtained in a variety of ways, e.g. directly eliciting probability values, probability ranges, or eliciting on qualitative or semi-quantitative scales [60, 61, 62, 63, 64, 65]. This widespread use of expert judgement calls for traceable and defendable approaches to elicit the judgments and incorporate them in HRA methods [66]. To address this need, this thesis follows a systematic and transparent approach to elicit and aggregate judgements to compute HEPs. In addition, it provides details on the design of the expert elicitation and the approach used to aggregate expert data.

A key element in the design of the expert judgment session has been to formulate the tasks and the performance conditions affecting them very specifically rather than on a generic formulation of tasks and conditions. This specificity allowed the experts contextualizing the analyzed situations and thinking of specific performance influencing conditions affecting the task. This would also reduce the cognitive effort of the experts in analyzing the abstract categories of GTTs and PIFs and foster discussions linking to daily personnel experience.

Finally, the thesis also performed convergent validation of the obtained HEP values from the expert judgment process against relevant, applicable values from existing HRA methods [67]. This convergence shows the consistency of the radiotherapy HEP values against applicable values from existing HRA methods from different domains.

1.3.4: Method application to workflow under-development

This thesis applied the developed HRA method to a workflow that is under development[(2)] with the aim of optimizing the workflow by systematically analyzing the risks in the workflow. 4D therapy, i.e. the time-dependent treatment planning and irradiation of a non-stationary tumor, is a relatively novel approach at PSI's CPT, with the 4D treatment planning workflow being currently under development.

The risk analysis performed in this thesis included:

- study of effectiveness of the checks when carried out by a different person compared to who performed the primary task,
- study of the impact of increased workload to characterize situations in which the number of 4D patients will increase.

Then, based on these analyses, the thesis proposed the implementation in the procedures of specific checks to improve the workflow robustness against failure events.

1.3.5: Applicability to other domains

The knowledge gained from the aforementioned contributions of the thesis is transferable to other domains. In particular, the following elements are directly transferable:

- the proposed systematic and traceable GTT-PIF structures formation methodology;

(2) Under development here means that the workflow with proper sequential tasks to achieve the final goal exists, yet this is under continuous improvement. The final optimized task flow will translate into the standardized operating procedure.

- the systematic and traceable elicitation and aggregation of expert judgments;
- failure sequences generation and method application.

Indeed, the developed approaches have been generically defined such that the transfer of knowledge to other domains is possible.

For example, the proposed GTT-PIF structure formation methodology can be directly applied to other domains to identify GTTs and domain-specific PIF influences on the GTTs. In addition, certain results obtained for radiotherapy domain are also transferable to other domains; these relate to tasks that are common in different domains like quality check tasks or simple execution tasks or communication tasks. It should be noted that the specific tasks will be different for other domains but the PIF influences are expected to be similar. This similarity is due to applicable generic failure mode and same macrocognitive function failure, which further links to the generic causal chain of identifying the PIFs. As another example, the elicitation process followed in the thesis and the aggregation of expert inputs can be used for other domains to quantify HEPs. However, the direct use is subject to the needs of the domain i.e. for example, a domain might require using a different scale for elicitation (as compared to the one used in this thesis). Similarly, for the direct use of HEP values from the radiotherapy results, the applicability depends on the similarity of the description of the negative conditions being studied as well as on the type of task. Influences like “less than adequate experience”, “high workload”, “low vigilance due to expectation”, have common descriptions across domains. Thus, the values for such influences on similar tasks can be directly taken.

1.4: Outline of the thesis

The remainder of this dissertation is divided into six chapters, which are described in the following paragraphs.

Chapter 2: HRA method- qualitative aspects, focusses on building the qualitative part of the HRA method, i.e. PIFs and GTTs. It first presents the scope and boundary conditions for the method development. Next, it presents the identified PIF taxonomy for radiotherapy. Then, it outlines the methodology proposed to systematically and traceably build the GTTs and the results obtained by applying the methodology to the radiotherapy process; this, is the first key contribution of this chapter. Another key contribution is the explicit use of an existing macrocognitive model to identify PIF sets affecting task-failure modes and use this piece of information to build the GTT taxonomy. Such approach gives a strong theoretical basis to the

GTT development and transparency to the process. Overall, this chapter addresses research objectives #1 and #2.

Once these two building blocks have been developed (GTTs and PIFs), **Chapter 3: HRA method- quantitative elements**, focusses on the quantitative part of the proposed HRA method. It first presents the development of the quantitative framework for the method, i.e. the decision trees. Subsequently, it applies expert judgment for gathering failure data on a qualitative scale and then it applies a Bayesian model to aggregate expert data to finally quantify the HEPs for the developed decision trees [68]. The aforementioned Bayesian aggregation methodology is based on [69]. Finally, it presents the convergent validation (defined in [67]) of the HEP results obtained from the application of the expert judgment method using consistent datasets obtained from existing HRA methods, i.e. HEART and CARA. This gives a consistency check to the obtained results and strengthen the credibility and applicability of the developed framework. Thus, this chapter addresses the research objectives #3 and #4.

Chapter 4: HRA method for radiotherapy: overview outlines the developed HRA method for radiotherapy including the assumptions and the limitations of the method. This is followed by the detailed exposition of the guiding principles for the deployment of the methodology. Thus, this chapter addresses the research objective #5.

Once the method is outlined, the focus in **Chapter 5: Identification of failure sequences as case studies for method application** shifts to the identification and characterization of the failure sequences to be studied for the application of the developed HRA method. In contrast to the nuclear domain where the accumulated experience on Probabilistic Safety Assessment (PSA) across facilities and decades can be used to identify the relevant human failure scenarios, the radiotherapy domain cannot benefit from such prior information. Therefore, the chapter first presents the step-by-step methodology adopted to identify failure sequences and, then, presents the various developed failure sequences of interest. In this thesis, four failure sequences have been developed and three of them have two additional variants, therefore, a total of ten failure sequences are developed to be assessed. Thus, this chapter addresses the research objective #6 (#6.1).

After the failure sequences have been identified, **Chapter 6: Application of HRA method to failure sequences**, focusses on the application of the radiotherapy HRA method to the developed failure sequences to systematically asses the risks and quantify the HEP values. It analyzes the results and proposes recommendations on how to improve the effectiveness of the checks and

improve task performance conditions. Remarkably, the analysis carried out in this chapter provided input to the development of the actual process workflow, i.e. identifying important check tasks to be included within the 4D therapy process workflow. Thus, overall, the chapter addresses the research objective #6 (#6.2).

Lastly, **Chapter 7: Conclusions and future work**, presents the conclusions, the key achievements of this thesis and outlines the future work.

Figure 3 presents the overall structure of the thesis with the main features of each chapter.

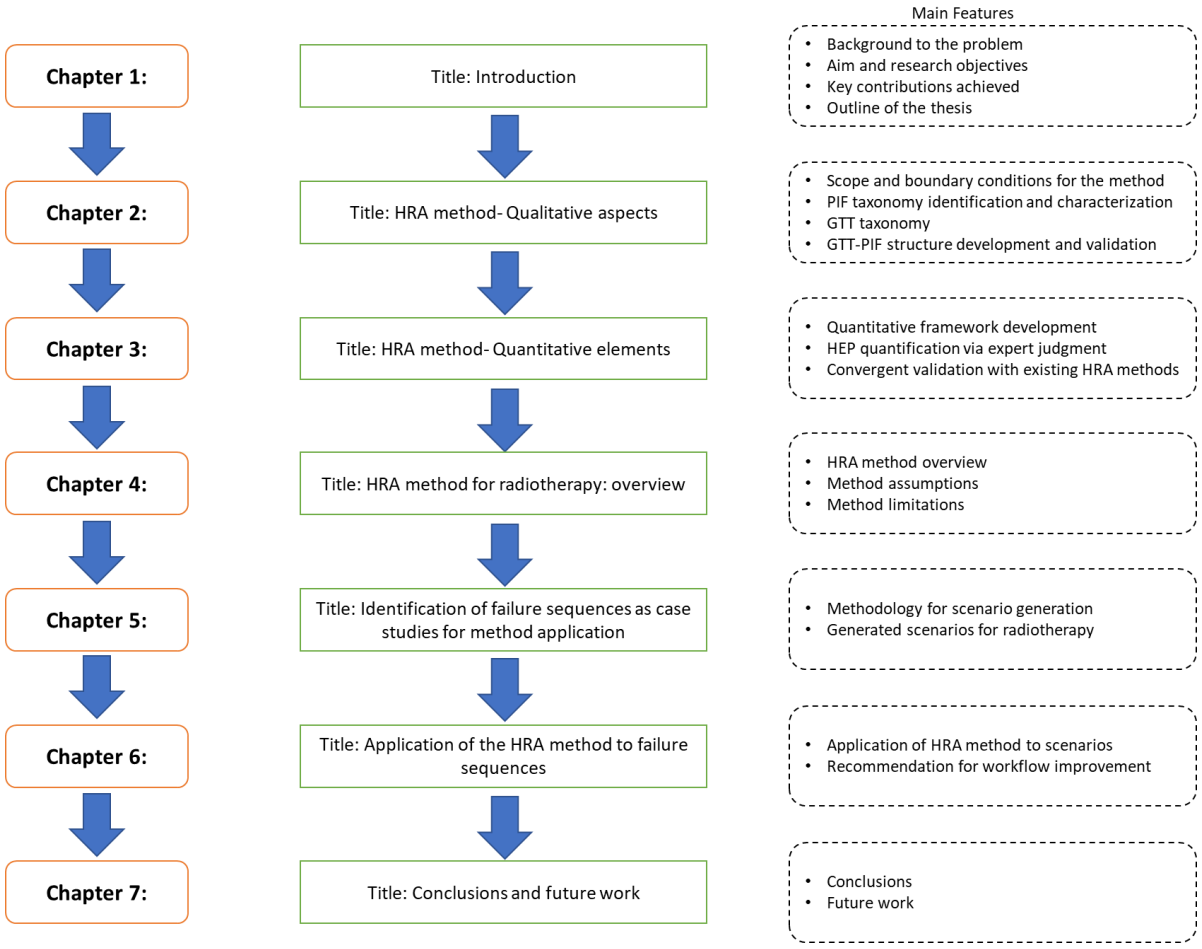


Figure 3: Thesis overview with key features in each chapter

References:

1. World Health Organization. Radiotherapy risk profile. Technical manual World Health Organization, 2008.
2. Webb S. Conformal and Intensity-Modulated Radiotherapy, in Mayles, P., Nahum, A., and Rosenwald, J.-C. (eds) Handbook of Radiotherapy Physics – Theory and Practice. 1st edition. Boca Raton (FL), 2007, pp:1 943–973.
3. Busherg JT. et al. The Essential Physics of Medical Imaging. 3rd edition. 2007.
4. Khan FM. (2010) The Physics of Radiation Therapy.
5. Khan FM, Gibbons JP and Sperduto PW. Treatment Planning in Radiation Oncology. 4th edition. 2016.
6. Durante M and Paganetti H. Nuclear Physics in Particle Therapy: A Review, Reports on Progress in Physics, 2016; 79: 1–59.
7. ICRU Report 78, Prescribing, Recording, and Reporting Proton-Beam Therapy 2007.
8. Engelsman M and Bert, C. Precision and Uncertainties in Proton Therapy for Moving Targets, in Paganetti, H. (ed.) Proton Therapy Physics. 1st edition. 2011, Boca Raton (FL), pp: 413–434. :10.1201/b11448-14.
9. Palta, J and Yeung D. Precision and Uncertainties in Proton Therapy for Nonmoving Targets, in Paganetti, H. (ed.) Proton Therapy Physics. 1st edition. 2011, Boca Raton (FL), pp. 413–434. doi: doi:10.1201/b11448-14.
10. Lee RC, Kelly KL, Newcomb C, Cooke D, et al, Quantitative approaches to patient safety, HTA Initiative #15 Health Technology Assessment Unit Alberta Heritage Foundation for Medical Research, 2004.
11. The American College of Radiology. Practice Guideline for Radiation Oncology. Practice Guidelines and Technical Standards. Reston, U.S.A., American College of Radiology, 2009.
12. Nath R, Biggs PJ, Bova FJ, Ling CC, Purdy JA, Gejin JVD and Weinhous MS. AAPM code of practice for radiotherapy accelerators. Report for the American Association of Physicists in Medicine, Task Group 45, Report number 047, September 1994.
13. Klein EE, Hanley J, bayouth J, Yin F-F, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B, and Liu C. Report for the American Association of Physicists in Medicine, Task Group 142 report: Quality assurance of medical accelerators, Medical Physics 2009; 36 (9): 4197-4212.
14. MED. European Directive 97/43 Euratom. Medical Exposure Directive. European Commisssion. http://ec.europa.eu/energy/nuclear/radiation_protection/medical/medlegislation_en.htm.
15. Malicki J, Bly R, Bulot M, Godet JL, Jahnen A, Krengli M, Maingon P, Martini CP, Przybylska K, Skrobała A, Valero M and Jarvinen H. Patient safety in external beam radiotherapy, results of the ACCIRAD project: current status of proactive risk assessment,

reactive analysis of events, and reporting and learning systems in Europe Radiother Oncol, 2017; 123, pp. 29-36. DOI: 10.1016/j.radonc.2017.02.016.
16. ROSIS, Radiation Oncology Safety Information System, http://rosis.ch/ge/rosis_daten1.asp
17. SAFRON, Safety reporting and learning system for radiotherapy, www.rpop.iaea.org/RPOP/RPoP/Modules/login/safron-register.htm
18. IAEA. Case studies in the application of probabilistic safety assessment techniques to radiation sources. Technical document for International Atomic Energy Agency -TECDOC no.-1494, Vienna, 2006.
19. International Atomic Energy Agency (IAEA). Lessons learned from accidental exposures in radiotherapy. Safety report- IAEA-Series No. 17, Vienna, 2000.
20. International Commission on Radiological Protection. Prevention of accidental exposures to patients undergoing radiation therapy. Report for International Commission on Radiological Protection, publication 86, volume 30, no. 3, 2000.
21. International Commission on Radiological Protection. Preventing accidental exposures from new external beam radiation therapy technologies ICRP REF: 32/147/07, 2009.
22. Yeung TK, Bortolotto K, Cosby S, Hoar M and Lederer E. Quality assurance in radiotherapy: evaluation of errors and incidents recorded over a 10 year period. Radiotherapy and Oncology 2005; 74(3): 283-291.
23. Derreumaux S, Etard C, Huet C, Trompier F, Clairand I, Bottolier-Depois JF, Aubert B and Gourmelon P. Lessons from recent accidents in radiation therapy in France. Radiation Protection Dosimetry 2008; 131(1): 130-135.
24. ESTRO 1995. Practical Guidelines for the Implementation of a Quality System in Radiotherapy, Physics for Clinical Radiotherapy Booklet No. 4, ESTRO, Brussels (1998).
25. European Council Directive 2013/59/Euratom on basic safety standards for protection against the dangers arising from exposure to ionising radiation and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom. OJ of the EU. L13; 57: 1–73 (2014).
26. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Medin PM, Mundt A, Mutic S, Palta JR, Thomadsen BR, Williamson JF and Yorke ED. A Method for Evaluating Quality Assurance Needs in Radiation Therapy. International Journal Radiation Oncology Biology Physics 2008; 71(1 suppl): 170-173.
27. Thomadsen B. Critique of traditional quality assurance paradigm. International Journal of Radiation Oncology Biology Physics 2008; 71(1 suppl) pp: 166-169.
28. Ekaette EU, Lee RC, Cooke DL, Kelly K-L and Dunscombe PB. Risk analysis in radiation treatment: Application of a new taxonomic structure. Radiotherapy and Oncology 2006; 80(3): 282-287.
29. Portalauri M, Fucilli F I.M., Bambace S, Castagna R, Chiara De Luca M, Pili G, Didonna V, Tramacere F, Francavilla MC, Leone A and Leo MG. Incidents analysis in radiation therapy:

application of the human factors analysis and classification system. Ann Ist Super Sanità 2009, 45(2), pp: 128-133.

30. Lam C, Medlam G, Wighton A, Breen SL, Bissonnette J-P, McGowan T S, Carlone M and Milosevic MF. A Practice-based Taxonomy for Radiation Treatment Errors. Journal of Medical Imaging and Radiation Sciences 2013, 44: 173-179.
31. Ford E, Santos LF de L, Pawlicki T, Sutlief S and Dunscombe PB. Consensus recommendations for incident learning database structures in radiation oncology. Medical Physics 2012; 39(12): 7272-7290.
32. Scorsetti M, Signori C, Lattuada P, Urso G, Bignardi M, Navarria P, Castiglioni S, Mancosu P and Trucco P. Applying failure mode effects and criticality analysis in radiotherapy: Lessons learned and perspectives of enhancement. Radiotherapy and Oncology 2010, 94: 367–374
33. Wilkinson, D and Koller A. Failure modes and effects analysis applied to high-dose-rate brachytherapy treatment planning. Brachytherapy 2013, Volume 12 (4); pp: 382 – 386. DOI https://doi.org/10.1016/j.brachy.2013.03.002
34. Cantone MC, Ciocca M, Dionisi F, Fossati P,Stefano S, Krengli M, Molinelli S, Orecchia R, Schwarz M, Veronese I and Vitolo V. Application of failure mode and effects analysis to treatment planning in scanned proton beam radiotherapy. Radiation Oncology 2013; pp: 8:127. doi:10.1186/1748-717X-8-127.
35. Torres A and Montes de Oca J. Nuevo algoritmo para análisis de riesgo en radioterapia. Nuclear 2015, pp: 39-46.
36. Marx DA and Slonim AD. Assessing patient safety risk before the injury occurs: an introduction to sociotechnical probabilistic risk modelling in health care. BMJ Quality and Safety 2003; 12(2 suppl): 33-38.
37. Ekaette E, Lee RC, Cooke DL, Iftody S and Craighead P. Probabilistic fault tree analysis of a radiation treatment system. Risk Analysis 2007; 27(6): 1395-1410.
38. Vilaragut JJ, Ferro R, Lozano B, De la Fuente A, Duménigo C, Troncoso M, Pérez Y, Alemañy J, León L, Amador R, Lazo R, Labrador F, Blanco A, Betancourt L, Crespo D and Silvestre I. Probabilistic safety assessment to the cobalt therapy process. In: 11th International Congress of the International Radiation Protection Association (IRPA), Madrid Spain, 23-28 May 2004.
39. Vilaragut JJ, Ferro R, Marti MR, López PO, Ramírez ML , Mulas AP , Montero MB, Somoano F, Rodriguez JMD, Papadópulos S, Pereira Jr PP, Morones RL, Cortina EL, Oliva JJR and Alemañy J. Probabilistic safety assessment of the radiotherapy treatment process with an electron linear accelerator for medical uses. In: 12th International Congress of the International Radiation Protection Association (IRPA), Buenos Aires, Argentina, 19-24 October 2008.
40. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR, Williamson JF and Yorke ED. The report of Task Group 100 of

the AAPM: Application of risk analysis methods to radiation therapy quality management, Medical physics 2016; 43 (7): 4209-4262.

41. Lyons M, Adams S, Woloshynowych M, Vincent C. Human reliability analysis in healthcare: a review of techniques. International Journal of Risk and Safety in Medicine 2004; 16: 223-237.
42. Turra F and Verbano C. A human reliability analysis approach to clinical risk management: first steps towards a new methodology. In: proceedings of the European Safety and Reliability Conference (Eds. BMD Sjoberg and T Aven), Stavanger, Norway, 25-27 June 2007, pp: 209-217, Taylor & Francis group.
43. McLeod JN, Baron J and Rivera S. Human Reliability Analysis in Cobalt-Therapy Process using an Adapted ATHEANA Prospective Approach. In Proceedings of Probability Safety Assessment and Management (Eds. C Spitzer, U Schmocker and VN Dang), Berlin, Germany, 14-18 June 2004.
44. U.S. Nuclear Regulatory Commission. Technical basis and implementation guidelines for A Technique for Human Error Analysis (ATHEANA). Technical report NUREG-1624, Rev 1, U.S. Nuclear Regulatory Commission, Washington DC, May 2000.
45. Chadwick L and Fallon EF. Human Reliability Assessment of a critical nursing task in a radiotherapy treatment process. Applied Ergonomics 2012; 43(1): 89-97.
46. Williams JC. HEART – a proposed method for assessing and reducing human error. In: 9th Advances in Reliability Technology Symposium, Bradford, U.K., 2-4 April 1986, University of Bradford.
47. Williams JC and JL Bell. Consolidation of the error producing conditions used in the human error assessment and reduction technique (HEART). Safety and Reliability 2016; 35(3): 26-76.
48. Gomes EC, Duarte JP and Frutuoso e Melo PF. Human reliability modeling of radiotherapy procedures by Bayesian networks and expert opinion elicitation. Nuclear Technology 2016; 194(1): 73-96.
49. Wreathall J, Brown WS, Militello L, Cooper SE, Lopez C and Franklin C. A risk-informed approach to understanding human error in radiation therapy. Technical report NUREG-2170, U.S. Nuclear Regulatory Commission, Washington DC, June 2017.
50. Chandler FT, Chang YH, Mosleh A, Marble J, Boring RL and Gertman DI. Human Reliability Analysis Methods selection guidance for NASA. Technical Report for National Aeronautics and Space Administration, U.S.A., July 2006.
51. Rasmussen M, Standal MI and Laumann K. Task complexity as a performance shaping factor: A review and recommendations in Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) adaption. Safety Science 2015; 76: 228-238.
52. Laumann K, Oien K, Taylor C, Boring RL, Rasmussen M. Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability

analysis methods. In: The proceedings of Probabilistic Safety Assessment and Management (Eds. C Smith and T Paulos), Honolulu, U.S.A., 22-27 June 2014.

53. Kirwan B, Gibson HW, Kennedy R, Edmunds J, Cooksley G and Umbers I. Nuclear Action Reliability Assessment (NARA): A Data-based HRA tool. In: Proceedings of Probability Safety Assessment and Management (Eds. C Spitzer, U Schmocker and VN Dang), Berlin, Germany, 14-18 June 2004, pp: 1206-1211.
54. Gibson HW, Mills A, Smith S and Kirwan B. Railway Action Reliability Assessment A Railway - Specific Approach to Human Error Quantification. In: International Rail Human Factors Conference, Rail Human Factors Supporting Reliability, Safety and Cost Reduction (eds. N Dadashi, A Scott, JR Wilson and A Mills), London, UK, 5-7 March 2013, pp: 671-676, Taylor & Francis Group.
55. Kirwan B and Gibson WH. CARA: A Human Reliability Assessment Tool for Air Traffic Safety Management– Technical Basis and Preliminary Architecture. In: The Safety of Systems (eds. F Redmill and T Anderson), Bristol, U.K., 13-15 February 2007, pp: 197-214, Springer.
56. Whaley AM, Xing J, Boring RL, Hendrickson SML, Joe JC, LeBlanc KL and Morrow SL. Cognitive basis for Human Reliability Analysis. Technical report NUREG-2114, U.S. Nuclear Regulatory Commission, Washington DC, January 2016.
57. Forester J, Dang VN, Bye A, Lois E, Massaiu S, Broberg H, Braarud PØ, Boring RL, Männistö I, Liao H, Julius J, Parry G and Nelson P. The International HRA Empirical Study –Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data. Technical report NUREG-2127, U.S. Nuclear Regulatory Commission, Washington DC, August 2014.
58. Chang YJ, Bley D, Criscione L, Kirwan B, Mosleh A, Madary T, Nowell R, Richards R, Roth EM, Sieben S and Zoulis A. The SACADA database for human reliability and human performance. Reliability Engineering System Safety. Reliability Engineering and System Safety 2014; 125: 117-133.
59. Ekanem NJ, Mosleh A and Shen S-H. Phoenix: A model based Human reliability analysis methodology: Qualitative analysis procedure. Reliability Engineering System Safety 2016; 145: 301-315.
60. Li PC, Chen GH, Dai LC and Zhang L. A fuzzy Bayesian network approach to improve the quantification of organizational influences in HRA frameworks. Safety Science 2012; 50 (7): 1569-1583.
61. Martins MR and Maturana MC. Application of Bayesian belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. Reliab Eng Syst Saf 2013; 110: 89-109.
62. Trucco P, Cagno E, Ruggeri F and Grande O. A Bayesian belief network modelling of organisational factors in risk analysis: a case study in maritime transportation. Reliab Eng Syst Saf 2008; 93 (6): 845-856.

63. Ale BJ, Bellamy LJ, Cooke RM, Goossens LHJ, Hale AR, Roelen ALC and Smith E. Towards a causal model for air transport safety-an ongoing research project. Saf Sci 2006; 44: 657-673.
64. Ale BJ, Bellamy LJ, Van der Boom R, Cooper J, Cooke RM, Goossens LHJ, Hale AR, Kurowicka D, Roelen ALC and Spouge J. Further development of a causal model for air transport safety (CATS): building the mathematical heart. Reliab Eng Syst Saf 2009, 94 (9): 1433-1441.
65. Musharraf M, Hassan J, Khan F, Veitch B, MacKinnon S and Imtiaz D. Human reliability assessment during offshore emergency conditions. Saf Sci 2013, 59: 19-27.
66. Mkrtchyan L, Podofillini L and Dang VN. Bayesian belief networks for human reliability analysis: A review of applications and gaps. Reliab Eng Syst Saf 2015; 139: 1-16. https://doi.org/10.1016/j.ress.2015.02.006.
67. B Kirwan. Validation of Human Reliability Assessment techniques: part 1 – validation issues. Safety Science 1997; 27(1): 25-41.
68. Podofillini L and Dang VN. A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. Reliab Eng Syst Saf 2013; 117: 52-64.
69. A Mosleh. Bayesian modeling of expert-to-expert variability and dependence in estimating rare event frequencies. Reliability Engineering and System Safety 1992; 38: 47-57.

Chapter 2: HRA method- qualitative aspects

This chapter reproduces the author's published article in the international scientific journal: "*Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*". The work in this article is to identify the set of factors (PIFs) influencing each of the safety-relevant tasks categorized as GTTs. Each GTT failure mode may be influenced by a different set of factors. The proposed method allows traceable identification of these sets of factors, developing a mapping between failure modes and influencing factors via failure causes and failure mechanisms. Traceability supports the completeness of the mapped set of factors as well as the review by others (see objectives 1 and 2 in **Section 1.2: Aims and objectives**).

The chapter first briefly outlines the overview of all the elements of HRA method, which include PIFs, GTTs and the underlying cognitive framework. Then, it presents the proposed methodology to systematically and transparently develop the Generic Task Types (GTTs) for an HRA method. Another key element of this methodology, apart from being systematic and transparent, is that it explicitly uses the described underlying macrocognitive model to identify sets of PIFs affecting task-failure mode and using that information to build GTTs and GTT (failure mode)-PIF structures. The proposed methodology is then applied to the radiotherapy domain workflow and the obtained results i.e. GTTs and GTT-PIF structures are then presented. Finally, validation of these GTT-PIF structures using existing FMEA from AAPM's Task group 100 is done.

Publication Details:

This chapter is reproduced with the permission of **Pandya D**, Podofillini L, Emert F, Lomax AJ and Dang VN. *Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy*. Proc IMechE Part O: J Risk and Reliability 2017; First published October 2: p. 1–35. https://doi.org/10.1177/1748006X17731903

Additional information relevant for this chapter:

- The article gives a brief overview of the radiotherapy workflow to which the proposed methodology is applied, a more detailed overview can be seen in **Appendix 1**.
- As described in the article, the identification and characterization of PIFs for radiotherapy were performed in a separate work by the author. Details are presented in **Appendix 2**.

Abstract

Most Human Reliability Analysis (HRA) methods have been developed for nuclear power plant applications; this challenges the application of the available techniques to other domains. Indeed, for application to a specific domain, an HRA method should address the relevant tasks and performance conditions. The aim of the paper is to propose a methodology to develop a Generic Task Type (GTT)-Performance Influencing Factor (PIF) structure, specific for application to a domain of interest and directly linked to an underlying cognitive framework of literature. The structure provides the foundation of an HRA method built on the GTT concept: it identifies the sector-specific PIF effects on the failure probability that the method needs to represent and quantify for each GTT. The methodology is intended to support a systematic and traceable process to develop the GTT-PIF structure, to ease the review of the process and of its results, and, in case, identify and implement changes to the structure. The proposed methodology is applied to the radiotherapy domain allowing the development of sector-specific taxonomies of representative critical tasks, their failure modes, underlying cognitive failure mechanism and influencing PIFs. This is part of a broader activity carried out by the Risk and Human Reliability Group at the Paul Scherrer Institute (PSI) of Switzerland to develop an HRA method, specific for the radiotherapy domain. The activity is conducted in close cooperation with PSI's Centre for Proton Therapy (CPT), where a first application of the method is foreseen.

2.1: Introduction

Most of the Human Reliability Analysis (HRA) methods have been developed for the nuclear industry, driven by the need to include the human component into the plant Probabilistic Safety Assessment (PSA). The results of HRA, and in general of PSA, are widely used to inform safety-related decisions, operational and regulatory. Typically, HRA may point to improvements in the plant procedures, in the human-machine interface, or in the level of automation of certain activities [1, 2, 3]. The quantitative perspective, related to the calculation of a risk figure-of-merit and of its contributors, is needed to prioritize which improvements are the most effective to reducing risk.

Outside the nuclear power domain, in industries for which PSA practice is established (typically, aerospace and oil and gas), the treatment of the human component has been generally simplified, often introducing bounding error probability values (e.g. 0.01 or 0.001), without performing the

detailed analyses underlying the typical HRA [4, 5, 6]. This practice strongly limits the use of PSA to inform safety-enhancing decisions. Some efforts towards extending detailed HRA applications beyond nuclear power plants have been undertaken by the space industry: in their 2011 revision of the PSA guide, the US National Aeronautics and Space Administration has included guidance for application of four HRA methods for space mission analysis [7]. Also, a very recent initiative is ongoing to develop an HRA method specific for application to the oil and gas industry, based on an adaptation of the nuclear-specific SPAR-H method [8, 9].

One HRA method that has been successfully tailored to various domains is the Human Error Assessment and Reduction Technique (HEART) [10, 11]. The most cited strengths of HEART are its easiness to use, its resource effectiveness compared to other HRA methods, and the empirical support from a number of validation studies [12, 13]. HEART is intended to be applicable to different industrial sectors and therefore it builds on broad definitions of the types of tasks and influencing factors it addresses; the motivation for developing sector-specific versions of this method has been to provide a set of tasks, influencing factors and error probability values that would be more representative of the specific domain; this is expected to improve the acceptability, representativeness, and consistency of applications across domains. These sector-specific methods are the Nuclear Action Reliability Analysis (NARA), the Railway Action Reliability Analysis (RARA), and the Controller Action Reliability Analysis (CARA), (for the air traffic control domain) [14, 15, 16].

These adaptations demonstrate the flexibility of the HEART framework. Indeed, each domain may be characterized by different sets of safety-relevant personnel tasks, influencing factors, and applicable human failure probabilities. However, the analysis framework made of Generic Task Types (GTTs) and Error-Producing Conditions (EPCs) can be applied to several domains, provided that representative taxonomies of task types and conditions are developed and appropriate relationships to the failure probabilities are established. This paper presents a methodology to develop the GTT-Performance Influencing Factor (PIF)[3] structure, directly linked to an underlying cognitive framework of literature [17]. The GTT-PIF structure indicates which PIFs are the most relevant for each GTT, and through which failure modes and mechanisms. Once this structure has been developed, the next step will be to determine the

[3] The structure presented in this paper will refer to PIFs, intended as the generic factors influencing performance (e.g. “adequacy of training”, “task complexity”), and not to specific EPCs, as in the original HEART framework. This allows direct link with the cognitive framework underlying the developed structure. The characterization of the PIFs in terms of specific conditions (error-producing, or, more generally, performance-influencing) will be addressed in a separate work (see **Section 2.2: Main elements of the HRA method under development**).

quantitative relationships linking GTTs and PIFs to human failure probability values (the main concepts underlying the method under development will be presented in Section 2). The aim of the methodology is to support a systematic and traceable process to develop the GTT-PIF structure. These properties are needed to ease the review of the process and of its results, and, in case, identify and implement changes to the structure. Indeed, the need for developing guidance to support the creation and definition (and, finally, usage) of the PIFs has been recently recognized [18, 19, 20]. The direct link to the cognitive framework is intended to allow for comprehensiveness of the developed taxonomies as well as to make sure that the important links among these taxonomies are captured in the GTT-PIF structure. The need to strengthen the link between HRA and cognitive models has also recently been raised [17, 20, 21, 22].

In this paper, a GTT-PIF structure suitable to the external beam radiotherapy domain is developed. This is part of a broader activity carried out by the Risk and Human Reliability Group at the Paul Scherrer Institute (PSI) of Switzerland to develop an HRA method, specific for the radiotherapy domain. The activity is conducted in close cooperation with PSI's Centre for Proton Therapy (CPT), where a first application of the method is foreseen.

In radiotherapy, assuring the safety of patients is a key concern [23]. Activities are continuously running at the international, national, and clinic levels to identify root causes from past incidents, near misses and accidents. International incident reporting systems are promoted, e.g. SAFRON, ROSIS etc. (the web links are provided in the references) [24, 25]. The vulnerabilities identified from past events have been addressed with guidelines and recommendations for assuring patient safety and reducing risk at the facility (clinic) level, which implement these in form of safety and quality assurance, control, and management measures [26, 27, 28, 29]. However, the disproportionate focus of these measures/guidelines on equipment is a concern as analyses of adverse events consistently indicate human errors as dominant contributors to clinical incidents [30, 31, 32, 33, 34, 35, 36, 37]. These references stress the need to assess risk and evaluate the effectiveness of the procedures, prospectively and at individual clinics.

For prospective risk analyses, the American Association of Physicists in Medicine (AAPM) recommends Failure Modes and Effects Analysis (FMEA), and its Task Group 100 developed a consensus-based guidance for integrating risk analysis techniques (FMEA and fault trees) within the quality management programs of radiotherapy clinics [38]; Probabilistic Safety Analysis (PSA) is also recommended, emphasizing its quantitative focus, its systems perspective, and its orientation to scenarios combining the interactions of equipment, personnel, and safety systems and measures [23, 34, 39]. These techniques are widely used in various industries, including

radiotherapy, to support a comprehensive and systematic analysis of hazards and safety barriers and for radiotherapy have produced useful results [23, 39, 40, 41, 42, 43, 44]. A recognized limitation to the development of PSA for radiotherapy is the lack of HRA methods directly applicable to this domain, i.e. addressing the relevant types of personnel tasks and influencing factors [23, 39, 45, 46, 47, 48]. A conscious use of screening can allow practical results; however, the development of a method to support a detailed HRA would allow better characterization of the risk and of its main contributors, and the identification and evaluation of specific improvements to the process [23, 49].

The paper is structured as follows; in **Section 2.2: Main elements of the HRA method under development**, the main elements of the HRA method under development are introduced: the method's building blocks, its technical basis and the underlying cognitive framework of reference. In **Section 2.3: Methodology for the development of the GTT-PIF structure**, the methodology to develop the GTT-PIF structure is presented, including the data sources used in the methodology. **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy** presents the application of the methodology to radiotherapy, along with the resulting six GTTs connected to eighteen GTT-PIF structures linking possible functional failures, failure mechanisms and PIFs. A validation of the GTT-PIF framework is presented in **Section 2.5: Validation of the GTT-PIF structure against Huq et al. [38]**; the validation is based on the generic FMEA developed in reference [38]: the process tasks, failure modes and potential causes identified in reference [38] are mapped to the GTT-PIF structure to evaluate whether the structure covers the sector-specific important tasks and influencing factors. Finally, **Section 2.6: Conclusion** gives concluding remarks.

2.2: Main elements of the HRA method under development

As said in the introduction, the methodology proposed by this paper to develop the GTT-PIF structure is intended to form the basis of an HRA method for application to the radiotherapy domain. This section presents the concept underlying the HRA method, to clarify the relationship of the GTT-PIF structure to the overall analysis and quantification approach.

The HRA method under development is intended to guide the analysis and the quantification of human failures. It is planned to include all personnel tasks in the patient handling process, from when the patient is admitted for treatment until he/she leaves the facility. The involved personnel are radiation oncologists, radiation therapist, medical physicists, planners, dosimetrists and medical assistants. Human failures may indeed contribute to initiating events and to additional

failures along the accident evolution. In general, for accidents to occur, quality checks should also fail: the scope of the method will cover these accident contributors as well. Two treatment phases are out of the method scope: the doctors' decision to treat with the associated dose prescription, and the post-treatment follow-up. This is in line with assumptions in other PSA studies for radiotherapy facilities [41, 44]. The development emphasizes the personnel tasks directly involved in the treatment process. In particular, machine calibration and commissioning tasks have been left out of the scope: these tasks are expected to have characteristics similar to other industrial domains, so that the applicability of available techniques (e.g. NARA and CARA as HEART-based examples) appears less problematic [14, 16].

2.2.1: Generic Task Types and Performance Influencing Factors

The building blocks of the method are Generic Task Types (GTTs) and Performance Influencing Factors (PIFs). The concept behind the GTTs is taken from the Human Error Assessment and Reduction Technique (HEART) and it is intended to guide the analyst to address the factors that may have specific influence for the particular GTT [10, 11]. A GTT is a representative critical task (within a domain) such that the enveloped tasks are characterized by similarities in the interactions with the system, in the cognitive requirements (e.g. decision-making vs. execution), and in the potentially influencing factors. In particular, influencing factors would have the same quantitative influence on the failure probability for all tasks enveloped by the GTT. These requirements have been achieved by explicitly building the GTT-PIF structure based on an underlying cognitive model, as it will be presented in **Section 2.3: Methodology for the development of the GTT-PIF structure** and **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy**. GTTs shall be defined such as to be mutually exclusive: given a task to be analysed, the applicable GTT shall be as clear as possible, without ambiguity among different GTTs.

As typical in HRA, PIFs are used to provide the analyst with a structured framework for assessing what may influence personnel performance during a particular task. The concept is to support their assessment with anchor questions, which guide the analyst to consider the possible existence of conditions that would negatively impact the personnel performance (the Error Producing Conditions, EPCs) – this concept is presented in reference [50]. The approach envisioned is to use the GTT-PIF structure for developing a decision tree quantification framework (**Figure 4**). In this framework, the heading of the decision trees would represent influencing PIFs, with branching rules based on the developed set of anchor questions; each decision tree pathway is then mapped to failure probabilities (the root path of the tree identifies

the base failure probability of the GTT-failure mode and the rest are failure probabilities with the effect of respective PIFs, see **Figure 4**). To quantify the failure probabilities, the use of data from existing HRA methods will first be explored, e.g. from A Technique for Human Error Rate Prediction (THERP [51] and HEART, with its sector-specific variants. Indeed, although HRA methods have been developed for generally different types of tasks and operating contexts, it may very well be that some specific data are relevant for the radiotherapy domain as well (for example, generic data on reading errors from display may be expected to be industry-independent). The CORE-DATA database, which includes probability data on human failures in safety-related tasks in diverse industries, will also be analyzed [52]. In addition to the input from existing methods and databases, sessions of expert elicitation at PSI's Center for Proton Therapy are planned. Nevertheless, it is expected that the data from existing methods and from CORE-DATA will not allow the quantification of all decision tree structures. Ranges of likelihood and factor influence importance will be elicited for specific tree branches to complement the information from the existing sources. The design of the expert elicitation sessions is on-going.

It is worth noting that the level of decomposition of the GTTs is chosen such to match the key tasks identified as critical though past incident analyses and literature studies, e.g. checking, software interactors, communication failures. At this rather low level of decomposition, it may also be feasible to use human performance data (from similar tasks in other industries) as well as to collect sector-specific data. Other approaches are possible, for example analyzing tasks at a more holistic level following the ATHEANA (A Technique for Human Event Analysis) method [1]. The ATHEANA approach demonstrated strong potential to produce rich qualitative insights [20, 53]; in ATHEANA, at present, however, quantification of failure probability is not linked to data, but based on expert judgment. For the present work, the more classical HRA framework based on GTTs and PIFs was preferred over the ATHEANA approach also in view of its easiness to use: this may indeed foster its application in a novel domain such as radiotherapy and healthcare in general. Indeed, rich qualitative analysis (possibly following the ATHEANA scheme) shall always accompany any HRA application to produce qualitative insights on performance; this can always be combined with a lower level quantification scheme such as the one under investigation in the present work. This aspect will be further addressed in the future case study application.

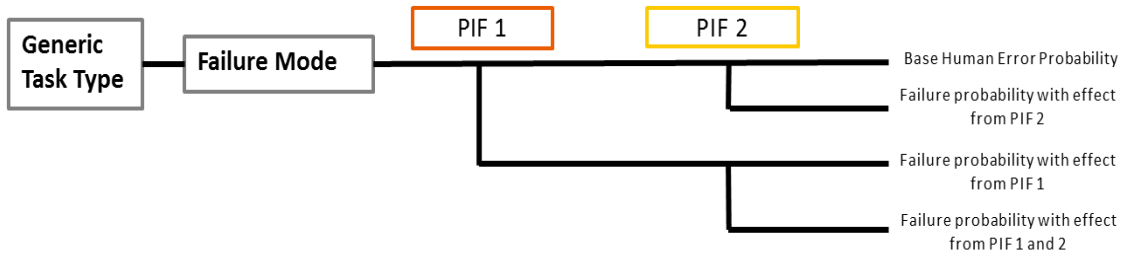


Figure 4: Concept for decision trees to quantify the failure probability of GTT failure modes

2.2.2: The underlying cognitive framework

A novelty of the method under development compared to HEART and its variants is its explicit foundation on a cognitive framework, in particular the one presented in reference [17]. It is worth noting that the need to provide a more direct link between HRA models and cognitive sciences has been raised recently for HRA in general, touching all HRA methods [17]. The cognitive foundation aims at better identifying and characterizing the performance influences: eventually, the goal would be to reduce the uncertainty and variability of HRA results. The need for this improvement gained new momentum following the International and US HRA Empirical Studies, which highlighted that at least part of the variability of HRA results can be attributed to a limited underlying theoretical basis [20, 53].

The cognitive framework developed by reference [17] includes two parts: the cognitive model (**Figure 5)**, consisting of five macrocognitive functions and the formulation of cognitive maps linking the function failure mechanisms to PIFs (**Figure 6**). The five functions are: "*Detecting and Noticing", "Sense-making and Understanding", "Decision-making", "Action" and "Team coordination"*. As **Figure 5** shows, there are no arrows connecting each macrocognitive function, to emphasize that the cognitive flow is complex, dynamic, and multi-directional. The identification of these functions draws on existing cognitive models of literature [54, 55, 56, 57, 58, 59]. Reference [17] highlights the general consensus across the models on the main functions, with some differences on the function boundaries, their aggregation and, most prominently, the representation of how they interact. Reference [17] recognizes that for their purposes it suffices to consider that these functions work largely independently (so that their failure modes can be considered independent), although the cognitive process is strongly parallel and cyclical.

Figure 5: Representation of the interacting macrocognitive functions [17]

The second part of the cognitive framework consists of maps of macrocognitive function failures linking to PIFs. **Figure 6** shows the concept of the mapping; the failure of each macrocognitive function is first linked to the set of possible causes of failure (termed "*Proximate Causes"*); then each "*Proximate Cause"* is linked to the respective "*Failure Mechanism"*, the means by which a function may fail. Reference [17] identifies 13 "*Proximate Causes"*:

Proximate causes for failure of *"Detecting and noticing"*:

- Cue/information not perceived
- Cue/information not attended to
- Cue/information misperceived

Proximate causes for failure of *"Sense-making and understanding"*:

- Incorrect data
- Incorrect integration of data, frames, or data with a frame[4]
- Incorrect frame

Proximate causes for failure of "*Decision-making*":

- Incorrect goals or priorities set
- Incorrect pattern[5] matching
- Incorrect mental simulation or evaluation of options

4 "*A frame encompasses the concepts of a mental representation, a mental model, a story, a map, a schema, a script, or a plan, and serves as a structure for explaining the data and guiding the search for more data*" [17].

5 In sense-making and understanding, the personnel forms a mental model of the situation. Pattern matching relates to the comparison of the mental model with any previously encountered situations in order to judge the situation and devise a plan.

Proximate causes for failure of "*Action*":

- Failure to execute desired action
- Execute desired action incorrectly

Proximate causes for failure of "*Team coordination*":

- Failure of team communication
- Failure in leadership/supervision

The failure mechanisms point to the specific cognitive failures that result in the function failure, for example: failure in attention and working memory, presence of conflicting priorities, cognitive biases. Finally, as typical for HRA, PIFs are the contextual factors that may influence human performance. Each failure mechanism is linked to the PIFs deemed to influence the likelihood of the failure (**Figure 6**).

The framework in reference [17] is intended to provide the foundation for an HRA method for the nuclear power plant domain: an important issue is its applicability to the radiotherapy domain. Indeed, after the review of the reference [17]'s sets of cognitive functions, proximate causes and failure mechanisms, it can be said to have general applicability to human performance in any complex, dynamic domain. Indeed, the psychological, cognitive, human factors and operational research literature reviewed in reference [17] to develop their model is not limited to nuclear power applications. Of course, the nuclear power plant application is evident in the specific characterizations of the macrocognitive functions, proximate causes, failure mechanisms and PIFs. For example, function *"Detecting and Noticing"* generally applies to the process of sensing and perceiving important information in the work environment. In nuclear power plant applications, this predominantly translates in sensing and perceiving control room indications of abnormal plant statues. For radiotherapy applications, basically it applies to sensing and perceiving information relevant for patient treatment, e.g. the tumour spread on computer tomography scans, but also identifying the patient identity and relevant treatment files. As another example, in the characterization of function *"Team Coordination"*, Reference [17] emphasizes the leadership and supervision roles required to coordinate the operating crew response to a situation of possible emergency (as shown by the choice of the corresponding proximate cause). Indeed, in the radiotherapy domain, teamwork manifests in a very different way: patient treatment planning and delivery are distributed across various phases, and so are most of the personnel interactions. Leadership and supervision also play important role,

especially in the coordination of the activity in each phase and in the performance of the several quality checks.

Anticipating more details in **Section 2.3: Methodology for the development of the GTT-PIF structure** and **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy**, the present work applies reference [17]'s cognitive framework to identify the set of PIFs influencing each failure mode characterizing each GTT: the concept is that the GTT-PIF structures should support the identification of GTTs influenced by the same set of factors, with the same impact on the task failure probability. The elements of the cognitive framework (the sets of cognitive functions, proximate causes, failure mechanisms and PIFs) are applied considering the different characterizations that these elements assume in the radiotherapy context. Also, reference [17]'s structures linking macrocognitive function failures to PIFs (depicted in **Figure 6**) are adapted for the specific radiotherapy application: the set of failure mechanisms influencing each proximate cause and the set of PIFs influencing each failure mechanism reflect application-specific influences. Some examples will be presented in **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy**.

Finally, indeed as above mentioned, various cognitive frameworks and models have been proposed in the literature, e.g. Endsley, Klein, O'Hara, Roth etc. along with those traditionally at the basis of the early HRA methods Rasmussen and Reason [54, 56, 57, 58, 59, 60, 61]. The cognitive framework from reference [17] has been used as the basis to develop the GTT-PIF structures, because it systematically links cognitive failure to PIFs via failure mechanisms: as presented in detail in **Section 2.3: Methodology for the development of the GTT-PIF structure** and **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy**, these structures are the basis for the GTT identification subject of the present work. Another option could be to adopt the cognitive framework underlying the Cognitive Reliability and Error Analysis Method (CREAM) [62], which also links cognitive functions to influencing factors (in the form of CREAM's Common Performance Conditions). The implications of using this different framework could be explored in future work.

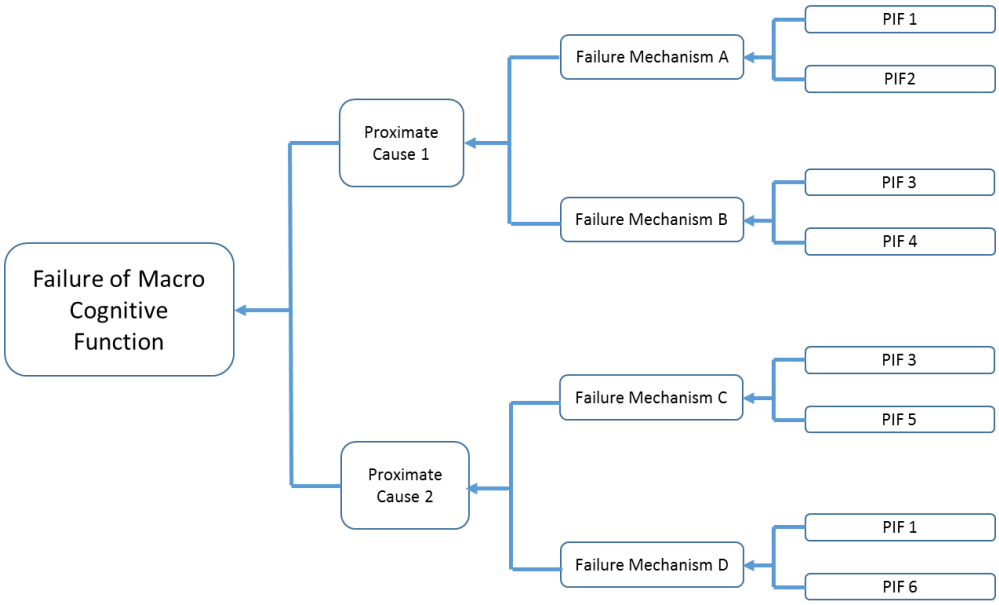


Figure 6: Generic cognitive framework: links between macrocognitive functions, proximate causes, failure mechanisms, and PIFs [17]

2.3: Methodology for the development of the GTT-PIF structure

The proposed methodology comprises of four steps. The first two steps aim at the identification of a comprehensive set of instances of tasks (so-called Example Tasks) to be performed by the personnel, characteristic of the radiotherapy process. This defines the set of tasks to be covered by the GTTs; in the next steps, these tasks are grouped to form the GTTs, based on the type of task, interaction with the system, and the set of influencing PIFs. Specific tasks are first introduced in step one of the methodology, typically at a higher level of task definition. Example Tasks are then defined in step two, including all information relevant to identify the cognitive functions predominantly characterizing each task. The third step applies the cognitive framework of reference [17] to identify, for each Example Task, the relevant failure modes, mechanisms and influencing PIFs. In step four, the Example Tasks are grouped, to define the GTTs. The final result is the set of GTTs, the influencing PIFs and the corresponding structure which connects GTTs and PIFs through failure modes, causes and mechanisms.

The overall steps of the proposed methodology are shown in **Figure 7**.

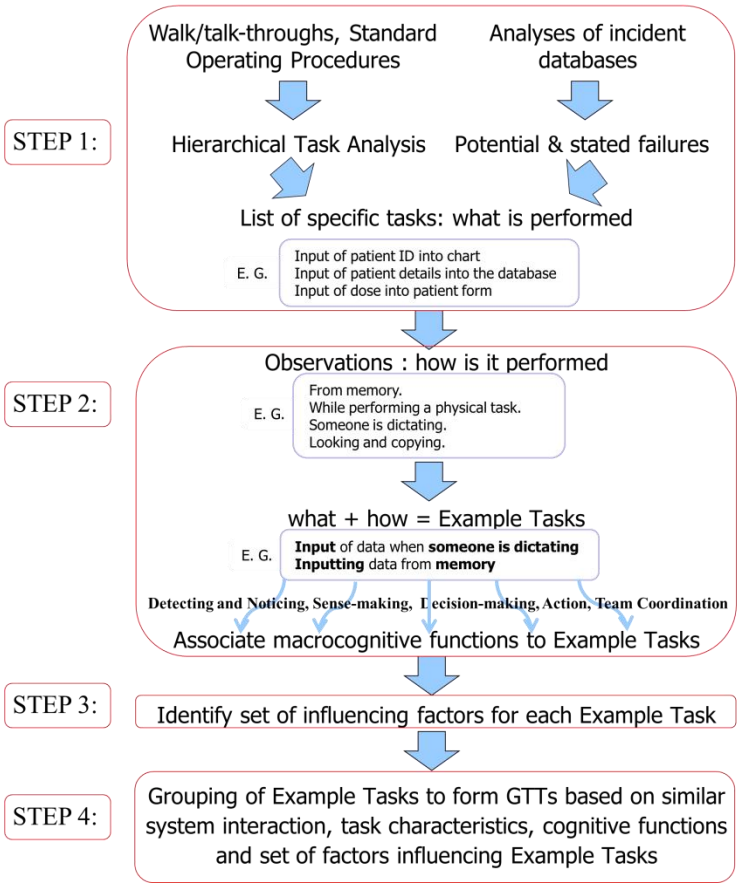


Figure 7: Overview of the methodology for the development of the GTT-PIF structure

Step 1: Identify Specific Tasks

The overall goal of this step is to obtain a collection of the personnel tasks that, if failed, may have safety-relevant implications. The term "specific" underscores the relevance of the tasks for the radiotherapy domain.

The main inputs for the identification of the specific tasks are Hierarchical Task Analyses (HTAs) and analyses of past incidents. As typical for this type of analysis, HTA breaks the radiotherapy process down into its constituent personnel tasks, providing a first list of specific tasks. The list is further refined with the analyses of past incident databases and literature: it is important to make sure that the failed tasks contributing to actual incidents are included in the list [24, 25, 32, 37, 63, 64, 65]. Indeed, this combination allows for completeness and relevance of the GTT definitions. Completeness should come from the HTAs, which address the whole therapy process. The combination with past incident analyses makes sure that the GTT definitions cover failures that actually resulted in undesired events.

Another important comment concerns the use of past incident information. Depending on the level of detail of the incident description, it may not always be possible to identify the specific failed task, at the level of decomposition matching that of the specific tasks from the HTA. In this case, it is possible to make assumptions on which tasks failed, and, eventually, add these tasks to the specific tasks list. This is possible because the goal of this step is to obtain the specific tasks list, not to perform an incident investigation. The failed tasks should be covered in the list, whether these are actual failures or assumed in lack of precise information.

The result of this step is a list of specific tasks related to the radiotherapy patient handling process. In the next steps of the methodology, these tasks will be further specified and grouped to define the GTTs.

Step 2: Derive Example Tasks

In this step, Example Tasks are derived by further detailing the specific tasks. In particular, the information included in the specific tasks obtained from the HTA generally relates to 'what' is performed, not necessarily to the level of detail required to determine the cognitive function involved (which is required to define GTTs consistently, see Step 3 of the proposed methodology). For example, a task such as "*inputting patient ID*" (as it would result from the HTA) could be done from "*memory, reading from a document/screen, or while someone is dictating*" etc. In this step, the information on 'how' tasks are performed is collected such to identify the cognitive functions active during the tasks. Along with the set of specific tasks, this step requires information from the field observations of the treatment process and, possibly, further interviews with the facility personnel.

The second part of this step is to identify for each Example Task the cognitive functions that would be deemed active when the task is performed. Indeed, in general, all macrocognitive functions would be active in a task, but only few would be necessary to characterize the cognitive failure of a task. These dominant cognitive functions are adjudged based on three criteria: the function that would stay active for most of the time; the one that addresses the most challenging parts of the task; the one that, if failed, would lead to the most severe consequences. Concerning the latter criterion, the severity of the consequences is judged based on the actual result of the failure as well as on the ease to detect the failure. Any function that satisfies at least one criterion would be considered as one of the dominant functions.

The dominant cognitive function informs the typical cognitive failure that can be expected when the task fails.

Step 3: Determining influencing factor set for each Example Tasks

The tasks covered by each GTT should be characterized by the same set of influencing factors, with the same quantitative influence. To achieve this, step 3 identifies the factors influencing the performance in each Example Task. This will provide the basis for grouping the Example Tasks into GTTs (in step 4). The process to map each Example Task to the corresponding set of PIFs is shown in **Figure 8**. The mapping follows the cognitive framework presented in reference [17]. For each cognitive function involved in the Example Task, failure modes are identified. Each failure mode is then associated to one or more proximate causes that are applicable from the list in reference [17]; then, failure mechanisms for each proximate cause are identified and finally the PIFs that are deemed to be affecting the failure mechanism are selected. In some cases, the sets of PIFs affecting each failure mechanism were adapted to reflect differences in the application domain. Examples of these will be presented in **Section 2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy**. The final result of this step is a map of Example Tasks and the set of Performance Influencing Factors affecting it.

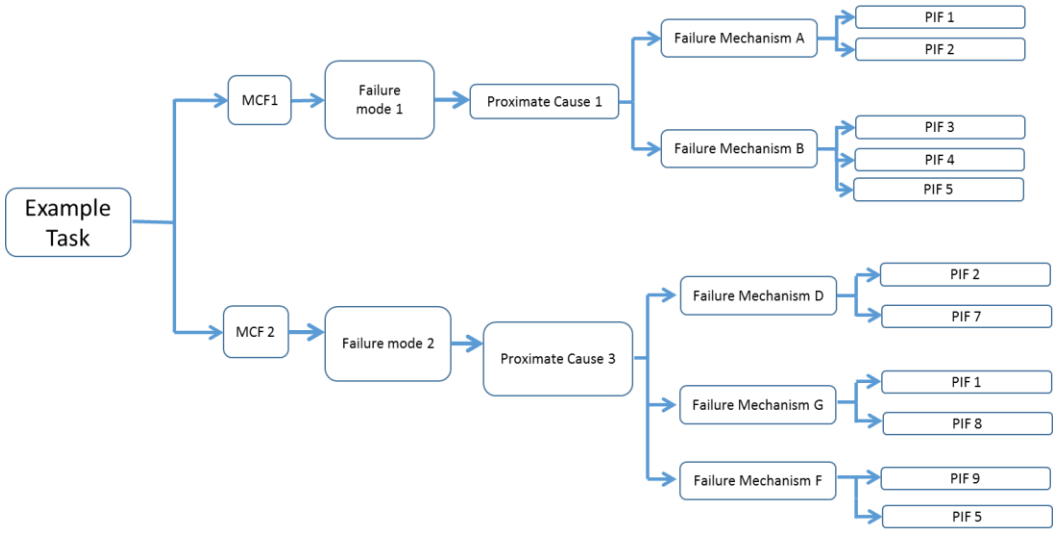


Figure 8: Example Task- Performance Influencing Factors mapping scheme (MCF: macrocognitive function)

Step 4: Formation of Generic Task Types

The overall goals of this step are to form the set of Generic Task Types, each associated with the corresponding set of Example Tasks and to develop the GTT definitions.

The inputs to this step are the Example Tasks obtained with their respective associated cognitive functions and set of influencing factors. The step is performed in two stages. First, the Example

Tasks are grouped based on similarities in the system interaction and task characteristics into an initial set of Generic Task Types. This grouping is based on the specific activity performed by the personnel, as suggested by the Example Task description: for instance, Example Tasks involving planning, identification of items, running software calculations would be grouped into different GTTs. The key criterion for the grouping is that the Example Tasks belonging to the same group would be characterized by very similar cognitive functions and PIF profiles. Based on this, the initial grouping outlines the definition of the GTT by transforming the task description and system interactions into its definition. Then, the groupings are further refined (including the definition) by homogenizing the associated cognitive functions and the set of influencing factors within each group; i.e. certain Example Tasks are moved to other groups, where they would be more applicable: each GTT would then comprise Example Tasks featuring task characteristics, similar system interaction, set of influencing factors and dominant cognitive function.

The result of this step is the final list of Generic Task Types with their definitions and their respective set of influencing factors. In addition, GTTs are backed with the associated dominant cognitive functions, set of failure modes and of Example Tasks.

2.4: Development of the GTT-PIF structure: application of the methodology to radiotherapy

Step 1: Identify Specific Tasks

The documents used for the development of the HTAs were the Centre's Work Flow Analyses and the Standard Operating Procedures. Several sessions of talk-throughs and walk-throughs were conducted, covering the whole patient handling process: this was to make sure that the HTAs reflect the actual practice at the centre.

Eleven HTAs were performed, covering the process phases to be addressed by the HRA method. As an example, **Figure 9** addresses the "*Moulaging and Computed Tomography Scan*" step. Moulaging is the process of designing a device such as a mould or mask to keep the patient in the same position during each treatment session such that the desired location is treated every time. Moulage is divided into six main tasks, namely, "*Anaesthesia check, Immobilization and positioning Checks, Marking of fixation devices or skin tattoos, putting ID on fixation device, and Taking photos of the devices*". These main tasks were further divided into specific tasks; for

example "*immobilization and positioning*" was further broken into "*moulage preparation, fix and adjust the moulage and adjust the fixation devices*".

The level of task decomposition in the HTA was based on the rule presented in reference [12]. Each task was decomposed until the breakdown provides no new useful information about the risk or the failure. For example, task "*Fix and adjust the moulage*" in **Figure 6** could be further broken down into "*shift or rotate the moulage*" and "*press or hold the moulage*" sub-tasks. However, the failure of interest for this application is that the moulage does not immobilize the patient in the desired position: failure of either sub-task would lead to such failure so that further decomposition was deemed unnecessary. The list of specific tasks from the HTAs is given in **Appendix 3**.

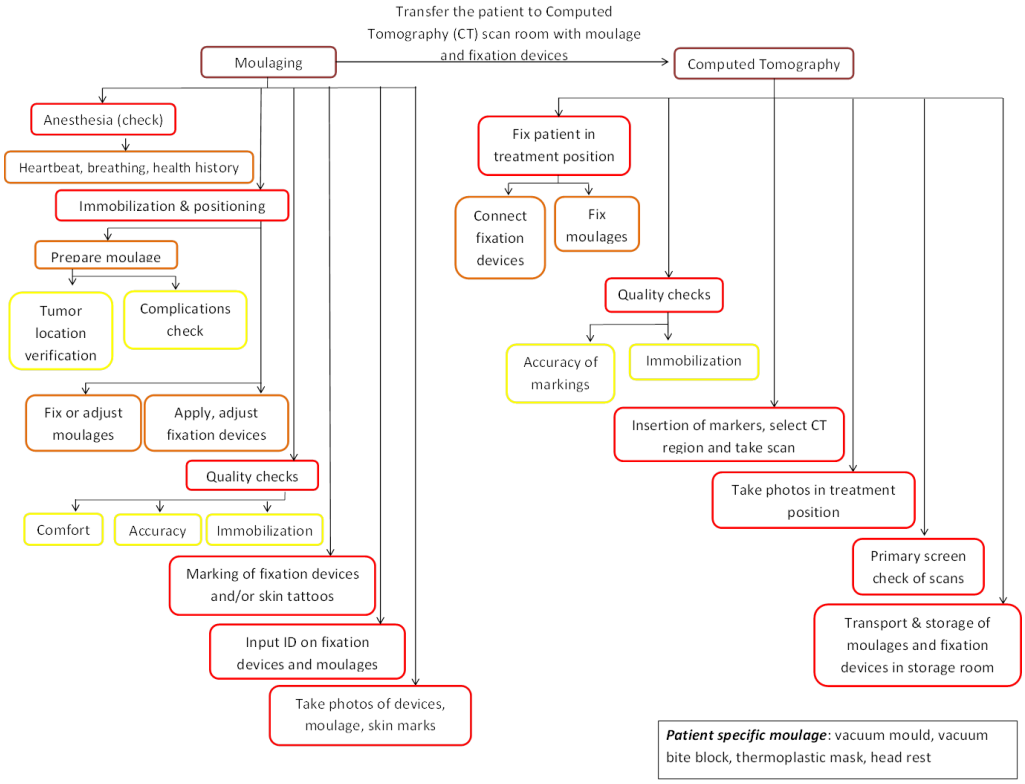


Figure 9: Example of HTA to support identification of the specific tasks: Moulaging and Computed Tomography (CT) scan:

The list was then further populated by analysing past incidents from databases and literature [24, 25, 32, 37, 63, 64, 65, 66]. From the listed failures the associated failed tasks were eventually added to the list of specific tasks. The perspective with which the past incidents are reviewed was to gather information on which tasks did, or may have, failed in the incidents. Indeed, in some cases the incident description was not detailed enough to identify which task had failed. In

such cases also assumed possible manifestations of failures were considered and included into the list of specific task types. The typical case is of failures in administrative checks. These are often not explicitly reported in incident descriptions; however, if in the accident review it was deemed that the course of the accident would have foreseen failed checks, then the corresponding tasks related to performing the specific checks were added to the list.

Step 2: Derive Example Tasks

As mentioned in **Section 2.3: Methodology for the development of the GTT-PIF structure**, the derivation of the Example Tasks requires information on the cognitive function characterizing these tasks. For some specific tasks, the level of detail of the HTA was not sufficient to determine the cognitive function. The needed information was acquired via observations and interviews with the plant personnel. This was generally needed for tasks that may be performed in different ways depending on the situation: typically tasks related to handling data, handling files and tools and quality checks. Examples of these "what + how" combinations are reported at the top of **Figure 10**: these combine what needs to be done (e.g. "*Inputting of patient data into patient positioning and verification software*"), with how it is performed (e.g. "*looking and copying from Therapy Planning System (TPS)"*.

Certain "what + how" combinations were grouped into a (more generic) single form Example Task. As an example, tasks like "*inputting the patient ID into Patient Positioning and Verification software reading from patient Therapy Planning System (TPS)"* and "*inputting the patient ID into patient chart while reading it from a document on the computer"* are grouped into the single Example Task "*Looking and copying/Inputting Data",* bottom of Figure 7. As another example, tasks like "*Detailed quality checks of data within High Resolution CT and CT*" and "*Detailed Quality assurance of plan*" are grouped into a single Example Task "*Detailed check of data within documents*". In other words, the Example Tasks define the activity performed by the personnel both in terms of interaction with the system/environment (e.g. inputting data, comparing two values) as well as detailing the key mental activity (e.g. looking and copying, understanding and noticing an error).

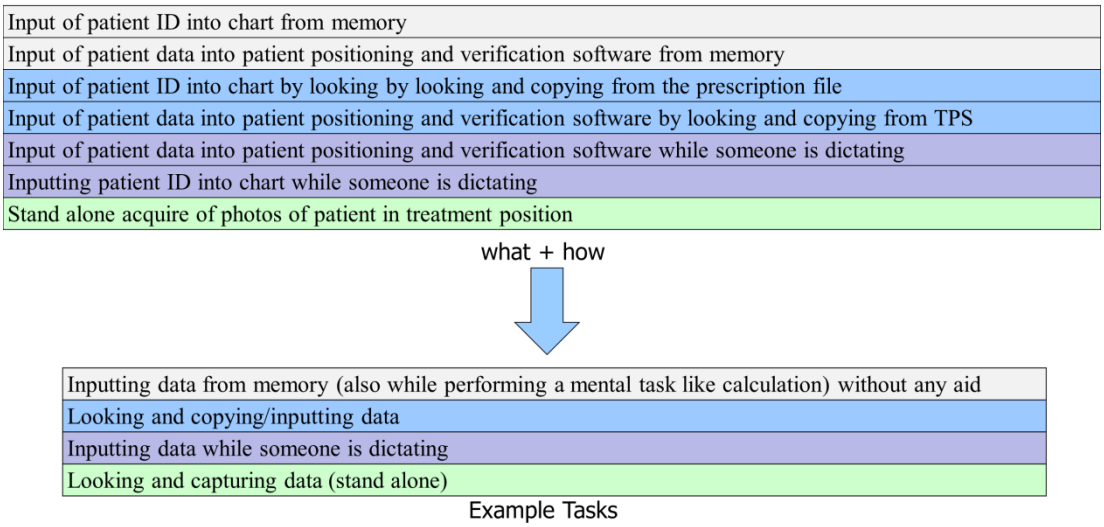


Figure 10: Formation of Example Tasks

Each Example Task is then associated the dominant cognitive function, adopting the cognitive model presented in reference [17] and based on the criteria mentioned in **Section 2.3: Methodology for the development of the GTT-PIF structure** (description of Step 2). As shown in **Table 1**, for some Example Tasks, more than one dominant function has been identified: this is because it was judged that the task is characterized by important aspects of multiple functions (as discussed in **Section 2.3: Methodology for the development of the GTT-PIF structure**, any macrocognitive function verifying at least one criterion would be considered as part of the dominant ones).

As example application of these criteria, consider the set of Example Tasks (from the HTAs "*Volumes of interest*" and "*Treatment planning*"), reported in **Table 1**. Task 1 is related to identification of the file from a set of files and would require to detect the desired ID number while reading it. This task is straightforward: the dominant cognitive function is "*Detecting and noticing"*. Relating it to the criteria for selection of the dominant cognitive function: (1) task is to actively seek the concerned file therefore "*Detecting and noticing"* will be active for more time; (2) noticing the cue or the information of the file while going through a list of files is more challenging than deciding the file based on reading the ID; (3) failure in "*Detecting and noticing"* may result in the wrong treatment file being used which is the negative outcome of interest for the analysis.

In other cases, more than one function was deemed as dominant: consider, for example, Task 7 of **Table 1**, related to deciding the plan mode, beam directions, fields etc. for the tumour. This

task requires understanding of the tumour, its location etc. and then deciding the type of plan, number of fields etc. to develop the plan. Thus, both, "*Sense-making and understanding"* and "*Decision-making"* are active in this task. Indeed, depending on the tumour type and location one of the two functions may take prominence. For example, in case of simple tumour locations, the personnel would most likely directly decide the treatment plan, without much need for "*Sense-making and understanding".* On the other hand, for complex tumour locations possibly closed to healthy vital organs, the decision would be preceeded by significant activity of "*Sense-making and understanding"*.

The complete list of Example Tasks identified and associated macrocognitive functions is presented in **Appendix 4**.

Table 1: Example Tasks and associated cognitive functions (derived from HTAs 'Volumes of interest' and 'Treatment planning')

#	Example Tasks	Dominant Macrocognitive Functions
1	Identification of file from a list of files while reading the ID	Detecting and noticing (D/N)
2	Identification of file by remembering the ID (memory based)	Sense-making and understanding or D/N
3	Opening the file	Action (A)
4	Drawing contours around tumour and OARs (organs at risk) for every slice (mechanical task)	Sense-making and understanding (SM/U) or Action (A)
5	Assigning names and keys to VOIs (volume of interest) and OARs	Action (A)
6	Export of data from one file to another (electronic)	Action (A)
7	Decide and select plan mode, beam directions, fields	SM/U or Decision-making (DM)
8	Iterative calculation of doses	DM
9	Check of clinical limits or data	D/N
10	Detailed Discussion with medical doctor and medical physicist	D/N and Team coordination

Step 3: Determining influencing factor set for each Example Task

In this step, each Example Task is mapped to the corresponding set of PIFs, to prepare for grouping the Example Tasks into the GTTs (Step 4). The PIF hierarchy used in this step to present the results is presented in Table 2 and is based on reference [19].

As an example, Figure 11shows the process for the Example Task "*Check of transferred data to a software/machine*". Four failure modes have been identified: the decision is made not to perform the quality check (involving the "*Decision-making*" cognitive function); the check is not performed, involuntarily (e.g. it is forgotten, a slip following a procedure, involving the *"Action"* cognitive function); the check is performed but the deviation from the requirement is not noticed

or recognized (*"Detecting and noticing"*) or slips occur during the check e.g. a checklist is not correctly gone through, involving the *"Action"* cognitive function.

Then, each failure mode is further specified by the expected failure causes, mechanisms and affecting PIFs. For example, for failure mode "*deviation from the requirement is not recognized*", the cause "*Misperception of information*" is identified as relevant. As for the failure mechanisms, the following are deemed as relevant: poor "*cue/information content"*, "*distraction* (*vigilance-attention*)", and "*low vigilance due to expectation"* (i.e. expectation that the information to be checked is correct). The affecting PIFs were identified as "*Training-Experience, Resources, HMI, Personal, Loads and Environment"*.

Note that by doing so, the PIFs influencing each failure mechanism are incorporated in the final method itself. For example, from **Figure 11** the failure mechanism "cue content" points to failures caused by how the information is presented: the factors that were identified as relevant for this failure mechanism are "*Resources*" and "*HMI*". These will be incorporated in the decision trees developed for each failure mode (ref to **Section 2.2: Main elements of the HRA method under development**). Indeed, in the end, many PIFs (possibly all) affect the overall task, depending on the context; this is visible from **Figure 11** where all PIFs affect the example task if one considers all failure mechanisms. The specific influences of the context are represented in the PIF evaluations, which determine the applicable failure probability through the developed decision trees (**Figure 4**).

Generally, as discussed in **Section 2.2.2: The underlying cognitive framework**, the taxonomies provided by reference [17] were found to be directly applicable to the radiotherapy domain: at their definition level, the sets of cognitive functions, proximate causes, failure mechanisms and PIFs are of generic nature. Indeed, domain-specific aspects characterize each element (see **Section 2.2.2: The underlying cognitive framework**). Also, the developed structures linking macrocognitive function failures to PIFs reflect specific influences. Some examples regarding the different characterization of the macrocognitive functions were already presented in **Section 2.2.2: The underlying cognitive framework**. Here, other examples of differences compared to reference [17] are presented: in the developed structures and in the characterization of the PIFs. Consider again the case presented in **Figure 11** concerning a check of transferred data. The Proximate Cause "*cue/information misperceived*" addresses the fact that the quality check does not recognize the presence of a deviation from the requirement (e.g. the dose in the patient chart is not the same as in the control document (i.e. the prescribed by the doctor). The structure

presented in reference [17] identifies five cognitive failure mechanisms affecting this proximate cause:

- Failure of attention – missing a change in cues
- Failure due to cue content – cues too complex ("*cue content*" in **Figure 11**)
- Failure of vigilance in monitoring – divided attention ("*vigilance – attention*" in **Figure 11**)
- Failure due to expectation – mismatch between expected and actual cues ("*expectation*" in **Figure 11**)
- Failure of working memory – working memory capacity overflow

As shown by **Figure 11**, two of the above mechanisms were not considered to be applicable for the specific task: failure of attention – missing a change in cues, and failure of working memory – working memory capacity overflow. Concerning the former, attention to changes in the cues over the course of an event is indeed an important cognitive activity in control room settings to deal with the evolution of an accident. This was not deemed relevant for the specific application because the information subject of the quality check does not evolve during the check itself. The latter also refers to a situation relevant to control room setting in which the operating crew may need to deal with large amounts of information (alarms, indications, plant parameter trends) accumulating over time and possibly imposing on the capacity of the operator working memory. Again, this situation was not deemed to be relevant for the present application: during the check, generally, the personnel do not have to mentally process large amount of information such to possibly reach the working memory capacity.

As said above, in some cases the specific influences relevant for the PIFs had to be adapted from reference [17]. An important example is the case of PIF "*Environment*". For the proximate cause "*cues/information misperceived"* (for instance relevant to characterize the failure of cognitive function *"Detecting and Noticing"* from the example in **Figure 11**), reference [17] emphasizes the possible effect of background noise on verbal communication, accounting for how humans recognize the words in environments and, for example, how background noise can possibly affect information perception. On the other hand, when considering the radiotherapy domain, the specific influence of PIF "*Environment*" in form of distractions and interruptions was observed as important negative performance contributor in past event analyses and databases [24, 25, 32, 37]. Indeed, their relevance for patient safety has been recognized for the overall healthcare sector [67]; recently, William and Bell have added these as a new EPC to the HEART method [68]. To highlight this difference in the specific influence, PIF "*Environment*" is marked with

symbol '○' in the structure shown in **Figure 11**. The above discussion is intended to give an idea of the adaptations required to apply reference [17]'s cognitive framework to the specific domain. A systematic comparison of the PIFs in the two domains is considered outside the scope of the present paper. Definitions of the PIFs will be developed in combination with the development of the decision trees supporting quantification (see **Section 2.2: Main elements of the HRA method under development**).

Finally, for each Example Task the final set of PIFs is collected from all its failure modes and mechanisms.

Table 2: PIF hierarchy used for GTT-PIF structures

Training-Experience	**Safety Culture**	**Resources**	**Team**	**Personal**	**HMI**	**Loads**	**Complexity**	**Environment**
Training, Knowledge, Experience, familiarity with situation	Organizational culture, Management activities, work processes	Procedures, Tools, Necessary information	Communication, Coordination	Fatigue, Bias, Morale, motivation, Physical abilities	Machine interface, System response	Workload, Time pressure, Time load, Other loads, Stressors	Task complexity	Distraction, Interruption, Noise, Lightning etc.

Example Task	Macrocognitive function	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Check of transferred data to a software/machine	Detection and noticing	Deviation from requirement not recognized	cues/information misperceived	cue content			√			√			
				vigilance- attention	√					√	√		o
				expectation	√				√				
	Decision-making	Check not performed (decision-based)	incorrect goals	goal conflict	√	√			√		√		
				incorrect goal selected	√	√			√		√		
				incorrect prioritization of goals	√	√			√		√		
				Incorrect judgment of goal success	√	√			√		√		
	Action	Execute desired action incorrectly	execute desired action incorrectly	motor learning	√						√		o
				automaticity control	√		√	√		√	√		o
				recognition error			√	√		√	√		o
		Failure to execute desired action (error of omission)	failure to execute desired action (error of omission)	divided attention							√		o

Figure 11: Example Task-PIF mapping structure- Check of transferred data to a software/machine. Symbol o: factors requiring modification compared to the original framework of [17]

Legend: MCF- Macrocognitive Function, TR-EX- Training-Experience, SC- Safety Culture, RE- Resources, TE- Team, PE- Personal, HMI- Human Machine Interface, Lo- Loads, Co-Complexity, ENV- Environment

<u>Step 4</u>: Formation of Generic Task Types

The last step is to form the Generic Task Types from the set of Example Tasks. As an example, **Table 3** shows an initial grouping of all the Example Tasks that relate to identification: of resources, patients etc. then, when the involved cognitive functions and set of influencing factors were considered, some Example Tasks of **Table 3** were moved to a different GTT group (to '*simple interaction with software or tools'*). In particular, these tasks included "*identification of patient using electronic devices*", such as "*barcode, or using magnetic strip, or embedded chips*". Indeed, because of the support from these devices, these tasks were judged to be better characterized by the cognitive function "*Action*" (A), as opposed to "*Detecting and noticing*" (D/N) which characterizes the rest of the Example Tasks of **Table 3**. The final list of Example Tasks characterizing the GTT "*Identification of patient or patient related items*" is shown in **Table 4**. Note that all Example Tasks in the group (**Table 4**) involve matching a value/code/image to identify whether a person, an object, or a value is the correct one. The way in which the initial approach to the person, object or value may be different, as exemplified by the first part of the Example Task description (with verbs such as call, ask, look), but for all tasks the key cognitive activity is related to the matching (therefore involving the Detecting and Noticing cognitive function).

Based on the Example Tasks and the general expected features of the overall GTT, its definition is then intuitively developed as "*identification of either the patient to be treated or the patient-specific items (like tools, moulages, files, charts etc.) required for the radiotherapy treatment process. The identification task is done by comparing the specified item like ID number, patient's face or any other unique feature of the patient (e.g. birthmark, patient full name etc.) with the number, photograph or the feature on the control document. The identification of patient takes place as many times in the process as the patient comes to the institution; starting from the first-time patient comes for Planning Computed Tomography scan till the delivery of the last fraction. The identification of patient's items occurs throughout the process whenever they are required*"

Table 3: Initial Grouping of Example Tasks into GTT – example of identification-related Example Tasks

Generic Task Types	**Example Task**
Identification of patient or patient related items	Calling name out loud (in front of a group) and matching with some ID
	Ask for the name to the patient and match with the ID card.
	Checking the patient-specific parameters like patient Identification number, weight, birthmarks etc.
	Look at the photograph and match
	Identification of file or tool from a list by looking at the ID

	Identification of file or tool by remembering the ID (memory based)
	Identification of file or tool while person is reading out the ID
	Identification of the patient using barcodes
	Identification of the patient using chips in the wrist of the patient
	Personalized magnetic strip cards based identification

Table 4: Final grouping of the Example Tasks into the GTT Identification of patient or patient related items

Generic Task Types	**Example Task**	**Macrocognitive Function**
Identification of patient or patient related items	Calling name out loud (in front of a group) and matching with some ID	Detecting and Noticing (D/N)
	Ask for the name to the patient and match with the ID card.	D/N
	Checking the patient-specific parameters like patient Identification number, weight, birthmarks etc.	D/N
	Look at the photograph and match	D/N
	Identification of file or tool from a list by looking at the ID	D/N
	Identification of file or tool by remembering the ID (memory based)	D/N
	Identification of file or tool while person is reading out the ID	D/N

As another example, Example Tasks that relate to input or transfer of data etc. were grouped into "*Simple interaction with software or tool*" GTT, see **Table 5**. The dominant cognitive function for this type of GTT is "*Action*", **Appendix 5**. No refinement was required for this GTT as the grouping was straightforward and the Example Tasks fully met the criteria of task description, similar cognitive function and similar PIF profiles affecting the tasks. For the definition of the GTT see **Appendix 6**.

The other Example Tasks were grouped following a similar process, obtaining the set of 6 GTTs and corresponding Example Tasks presented in **Table 5**. The final GTT-PIF structure with failure causes, mechanisms and PIFs is presented in **Appendix 5**. For the detailed definitions of GTTs with descriptions refer to **Appendix 6**. Sub-groups within each GTT were formed for convenience of the presentation and are shown in **Table 5**.

Concerning the GTT "*Quality Check*", the subgrouping is intended to distinguish between short and detailed check, which is aimed to capture difference in the depth and the scope of checking tasks, and thus the effectiveness of the checks. The additional cognitive function in a "*detailed check*" case is "*Sense-making and understanding and Team coordination*" and the PIFs included in this would model the variation (**Table 5**). The aspect of dependency between checkers, as

analysed for example in A Technique for Human Error Rate Prediction (THERP), will be considered and modelled at the quantification stage of the method [51].

Indeed, the list of Example Tasks in **Table 5** is not exhaustive. For example, in GTT "*Identification of patient or patient related items*", a different Example Task could be the combination of the 1st and 4th Example Task, i.e. "*calling the name out loud and matching photograph*". Furthermore, for GTT "*Complex interaction with software or tool*" GTT, a different Example Task could be formed as "*Positioning the patient with photos while adjusting the patient*". Indeed **Table 5** is not meant to cover all possible variants of personnel interactions, but reference tasks to better define each GTT.

Table 5: GTTs with their Example Tasks and macrocognitive functions

Generic Task Types	**Further distinction- GTT Subgroup**	**Example Task**	**Macrocognitive Functions**
Identification of patient or patient related items	Patient Identification	Calling name out loud (in front of a group) and matching with some ID	D/N
		Ask for the name to the patient and match with the ID card.	D/N
		Checking the patient-specific parameters like patient Identification number, weight, birthmarks etc.	D/N
		Look at the photograph and match	D/N
	File/tool identification	Identification of file or tool from a list by looking at the ID	D/N
		Identification of file or tool by remembering the ID (memory based)	D/N
		Identification of file or tool while person is reading out the ID	D/N
Quality Check	Short quality check	Check of transferred patient data to a software/ machine	D/N
		Air gap check	D/N
		Check of clinical limits or data	D/N
		Immobilization and positioning checks	D/N
		Two personnel read aloud cross check	D/N
	Detailed Quality Check	Detailed check of the plan by doing back calculation	D/N, Sense-making and understanding (SM/U)
		Detailed check of data within documents (e.g. CT scans)	D/N, SM/U
		Detailed check of the plan between doctor, medical physicist and planner	D/N, SM/U, Team coordination
Iterative determination of optimum parameter	Treatment Plan Calculation	Deciding and assigning plan mode, beams, directions and other parameters like beam energy, iso-centers	SM/U, Decision-making (DM)
		Performing iterative calculation of doses	DM

Generic Task Types	Further distinction- GTT Subgroup	Example Task	Macrocognitive Functions
		Hand calculation of dose or weights of each field Or any other calculation	SM/U, DM
		Consider organ motion effect and other special effects	SM/U, DM
	Planning the patient immobilization and positioning	Moulage preparation (with planning)	DM
		Decision on devices like head support, table type etc.	SM/U, DM
		Adjusting the immobilization devices	DM
		Marking tattoos on skin	SM/U
		Planning positioning using laser coordinates	SM/U
Complex interaction with software or tools	Replicating the patient immobilization and positioning	Positioning with photos, scans or any other document	SM/U, Action (A)
		Fixing the prepared moulages (matching the marks)	SM/U, A
		Positioning with tattoos on skin	SM/U, A
		Using laser coordinates to position	SM/U, A
	Standalone Selection/ capture/ acquire of data	Looking and capturing data (with understanding what is required)	SM/U, A
		Selecting data while simultaneously performing some mental task	SM/U, A
		Acquiring data while someone is dictating	SM/U, A
		Contouring of Volume Of Interests (VOIs) like tumour or Organs At Risk (mechanical task)	SM/U, A
Simple interaction with software or tool	Input/ entry of data	Simple looking and copying/inputting data	Action
		Simple input of data from memory (also while performing a mental task like calculation) without any aid	Action
		Simple data input while someone is dictating (purely executing the task)	Action
	Transfer of data	Import or export of data to another location (electronically or manually)	Action
	Patient Identification using machine	Patient identification using barcodes	Action
		Patient identification using chips in the wrist of the patient	Action
		Personalized magnetic strip cards based identification	Action
	Selection and transfer file or tool	Selection/ pick up of chart/file or tool	Action
		Transfer of the file or tool (including storing)	Action
Verbal Communication	Verbal Communication	Give instructions	Team coordination
		Give clinical or identification data / information	Team coordination
		Any kind of update	Team coordination

2.5: Validation of the GTT-PIF structure against Huq et al. [38]

This section presents an attempt to validate, at least partially, the developed GTT-PIF framework. The validation considers recent work by the American Association of Physicists in Medicine (AAPM) to provide guidance on the application of a combination of FMEA-FTA techniques to radiotherapy, in support of quality management programs for specific clinics [38]. Field experts belonging to various clinics developed a generic FMEA and FTA for "Intensity modulated radiation therapy". The concept for validation has been to map the process tasks, failure modes and potential causes identified in reference [38] to the GTT-PIF structure developed in the present work: a correspondence would indicate that the GTT-PIF structure addresses tasks and influencing factors relevant for the specific application. It is important to mention that the FMEA from reference [38] was not used to develop the GTT-PIF structure: in this sense, it represents new information to assess the coverage of the GTT-PIF structure.

As an example, the mapping of one process task failure mode from reference [38] is presented in detail in **Table 6**. Task number 31 of reference [38] addresses imaging of the clinical target volume: this volume covers the visible tumor volume plus the possible extent of microscopic tumor spread i.e. the primary tumor and its surrounding tissues [38]. According to the FMEA in reference [38] (see **Table 6**), one associated failure mode is the incorrect interpretation of this treatment volume; the potential causes identified by reference [38] are inadequate training and lack of communication (the latter intended as missing necessary information as opposed to failures in verbal communication). When mapping this task to the developed GTT-PIFs structure, the GTT chosen for the task is "*Complex interaction with software or tool*", as the task involves understanding of the tumor and of its surrounding volume before the image can be taken, interacting with the software in use (Example Task "*Looking and capturing data*" represents well this task, see **Table 5**). As shown in **Table 6**, the failure mode "*misinterpretation of data*" was found to correspond to the one identified in the FMEA (Incorrect interpretation of tumor or tissue for images). In this context, the "data" is the tumor and the surrounding tissues, such that the clinical target volume can be identified and captured in the imaging.

The factors identified in reference [38], inadequate training and lack of communication, are both covered in the developed GTT-PIF structure, by factors "*Training / Experience*" and "*Resources (necessary information)*", respectively. As shown in **Table 6**, "*Training / Experience*" influences the two cognition-related failure causes associated to the failure of the "*Sense-making and*

understanding" cognitive function(6): "*Incorrect frame*" and "*Incorrect integration of data and frames*". In this case, for example considered in **Table 6**, incorrect frame may represent the lack of knowledge that a specific type or location of tumor would possibly spread in certain body locations or that delivering dose in a specific location may involve the presence of organs at risk. As another example, "*incorrect integration of data and frames*" may represent that, although the therapist has the above mentioned knowledge, he/she fails to match the case under analysis to those specific tumor type and location. Similarly, "*necessary information*" in the "*Resources*" PIF influences the data-related failure cause associated to the failure of "*Sense-making and understanding*" cognitive function[6]: "*Incorrect data*". In this case, for example, if the necessary information like patient-specific tumor location, its size, spread and other clinical details etc. is not provided to the personnel performing the imaging, then the lack of necessary information ("incorrect data" as mentioned in reference [17]) may lead to incomplete or incorrect understanding of the situation and thus incomplete or incorrect imaging of the required clinical target volume.

The FMEA cases involving human failures with the highest severity and highest Risk Priority Number (RPN) were selected for validation of the GTT-PIF structure. Overall, about 20 cases were selected; **Table 7** shows the summary of the matching for the top 10 RPN cases (Full table in **Appendix 7**). As shown by the analysis of case 31 above, a failure mode is influenced by a PIFs via different pathways over the GTT-PIF structure (e.g. in **Table 6** "*Training / Experience*" influences the failure mode "*misinterpretation of data*" through the two failure mechanisms "*Incorrect frame*" and "*Incorrect integration of data and frames*"). For brevity, for each pair of failure mode and potential cause in the FMEA, **Table 7** shows only one of the possible pathways through the failure mechanisms. Overall, it can be seen from **Table 6** that in practice all the potential causes from reference [38] were mapped to the produced GTT-PIF structures at the PIF level. The recurrent potential causes such as "*Inadequate training*", "*lack of standardized procedures*", "*inadequate design specification"*, "*lack of staff*" and "*Lack of communication (intended as missing necessary information)*" etc. can be easily mapped to PIFs

6 The model adopted in reference [17] to represent "*Sense-making and understanding*" is Klein's data-frame theory [17, 69]. Data is the information coming into the sense-making process; the data is integrated with an existing frame, which is a mental representation that serves as a structure for explaining the data and guiding the search for more data [69]. "A frame encompasses the concepts of a mental representation, a mental model, a story, a map, a schema, a script, or a plan, and serves as a structure for explaining the data and guiding the search for more data"[17]. Accordingly, three primary sources of failure are identified for this cognitive function: the data (e.g., wrong information is used), the frame (e.g., an incorrect model to understand the situation is used), or the integration of the two (e.g., new information is not properly integrated with the frame) [17].

"*Training/Experience*", "*Resources (Necessary information)*", "*Human machine interface*" and "*Loads (Workload, time pressure)*" and "*Resources (Necessary information)*", respectively.

For some cases, potential causes were not matched to PIFs, but to failure mechanisms or to other GTTs. For example, potential cause "*inattention*" from reference [38] case 58 is rather a failure mechanism in the GTT-PIF structure (**Appendix 7**) as opposed to a PIF: indeed, for the purposes of a human reliability analysis, inattention is regarded as the result of other underlying causing factors, such as loads and environment, as exemplified in **Appendix 7** [38]. Other potential causes from reference [38] instead relate to failures in other task, e.g. "*failure to review work*", "*tool used incorrectly*" or "*incorrect procedure used*" (e.g. cases 59 and 137) [38]. In the GTT-PIF structure, these would correspond to failures in other task types: indeed, the fact that these could be matched to other GTTs gives further assurance of the comprehensive coverage of the proposed structure. For example, in case no. 59 "*tools incorrectly used*" will be assigned to "*Simple interaction with software or tools*" GTT and in case no. 126 "*incorrect final prescription*" will be assigned to "*iterative evaluation of optimum parameters*" GTT.

The above discussed mismatches in mapping failure causes originate from differences in the aims of the FMEA and GTT-PIF structure. The "potential cause" entry of the FMEA is intended to identify any negative condition potentially causing the failure, independently on whether this reflects the result of another failure, the way in which the failure occurs, or the underlying factor influencing the performance: the important element for the FMEA is that the "potential cause" identifies some direct improvement in the process to decrease the chances for occurrence of the failure. The proposed GTT-PIF structure instead aims at a comprehensive and structured identification of the failure causes and influencing factors: in this aim, it is instead important to distinguish the different levels at which failure modes, mechanisms and factors operate. Indeed, for the purposes of the present validation, it was deemed as important that the failure causes from reference [38] are actually covered by the GTT-PIF structure, regardless of the level at which this occurs.

Table 6: Mapping the developed GTT-PIF structure to case 31 of [38]

FMEA from Huq et al. [38]				**GTT-PIF structure proposed in the present work (Appendix 5)**			
No	**Major Process**	**Failure Modes**	**Potential causes (factors)[(1)]**	**Corresponding GTTs (Failure Modes)**	**Proximate Cause**	**Failure Mechanism**	**Corresponding PIF[(2)]**
31	Other pre-treatment imaging for Clinical Target Volume localization	Incorrect interpretation of tumor or tissue for images	Inadequate training (user not familiar with modality)	Complex interaction with software /tool (misinterpretation of data)	Incorrect frame[(3)]	Incorrect or inappropriate frame used to interpret information	Training/ Experience
						Incorrect or inappropriate frame used to attend to information	
					Incorrect integration of data and frames	Improper integration of frames/information	Training/ Experience
						Incorrect frame selected for data comparison	
						Improper control of attention	
			Lack of communication (intended as missing necessary information)		Incorrect data	Information available in the environment is not complete or incorrect or otherwise sufficient to understand the situation	Resources (necessary information)

(1) One of the causes mentioned by reference [38] for many failure cases is "*Failure to review work*": in the presented GTT-PIF structure this corresponds to a GTT failure rather than an influencing factor (specifically, it would be covered in the "Quality Check" GTT). This failure cause is not reported in the above table for brevity [38].

(2) The adopted PIF hierarchy is presented in **Table 2**

(3) "A frame encompasses the concepts of a mental representation, a mental model, a story, a map, a schema, a script, or a plan, and serves as a structure for explaining the data and guiding the search for more data". See text in Section 5 for an example interpretation of frame in this context [17].

Table 7: Summary of validation of the GTT-PIF structure against FMEA for radiotherapy [38]

FMEA from Huq et al. [38]			**GTT-PIF structure proposed in the present work (Appendix 5)**	
Major Process	**Failure Modes**	**Potential causes (factors)[1, 2]**	**Allotted GTTs (Failure Modes)**	**Corresponding PIF[3]**
Other pre-treatment imaging for Clinical Target Volume localization	Incorrect interpretation of tumor or tissue for images	Inadequate training (user not familiar with modality)	Complex interaction with software /tool (misinterpretation of data)	Training/ Experience
		Lack of communication (intended as missing necessary information)		Resources (necessary information)
Radiation Treatment Planning anatomy	3*sigma contouring error: wrong organ, wrong site, wrong expansions	Lack of standardized procedures	Complex interaction with software /tool (misinterpretation of data)	Resources (procedures)
		Inadequate design specification,		Human Machine Interface
		inadequate assessment of operational capabilities		Training/Experience
		Lack of staff (rushed process, lack of time, fatigue)		Loads (Work load, time pressure)
Initial Treatment planning Directive (from Medical doctor)	Wrong summary of other treatments, other treatments not documented	Lack of staff (rushed process, lack of time, fatigue)	Iterative determination of optimum parameters (misinterpretation of information)	Loads (Work load, time pressure)
		Lack of communication, wrong information obtained, information not available		Resources (necessary information)
		Wrong reconstruction of previous events		This is not a PIF but a failure mechanism[4]
Radiation Treatment Planning anatomy	Excessive delineation errors	Lack of standardized procedures, availability of defective tool	Complex interaction with software /tool (misinterpretation of data)	Resources (Procedures, tools)
		Inadequate design specification,		Human Machine Interface
		Lack of staff (rushed process, lack of time, fatigue)		Loads (Work load, time pressure)
		Inadequate assessment of operational capabilities, Inadequate assessment of tool for task, inadequate training, *tool used incorrectly*[5]		Training/ Experience
Radiation Treatment Planning anatomy	Margin width protocol for PTV is inconsistent	Lack of standardized procedures, lack of communication (intended as missing necessary information)	Complex interaction with software /tool (failure to execute desired action)	Resources (Procedures, necessary information)
		Lack of staff (rushed process, lack of time, fatigue)		Loads (Work load, time pressure)
		Inadequate training		Training/ Experience

FMEA from Huq et al. [38]			**GTT-PIF structure proposed in the present work (Appendix 5)**	
Major Process	**Failure Modes**	**Potential causes (factors)[1, 2]**	**Allotted GTTs (Failure Modes)**	**Corresponding PIF[3]**
Plan approval	Bad plan approved	Miscommunication	Quality Check (Execute desired action incorrectly)	Team (communication)
		Lack of procedures, inadequate procedures	Quality Check (deviation from requirement not recognized)	Resources (Procedures)
		Inadequate training, *incorrect procedure used*[5]		Training/ Experience
		Procedures not followed	Quality Check (Check not performed)	Safety Culture, Loads (Work load, time pressure)
Day N treatment	Special motion methods not applied or incorrectly applied	Poor software or hardware design	Complex interaction with software /tool (misinterpretation of data)	Human Machine Interface
		Inadequate training	Complex interaction with software /tool (Mismatch inconsistency not recognized)	Training/ Experience
Initial treatment planning directive (from Medical Doctor)	Special instructions not given or wrong instructions given	Documentation not there, lack of communication	Iterative determination of optimum parameters (Misinterpretation of information)	Resources (Necessary information)
		Lack of staff (rushed process, lack of time, fatigue)		Loads (Work load, time pressure)
Treatment planning	Inadequate evaluation of plan	Not enough time	Quality check (Check not performed)	Loads (Work load, time pressure)
		Inadequate training, poor evaluation strategy, *incorrect final prescription*[5]		Training/ Experience

(1) One of the causes mentioned by reference [38] for many failure cases is "*Failure to review work*": in the presented GTT-PIF structure this corresponds to a GTT failure rather than an influencing factor (specifically, it would be covered in the "Quality Check" GTT). This failure cause is not reported in the above table for brevity.

(2) A recurrent potential cause in reference [38] is "*inattention*", this has been mapped to the "*vigilance*" failure mechanism, see **Appendix 7** (not reported here for brevity as it applies to all cases).

(3) The adopted PIF hierarchy is presented in **Table 2**

(4) This has been mapped to failure mechanism "*mental manipulation of the information is inadequate, inaccurate or otherwise inappropriate*".

(5) These causes as presented by reference [38] correspond to GTT failures in the proposed GTT-PIF structure, rather than PIFs, as described in the text.

2.6: Conclusion

This paper has presented a methodology to develop a GTT-PIF structure, as the causal mapping foundation for a new HRA method based on the GTT and PIF notions. The structure is directly linked to a cognitive model of literature to ensure that the relevant failure modes and influencing factors are covered.

The proposed methodology emphasizes the traceability of the process underlying the formation of the structure. The process is traceable i.e. it is possible to follow how tasks are identified, transformed, and finally included in the GTT definitions. The tasks are sequentially processed, from the initial HTAs and incident analyses, as specific tasks, Example Tasks and finally Generic Task Types, through different levels of abstraction. All steps of the process can indeed be easily followed. The inclusion of the Example Tasks in the GTT definition (1) is intended to make GTT descriptions transparent: it shows which tasks are in the scope of each GTT; (2) helps in assessing the orthogonality for the GTTs when defining them; Example Tasks can be compared and overlapping components can be modified; and (3) is expected to improve the usability of the GTT taxonomy: an analyst would clearly know what tasks belong to a specific GTT, and thus would reduce the practitioner's effort.

The properties of traceability and transparency were sought to facilitate the review of the process by external people, not involved in the GTT formation in the first place. This is expected to foster the consensus and acceptability of the HRA method under development, as a whole. Traceability and transparency facilitate the incorporation of new tasks (or new ways to carry out tasks) into the GTT definitions. This may become necessary to reflect modifications in the work processes in a specific facility or application of the HRA method to a different facility.

The methodology has been applied to radiotherapy domain allowing formation of definitions of sector-specific representative critical tasks and development of GTT-PIF structure which provide structured framework for assessing what may influence the personnel performance of a particular task. A total of six GTTs were formed and based on the number of identified failure modes a total of eighteen GTT-PIF structures were developed using the framework from reference [17] which included thirteen "*proximate causes*" and more than fifteen "*failure mechanisms*". Conditions like "*Interruptions"* and *"distractions*" were included in the GTT-PIF structures to capture the domain-specific characteristics.

The GTT-PIF structure was validated against a consensus FMEA developed by the American Association of Physicists in Medicine (AAPM) by field experts belonging to various clinics [38]. The concept for validation has been to map the process tasks, failure modes and potential causes identified in reference [38] to the GTT-PIF structure developed in the present work [38]. The correspondence indicates that the GTT-PIF structure addresses tasks and influencing factors relevant for the specific application. It is important to mention that the FMEA was not used to develop the GTT-PIF structure: in this sense, it represents new information to assess the coverage of the GTT-PIF structure.

The next step is to develop the quantification model for the under development HRA method for radiotherapy. The aim is to use the GTT-PIF structures as the base for quantification; Nominal Human Error Probabilities (HEPs) at the failure mode level of the GTTs and then systematically incorporate the effect of applicable PIFs (from the GTT-PIF structures). The primary sources to quantify the HEPs will be from existing HRA methods (Like THERP, CARA, NARA etc.), cross-sector human failure incident database and incident database of PSI's CPT. Indeed, the applicability of the data from the first two sources will have to be studied and is currently underway. The unquantifiable HEPs from the primary source of data will be elicited by experts. The design of expert elicitation exercise is also currently on going. Finally, the method once developed will be applied to PSI's CPT to study potential failure scenarios.

References:

1. U.S. Nuclear Regulatory Commission. Technical basis and implementation guidelines for A Technique for Human Error Analysis (ATHEANA). Technical report NUREG-1624, Rev 1, U.S. Nuclear Regulatory Commission, Washington DC, May 2000.
2. Mosleh A. PRA: A perspective on strengths, current limitations and possible improvements. Nuclear Engineering and Technology 2014; 46(1): 1-10.
3. Johanson G, Jonsson S, Bladh K, Iseland T, Karlsson K-H, Karlsson A, Ljungbjork J, Becker G, Tunturivuori L, Porthin M, Olsson A and Bohm J. Evaluation of Existing Applications and Guidance on Methods for HRA- EXAM-HRA. Report for Nordic nuclear safety research, NKS-305, March 2014.
4. Chandler FT, Chang YH, Mosleh A, Marble J, Boring RL and Gertman DI. Human Reliability Analysis Methods selection guidance for NASA. Technical Report for National Aeronautics and Space Administration, U.S.A., July 2006.
5. Seljelid J, Haugen S, Sklet S and Vinnem JE. Operational risk analysis- total analysis of physical and non-physical barriers BORA Handbook- Rev.00. Report no. 200254-08, Norway, June 2007.

6. Skogdalen JE and Vinnem JE. Quantitative risk analysis offshore—Human and organizational factors. Reliability Engineering and System Safety 2011; 96: 468–479.
7. Stamatelatos M and Dezfuli H. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Technical Report for National Aeronautics and Space Administration, report no. NASA/SP-2011-3421, Second Edition, United States, December 2011.
8. Laumann K, Oien K, Taylor C, Boring RL, Rasmussen M. Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability analysis methods. In: The proceedings of Probabilistic Safety Assessment and Management (Eds. C Smith and T Paulos), Honolulu, U.S.A., 22-27 June 2014.
9. Rasmussen M, Standal MI and Laumann K. Task complexity as a performance shaping factor: A review and recommendations in Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) adaption. Safety Science 2015; 76: 228-238.
10. Williams JC. HEART – a proposed method for assessing and reducing human error. In: 9th Advances in Reliability Technology Symposium, Bradford, U.K., 2-4 April 1986, University of Bradford.
11. Williams JC. HEART – a proposed method for achieving high reliability in process operation by means of human factors engineering technology. In proceedings of a symposium on the achievement of reliability in operating plant, Safety and Reliability Society 2015; 35 (3): 5-25.
12. Kirwan B. A guide to practical human reliability assessment. CRC press, 1994.
13. Kirwan B. The validation of the three Human Reliability Quantification Techniques-THERP, HEART and JHEDI: part 1- Technique description and validation issues. Applied Ergonomics 1996; 27(6): 359-373.
14. Kirwan B, Gibson HW, Kennedy R, Edmunds J, Cooksley G and Umbers I. Nuclear Action Reliability Assessment (NARA): A Data-based HRA tool. In: Proceedings of Probability Safety Assessment and Management (Eds. C Spitzer, U Schmocker and VN Dang), Berlin, Germany, 14-18 June 2004, pp: 1206-1211.
15. Gibson HW, Mills A, Smith S and Kirwan B. Railway Action Reliability Assessment A Railway - Specific Approach to Human Error Quantification. In: International Rail Human Factors Conference, Rail Human Factors Supporting Reliability, Safety and Cost Reduction (eds. N Dadashi, A Scott, JR Wilson and A Mills), London, UK, 5-7 March 2013, pp. 671-676, Taylor & Francis Group.
16. Kirwan B and Gibson WH. CARA: A Human Reliability Assessment Tool for Air Traffic Safety Management– Technical Basis and Preliminary Architecture. In: The Safety of Systems (eds. F Redmill and T Anderson), Bristol, U.K., 13-15 February 2007, pp. 197-214, Springer.
17. Whaley AM, Xing J, Boring RL, Hendrickson SML, Joe JC, LeBlanc KL and Morrow SL. Cognitive basis for Human Reliability Analysis. Technical report NUREG-2114, U.S. Nuclear Regulatory Commission, Washington DC, January 2016.
18. Park J. Scrutinizing inter-relations between performance influencing factors and the performance of human operators pertaining to the emergency tasks of nuclear power plants – an explanatory study. Annals of Nuclear Energy 2011; 38(11): 2521-2532.

19. Groth KG and Mosleh A. A data informed PIF hierarchy for model-based Human Reliability Analysis. Reliability Engineering and System Safety 2012; 108: 154-174.
20. Forester J, Dang VN, Bye A, Lois E, Massaiu S, Broberg H, Braarud PØ, Boring RL, Männistö I, Liao H, Julius J, Parry G and Nelson P. The International HRA Empirical Study –Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data. Technical report NUREG-2127, U.S. Nuclear Regulatory Commission, Washington DC, August 2014.
21. Chang YJ, Bley D, Criscione L, Kirwan B, Mosleh A, Madary T, Nowell R, Richards R, Roth EM, Sieben S and Zoulis A. The SACADA database for human reliability and human performance. Reliability Engineering System Safety. Reliability Engineering and System Safety 2014; 125: 117-133.
22. Ekanem NJ, Mosleh A and Shen S-H. Phoenix: A model based Human reliability analysis methodology: Qualitative analysis procedure. Reliability Engineering System Safety 2016; 145: 301-315.
23. IAEA. Case studies in the application of probabilistic safety assessment techniques to radiation sources. Technical document for International Atomic Energy Agency -TECDOC no.-1494, Vienna, 2006.
24. ROSIS, Radiation Oncology Safety Information System, http://rosis.ch/ge/rosis_daten1.asp
25. SAFRON, Safety reporting and learning system for radiotherapy, www.rpop.iaea.org/RPOP/RPoP/Modules/login/safron-register.htm
26. Nath R, Biggs PJ, Bova FJ, Ling CC, Purdy JA, Gejin JVD and Weinhous MS. AAPM code of practice for radiotherapy accelerators. Report for the American Association of Physicists in Medicine, Task Group 45, Report number 047, September 1994.
27. Klein EE, Hanley J, bayouth J, Yin F-F, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B, and Liu C. Task Group 142 report: Quality assurance of medical accelerators, Medical Physics 2009; 36 (9): 4197-4212.
28. The American College of Radiology. Practice Guideline for Radiation Oncology. Practice Guidelines and Technical Standards. Reston, U.S.A., American College of Radiology, 2009.
29. Leer JWH, McKenzie AL, Scalliet P and Thwaites DI. Practical Guidelines for the Implementation of a Quality System in Radiotherapy. Report for European Society for Radiotherapy and Oncology (ESTRO), Physics for Clinical Radiotherapy, Booklet No. 4, Brussels, 1998.
30. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Medin PM, Mundt A, Mutic S, Palta JR, Thomadsen BR, Williamson JF and Yorke ED. A Method for Evaluating Quality Assurance Needs in Radiation Therapy. International Journal Radiation Oncology Biology Physics 2008; 71(1 suppl): 170-173.
31. Thomadsen B. Critique of traditional quality assurance paradigm. International Journal of Radiation Oncology Biology Physics 2008; 71(1 suppl): 166-169.
32. International Atomic Energy Agency (IAEA). Lessons learned from accidental exposures in radiotherapy. Safety report- IAEA-Series No. 17, Vienna, 2000.
33. International Commission on Radiological Protection. Prevention of accidental exposures to patients undergoing radiation therapy. Report for International Commission on Radiological Protection, publication 86, volume 30, no. 3, 2000.

34. Lopez PO, Cosset J-M, Dunscombe P, Holmberg O, Rosenwald J-C, Ashton LP, Llanes JJV and Vatnitsky S. Preventing accidental exposures from new external beam radiation therapy technologies. Report for International Commission on Radiological Protection, publication 112, volume 39, no. 4, 2009.
35. Yeung TK, Bortolotto K, Cosby S, Hoar M and Lederer E. Quality assurance in radiotherapy: evaluation of errors and incidents recorded over a 10 year period. Radiotherapy and Oncology 2005; 74(3): 283-291.
36. World Health Organization. Radiotherapy risk profile. Technical manual World Health Organization, 2008.
37. Derreumaux S, Etard C, Huet C, Trompier F, Clairand I, Bottolier-Depois JF, Aubert B and Gourmelon P. Lessons from recent accidents in radiation therapy in France. Radiation Protection Dosimetry 2008; 131(1): 130-135.
38. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR, Williamson JF and Yorke ED. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management, Medical physics 2016; 43 (7): 4209-4262.
39. Marx DA and Slonim AD. Assessing patient safety risk before the injury occurs: an introduction to sociotechnical probabilistic risk modelling in health care. BMJ Quality and Safety 2003; 12(2 suppl): 33-38.
40. Ekaette EU, Lee RC, Cooke DL, Kelly K-L and Dunscombe PB. Risk analysis in radiation treatment: Application of a new taxonomic structure. Radiotherapy and Oncology 2006; 80(3): 282-287.
41. Ekaette E, Lee RC, Cooke DL, Iftody S and Craighead P. Probabilistic fault tree analysis of a radiation treatment system. Risk Analysis 2007; 27(6): 1395-1410.
42. Gomes EC, Duarte JP and Frutuoso e Melo PF. Human reliability modeling of radiotherapy procedures by Bayesian networks and expert opinion elicitation. Nuclear Technology 2016; 194(1): 73-96.
43. Vilaragut JJ, Ferro R, Lozano B, De la Fuente A, Duménigo C, Troncoso M, Pérez Y, Alemañy J, León L, Amador R, Lazo R, Labrador F, Blanco A, Betancourt L, Crespo D and Silvestre I. Probabilistic safety assessment to the cobalt therapy process. In: 11th International Congress of the International Radiation Protection Association (IRPA), Madrid Spain, 23-28 May 2004.
44. Vilaragut JJ, Ferro R, Marti MR, López PO, Ramírez ML , Mulas AP , Montero MB, Somoano F, Rodriguez JMD, Papadópulos S, Pereira Jr PP, Morones RL, Cortina EL, Oliva JJR and Alemañy J. Probabilistic safety assessment of the radiotherapy treatment process with an electron linear accelerator for medical uses. In: 12th International Congress of the International Radiation Protection Association (IRPA), Buenos Aires, Argentina, 19-24 October 2008.
45. Chadwick L and Fallon EF. Human Reliability Assessment of a critical nursing task in a radiotherapy treatment process. Applied Ergonomics 2012; 43(1): 89-97.
46. Lyons M, Adams S, Woloshynowych M, Vincent C. Human reliability analysis in healthcare: a review of techniques. International Journal of Risk and Safety in Medicine 2004; 16: 223-237.

47. McLeod JN, Baron J and Rivera S. Human Reliability Analysis in Cobalt-Therapy Process using an Adapted ATHEANA Prospective Approach. In Proceedings of Probability Safety Assessment and Management (Eds. C Spitzer, U Schmocker and VN Dang), Berlin, Germany, 14-18 June 2004.
48. Turra F and Verbano C. A human reliability analysis approach to clinical risk management: first steps towards a new methodology. In: proceedings of the European Safety and Reliability Conference (Eds. BMD Sjoberg and T Aven), Stavanger, Norway, 25-27 June 2007, pp: 209-217, Taylor & Francis group.
49. Wreathall J, Brown WS, Militello L, Cooper SE, Lopez C and Franklin C. A risk-informed approach to understanding human error in radiation therapy. Technical report NUREG-2170, U.S. Nuclear Regulatory Commission, Washington DC, June 2017.
50. Pandya D, Podofillini L, Emert F, Lomax AJ and Dang VN. A method for human reliability analysis in radiotherapy: identification and characterization of influencing factors. In: proceedings of European Safety and Reliability Conference(Eds. L Podofillini, B Sudret, B Stojadinovic, E Zio and W. Kroger), Zurich, Switzerland, 7-10 September 2015, Taylor & Francis group.
51. Swain AD and Guttmann HE. Handbook of human reliability analysis with emphasis on nuclear power plant applications. Final report NUREG/CR-1278, U.S. Nuclear Regulatory Commission, Washington DC, August 1983.
52. Gibson HW, Basra G and Kirwan B. Development of the CORE-DATA database. Safety and Reliability Journal, Safety and Reliability society 1999; 19(1): 6-20.
53. Forester J, Liao H, Dang VN, Bye A, Lois E, Presley M, Marble J, Nowell R, Broberg H, Hildenbrandt M, Hallbert B and Morgan T. The U.S. HRA Empirical Study- Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator. Technical Report NUREG-2156, U.S. Nuclear Regulatory Commission, Washington, DC, June 2016.
54. Endsley MR. Toward a theory of situation awareness in dynamic systems. Human Factors 1995; 37(1): 32-64.
55. Endsley MR. The role of situation awareness in naturalistic decision making. In: Naturalistic decision making (eds. CE Zsambok and G Klein), Dayton, Ohio, U.S., 1994, pp. 269-284, Lawrence Erlbaum Associates Publishers.
56. Klein GA, Ross KG, Moon BM, Klein DE, Hoffman RR and Hollnagel E. Macrocognition. Intelligent Systems- IEEE 2003; 18(3): 81-85.
57. O'Hara JM, Higgins JC, Brown WS, Fink R, Persensky JJ, Lewis PM and Szabo A. Human Factors Considerations with Respect to Emerging Technology in Nuclear Power Plants. Technical report NUREG/CR-6947, U.S. Nuclear Regulatory Commission, Washington DC, October 2008.
58. Roth EM. Generic List of Complicating Situational Factors, Individual and Team 'Macro-Cognitive' Performance Problems, and Unsafe Actions. Internal project report prepared for the HRA Data project, Office of Research, U.S. Nuclear Regulatory Commission, Washington DC, 2010.
59. Patterson ES and Hoffman RR. Visualization Framework of Macrocognition Functions. Cognition, Technology & Work 2012; 14(3): 221-227.

60. Rasmussen J, Duncan K and Leplat J. New Technology and Human Error. John Wiley and Sons, London & New York, 1987.
61. Reason J T. Human Error. Cambridge: Cambridge university press, 1990.
62. Hollnagel E. Cognitive Reliability and Error Analysis Method (CREAM), Elsevier, Amsterdam, The Netherlands, 1988.
63. Ford E, Santos LF de L, Pawlicki T, Sutlief S and Dunscombe PB. Consensus recommendations for incident learning database structures in radiation oncology. Medical Physics 2012; 39(12): 7272-7290.
64. Fraass B, Doppke K, Hunt M, Kutcher G, Starkschall G, Stern R and Dyke JV. American Association of Physicists in Medicine radiation therapy committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning. Medical Physics 1998; 25(10): 1773-1829.
65. Toft B. Independent review of the circumstances surrounding a serious adverse incident that occurred in the.. (Confidential Report, Redacted version). Report for World Health Organization, 2004. http://www.who.int/patientsafety/news/Radiotherapy_adverse_event_Toft_report.pdf
66. Verhey LJ. Immobilization and positioning patients for radiotherapy. Seminars in radiation oncology 1995; 5(2): 100-114.
67. Dekker S. Patient Safety: A Human Factors Approach. CRC Press, 2011.
68. Williams JC and JL Bell. Consolidation of the error producing conditions used in the human error assessment and reduction technique (HEART). Safety and Reliability 2016; 35(3): 26-76.
69. Klein GA, Phillips JK, Rall EL and Peluso DA. A data-frame theory of sensemaking. In: Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Pensacola Beach, U.S., 2003, pp. 113-155, Lawrence Erlbaum Associates Publishers.

Chapter 3: HRA method- quantitative elements

This chapter reproduces the author's submitted article to Reliability Engineering and System Safety journal. The work in this article focusses on the thesis's objective of development of the quantitative part of the HRA method. This includes: the concept development of the overall quantification approach, development of the error representational model (i.e. Decision Trees) and quantification of the HEPs for the method (see objective 3 and 4 from **Section 1.2: Aims and objectives**).

The chapter first outlines the overview of the approach taken towards quantification for the HRA method in development. Then, it presents the quantitative modelling framework (Decision Tree) chosen for the method. It is a tree structure with decision points (developed for each GTT-Failure mode, with decision points as PIFs coming from results in chapter 2) with questions that an analyst uses to determine if the given PIF is affecting the GTT-failure mode. The end of a DT path in a tree is the HEPs of the GTT-failure mode given a certain PIF(s) is(are) affecting it. In terms of method development a DT structure serves dual purpose of (a) representing error framework and as well as (b) to systematically incorporate the effect of PIFs on task failure probabilities. The development of DT is based on the identified GTT(failure mode)- PIF structures from chapter 2. Then, some examples of developed decision trees for the HRA method are shown. All the developed decision trees for the method can be seen in **Appendix 8**. Then, the chapter presents the approach for quantification of the Human Error Probabilities (HEPs) for the developed decision trees. This includes the expert judgment-based quantification of the HEPs. Finally, it ends with the "convergent" validation of the quantified HEPs using applicable data from the existing HRA methods.

Publication Details:

This chapter is reproduced with permission from **Pandya D**, Podofillini L, Emert F, Lomax AJ, Dang VN and Sansavini G. *Quantification of a human reliability analysis method for radiotherapy applications based on expert judgment aggregation*. Submitted to Reliability Engineering and system safety 2018, special issue.

Abstract

This paper develops a quantification framework of a new Human Reliability Analysis (HRA) method for application to the radiotherapy domain. The whole development process is presented, from the definition of the structure of the quantification model (i.e. decision trees), to its use for producing the probability values via expert judgment. To avoid shortcomings of directly eliciting probability values, experts are asked to assess the importance of specific factors for the failure probability, elicited on a qualitative scale. Each expert assessment is converted into statements about the order of magnitude of the probability value. The values from different experts are combined via an expert aggregation method, developed specifically for HRA. Although validation is typically challenging for HRA due to the low probabilities and to shortcomings in the available data, the present paper includes an attempt to validate, at least partially, the results obtained from the elicitation. Applicable Human Error Probability (HEP) values from existing HRA methods are identified. Indeed, some tasks are generic in nature and data can be assumed to be sector-independent (e.g. checking activities, interacting with interfaces, simple tasks such as identifying objects or characters/numbers). Differences in the values are identified and, when possible, linked to differences in the performance context characteristic of the field of application of the different methods.

3.1: Introduction

Human Reliability Analysis (HRA) is based on sets of methods and tools to identify and analyze potential failures in the human interactions with complex socio-technical systems [1, 2, 3]. The systematic analysis and management of human failures contributes to reaching and maintaining the very high level of safety typical of several industrial sectors, such as nuclear power and air traffic control [4, 5].

HRA typically includes both qualitative and quantitative analyses. Generally, the former aims at understanding how tasks are performed by the personnel, how personnel can fail and the factors that influence performance [1, 2, 3, 4, 5, 6]. Quantification relates to assessing the failure probabilities. Depending on the HRA method, quantification involves characterizing the type of task and the performance influences through categories and/or rating scales. These methods provide values of human failure probabilities with associated ranges of uncertainty. The type of data feeding these methods is dependent on the method itself and, especially, on the application domain. The data is used to assess the reference failure probability values characteristic of the method (e.g. values for specific types of tasks performed under specific context conditions); also,

the data is used to inform the parameters of the failure probability calculation model. Such parameters include, for example, the multipliers used to capture the effect of the difference between the performance and reference conditions on the failure probability. For the nuclear power domain, data have been sought from the early beginning of the HRA field and comprises operational data, data from simulated environment, and human performance studies [2, 7, 8, 9, 10]. Furthermore, there are on-going, large-scale, data collection initiatives [11, 12]. In all cases, the importance of expert judgment is to be highlighted in the evaluation of the applicability of the data as well as in compensating for the lack of data. In particular, HRA applications are recently focusing on very challenging performance situations, combining multiple equipment failures as well as possibly harsh environmental conditions (e.g. addressing human performance in response to seismic events or to mitigate severe accidents). Empirical data to quantify failure probabilities in these situations, in practice, do not exist – and will be very challenging to collect even for the mentioned large-scale data collection initiatives [11, 12]. For a detailed overview of the status of HRA data, refer to [2, 10].

For other application domains, the availability of data relevant for HRA is generally much reduced compared to the nuclear power domain. Various human factor studies and the CORE-data database [13] are at the basis of the data feeding the Nuclear Action Reliability Analysis (NARA) [14], the Railway Action Reliability Analysis (RARA) [15], and the Controller Action Reliability Analysis (CARA, for the air traffic control domain) [16] methods. In their development, they adopt the framework of Human Error Assessment and Reduction Technique (HEART) [7, 8]. Another example is the PETRO-HRA method [17], recently developed for application to the oil and gas industry, adopting the framework of the Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) method [18]. The values of the multipliers featured in PETRO-HRA are determined by adapting those of the original SPAR-H method, based on the judgment of the method's developers. In the majority of the HRA applications outside the nuclear power domain (but also including cases from the nuclear power), models are quantified by expert judgment, elicited in variety of ways, e.g. directly probability values or ranges, or on qualitative or semi-quantitative scales, or eliciting probability rankings [19, 20, 21, 22, 23, 24].

The widespread use of expert judgment to quantify HRA models calls for traceable and defendable approaches to elicit the judgments and incorporate them in the models, as reference data as well as parameters of the calculation model, as advocated in [25]. Along these lines, this paper addresses the development of the quantification framework of a new HRA method,

intended for application to the radiotherapy domain. The design of the expert elicitation sessions is presented, along with the approach to aggregate the judgments. Comparison of the obtained values with HRA data from existing methods is performed for partial validation of the elicitation procedure. The work presented in this paper is part of the research activity to develop an HRA method for application to radiotherapy carried out at the Paul Scherrer Institute. The activity addresses the need to analyze possible human failures with the quantitative perspective typical of HRA, such to complement the safety assessment practice already carried out within the radiotherapy sector [26, 27, 28]. In previous work [29], the authors have identified six groups of critical task types, denoted as Generic Task Types (GTTs), following the nomenclature of the HEART method [7, 8]. A traceable and systematic process is presented in [29], which, for each GTT, progressively identifies the involved cognitive functions, their failure modes and causes, failure mechanisms and Performance Influencing Factors (PIFs). In this paper, decision trees are first developed from these GTT-PIF structures, and then quantified via the presented elicitation and aggregation procedure.

Figure 12 outlines the method development with reference to the relevant sections in the present paper. In particular, Section 2 presents the GTT-PIF structures developed in [29] and the development of the Decision Tree (DT) models, i.e. the identification and definition of the DTs and of the branching points (the headings of each decision node). **Section 3.2: Quantitative framework: background and DT development** presents both the underlying concepts for identification and definition as well as examples of the results. **Section 3.3: Quantification approach** presents the approach to quantification, i.e. the design of the elicitation sessions and the method for aggregation of the HEP estimates by the different experts. The sessions are aimed at identifying and assessing the negative conditions defining the DT branch points. Preliminary work for identification of these conditions has been carried out as part of the GTT-PIF structure development (**Figure 12**), and the elicitation sessions served to confirm and possibly complement the set of conditions. **Section 3.4: Expert elicitation: results and discussion** presents the results of the elicitation. **Section 3.5: Convergence validation** presents the validation of the results against HEP values from existing methods (convergent validation following the definition of [30]). Conclusions are drawn at closure.

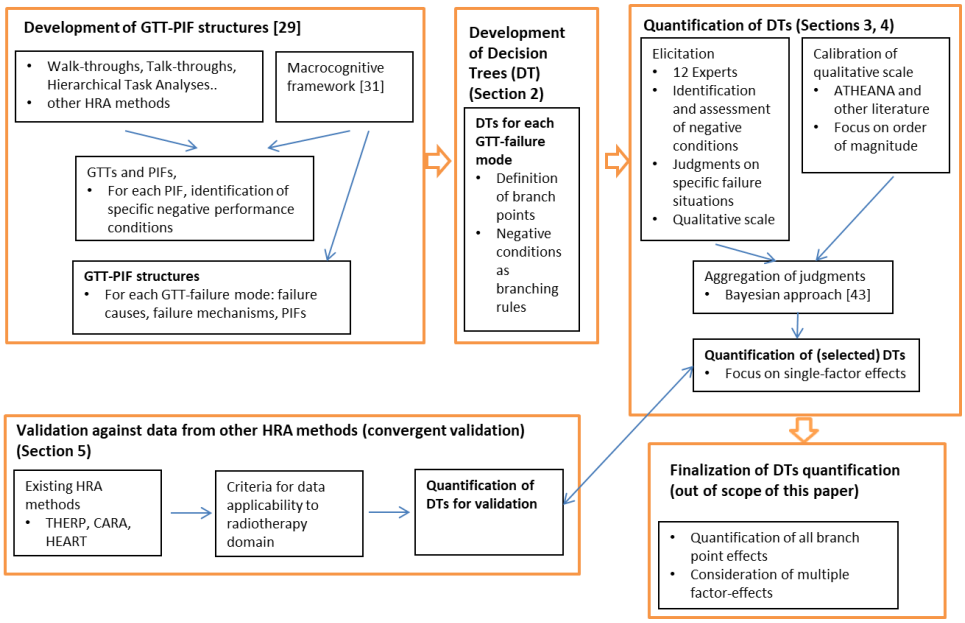


Figure 12: Overall process of HRA method development

3.2: Quantitative framework: background and DT development

3.2.1: Decision Trees development - concepts

The Generic Task Types (GTTs)- Performance Influencing Factors (PIFs) structures developed in [29] (**Figure 13**) link each GTT to the set of PIFs that influence the failure probability (the sets of PIFs relevant for each GTT failure mode are reported in **Table 8**). These structures are developed via a systematic and traceable process, which, for each GTT, progressively identifies the involved cognitive functions, their failure modes, failure causes, failure mechanisms and PIFs, following the cognitive framework presented in [31]. The framework identified by the GTT-PIF structures is causal, i.e. explanations are searched for the causes of failures of each GTT. Each GTT is first linked to the cognitive functions involved in its performance: "Detecting and Noticing", "Sense-making and Understanding", "Decision-making", "Action" and "Team coordination" [31]. Failure modes are then identified for each function, representing the "observable" effect of the failure (see **Table 8** for an overview of the identified failure modes). Each failure mode is then associated to the failure mechanisms, which point to specific cognitive faults that are generally not observable, e.g. memory failure, loss of attention, expectation biases, missed perception, etc. In analogy to the approach in [31], the failure mechanisms are grouped

into “proximate causes” which represent a higher-level explanation for the failure mode (**Figure 13**).

Decision Trees (DTs) are chosen as the quantification model because they can represent the causal influences on failures identified by the GTT-PIF structures, i.e. the DTs identify the causes possibly leading to the GTT failure. Similar to HRA methods [32, 33], each decision tree addresses a failure mode with branching points representing the effects of PIFs (**Figure 13**). An important observation on the structures from [29] justifies the development of decision trees at the level of failure modes, i.e. each different failure mode of the same GTT is affected by a different set of PIFs, while similar failure modes across different GTTs are affected by same sets of PIFs – these patterns are reproduced in **Table 8**. For example, **Table 8** shows that the same PIF patterns affect the failure modes “Patient information incorrectly matched”, “Deviation from requirement not recognized”, and “Mismatch or inconsistency not recognized”, all characterized by failure of the “Detection and Noticing” macrocognitive function (See ‘#’ in **Table 8**). Indeed, the failure to identify that some information is incorrect is expected to be driven by factors like human machine interface (HMI), resources, loads, and environment. Conversely, different factor sets affect the failure modes that relate to “Decision-making” macrocognitive function, e.g. “Identification check not performed (decision-based)”, “Check not performed (decision-based)”, “Inappropriate decision on strategy selection”, and “not communicated (decision-based)” (See ‘■’ in **Table 8**). Indeed, for these failure modes, the driving factors are expected to be safety culture, loads, complexity, training and experience.

The GTT-PIF structures from [29] are the primary sources for the definition of the decision tree branching points. Mainly, the PIFs lie at the branching points as conceptually shown in **Figure 14**. To reduce the number of DT branches, different PIFs are combined when they affect performance in a similar way; in these cases, the branching point represents the higher-level failure mechanism from the GTT-PIF structures. For example, the PIF “Loads” (that includes cognitive workload, simultaneous tasks, distractions and interruptions etc.) and the PIF “Environment” (that includes noise, lighting, temperature etc.) both affect the attention of the person while performing a task, and, therefore, these have been combined into the single branch point (“Distractions/interruptions and excessive workload”). More examples are provided in **Section 3.2.2: Decision Trees development - results**.

Each branch point is associated a set of domain-specific conditions that would negatively impact the performance of the personnel (**Figure 14**). These negative conditions assist the analyst in deciding which path of the DT to follow, i.e. if at least one of the conditions is verified, then the

analyst would proceed on the lower branch. The identification of these negative conditions is not straightforward. They have to be easily recognizable by field domain practitioners eventually applying the DTs. They should be homogeneous such that the expected quantitative impact is similar for all. These are identified from various sources, i.e. past event analysis from incident databases [34, 35], review of causal factor hierarchies developed for radiotherapy [36] and observations at Paul Scherrer Institut's Center for Proton Therapy [37]. As presented in **Section 3.3: Quantification approach**, one of the goals of the elicitation sessions is to validate the set of these conditions.

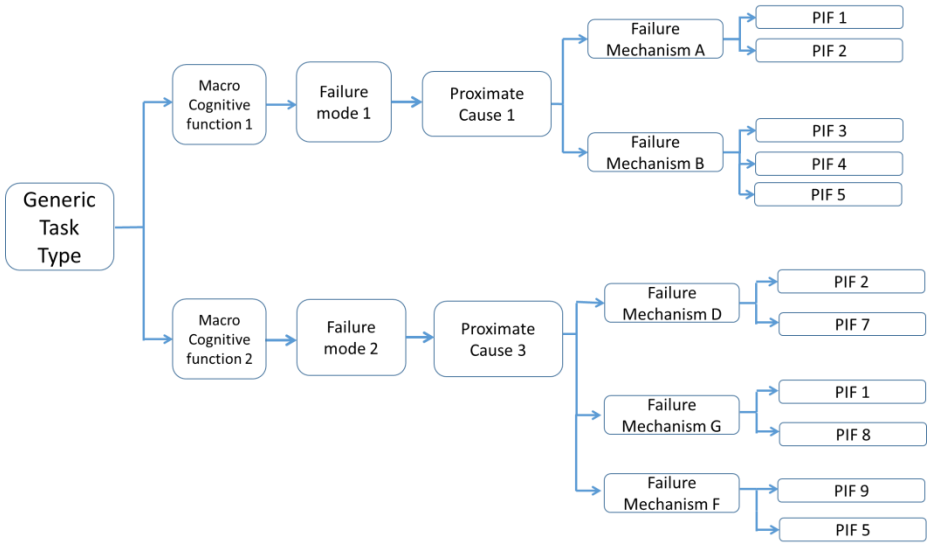


Figure 13: Concept for the GTT-PIF structures from [29]

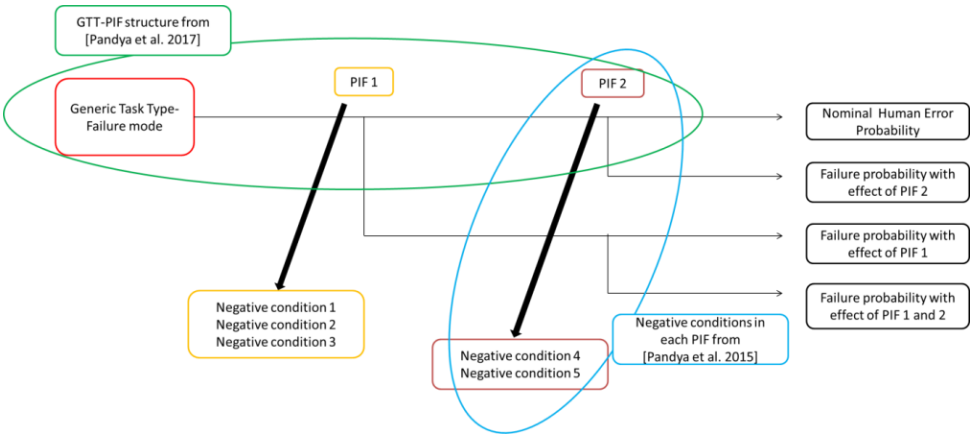


Figure 14: Example of a decision tree with two PIF branch points and negative conditions characterizing each branch point

3.2.2: Decision Trees development - results

Table 9 provides the overview of the DTs developed, one for each of the eighteen GTT failure modes. The first two columns show the corresponding GTT and failure modes; the rest of the columns gives the branch points included in each DTs, for example "*Information unclear*", "*tumor complexity*", "time *pressure*", "*Information content unclear-verbal*" etc. For each DT (i.e. each row), a shaded area indicates that the branch point is included in the DT and the text in the area specifies the PIFs modeled in the branch point. While the headings of the branch points are generic (i.e. applicable to different DTs), the included negative conditions can be different because they are specific to the corresponding GTT failure mode. Some examples of DT formation are explained in the following.

Table 8: GTT-Failure modes and the associated PIFs [29]

No	GTT- Failure modes		PIFs								
			HMI	RE	TE	SC	TR-EX	PE	CO	LO	ENV
1	Simple interaction with software or tool	Execute desired action incorrectly	**O**				**O**	**O**		**O**	**O**
		Failure to execute desired action								**$**	**$**
2	Quality Check	Deviation from requirement not recognized	**#**	**#**			**#**	**#**		**#**	**#**
		Inappropriate understanding of underlying principles	**Δ**	**Δ**			**Δ**		**Δ**	**Δ**	
		Check not performed (decision-based)				**■**	**■**	**■**		**■**	
		Execute desired action incorrectly	**O**				**O**	**O**		**O**	**O**
		Failure to execute desired action								**$**	**$**
		Coordination failure			**&**	**&**				**&**	
3	Identification of patient or patient related items	Patient information incorrectly matched(1)	**#**	**#**			**#**	**#**		**#**	**#**
		Identification check not performed (decision-based)				**■**	**■**	**■**		**■**	
		Failure to execute desired action								**$**	**$**
4	Complex interaction with software or tool	Misinterpretation of data	**Δ**	**Δ**			**Δ**		**Δ**	**Δ**	
		Execute desired action incorrectly	**O**				**O**	**O**		**O**	**O**
		Mismatch or inconsistency not recognized	**#**	**#**			**#**	**#**		**#**	**#**
5	Iterative determination of optimum parameters	Misinterpretation of information	**Δ**	**Δ**			**Δ**		**Δ**		
		Inappropriate decision on strategy selection		**■**			**■**		**■**	**■**	
6	Verbal communication	Communication failure			**&**	**&**	**&**		**&**	**&**	**&**
		Not communicated (decision-based)				**■**	**■**			**■**	

(1) – Same symbols identify same or very similar PIF patterns. Correspondingly, same or very similar DTs are developed (see **Table 9**).

Legend: TR-EX- Training-Experience, SC- Safety Culture, RE- Resources, TE- Team, PE- Personal, HMI- Human Machine Interface, LO- Loads, CO- Complexity, ENV- Environment

Table 9: Overview of the developed decision trees, their branch point headings and PIFs affecting each branch point

#	GTT - Failure modes (Decision Tree)		Decision Tree branch points with affecting PIFs							
			B1a, B1b	B1c	B2	B3	B4	B5	B6	B7a, B7b, B7c
			Information unclear	Information content unclear-Verbal	Low vigilance due to expecting no error	Tumor complexity	Lack of training or experience	Software or resource issues	Time pressure	Distractions/ interruptions and excessive workload
1	Simple interaction with software or tool	Execute desired action incorrectly[(1)]	HMI		TR-EX, PE					LO, ENV
		Failure to execute desired action [(2)]								LO, ENV
2	Quality Check	Deviation from requirement not recognized[(1)]	HMI, RE		TR-EX, PE					LO, ENV
		Inappropriate understanding of underlying principles [(3)]	HMI			CO	TR-EX	RE	LO	
		Check not performed (decision-based)			SC,TR-EX, PE					LO
		Execute desired action incorrectly [(3)]	HMI		TR-EX, PE					LO, ENV
		Failure to execute desired action [(2)]								LO, ENV
		Coordination failure [(3)]		TE, SC						LO
2	Identification of patient or patient related items	Patient information incorrectly matched[(1)]	HMI, RE		TR-EX, PE					LO, ENV
		Identification check not performed (decision-based)			SC, TR-EX, PE					LO
		Failure to execute desired action [(2)]								LO, ENV
4	Complex interaction	Misinterpretation of data[(1)]	HMI			CO	TR-EX	RE	LO	

#	GTT - Failure modes (Decision Tree)		Decision Tree branch points with affecting PIFs							
			B1a, B1b	B1c	B2	B3	B4	B5	B6	B7a, B7b, B7c
			Information unclear	Information content unclear- Verbal	Low vigilance due to expecting no error	Tumor complexity	Lack of training or experience	Software or resource issues	Time pressure	Distractions/ interruptions and excessive workload
	with software or tool	Execute desired action incorrectly (3)	HMI		TR-EX, PE					LO, ENV
		Mismatch or inconsistency not recognized (3)	HMI, RE		TR-EX, PE					LO, ENV
5	Iterative determination of optimum parameters	Misinterpretation of information(1)	HMI			CO	TR-EX	RE		
		Inappropriate decision on strategy selection (2)				CO	TR-EX	RE	LO	
6	Verbal communication	Communication failure(1)		TE,SC,CO						LO, ENV
		Not communicated (decision-based)					SC, TR-EX		LO	

(1)- Failure situations related to this GTT-failure mode were addressed in the elicitation sessions

(2)- Some DT branch points for this failure mode were addressed in the elicitation sessions.

(3)- The DTs associated to this GTT failure modes can be taken from those directly addressing elicited situations, i.e. those from note (1).
Note also that the specific conditions underlying the same decision tree headings across different GTTs may differ; see **Section 3.2.2: Decision Trees development - results**

Legend: TR-EX- Training-Experience, SC- Safety Culture, RE- Resources, TE- Team, PE- Personal, HMI- Human Machine Interface, LO- Loads, CO- Complexity, ENV- Environment

For the DT "*Quality Check- deviation from requirement not recognized*" (**Table 9**, and **Figure 15**), six PIFs are identified tied to three failure mechanisms, see **Figure 16** [29]. From **Figure 15**, the first branch point, "*Information unclear*", captures possible ambiguous and unclear way of presenting the information, either on the interface or in the documents needed to assist the personnel. The branch point joins the PIFs "*Human-Machine Interface*" and "*Resources*". The latter refers to the clarity of the available documentation, and influences perception of the values to be compared as part of the quality check. Other aspects of the PIF "*Resources*" are evaluated for other failure modes, e.g. the availability and adequacy of the information provided by the documents would impact the failure mode "*inappropriate understanding of the check*". The conditions identified as relevant for this branch points are given in **Table 10**, and include difficulty in readability of the information, physical cluttering of information on the screen, counter-intuitive order of information on screen in comparison with order on the document etc. The second branch point, "*Low vigilance due to expecting no error*", is formed by joining PIFs that relate to conditions that can lead to bias due to expectation or familiarity and hence decreasing the vigilance while performing the task, i.e. "*Training-Experience*" (TR-EX: includes situation familiarity) and "*Personal*" (PE: includes biases). Similarly, the last branch point "*Distractions/Interruptions and excessive workload*", is formed by joining PIFs that relate to conditions that distract the focus of the personnel and hence leading to attention loss i.e. "*loads*" and "*environment*". Representative conditions for this branch point include cognitive overload from work, workplace distractions like interruptions from colleagues, background noise. The conditions falling in the three branch points are reported in **Table 10**.

Additionally, for the "Simple interaction with software or tool-executed desired action incorrectly" GTT-failure mode (**Table 9** (#1) and **Figure 17**), reference [29] identifies the following relevant PIFs, i.e. "*training-experience*" (includes training program and individual knowledge for automaticity of the task), "*Human Machine Interface*", "*Personal*" (includes bias due to monotonous nature of the task), "*loads*" (includes workload, time pressure) and "*Environment*" (including distractions, noise) PIFs as demonstrated in **Figure 18**. Similarly to the aforementioned GTT-PIF mapping, PIFs and failure mechanisms with similar effects are grouped into single branching points. The three branch points formed are "*Information unclear*", "*Low vigilance due to expecting no error*", and "*Distractions/ interruptions and excessive workload*". As previously mentioned, branch points with the same headings may not necessarily involve the same set of conditions. For example, the branch point "*Distraction/ interruptions, excessive workload*" (B7a) in **Table 11** captures the possible loss of focus due to simultaneous tasks being performed, distractions from colleagues and reduced time to carry out the task. The

corresponding branch point in **Table 10** does not include the effect of multiple tasks being performed simultaneously, because, given their safety importance, quality checks are not typically performed with other tasks, but assigned dedicated time.

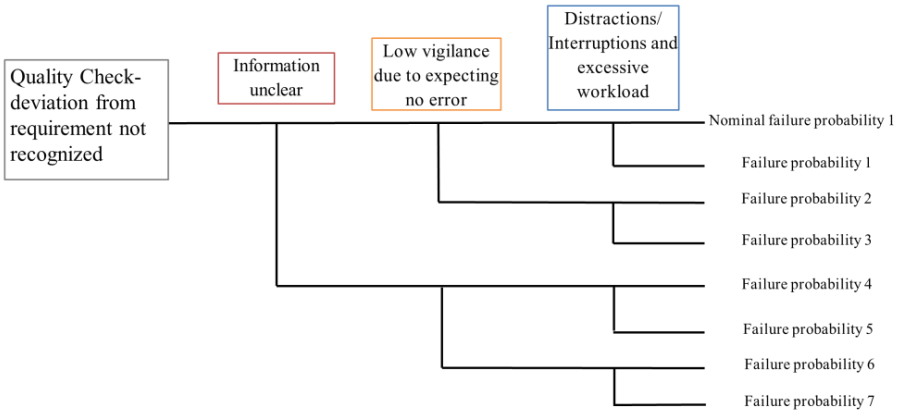


Figure 15: Decision tree of Quality check- deviation from requirement not recognized

Example Task	Macrocognitive function	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Check of transferred data to a software/machine	Detection and noticing	Deviation from requirement not recognized	cues/information misperceived	cue content			√			√			
				vigilance- attention	√					√	√		o
				expectation	√				√				

Figure 16: Example of GTT-PIF mapping for Quality Check GTT from [29]

Table 10: Negative conditions defining branch points in "Quality check- deviation from requirement not recognized" (GTT-failure mode)

Branch point		**Affecting negative conditions**
B1a	Information unclear	The values on the interface are not easily readable
		The values look alike
		Too much information on the software screen leading to confusion
		The ordering of the values on the control document and on the screen do not match (i.e. X,Y,Z on screen and Z,X,Y in the document)
B2	Low vigilance due to expecting no error	The check was performed recently and you trust it was performed correctly
		The task to produce the output is simple - no error is expected
		Expecting no error as the check is performed by the same person doing the initial task
B7a	Distractions/ interruptions and excessive workload	Overloaded with other work
		There is little time to do the task
		The background noise level is too high, it distracts the focus
		Interruptions from colleagues while doing the task

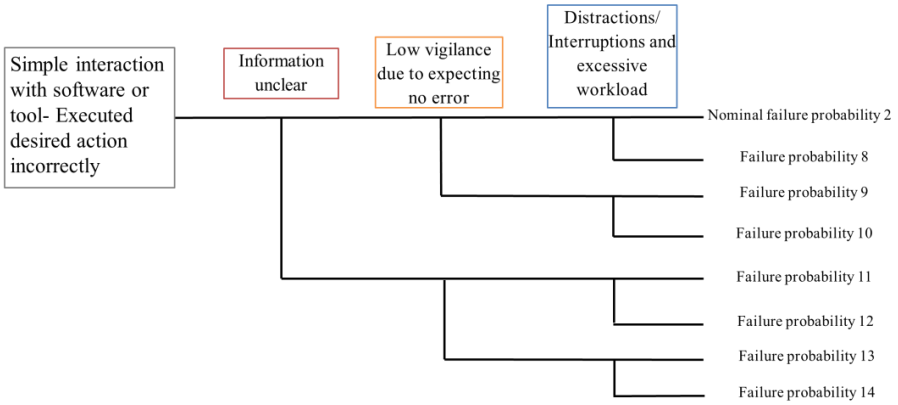


Figure 17: Decision tree of Simple interaction with software/tool- Executed desired action incorrectly

Example Tasks	Macrocognitive function	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Looking and copying/inputting data.	Action	Execute desired action incorrectly	Executed desired action incorrectly	motor learning	√				√				√
		-Slip in data input		manual control issues						√			
		-Wrong location		automacity control	√					√	√		
		-Default value partially overwwritten		recognition error	√				√	√	√		√
		-Look at wrong data		dual task interference						√	√		

Figure 18: GTT-PIF mapping from [29]

Table 11: Negative conditions defining branch points in "Simple interaction with software/tool- Executed desired action incorrectly" (GTT-failure mode)

Branch point		**Affecting negative conditions**
B1a	Information unclear	The indications for the input boxes on the screen are not readable
		The layout of the input interface is unusual or confusing
		Too much information on the screen leading to confusion
		Touch screen or input keyboard is very sensitive (undetected mis-selection is possible)
B2	Low vigilance due to expecting no error	Task is mechanical and repetitive
B7a	Distractions/ interruptions and excessive workload	The background noise level is too high, it distracts the focus
		Simultaneously doing another urgent task
		Interruptions from colleagues while doing the task
		There is less time to do the task (excessive workload)
		Simultaneously tracking/supervising another task

3.3: Quantification approach

This section presents the methodological approach to quantify the Human Error Probabilities (HEPs) for the developed decision trees. The design of the sessions is presented in **Section 3.3.1: Expert elicitation**, and the aggregation of the judgments is presented in **Section 3.3.2: Expert data aggregation**.

3.3.1: Expert elicitation

The aim of the expert elicitation sessions is to identify potentially negatively influencing conditions in specific failure situations and to quantify this influence. Six failure situations are identified for the elicitation, each involving one task related to each GTT (i.e. one failure situation per GTT, **Table 12**). Besides covering all GTTs, the failure situations are selected after discussions with proton therapy experts, prioritizing those with possibly high impact on patient safety as well as including tasks involved in past undesired events. Each failure situation corresponds to one GTT failure mode (**Table 9**). The selection also intended to capture failures in the GTT cognitive function identified as dominant in earlier work [29] as well as also covering as many macrocognitive functions as possible [31]. Dealing with specific tasks allowed experts to contextualize the analyzed situations and to think of the specific performance influences affecting the tasks. An alternative approach could be to analyze the performance influences on the GTTs directly, without referring to specific tasks (or possibly providing a set of example specific tasks). The former approach is preferred to avoid that experts have to deal with the abstract categories of GTTs and PIFs; this allows linking the discussions to the daily experience of the personnel.

Table 12: GTT-Failure modes quantified using expert elicitation

Failure situations	**Failure mode [Macro-Cognitive function]**	**Generic Task Type**
To transfer incorrect patient plan information (Offset data) from Patient Positioning and Verification to PatBase.	Execute desired action incorrectly [Action]	Simple interaction with software or tool
Given the correct reference values on the control document, failure to recognize any error in the “X, Y and Z” coordinates of the computed tomography in comparison with the given reference values.	Deviation from requirement not recognized [Detection and Noticing]	Quality check
Failure to identify correct ID from control document on the bite-block, couch or file etc. such that incorrect item is picked up	Patient information incorrectly matched [Detection and Noticing]	Identification of patient or patient related items

Draw suboptimal (incorrect or incomplete) contours around volumes of interest for every slice due to misunderstanding of the data	Misinterpretation of data [Sense-making and Understanding]	Complex interaction with software or tool
Selection of incorrect or suboptimal number of fields and their angles for a patient plan due to misunderstanding of the information	Misinterpretation of information [Sense-making and Understanding]	Iterative determination of optimum parameters
Give incorrect or incomplete critical data verbally over the phone	Communication failure [Team Coordination]	Verbal communication

Table 8 shows that certain failure modes are repeated across various GTTs, e.g. "Execute desired action incorrectly" and that some branching points are repeated across different DTs, "Distractions/interruptions and excessive workload" and "Information unclear". This may allow to reuse the elicited information across the different GTTs and DTs, thus further populating the whole set of DTs. Based on this, apart from the six failure modes directly elicited, five additional failure modes can be quantified because of repeated failure modes and four because of repeated branch points (**Table 9**). Indeed, future work may address the quantification of the remaining DTs.

Twelve experts with expertise in diverse aspects of radiotherapy took part in the elicitation: medical doctors, medical physicists, dosimetrists and radiation technologists. Each expert generally dealt with three failure situations, involving tasks that are part of his/her daily job. Single-expert sessions were designed. Prior to the sessions, a test run was performed with three experts, to get general feedback on the session design, organization and clarity. Each session lasted about two hours, including initial explanation, signing of a data confidentiality agreement and elicitation for the failure situations.

Each expert session was divided into two parts. The first part focusses on the identification of the negative conditions associated to each DT branching point, to validate or eventually complement the set developed in previous work [37]. For each task and for each branch point, experts are asked to think about the conditions that challenge the performance of the task. As noted above, these conditions are basically the PIF manifestations associated to each branch point. For example, negative conditions such as "distraction due to phone calls" and "distraction due to people coming in the room" would be manifestations of the PIF "Environment" and are associated to the "Distraction/ interruption and excessive workload" branch point.

The second part of the session aims at eliciting how strongly each negative condition in the branch point impacts the failure probability of the concerned GTT failure mode. To avoid direct

elicitation of failure probability values, the impact is elicited on a qualitative scale that verbally describes the likelihood of failure, given the negative condition is present (**Figure 19**). The probability values associated to each element of the scale are not presented to the experts and are used to process the expert judgments into probability distributions as explained in **Section 3.3.2: Expert data aggregation**. The use of the qualitative scale avoids known shortcomings of directly eliciting probability values, typically biases such as overconfidence, anchoring, availability [38, 39]. For the present case, elicitation of probability would even be more challenging as conditional probabilities are involved. The qualitative scale used for this study presented in **Figure 19** is adapted from the one presented as part of the ATHEANA HRA method [40]. Other literature sources like [13, 41, 42] are used to confirm the lowest value in the scale; for example, from reference [41, 42], it can be inferred that the base/nominal error rates in patient identification and data entry in healthcare would lie in the range of 1e-3 and 3e-3. The scale allows interpreting each expert assessment as a statement regarding the order of magnitude for the probability value for the HEP at the DT path for the GTT addressed by the elicited failure situation (**Table 12**).

Table 13 shows an example of the elicitation data. For each negative condition, the experts are asked to place placards on the appropriate level on the scale. In the case shown in the **Table 13**, five experts assess the impacts of each of the negative conditions falling in the three branch points of the failure situation representing "Simple interaction with software or tool GTT-Execute desired action incorrectly failure mode". These data are then aggregated using the methodology presented in **Section 3.3.2: Expert data aggregation**.

The expert elicitation only focuses on estimating HEPs for single factor effect, i.e. of single branch points (for example paths # 1, 2, 4 in **Figure 15**). Joint factor effects are an important element for an HRA method [25], and their incorporation in the model is planned in future work.

Impact	Descriptor	Meaning	Order of magnitude of probability
Low impact	Failure is not expected to happen, although I see how it could happen.	Given the negative condition, the desired task is still so easy that it is inconceivable that any personnel would fail if they were to experience this condition.	~1/1000
Moderate impact	Failures happen occasionally/ sometimes with such conditions	Given the negative condition, the desired task becomes moderately difficult that it is possible so that personnel would occasionally/ sometimes fail if they were to experience this condition.	~1/100
High impact	Failures happen often with such conditions	Given the negative condition, the desired task becomes highly difficult that is expected so that personnel would often fail if they were to experience this condition.	~1/10
Extreme impact	Failure is almost unavoidable	Failure is almost unavoidable. Almost all personnel would not be able to perform the desired task.	~1

Figure 19: Qualitative scale used during the exercise (probability values are not shown to the experts)

Table 13: Representation of the data obtained from the expert elicitation (example assessments)

Failure situation	**Branch point**	**Negative conditions**	**Exp1**	**Exp2**	**Exp3**	**Exp4**	**Exp5**
To transfer incorrect patient plan information (Offset data) from Patient Positioning and Verification to PatBase.	Information unclear	The indications for the input boxes on the screen are not readable	L	M	H	L	L
		The layout of the input interface is unusual or confusing	M	M	M	M	M
		Too much information on the screen leading to confusion	L	M	M	L	L
		Touch screen or input keyboard is very sensitive (unknown mis-selection is possible)	L	L	L	L	M
	Distractions/ interruptions and excessive workload	The background noise level is too high to distract the focus	M	M	M	L	M
		Simultaneously doing another urgent task	M	M	L	L	L
		Interruptions from colleagues while doing the task	L	M	M	M	M
		There is less time to do the task (excessive workload)	L	M	H	M	M
		Simultaneously tracking/supervising another task	H	H	L	L	M
	Low vigilance due to expecting no error	Task is mechanical and repetitive	H	H	E	E	H

Legend: L= Low impact, M= Moderate impact, H= High impact, E= Extreme impact

3.3.2: Expert data aggregation

The approach from reference [43] is used to aggregate the expert statements, e.g. from **Table 13**. This approach represents the HEPs produced by an HRA model as an inherently variable quantity, following the concept originally developed by reference [44] for general application to probabilistic safety assessment. This variability is due to the fact that the HEP values are associated to categories of tasks types and performance conditions, e.g. a method-specific PIF scale. Indeed, variability exists within the conditions covered by each category: for example, a specific PIF rating would manifest in different ways as specific conditions realize. Likewise, the HEPs to be assessed are associated to GTTs and DT branching points, which both represent categories of task types (the GTTs) and performance influences (the branching points), aggregating various specific conditions that result in the inherent variability of the HEP. The elicitation sessions address specific manifestations of these categories, i.e. specific tasks and performance conditions as shown in **Table 13**.

The expert data aggregation is based on a Bayesian model. Mathematically, the HEP is assumed to be lognormally distributed, with unknown median to be assessed based on the information from the experts, expressed as belief on the levels on the scale. The error factor (square root of the 95^{th} and 5^{th} percentiles) is assumed to be known and equal to three. The latter assumption is not a requirement of the approach, but largely decreases the amount of data required to be elicited to inform on the inherent variability of the HEP. It is indeed a typically used and accepted value in HRA.

The model requires to assign confidence in the experts being able to provide the correct value of the probability. This confidence is expressed in terms of a conditional probability that, given the real order of magnitude of the probability is one of the four in **Figure 19**, the experts would either assess the correct one or will be off by one or more orders of magnitude (G matrix in reference [43]). In this work, it is assumed that experts have about 80% probability to provide the correct order of magnitude, 10% of being one order of magnitude off, 5% of being two or more orders of magnitude off (these values are assumed equal for all experts). The exact values of these probabilities depend on the position of the interval with respect to the lower and upper bounds to have them normalized to a probability distribution. The assumed values appear to represent reasonable assumptions on the ability of the experts to provide correct estimates in this context.

The Bayesian model is used to aggregate the expert statement on each condition, thus obtaining the distribution reflecting the expert beliefs (i.e. for each row of **Table 13**). Then, the distribution for each condition needs to be combined to obtain the distribution for the branching point. Since the negative conditions are assumed equally likely, the final distribution is obtained as the average distribution across the negative conditions. In particular, as presented in reference [45], for each of the levels in **Figure 19**, the final degree of belief is the average degree of belief across the corresponding negative conditions. The aggregated distributions provide the degree of belief on which of the four values of the median characterizes the HEP distribution for the particular branch point. Then, the expected distribution of the HEP is obtained by weighting the four possible distributions.

3.4: Expert elicitation: results and discussion

3.4.1: Overall results from the elicitation

Generally, the conditions underlying each branching point presented to the experts represent well the possibly relevant performance issues. Few additional ones are identified by the experts to represent conditions specific to PSI's CPT.

With reference to the second part of the elicitation, eliciting the strength of each negative condition, **Figure 20** presents the posterior conditional HEP values obtained from the expert elicitation when each branch point is affecting the task, one branch at a time. The strongest impact on the HEP is determined by the "*Lack of training or experience*" branching point on the "*Complex interaction with software/tool*" (this is the branch point associated the highest median conditional HEP, in **Figure 20**). This is followed by "*Time pressure*" and "*Information unclear*" on "*Complex interaction with software/tool*" and "*Information content unclear-verbal*" on the "*Verbal communication*" GTT. According to the posterior HEP distribution, the HEP has more than 10% probability of lying in the 1e-1 region, if negatively affected by these branching points. The higher HEP values associated to the branching points generally reflect expert's view on the higher impact of these factors on the personnel performance. "*Training-experience*", and "*time pressure*" have key influence on tasks that require high cognitive effort and attention such as those enveloped by the two aforementioned GTTs. For "*Information unclear*", the experts identify the intensive use and interaction with multiple pieces of software in that specific GTT, which, if characterized by inadequate interface, leads to higher failure rates. Furthermore, for "Information content unclear", unclear verbal instruction or information given over the phone are very critical for the success of the verbal communication task.

Remarkably, similar branching points have different impact on different GTTs. As an example, the median failure probability related to the "Information unclear" branching point spans one order of magnitude across the GTTs. The lowest HEPs are expected for the "*Simple interaction with software or tool*" and "*Identification of patient or patient-related items*" GTTs, in the 1e-3 range. Again, this reflects the experts' view on the influence of the factor on the corresponding task. For the "Simple interaction with software or tool" GTT, the lower HEP reflects the fact that the task is generally simple with little interaction with the interface (e.g. inputting one value on a software input box). For this type of task, the influence of interface issues is perceived to be smaller than for the more complex tasks enveloped by the "*Complex interaction with software/ tool*" GTT (e.g. interaction with multiple pieces of software, investigation of images to find relevant tumor cells). Similarly, for the "*Identification of patient or patient-related items*" branching point, the experts motivate the low impact of the factor on the fact that the personnel verifies both the patient ID and, the name of the patient verbally, such that performance is generally robust to interface issues (which would only affect the recognition of the patient ID).

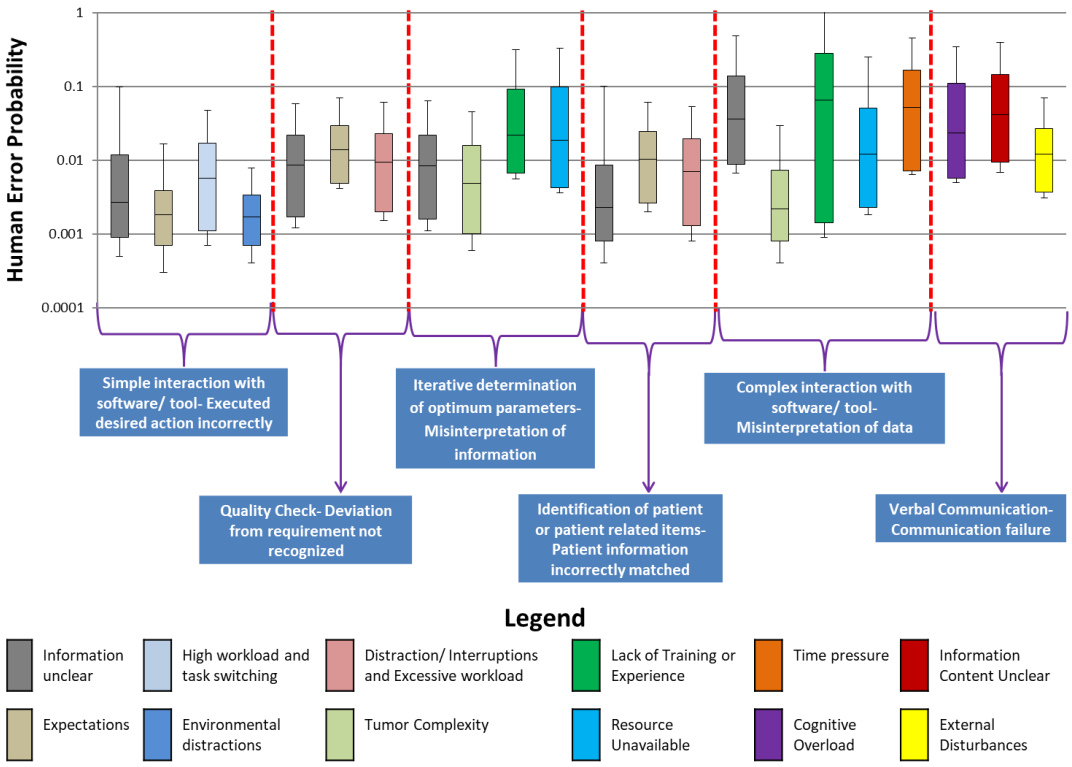


Figure 20: Box plot for human error probability from elicitation (median, 25th, 75th and the extremes as 5th and 95th percentiles); single branch point on GTT failure modes

3.4.2: Aggregation results from example GTT failure modes

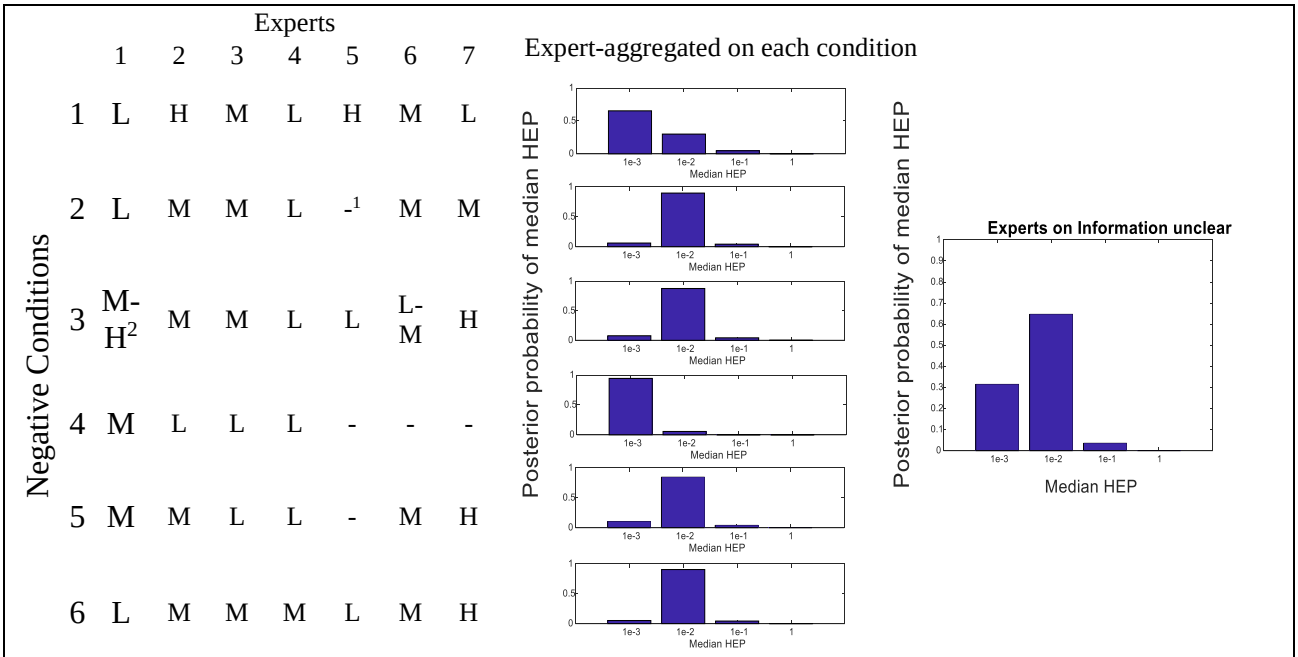
This section presents selected examples of the application of the aggregation approach outlined in **Section 3.3.2: Expert data aggregation** utilized to derive the failure probability distributions for **Figure 20**, specifically for the "Simple interaction with software/ tool" GTT – other examples can be found in [45].

Figure 21 and **Figure 22** show the processing of the expert evaluations to obtain the posterior probability distribution of the median HEP, conditional on the different branch points. The left part of **Figure 21** and **Figure 22** shows the experts judgments for each negative condition. The middle part shows the posterior probability distribution of the HEP as affected by each negative condition (represented on each row). Each of the distributions in the middle represents the aggregation of the expert judgments for each negative condition. The right part shows the final aggregation across all conditions.

Figure 21 addresses the branching point "Information unclear" of the GTT failure mode "Quality Check- Deviation from requirement not recognized". Seven experts are involved. For the first negative condition, three experts provided the judgment of "Low", two of "Moderate", two of "High" (#1 on the left of **Figure 21**). Correspondingly, the aggregated distribution in the middle shows a maximum on the median HEP associated to the "Low" judgment (i.e. 1e-3), some probability mass for "Moderate" (1e-2), with some tail on "High" (1e-1). Differently, for the second negative condition, two experts assess "Low" impact; while four experts assess "Moderate" impact (one expert does not provide any judgment). The corresponding aggregated distribution reflects the stronger agreement across experts compared to the previous case, with a more prominent maximum on "Moderate" (1e-2). Similar considerations can be made for the other conditions. Finally, the aggregated distribution for the branch is obtained as the average probability value across the conditions for each level of the assessment scale and it reflects the generally larger mass on "Moderate", then on "Low" and finally on "High" of the six condition-specific distributions.

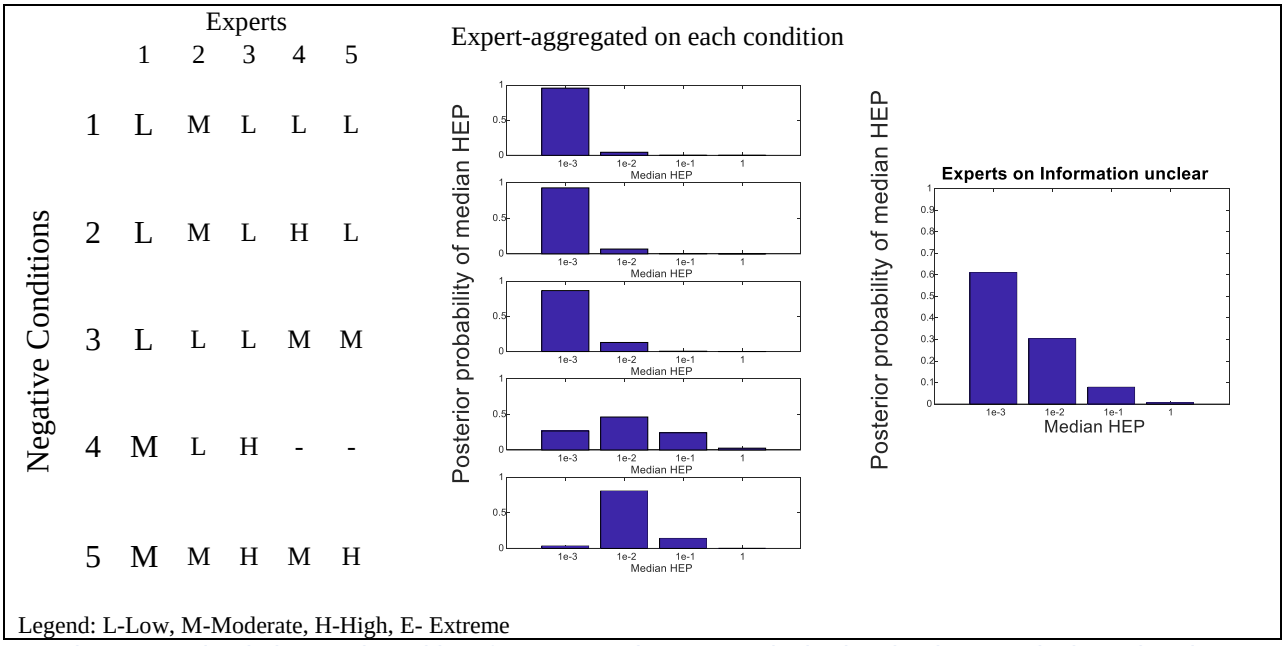
Figure 22 shows the results for the GTT failure mode "Simple interaction with Software/tool-executed desired action incorrectly" and the branching point "Information unclear". **Figure 22** shows that different levels of agreement across the five experts (left pane) convert into the aggregated distributions (center pane). This is evident comparing the judgments and the corresponding distribution for the first condition and those for the fourth. The aggregated distribution in the right pane reflects the generally larger mass on "Low", then on "Moderate",

and finally on “High”. The distribution comes from the experts’ belief that interface issues do not have large impact on the personnel reliability on the tasks characterized by interactions with simple interfaces.



1- Expert did not provide a judgment
2- Expert could not decide on a single judgment, so elicited two levels
Legend: L-Low, M-Moderate, H-High, E- Extreme

Figure 21: Quality Check- Deviation from requirement not recognized, branch point: Information unclear Left: judgements from experts, Middle: expert-aggregated posterior distribution of median HEP for each condition, Right: posterior probability distribution of median HEP for the branch point



Legend: L-Low, M-Moderate, H-High, E- Extreme

Figure 22: Simple interaction with software or tool- Execute desired action incorrectly, branch point: Information unclear. Left: judgments from experts, Middle: Expert-aggregated posterior probability distribution of median HEP for each of the five negative conditions, Right: Probability distribution of HEP of the complete branch point

3.4.3: Lessons learned from the expert elicitation exercise

This section presents some insights on the expert elicitation exercise. Given the number of experts involved in the elicitation, single expert sessions were preferred to group sessions. The single-expert sessions allowed each expert to have adequate time to provide and discuss his/her judgments. The sessions were facilitated by one or two human reliability analysts and a medical physicist, well familiar with both the therapy process and HRA method development work. The presence of the latter, a coworker at the therapy center, was very helpful to connect the elicited situations to the daily experience of the experts. All experts deeply engaged in the sessions; no signs of reluctance to share their views were noted.

As mentioned earlier, the main goal of the first part of the elicitation was to identify the relevant negative conditions underlying each branching point. To avoid any potential bias to the experts, the initial strategy for the elicitation was to let the experts themselves identify the conditions, without showing the set of the pre-identified conditions [37] – indeed few conditions were provided just as examples for the experts. This "white sheet" strategy turned out to be very tiring for the experts and of little efficiency: much of the conditions provided by the experts overlapped so that little new information was coming after 2-3 experts had evaluated the same branching point. It was then decided to adopt the "white sheet" strategy only for the first couple of experts for each branching point; then the rest of the experts were provided with the consolidated set of conditions and asked for possible integration. This allowed moving more efficiently to the second part of the sessions, with longer time dedicated to discuss the rationale behind the impact judgments.

An important feature of the elicitation of the factor impact has been to address the impact of specific negative conditions of the performance of specific tasks. For example, experts were asked to assess the impact of high background noise (negative condition #3 in **Table 10**), while transferring patient plan information on software (failure situation #1 of **Table 12**). A different approach would have been to ask about the impact at the branching point and GTT levels; this would have entailed referring to the impact of distractions and interruptions (branching point #3, **Table 11**), while performing a simple interaction task with software or tool (GTT 1 of **Table 8**). The former approach was preferred for the session design and turned out to allow experts to easily recognize daily practices, relate situations to occurred events, and specific performance issues. A critical aspect of this approach is how the specific situations generalize to the overall category they are called to represent. Concerning the branching points, as shown in **Table 11**, a set of about four conditions is addressed per branching point, indeed to ensure that the

performance issues underlying each branching point are adequately covered. Concerning the tasks, besides the arguments presented in **Section 3.3.1: Expert elicitation**, indeed the choice felt on those tasks deemed as well representative of the corresponding GTT. This aspect may be subject to confirmation in future work, ideally by organizing additional elicitation sessions to address other representative tasks. Finally, on this issue, it is important to mention that the expert aggregation approach adopted (**Section 3.3.2: Expert data aggregation**) does consider mathematically the possible variability of the specific tasks and conditions within their categories, giving proper credit to the uncertainty that comes from addressing specific instances of those.

3.5: Convergence validation

According to [30], different types of validation can be distinguished according to the quality of the data used for validation. Absolute validation refers to the use of unbiased data, from real operational experience. Approximate validation denotes the use of other data sources such as simulator, literature, expert judgment. For applications for which known data are not available, comparison of the model prediction with other model predictions may give an idea of the convergence of the models [30]. This type of validation is carried out in the present paper. **Section 3.5.1: Identification of relevant data: criteria** presents the criteria to select applicable data from existing HRA methods, **Section 3.5.2: Data applicability: results** presents the data used for validation and provides some general considerations on their applicability, and **Section 3.5.3: Comparison with HRA data** compares the data with the results from the elicitation.

3.5.1: Identification of relevant data: criteria

Two HRA methods are selected for identification of relevant data, i.e. the Technique for Human Reliability Analysis (THERP) [9] and CARA [16]. The former is selected because it represents the fundamental data basis for most of the subsequent HRA methods; also, most of the data underlying THERP refer to detailed manipulation tasks and interactions with human-machine interfaces, which arguably are mostly dependent on the interfaces themselves, as opposed to the industrial sector and on the type of task. CARA is selected because its GTTs are defined based on the involved cognitive functions, in a similar way as done in the present work. This simplifies the identification of relevant data and, ideally, allows verifying if and how the different application field has implications on the relevant influencing factors and on their effect on the HEPs.

Two sets of criteria are defined to identify relevant data. For nominal HEPs, i.e. HEPs that are intended to be applicable to the failure of the task under the nominal, expected performance conditions (all PIFs rated as nominal, with no specific error-forcing influence):

(1) Involved cognitive function and associated failure mode and,
(2) Context, broadly defined as including interactions with the system and general performance conditions.

Criterion (1) addresses the essential cognitive features of the task and of its failure. Indeed, a prerequisite for the applicability of an HEP value is that it addresses the same cognitive failure. Criterion (2) broadly addresses all other task features beyond the pure similarity of the cognitive function and its failure mode, i.e. beyond criterion (1). Criterion (2) flags any features of the radiotherapy task or of the task associated to the THERP/CARA value that may not suggest convincing relevance (for example due to the use of very different human-machine interfaces, or to the difference scope of the procedural guidance).

The second set of criteria is used to evaluate data for the DT branching points:

(3) Similarity in the manifestation of the influencing factors and,
(4) Similarity in strength of the impact of the specified condition.

Criterion (3) addresses the similarity between the branching point conditions and the performance influences reflected in the HRA data. These influences may be represented as multipliers or absolute HEP values depending on the HRA method. Indeed, criterion (3) reflects a fundamental prerequisite for the applicability of the data. Criterion (4) assesses the strength of the impact, which also has to be aligned for the condition to be applicable. For example, some CARA error-producing conditions reflect very strong negative performance situations that are not expected in the radiotherapy domain (e.g. stress facing a terrorist attack). The effect of these conditions need to be taken into consideration to assess the applicability.

All criteria are evaluated based on the same scoring guidance:

A (Acceptable) - Good match between the two, similarities are convincing,

P (Pass) - Differences are identified, but not such to reject the data point

N (Not acceptable) - Differences are too large to allow use of the data point.

If a method's data point is assigned at least a score of "P" for the criterion (1) or (3), only then it is evaluated based on criterion (2) or (4), respectively. A value is relevant if assigned at least a

score of "P" to criterion (2) or (4), respectively. In other words, the value is rejected if assigned a score of "N" in any criteria.

3.5.2: Data applicability: results

A total of 14 out of the 18 nominal HEPs (one for each DT) are quantified using HRA methods. The failure modes for which no data are applicable, are related to the "sense-making and understanding" and "decision-making" macrocognitive functions. The lack of applicable data for these functions is to some degree not surprising because these tasks are strongly domain-specific as well as depend on the training and experience of the performers.

Concerning data from THERP, HEP values are found from THERP Tables 20-10, 20-11 and 20-12 for errors of commission (EOCs); and 20-6 for errors of omission (EOOs) [9]. The typical tasks for which data can be found include checking and comparing items or objects, interacting with man-machine interfaces and instrumentation, and selecting and manipulating physical controls. The typical HEP values from THERP are in the order of 10^{-3} for both EOCs and EOOs. For certain task types, no THERP data are relevant. These tasks relate to verbally communicating instructions, planning and decision-making, i.e. the following GTTs: "Verbal communications", "Iterative determination of optimum parameters". For detailed selection of data from THERP, the readers can refer to [46].

Concerning CARA, a good match is identified for CARA GTT G2 [47], related to physical slips while interacting with the human-machine interface, well applicable to the radiotherapy domain as well (Item # 2 in **Table 14**, hence the score "AA"). For the other HEP values, less convincing matches are found, resulting in the "AP" scores given in **Table 14**. For example, GTT B1, related to visual slips in reading data, is also applicable, although the CARA value appears to refer to more complex interfaces and larger information content than in the corresponding radiotherapy tasks (Item # 1 in **Table 14**). Additionally, GTT F relates to routine elements of air traffic sector management typically involving decisions based on established, well-known rules. A similar situation exists for radiotherapy, where in most situations default therapy plan options exist based on defined rules for different tumors. On the other hand, the number of decisions required to customize a radiotherapy plan, even for routine situations, seems to be larger than what is implied in the description of the CARA GTT. In radiotherapy, the generation of the optimal plan typically involves problem-solving and planning, depending on the possible beam angles or constraints of the tumor (like its location etc.). Generally, the HEP values from CARA range from $2 \cdot 10^{-3}$ to $5 \cdot 10^{-3}$.

The quantification of the relevant branching points is presented in **Appendix 9**. From THERP, data are found for the three branching points: "*Information unclear*" (B1), "*low vigilance*" (B2), and "*Distractions/interruptions*" (B7); from CARA, relevant data are found for all the branching points. A data point is also added from the recent consolidated version of the HEART EPCs [8], concerning distractions/interruptions due to poor working environment. For branch points B1 and B7, three variants are distinguished to reflect different manifestations of the branching points for different GTTs (**Appendix 9**). Concerning the former, B1 in general addresses issues with the way in which information is presented: B1a models various interface issues; B1b focusses on indication clarity (therefore, it has a reduced scope compared to B1a); B1c deals with issues in the information delivered verbally. Concerning the distraction/interruptions branch point (B7), the three variants distinguish the combined effect of distractions due to the environment and task switching (B7a) from the two effects taken singularly for application to the "Simple interaction with software or tool" GTT (B7b and B7c).

Overall, fifteen data points (multipliers or directly HEP values depending on the method) are found relevant from CARA, HEART and THERP for the seven branch points. Out of the fifteen, eight receive "AA" scores, five receive "PA" or "AP" scores and two "PP" scores. The most convincing matches are those related to the "*Information unclear*" ("*Problematic interface*", "*problematic indicators*", and "*information content clarity*", B1a, B1b and, B1c respectively in **Appendix 9**), "*lack of training or experience*" (B4) and "*Software or resource issues*" (B5). In general, the good match may be attributed to the similarity of the issues concerning these factors across different industries – specific justifications are given in **Appendix 9**. For other branching points, the data applicability is less convincing. Generally, the performance influences associated to these branching points are more specific to the application field. As an example, consider the "*Task complexity*", B3, branching point. The way in which complexity of the tasks manifests is closely linked to the application domain. For the air traffic domain, complexity relates to the air traffic dynamics, the number and the routes of the aircrafts involved, possibly evolving quickly over time. For radiotherapy, complexity relates to how the type, position, shape and size of tumors affect the definition of the therapy plan. These very different situations would not warrant a convincing match.

In other cases, the partial match is generally due to the difference in the factor scope. For example, for the "*Distractions/ interruptions and excessive workload*" (B7a) branch point, the THERP data point only represents the effect of heavy workload, and not of distractions by the environment (e.g. excessive background noise, distracting activities carried out in the

surroundings) or interruptions by other personnel. Similarly, the CARA datum refers to the effect of workload, distractions and interruptions due to task switching and personnel interference, but does not cover the environment effect.

Note that in the CARA and HEART methods, the impact of each Error Producing Condition (EPC) is scored depending on the strength of the influence of the condition. In most cases, the data from these methods is found to be applicable at the "maximum affect" of the EPC. In some cases, the maximum affect is judged to reflect too strong negative conditions with respect to those covered by the corresponding branching points or the EPC scope is larger than the one of the branching points. For example, the "*Unfamiliarity and adequacy of training/experience*" EPC from CARA covers rare events and scenarios never seen in training or experience like terrorist events or total radar loss. These extreme cases of unfamiliar situations are not expected for the present application; therefore, it is deemed that the branching point situations are closer to the graded affect (0.5 proportion, see B4 in **Appendix 9**). Concerning difference in the scope, an example would be the applicability of the CARA EPC "*Poor, ambiguous or ill-matched system feedback*" for branching point B1b "Problematic indicators" the latter has larger scope including layout issues, information availability etc., compared to the former which is focused on indicators. For each data point, the applicability of the strength of the impact is rated by second criterion (justifications in **Appendix 9**).

Table 14: Relevant CARA data to GTT-failure mode

		Criterion 1: cognitive function & failure mode		**Criterion 2: context**	
#	**HEP (CARA GTT)**	**Score**	**Comment**	**Score**	**Comment**
1	0.005 (GTT B1, active search of radar, assuming some confusable information on display) [47]	A	[47]'s GTT basically refers to visual slips in reading data, related to text, colours, strings of numbers. All this applies well to the GTT failure modes.	P	The quantity of information (and therefore the potential for confusion) expected in Air Traffic Control radar screens may be much larger than what expected for patient identification
2	0.002 (GTT G2, physical slips) [47]	A	[47]'s GTT refers to physical slips while inputting data and more generally interacting with man-machine interfaces	A	[47]'s GTT is not expected to depend on the specific industry but is driven by the quality of the interface. The CARA GTT represents a generic value applicable to different controls and interfaces, keypads, graphical user interfaces. This covers well the variety expected in radiotherapy domain

3	0.003 (GTT F, Routine element of sector management (e.g. rule-based selection of routine plan for an aircraft or omission of clearance) [47]	A	Typically, in radiotherapy, there is a default plan option which can be selected based on the defined rules for different tumors	P	Although rules exist to select a default plan option, the amount of customization required to generate the optimal plan appears higher than for the air traffic control domain, even in routine situations
4	0.002 (GTT G1, Verbal slips) [47]	A	The tasks in this belong to verbal slips and misses. These two GTTs completely match.	A	[47]'s GTT completely matches with the radiotherapy GTT

3.5.3: Comparison with HRA data

Figure 23 and **Figure 24** compare the expert judgment results and the data from HRA methods. **Figure 24** shows the deviation across the GTTs and the branch points associated to the elicited failure situations. The deviation factors are summarized in **Table 15**. The median values are provided in **Appendix 10**, with details on their calculation for the CARA and THERP methods. In view of the large uncertainties in the values and of the different application domains, a detailed quantitative comparison or statistical analysis of the differences between the values based on expert judgment and the HRA data could overstate the strength of this evidence. Some general considerations on the comparison seem more appropriate, with the overall aim to understand these differences and trace those to arguments by the experts, to verify if these are plausible in view of the differences in the tasks and in the application domains.

Table 15: Deviation factor in the 32 comparisons (10 with data from THERP, 22 with data from CARA) vs expert judgment results

Deviation between median values	#cases (32 total)	In Figure 23
Below a factor of 3	19	points within the dotted lines
Between 3 and 5	4	points between the continuous and dotted lines
Between 5 and 10	7	points between the dashed and dotted lines
Larger than a factor of 10	2	beyond dashed lines

Much of the largest deviations (e.g. above a factor 5) are located in the upper left part of **Figure 23**, with values from expert judgment lower than the HRA data. In general, these are cases in which experts assess low impact for the negative conditions, i.e. for various reasons, this low impact is not reflected in the HRA data. For example, the case with the largest deviation, i.e. Triangle (5) in **Figure 23**, relates to the branching point "*Tumor complexity*", applied to GTT "*Complex interaction with software/tool*". As discussed in **Section 3.5.3: Comparison with HRA data**, the related data point from CARA is assigned "PA" score, in view of the very

different way in which complexity manifests in the two application sectors (see also the detailed evaluation in **Appendix 9**, branch B3). In particular, experts do not feel that complexity would have large impact on the failure probability, i.e. the development of a complex therapy plan would be dedicated more time and resources. On the other hand, the much larger failure probability from the CARA method possibly reflects the much larger impact of complex traffic dynamics, which has to be dealt with within a generally tight timeframe. Similar considerations can be made for Triangle (3) in **Figure 23**, again related to branch point "*Tumor complexity*" applied to GTT "*Complex interaction with software or tool*". For Triangle (12) in **Figure 23**, "*Information unclear*" applied to GTT "*Identification of patient or patient-related item*", interface issues are considered not problematic because of the diverse ways of checking the patient identity (not just by ID, but also verbally and with picture). For other points, such as (1), (29), (18), (31) the experts generally agree that these tasks are very simple with low cognitive requirements and simple interactions, therefore, they are less impacted by attention losses, interruptions, and interface issues. Remarkably, the same branch points are assessed to have larger impacts for other types of tasks. For example, for the two branching points "*Information unclear*" and "*Low vigilance due to expectation*" the CARA data is around 10^{-2} for all applicable GTTs, with little discrimination (**Figure 23**, Triangles (6), (9), (12), (14) and (13), (15), (19), respectively; **Figure 24**, along columns "*Information unclear*" and "*Low vigilance due to expectation*"). Indeed, the deviating cases reflect situations that are specific to the particular radiotherapy tasks and contexts analyzed by the experts. It can then be expected that these sector-specific influences are not fully reflected by the HRA data.

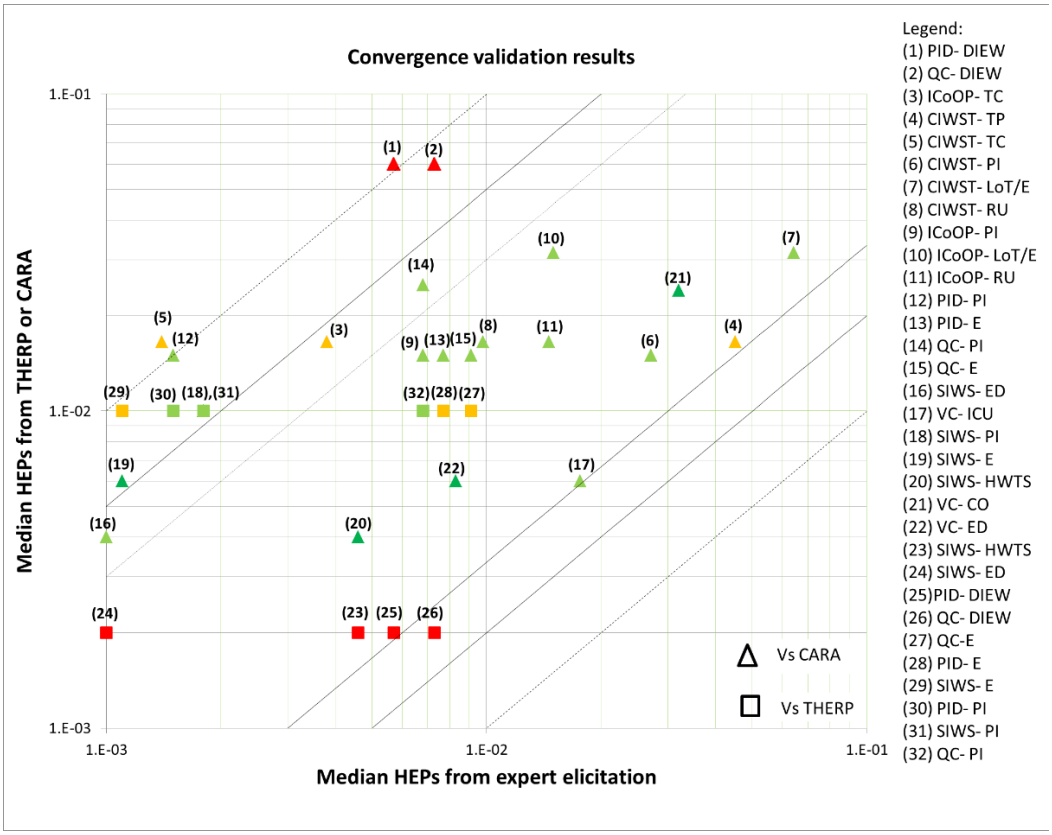


Figure 23: Convergence validation results of elicitation and HRA data

Abbreviations for **Figure 23**:

- for GTTs, PID: Identification of patient or patient related items, QC: Quality Check, IDoOP: Iterative determination of optimum parameters, CIWST: Complex interaction with software or tool, SIWS: Simple interaction with software or tool, VC: Verbal communication;
- for branch points, DIEW: Distractions/interruptions or excessive workload, TC: Tumor complexity, TP: Time pressure, PI: Problematic interface, LoT/E: Lack of training/experience, RU: Resources unavailable, E: Expectations, ED: Environmental distractions, ICU: Information content unclear, HWTS: High workload and task switching, CO: Cognitive overload.

Color code for **Figure 23**: "Dark green" when both branch point and base failure probability received "AA" score, "Light green" when one of them received "AA" and the other either "AP" or "PA", "Yellow" when both receive either "AP" or "PA" and "Red" when one of them received any "PP".

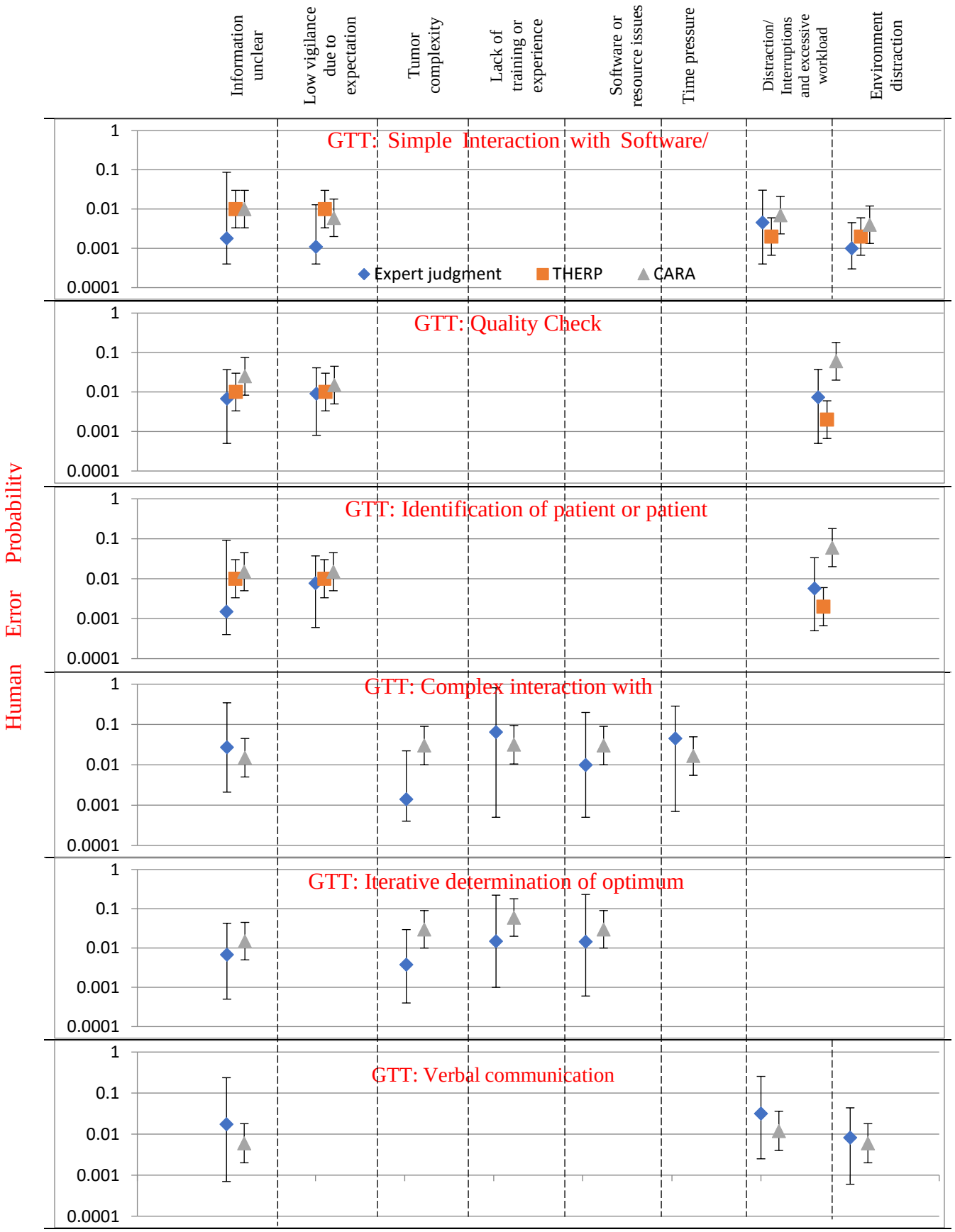


Figure 24: Comparison of results from elicitation with HRA data. Symbols identify medians; error bars identify 5th and 95th percentiles of the uncertainty distribution.

3.6: Conclusions

This paper presents the quantification structure of an original HRA method for application to the radiotherapy domain. It includes the development of the quantification framework (decision trees), quantification of the Human Error Probabilities (HEPs) using expert judgment and a partial validation of the results against data from existing HRA methods. The approach to develop the quantification structure is not domain-specific and can be applied for the development of HRA methods for novel domains.

Eighteen decision trees are developed, built on the GTT (failure mode)-PIF structures presented in reference [29]. Specific negative performance conditions for each branch point of the DTs serve as guidance for the analyst on the choice of DT path. To quantify the HEPs for each DT paths, an expert judgment elicitation approach is adopted, with the aim to elicit the order of magnitude of the HEP. The elicitation sessions are designed encompassing two key features: (1) information on probabilities is asked to experts on a qualitative scale, with the goal of getting evidence on the order of magnitude for the probability; and (2) specific situations are presented to the expert, i.e. specific failure scenarios influenced by specific negative conditions. The latter is incorporated to avoid that the experts deal with abstract categories such as tasks types and influencing factors. Using the Bayesian model presented in reference [32], the qualitative expert data are aggregated in a systematic and traceable way to determine the degrees of belief on the correct values of the HEPs.

According to the elicitation results, the strongest effects on the HEP come from the "*Lack of training or experience*" branching point on the GTT "*Complex interaction with software/tool*", followed by "*Time pressure*" and "*Information unclear*" on "*Complex interaction with software/tool*" and "*Information content unclear-verbal*" on the "*Verbal communication*" GTTs. Indeed, "*Training*", "*experience*", and "*time pressure*" have key influence on tasks that require high cognitive effort and attention such as those enveloped by the two aforementioned GTTs. For "*Information unclear*", the experts identify the intensive use and interaction with multiple pieces of software in that specific GTT, which, if characterized by inadequate interface, leads to higher failure rates. Furthermore, for "*Information content unclear*", unclear verbal instruction or information given over the phone are very critical for the success of the verbal communication task.

Relevant data from other HRA methods is also sought to quantify the DTs and provide validation to the expert judgment results. The data applicability is convincing for tasks involving simple interactions with interface and instruments. The failure modes for which no data are found, are related to the "*sense-making and understanding*" and "*decision-making*" macrocognitive functions. In case of the branch points, relevant data are found for all the branch points, with most convincing matches for data concerning interface and training issues.

When comparing the data from the elicitation and the HRA methods; out of 32 comparisons, in 17 cases the deviation between the median values is below a factor of 3, in 5 cases it is between 3 and 5, in 8 cases between 5 and 10, and in 2 cases it is larger than 10. The cases with larger deviation reflect situations that are specific to the particular radiotherapy tasks and contexts analyzed by the experts. It is expected that these sector-specific influences be not fully reflected by the data. On the one hand, the satisfactory agreement between the expert judgment and method data is a suggestion that HRA data may be used across industries for tasks and influences that are not specific to one industry. On the other hand, it further confirms that expert judgment can provide valuable input for quantification of HRA methods. This encourages the continuous effort to develop methods to elicit judgments, rigorously and traceably, to supplement for missing data.

It is important to mention that quantification of the influence of multiple factors on the HEP is left out of the scope of the present paper. Future work may be devoted to the quantification of this important aspect. Additional expert elicitation sessions may be developed to decrease the large uncertainty ranges for some of the branch points. Lastly, future work will apply the developed HRA method to hypothetical accident scenarios at the institute's Center for Proton Therapy.

References:

1. Kirwan B. A guide to practical human reliability assessment. Boca Raton: CRC Press, 1994.
2. Spurgin AJ. Human reliability assessment theory and practice. Boca Raton: CRC Press, 2010.
3. Podofillini L. Human reliability analysis. In: Moller N, Hansson SO, Holmberg JE and Rollenhagen C, editors. Handbook of safety principles, Wiley, 2018.
4. Mosleh A. PRA: A perspective on strengths, current limitations and possible improvements. Nuclear Engineering and Technology 2014; 46(1): 1-10. DOI: http://dx.doi.org/10.5516/NET.03.2014.700

5. Kirwan B, Gibson WH and Hickling B. Human error data collection as a precursor to the development of a human reliability assessment capability in air traffic management. Reliability Engineering & System Safety 2008; 93(2): 217-233.
6. Kirwan B and Ainsworth LK. A guide to task analysis. Boca Raton: CRC press, 1992.
7. Williams JC. HEART—a proposed method for assessing and reducing human error. In: Proceedings of the 9th advances in reliability technology symposium, Bradford; 1986. National Center for Systems Reliability.
8. Williams JC. HEART—a proposed method for achieving high reliability in process operation by means of human factors engineering technology. Safety and Reliability 2015; 35(3): 5–25.
9. Swain AD and Guttmann HE. Handbook of human reliability analysis with emphasis on nuclear power plant applications. Final report NUREG/CR-1278, August 1983. Washington, DC: US Nuclear Regulatory Commission.
10. Prvakova S and Dang VN. A review of the current status of HRA data. In: Steenbergen, RDJM, Gelder PHAJM, van Miraglia S and Vrouwenvelder ACWM, editors. Safety, reliability and risk analysis: beyond the horizon, European Safety and Reliability Conference, ESREL CRC press; 2013. p. 585-593.
11. Chang JY, Bley D, Criscione L, Kirwan B, Mosleh A, Madary T, Nowell R, Richards R, Roth EM, Sieben S and Zoulis A. The SACADA database for human reliability and human performance. Reliability Engineering & System Safety 2014; 125: 117-133.
12. Kim Y, Park J and Jung W. A classification scheme of erroneous behaviors for human error probability estimations based on simulator data, Reliability Engineering & System Safety 2017; 163: 1-13.
13. Gibson HW, Basra G and Kirwan B. Development of the CORE-DATA database. Safe Reliability 1999; 19(1): 6–20.
14. Kirwan B, Gibson HW, Kennedy R, Edmunds J, Cooksley G and Umber I. Nuclear action reliability assessment (NARA): a data-based HRA tool. Safety and Reliability 2005; 25(2): 38-45.
15. Gibson HW, Mills A, Smith S and Kirwan B. Railway action reliability assessment, a railway-specific approach to human error quantification. In: Dadashi N, Scott A and Wilson JR editors. Proceedings of the international rail human factors conference, rail human factors supporting reliability, safety and cost reduction, London, Abingdon: Taylor & Francis Group 2013; pp: 671–676.
16. Kirwan B and Gibson WH. CARA: a human reliability assessment tool for air traffic safety management— technical basis and preliminary architecture. In: Redmill F and Anderson T, editors. The safety of systems. London: Springer; 2007, pp:197–214.
17. Rasmussen M, Standal MI and Laumann K. Task complexity as a performance shaping factor: a review and recommendations in Standardized Plant Analysis Risk- Human Reliability Analysis (SPAR-H) adaption. Safety Science 2015; 76: 228–238.
18. Laumann K, Oien K, Taylor C, Boring RL and Rasmussen M. Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability analysis methods. In: Smith C and Paulos T, editors. Proceedings of the probabilistic safety assessment and management, Honolulu: CreateSpace Independent Publishing Platform;2014.

19. Li PC, Chen GH, Dai LC and Zhang L. A fuzzy Bayesian network approach to improve the quantification of organizational influences in HRA frameworks. Safety Science 2012; 50 (7): 1569-1583.
20. Martins MR and Maturana MC. Application of Bayesian belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. Reliab Eng Syst Saf 2013; 110: 89-109.
21. Trucco P, Cagno E, Ruggeri F and Grande O. A Bayesian belief network modelling of organisational factors in risk analysis: a case study in maritime transportation. Reliab Eng Syst Saf 2008; 93 (6): 845-856.
22. Ale BJ, Bellamy LJ, Cooke RM, Goossens LHJ, Hale AR, Roelen ALC and Smith E. Towards a causal model for air transport safety-an ongoing research project. Saf Sci 2006; 44: 657-673.
23. Ale BJ, Bellamy LJ, Van der Boom R, Cooper J, Cooke RM, Goossens LHJ, Hale AR, Kurowicka D, Roelen ALC and Spouge J. Further development of a causal model for air transport safety (CATS): building the mathematical heart. Reliab Eng Syst Saf 2009, 94 (9): 1433-1441.
24. Musharraf M, Hassan J, Khan F, Veitch B, MacKinnon S and Imtiaz D. Human reliability assessment during offshore emergency conditions. Saf Sci 2013, 59: 19-27.
25. Mkrtchyan L, Podofillini L and Dang VN. Bayesian belief networks for human reliability analysis: A review of applications and gaps. Reliab Eng Syst Saf 2015; 139: 1-16. https://doi.org/10.1016/j.ress.2015.02.006
26. International Atomic Energy Agency (IAEA). Lessons learned from accidental exposures in radiotherapy. Safety report—IAEA-Series no. 17, 2000. Vienna: IAEA.
27. World Health Organization (WHO). Radiotherapy risk profile (Technical manual), 2008. Geneva: WHO.
28. Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. Med Phys 2016; 43(7): 4209–4262.
29. Pandya D, Podofillini L, Emert F, Lomax AJ and Dang VN. Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy. Proc IMechE Part O: J Risk and Reliability 2017; First published October 2: 1–35. https://doi.org/10.1177/1748006X17731903
30. B Kirwan. Validation of Human Reliability Assessment techniques: part 1 – validation issues. Safety Science 1997; 27(1): 25-41.
31. Whaley AM, Xing J, Boring RL, Hendrickson SML, Joe JC, Le Blanc KL and Morrow SL. Cognitive basis for human reliability analysis. Technical report NUREG- 2114, 2016. Washington, DC: US Nuclear Regulatory Commission.
32. Electric Power Research Institute. An approach to the analysis of operator actions in Probabilistic Risk Assessment, Technical report-100259, EPRI 1992; Palo Alto, CA.
33. Xing J, Parry GW, Presley M, Forester J, Hendrickson S and Dang V. An Integrated Human Event Analysis System (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application (NUREG-2199, Volume 1), 2017. Washington, DC: US Nuclear Regulatory Commission.
34. Radiation Oncology Safety Information System (ROSIS), http://rosis.ch/ge/rosis_daten1.asp

35. Safety reporting and learning system for radiotherapy (SAFRON), https://rpop.iaea.org/RPOP/RPoP/Modules/login/safron-register.htm
36. Ford E, Fong de Los Santos L, Pawlicki T, Sutlief S and Dunscombe P. Consensus recommendations for incident learning database structures in radiation oncology. Med Phys 2012; 39(12): 7272–7290.
37. Pandya D, Podofillini L, Emert F, Lomax AJ and Dang VN. A method for human reliability analysis in radiotherapy: identification and characterization of influencing factors. In: Podofillini L, Sudret B, Stojadinovic B et al., editors. Proceedings of the European safety and reliability conference; Abingdon: Taylor & Francis Group 2015.
38. Meyer MA and Booker JM. Eliciting and analyzing expert judgment: a practical guide. London: Academic press, 1991.
39. Kahneman D, Slovic P and Tversky A. Judgment under uncertainty: heuristics and biases. Cambridge: Cambridge university press, 1982.
40. US Nuclear Regulatory Commission. Technical basis and implementation guidelines for A Technique for Human Error Analysis (ATHEANA). Technical report NUREG-1624, rev 1, May 2000. Washington, DC: US Nuclear Regulatory Commission.
41. Wahi MM, Parks DV, Skeate RC and Goldin SB. Reducing errors from the electronic transcription of the data collected on paper forms: a research data case study. J Am Med Inform Assoc 2008; 15(3): 386-389. doi: 10.1197/jamia.M2381.
42. Salinas M, Lopez-Garrigos M, Lillo R, Gutierrez M, Luga J and Leiva-Salinas C. Patient identification errors: The detective in the laboratory. Clinical Biochemistry 2013; 46: 1767-1769.
43. Podofillini L and Dang VN. A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. Reliab Eng Syst Saf 2013; 117: 52-64.
44. A Mosleh. Bayesian modeling of expert-to-expert variability and dependence in estimating rare event frequencies. Reliability Engineering and System Safety 1992; 38: 47-57.
45. Podofillini L, Pandya D, Emert F, Lomax AJ, Dang VN and Sansavini G. Bayesian aggregation of expert judgment data for quantification of human failure probabilities for radiotherapy. In: Proceedings of European safety and reliability conference; Norway, Taylor & Francis Group 2018 (under publication)
46. Pandya D, Podofillini L, Emert F, Lomax AJ, Dang VN and Sansavini G. Quantification of human failure probabilities for radiotherapy: relevance of THERP's values. In: Cepin M and Bris R, editors. Proceedings of European safety and reliability conference; Slovenia, Taylor & Francis Group 2017.
47. Gibson H and Kilner A. Controller Action Reliability Assessment (CARA) user manual. Technical report for European organization for the safety of air navigation.7th August 2009.

Chapter 4: HRA method for radiotherapy: overview

The goal of this chapter is to present the overview of the HRA method. In previous chapters, the elements of the method were developed separately, thus, this chapter aims to combine the qualitative and quantitative results obtained in **Chapter 2: HRA method- qualitative aspects** (i.e. PIFs with negative conditions, identification of GTTs and GTT-PIF structures) and **Chapter 3: HRA method- quantitative elements** (i.e. decision trees (DTs) developed from GTT-PIF structures and the HEPs of selected DTs obtained from experts) respectively, and present them as a compact HRA method. This includes the method structure, the assumption and limitations of the method. More precisely, the overview of the method aims to additionally cover parts of the method that were not addressed in previous chapters. Indeed, quantification of more than single branch point impacts was not covered in **Chapter 3: HRA method- quantitative elements**; also, not all DT branches were quantified.

The Chapter closes with a short method application guidance (i.e. the steps to be followed by an analyst for the application of the method).

4.1: Outline of the radiotherapy HRA method

The HRA method for radiotherapy has been developed based on a quantification model that can represent the causal influences on failures identified by the GTT-PIF structures, i.e. the DT (see **Chapter 3: HRA method- quantitative elements** for DTs) and with an underlying direct link to a cognitive model [1]. It consists of two main parts:

1. Six Generic Task Types with definitions, specified by a set of Example Tasks for each GTT (total 44 Example Tasks, reported in **Appendix 8**); possible failure modes identified for each GTT (total eighteen failure modes, reported in **Table 16**).
2. Eighteen Decision Trees developed for each of the eighteen failure modes identified for the set of GTTs, with
    a. Branch points characterized by PIFs (aided with negative conditions influencing performance that would help the analyst in deciding which DT path to choose) derived from the GTT (failure mode)-PIFs structures.
    b. HEPs associated to each DT path, which are the conditional failure probability of the GTT-failure mode given the set of branch points affecting it.

The six GTTs and the eighteen failure modes (each corresponding to a DT) are reported in **Table 16**. Shaded boxes indicate that the decision tree includes the corresponding branch point; the value in the box is the median HEP of the DT path when the branch point is affecting the GTT. The ‘black’ color median HEPs represent the directly quantified HEPs from experts and ‘red’ color median HEPs represent the reused applicable values from the directly quantified. One

should note that the matrix structure in **Table 16** shows only single branch point impacts. It does not show HEP values for those DT paths where more than one branch point is affecting the GTT-failure mode (for the complete DT structure of a single GTT-failure mode see **Figure 25** as an example). Refer to **Table 16** for the quantified HEPs for the DTs, and for further details on each DT refer to **Appendix 8.**

Figure 25 presents the DT for "Quality Check" GTT and "Deviation from requirement not recognized" failure mode. **Table 17** complements the DT with the negative conditions supporting the evaluation of each branch point. Similarly, **Figure 26** and **Table 18** presents the DT for "Simple interaction with software or tool" GTT and "Execute desired action incorrectly" failure mode and the negative conditions supporting the evaluation of each branch point. For details of all the DTs refer to **Appendix 8.**

Table 16: GTT-Failure mode and Decision Tree Branch Point: HEPs representing single branch point effects

#	GTT - Failure modes (Decision Tree)		Decision Tree branch points with HEPs								
			IU	ICU-V	LVDTENE	TC	LTE	SRU	TP	DIEW	ED
1	Simple interaction with software or tool	Execute desired action incorrectly(1)	0.0018		0.0011					0.0046	0.001
		Failure to execute desired action(2)								0.0046	0.001
2	Quality Check	Deviation from requirement not recognized(1)	0.0068		0.0091					0.0073	
		Inappropriate understanding of underlying principles (2)				0.0014	0.0642	0.0098	0.0451		
		Check not performed (decision-based)									
		Execute desired action incorrectly (2)	0.0018		0.0011					0.0046	0.001
		Failure to execute desired action(2)								0.0046	0.001
		Coordination failure (2(		0.0176						0.032	0.0083
2	Identification of patient or patient related items	Patient information incorrectly matched(1)	0.0015		0.0077					0.0057	
		Identification check not performed (decision-based)									
		Failure to execute desired action(2)								0.0046	0.001
4	Complex interaction with software or tool	Misinterpretation of data(1)	0.0271			0.0014	0.0642	0.0098	0.0451		
		Execute desired action incorrectly(2)	0.0018		0.0011					0.0046	0.001
		Mismatch or inconsistency not recognized (2)	0.0068		0.0091					0.0073	
5	Iterative determination of optimum parameters	Misinterpretation of information(1)	0.0068			0.0038	0.015	0.0146			
		Inappropriate decision on strategy selection									
6	Verbal communication	Communication failure(1)		0.0176						0.032	0.0083
		Not communicated (decision-based)									

(1)- GTT-failure mode was quantified **(2)**- The DTs associated to this GTT failure modes can be taken from those directly addressing elicited situations

Legend: IU= Information unclear, ICU-V= Information content unclear- Verbal, LVDTENE= Low vigilance due to expecting no error, TC=Tumor complexity, LTE=Lack of experience and training, SRU= Software or resource unavailable, TP= Time pressure, DIEW= Distraction/interruptions and excessive workload, ED= Environmental distractions

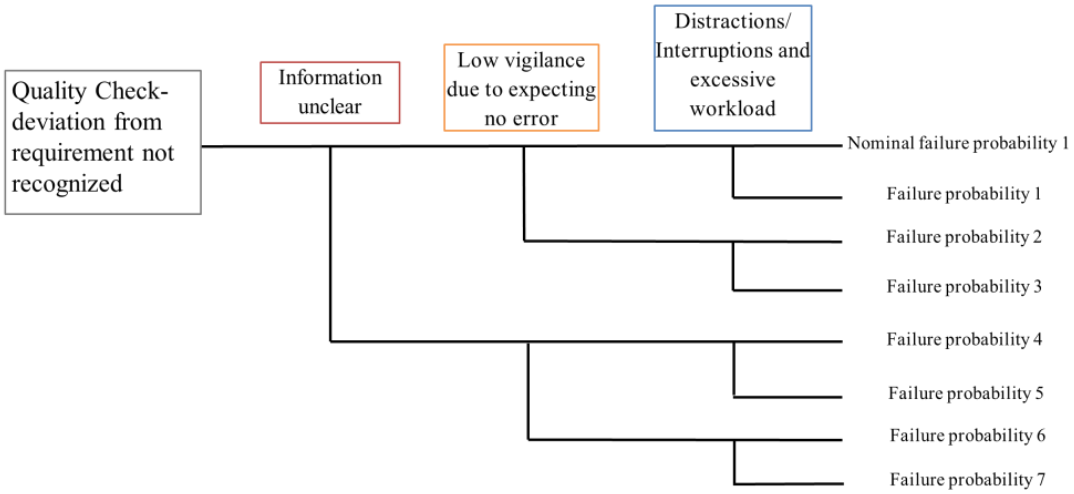


Figure 25: DT for Quality Check GTT- deviation from requirement not recognized failure mode

Table 17: Negative conditions falling in branch point of Quality Check-Deviation from requirement not recognized DT

Branch point	**Affecting negative conditions**
Information unclear	The values on the interface are not easily readable
	The values look alike
	Too much information on the software screen leading to confusion
	The ordering of the values on the control document and on the screen do not match (i.e. X,Y,Z on screen and Z,X,Y in the document)
Low vigilance due to expecting no error	The check was performed recently and you trust it was performed correctly
	The task to produce the output is simple - no error is expected
	Expecting no error as the check is performed by the same person doing the initial task
Distractions/ interruptions and excessive workload	Overloaded with other work
	There is little time to do the task
	The background noise level is too high, it distracts the focus
	Interruptions from colleagues while doing the task

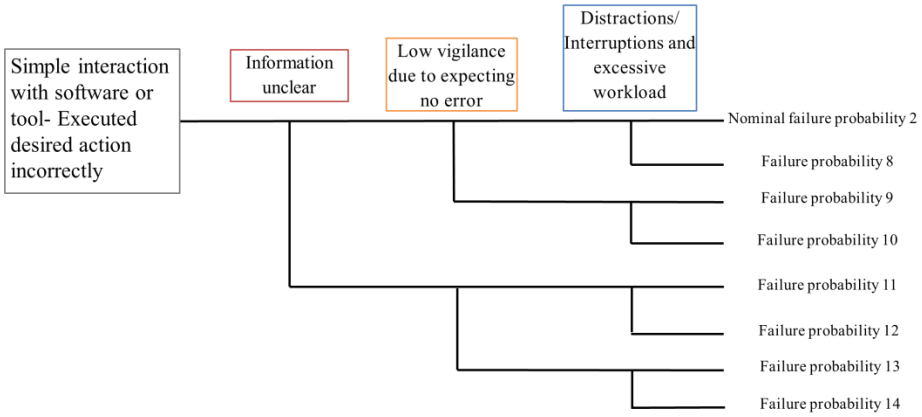


Figure 26: DT for Simple interaction with software or tool GTT- Executed desired action incorrectly failure mode

Table 18: Negative conditions falling in branch points of Simple interaction with software or tool- executed desired action incorrectly DT

Branch point	**Affecting negative conditions**
Information unclear	The indications for the input boxes on the screen are not readable
	The layout of the input interface is unusual or confusing
	Too much information on the screen leading to confusion
	Touch screen or input keyboard is very sensitive (undetected mis-selection is possible)
Low vigilance due to expecting no error	Task is mechanical and repetitive
Distractions/ interruptions and excessive workload	The background noise level is too high, it distracts the focus
	Simultaneously doing another urgent task
	Interruptions from colleagues while doing the task
	There is less time to do the task (excessive workload)
	Simultaneously tracking/supervising another task

The main assumptions for method quantification are:

- The negative conditions falling in the branch point aim to characterize the branch point. It is assumed that if more than one negative condition of the same branch point is affecting the task in a scenario then HEP of the whole branch point is applied modeling all the aspects of the branch point (Refer to **Appendix 8** for HEP for each negative condition in each branch point and the whole branch point for each DT).
- DTs corresponding to same failure modes (e.g. "*execute desired action incorrectly*", **Table 16**) across different GTTs will have the same branch point impacts and thus, the HEP values quantified for one DT of a GTT-failure mode can be used for other DT for another GTT but same failure modes and branch point. (see Failure modes with

superscript 2 in **Table 16** for DTs that are quantified from those directly elicited from experts in **Chapter 3: HRA method- quantitative elements**)

- If two branch points are affecting the GTT-failure mode then the HEP for the joint impact is assumed to follow the hypothesis presented in **Table 19**. As mentioned before, the aim is to find out the order of magnitude of the failure probability rather than the exact failure probability. The principle is as follows: (1) If for one of the performance conditions level is Low and other has any of the other levels i.e. Low/Moderate/High/Extreme then the joint impact HEP is taken to be of the higher level (#1 in **Table 19**). The reasoning is that that Low impact level in combination with any other level does not amplify the joint impact; (2) If both performance conditions have Moderate level then the joint impact HEP jumps one order of magnitude higher to 0.1 i.e. amplification is considered (#2 in **Table 19**); (3) If one performance condition has Moderate level and the other is High then the joint impact HEP jumps to one order of magnitude higher to 1 i.e. amplification is considered (#3 in **Table 19**); (4) and if both performance conditions have High level then the joint impact HEP jumps to 1 i.e. amplification is considered (#4 in **Table 19**).
- If three branch points affect the GTT-failure mode then the HEP for that is assumed to follow the hypothesis presented in **Table 20**. In that it is assumed that Low level still has no amplification effect, except when all three branch points are at low level.
- For more than three branch points affecting the GTT-failure mode then the HEP for that is assumed to be 1.
- Lastly, a failure sequence is a chain of task failures (called failure event) in which typically there is a scenario that transforms in a task failure and then failure of the corresponding check task. When multiple failures are studied then **dependence** between the two tasks needs to be incorporated. Where, **dependence** refers to assessing the influence of the failure of the operators to perform one task on the failure probabilities of subsequent tasks. The current version of the method only covers dependence due to bias in checking tasks; it does not cover all the possible dependencies between them like working on similar interface/software, following similar procedure etc. Thus, in method application if the method calculates HEP for a failure sequence to be less than 10^{-4} then a **cut-off HEP of 10^{-4}** is to be applied. Where, 10^{-4} is an accepted value for very good performance conditions for a single task [2].

Table 19: Hypothesis for joint branch point impact estimation

#	Median level of branch point 1	Median level of branch point 2	HEP of joint impact
1	Low	Low/ Moderate/ High/ Extreme	Higher HEP out of the two
2	Moderate	Moderate	0.1
3	Moderate	High	1
4	High	High	1

Where, the Low, Moderate, High and Extreme are the levels used in elicitation scale for experts. These represent the four orders of magnitude of 10^{-3}, 10^{-2}, 10^{-1} and 1 respectively (See **Chapter 3: HRA method- quantitative elements**).

Table 20: Hypothesis for three branch point impact estimation

#	Median level of branch point 1	Median level of branch point 2	Median level of branch point 3	Median HEP
1	Low	Low	Low	0.01
2	Low	Low	Moderate	0.1
3	Low	Moderate	Moderate	0.1
4	Low	High/Extreme	High/Extreme	1
4	Moderate/High/Extreme	Moderate/High/Extreme	Moderate/High/Extreme	1

Where, the Low, Moderate, High and Extreme are the levels used in elicitation scale for experts. These represent the four orders of magnitude of 10^{-3}, 10^{-2}, 10^{-1} and 1 respectively (See **Chapter 3: HRA method- quantitative elements**).

4.2: HRA method application guidance

This sub section presents the stepwise guidance to apply the method and obtain the HEP for failure events (failure event means how a task fails i.e. task plus its failure mode, for example incorrect input of patient name and surname in the database) in a failure sequence (this is a chain of task failure that finally result in a failure)

Step 1: Failure event to GTT-failure mode matching

Each failure event is first matched to its respective closest Generic Task Type (GTT) by comparing it with the list of Example Tasks (see **Appendix 4** for list of Example Tasks) given as an aid with each GTT. Then the failure mode (FM) is selected for the given GTT based on the failure event. This matching identifies which GTT-FM decision tree is to be used to assign failure probability. It also estimates the nominal HEP for the given failure event as each tree ha1s a nominal HEP.

Step 2: Contextual factors matching with branch point- negative condition

The second step is to identify which conditional HEP is to be selected from the selected decision tree for the task failure i.e. to identify the failure probability distribution conditioned to the given branch point. The factor (which means the error producing conditions) defined in the failure sequence for the given failure event is matched with the branch points and the negative

conditions falling in it. If the factor is precise, then the closest negative condition listed within the given branch point is selected and the conditional failure probability distribution (median, 95th and 5th percentile) of that specific negative condition is taken. Otherwise, for generic context description or when two negative conditions falling the same branch point are found to be impacting together then the aggregated conditional failure probability for all the conditions is used (See **Chapter 3: HRA method- quantitative elements** for aggregation details)

Finally, if two or more factors are affecting the failure event then **Table 19** or **Table 20** is used to compute the HEP. For more than 3 factors the HEP is assumed to be 1.

References:

1. Whaley AM, Xing J, Boring RL, Hendrickson SML, Joe JC, LeBlanc KL and Morrow SL. Cognitive basis for Human Reliability Analysis. Technical report NUREG-2114, U.S. Nuclear Regulatory Commission, Washington DC, January 2016.
2. Presley M, Parry G, Julius J, Gunter K, Grobbelaar J and Hirt M. Use of minimum bounds for joint human error probabilities in PRA. In: 13th International Conference on Probabilistic Safety Assessment and Management; Seoul Korea, 2016.

Chapter 5: Identification of failure sequences as case studies for method application

One of the aims of the PhD work is to analyze possible human failure at PSI's Center for Proton Therapy (CPT), using the developed HRA method. The failure sequences were selected with focus on those expected to have largest likelihood and consequences as well as on personnel tasks of specific interest for CPT. The first section of the chapter presents the steps followed to identify the failure sequences of interest, while the second section reports on the application of these steps. The third section presents the failure sequences themselves, along with the scenarios of interest and the characterized performance context. These elements provide the input for the method application (presented in **Chapter 6: Application of HRA method to failure sequences**).

5.1: Steps for identification and characterization of failure sequences, failure events, scenarios and context

The steps for the identification of the failure events and sequences, of the associated scenarios as well as for the context characterization were as follows.

Step 1: Identification of the radiotherapy process activities of interest

The identification of the failure events as case studies focused on the most critical process activities. Criticality was assessed based on information from the FMEA produced by the task group-100 [1], of general applicability to radiotherapy. Two process activity lists were populated, based on the characterization of the associated failures from [1]: one of the most frequently occurring failures and one of the most severe. The criteria chosen to select the failure events were based on the rating scale used in the FMEA; Occurrence (O) frequency was chosen to be > 6 (where, 6= occasional failures, 0.5% failure probability in the rating scale) and Severity (S) was chosen to be >8 (where, 8= potentially serious toxicity or tumor under-dose) [1]. The resulting lists were reviewed with the CPT personnel to further narrow down based on criteria such as: low potential for discovering failures, possible introduction of systematic issues (across multiple treatment sessions by the same or different patients), a specific interest by the facility

(see **Section 5.2: Failure events, scenarios and contexts: intermediate results before failure sequence formation** for this last point).

Step 2: Identification of failure events of interest

For each of the activities from Step 1, detailed workflows were developed to identify the specific tasks performed as part of the activity. Then with the help of the facility personnel, possible errors on each task were identified on the basis of the failure mode taxonomy developed in **Chapter 2: HRA method- qualitative aspects** (see **Table 4**). As typical in error identification, each task is associated keywords to define a set of failure modes (see Systematic Human Error Reduction and Prediction Approach (SHERPA) [2] or Critical Task Analysis [3]). For each failure mode of each task, the severity of the impact was assessed along with the possible barriers to recognize and recover the failure. Two types of barriers were considered: (1) proceduralized: i.e. a check task allowing to notice the failure is explicitly part of the workflow (or of the applicable procedure). (2) non-proceduralized: a different task in the workflow may allow recovery of the failure, although not explicitly requiring to check the correctness of previous tasks (for example because a value would be used as input for a subsequent task and the incorrectness of the value may be noticed). The output of this step was a list of human failure events with highest severity and lowest recovery potential. This set would be subject to detailed analysis.

Step 3: Task analysis: scenario and context characterization

Next, for each of the failure events identified from step 2, Task analysis [4] was performed with the facility personnel to characterize typical performance contexts in which the tasks are performed, including contexts for the barrier tasks. Also, information was gathered on the scenarios that may perturb the daily working routine (scenario is intended here as the combination of events preceding the performance of the task, e.g. the tumor is aggressive, there is a delay in the delivery of the patient documentation, there is an increase in workload due to specific reasons, and the like); the scenarios reflect deviations from the routine, reflecting the fact that, as discussed in [5], probability of the error increases following perturbance to the routine.

Step 4: Failure sequence formation

Finally, failure sequences were formed combining scenario information, failure events and the characterized contexts. A failure sequence is intended here as the combination of:

- the scenario as the initial perturbance in the workflow,
- the failure event under the given performance contexts,
- the failure of the barrier tasks (if present) under their respective contexts.

5.2: Failure events, scenarios and contexts: intermediate results before failure sequence formation

Step 1: Identification of the radiotherapy process activities of interest

Table 21 lists the process areas and the process activities that were narrowed down from the complete FMEA of [1], using the defined Severity and Occurrence criteria. Then, based on input from PSI CPT's experts, two process activities from the derived list were selected with priority: motion and uncertainty management in treatment planning, e-chart preparation (other process activities can be addressed in future work). Motion and uncertainty management in treatment planning is intended as the management of the moving tumor in the treatment plan and of the uncertainty in its position, referred to as "*4D treatment planning*" throughout the thesis, with the 4th dimension referring to the time. This was chosen because it is a current and relevant area of research at CPT and a workflow is currently under development at the facility; identifying potential critical task failures would be important input for its development. On the other-hand, the e-chart preparation (which generally involves calculating and transferring total dose, fractionations of patient or plan etc.) was chosen as it at the heart of the radiotherapy workflow: a failure during this could result in an error that would remain unnoticed through multiple fractions leading to high severity consequence. These two are highlighted by yellow color in **Table 21**. This thesis only addressed the 4D treatment planning process activity as that was the primary interest for the facility. Future work may be devoted to e-chart preparation activity. The next steps were followed for 4D treatment planning process activity only.

Table 21: Results of step 1: radiotherapy process areas, activities and failure modes with highest occurrence and severity from [1]. Highlight indicates process activities selected for case studies

#	Process areas	Process activity	Failure mode
Most frequently occurring failures			
8	Immobilization and positioning	Radiological properties of positioning aids are known to planners	Properties of device not consistent with accurate dose delivery
12	CT/Simulation	Immobilized patient setup on CT simulator	Immobilization aids incorrectly applied
26	Other pretreatment imaging for CTV	Patient advised for special requirements	Special requirements not respected
27	Other pretreatment	Patient setup for imaging	Poor positioning

	imaging for CTV		
31	Other pretreatment imaging for CTV	Images correctly interpreted	Incorrect interpretation of tumor or normal tissue
41	Initial treatment planning directive from MD	Specify protocol for delineating target and structure	Incomplete/incorrect list of specified structures and corresponding images
42	Initial treatment planning directive from MD	Specify image registration goals	Specify inappropriate protocol, tolerances for registration
44	Initial treatment planning directive from MD	Motion management	Specify wrong motion-compensated treatment protocol
60	RTP anatomy	Delineate tumor volumes and other structures	Poorly drawn contours (spikes, sloppy)
64	RTP anatomy	Tumor volume construction	Use margin width protocols that are inconsistent with dept. procedures
65	RTP anatomy	Tumor volume construction	Margin width protocol for PTV is inconsistent with actual distribution of patient setup errors
85	Treatment planning	Specify regions of interest for optimization process	Inconsistent length of regions of interest
92	Treatment planning	Enter prescription and planning constraints	Incomplete or incorrect set of objectives and constraints
143	Plan approval	Completion of formal prescription after planning	Not signed when appropriate
190	Day 1 treatment	Monitor treatment	Failure to see patient move
205	Day N treatment	Set treatment parameters	Special motion management methods not or incorrectly applied
Failure with Highest Severity			
3	Patient database info	Entry of patient data in electronic database or chart	Incomplete or incorrect treatment history
18	CT/Simulation	Patient position properly represented by image transfer software	Unusual patient position not handled by image transfer software (L and R labels exchanged)
32	Transfer images and DICOM data	Transfer primary CT data	Incorrect CT data set associated with patient
46	Initial treatment planning directive from MD	Specify special instructions, pacemaker, voiding etc.	Wrong or special instructions not given
48	Initial treatment planning directive from MD	Retreatment, previous treatment etc.	Wrong treatment summary of other treatments
133	Plan approval	Plan OK to go to treatment	Wrong patient plan imported
138	Plan approval	Completion of formal prescription after planning	Wrong total dose/ fractionation
144	Plan preparation	Entry of demographic information	Bad info entered, critical patient info not recorded
167	Plan preparation	Prepare e-chart	Incorrect Tx, Rx or wrong patient/plan
171	Plan preparation	Download complete delivery plan to delivery system	Incorrect plan info, connect wrong patient plan in RTP with plan in TX delivery system
176	Day 1 treatment	Gather patient treatment information	Incorrect patient record/chart used
180	Day 1 treatment	Gather patient treatment information	Incorrect treatment data
181	Day 1 treatment	Position patient for treatment	Incorrect treatment isocenter
210	Day N treatment	Treatment delivery	Gantry collides with patient

Step 2: Identification of failure events of interest

The workflow of 4D treatment was detailed with the help of the PSI CPT experts (see **Figure 27**). Error identification with the radiotherapy expert identified the failures with the highest severity impact and lowest recovery potential (the corresponding tasks in the workflow are marked green in **Figure 27**). **Table 22** reports the associated failure events, severity ratings, and the recovery barriers.

The remainder of this section will provide background information on the 4D treatment and on the developed workflow as well as details on the error identification.

4D radiotherapy and workflow

Tumors located in sites like thorax or abdomen move in time mainly due to respiration. Such movement leads to distorted images, incorrect anatomical positions, volumes or shapes for a conventional 3D therapy; and, if not corrected, lead to a potentially wrong treatment delivery. 4D treatment approaches are relatively new in radiotherapy and try to overcome the above-mentioned issues during the formation of the therapy plan to be delivered to a moving tumor. Worldwide research is on-going in 4D treatments to find solutions for optimizing dose distributions (e.g. to reduce margins) for a moving tumor.

In the last few years at Paul Scherrer Institute research has been conducted on the 4D treatment approach to develop motion mitigation strategies. Currently the most advanced method is rescanning, a technique applying an averaged dose distribution over all respiratory phases of the patient. After its initial application to certain motion-sensitive tumors, work is going on to routinely implement rescanning into clinical practice. The workflow is currently under implementation, not finally developed and there is no complete Standard Operating Procedure. At the moment certain tasks can only be performed by one trained personnel. In near future the process will be performed by other personnel. The course of activities of the rescanning concept at CPT can be described as follows:

The treatment planning phase starts after the acquisition of a 4D planning Computed Tomography (CT). During 4D-CT acquisition the patient's breathing cycle is monitored. The registered respiratory pattern is finally manually subdivided into different phases. This binning is used to order the acquired, time-resolved CT projections into 8 to 10 'Sub'-CTs each related to a

defined breathing phase. After CT reconstruction Gross Tumor Volume (GTVs)/ Clinical Tumor Volumes (CTVs) are contoured in each 'Sub'-CT following a semi-automated procedure. Out of these 'Sub'-CTs a so called "*Mid-ventilation phase*" CT is selected. It represents the time-averaged CT over all breathing phases regarding the minimization of distances between the tumor center in each 'Sub'-CT and their common center of gravity (computed by a MATLAB script). Once selected, the set of contours (target volumes and OARs) is fully completed and checked within the "*Mid-ventilation phase*" CT. After extending it to a complete reference CT using the newly developed "*4D treatment planning helper*" software, this reference CT is imported into the TPS PSIPlan (followed by the planning CT import QA) together with the previously completed structure set. In parallel to additional CT conversion steps within the 4D Graphical User Interface (GUI), a conventional, static 3D treatment plan is created in PSIPlan based on the reference CT and its structures. Subsequently, the time-dependent displacement of all structures in all CTs in relation to the reference CT (mid-ventilation phase) is determined (in form of motion vectors) to calculate the dose grid deformation. Based on these computations a time-dependent 4D Dose Calculation is performed including the selection of a starting phase and the determination of the number of rescans. Finally, as a QA measure, all CT-phase dependent plans are slice-by-slice compared to the static plan (on the reference CT) and the corresponding steering files are created.

Figure 27 shows the detailed tasks of the 4D treatment planning process (including 4D-CT acquisition) at CPT. Although 4D treatment planning process is much more complex, however currently at CPT, it is basically an extension of the 3D workflow, in which tasks dealing with the tumor movement, respectively rescanning, are added.

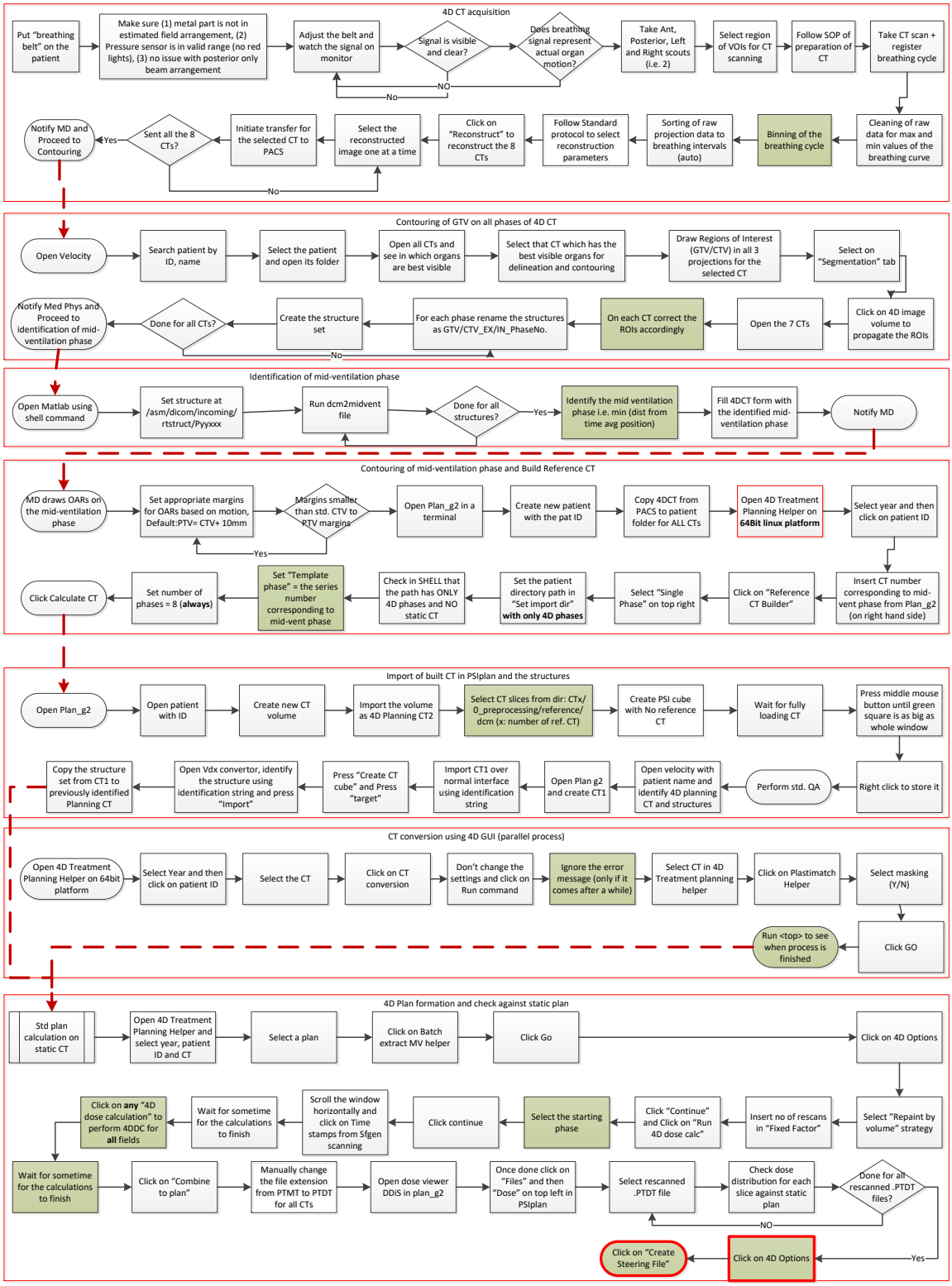


Figure 27: Detailed workflow of 4D treatment planning at PSI

Error identification from the 4D treatment planning workflow

The key tasks (high severity and low recovery potential) identified by the CPT personnel are listed in **Table 22,** with their failure modes (i.e. the failure events) and barriers. During the recovery potential evaluation of the tasks, one of the key observations was that some of the failures may be recovered while performing other activity tasks allowing indirect check of the correctness of previous tasks. These are not explicit checks that are part of the process Quality Assurance (QA), but subsequent tasks that may allow noticing earlier failures. For example, the "*movie maker*" feature of the 4D treatment planning software allows seeing the movement of the contours while breathing; this would help in visually identifying if the contours are drawn correctly or not. The analysis made in this thesis may provide useful information as to whether to include the movie maker as additional QA check in the process workflow.

Table 22: results of step 2: human failure events associated to the 4D treatment planning with highest severity and lowest recovery potential; potential failure consequences and barriers (i.e. tasks allowing error detection)

#	Failure event	Consequences	Barrier tasks (other tasks allowing error detection)
1	Incorrect understanding of the breathing cycle binning, leading to non-homogenous binning of the cycle (i.e. incorrect binning)	Wrong volume	Currently no proceduralized check task. Inappropriate binning of breathing cycle could be detected via: 1. Breathing phases can be seen as a movie. The breathing phases would smear over each other if binning is inappropriate 2. In the contour check task for all the phases
2	Forgetting (slip) to correct the contour of gross tumor volume	Wrong volume, wrong location	Proceduralized check: 1. Contour review group meetings (This task of correction takes the whole day)
3	Incorrect selection of the mid-ventilation phase	Wrong dose distribution	Proceduralized checks: 1. Check of the 4D Computed Tomography (CT) form by the person who performed the selection of mid-ventilation phase task 2. Check and signing of the 4D CT form by different person
4a	Input of "CT no." rather than "Series no." in "Template phase" from the 4D form	Wrong dose distribution	Currently no proceduralized check task. Possible check task: 1. Visual comparison possible of the series number and respiratory phase (CT number)
4b	Input of a different (incorrect) "Series no." from the Shell command in the "Template phase"		
5	Incorrect refence CT file selected	Wrong dose distribution, wrong dose	Currently no proceduralized check task. Possible check task: 1. At the step of "*open velocity with patient name and identify 4D Planning CT and structures*" personnel can see the error from the list of CT files
6	Ignore the error message too early and proceeded with the subsequent task	Wrong dose distribution	Proceduralized check: 1. Check of dose distribution with static plan
7	Incorrect selection of starting phase for a single field plan.	Wrong dose distribution	Currently no proceduralized check. In general, the selection of starting phase is not important as the "rescanning" technique reduces the impact of wrong selection. However, this does not apply for a single field as no rescanning is done.
8	Incorrect creation of 4D steering file i.e. by using the 3D method	Wrong dose, wrong dose distribution	Proceduralized checks: The difference between 4D and non 4D plan is the inclusion of rescans. 1. Check in steering file if "rescans" are included or not 2. Check in info file if "rescans" are included or not

Step 3: Task analysis: scenario and context characterization

For a subset of the identified 4D treatment failure events, the scenarios and contexts for analysis were defined with the CPT personnel as well as on-the-job-observations. The identified scenarios were intended to perturb the routine performance conditions such to address more likely failures. To aid discussion with the center personnel, the scenarios were defined very specific, helping the imagination of real situations (e.g. aggressive nature of the tumor, delay of receipt of required documentation, a child to be treated).

The scenarios and contexts for analysis were defined to reflect high workload situations, one of the common features of error events as reported in [5]. The assumed load is partially characterized as the result of urgency to treat (e.g. because of the aggressive nature of a tumor). The other reason to choose high workload scenarios and contexts was due to the interest of the facility as well: the facility plans to increase the number of patients treated with 4D workflow and thus, was interested in studying the potential fallbacks on workflow.

Typical contexts identified with the help of radiotherapy experts included: increase in workload for new type of treatment, inadequate interface issues, same person performing the primary task and the corresponding check task etc. The basis of identification is given below.

With radiotherapy tasks requiring heavy interaction with software, a natural choice was to focus on the software used during respective failure events. During task analysis, it was observed that certain tasks in 4D treatment plan workflow, which were earlier performed using 3D software are now to be performed on a new window in the same 3D software (the person has to click on a new tab to open that window). The concern raised was that these tasks for a 3D plan are routinely performed on 3D software and for a 4D treatment plan the personnel will have to break the routine and use a different window in the same software to correctly do the tasks. Thus, it seemed relevant to study the factor of mechanically performing the task on the 3D software rather than the 4D dedicated window. Lastly, dependency assessment between checkers was also identified to be studied. Some tasks required the same person to check the output of the task they recently performed themselves. This was studied in contrast with a different person checking the same output.

Step 4: Failure sequence formation

Based on all the information obtained in steps 1, 2 and 3, four failure sequences were formed with some variants of each failure sequence making ten failure sequences in total. These are presented in next section.

5.3: Failure sequences

Failure sequences 1 and 2 (and variants):

Currently, the time from patient admission to treatment delivery is 2-3 weeks, giving ample time to perform the tasks required to prepare the 4D treatment plan. However, in cases when a tumor grows aggressively, there may be the need to treat with urgency, within about one week after patient admission. As an additional source of workload, it is assumed that the relevant patient information from an external hospital reaches Center for Proton Therapy (CPT) late during the week preceding the treatment, thus leaving little margin for treatment preparation. It is therefore assumed that the 4D Computed Tomography (CT) acquisition process starts late during the week preceding the treatment. The other pre-planning tasks followed for the 4D treatment are done on Friday morning to afternoon and the corresponding checks are done late afternoon.

The failure sequence entails the failure event of a task done during the 4D preparatory work (i.e., building the reference CT), using the 4D Graphic User Interface (GUI). During this, the personnel is to enter the "*Series No*." of the corresponding "*Mid-ventilation phase*" derived previously from a MATLAB script into the "*Template Phase*" input field. This MATLAB script computes the distances between the tumor center in each ventilation phase CT and their common center of gravity. The "*Mid-ventilation phase*" is characterized by the shortest distance, i.e. it is the most representative CT regarding a full respiratory cycle.

The failure event under analysis is that the personnel enters an incorrect "*Series No.*" in the "*Template phase*" field of the 4D GUI. It is assumed that this can happen in two ways: first, the personnel enters the "*CT No.*" from the form instead of the "*Series No.*" while inserting the value from the 4D form (the values are close in the GUI and one value may be taken for the other); the other way is to enter the wrong "*Series No.*" from the MATLAB prompt while looking at it. These will be treated as two different failure events leading to two different failure sequences i.e. failure sequence 1 and failure sequence 2.

In the current workflow, there are no proceduralized checks to prevent such failure to propagate; however, a possibility to detect the error exists by comparing the series number and the selected

respiratory phase. Thus, two more variants of each failure sequence 1 and 2 will also be studied which will characterize the hypothetical situations if the check will be proceduralized in future. Variant A will represent when the check is done by different person than the one inputting on the 4D GUI (termed as failure sequence 1A and 2A); Variant B will represent the situation when the check task is performed by the same person who did the transfer task (termed as failure sequence 1B and 2B). Failure event in those variants will be the failure of the proceduralized check task. **Figure 28** shows this concept with failure sequence 1.

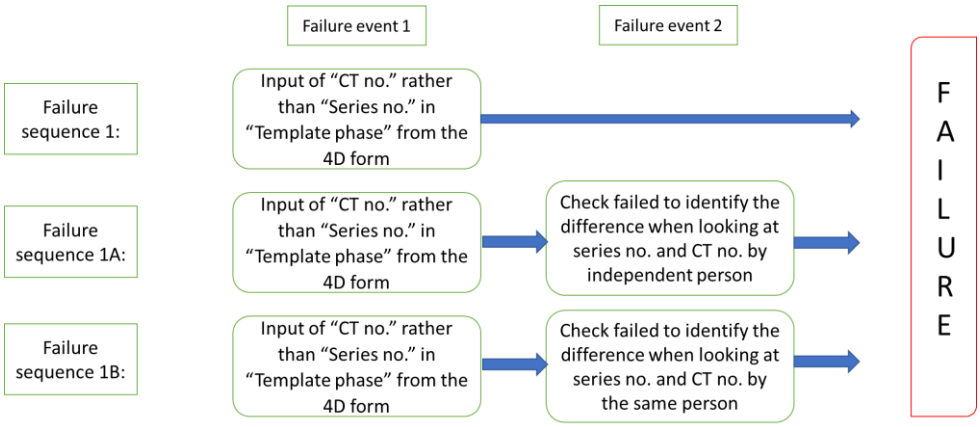


Figure 28: Failure sequence progression concept

Concerning the interface, the task in failure event 1 for failure sequences 1, 1A and 1 B, is performed using the 4D form. Whereas, the task in failure event 1 for failure sequences 2, 2A and 2B, is performed using the MATLAB command tab or workspace tab.

Lastly, the consequence of the failure sequences would be a delivery of a wrong dose distribution to the patient. According to reference [1] the rating means: "*the failure would lead to increase in adverse clinical outcomes like reduced tumor controls or increased likelihood of moderate grade toxicity; and the variation in the dose is expected to be between 5-10%*". In the present case it is expected that the error will not be corrected for any of the fractions.

The two base failure sequences 1 and 2 (See **Figure 29** and **Figure 32**) and variants (1A, 1B, 2A, 2B) of each failure sequences are presented below (**Figure 30, Figure 31, Figure 33** and **Figure 34**). The bubbles with the "lightening" type symbol indicate the contextual factors affecting the failure event. This form of representation is replicated in other failure sequences.

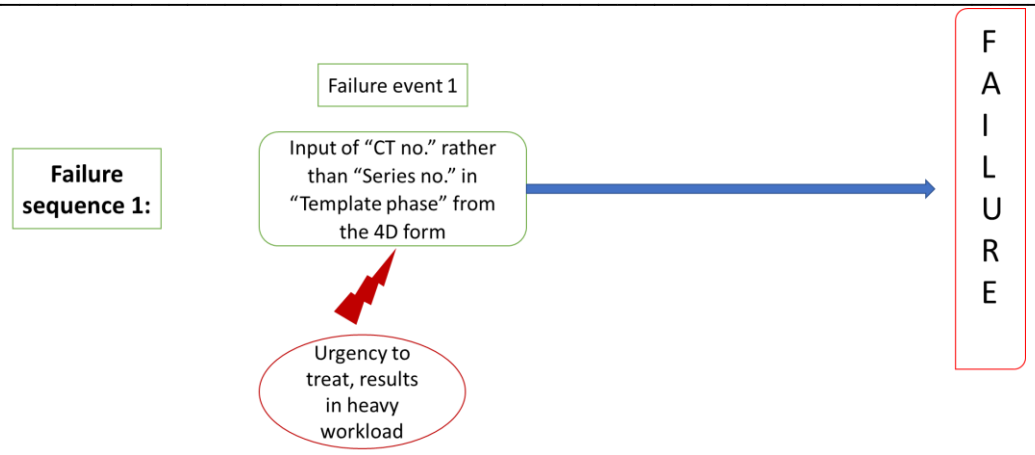


Figure 29: Failure sequence 1

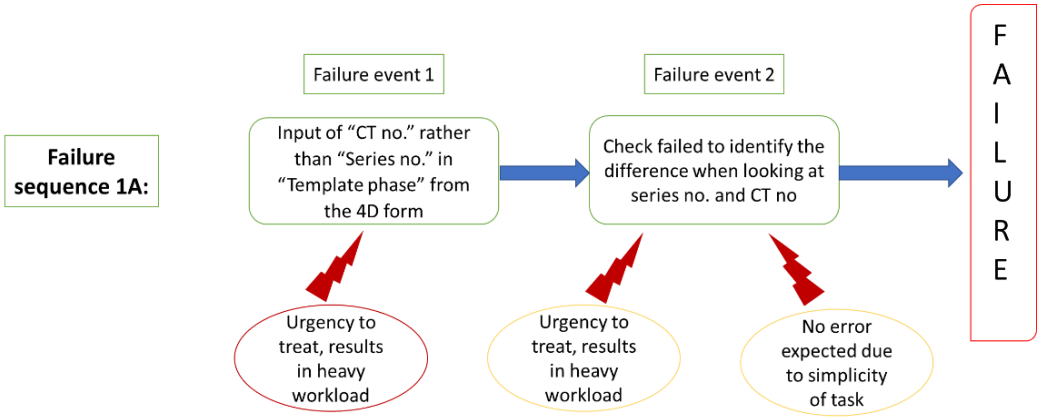


Figure 30: Failure sequence 1A variant

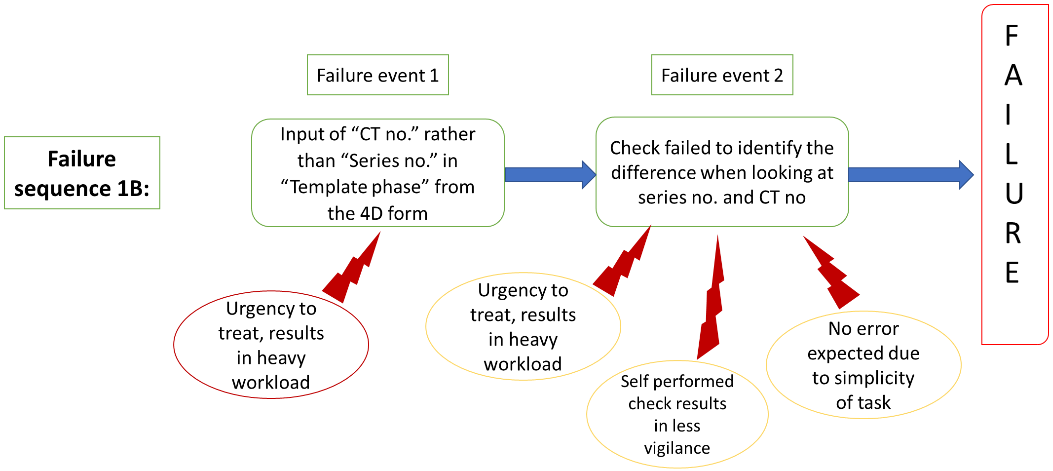


Figure 31: Failure sequence 1B variant

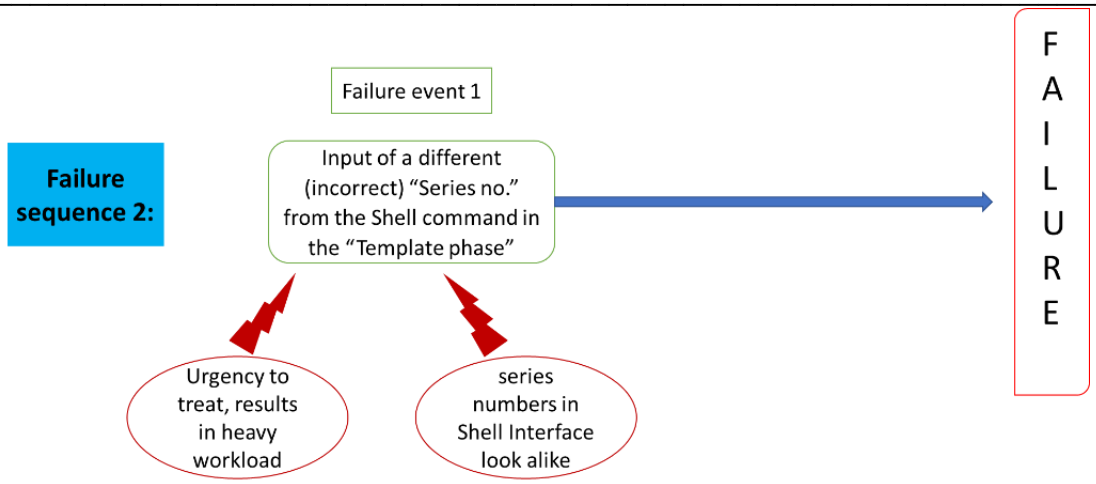


Figure 32: Failure sequence 2

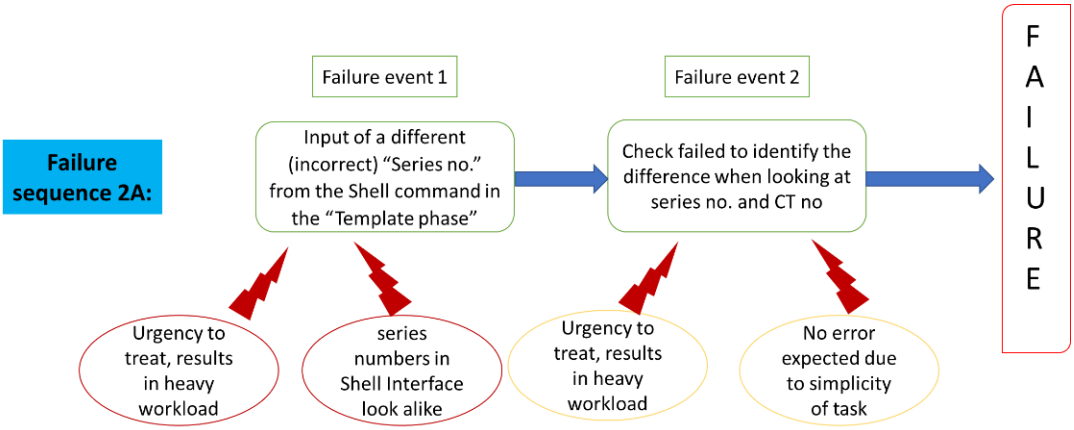


Figure 33: Failure sequence 2A variant

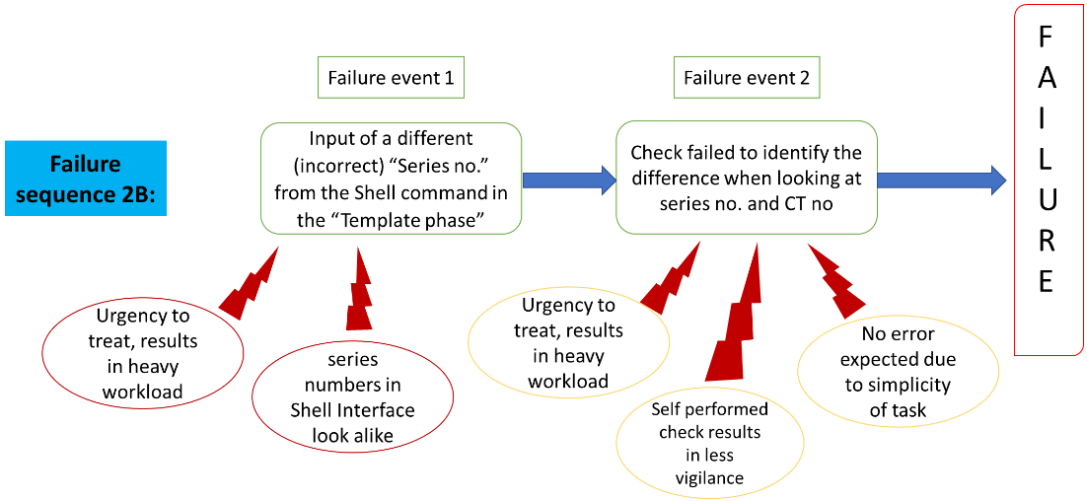


Figure 34: Failure sequence 2B variant

Failure sequence 3:

Let us assume a scenario with more than the normal number of patients to be treated on the same day the following week. Their plans are being prepared in parallel by the same dosimetrist. One patient requires a 4D treatment while the rest require 3D. In addition, the plans are in delay, resulting in high workload. Before introduction of the 4D treatment at CPT, the steering files were generated using the available software for 3D treatment planning (PSIPlan). With the introduction of 4D treatment, a new software component specific for 4D treatment planning has been added to PSIplan. The planner must switch between similar windows in PSIPlan in order to create the 4D steering file.

In the workflow (see **Figure 27**) the tasks before creating the steering files for the 4D plan are done in the 3D software (where the create steering file button is visible) and then the planner must switch to the 4D tab in the same software window to create the 4D steering file.

The two other failure events are the two proceduralized checks of the created steering file. The first check is performed in a sub window (for the 4D option) of the 3D main interface. This 3D main interface also includes the button for 3D steering file generation. The first check is performed by the planner who created the steering file. The sub window on which the check is performed is a basic "*table*" in an "*excel*" like format and the person should check "*repainting*" or "*no repainting*" in '*rescan*' column to distinguish if it is a 3D or 4D plan. The second check is performed on the 'Info' file, a parameter list produced for each steering file. This task is performed by a different person, mostly a medical physicist.

The consequences of this error would be the delivery of a wrong absolute dose or a wrong dose distribution to the patient. According to reference [1] the rating means "*variation in the dose is expected to be between 5-10%*".

Figure 35 shows the derived failure sequence below.

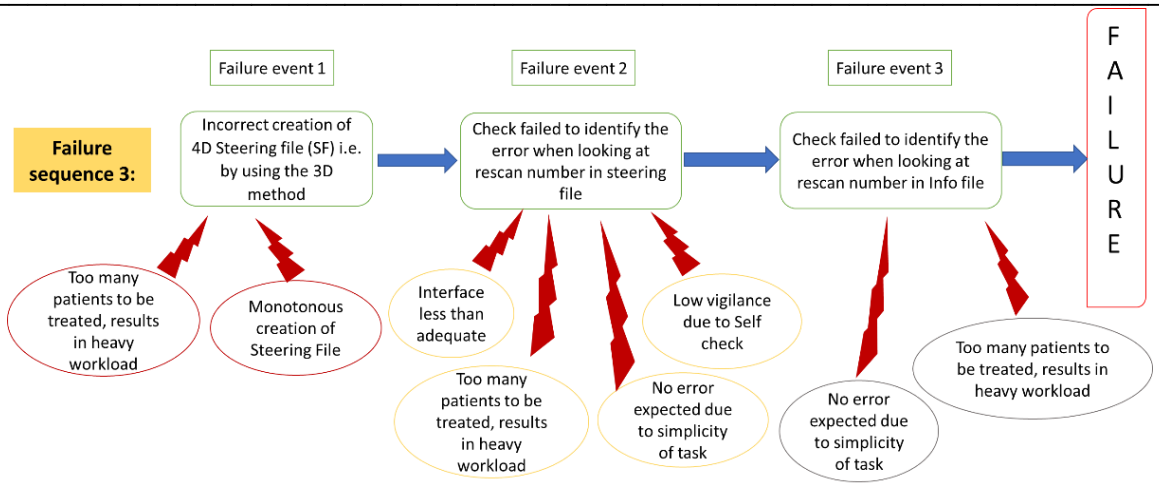


Figure 35: Failure sequence 3

Failure sequence 4 (and variants):

Let us assume there is a child or a very old patient with unusual breathing pattern (i.e. either very fast or slow) and the planner countering the problem has lack of experience in tackling such a problem. Typically there is the possibility to consult another planner with high experience in 4D tasks. The scenario assumes that this possibility cannot be pursued, due to unavailability of the specific planner.

The task in this scenario is to bin (i.e. to divide) the breathing cycle of the patient such that each part represents the percentage of inhalation or exhalation of the breathing cycle. The failure that is being modelled is that the planner does not completely understands the binning methodology and how to bin unusual breathing pattern thus, ends up in incorrectly binning the breathing cycle. The consequence of such a failure would be wrong volumes being treated.

There are no proceduralized tasks to check whether the breathing cycle was binned correctly. However, two possibilities serving as check were identified. The first, to use the function in the software where the output can be seen as a movie. The second, when different person performs the contouring check later in the workflow; there the person can be directed to check smearing of the contours. Thus, as done before, two more failure sequence variants were studied. Variant A when only the first check is proceduralized; Variant B when both checks are proceduralized.

Figure 36 shows the derived base failure sequence and **Figure 37** and **Figure 38** the 2 variant failure sequences.

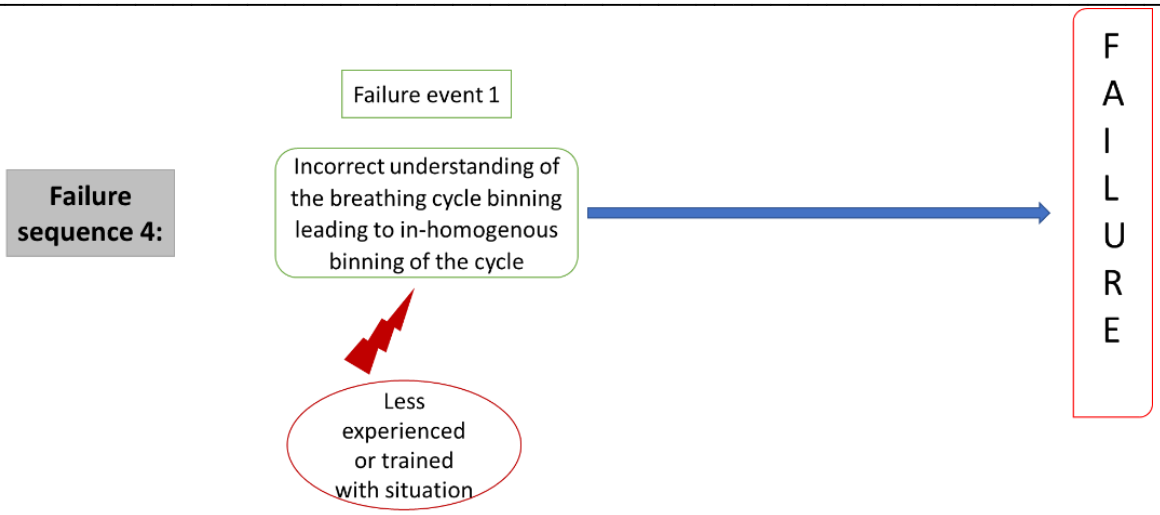


Figure 36: Failure sequence 4

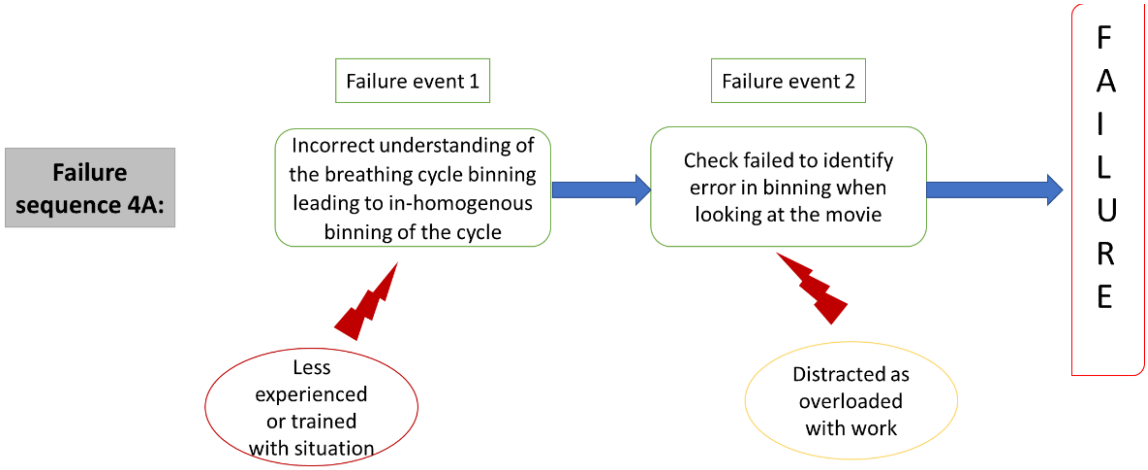


Figure 37: Failure sequence 4A

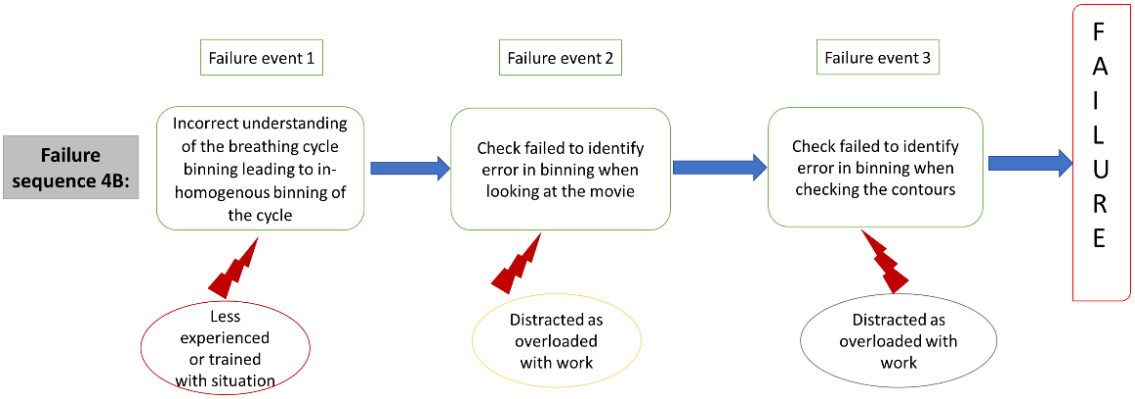


Figure 38: Failure sequence 4B

References:

1. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR, Williamson JF and Yorke ED. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management, Medical physics 2016; 43 (7): 4209-4262.
2. Embrey D. SHERPA: A systematic human error reduction and prediction approach. Contemporary Ergonomics 1984-2008: Selected Papers and an Overview of the Ergonomics Society Annual Conference 2009. pp: 113-119.
3. Energy Institute. Guidance on Human Factors Safety Critical Task Analysis (1st Ed.). London: Energy Institute, 2011.
4. Kirwan B and Ainsworth LK. A guide to task analysis. Boca Raton: CRC press, 1992.
5. Wreathall J, Brown WS, Militello L, Cooper SE, Lopez C and Franklin C. A risk-informed approach to understanding human error in radiation therapy. Technical report NUREG-2170, U.S. Nuclear Regulatory Commission, Washington DC, June 2017.

Chapter 6: Application of HRA method to failure sequences

This chapter presents the method application to the four failure sequences with their variants identified in **Chapter 5: Identification of failure sequences as case studies for method** application. It should be noted that, certain key tasks in 4D treatment planning workflow were observed to not have proceduralized barrier check tasks. Thus, variants were introduced to study the impact of proceduralizing check tasks for those key tasks. An overview of the analysis is given in **Table 23** and **Table 24**. **Table 23** presents the assessed Human Error Probabilities (HEPs) of the failure events when the identified performance factors (like interface, workload etc.) are affecting it (one at a time). **Table 24** presents the HEP of the total failure sequences. The effect of multiple factor on the HEP is determined based on **Table 19** and **Table 20**.

6.1: Failure sequences 1, 1A and 1B

Failure sequence 1 includes only one failure event (See **Figure 29, Figure 30, Figure 31**): Input of "*CT No*." rather than "*Series No*." in "Template phase" from the 4D form. The two variants 1A and 1B include the additional failure of the proceduralized check: failure to identify the difference when comparing the entered "*Series No*." with the required series number (1A when the check is performed by a different person from whom inputs the 4D form, 1B when the check is performed by the same person).

<u>Step 1</u>: Failure event to GTT-failure mode matching

The first failure event of failure sequence 1 (see **Figure 29** and **Table 23**) matches to the "Simple interaction with software or tool" GTT description, as the task involves inputting information from paper into computer. The failure mode in this case is to input an incorrect value into the "*Template phase*" field i.e. inputting CT no. rather than Series no., which matches to the "*execute desired action incorrectly*" failure mode in the given GTT. Thus, the decision tree selected for this task failure is "*Simple interaction with software or tool*"- "*execute desired action incorrectly*".

The second failure event (in failure sequences 1A and 1B, see **Figure 30, Figure 31**, **Table 23**) is about not recognizing an existing error made in previous task when comparing it with a control value, which matches the description of the "*Quality Check*" GTT. The failure mode in this case is not to recognize the error, thus matching with the

"*Deviation from requirement not recognized*" failure mode belonging to this GTT. Thus, the decision tree selected for this task failure is "Quality Check"- "deviation from requirement not recognized".

<u>Step 2</u>: Contextual factors matching with branch point- negative condition

For the first failure event (i.e. failure sequence 1, see #1 in **Table 23**), the performance context is characterized by urgency to treat, leading to heavy workload (ref. to **Figure 29**). This situation is represented by the negative condition "*there is less time to do the task (excessive workload)*" in the branch point "Distractions/ interruptions and excessive workload" in the above identified GTT-Failure mode "*Simple interaction with software or tool*"- "*execute desired action incorrectly*" (See **Appendix 8** for the negative conditions in each GTT-FM decision tree).

For the failure event 2 (in failure sequences 1A and 1B, ref. to **Figure 30, Figure 31**, see **Table 23**), in addition to the urgency to act, the context is further characterized by low vigilance of the person performing the task, due to the very simple nature of the task. This additional factor stems from the fact that most of the execution tasks are simple so low vigilance due to bias in checking simple tasks is assumed. This situation is represented by negative condition "*Task to produce output is simple- no error expected*" in the branch point "*Low vigilance due to expecting no error*". In addition, for variant B the impact of bias due to check performed by same person who did the primary was further incorporated as a contextual factor. This translates in the additional negative condition for 1B "*Expecting no error as the check is performed by the same person doing the initial task*" under the branch point "*Low vigilance due to expecting no error*".

Based on the above steps, **Table 23** shows the matched negative conditions to the defined contexts and presents the median HEP for the given failure events and the identified contexts. On the other hand, **Table 24** gives the median HEPs of the complete failure sequences.

6.2: Failure sequences 2, 2A, 2B, 3, 4, 4A and 4B

Similarly, failure events in failure sequences 2 (variant A and B), 3 and 4 (variants A and B) (See **Figure 32** to **Figure 38**) were matched with their respective GTT-Failure modes and the contexts to the respective branch point-negative conditions. **Table 23** shows the matchings, in all the cases the contexts could be captured by the branch point-negative conditions in the corresponding GTT-FM decision tree (See **Appendix 8** for negative conditions in each branch

point of the GTT-FM DT). See **Table 24** for Median HEP for the identified failure sequence. For detailed discussion on results see **Section 6.3: Discussion**.

A key point to note is that the scenario 4 is only case out of the four cases where the initial task failure is matched to a sense-making failure mode. Whereas, in the rest of the cases the initial task failures are mapped to execution failure. Thus, in this case the bias due to task being simple that leads to less vigilance is not considered.

Table 23: Matching scenarios to applicable GTT-Failure mode and negative conditions from the method

Failure Sequence	Failure event	Decision tree used for quantification	Context (negative) (1)	Branch point with negative impact	Applicable negative condition	Judgments from elicitation	Median HEP
1 (Base case)	Input of "CT no." rather than "Series no." in "Template phase" from the 4D form	Simple interaction with software or tool-execution failure	Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, L, M, M, M	0.0088
1A, failure of proceduralized check, different person	Failure to identify the difference when checking the Series no.	Quality Check - deviation from requirement not recognized	Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
1B, failure of proceduralized check, same person	Failure to identify the difference when checking the Series no.	Quality Check - deviation from requirement not recognized	Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple**Error! Bookmark not defined.**	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
			Self performed check results in less vigilance**Error! Bookmark not defined.**	Low vigilance due to expecting no error	Expecting no error as the check is performed by the same person doing the initial task	M, M, H, M, M, M, H	0.01

Failure Sequence	Failure event	Decision tree used for quantification	Context (negative) [1]	Branch point with negative impact	Applicable negative condition	Judgments from elicitation	Median HEP
2, (base case)	Input of a different (incorrect) "Series no." from the Shell command in the "Template phase"	Simple interaction with software or tool-execution failure	Series numbers in Shell Interface are not readable	Information unclear	The indications for the input boxes on the screen are not readable	L, L, L, M, H	0.0011
			Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, L, M, M, M	0.0088
2A, failure of proceduralized check, different person	Failure to identify the difference when checking the Series no.	Quality Check - deviation from requirement not recognized	Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
2B, failure of proceduralized check, same person	Failure to identify the difference when checking the Series no.	Quality Check - deviation from requirement not recognized	Urgency to treat, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple **Error! Bookmark not defined.**	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
			Self performed check results in less vigilance **Error! Bookmark not defined.**	Low vigilance due to expecting no error	Expecting no error as the check is performed by the same person doing the initial task	M, M, H, M, M, M, H	0.01

Failure Sequence	Failure event	Decision tree used for quantification	Context (negative) (1)	Branch point with negative impact	Applicable negative condition	Judgments from elicitation	Median HEP
3	Incorrect creation of 4D Steering file (SF) i.e. by using the 3D method	Simple interaction with software or tool-execution failure	Monotonous creation of steering files with one exception of 4D plan	Low vigilance due to expecting no error	Task is mechanical and repetitive	L, M, L, L, M	0.0011
			Too many patients to be treated, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, L, M, M, M	0.0088
	Failure to identify the difference when checking the rescan in SF screen	Quality Check - deviation from requirement not recognized	The interface is not clear to distinguish 3D and 4D plans	Information unclear	The values on the interface not easily readable	L, L, M, M, M, M	0.0098
			Too many patients to be treated, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple**Error! Bookmark not defined.**	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
			Self performed check results in less vigilance**Error! Bookmark not defined.**	Low vigilance due to expecting no error	Expecting no error as the check is performed by the same person doing the initial task	M, M, H, M, M, M, H	0.01

Failure Sequence	Failure event	Decision tree used for quantification	Context (negative) [(1)]	Branch point with negative impact	Applicable negative condition	Judgments from elicitation	Median HEP
	Failure to identify the difference when checking the rescan in Info screen	Quality Check - deviation from requirement not recognized	Too many patients to be treated, results in heavy workload	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
			Bias due to task being simple	Low vigilance due to expecting no error	The task to produce the output is simple - no error is expected	M, L, M, M, H, L, M	0.0099
4, (base case)	Incorrect understanding of the breathing cycle binning leading to in-homogenous binning	Iterative determination of optimum parameters- Mis-interpretation of information	Less experienced or trained with situation	Lack of training or experience	Training program does not cover information on specific constraints for special tumors	M, M, M, M, H, H, H, H	0.0303
4A, proceduralized check 1	Failure to identify the smearing when checking the breathing phases as a movie	Quality Check - deviation from requirement not recognized	Distracted as overloaded with work	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01
4B, proceduralized check 2	Failure to identify the smearing during contour check of all the phases	Quality Check - deviation from requirement not recognized	Distracted as overloaded with work	Distractions/ interruptions and excessive workload	There is less time to do the task (excessive workload)	L, M, M, M, M, H, H	0.01

(1)- This column summarizes the most important contextual factors negatively influencing performance. For details on the overall performance context see Appendix 2

(2)- These values are taken for the whole branch point as more than one negative condition are impacting within the branch point (as per the assumption)

Table 24: Median HEPs for the identified failure sequences

Failure sequence	Failure event median HEPs (no of factors applying)			Median HEP of the failure sequence
	Task 1	Task 2	Task 3	
1	0.0088 (1)	-	-	0.0088
1A	0.0088 (1)	0.1 (2)	-	0.0001*
1B	0.0088 (1)	1 (3)	-	0.0088
2	0.01 (2)	-	-	0.01
2A	0.01 (2)	0.1 (2)	-	0.001
2B	0.01 (2)	1 (3)	-	0.01
3	0.01 (2)	1 (4)	0.1 (2)	0.001
4	0.0303 (1)	-	-	0.0303
4A	0.0303 (1)	0.01 (1)	-	0.0001*
4B	0.0303 (1)	0.01 (1)	0.01 (1)	0.0001*

(1) - it should be noted for these failure sequences the calculated probabilities are lesser than the cut-off of 10^{-4} and therefore cut-off value has been used.

6.3: Discussion

6.3.1: Failure sequences 1 and 2 (1A, 1B and 2A and 2B variants)

Variants A and B were considered to analyze the effect of introducing a check to ensure that correct values are entered in the 4D form (variant A: check by a different person from whom is inputting the 4D form; variant B: check by same person). When comparing the failure sequences 1 and 2 with their variants A and B in **Table 24**, it can be seen that the joint HEP of the failure sequence reduces by one order of magnitude for variant A (i.e. from 0.0088 for the base case 1 to the cut-off value of 0.0001 for 1A; and from 0.001 to 0.0001 for 2 and 2A, see **Table 24**). However, in case of variant B, the addition of the proceduralized check task does not reduce the joint HEP. In this case, the effectiveness of the check is much reduced because it is performed by the same person who did the task of entering the value. This is attributed to the fact that: when the same person performs the check on a very simple task that he/she performed right before, he/she would do this very quickly with paying little attention (task simple and he/she did it recently and thus everything must be correct). This is a key input for the workflow improvement when introducing a proceduralized check. Proceduralized checks on simple execution tasks should be performed by a different person because they have higher effectiveness than compared to the check task being performed by the same person who did the execution task.

When comparing the interface (MATLAB or 4D form) used to input the "*Series No*." into the 4D GUI, the analysis showed no relevant difference between the two (HEP for failure sequence 1 is 0.0088 and for failure sequence 2 is 0.001, see **Table 24**). The little sensitivity of the decision tree towards the interface is again because the task and the interface are very simple, with little possibility of confusion. Indeed, a slightly lower failure probability may be achieved if the task is performed by using the 4D "*template form*", as opposed to the MATLAB window.

Workload analysis

Another key element of the analysis was to study the impact of workload on the important tasks in a new process. It was studied because of its importance from human factor studies in radiotherapy [5]. Based on the analysis done for the failure events across all failure sequences, the analysis confirms that workload may have important influence on the performance (HEP increase by one order of magnitude). High workload hinders the effectiveness of the barriers. As explained before, the failure of the check task on a simple execution task, if performed by the same person under heavy workload is almost certain (HEP increases by 3 orders of magnitude).

As the number of patients treated with 4D technique will increase, it becomes important that the process steps are proceduralized along with the important checks as well as which parts of the process shall be subject to checks by different person. It would be advisable to distribute the workload so as to reduce its effect on task performance. In addition, the option of formalizing proceduralized checks by different person, changing porceduralized checks done by the same person to checks by different persons should be considered. These would result in reduction in the likelihood of failure sequence as well as improving the effectiveness of the barrier tasks.

6.3.2: Failure sequence 3

Currently, as mentioned in **Chapter 5: Identification of failure sequences as case studies for method application**, the 4D tasks are overlapping across the 3D task flow. Where the person must switch from the daily 3D task flow and use specific 4D software or a new window in the 3D software to do the 4D task. One of them is steering file generation which is the focus of this failure sequence. The failure studied is that the person uses the 3D software to generate steering file for 4D treatment plan, instead of using the 4D window.

Creation of steering file is a simple and fast 'click a button' task, and with high number of plans in line the task becomes repetitive and mechanical. In addition, the button to click to create steering file (for 3D plans) is on the same window where the tasks prior (that are common to 3D and 4D workflow) to steering file creation are performed. However, for 4D steering file creation a new window must be opened and button on that window should be clicked. Thus, these circumstances of mechanically performing a simple task with button in view of the person and under time pressure together result in the HEP of creating steering file with 3D software window for a 4D treatment plan as 0.01 (**Table 24**). This analysis points to the fact that this overlap of tasks between 3D and 4D workflow increases the failure of a key task (increases HEP to one order of magnitude higher). It would be worthwhile considering separating the two workflows. Possible solutions could be:

(1) having a dedicated software for all the tasks in the 4D workflow i.e. to linearize the workflow with only one dedicated software,
(2) to have a dedicated workstation where the all the tasks of the 4D patients are performed.

In both the solutions, separating the workflows would reduce this accidental creation of steering file with 3D software window. The selection of strategy would be done after discussions with CPT experts, pondering pro and cons of each option.

Generation of steering file is an important task as after that the tasks are mostly automatic and delivery is done automatically using the contents of the file. Thus, two checks are proceduralized to see if the steering file is generated using the 4D window for the plan. Both are the same check but in two different software windows: steering file window and info file window. The first check involves looking at a list of parameters and confirming if the "no of re-scans" field mentions "repainting" or "no repainting". In this check, the interface is very basic, the parameters in the interface are listed sequentially and the person has to go through and then see if the "no of rescans" row has "repainting" or "no repainting"; which renders the interface to be less than adequate. The HEP for inadequate interface is 0.0098. Then, the HEP of the task under heavy workload is 0.01. And finally, the last two conditions are low vigilance due to bias of task of producing the steering file is simple and low vigilance due to the check task is performed by the same person who created the steering file. Overall, the HEP of such a check under the identified contexts is 1. The second check is done by a different person under high workload, with the HEP as 0.1. Overall, the HEP for the two proceduralized check tasks to fail is 0.1 and the overall HEP of the failure sequence 3 is 0.001 (**Table 24**).

The analysis points to the contexts that should be improved in order to increase the effectiveness of the barrier tasks; and a few proposals are listed:

(1) both the checks should be done by a different person (as explained before in workload analysis, different person increases effectiveness of checks)
(2) the interface in the steering file could be improved, by highlighting the 4D parameters with colors, this way it would easier to spot the 4D parameters.

By doing these the HEP would reduce to 0.0001, i.e. the cut-off value (**Table 24**). Actual strategy would be selected after discussions with the CPT experts.

6.3.3: Failure sequences 4, 4A and 4B

Binning of the breathing cycle is a new process step, specific to 4D process and requires cognitive effort in determining on how to bin the cycle correctly. For the chosen case study, the context applicable in the failure event is characterized by less than adequate training or experience in doing the task, this has a HEP of 0.03 (**Table 24**). And as explained in **Chapter 5: Identification of failure sequences as case studies for method application**, this task does not have a procedure-directed check. The analysis showed that two tasks could be proceduralized as checking tasks. The first one is already part of the workflow; the other one is the feature of the software to see the result as a movie. When only one check is proceduralized, the HEP already

reduces to 0.0001 (cut-off value of 0.0001 used, **Table 24**) (4A). And if both are proceduralized in the process and if performed under a high workload, then the HEP for that scenario (4B) will be 0.0001 (cut-off value of 0.0001 used, **Table 24**). This would significantly reduce the HEP of the failure sequence and thus it would be suggested to consider to introduce them in the process. Adding the checks would increase the workload, that effect has already been taken into account when including the less time to do the task.

Chapter 7: Conclusions and future work

This chapter presents the main aims, requirements and the objectives achieved with this PhD research work for the fields of human reliability analysis and radiotherapy. Each of the objectives formulated in **Chapter 1: Introduction** is addressed and the corresponding answers/findings and main contributions are presented. The chapter then ends with research ideas for the future and the list of publications produced by the PhD work (articles in international scientific journals and in conference proceedings).

7.1: Research objectives

The main aims, its three requirements and the further broken objectives presented in **Chapter 1: Introduction** are re-stated here to review and assist in understanding the main thesis findings. The main aims of the thesis were to develop a radiotherapy-specific Human Reliability Analysis method to address human failures in radiotherapy and then to apply the developed method to the CPT at PSI for an assessment of risks in specific scenarios and, possibly, for suggesting safety-enhancing measures. The four main requirements were: (1) the developed method should address a set of tasks representative of the radiotherapy treatment process and the relevant failure modes; (2) it should identify the factors that influence the reliability of the performance of these tasks in different situations and guide their evaluation; (3) it should support the estimation of task failure probabilities, accounting for the influence of the identified performance influencing conditions; and (4) its development should be based on systematic and traceable approaches: i.e. the set of representative tasks, associated influencing factors and the use of expert judgement for elicitation of the method's data.

The aims were broken into five smaller objectives, which are presented below:

1. Identification of factors that influence the performance of the personnel working in radiotherapy. (**Chapter 2: HRA method- qualitative aspects**)
2. A systematic and traceable procedure for developing the taxonomy of the representative set of tasks carried out by the facility personnel as part of the patient handling process in a radiotherapy facility for the radiotherapy HRA method. (**Chapter 2: HRA method- qualitative aspects**)
3. Development of a quantification framework to systematically incorporate the quantitative impact of PIFs on the task failure probabilities (**Chapter 3: HRA method- quantitative elements**)

4. Quantification of human error probabilities (HEPs) of tasks given the negative performance conditions in traceable and systematic way using expert judgment (**Chapter 3: HRA method- quantitative elements**)
5. HRA method description with assumptions, limitations and application guidance (**Chapter 4: HRA method for radiotherapy: overview**)
6. Application of the developed method to potential failure sequences (**Chapter 5: Identification of failure sequences as case studies for method application** and **Chapter 6: Application of HRA method to failure sequences**)

7.2: Conclusions

This thesis successfully developed the first HRA method specific for radiotherapy applications, with Generic Task Types (GTTs) and Performance Influencing Factors (PIFs) as the building blocks of the method. The method development included (a) identification and characterization of taxonomies of representative tasks (termed GTTs) and factors influencing performance (termed PIFs) for radiotherapy, (b) systematic and transparent identification of sets of PIFs for GTTs via failure mode, failure cause and failure mechanisms, (c) formation of decision trees (DTs) as the quantification framework, (d) and quantification of HEPs for DTs via aggregation of elicited expert judgments. A step-by-step process was followed during the development of the method and estimation of HEPs from experts; the results of each step were also documented. In this way, **satisfying the requirement of systematic and traceable method formation**.

The thesis validated the main results of the thesis using existing literature: GTT-PIF structures and estimated HEPs from experts.

Once developed and its results validated, **the method was applied to CPT's 4D treatment planning workflow to systematically assess and quantify the failure probabilities of ten developed failure sequences in this thesis**. The analysis transferred into **safety-enhancing proposals related to the implementation of checks and to the improvement of their effectiveness**.

Lastly, **the developed method is generic for the domain** (i.e. not CPT specific) and opens its application to other radiotherapy centers.

The rest of section details the aforementioned major achievements and aligns them to the declared research objectives.

7.2.1: Qualitative building blocks: GTTs and PIFs

The achievements of this section cover the objectives #1 and #2 presented in **Section 7.1: Research objectives**.

Performance Influencing Factors (PIFs)

Chapter 2: HRA method- qualitative aspects of the thesis dealt with the identification and characterization of factors influencing performance of radiotherapy personnel. The first achievement has been the development of a taxonomy of 9 PIFs, further specified into twenty-nine influencing factors (categorization done following [1]). In addition, the second achievement was the identification of forty-four negative conditions (similar to Error Producing Condition in HEART) falling in the identified nine PIFs, these characterize how a PIF e.g. "*Human machine interface*" may manifest negatively (e.g. "*Values on interface not easily readable*") when performing a task in radiotherapy.

Generic Task Types (GTTs)

The chapter further dealt with the systematic and traceable identification and characterization of representative tasks specific to radiotherapy. The first achievement was the proposal of a traceable and systematic methodology to develop GTT definitions via identification of the sets of PIFs affecting each GTT. The methodology has strong theoretical foundation in the causal mapping of PIFs to the respective GTT through progressive identification of failure modes, failure causes and failure mechanisms. For this purpose, the mappings were directly linked to a cognitive model of literature [2], which has ensured that the relevant failure modes and influencing factors are covered. This achievement includes traceability feature which was achieved by ensuring that the GTT development was sequential and one could follow how tasks are identified from workflow, transformed, and finally included in the GTT definitions. This traceability property was deemed helpful for future review of the GTTs to incorporate any changes that might occur due to fast changing technology.

The second achievement in this was the successful application of the proposed methodology to radiotherapy domain to construct GTT definitions and identifying PIF sets for each GTT-failure mode. This achievement includes identification of six GTTs for the HRA method and a total of forty-four Example Tasks belonging in them. It further includes identification of eighteen failures modes leading to construction of a total of eighteen GTT-PIFs structures using the framework presented in reference [2] as shown in reference [3]. Example Tasks were a key

contribution to GTTs as they (1) keep the GTT descriptions transparent: it shows which tasks are in the scope of each GTT; (2) assist in assessing the orthogonality for the GTTs when defining them; Example Tasks can be compared and overlapping components can be modified; and (3) improve the usability of the GTT taxonomy: an analyst would clearly know what tasks belong to a specific GTT, and thus would reduce the practitioner's effort [3].

The results obtained were validated against a consensus-based FMEA developed by the AAPM's field experts belonging to various clinics [4]. The comparative results indicate that the GTT-PIF structures address the tasks and influencing factors relevant for the specific application.

7.2.2: Quantitative framework and quantification of HEPs

The achievements of this section cover the objectives #3 and #4 presented in Section **7.1: Research objectives**.

Chapter 3: HRA method- quantitative elements of the thesis dealt with the third challenge of developing HRA method's quantitative structure. It includes four main achievements:

- formation of the quantitative framework i.e. Decision Trees,
- the quantification of the HEPs for the method,
- validation of the generated radiotherapy HEPs against the matching HEPs from existing HRA methods,
- quantification approach not domain-specific (can be transferred to other application domains [5]).

The first main achievement includes development of eighteen decision trees for the method based on the GTT-PIF structures from **Chapter 2: HRA method- qualitative aspects** [3]. The development included key contribution of formation of branch points by either using a single PIF or combination of PIFs and to aid them with the identified specific manifestations of the PIFs (i.e. negative conditions) to assist the analyst in selecting which DT path to choose, consequently the HEP. Inclusion of negative conditions is a key contribution to ease the use of the method.

The second main achievement includes the successful development of the elicitation sessions and aggregation of data to quantify of order of magnitude of the HEPs for the DTs of HRA method using expert judgment. The two key features that were adopted in the expert elicitation sessions were: (1) experts were asked information about the failure probabilities on a qualitative scale, with the goal of getting evidence on the order of magnitude for the probability; and (2) specific situations were presented to the expert, i.e. specific failure scenarios influenced by specific

negative conditions. The latter feature was included for experts to avoid dealing with abstract categories such as tasks types and performance influencing factors. The second achievement was the adoption of systematic and traceable way to aggregate the elicited data from experts using a Bayesian model presented in [6] to determine degrees of belief on the correct values of the HEPs [5]. A total of six out of the eighteen DTs were directly quantified using combination of expert judgment and Bayesian data aggregation approach. The remaining DTs, except four, were quantified by using applicable data from the six DTs.

According to the results obtained from the quantification, "*the strongest effects on the HEP come from the "Lack of training or experience" branching point on the GTT "Complex interaction with software/tool", followed by "Time pressure" and "Information unclear" on "Complex interaction with software/tool" and "Information content unclear-verbal" on the "Verbal communication" GTTs.*" [5]. This showcased the prioritization of effects of specific factors on specific tasks in radiotherapy domain, indicating that these factors would drive the HEP of these tasks when impacting alone and a need to avoid such situations.

The third main achievement included the validation of the results obtained from expert elicitations using relevant applicable data from existing HRA methods. A data comparison between the two sources showed that out of 32 comparisons, in 17 cases the deviation between the median values is below a factor of 3, in 5 cases it is between 3 and 5, in 8 cases between 5 and 10, and in 2 cases it is larger than 10. The cases with larger deviation were found to reflect situations that are specific to the radiotherapy tasks and contexts analyzed by the experts. It can then be understood that these sector-specific influences are not fully reflected by the data. Two key conclusions were derived from this validation exercise: (a) the satisfactory agreement between the expert judgment data and HRA method data provides a first suggestion that HRA data can be used across different domains for tasks and influences that are not specific to a specific industry. (b) and it re-affirms that expert judgment can provide valuable input for quantification of HEPs for HRA methods [5].

7.2.3: Method application to failure sequences

The achievements of this section cover the objectives #5 and #6 presented in **Section 7.1: Research objectives.**

This section includes three key achievements of this thesis:

- Systematic identification of ten failure sequences

- Successful method application to failure sequences
- Suggested workflow improvements based on results

The first achievement includes use of a systematic three-step approach to identify and characterize the failure sequences. It further includes, development of ten failure sequences for an underdevelopment workflow: 4D treatment planning at CPT; with an aim to provide recommendations on how to improve the procedure of it.

An important result observed during identification process indicates that certain important tasks do not have a follow-up proceduralized checking tasks. For example: "*input of series number into "template phase"*" or "*Binning of breathing cycle*". These tasks though being important for the delivery but do not have barrier tasks to reduce potential error propagation. Thus, these tasks were included in the failure sequences for analysis.

A key contribution of the thesis was inclusion of variant failure sequences to study the effect of proceduralizing a check task on the HEP of the failure sequence. These variants also inform about effectiveness of a proceduralized check when performed by the same person who did the primary task or by a different person. And lastly, the failure sequences include study of impact of "*high workload*" on task performance, especially on barrier tasks.

The HRA method allowed a systematic analysis of the 4D treatment planning workflow. Results showed that proceduralization of check task would reduce the failure probability of the sequence by one order of magnitude, but this effectiveness is only when the check task is performed by a different person. A second result was that a self-check performed under high workload has no effectiveness. This ineffectiveness was attributed to the fact that high workload can reduce the person's attention on the checking task, which gets further reduced by the fact that the primary task is easy, and the same person had done it recently. Thus, the results provided first key recommendation on where and what type of barrier tasks should be proceduralized such that even when 4D patient treatment workload increases these barrier tasks remain effective.

Lastly, another key result from the methods application was the analysis of overlap 3D and 4D tasks in 4D treatment planning workflow. The results of analysis of failure event: "*creation of 4D steering file (SF) using 3D software*" showed that the combination of heavy workload with mechanically creating the SF and the overlap of use of 3D software for 4D purposes leads to increase of the failure probability dramatically by one order of magnitude. Further analysis of the two barrier tasks showed that inclusion of them would improve the HEP by one order of magnitude. To reduce the failure probability, it was suggested to consider separating the two

workflows. This would drastically reduce the failure probability as the condition of using 3D software window to create SF would not be present.

7.3: Future work and recommendations

The research performed in this thesis involves the first steps towards developing an HRA method specific to radiotherapy and applying it to study human failures qualitatively and quantitatively. These initial steps further raised a number of possible research directions. These are listed below:

- More failure sequences should be analyzed at CPT and similar studies should be conducted for other radiotherapy centers. A comparative analysis of qualitative and quantitative results between centers would foster learning between centers and the use of quantitative models to study human failures.
- Failure and success (near miss) data collection for the purpose of HRA quantification should be promoted in radiotherapy. A possible approach would be to use the GTT-failure mode classification and the branch points classification as a template to collect data.
- Joint impact of two branch points has been assumed to follow a certain hypothesis. A natural way forward is to conduct a systematic expert elicitation exercise to validate/test this hypothesis with the same conditions and same task failure modes that were used in the first elicitation session. This testing would provide first scientific feedback from expert data on joint impacts of branch points for radiotherapy domain.
- Further research should be performed on modelling task dependencies. Currently, the method is too simplistic to consider all types of dependencies between tasks. Further study on modelling uncaptured dependencies between tasks should be studied.
- The unquantified DTs should be quantified for the method.

7.4: Publications

A total of 5 peer reviewed papers were submitted out of which 4 have been published and one is under review. Out of the 5 publications 3 are conference papers and 2 are journal papers. The list of these is presented below:

Journal Papers:

- 2018- **Pandya D**, Podofillini L, Emert F, Lomax AJ, Dang VN and Sansavini G. Quantification of a human reliability analysis method for radiotherapy applications based on expert judgment aggregation. Submitted to Reliability Engineering and system safety 2018, special issue.
- 2017- **Pandya D**, Podofillini L, Emert F, Lomax AJ and Dang VN. Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy. Proc IMechE Part O: J Risk and Reliability 2017; First published October 2: p. 1–35. https://doi.org/10.1177/1748006X17731903

Conference papers:

- 2018- Podofillini L, **Pandya D**, Emert F, Lomax AJ, Dang VN and Sansavini G. Bayesian aggregation of expert judgment data for quantification of human failure probabilities for radiotherapy. In: Proceedings of European safety and reliability conference; Norway, Taylor & Francis Group 2018
- 2017- **Pandya D**, Podofillini L, Emert F, Lomax AJ, Dang VN and Sansavini G. Quantification of human failure probabilities for radiotherapy: relevance of THERP's values. In: Cepin M and Bris R, editors. Proceedings of European safety and reliability conference; Slovenia, Taylor & Francis Group 2017.
- 2015- **Pandya D**, Podofillini L, Emert F, Lomax AJ and Dang VN. A method for human reliability analysis in radiotherapy: identification and characterization of influencing factors. In: Podofillini L, Sudret B, Stojadinovic B et al., editors. Proceedings of the European safety and reliability conference; Abingdon: Taylor & Francis Group 2015

References:

1. Groth, K. G. and Mosleh, A. A data informed PIF hierarchy for model-based Human Reliability Analysis. Reliability Engineering and System Safety 2012; 108, pp: 154-174.
2. Whaley AM, Xing J, Boring RL, Hendrickson SML, Joe JC, LeBlanc KL and Morrow SL. Cognitive basis for Human Reliability Analysis. Technical report NUREG-2114, U.S. Nuclear Regulatory Commission, Washington DC, January 2016.
3. Pandya D, Podofillini L, Emert F, Lomax AJ and Dang VN. Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy. Proc IMechE Part O: J Risk and Reliability 2017; First published October 2: p. 1–35. https://doi.org/10.1177/1748006X17731903
4. Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr. JP, Ibbott GS, Mundt AJ, Mutic S, Palta JR, Rath F, Thomadsen BR, Williamson JF and Yorke ED. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management, Medical physics 2016; 43 (7): 4209-4262.
5. Pandya D, Podofillini L, Emert F, Lomax AJ, Dang VN and Sansavini G. Quantification of a human reliability analysis method for radiotherapy applications based on expert judgment aggregation. Submitted to Reliability Engineering and system safety 2018, special issue
6. Podofillini L and Dang VN. A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. Reliab Eng Syst Saf 2013; 117: p. 52-64.

Appendix 1

For the development of the HRA method the workflow followed and observed is of Paul Scherrer institute's (PSI) Center for Proton Therapy (CPT). At PSI there are 9 stages in the workflow: (1) Case evaluation, (2) Moulaging and Planning Computed Tomography (CT), (3) Volume definition, (4) Treatment planning, (5) Plan approval, (6) Import of reference images, (7) Plan QA, (8) Treatment delivery, and (9) Follow up. **Figure A** below shows this workflow. Though it shows a linear process it should be noted that it is not and interaction (feedback loops) between stages is possible. In these stages a total of five expertise are involved, namely, radiation oncologists, radiation therapists, medical physicists, dosimetrists (planners), and medical assistants. All these expertise are included in the scope and the tasks performed by them are included in the method development. No particular expertise is solely involved in a single stage and there is a strong interaction between various expertise during each stage.

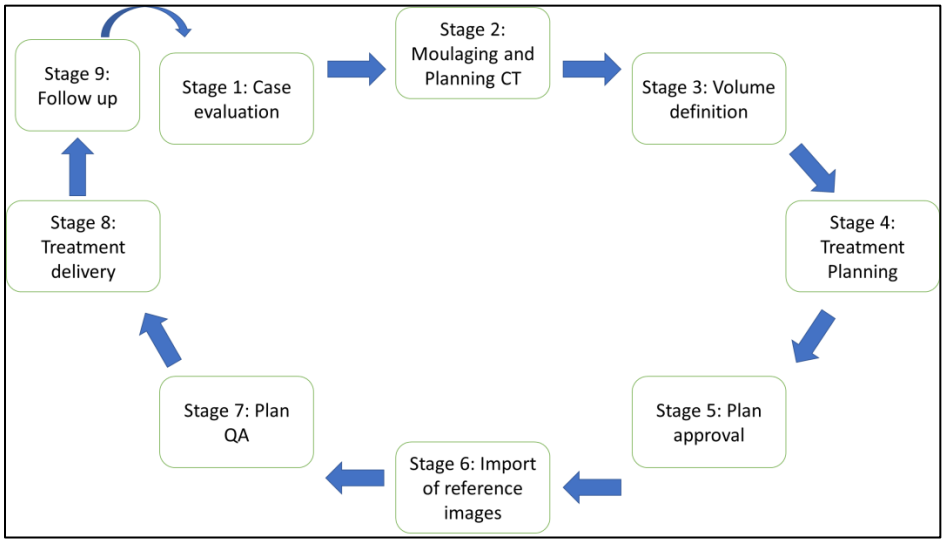


Figure A

In Stage 1, Case evaluation, each patient case is initially evaluated according to his diagnosis and prior treatment history. If proton therapy at CPT could be a treatment option the case is presented in the weekly, interdisciplinary tumor board. After the decision is made if the patient will be treated or not important treatment constraints (expected total dose, treatment position and immobilization, start date) are defined. In Moulaging and Planning CT, the patient gets

immobilized in his final treatment position according to the tumor location and in relation to the expected beam arrangement during irradiation. In this position a planning CT is acquired delivering the major input for the upcoming treatment planning phase. In the Volume definition phase, the physician contours the Volumes of Interests (VOIs), i.e. tumors, critical organs (Organs At Risk - OARs), etc. on the acquired planning CT taking into account all available information (e.g. MRIs) and finalizes the patient's dose prescription. In Treatment planning, the planner then determines how the prescribed dose will be delivered best to the tumor while sparing healthy tissue and OARs using the treatment plan system (TPS). Subsequently, the developed plan is discussed between the involved experts (planner, radiation oncologist, medical physicist) in the Plan approval phase and gets accepted. Based on the accepted plan two sets of positioning images defining the daily treatment position are imported into the Patient Position and Verification (PPV) system which is used to verify daily positioning before each irradiation fraction and to correct for daily positioning uncertainties. This happens in Import of reference images. In Plan QA, the detailed, mainly patient specific checks of the plan, the prescribed dose, and the transferred data are performed. An important dosimetric part of the plan quality assurance are independent dose calculations. At the end of a successful plan QA the plan is finally approved. In Treatment delivery, the patient is positioned daily in the CT according to reference points, transferred to the gantry (the actual radiation delivery device) and delivery of the prescribed fraction dose is performed. In Follow up, the patient is monitored in regular intervals after completion of the entire treatment to observe tumor control and possible side effects.

For the thesis, the main personnel tasks to be in scope of the HRA method are within the stage 1 to stage 8, i.e. from when the patient is admitted to CPT to the time the last fraction dose is delivered. Stage 9 of Follow up is not in scope of the HRA method. As well as in stage 1, only the tasks of patient identification and the expected dose prescription are included; the steps of patient acceptance, how the dose prescription was reached, and the coordination between external hospitals and CPT for information are not included in the scope. This is in line with assumptions in other PSA studies for radiotherapy facilities. Furthermore, the emphasis in the development of the task types for the method is on tasks involved in the treatment process. The

tasks involved in machine calibration and commissioning tasks have been left out of the scope. These tasks are expected to have similar characteristics as of found in other industrial domains as railways, air-traffic and nuclear; if required such tasks can be directly taken from existing HRA methods after scrutinizing their applicability.

Appendix 2

The first corner stone of this HRA method is the set of Performance Influencing Factors (PIFs). These are the conditions that, when present during the task, influence the performance of the personnel on task. For the thesis, the PIF refers to the generic factors influencing performance and not the specific Error Producing Conditions (EPCs) as done in HEART, CARA, NARA. This is performed in such a way that these PIFs can be directly linked to the underlying macro cognitive model used to develop GTT-PIF structures presented. In this thesis the method only captures the negative influence of a PIF following HRA methods e.g. HEART, CARA, NARA etc.

Sources used for the identification of the PIFs are:

- Past events from generic databases like ROSIS, SAFRON, facility specific like CIRS,
- Reports like [1, 19, 20, 21, 23], etc.
- Existing taxonomies built to classify errors including human like [1, 28, 29, 31].
- Observations performed at CPT
- Discussions with experts working in the radiotherapy field.

PIF Taxonomy for radiotherapy:

A total of 29 Influencing factors were identified using the sources mentioned. These were further categorized into 9 high level PIF "labels" based on similar characteristics, similar influences on performance. For example, training, knowledge, experience, and familiarity of situation have been grouped into one PIF label of "Training-Experience" as their description and their manner of influence are similar. This grouping is necessary to have a limited number of PIFs to be used for an HRA method and also to understand and ensure the orthogonality of the PIF definitions. **Table A** below shows the nine PIF with their definitions. Then Table B presents forty-four identified negative conditions falling in each PIF for the radiotherapy domain. One should note that the negative conditions identified and listed may not be exhaustive but aim to cover various forms of the manifestations specific to radiotherapy.

\

Table A

PIF Definitions
Training-Experience: It is defined as the knowledge gained over a certain period of time by either receiving it as training from the center/hospital or as experience by working. The training is mostly provided when a person joins a new center or when a new equipment or technique is installed at a center. This PIF covers three aspects: availability, quality and recency. Availability adjudges if there is a training program available or if the task is part of the training or if the person is experienced or not. Quality adjudges the adequacy of the content, maintenance and how it is delivered of both training on tools and on tasks and also the adequacy of the previous experiences on the same. Recency adjudges how recent the training is provided or if there is periodic competency evaluation program or how recent was the experience. **Resources**: It includes procedures, software, hardware, tools and any additional necessary documents/information required to perform the task or help in performing the task. The procedures represent the set of tasks to be performed by the personnel. These are mostly found in written/documented format. Similarly, software, hardware and tool are the equipment that are used in process. And lastly, Necessary information is the information about the system or which is needed to perform the task . The information includes equipment manuals, logs from previous fractions or task, and maintenance logs. It specifically does not include information in the procedures (which is covered in procedures). Overall, the PIF covers aspects of availability and quality of the mentioned resources. Where, availability adjudges the availability of a concurred resource and Quality checks the overall adequacy of the resource like clarity in content, completeness, proper calibration of the equipment, and its maintenance/update. One should note that in radiotherapy there are procedures which are orally accepted and need not have a document. Such orally accepted procedures are also included in the PIF procedures. **Complexity**: It is defined as cognitive or physical demands required by a task. It also deals with characteristics that can make a task complex. The PIF covers confusion/ambiguity and difficulty. Each dimension has further been split into two different sub-dimensions, execution and cognitive. Difficulty could be execution difficulty or cognitively difficult, with both being equally dominant; ambiguity would mostly be cognitive rather than execution. Confusion characterizes those conditions that can confuse the person for example similar identification marks of the body of the patient. Difficulty characterizes conditions that can make the task difficult for example more information to be remembered to perform a task, more number of sub tasks, some unusual calculation etc. **Environment**: This factor deals with the environment that are external to both the machine and the person. It refers to working environment which includes temperature, noise, lighting, humidity and other external weather factors. The working environment in radiotherapy is not controlled; the work process is conducted in many different working places (like office, office-reception, treatment room, storage room etc.) and were observed to be affected by lighting and more importantly noise. Thus, this PIF deals with Noise in detail; Noise has been further characterized into distractions and interruptions that take place in the workspace for example phone calls, relatives being present etc. Interruptions are predictable and frequent and

practitioners can develop a working way around them and distractions can be lengthier and lead to loss of attention. Interruptions could be due to phone calls or colleagues or even patient's relatives. Distractions would be due to long alarms or long disturbances that reduce attention/ cognition required for the task.

Human Machine Interface: This deals with the ergonomics, usability and physical access: which include design, quality and quantity of information displayed and layout of software/tool and displays. Both input and output have been considered. Difference between equipment and software interfaces are emphasized by the present taxonomy. Here it would also incorporate the home-made tools and softwares. Thus, the definition of this PIF has been expanded to include aspects of usage of in-house software and tools and not just purchased. Many radiotherapy institutes develop their own software and tools to perform tasks and thus the implications need to be considered in the analysis.

Team: Team is any group of people expected to work together to complete a task. They are expected to interact directly either in person or over phone. The defining characteristic of the team is that people work together to achieve a common goal. This PIF covers communication and coordination, which includes the quality and clarity of the information communicated or the language/jargon used in communication. Availability and quantity of communication is not covered in this but would rather relate to safety culture as the person who communicates should know what all information should be given.

Personal: This includes person's internal factors that affect each individual. It encompasses the worker's state of mind, temperament, and various intrinsic characteristics like bias, motivation, morale etc. It also includes the person's physical and mental fitness. It also includes the monotonous nature of task that can form bias leading to loss of attention (than required).

Loads: It characterizes conditions which can lead to stress while performing a task. It does not characterize stress due to difficult or confusing task; these have been already modeled in complexity. It captures stress which is caused by work load, time pressure and stress due to patient's medical condition. It includes task load, time load, other loads etc.

Safety Culture: This characterizes the organizational attitude, values, and beliefs toward worker and public safety. The PIF is not inherently observable but one can observe how safety culture is implemented. So, in the present taxonomy the PIF only covers the direct and observable influences on work processes e.g. adherence to procedures and quality assurance, awareness in what and how much information to pass and if simultaneous works have been assigned.

Table B

PIFs	Sub category of PIFs	Negative conditions
Training-Experience	Training, Knowledge, Experience, Familiarity with situation	• Lack of familiarity with the case or situation • Lack of training or experience on treating special tumor locations • Training program (facility or vendor) does not cover information on specific constraints for special tumor locations • New software given without training (lack of training for new software) • Lack of training or experience or knowledge on proton physics given to a newcomer who does not have a background in it • No in-depth training on generic parameters of TPS for newcomers • Training program (lack of experience) does not cover guidance on selecting specific methodology for special tumor situation. • Training does not cover or not experienced to know all key information is to be given for specific or special treatments/ tumors or new or adapted techniques
Safety Culture	Organizational culture, Management activities, Work processes	• Person not aware of importance of the check or importance of check to be performed completely and not partially • Simultaneous work assigned • High volume of information is communicated without possibility to write • Not complete information given on the background
Resources	Procedures, Tools, Necessary information	• Some information incomplete • Procedure not available • Information not available • Specific information required for planning not available • Update on rarely used software not known • Procedure not clearly understandable • Software or tool not available • Bad phone reception
Team	Communication, coordination	• Language barrier between personnel • Jargon or technical lingo used not familiar • Information communicated not clear

		• Caller is simultaneously occupied with another work • Person receiving the data is busy with other work • The information receiver is busy with monitoring other task or a patient
Personal	Fatigue, Bias, Morale, motivation, Physical abilities	• Task is mechanical and repetitive • The check was performed recently and you trust it was performed correctly • The task to produce the output is simple - no error is expected • Monotonous behavior of the task- sequentially checking many times • Expecting no error as the check is performed by the same person doing the initial task • Expecting no error as the task was self-performed • The file or device was left here yesterday or last time • Items were put by another personnel and if one ID matches, all the other IDs must match
HMI	Machine interface, System response	• The values written look alike • The value on the label or file are not easily readable • No Id on the support device • The indications for the input boxes on the screen are not readable • The layout of the input interface is unusual or confusing • Touch screen or input keyboard is very sensitive (unknown mis-selection is possible) • Logic of the interface not adequate (pressing 'Enter' necessary after value insertion) • Mental calculation required due to ineffective logic of interface • Too much information on the software screen leading to confusion • The ordering of the values on the control document and on the screen do not match (i.e. X, Y, Z on screen and Z, X, Y in the document) • Swapping through the pages or scrolling through screen is necessary • Lead marks are small and difficult to find • Full scan is not visible on the screen (needs to be scrolled) • Necessary information appears on multiple screens • Bad quality of imaging • The device looks the same as another device • Scans on the interface are not readable or clear

Loads	Workload, Time pressure, Time load, Other loads, Stressors	• Simultaneously doing another urgent task • There is less time to do the task (excessive workload) • Simultaneously tracking/supervising another task • Overloaded with other work • Time pressure due to information not coming in time • Time pressure due to workload • Stress due to patient's bad health
Complexity	Task complexity	• Tumor present where it can move (mobile tumor) • Tumor located where air cavities can change over time • Multiple tumor parts physically close to each other but not joined • Tumor located next to multiple critical organs • Multiple artefacts present • Non-usual way to calculate the dose distribution • Physical changes to the patient occur frequently • Manual conversion of units required • Too much information given
Environment	Distraction, Interruption, Noise, Lightning etc.	• The background noise level is too high, it distracts the focus • Interruptions from colleagues while doing the task • The lighting and other environmental conditions are not adequate

Appendix 3

HTA of the process

HTA- Moulaging and CT scan	
1	Preparation of moulage (moulage formation)
2	Fix and adjust the moulage
3	Adjusting fixation devices
4	Marking of fixation devices
5	Input of ID on fixation devices and moulages
6	Taking photos of fixation devices and moulages
7	Accuracy checks for positioning
8	Selecting region for Computed Tomography (CT) scan (including lead marks)
9	Inserting markers on the CT scan (orientation)
10	Taking Photos in treatment position
11	Transport and storage of devices and moulages
CT import into Treatment Planning System	
12	Identification of CT files to import based on ID
13	Selection of data in High Resolution CT
14	Saving of files electronically
15	Detailed Quality checks of data within High Resolution CT (HRCT) and CT
16	Recording of coordinate data onto the form
17	Sign and send form
File fusion	
18	Identification of files based on ID
19	Quick check of tumor location
20	Copy of data from one software to another
21	Iterative task of achieving an optimum fused file by adjusting parameters and with an aid of visual output
22	Saving the file into the folder electronically
Volumes of Interest (VOI) determination	
23	Identification of file from a list of files based on ID.
24	Drawing contours around (located) tumor and Organs at Risk (OAR) for every slice
25	Assigning names and keys to VOIs and OARs
Treatment planning	
26	Copy of data from one file to another
27	Decide and select plan mode, beam directions, fields
28	Iterative calculation of doses with visual output checks and practical considerations
29	Short Quality check of prescribed dose limits to OARs and VOIs
30	Quality check discussion with medical doctor and medical physicist
Pre-Plan Quality Assurance (QA)	
31	Detailed QA of plan, reverse/ back calculation and matching

32	Create steering file
33	Correction of Nuclear interactions
34	Printing of the Steering file
Airgap measurement	
35	Printing of file based on ID
36	Sending of file to Medical Technical Radiology Assistant (MTRA)
37	Identify and bring patient equipment
38	Fixation of moulages and other fixation device
39	Quality check of gantry angles and other parameters
40	Position of patient with given coordinates and check of air gap for any collision possibility for each field
41	Leaving the equipment for MTRA to store it
Digital Reconstructed Radiographs (DRRs) generation and Steering File import to Gantry	
42	Create new file with patient ID
43	Data input from paper
44	Calculate DRR
45	Adjust window level for DRR
46	Save the file
47	Create patient Folder in Patient Positioning Verification (PPV)
48	Input patient data into PPV
49	Select table and head support in the file
50	Select the folder where DRR is to be imported
51	Import DRR
52	Check the imported file
53	Selecting the Plan Steering File (SF) to be sent to Gantry and send
54	Check the arrival of SF in Gantry
55	Print the documentation

Appendix 4

List of Example Tasks with macrocognitive functions

Example Task	**Macrocognitive Functions**
Calling name out loud (in front of a group) and matching with some Identification	Detecting and noticing (D/N)
Ask for the name to the patient and match with the ID card.	D/N
Checking the patient-specific parameters like patient identification number, weight, birthmarks etc.	D/N
Look at the photograph and match	D/N
Identification of file or tool from a list by looking at the ID	D/N
Identification of file or tool by remembering the ID (memory based)	D/N
Identification of file or tool while person is reading out the ID	D/N
Check of transferred patient data to a software/ machine	D/N
Check of clinical limits or data	D/N
Air gap check	D/N
Two personnel read aloud cross check	D/N
Detailed check of the plan by doing back calculation	D/N, Sense-making and understanding (SM/U)
Detailed check of data within documents (e.g. CT scans)	D/N, SM/U
Detailed check of the plan between doctor, medical physicist and planner	D/N, SM/U, Team coordination
Deciding and assigning plan mode, beams, directions and other parameters	SM/U, Decision-making (DM)
Performing iterative calculation of doses	DM
Hand calculation of dose or weights of each field or any other calculation	SM/U, DM
Consider organ motion effect and other special effects	SM/U, DM
Moulage preparation (with planning)	DM
Decision on devices like head support, table type etc.	SM/U, DM
Adjusting the immobilization devices	DM
Marking tattoos on skin	SM/U
Planning positioning using laser coordinates	SM/U
Positioning with photos, scans or any other document	SM/U, Action
Fixing the prepared moulages (matching the marks)	SM/U, Action
Positioning with tattoos on skin	SM/U, Action
Using laser coordinates to position	SM/U, Action
Looking and capturing data (with understanding what is to be done)	SM/U, Action
Selecting data while simultaneously performing some mental task	SM/U, Action
Acquiring data while someone is dictating including understanding	SM/U, Action
Contouring of volume of interests (VOIs) like tumours or Organs At Risk (OAR) (mechanical task)	SM/U, Action

Looking and copying/inputting data	Action
Simple input of data from memory (also while performing a mental task like calculation) without any aid	Action
Simple data input while someone is dictating (purely execution)	Action
Import or export of data to another location (electronically or manually)	Action
Patient identification using barcodes	Action
Patient identification using chips in the wrist of the patient	Action
Personalized magnetic strip cards based identification	Action
Selection/ pick up of chart/file or tool	Action
Transfer of the file or tool (including storing)	Action
Give instructions	Team Coordination
Give clinical or identification data /information	Team Coordination
Any kind of update	Team Coordination

Appendix 5

GTT-PIF structures

GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Identification of patient or patient related items	Patient identification by human, Identification of file/tool	Calling name out loud (in front of a group) and matching with some identification card, ask for the name to the patient and match with the ID card, look at the photograph and match, identification of file or tool from a list by looking at the ID	D/N	Patient information incorrectly matched	Cues/information misperceived	Cue content, vigilance- attention, expectation	√		√		√	√	√		√
			DM	Identification check not performed (decision-based)	incorrect goals	Goal conflict, incorrect goal selected, incorrect prioritization of goals, incorrect judgment of goal success	√	√			√		√		
			Action	Failure to execute desired action (error of omission)	failure to execute desired action	Divided attention							√		√
GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Quality Check	Short quality check, Detailed quality check	Check of transferred patient data to a software/ machine, air gap check, check of clinical limits or data, immobilization and positioning checks, two personnel read aloud cross check, detailed check of the plan by doing back calculation, detailed check of data within documents (e.g. CT scans), detailed check of the plan	D/N	Deviation from requirement not recognized	Cues misperceived	Cue content, vigilance- attention, expectation	√		√		√	√	√		√
			SM/U	Inappropriate understanding of underlying principles (for detailed quality check)	Incorrect integration of data, frames	Data not properly recognized, improper integration of information, incorrect match of data selected for comparison, improper control of attention	√		√			√	√	√	
					Incorrect frame	incorrect frame used to attend to information	√							√	
			DM	Check not performed (decision-based)	incorrect goals	Goal conflict, incorrect goal selected, incorrect prioritization of goals, incorrect judgment of goal success	√	√			√		√		
			Action	Execute desired action incorrectly	execute desired action incorrectly	motor learning, automaticity control, recognition error	√		√			√	√		√

		between doctor, medical physicist and planner		Failure to execute desired action (error of omission)	failure to execute desired action (error of omission)	divided attention						√		√
			TC	Coordination failure	Error in leadership/ supervision	Failure to consider information communicated by an individual		√		√			√		
GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Complex interaction with software or tools	Standalone selection of data, replication of patient immobilization and positioning	Looking and capturing data with understanding the data, selecting data while performing a mental task, contouring of volume of interests like tumors or Organs At Risk (mechanical task), positioning with tattoos on skin, using laser coordinates, positioning with photos, scans or any other document	SM/U	Misinterpretation of data --Focus on incorrect or incomplete data --Patient positioned out of tolerance limits due to misinterpretation of correct data	Incorrect integration of data and frames	Data not properly recognized, improper integration of frames/ information, incorrect match of data selected for comparison, improper frame selected for data comparison, improper control of attention	√		√			√		√	√
					Incorrect frame to understand the situation	incorrect frame used to attend to information, incorrect frame used to interpret information, attend to wrong information	√		√			√		√	
					Incorrect Data	Information available in the environment is not complete			√			√		√	
			Action	Execute desired action incorrectly	Executed desired action incorrectly	manual control issues, automaticity control, dual task interference	√					√	√		√
			D/N	Mismatch or inconsistency not recognized	Cues/information misperceived	vigilance, cue content	√		√		√	√	√		√
GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV

Simple interaction with software or tool	Input/ entry of data, transfer of data, patient Identification using machine, selection and transfer file or tool	Looking and copying/inputting data, inputting data while someone is dictating, inputting data from memory, export of data to another location, acquiring data when someone is dictating, selection/ pick up of chart/file or tool, transfer of the file or tool (including storing)..	Action	Execute desired action incorrectly -Slip in data input, wrong location, default value partially overwritten, look at wrong data	Executed desired action incorrectly	motor learning, manual control issues, automaticity control, recognition error, dual task interference	√				√	√	√		√
				Failure to execute desired action (error of omission)	Failure to execute desired action	Working memory failure, prospective memory failure, divided attention	√						√		
			D/N	Failure to identify relevant information	information not attended to	vigilance			√		√		√		√
					information not perceived	Vigilance			√		√	√	√		√
GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV
Iterative determination of optimum parameters	Planning patient immobilization & positioning, treatment plan calculation	Moulage preparation (with planning), decision on devices like head support, table type etc., adjusting the immobilization devices for final position, assigning plan mode, beams, directions and other parameters and perform calculations, hand calculation of dose or weights of each field or any other calculation	SM/U	Misinterpretation of information	Incorrect frame	incorrect frame used to attend to information, incorrect frame to interpret information	√							√	
					incorrect integration of data and frame	improper integration of information, improper control of attention, mental manipulation of the information is inadequate, data not properly recognized or distinguished	√		√			√	√	√	√
					Incorrect Data	Information available in the environment is not complete			√			√		√	
			DM	Inappropriate decision on strategy selection (tumor-specific)	Incorrect internal pattern matching	incorrectly comparing the mental model to previously encountered situation, incorrect recall of previous experience	√						√	√	
					incorrect mental simulation	incorrect inclusion of alternatives, misinterpretation of procedures, cognitive biases	√		√				√	√	
GTT	GTT- sub group	Example Tasks	MCF	MCF failure	Proximate cause	Failure mechanisms	TR-EX	SC	RE	TE	PE	HMI	LO	CO	ENV

Verbal communication	Verbal Communication	Give instructions Give clinical or identification data	TC	Communication failure	Failure of team communication	Source error of commission	√				√		√	√	√
					Failure of team communication	Source error of commission, target error of commission	√			√			√	√	√
					Failure of team communication	Source error of commission, target error of commission		√		√			√		
					Failure of team communication	Source error of omission	√	√		√			√		
					Failure of team communication	incorrect timing of communication	√	√					√		
			DM	Not communicated (decision based)	Incorrect goals	incorrect goal selected	√	√					√		

Legend: MCF- Macrocognitive Function, TR-EX- Training-Experience, SC- Safety Culture, RE- Resources, TE- Team, PE- Personal, HMI- Human Machine Interface, Lo- Loads, Co-Complexity, ENV- Environment

Appendix 6

Definitions of GTTs (selected)

Identification of patient or patient related items

This Generic Task Type deals with the identification of either the patient to be treated or the patient-specific items (like tools, moulages, files, charts etc.) required for the radiotherapy treatment process. The identification task is done by comparing the specified item like ID number, patient's face or any other unique feature of the patient (e.g. birthmark, patient full name etc.) with the number, photograph or the feature on the control document. The identification of patient takes place as many times in the process as the patient comes to the institution; starting from the first time patient comes for Planning Computed Tomography scan till the delivery of the last fraction. The identification of patient's items occurs throughout the process whenever they are required.

Typical tasks for the identification of the patient could be "c*alling the name out loud, asking the patient for his/her name, looking at the photograph, and checking an identity card or a combination of two tasks"*. It is to be noted that calling out name and asking for name is included in this task type and is not part of verbal communication task type as this specific communication is done during identification of the patient only. In case of identification of the patient related items the tasks include "*Identification of file or tool from a list by looking at the ID, Identification of file or tool by remembering the ID (memory based), Identification of file or tool while person is reading out the ID"*.

Iterative determination of optimum parameter

This Generic Task type deals with the tasks that are aimed to determine an optimum solution (final parameters or final products) by iteratively engaging in the task. The task typically includes some problem-solving or planning aspect that is to be based on the inputs or constraints of the tumor (like its location etc.).

One of the common characteristics of these tasks is that most tasks are repeatedly performed to achieve the optimum final plan. These iterations of the same task are coupled with fast checks (mostly visual) to see how good the intermediate output is. They might also include iteratively inputting of data like beam angle, size etc. to achieve the optimum solution. The example tasks include: *"Deciding and assigning plan mode, beams, directions and other parameters, Performing iterative calculation of doses, Hand calculation of dose or weights of each field Or any other calculation, Consider Organ motion effect and other special effects, Moulage preparation (with planning), Decision on devices like head support, table type etc.,*

Adjusting the immobilization devices (matching the marks), Marking tattoos on skin, Planning positioning using Laser coordinates, Re-calculation etc.".

Complex interaction with software or tool

This Generic Task Type deals with those tasks that require some level of understanding (cognitive activity) of what is to be done or required of the situation while interacting with the software or tool. For example, it includes tasks that are related to standalone tasks of acquire/selection/capture of data (standalone i.e. without transfer) that require some understanding of what is to be captured for the specific situation or tasks that are to be done by understanding a document etc. These are not simple tasks and require some understanding of the situation or task apart from execution. It should be noted that these tasks exclude the iterative planning tasks or problem solving tasks, also tasks that are simple execution tasks which include no understanding of the situation.

The examples of such task include "*Positioning with photos, Scans or any other document, Fixing the prepared moulages (matching the marks), Positioning with tattoos on skin, Using laser coordinates to position, Looking and capturing data with understanding, Selecting data while simultaneously performing some mental task, Acquiring data while someone is dictating, Contouring of volume of interests like tumors or Organs At Risk (mechanical task)".*

Simple Interaction with Software or tool

This Generic Task type deals with the simple execution tasks and do not require major cognitive effort. These tasks include aspects of the management of the (optimum) data, numerical or photographic, tools files and any other items and machine based identification of patient or tool etc. The characteristics of it are "*Input* or *Enter*, *Copy* or *Transfer and identification using machine"* of data, file or tool or patient. The word "optimum" clarifies that this Task Type does not include iterative insertion of data for calculation of clinical data to reach optimum values via software (inserting a data field again and again to see the output). Data could be a number, text, image, points, and files include Charts, prescriptions, documents etc. In other words these would be technical parameters, personal credentials, Scans (MRI, CT scan), patient identification numbers, charts, prescription documents, other documents etc.

Appendix 7

Validation of the GTT-PIF structure against FMEA for radiotherapy FMEA from Huq et al.				**GTT-PIF structure proposed in the present work**			
No	**Major Process**	**Failure Modes**	**Potential causes (factors)**	**Applicable GTTs (Failure Modes)**	**Proximate Cause**	**Failure Mechanisms**	**Corresponding PIFs**
31	Other pre-treatment imaging for Clinical Target Volume localization	Incorrect interpretation of tumor or tissue for images	Inadequate training (user not familiar with modality)	Complex interaction with software /tool (misinterpretation of data)	Incorrect frame	Incorrect or inappropriate frame used to attend to information	Training/ Experience
			Lack of communication (intended as missing necessary information)		Incorrect data	Information available in the environment is not complete to understand the situation	Resources (necessary information)
58	Radiation Treatment Planning anatomy	3*sigma contouring error: wrong organ, wrong site, wrong expansions	Lack of standardized procedures	Complex interaction with software /tool (misinterpretation of data)	Incorrect integration of data and frames	Improper integration of frames/information	Resources (procedures)
			Inadequate design specification,		Incorrect integration of data and frames	Data not properly recognized	Human Machine Interface
			inadequate assessment of operational capabilities		Incorrect frame	Incorrect or inappropriate frame used to attend to information	Training/ Experience
			Inattention		Incorrect integration of data and frames	Improper control of attention	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
			Lack of staff (rushed process, lack of time, fatigue)		Incorrect integration of data and frames	Improper integration of frames/information	Loads (Work load, time pressure)
48	Initial Treatment planning	Wrong summary of other	Lack of staff (rushed process, lack of time, fatigue)	Iterative determination of optimum parameters	incorrect integration of data and frames	Improper integration of frames/ information	Loads (Work load, time pressure)

	Directive (from Medical doctor)	treatments, other treatments not documented	Inattention	(misinterpretation of information)	Incorrect integration of data and frames	Improper control of attention	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
			Lack of communication, wrong information obtained, information not available		Incorrect data	Information available in the environment is not complete to understand the situation	Resources (necessary information)
			Wrong reconstruction of previous events		Incorrect integration of data and frames	mental manipulation of the information is inadequate, inaccurate or otherwise inappropriate	Training/ Experience
59	Radiation Treatment Planning anatomy	Excessive delineation errors	Lack of standardized procedures, availability of defective tool	Complex interaction with software /tool (misinterpretation of data)	Incorrect integration of data and frames	Incorrect match of data selected for comparison	Resources (Procedures, tools)
			Inadequate design specification		Incorrect integration of data and frames	Data not properly recognized	Human Machine Interface
			Lack of staff (rushed process, lack of time, fatigue)		Incorrect integration of data and frames	Improper integration of frames/information	Loads (Work load, time pressure)
			Inattention		Incorrect integration of data and frames	Improper control of attention	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
			Inadequate assessment of operational capabilities, Inadequate assessment of tool for task, inadequate training, *tool used incorrectly*		Incorrect frame	Incorrect or inappropriate frame used to attend to information	Training/ Experience

65	Radiation Treatment Planning anatomy	Margin width protocol for PTV is inconsistent	Lack of standardized procedures, lack of communication (intended as missing necessary information)	Complex interaction with software /tool (failure to execute desired action)	Cues/information misperceived	Cue/ information content	Resources (Procedures, necessary information)
			Lack of staff (rushed process, lack of time, fatigue)		Cues/information misperceived	Vigilance	Loads (Work load, time pressure)
			Inattention		Cues/information misperceived	Vigilance	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
			Inadequate training		Cues/information misperceived	Vigilance	Training/ Experience
137	Plan approval	Bad plan approved	Miscommunication	Quality Check (Execute desired action incorrectly)	Execute desired action incorrectly	Recognition error	Team (communication)
			Inattention	Quality Check (deviation from requirement not recognized)	Cues/information misperceived	Vigilance	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
			Lack of procedures, inadequate procedures		Cues/information misperceived	Cue/ information content	Resources (Procedures)
			Inadequate training, *incorrect procedure used*[4]		Cues/ information misperceived	Expectation and Vigilance- attention	Training/ Experience
			Procedures not followed	Quality Check (Check not performed)	Incorrect goals	Incorrect prioritization of goals	Safety Culture, Loads (Work load, time pressure)
205	Day N treatment	Special motion methods not applied or incorrectly	Poor software or hardware design	Complex interaction with software /tool (misinterpretation of data)	Incorrect integration of data and frames	Data not properly recognized	Human Machine Interface

		applied	Inadequate training	Complex interaction with software /tool (Mismatch inconsistency not recognized)	Cues/ information misperceived	Vigilance	Training/ Experience
46	Initial treatment planning directive (from Medical Doctor)	Special instructions not given or wrong instructions given	Documentation not there, lack of communication	Iterative determination of optimum parameters (Misinterpretation of information)	Incorrect data	Information available in the environment is not complete to understand the situation	Resources (Necessary information)
			Lack of staff (rushed process, lack of time, fatigue)		Incorrect integration of data and frame	Improper integration of information, data not properly recognized	Loads (Work load, time pressure)
			Inattention		Incorrect integration of data and frame	Improper control of attention	Loads (Work load, time pressure), Environment (distractions, interruptions, noise)
126	Treatment planning	Inadequate evaluation of plan	Not enough time	Quality check (Check not performed)	Incorrect goals	Incorrect judgment of goal success	Loads (Work load, time pressure)
			Inadequate training, poor evaluation strategy, *incorrect final prescription*		Incorrect goals	Goal conflict, incorrect goal selected	Training/ Experience

Appendix 8

This Appendix presents all the 18 developed decision trees for the HRA method. It explains what each branch point is aiming to capture in each DT. One DT represents one GTT-Failure mode derived from the GTT-PIF structures. The appendix also presents raw expert judgment data and the calculated HEP for the quantified DTs. As explained before during the assumptions of the method the raw data (expert judgment and HEP) from one GTT-FM can be transferred to another GTT-FM only where the failure mode and the branch points are the same; this exchange has been color coded by red in the DT.

GTT 1: Simple Interaction with software or tool

Failure mode: Executed desired action incorrectly

This has 4 Branch points. The first branch point "Information unclear" aims to capture the inadequate aspects of the machine interface, like indication boxes, proximity of input boxes and logic of interface, that can lead to simple execution failure. The second "Distractions/ interruptions and excessive workload" focusses on capturing attention failure due to workload or personnel interruptions during the work. The third "Low vigilance due to expecting no error" aims to capture the aspect of monotonous nature of the task or monotonous nature of the information itself. The last "Environmental distraction" aims to capture the aspect of attention loss but now due to the surrounding environment. The conditions falling in each branch point are given below.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Information unclear	The indications for the input boxes on the screen are not readable	L M L H L	0.0011
	The layout of the input interface is unusual or confusing	L M L L L	0.001
	Touch screen or input keyboard is very sensitive (unknown mis-selection is possible)	L L L M M	0.0011
	Logic of the interface not adequate (pressing enter)	M L - H -	0.0098

	Mental calculation required due to ineffective logic of interface	M M - H M H	0.0112
Distractions/ interruptions and excessive workload	Simultaneously doing another urgent task	M H L M L M	0.0094
	Interruptions from colleagues while doing the task	L M L L M M	0.0024
	There is less time to do the task (excessive workload)	L M L M M	0.0088
	Simultaneously tracking/supervising another task	L M L L L M	0.001
Low vigilance due to expecting no error	Task is mechanical and repetitive	L M L L M	0.0011
Environmental Distraction	The background noise level is too high to distract the focus	L L L L M	0.001

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0018
B	Distractions/ interruptions and excessive workload	0.0046
C	Low vigilance due to expecting no error	0.0011
D	Environmental distraction	0.001
E	A and B	0.0046
F	A and C	0.0018
G	A and D	0.0018
H	B and C	0.0046
I	B and D	0.0046
J	C and D	0.0011
K	A and B and C	0.01
L	A and B and D	0.01
M	A and C and D	0.01
N	B and C and D	0.01
O	A and B and C and D	1

Failure mode: Failure to execute desired action

It should be noted that the two branch points and their respective negative conditions in this DT are similar to the first failure mode. Thus, their descriptions and their values from the first failure mode, above, can be used.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Distractions/ interruptions and excessive workload	Simultaneously doing another urgent task	M H L M L M	0.0094
	Interruptions from colleagues while doing the task	L M L L M M	0.0024
	There is less time to do the task (excessive workload)	L M L M M	0.0088
	Simultaneously tracking/supervising another task	L M L L L M	0.001
Environmental Distraction	The background noise level is too high to distract the focus	L L L L M	0.001

Median HEP of the branch points

#	Branch point	Median HEP
A	Distractions/ interruptions and excessive workload	0.0046
B	Environmental distraction	0.001
C	A and B	0.0046

GTT 2: Quality Check

Failure mode: Deviation from requirement not recognized

This has 3 branch points. The first “Information unclear” focusses on readability, ordering and presentation of the information on the interface (screen or paper). The second “Low vigilance due to expecting no error” focuses at the aspects of bias and dependency that will reduce the attention in doing the task, these are mostly captured by looking the characteristics of the task itself and the expectation of the personnel with the task. The last, “Distractions/ interruptions and excessive workload” focusses on capturing attention failure due to workload or personnel interruptions or environmental distractions during the work. It does not capture simultaneous work as quality checks are not performed in simultaneously.

Branch point	Negative conditions	Expert judgment data	Median HEP
Information unclear	The values on the interface are not easily readable	L M M L M M	0.0088
	The values look alike	L H M L H M L	0.0016
	Too much information on the software screen	M M L L M H	0.0094

	leading to confusion		
	The ordering of the values on the control document and on the screen do not match (i.e. X,Y,Z on screen and Z,X,Y in the document)	M H M M L L L M H	0.0097
	Swapping through the pages or scrolling through screen	M M L L M M H	0.0099
	Lead marks are small and difficult to find	M L L L	0.001
Low vigilance due to expecting no error	The check was performed recently and you trust it was performed correctly	M M L L H M M	0.0099
	The task to produce the output is simple - no error is expected	M L M M H L M	0.0099
	Monotonous behavior of the task- sequentially checking many times	M M L H L L M	0.0056
	Expecting no error as the check is performed by the same person doing the initial task	M M H M M M H	0.01
Distractions/ interruptions and excessive workload	Overloaded with other work	M L M L M M H	0.0099
	There is little time to do the task	M M M L H M H	0.01
	The background noise level is too high, it distracts the focus	L M L L L M M H	0.0012
	Interruptions from colleagues while doing the task	L M M L L H M H	0.0078

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0068
B	Distractions/ interruptions and excessive workload	0.0073
C	Low vigilance due to expecting no error	0.0091
D	A and B	0.1
E	A and C	0.1
F	B and C	0.1
G	A and B and C	1

Failure mode: Inappropriate understanding of underlying principles

This failure mode is for detailed quality check type of task. The failure mode captures the aspect of misunderstanding of underlying radiotherapy/technical principles required to perform a detailed calculation check. It has 5 branch points. The first “Unclear information” focusses on

capturing the readability/ambiguity and the presentation of the information on the interface. The second, "Tumor complexity" aims to capture tumor difficulties like location, proximity, size, movement etc. that impact the understanding of the issue. The third, "Lack of training or experience" aims to capture situations where the personnel is not familiar with the case or has no training or experience in handling special tumor locations. The fourth "Time pressure" aims to capture lack of time to perform the task which could be due to workload or due to the late arrival of the relevant information. The last "Resource unavailable" aims to capture the aspect of incomplete information available to the understand the whole situation.

The data for this DT has been taken from Complex interaction with software/tool – mis-interpretation of Data DT.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Information unclear	Full scan is not visible on the screen (needs to be scrolled)	M M L	0.0089
	Necessary information on multiple screens	M M M	0.01
	Bad quality of imaging	H H E	0.103
	Scans on the interface not readable or clear	E H H	0.103
Tumor Complexity	Tumor present where it can move	L M L	0.0011
	Tumor located where air or water cavity can change over time	M L L	0.0011
	Multiple tumor physically close to each other but not joined	M L L	0.0011
	Tumor located next to multiple critical organs	M L M	0.0089
	Multiple artefacts present	L M L E	0.0012
Lack of training or experience	Lack of familiarity with the case	L H L	0.0011
	Lack of training or experience on treating special tumor locations	H E H	0.103
Time pressure	Time pressure due to information not coming in time	M L H	0.0098
	Time pressure due to workload	H L H	0.0897
Resource unavailable	Some information incomplete	H L M	0.0098

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0271

B	Tumor complexity	0.0014
C	Lack of training or experience	0.0642
D	Time pressure	0.0451
E	Resource issues	0.0098
F	A and B	0.0271
G	A and C	0.1
H	A and D	0.1
I	A and E	0.1
J	B and C	0.0642
K	B and D	0.0451
L	B and E	0.01
M	C and D	0.1
N	C and E	0.1
O	D and E	0.1
P	A and B and C	0.1
Q	A and B and D	0.1
R	A and B and E	0.1
S	A and C and D	1
T	A and C and E	1
U	B and C and D	1
V	B and C and E	1
W	B and D and E	1
X	C and D and E	1
Y	A and B and C and D	1
Z	A and B and C and E	1
A1	A and B and D and E	1
B1	A and C and D and E	1
C1	B and C and D and E	1
D1	A and B and C and D and E	1

Failure mode: Check not performed (Decision based)

This failure mode covers the situation when the person decides not to perform the checking task. It has 2 branch points. The first “Working culture and expectation” aims to capture the aspect of safety culture in form of knowing if the task is important or not or bias due to daily working performance. The second, “Distractions/ interruptions and excessive workload” only focusses on the important simultaneous work being done that can make the personnel decide in not doing the task.

This DT was not quantified and its quantification cannot be used from other DTs

Branch point	Negative conditions	Expert judgment data	Median HEP
Working culture and expectations	Person not aware of importance of the check or importance of check to be performed completely and not partially	-	-
	The check was performed recently and you trust it was performed correctly	-	-
	The task to produce the output is simple - no error is expected	-	-
	Expecting no error as the check is performed by the same person doing the initial task	-	-
Distractions/ interruptions and excessive workload	Simultaneously doing another urgent task	-	-

Median HEP of the branch points

#	Branch point	Median HEP
A	Working culture and expectation	-
B	Distractions/ interruptions and excessive workload	-
C	A and B	-

Failure mode: Execute desired action incorrectly

This failure mode and its branch points are similar to the one in GTT "Simple interaction with software or tool".

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Indicator unclear	The indications for the input boxes on the screen are not readable	L M L H L	0.0011
	Too much information on the software screen leading to confusion	M M L L M H	0.0094
	The ordering of the values on the control document and on the screen do not match (i.e. X,Y,Z on screen and Z,X,Y in the document)	M H M M L L L M H	0.0097
Distractions/ interruptions and	Simultaneously doing another urgent task	M H L M L M	0.0094
	Interruptions from colleagues while doing	L M L L M M	0.0024

excessive workload	the task		
	There is less time to do the task (excessive workload)	L M L M M	0.0088
	Simultaneously tracking/supervising another task	L M L L L M	0.001
Low vigilance due to expecting no error	Task is mechanical and repetitive	L M L L M	0.0011
Environmental Distraction	The background noise level is too high to distract the focus	L L L L M	0.001

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0018
B	Distractions/ interruptions and excessive workload	0.0046
C	Low vigilance due to expecting no error	0.0011
D	Environmental distraction	0.001
E	A and B	0.0046
F	A and C	0.0018
G	A and D	0.0018
H	B and C	0.0046
I	B and D	0.0046
J	C and D	0.0011
K	A and B and C	0.01
L	A and B and D	0.01
M	A and C and D	0.01
N	B and C and D	0.01
O	A and B and C and D	1

Failure mode: Failure to execute desired action

It should be noted that the branch points and their respective negative conditions in this DT are similar to the fourth failure mode. Thus, the values from the fourth failure mode, can be used.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Distractions/ interruptions and	Simultaneously doing another urgent task	-	-
	Interruptions from colleagues while doing	-	-

excessive workload	the task		
	There is less time to do the task (excessive workload)	-	-
	Simultaneously tracking/supervising another task	-	-
Environmental Distraction	The background noise level is too high to distract the focus	-	-

Median HEP of the branch points

#	Branch point	Median HEP
A	Distractions/ interruptions and excessive workload	-
B	Environmental distraction	-
C	A and B	-

Failure mode: Coordination failure

This failure mode models the failure in coordination when two ore more people are interacting while performing a check. It is affected by three branch points. The first "Information unclear" aims to capture the aspects of unclear language, unclear jargon used, information without background etc. The second "External and environmental distractions" aims to capture the environmental and bad phone reception leading to confusion in coordination. The last "Cognitive overload" aims to capture the aspects of too much information given verbally, the personnel are involved in another task etc.

Branch point	Negative conditions	Expert judgment data	Median HEP
Information unclear	Language barrier between personnel	M H L M H H H	0.095
	Jargon or technical lingo used not familiar	M M L M H H L	0.01
	Information communicated not clear	H L L M H H H	0.0968
	Not complete information on the background	M L M H	0.01
External and environmental distractions	Too much noise in the room	M M L M M M	0.01
	Bad phone reception	M H L L M	0.0061
Cognitive	Too much information given	M M L	0.0089

overload	Caller is simultaneously occupied with another work	M L L M M H	0.0094
	Person receiving the data is busy with other work	H L L H	0.0046
	The information receiver is busy with monitoring other task or a patient	H M L H M H	0.0771
	High volume of information is communicated without possibility to write	H M L H H L	0.0859

Median HEP of the branch points

#	Branch point	Median HEP
A	Information unclear	0.0176
B	Cognitive overload	0.032
C	External and environmental distractions	0.0083
D	A and B	0.1
E	A and C	0.1
F	B and C	0.1
G	A and B and C	1

GTT 3: Identification of patient or patient related items

Failure mode: Patient information incorrectly matched

This failure mode models the incorrect matching of patient ID/information to identify the patient or patient's item. It has 3 branch points. The first "Indicator unclear" aims to capture the aspects of clarity and unambiguity of the indicators/patient id. The second "Low vigilance due to expecting no error" aims to capture the bias or the dependency aspects in routine checking of the items or patients. The last "Distractions/ interruptions and excessive workload" aims to capture the aspect of attention loss due to environment, colleague's interruptions or due to lack of time.

Branch point	Negative conditions	Expert judgment data	Median HEP
Indictor unclear	The values written look alike	L M M L L	0.0011
	The value on the label or file not easily readable	L L L L L	0.001
	No Id on the support	H M L	0.0098
Low vigilance due to expecting	The device looks the same	L M L L L	0.001
	Expecting no error as the task to leave the file was self-performed	L M M M -	0.0098
	The file or device was left here yesterday or last	M M M L M	0.01

no error	time		
	Task done recently by another personnel	M M L - -	0.0089
	Items were put by another personnel and if one ID matches, all the other IDs must match	M M - - M	0.01
Distractions/ interruptions and excessive workload	Level of background noise is too high to distract the focus	L L M L L	0.001
	Phone call asking patient ID of another patient	L M L M	0.0026
	Interruptions from colleagues while doing the task	M L M H L	0.0061
	Not enough time (lack of time)	M H L M M M	0.01
	Simultaneous work assigned	M M M H M	0.01

Median HEP of the branch points

#	Branch point	Median HEP
A	Indicator unclear	0.0015
B	Distractions/ interruptions and excessive workload	0.0057
C	Low vigilance due to expecting no error	0.0077
D	A and B	0.01
E	A and C	0.01
F	B and C	0.1
G	A and B and C	0.1

Failure mode: Identification check not performed (decision based)

This failure mode and its branch points are same to the one in GTT "Quality check" #4 failure mode. The information can be directly taken from there.

Branch point	Negative conditions	Expert judgment data	Median HEP
Working culture and expectations	Person not aware of importance of the check or importance of check to be performed completely and not partially	-	-
	The check was performed recently and you trust it was performed correctly	-	-
	The task to produce the output is simple - no error is expected	-	-
	Expecting no error as the check is performed by the same person doing the initial task	-	-
Distractions/	Simultaneously doing another urgent task	-	-

interruptions and excessive workload			

Median HEP of the branch points

#	Branch point	Median HEP
A	Working culture and expectation	-
B	Distractions/ interruptions and excessive workload	-
C	A and B	-

Failure mode: Failure to execute desired action

It should be noted that the branch points and their respective negative conditions in this DT are similar to the fourth failure mode of the GTT Quality Check. Thus, the values from the fourth failure mode, can be used.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Distractions/ interruptions and excessive workload	Simultaneously doing another urgent task	M H L M L M	0.0094
	Interruptions from colleagues while doing the task	L M L L M M	0.0024
	There is less time to do the task (excessive workload)	L M L M M	0.0088
	Simultaneously tracking/supervising another task	L M L L L M	0.001
Environmental Distraction	The background noise level is too high to distract the focus	L L L L M	0.001

Median HEP of the branch points

#	Branch point	Median HEP
A	Distractions/ interruptions and excessive workload	0.0046
B	Environmental distraction	0.001
C	A and B	0.0046

GTT 4: Complex interaction with software or tool

Failure mode: Misinterpretation of data

This failure mode models the understanding failure in performing a task that requires some level of cognitive effort. This has 5 branch points. The first "Unclear information" focusses on capturing the readability/ambiguity and the presentation of the information on the interface. The second, "Tumor complexity" aims to capture tumor difficulties like location, proximity, size, movement etc. that impact the understanding of the issue. The third, "Lack of training or experience" aims to capture situations where the personnel is not familiar with the case or has no training or experience in handling special tumor locations. The fourth "Time pressure" aims to capture lack of time to perform the task which could be due to workload or due to the late arrival of the relevant information. The last "Resource unavailable" aims to capture the aspect of incomplete information available to the understand the whole situation.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Information unclear	Full scan is not visible on the screen (needs to be scrolled)	M M L	0.0089
	Necessary information on multiple screens	M M M	0.01
	Bad quality of imaging	H H E	0.103
	Scans on the interface not readable or clear	E H H	0.103
Tumor Complexity	Tumor present where it can move	L M L	0.0011
	Tumor located where air or water cavity can change over time	M L L	0.0011
	Multiple tumor physically close to each other but not joined	M L L	0.0011
	Tumor located next to multiple critical organs	M L M	0.0089
	Multiple artefacts present	L M L E	0.0012
Lack of training or experience	Lack of familiarity with the case	L H L	0.0011
	Lack of training or experience on treating special tumor locations	H E H	0.103
Time pressure	Time pressure due to information not coming in time	M L H	0.0098
	Time pressure due to workload	H L H	0.0897
Resource unavailable	Some information incomplete	H L M	0.0098

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0271
B	Tumor complexity	0.0014
C	Lack of training or experience	0.0642
D	Time pressure	0.0451
E	Resource unavailable	0.0098
F	A and B	0.0271
G	A and C	0.1
H	A and D	0.1
I	A and E	0.1
J	B and C	0.0642
K	B and D	0.0451
L	B and E	0.01
M	C and D	0.1
N	C and E	0.1
O	D and E	0.1
P	A and B and C	0.1
Q	A and B and D	0.1
R	A and B and E	0.1
S	A and C and D	1
T	A and C and E	1
U	B and C and D	1
V	B and C and E	1
W	B and D and E	1
X	C and D and E	1
Y	A and B and C and D	1
Z	A and B and C and E	1
A1	A and B and D and E	1
B1	A and C and D and E	1
C1	B and C and D and E	1
D1	A and B and C and D and E	1

Failure mode: Execute desired action incorrectly

This has 4 Branch points. The first branch point “Information unclear” aims to capture the inadequate aspects of the machine interface, like indication boxes, proximity of input boxes and logic of interface, that can lead to simple execution failure. The second “Distractions/ interruptions and excessive workload” focusses on capturing attention failure due to workload or personnel interruptions during the work. The third “Low vigilance due to expecting no error” aims to capture the aspect of monotonous nature of the task or monotonous nature of the

information itself. The last "Environmental distraction" aims to capture the aspect of attention loss but now due to the surrounding environment. The conditions falling in each branch point are given below.

The data for this DT has been taken from Simple interaction with software/tool- executed desired action incorrectly DT.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Information unclear	The indications for the input boxes on the screen are not readable	L M L H L	0.0011
	The layout of the input interface is unusual or confusing	L M L L L	0.001
	Touch screen or input keyboard is very sensitive (unknown mis-selection is possible)	L L L M M	0.0011
	Mental calculation required due to ineffective logic of interface	M L - H -	0.0098
	Too much information on the screen leading to confusion	M M - H M H	0.0112
Distractions/ interruptions and excessive workload	Simultaneously doing another urgent task	M H L M L M	0.0094
	Interruptions from colleagues while doing the task	L M L L M M	0.0024
	There is less time to do the task (excessive workload)	L M L M M	0.0088
	Simultaneously tracking/supervising another task	L M L L L M	0.001
Low vigilance due to expecting no error	Task is mechanical and repetitive	L M L L M	0.0011
Environmental Distraction	The background noise level is too high to distract the focus	L L L L M	0.001

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0018
B	Distractions/ interruptions and excessive workload	0.0046
C	Low vigilance due to expecting no error	0.0011

D	Environmental distraction	0.001
E	A and B	0.0046
F	A and C	0.0018
G	A and D	0.0018
H	B and C	0.0046
I	B and D	0.0046
J	C and D	0.001
K	A and B and C	0.01
L	A and B and D	0.01
M	A and C and D	0.01
N	B and C and D	0.01
O	A and B and C and D	1

Failure mode: Mis-match or inconsistency not recognized

This failure mode models the noticing of discrepancies when some data is missing or noticing that incomplete task has been performed. The branch points are same as in "Quality check" GTT #1 failure mode and thus the information can be taken from there.

Branch point	Negative conditions	Expert judgment data	Median HEP
Information unclear	The values on the interface are not easily readable	L M M L M M	0.0088
	The values look alike	L H M L H M L	0.0016
	Too much information on the software screen leading to confusion	M M L L M H	0.0094
	The ordering of the values on the control document and on the screen do not match (i.e. X,Y,Z on screen and Z,X,Y in the document)	M H M M L L L M H	0.0097
Low vigilance due to expecting no error	The task to produce the output is known - no error is expected	M L M M H L M	0.0099
Distractions/ interruptions and excessive workload	Overloaded with other work	M L M L M M H	0.0099
	There is little time to do the task	M M M L H M H	0.01
	The background noise level is too high, it distracts the focus	L M L L L M M H	0.0012
	Interruptions from colleagues while doing the task	L M M L L H M H	0.0078

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0086
B	Distractions/ interruptions and excessive workload	0.0099
C	Low vigilance due to expecting no error	0.0091
D	A and B	0.1
E	A and C	0.1
F	B and C	0.1
G	A and B and C	1

GTT 5: Iterative determination of optimum parameters

Failure mode: Misinterpretation of information

This failure mode models the misunderstanding of the data or information to assess the situation at hand. It has 4 branch points, but does not include time pressure as tasks in this GTT are mostly performed in adequate time. The first aims to capture the confusion or misunderstanding caused by the inadequacy of the interface, it includes the readability issues, ambiguity of presented information etc. The second “Tumor complexity” aims to capture the aspects of complexity of the tumor location, size, proximity to multiple critical organs and other complexities that can impact the understanding of the problem at hand. The third “lack of training or experience” aims to capture the aspects of training or experience issues like training not covering special tumor situations or not experience in handling specific tumor situations. The last, “Resourse or software unavailable” aims to capture the aspects of availability of software, procedures etc. to assist in solving the problem at hand.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Information unclear	Interface does not display all required information (mental effort required)	M M M H H M	0.01
	Information for special tumors difficult to locate on the screens	M L L M L H H	0.0016
	Critical information on the screen is not readable	M H M M L H L	0.01
	Interpreting of information due to unavailability	L M M L	0.0026

	of straightforward information (No 3D view)		
Tumor Complexity	Tumor located to multiple critical organs	M L L L M M L	0.0011
	Tumor located where air or water cavity can change over time	M M L L M M H	0.01
	Non-usual way to calculate the dose distribution	M M L M M L H	0.01
	Physical changes occur frequently	H M L L L M H L	0.0011
	Tumor present where it can move	H L L L L M H M	0.0011
	Manual conversions required	L H M M M L M H	0.01
Lack of training or experience	Training program does not cover information on specific constraints for special tumor locations	M M H H M H M H	0.0303
	No training or experience on treating special tumor locations	M H H H M H M H	0.0965
	New software without training (no training for new software)	M M H H L M E	0.0112
	No training on proton physics given to a new comer who does not have a background in it	M L M H L	0.0061
	No in depth training on generic parameters of TPS for new comers	E L M M H	0.01
Resource and software unavailable	Procedure not available	M M M H H M	0.01
	Information not available (metal information)	M H M H E	0.0628
	Specific information required for planning not available	M H H M H M H	0.0873
	Update on software for rarely used software not known	H L L M M L	0.0013

Median HEP of the branch points

#	Branch point	Median HEP
A	Interface unclear	0.0068
B	Tumor complexity	0.0038
C	Lack of training or experience	0.015
D	Resource and software unavailable	0.0146
E	A and B	0.01
F	A and C	0.1
G	A and D	0.1
H	B and C	0.015
I	B and D	0.0146
J	C and D	0.1
K	A and B and C	0.1

L	A and B and D	0.1
M	A and C and D	1
N	B and C and D	0.1
O	A and B and C and D	1

Failure mode: In-appropriate decision on strategy selection

This failure mode models the failure in deciding the correct strategy to solve the problem. It has 4 branch points. The first "Tumor complexity" aims at capturing the aspects of difficulty of the task or the adequacy of the training provided which can affect the decision making while selecting a strategy for the tumor. The second "Lack of training or experience" aims to capture the aspect of guidance not provided in training program for selecting strategies for special tumors. The third "Resource and software unavailable" aims to capture the aspects of availability of resources like procedures, additional software/information to aid in making a decision. The last "time pressure" aims to capture the aspects of lack of time in making the decision.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Tumor Complexity	Tumor located to multiple critical organs	-	-
	Tumor located where air or water cavity can change over time	-	-
	Physical changes occur frequently	-	-
	Tumor present where it can move	-	-
Lack of training or experience	Training program does not cover guidance on selecting specific methodology for special tumor situation	-	-
	No training or experience on treating special tumor locations	-	-
Resource and software unavailable	Procedure not available	-	-
	Information not available (metal information)	-	-
	Specific information required for planning not available	-	-
	Update on software for rarely used software not known	-	-
Time pressure	There is very less time to take a decision	-	-

Median HEP of the branch points

#	Branch point	Median HEP
A	Tumor Complexity	-
B	Lack of training or experience	-
C	Resource and software unavailable	-
D	Time pressure	-
E	A and B	-
F	A and C	-
G	A and D	-
H	B and C	-
I	B and D	-
J	C and D	-
K	A and B and C	-
L	A and B and D	-
M	A and C and D	-
N	B and C and D	-
O	A and B and C and D	-

GTT 6: Verbal Communication

Failure mode: Communication failure

The failure mode models the communication failure either at the originator or receiver's end. It has 3 branch points. The first "Information unclear- verbal" aims to capture the aspects of clarity or quality of information provided for example language issues, jargon usages, incomplete background given etc. The second "External and environmental distraction" aims to capture the background noise or the noise in the medium of communication that can affect the transfer of information. The last "Cognitive overload" aims to capture the aspect of amount of information that is communicated verbally to study the cognitive workload in the communication tasks.

Branch point	Negative conditions	Expert judgment data	Median HEP
Information content unclear-verbal	Language barrier between personnel	M H L M H H H	0.095
	Jargon or technical lingo used not familiar	M M L M H H L	0.01
	Information communicated not clear	H L L M H H H	0.0968
	Not complete information on the background	M L M H	0.01
External and	Too much noise in the room	M M L M M M	0.01

environmental distractions	Bad phone reception	M H L L M	0.0061
Cognitive overload	Too much information	M M L	0.0089
	Caller is simultaneously occupied with another work	M L L M M H	0.0094
	Person receiving the data is busy with other work	H L L H	0.0046
	The information receiver is busy with monitoring other task or a patient	H M L H M H	0.0771
	High volume of information is communicated without possibility to write	H M L H H L	0.0859

Median HEP of the branch points

#	Branch point	Median HEP
A	Information content unclear- verbal	0.0176
B	Cognitive overload	0.032
C	External and environmental distractions	0.0083
D	A and B	0.1
E	A and C	0.1
F	B and C	0.1
G	A and B and C	1

Failure mode: Not communicated (decision based)

This failure mode models the decision based failure of not communicating some information. It has 2 branch points. The first "Lack of training or experience" aims to capture the aspects of lack of training or experience that do not inform the trainee what all key information has to be given in specific situations and in addition it covers the aspects of bias or expectation that the receiver would know some information already. The last "Time pressure" aims to capture the lack of time due to performing an urgent task simultaneously.

Branch point	Negative conditions	Expert Judgment Data	Median HEP
Lack of training or experience	Training does not cover or not experienced to know all key information is to be given for specific or special treatments/ tumors or new or adapted techniques	-	-
	The information simple and straight forward that the target person is expected to know	-	-
Time	There is less time as there is an urgent competing	-	-

pressure	task to be done with high priority		

Median HEP of the branch points

#	Branch point	Median HEP
A	Lack of training or experience	-
B	Time pressure	-
C	A and B	-

Appendix 9

Quantification of branching points for validation of the expert assessments

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
B1a	Problematic interface (ambiguous, unclear information, similar-looking numbers, names etc., confusing layout)	Quality Check-[Deviation from requirement not recognized], Complex interaction with software or tool-[Misinterpretation of data], Simple interaction with software or tool-[Execute desired action incorrectly], Iterative determination of optimum parameter-[Misinterpretation of information]	CARA (Poor, ambiguous system feedback- general adequacy of HMI): 5	A	As mentioned in CARA, the Error Producing Conditions (EPCs) model quality of feedback from interface. This includes poor display, unclear equipment/information etc.	A	The conditions listed in CARA's EPC (e.g. poor display, high false alarm rate, unclear equipment etc.) and the anchors associated with its "Maximum effect" capture the ergonomic issues being modelled by the branch point
			THERP (20-13, HEP for selection errors for locally operated valve, #5): 0.01	A	THERP table models HEP values for selection errors due to unclear, ambiguous indicators, and similarity issues like shape, size, state, and tags. These aspects match with the mentioned aspects in the branch point	A	The performance conditions underlying the THERP data point well match the branch point conditions

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
B1b	Problematic indicators (ambiguous, unclear indicators, similar-looking numbers and names, etc.)	Identification of patient or patient related items- [Patient information incorrectly matched]	CARA (Poor, ambiguous system feedback- general adequacy of HMI): 3	A	As mentioned in CARA the EPC models quality of feedback from interface. This includes poor display, unclear equipment/information etc.	A	In CARA's EPC's "0.5 effect" well matches the conditions modelled by this branching point: font size too small, use of complex symbology patterns. Note this branch point (B1b) focuses on indicator issues, as opposed to B1a which addresses issues concerning the whole interface layout (hence the reduced effect is applied to B1b)
B1c	Information content clarity (ambiguous, unclear, redundant, language problems, use of jargons)	Verbal communication- [Communication failure]	CARA (Communications quality): 3	A	As described in CARA the EPC models quality of the speech, which includes language problems, clipped speech, situation not completely conveyed, complexity of air traffic control transmission etc. This all matches with the conditions in the branch point.	A	According to CARA the "Maximum effect" models interference sufficient enough to make communication comprehension unreliable. This can be interpreted as the impact being modelled by the given branch point.
B2	Low vigilance (expectation bias due to either the task being very simple and mostly error-free or due to general error-free performance	Quality Check- [Deviation from requirement not recognized], Simple interaction with software or tool- [Execute desired action incorrectly],	CARA (Low vigilance): 3	A	In CARA, one of the conditions for this EPC is "suffering from complacency". This is the condition in the EPC that closely models the situation of expectation bias in high reliability condition, which is being modelled in the branch condition.	A	The conditions underlying the "0.5 effect" of the CARA's EPC: (Long time on position, after sustained busy period followed by medium or low traffic) well capture the strength of complacency being modelled by the branch point (i.e. long time on the same task or simple task that generally has no error).

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
	by other personnel)	Identification of patient or patient related items- [Patient information incorrectly matched]	THERP (20-22, estimated probabilities that a checker will fail to detect errors made by others, #4): 0.01	A	The THERP table is applicable to failure in checking errors in normal operation. The item #4 models the checking task that involves active participation. Thus, it can be interpreted that it models dependency between checkers when active participation is there, which is being modelled here.	P	Given the very different fields of application, it is very hard to judge how different is the strength of expectations in the radiotherapy domain compared with the conditions underlying the THERP data point. Thus, the score of P is given.
B3	Task complexity (Complexity due to the tumour's size, shape, location, additional constraints like metal implants, air gaps changing over time etc.)	Complex interaction with software or tool- [Misinterpretation of data], Iterative determination of optimum parameter- [Misinterpretation of information]	CARA (Traffic complexity): 5.5	P	EPC in CARA covers generically the difficulties when a complex range of inputs are present to be identified and understood. The context is specific to Air Traffic Management; it includes complex dynamics of the traffic, possibly changing over time. This dynamic aspect is not present in the radiotherapy domain: this is why full applicability A score is not given.	A	The categories (excessive traffic load, complex traffic mix etc.) mentioned in CARA, and the underlying problem solving conditions in the "0.5 effect" anchor (non-routine traffic, non-routine conflicts requiring constrained solutions, conflict resolution leading to a secondary conflicts) well match with the strength of the complexity of the branch point i.e. complexity due to non-routine tumor or constraints.

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
B4	Lack of training or experience (lack of training or experience for specific tasks or for specific requirements for a treatment technique or for a specific tumour)	Complex interaction with software or tool- [Misinterpretation of data], Iterative determination of optimum parameter- [Misinterpretation of information]	CARA (Unfamiliarity and adequacy of training): 10.5	A	The EPC models both unfamiliarity (of a potentially important and infrequent situation) and adequacy of training. It includes experience and knowledge on the specific tumor position, unfamiliar task in routine operations, novel situations, emergency training etc.	A	As explained in CARA the "Maximum effect" covers rare event/scenario never seen in training or experience. The medical personnel have some background or are trained once to deal with such situations, the "0.5 effect" (rare event covered once in training) covers this aspect and is thus more applicable.
B5	Resource or software unavailable (unavailability of documents, software, and information for the required task leading to increase in cognitive workload)	Complex interaction with software or tool- [Misinterpretation of data], Iterative determination of optimum parameter- [Misinterpretation of information]	CARA (unavailable equipment/degraded): 5.5	A	The example provided for the CARA's EPC (information on flight plan not available) captures the conditions in the branch point, which cover information unavailability and software unavailability.	A	The "0.5 effect" anchor of CARA's EPC covers situations in which due to loss of information the operator has to call on skills which are no longer fluent. The description captures the impact that the branch point is aiming at i.e. activation of cognitive effort when resource is not available, which is generally not needed.

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
B6	Time pressure (could be due to workload or any pressure to complete the task quickly)	Complex interaction with software or tool- [Misinterpretation of data]	CARA (Time pressure due to inadequate time to complete the task): 5.5	P	According to CARA the EPC models inadequate time to complete the task. The example provided covers time pressure during emergency and time pressure due to late identification of a conflict and requirement to recover. The conditions in the branch point cover time pressure due to workload and late arrival of information (not emergency). Thus, the match is partial.	A	The time pressure being modelled in the branch conditions relates to radiotherapy personnel working at fast pace to finish the task with little checking. The description of "0.5 effect" of the CARA's EPC (working at a fast pace with little time for checking) is capturing this aspect.
B7a	Distractions/ interruptions and excessive workload (cognitive overload or repeated interruptions during a task)	Quality Check- [Deviation from requirement not recognized], Identification of patient or patient related items- [Patient information incorrectly matched], Verbal communication-	CARA (cognitive overload due to simultaneous presentation of non-redundant information): 6	P	As mentioned in CARA, the EPC models high/excessive workload and distraction/ interruptions. However, the distractions are only due to task switching or personnel interference but do not cover environment-related distraction. Thus, a rating of P is given.	P	An emergency event, which is mentioned in the description of the "Maximum effect" of the EPC in CARA, is not expected in radiotherapy, but the conditions described in the 'full effect' anchor are the very similar to those covered in the branching point: "multiple demands from aircraft via phone, alarms and other controllers simultaneously", representing a possibly noisy, high workload hospital environment. A rating of P is given.

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
		[Communication failure]	HEART (poor environment): 2	P	The HEART EPC and the studies underlying it indicate that they cover only the aspect of environmental noise. Since it does not cover distraction due to interruptions, the rating of P is given.	A	The HEART EPC description indicates that really harsh environment is not being studied. It can be interpreted that "poor environment" is capturing degraded working environment that can cause distractions, closer to the situations covered by the branch point.
			THERP (20-16, modifiers for the effects of stress, #4 (b)): 2	P	Among the different stress levels addressed by THERP, the one closest to the branch point of interest is the one of moderately high stress. This focusses on the stress induced by heavy workload and not environment and interruptions (at least not explicitly). Thus, the rating of P is given.	P	Given the partial coverage of the THERP multiplier (heavy task load without effect of environment and interruptions), the rating of P is given.

#	Branch points	GTT- [Failure Mode]	Relevant data[1]: multiplier/ HEP[2]	Criterion 1: Similarity of manifestation		Criterion 2: Strength of impact	
				Score	Comment	Score	Comment
B7b	High workload and task switching (distractions during the task due to high workload, performing simultaneous tasks and interruption from colleagues)	Simple interaction with software or tool- [Execute desired action incorrectly]	CARA (cognitive overload due to simultaneous presentation of non-redundant information): 3.5	A	As mentioned in CARA, the EPC models high/excessive workload and distraction/ interruptions.	A	The "0.5 effect" of the EPC can be given a rating of A as there the description of the condition "Information required to be assimilated from sources but prioritization is possible" (e.g. dealing with complex sector management issues while also discussing other issues via phone)." matches with the description of the branch point modelling of high workload, task switching and distractions due to personnel. The "Maximum effect" is not applicable as it focusses on excessive workload and distractions and for the given task the branch point is not modelling that.
B7c	Environmental distractions (Distractions due to noisy environment)		HEART (poor environment): 2	P	The HEART EPC and the studies underlying it, indicate that they cover only the aspect of environmental noise. It does not cover distraction due to personnel, so a rating of P is given.	A	The HEART EPC description indicates that really harsh environment is not being studied. It can be interpreted that "poor environment" is capturing degraded working environment that can cause distractions.

1- Relevant data: the data that is associated from the HRA methods to the given branch point. For the cases when the multiplier is associated then it can be either the "maximum effect" from the HRA method or the associate proportion of the maximum effect. Different levels or proportion of effect correspond to different strengths of the impacting conditions. Maximum effect corresponds to the maximum impact the condition can make on the task failure.
2- Two sources of data were evaluated, which are multiplier values from the CARA method and absolute HEPs given the performance condition from THERP. For multipliers from CARA, the HEP was obtained by multiplying the value with the base HEP.

Appendix 10

This appendix reports the values of the HEPs used for the convergence validation analysis.

GTT-Failure Mode		Nominal HEP	Branch points, [Multiplier value]= HEP: Information unclear	Low vigilance due to expectation	Tumor complexity	Lack of training or experience	Resource or software issues	Time pressure	Distraction/interruptions and excessive workload	Environmental distraction
Simple interaction with software or tool-[Execute desired action incorrectly]	CARA	0.002[1]	[5[5]]= 0.01	[3[8]]= 0.006					[1.5[13]]= 0.003	[2[13]]= 0.004
	THERP	0.001[1]	0.01[5]	0.01[8]					[2[14]]= 0.002	[2[13]]= 0.002
	EE[15]	-	0.0018	0.0011					0.0046	0.001
Quality Check-[Deviation from requirement not recognized]	CARA	0.005[2]	[5[5]]= 0.025	[3[8]]= 0.015					[6x2[14]]= 0.06	
	THERP	0.001[2]	0.01[5]	0.01[8]					[2[14]]= 0.002	
	EE	-	0.0068	0.0091					0.0073	
Iterative determination of optimum parameter-[Misinterpretation of information]	CARA	0.003[3]	[5[5]]= 0.015		[5.5[9]]= 0.0165	[10.5[10]]= 0.0315	[5.5[11]]= 0.0165			
	THERP	-								
	EE	-	0.0068		0.0038	0.015	0.0146			
Identification of patient or patient related items-[Patient information incorrectly matched]	CARA	0.005[2]	[3[6]]= 0.015	[3[8]]= 0.015					[6x2[14]]= 0.06	
	THERP	0.001[2]	0.01[5]	0.01[8]					[2[14]]= 0.002	
	EE	-	0.0015	0.0077					0.0057	
Complex interaction with software or tool-[Misinterpretation of data]	CARA	0.003[3]	[5[5]]= 0.015		[5.5[9]]= 0.0165	[10.5[10]]= 0.0315	[5.5[11]]= 0.0165	[5.5[12]]= 0.0165		
	THERP	-								
	EE	-	0.0271		0.0014	0.0642	0.0098	0.0451		
Verbal communication-[Communication failure]	CARA	0.002[4]	[3[7]]= 0.006						[6x2[14]]= 0.024	[3[7]]= 0.006
	THERP	-								
	EE	-	0.0176						0.032	0.0083

1- #2 in Table 14
2- #1 in Table 14
3- #3 in Table 14
4- #4 in Table 7
5- B#1a in Appendix 9
6- B#1b in Appendix 9
7- B#1c in Appendix 9
8- B#2 in Appendix 9
9- B#3 in Appendix 9
10- B#4 in Appendix 9
11- B#5 in Appendix 9
12- B#6 in Appendix 9
13- B#7b in Appendix 9
14- B#7a in Appendix 9
15- EE: Expert Elicitation

Curriculum Vitae

Feb 2018 – Till date	GxP Consulting AG, Swtizerland. Human Factors Engineer at Novartis, Basel.
Feb 2014 – Jan 2018	Paul Scherrer Institut, Switzerland. PhD studies at the Risk and Human Reliability Group, Laboratory of Energy Systems Analysis, enrolled at Department of Mechanical and Process Engineering ETH Zurich.
Nov 2012 – Jan 2014	Indian Institute of Technology Delhi, India. Scientific Assistant, Photovoltaics Laboratory.
Aug 2012 – Oct 2012	Paul Scherrer Institut, Switzerland Research assistant, Risk and Human Reliability Group, Laboratory of Energy Systems Analysis.
Sept 2010 – July 2012	EPFL- ETHZ, Switzerland. Masters in Nuclear Engineering.
July 2006 – July 2009	Sri Venkateswara College, Delhi University, India. Bachelors in Physics (Hons).