SYSTEMS THEORETIC HAZARD ANALYSIS (STPA) APPLIED TO THE RISK REVIEW OF COMPLEX SYSTEMS: AN EXAMPLE FROM THE MEDICAL DEVICE INDUSTRY

by
Blandine Antoine

M. Sc. Nuclear Engineering, University of California Berkeley, 2005
Dipl. Ing. Ecole Polytechnique, 2006
M.P.A. Ecole Nationale des Ponts et Chaussées, 2007

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 2013

Signature of Author…….
Engineering Systems Division,
October 17th, 2012

Certified by……..
Prof. Nancy Leveson
Professor of Engineering Systems and Aeronautics and Astronautics
Thesis Committee Chair

Certified by……..
Prof. Olivier de Weck
Associate Professor of Aeronautics and Astronautics and Engineering Systems
Thesis Committee Member

Certified by……..
Prof. Joseph Sussman
JR East Professor of Civil and Environmental Engineering and Engineering Systems
Thesis Committee Member

Certified by……..…
Dr. Christian Hilbes
Lecturer at the School of Engineering of the Zurich University of Applied Sciences (ZHAW)
Thesis Committee Member

Accepted by…….....
Prof. Olivier de Weck
Associate Professor of Aeronautics and Astronautics and Engineering Systems
Chair, Engineering Systems Division Education Committee

PAGE INTENTIONALLY LEFT BLANK

To Christophe and our children

Blessed be the light,
your smiles,
and our learning journeys.

PAGE INTENTIONALLY LEFT BLANK

Systems Theoretic Hazard Analysis (STPA) Applied to the Risk Review of Complex Systems: an Example from the Medical Device Industry

by
Blandine Antoine

Submitted to the Engineering Systems Division on December 14th, 2012,
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

Traditional methods to identify and document hazards, and the corresponding safety constraints, are lacking in their ability to account for human, software and sub-system interactions in highly technical systems. STAMP, a systems-theoretic accident causality model, was created to overcome these limitations.

The application of STAMP hazard analysis method STPA to five sub-systems of the Paul Scherrer Institute's experimental PROSCAN proton therapy system demonstrated how STPA can augment design and risk review of existing complex systems. Two of the five human controllers active in treatment delivery, two of the four process attributes controlled by the PROSCAN facility, and one of the four control loops that control the beam to target alignment attribute were analyzed. In doing so, the following contributions were made:

- Analyzed the regulations currently in place in the US and Europe for the marketing of external beam radiotherapy devices and, more generally, medical devices that do not contain radioactive materials, concluding that STPA would be acceptable in both regulatory systems;
- Provided experience in applying STPA to a complex device. Information on efficacy was derived by comparing STPA results with an existing safety assessment but a more formal counterpart is needed for stronger evidence. Information on learnability and usability was obtained when an informal workshop showed that system designers, in the course of one day, could be taught to use STPA to push their thinking about yet to be designed system elements;
- Demonstrated the applicability of STPA to an experimental radiotherapy facility and, through this feasibility check, potentially influenced the state of the art in hazard analysis of medical devices and health care delivery;
- Advanced the STPA methodology by creating notations and a process to document, query and visualize the possibly large number of hazardous scenarios identified by STPA analyses, with the goal of facilitating their review and use by their intended audience;

Showed how STPA is complementary to more traditional hazard analysis techniques such as fault and event trees. Their respective strengths can be summoned when STPA is used to identify areas on which to focus the investigation lens of traditional hazard analysis techniques.

Keywords: STAMP, STPA, hazard analysis, risk analysis, risk management, proton therapy, medical devices, safety, certification

PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

"Travaillez, prenez de la peine: C'est le fonds qui manque le moins.
[...]Un trésor est caché dedans."

"Work hard, sweat all you can: Riches is what counts the least.
[...]A treasure is hidden in [the field]. "

Jean de la Fontaine, 1668 - *Le Laboureur et ses Enfants*

My journey at MIT's Engineering Systems Division has been a humbling experience, a quest whose treasure, as in the ploughman's fable, was not the one I had set off to seek. *γνῶθι σεαυτόν* truly is a difficult task! And one that I now understand to still be far from completion.

My recovery I owe to Professor Nancy Leveson. She offered me a shelter when I wanted to quit it all, confidence to undertake a new project, incredibly prompt comments on all thoughts submitted to her review, frequent refocusing ordeals, a desk with a window (precious!) and the opportunity to work with a magnificent team of colleagues.

I am immensely grateful to Professor Olivier de Weck for the generous attention and respectful consideration he has for people in general, and students in particular. I thank Professor Joseph Sussman and him for their faith in this project, for their patience and the wisdom of their guidance, for being the best committee members one could dream of. Dr. Christian Hilbes, from the School of Engineering of the Zurich University of Applied Sciences (ZHAW) initiated and coordinated the research project on which this dissertation builds. He is gladly credited for having pushed my work to higher levels of rigor thanks to his sharp comments, insatiable curiosity, and encyclopedic knowledge of the issues associated with certifying complex systems in the EU and investigating the safety issues they are associated with: thank you!

None of the results presented in this dissertation could have been obtained without the dedication of Paul Scherrer Institute's Martin Rejzek. For his help, his friendliness, and his availability to work with me on the analysis of the Gantry-2 user area, I am truly grateful. I also warmly thank Dr. Martin Grossman for the support he has provided to this research project, arranging for its funding by the Paul Scherrer Institute, providing detailed logistical support, organizing for the PROSCAN design and operation teams to provide input to our analysis, and being such a pleasant person to interact with. I gratefully acknowledge funding from the Paul Scherrer Institute, warm encouragements from PSI Vice-Director Dr. Martin Jermann, and very generous welcome from all the PROSCAN staff to this research.

To those who never failed to believe that this project could be honorably completed: Dr. Christophe Antoine who shouldered a large share of my household responsibilities during these five years, the Laurenty, Antoine, Moussallieh and Pointin families who never hesitated to send

thoughts and more across the Atlantic ocean as they understood that more help was needed than I wished to acknowledge, old and new friends in France and the USA, Dr. Franck Carré, Professor Jessika Trancik who also provided TA-ing opportunities that were providential in bridging funding gaps, Francisco Llemos, Eric Pité, Dr. David Opolon, Dr. Philippe Bonefoy, members of the Fondation Carnot, Dr. Ioannis Simaiakis, Dr. Claire Cizaire, Dr. Noemie Chocat, Mario Bernhart.

To those who took such great care of my family while I was working on this project: Mrs. Nilda Marquez, Mrs. Leonidas Rodriguez, as well as the teaching and management staff at the MIT Technology Children Center.

To the women who nurtured me, helped me grow and made me realize that, yes, there are different things to be learned from women than there are from men: Claire de Mazancourt, Dr. Lisa d'Ambrosio, Dr. Caroline Brun, Dr. Lynette Cheah, Dr. Karen Tapia-Ahumada, Jacqueline Donoghue, Erica Bates, Professor Hamsa Balakrishnan, visitors to the women restrooms who shared their motherly experiences when they would see me dutifully pumping for Clélia and Paul-Hector, and the Boston Team Handball's ladies, especially Dr. Sonja Hansen, Erin Kitzler, Anne Coulter and Annie Felix.

To the diverse, unique and beautiful CSRL team that has made coming to campus a joy every morning of the past two years, was always available for a quick (exception made of food safety and political topics!) chat or a longer theoretical debate, offered their intelligent minds as frequent ideas sounding boards and regularly shared fuel for thoughts and for chocolate-craving guts: John Thomas, Cody Fleming, John Helferich, Melissa Spencer, Ibrahim Khawaji, Takuto Ishimatsu, Dr. Qi Hommes, visitors to the CSRL lab and all CSRL affiliates that lived outside of building 33.

To the administrators of Corps des Ponts et Chaussées who generously agreed to me taking a leave of absence to study at MIT.

To all those who have always kindly offered me their time, help and advice along the way: Elizabeth Milnes (thank you Beth!), Erica Bates, Jennifer Kratochwill, Professor Richard de Neufville, Dr. John Reilly, Dr. Angelo Gurgel, Professor John Heywood, Professor Daniel Frey, Professor Michael Golay, Professor Mujid Kazimi, Professor Kenneth Oye, Professor George Apostolakis, Professor Mort Webster, Dr. Valerie Karplus, Dr. Lara Pierpoint, the Joint Program student group, Professor Zoe Szajnfarber, Dr. Danielle Wood, Paul Jones, Dr. Richard Lanza, Dr. Jacob Flanz.

Except for its errors and deficiencies, the completion of this project is your achievement as much as it is mine: thanks for all that you have given me!

Table of Contents

Table of Tables

Table of Figures

PAGE INTENTIONALLY LEFT BLANK

1 Introduction

"Among them stood up Kalchas, Thestor's son,
Far the best of the bird interpreters,
Who knew all things that were, the things to come and the things past,
Who guided into the land of Ilion the ships of the Achaians
Through that seercraft of his own that Phoibos Apollo gave him."

Homer - *Iliad, 1.68-72, approx. 8th century BC*

1.1 Motivation: improve safety as complex systems grow in number and in scale

Kalchas, "*who knew all things that were, the things to come, and the things past*", is long gone. In his absence, "*making predictions is hard, especially about the future*[1]". Yet, how can engineers, designers, creators of all trades release products in the hands of their fellows if they have no clue how these systems will behave after they have left their labs and factories? Short of being blessed with Kalchas' gifts, methods and tools are needed that can help designers ensure a safe fate to their inventions.

1.1.1 New risks in an increasingly automated world

As stressed in (Leveson 1995) and (Lutz, 2000), modern systems have achieved greater capabilities through growing reliance on increasingly capable software and digital systems. Short lifetime and low upgrade costs, versatility and flexibility enabled by the ease of their reconfiguration, replacement and reuse, compact size and continuous improvement in computational power explain the extent to which they have been integrated into modern systems, from large infrastructure projects (think GPS-enabled navigation tools or plans for a smarter

[1] This quote appears to be equally attributed to Mark Twain and Niels Bohr.

grid) to small intelligent devices (e.g. climate controllers in high energy efficiency buildings). By outperforming the computational ability of the human mind and ridding both routine and emergency decision making of situation-dependent cognitive biases and latencies[2], they have become increasingly attractive to system designers who wish to improve performance while maintaining high degrees of process reliability and predictability in increasingly complex systems.

Following a parallel trajectory, and as highlighted by Lutz (2000), an increasing number of safety-critical systems rely on software to achieve their purposes. Further, miniaturization and processing improvements have enabled the spread of safety-critical systems from nuclear and defense applications to domains as diverse as implantable medical devices, traffic control, smart vehicles, and interactive virtual environments. Future technological advances and consumer markets can be expected to produce more software dependent safety-critical applications, especially in the medical field.

While this software enabled increase in capability has been enthusiastically embraced by society, the means through which it has been realized and the change of performance scale that has accompanied it have given birth to new risks. The International Commission on Radiation Protection offers a concrete illustration to this trend (emphasis added): "*New types of accidents have been encountered due to: the complexity of the present treatment preparations; the increased sophistication of the whole treatment process (with an increasing number of steps and more people involved); the omnipresence of computers with frequent and regular upgrading of more and more complicated software; and the difficulty of regularly and correctly training all the physicians, physicists, dosimetrists, engineers, etc. involved in a busy radiotherapy unit*" (ICRP, 2009).

The emergence of these new and often high consequence risks are largely responsible for persisting, albeit somewhat schizophrenic, social defiance against technological endeavors. Perhaps best illustrated in Western philosophy by the myth of Prometheus and the Greeks'

[2]An example, from the field of radiation therapy: "*The most important feature related to the complexity and sophistication of 'new technologies' is the requirement for computer control. Computers are increasingly used at each stage of the process, from prescription to completion of the treatment*." (ICRP, 2009, p 26)

condemnation of the sin of hubris, enshrined as a legal principle in the French Constitution and other important legislative texts through the concept of precautionary principle, and recently resurfaced in post-Fukushima debates about the role of nuclear power in regional energy mixes, warnings against the risks created by technological progress and the power of the human mind are as old as human civilizations. Nonetheless, catastrophic accidents and irreversible losses caused by our search for higher speed, increased scale, improved performance, larger concentrations of activities and energy can't seem to be eliminated. The overdose of 145 radiotherapy patients at Rangueil (2006), loss of flight AF 447 (2009), the Deepwater Horizon oil spill (2010) and the Fukushima nuclear accident (2011) are only but a few among too many recent reminders that we are still in want of adequate constraints to control the risks created by our very own powers and appetites.

It indeed appears that digitally increased system capabilities often come at the cost of intellectual manageability by system designers and operators. This threat is well recognized. For example, in the field of radiation therapy, the International Commission on Radiation Protection notes (ICRP) in a report on accidents in the field of radiotherapy: "*New technologies are meant to bring substantial improvement to radiation therapy. However, this is often achieved with a considerable increase in complexity, which in turn brings opportunities for new types of human error and problems with equipment*." (ICRP, 2009). Not only do automation and the multiplication of process communication channels enabled by the integration of software in physical systems potentially increase the scale of accidental damage by making it possible to operate larger and more potent systems but they also challenge our ability to design and operate them safely.

The design of digitally intense systems indeed requires the collaboration of culturally distinct professions whose actions towards creating a whole are orchestrated through the establishment and communication of performance and design requirements. That requirements be necessary to coordinate designer contributions creates an avenue for design error resulting from incomplete or incorrect requirements capture. Further, digital automation, meant to alleviate the cognitive burden of human operators and to go beyond human performance limitations (e.g. speed in actuation or decision making), often creates detrimental distance between human operators and the systems that they are tasked with controlling as is well illustrated by the following ICRP

statement: "*As a result of the complexity of many newer treatment strategies [enabled by computer technology], 'common sense' and intuition may no longer be as effective a mechanism to perceive 'when something may be wrong' as it is with conventional radiation therapy*." (ICRP, 2009).

When digital sensors feed their inputs to signal processing equipment that enables computerized controllers to develop a perception of what state the system is in so that it may be displayed to the operator, direct sensory input from the system is lost to the operator. A discrepancy may thus build over time between his model for system evolution and the actual processes through which change occurs within the system. In the event that an abnormal situation prevents the salient display of critical system information (e.g. screen freeze, loss of sensor channel, inadequate ergonomics ...), adequate mitigation by the operator is made less likely by this cognitive distance and, perhaps more importantly, by the gap built over time between human abilities and those that are needed to effectively prevent the system from creating harm[3].

1.1.2 Identifying and analyzing hazards

Ensuring the safe[4] operation of complex systems in these conditions requires that both the powerful potentialities and the commensurate risks associated with increased reliance on digital systems be well-understood so that they can be adequately controlled. To control risks, one first needs to identify and analyze hazards[5].

Society aims to control technological risks by submitting the creation and introduction of new systems to regulatory processes and peer reviews that follow precise guidelines. Regulatory agencies can define procedures to be used at the design stage, or set performance bounds that the system will have to operate within to be eligible for market approval. Once design is completed, regulatory reviewers scrutinize how well the proposed system meets safety requirements and goals defined by public and private authorities (e.g. through the performance of car crashes,

[3] For example, it took only 3.5 minutes for Air France flight AF 447 to plunge 11 000 meters into the Atlantic, meaning that the crew would have had less than 2.5 minutes to correctly identify the situation they were in and implement the corrective actions needed to avoid disaster (Spiegel, 2011)

[4] This dissertation will follow (Leveson, 1995) in defining system safety as freedom from loss.

[5] A hazard is defined as a system **state or set of conditions** that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss) (Leveson, 1995)

flight tests, clinical trials). Then, if the technical system has been approved for release on the market, its potential users will have to be trained and officially qualified before they are authorized to use it (e.g. medical diploma delivery, pilot certification). Systems (the technical device, its users and their organization) in use are often required to implement regular maintenance reviews and quality assurance processes. Finally, inspections are meant to ensure the enforcement of these organizational prevention measures during the operational lifetime of the system, and licensing updates are required when specific changes are brought to the systems' technical or organizational features.

These external controls on system safety rely on preliminary hazard identification and the reviewer's assessment of how effective the mitigation measures present in the system (including both technical and organizational features) are. Several methodologies have been developed across fields and over time to help identify hazards, document hazardous scenarios, and assess the effectiveness of risk prevention and mitigation measures. The quality of their output is strongly dependent on the reviewer's detailed knowledge of the system under consideration and on his experience using hazard analysis methodologies (the guidance they provide for exploring the system's hazardous behavior is indeed often limited). The resulting analyses can take the form of "forward analyses" (starting from component/sub-system malfunctions and evaluating how the combinations of these events might lead to an accident) or "backward searches" (starting from the accidents that one desires to avoid, understand what combination of factors would lead to their occurrence). These techniques aim to provide quantitative risk metrics (evaluating the probability of a loss event happening) or qualitative definitions of safety requirements.

1.1.3 A new hazard analysis technique: STPA

Noting the limitations faced by traditional hazard analysis techniques in identifying and evaluating the hazards born from component interactions and increased reliance on software, Leveson (2003, 2012) developed a new accident model built on systems theory and control theory: STAMP, which stands for Systems Theoretic Accident Model and Processes.

Based on STAMP, hazard identification technique STPA (System-Theoretic Process Analysis) has the potential to uncover more causes of hazards than traditional techniques because it goes beyond chain of event accident models and conventional emphasis on system reliability. At the

other end of a system's lifetime, another STAMP-based method can be used to elicit useful knowledge from accidents and incidents in order to inform system and sector regulation redesign: CAST (Causal Analysis based on STAMP) provides a framework to help understand the entire accident process and identify the most important systemic causal factors involved (Leveson, 2012).

Reliability is defined by Wikipedia (2011) as "the ability of a person or system to perform and maintain its functions in routine circumstances, as well as hostile or unexpected circumstances". Should a component not perform its intended function, intuition warrants that undesired behavior and, possibly, an accident is bound to take place. Reciprocally, the post-event search for factors having caused undesired behavior usually, and quite spontaneously, starts with a search for components that may have failed to act as intended. Under these conditions, designing for maximum reliability is a logical step for anyone interested in designing safe systems. This intuition is based on the assumption that system designers can know, at the time of conception, all the behaviors that a well-functioning system can take once in the hands of its users, once in its deployment field. Since they can predict this behavior set, the thinking goes, they will design against and eliminate all possibility of unsafe behavior. Under these premises, no undesired behavior can result from the normal operations of system elements and, reciprocally, an accident can only be explained by the unexpected ill-function of a system component.

Unfortunately, especially as systems grow in scale and complexity, as systems live longer lives through multiple generations of users and upgrades, and unless stringent control measures are taken to constrain their behavior, this intuitive assumption does not hold: now that sub-systems are designed by separate teams, now that their control involves increasing amounts of software, now that component interactions through information, mass and energy exchanges knit a complicated and, oftentimes, unknown web, now that replacements and partial upgrades can introduce asynchronous evolution in the system's DNA, perfectly functioning components may interact in a way that can be both unexpected and unsafe. As a consequence, and as emphasized in the very first paragraph of Engineering a Safer World, "high reliability is neither necessary nor sufficient for safety" (Leveson, 2012)

STPA gives consideration to unsafe interactions between components that may not have failed, including organizations, software and humans whose ability to process the information they receive from a system may be so constrained, by design, experience or environmental factors, that unsafe decisions are made. Previous uses of STPA have found that its potential for enhanced hazard analysis is real (e.g. Pereira et al. 2006 on the Ballistic Missile Defense System, Ishimatsu et al. 2010 on the Japanese HTV spacecraft), opening the way to safer system design. It however remains to be seen whether the nature of the results that STPA can provide correspond to the kind of assurances deemed acceptable by authorities responsible for allowing new complex systems into the market. If they can, then STPA could be used by complex system manufacturers to meet their pre-market regulatory obligations. If they cannot, the evolution that would bring regulatory agencies towards a position that would accept STPA could be evaluated. Such evolution of regulations can indeed not be discounted: albeit impacted by regulatory delays, authorization and certification processes are meant to mirror public perception of safety, itself a concept that has much changed over time.

Informing this attempt to assess how useful STPA could be to the evaluation of hazards and whether it could be introduced in regulatory toolboxes is a case study performed in the area of nuclear medicine. Five example sub-systems of an experimental proton-therapy facility were analyzed to evaluate what hazards the facility's design and proposed operations were creating for patients, personnel and the facility. Performing this analysis provided experience in applying STPA to a complex device. In so doing, lessons were learned about how to use STPA and new techniques were developed to overcome the things that were found difficult to do.

1.2 A case study in the field of nuclear medicine

1.2.1 Radiotherapy: a powerful yet dangerous treatment option

Cancer appears in one out of three human lives (Tubiana, 2009). Current therapeutic options are often used in association one with the other. They consist in:

- Surgery: removal of unhealthy tissue
- Chemotherapy: delivery of toxic drugs to the whole body through the blood stream. Aimed at killing malignant cells, it has important side-effects due to limited selectiveness
- Radiotherapy: location-specific delivery of radiation energy to the body aimed at killing malignant cells by causing damage to cell's vital organelles and DNA.

50% to 60% of cancer patients are treated at least in part with radiotherapy (Procertus, 2012) with either palliative or curative intent (NCI, 2012). Radiotherapy makes use of either high-energy photons (X-rays, gamma-rays) or accelerated particle beams (electrons, protons, heavy ions such as carbon) to impair the metabolism of tumor cells. The range of these energetic particles in tissue depends on the mechanisms through which they lose energy by interacting with matter. These vary according to the particles' nature (photons penetrate further than charged particles; electrons themselves penetrate further than the much more massive protons) and energy, as well as to the density of the penetrated material.

Unfortunately, radiation is also very dangerous to healthy cells. As a consequence, errors in dose delivery position can have dramatic consequences, especially when critical structures such as nerves, the stem chord, the retina, reproductive organs etc. are affected. Further, even when delivery position is right, tight control of the amount of dose delivered is critical as dose defines the ability to control the target tumor. While underdose can lead to failure in controlling the tumor, overdoses can lead to tissue necrosis that often irreversibly impairs the functions of organs and may induce chronic complications such as secondary cancers (ICRP, 2000, chap. 3).

The past twenty-five years have witnessed the occurrence of horrific radiological accidents in therapeutic settings where the Hippocratic oath “do no harm” was seriously challenged, patients facing at times devastating consequences such as documented by (Bogdanich, 2010 and 2011) as a result of radiation overdose. (ICRP, 2000), a retrospective analysis of radiotherapy accidents

that occurred worldwide between 1974 and 2000, included 79 cases of radiotherapy accidents (external beam and brachytherapy). The treatment delivery steps and causal factors that were identified as having contributed to these 79 accidents are listed in Table 1 below.

Table 1- Classes and frequencies of accidental exposure in radiotherapy (ICRP, 2000)

Accidental exposures in external beam therapy	No. of cases	Percentage of cases (rounded)
Equipment problems	3	6.5
Maintenance	3	6.5
Calibration of the beams	14	30
Treatment planning and dose calculations	13	28
Simulation	4	9
Treatment set-up and delivery	9	20 (*)
Total	46	100
Accidental exposures in brachytherapy	No. of cases	Percentage of cases (rounded)
Equipment and source problems	5	15
Source order and delivery, calibration, and acceptance	3	9
Source storage and preparation for the treatment	5	15
Treatment planning and dose calculation	6	18
Treatment delivery	11	34
Source removal and return	3	9
Total	33	100

* It is likely that errors in the treatment set-up are more frequent than tabulated, since many instances probably remain unreported, especially if the consequences are moderate, i.e., affecting one or a few fractions.

While (ICRP, 2000) emphasized the inadequacy of personnel training and their lack of familiarity with the machine's operations (esp. calibration, treatment planning and dose calculation), the follow-up study (ICRP, 2009) adopted a very different tone, bringing to light the contribution of both treatment and device complexity to the creation of hazardous situations, with added emphasis on the important tasks delegated to software.

Understanding the underlying causes of past accidents (ICRP, 2000 and 2009) and proposing fixes that will prevent them from contributing to new ones (The Advisory Board Company, 2011) are important steps in improving radiotherapy safety. However, and as evidenced by the discrepancies observed between the findings of the two ICRP studies, technical innovation continuously introduces new features to the therapy scene, each one carrying risks that may have never been assessed before. It is therefore indispensible that the focus be broadened beyond

preventing the re-occurrence of past accident, i.e. retrospective improvement to system design. Accident prevention must also take a prospective stance so that new design features, both organizational and technical, are designed to be safe in the first place, i.e. based on a prospective and systemic analysis of potential safety issues. Improvements to human factors, ergonomics and organizational features of systems will not reap their full fruits until safety is allowed to become a primary technical and organizational design criterion.

1.2.2 Proton therapy at the PROSCAN facility

Proton therapy can achieve conformal dose distribution with less collateral irradiation of neighboring healthy tissues than can photon beams. It is therefore especially attractive when dose must be delivered in volumes close to radiation-sensitive structures (such as the optic nerve or the stem chord when treating brain tumors) (Allen et al, 2009). This benefit does not come cheap: with installation costs in the $125-250 million range and running costs around $20 million per year, proton accelerating facilities are as large as a football field, costly to build and expensive to operate[6].

Hosted at the Paul Scherrer Institute (PSI) in Switzerland, PROSCAN is an engineered facility that delivers an accelerated proton beam to four user areas, three of which are used to treat cancer patients. Its design has changed over time to accommodate evolving user requirements and to integrate state of the art technological advances. After several years of successfully delivering radiation in the form of "spots" that are moved one step at a time across a fixed tumor volume, it is being upgraded to provide faster "continuous scanning" in a new user area: Gantry-2. Continuous scanning would allow treatment of mobile tumors such as those attached to mobile organs like the lungs (Nichols et al. 2012) and, to a lesser extent, the prostate (Allen et al, 2009). The Gantry-2 system can scan a volume of one liter eight times in less than a minute, a powerful

---

[6] (Huff, 2007), who quotes figures from several US projects, aptly summarizes the gigantism of these investments: "*At a [investment] cost of $125 million [$15-25 million yearly operating costs], and sometimes much more, today's centers usually fill a city block, with treatment rooms and a phalanx of physicians, therapists and physicists to operate the proton-charged beam that enthusiasts say can treat tumors with astounding precision.*" A specific illustration is provided by the website of the MD Anderson Proton Therapy Center at the University of Texas, where the following description of the proton facility can be found: "*Within our 96,000-square-feet of space, the MD Anderson Proton Therapy Center houses four treatment rooms that include one fixed beam room and three equipped with gantries. Each gantry is three stories tall, 35 feet in diameter, weighs 190 tons and rotates around a patient to direct the proton beam precisely at the cancerous tumor.*" (MD Anderson, 2012)

capability that puts tight constraints on the system components' response times and spatial delivery precision.

1.2.3 The PROSCAN STPA study

A research project was undertaken to apply STPA to the PROSCAN facility's most recent user area in order to evaluate the applicability of STPA to the safety review of a large and complex design close to completion. Started in April 2011 and documented in (PSI, 2012b), the study evaluated Gantry-2's hazardous states and compared them to those documented in the draft PSI Safety Report (PSI, 2012a) that is being put together to support Gantry-2's commissioning application. An informal workshop was also held during which PROSCAN designers applied STPA to one of the PROSCAN subsystems.

While performing this analysis, several methodological questions were raised. They led to suggestions for additional guidance and complementary interpretations of the STPA process that are documented throughout this thesis.

1.3 Contributions

Engineering Systems, as a field, concerns itself with complex sociotechnical systems. In the words of J. Sussman (2012), they are technology-enabled networks that transform, transport, exchange and regulate mass, energy and/or information. They include large numbers of components and interactions. Their existence is accompanied by important sociotechnical aspects, which can be social, economic, political. They may exhibit nested complexity, where technical complexity is doubled by institutional complexity. Finally, they are dynamic, involving multiple time scales, uncertainty and life-cycle issues.

Applying STPA to the PROSCAN's facility Gantry-2 user area and deriving generalizable methodological findings from this case study contributes to the Engineering Systems' bodies of knowledge pertaining to risk evaluation, assessment of the interactions between different system elements and, through considerations of the changes in behavior to be expected from individuals, teams, and physical systems over time, understanding of life-cycle issues faced by complex

systems. Lessons were derived on how to perform STPA and new techniques were developed to overcome the difficulties that can be met in that process.

Albeit rooted in the engineering sciences, this contribution to advancing the study of safety, an emergent system property with inherent social relevant, includes an assessment of the socio-economic landscape in which design decisions are made: the regulatory frameworks applying to radiotherapy devices are discussed as they pertain to hazard analysis and risk evaluation.

Overall, the contributions presented in this dissertation can be summarized as follows:

1. Provide experience in applying STPA to a complex device. Some information on efficacy was derived by comparing STPA results with an existing safety assessment although a more formal counterpart is needed for stronger evidence. Some information on learnability and usability was obtained from the outcome of an informal workshop held at PSI.
2. Advance the STPA methodology by creating notations and a process to document, query and visualize the possibly large number of hazardous scenarios identified by STPA analyses, with the goal of facilitating their review and use by their intended audience.
3. Introduce the notion that STPA can be used to guide hazard analyses using more traditional techniques such as fault and event trees.
4. Introduce a new hazard analysis technique to the fields of nuclear medicine safety and medical devices, thus potentially influencing the state of the art in these domains.
5. Analyze the regulatory structure that currently exists for the market authorization of external beam radiotherapy and, more generally, medical devices that do not contain radioactive materials in the USA and the European Economic Area, and conclude whether STPA would fit well in it.

1.4 Organization of this thesis

Chapter 2 describes the regulatory landscape faced by manufacturers of external beam radiotherapy devices and other medical devices with respect to pre-marketing hazard analysis requirements in the USA and Europe. It describes the hazard analysis techniques currently used to meet these obligations and introduces the reader to STAMP and STPA.

Chapter 3 starts with an introduction to radiotherapy in general and proton therapy in particular. It continues with a short description of the PROSCAN facility before presenting the results of the STPA study. If finally compares the results of the STPA study with those of the Safety Report written by the PROSCAN design team for the Swiss commissioning authorities.

Chapter 4 presents the lessons learned from using STPA to perform the hazard analysis of the PROSCAN facility. It documents the methodological questions that were raised in this process and offers heuristics to answer them. It especially addresses the challenge of keeping track of STPA results by suggesting that their hierarchical nature lends itself to being organized in a causal taxonomy. This classification can be used to evaluate how powerful preventive and protective measures are at eliminating, reducing and mitigating hazards. Along the way, proposals are made for a visual representation of these causal dependencies and for the creation of a software tool that can map this information to, ultimately, support the analyst's task.

Chapter 5 summarizes the contributions associated with this work and explores avenues for future work.

Appendix 1 is a glossary of STAMP, meant to be especially useful to the reader new to STAMP and STPA.

Appendix 2 provides more technical information about the design of the PROSCAN facility.

Appendix 3 offers some background on the use of probabilistic and quantitative risk assessment metrics by regulatory authorities in different industrial sectors.

1.5 References

Allen A., Pawlicki T., Bonilla L., Bucci M.K., Buyyounouski M., Cengel K., Dong L., Fourkal E., Plastaras J., Yock T., Harris E., Price R., *An Evaluation of Proton Beam Therapy by the Evaluation Subcommittee of ASTRO's Emerging Technologies Committee*, 2009 https://www.astro.org/uploadedFiles/Content/Clinical_Practice/ProtonBeamReport.pdf

Black J., *System Safety as an Emergent Property in Composite System*", International Conference on Dependable Systems and Networks (DNS09), 2009

Bogdanich W., Radiation Boom, series of 7 articles published in the New York Times between January 2010 and February 2011, http://topics.nytimes.com/top/news/us/series/radiation_boom/index.html, last accessed July 17, 2012

Huff C., *Catching the Proton Wave*, Hospital and Health Network, 2007, accessed August 7th, 2012 at http://www.hhnmag.com/hhnmag/jsp/articledisplay.jsp?dcrpath=HHNMAG/Article/data/03MAR2007/0703HHN_FEA_Protonwave&domain=HHNMAG

ICRP, *Prevention of Accidents to Patients Undergoing Radiation Therapy*, ICRP Publication 86, Ann. ICRP 30 (3), 2000

ICRP, *Preventing Accidental Exposures from New External Beam Radiation Therapy Technologies*. ICRP Publication 112. Ann. ICRP 39 (4), 2009.

Ishimatsu T., Leveson N., Thomas J., Katahira M., Miyamoto Y., Nakao H., *Modeling and Hazard Analysis using STPA*, Presented at the Conference of the International Association for the Advancement of Space Safety, Huntsville, Alabama, May 2010

Leveson N.G, *Safeware - System Safety and Computers*. Addison-Wesley, Reading, MA, 1995

Leveson N.G., *A New Accident Model for Engineering Safer Systems*, ESD-WP-2003-01.19, MIT Engineering Systems Division Internal Symposium, May 29-30 2012, accessed at http://esd.mit.edu/WPS/internal-symposium/esd-wp-2003-01.19.pdf

Leveson N.G., Daouk M., Dulac N., Marais K., *A Systems Theoretic Approach to Safety Engineering*, Engineering Systems Monograph, 2004

Leveson N., Engineering a Safer World: Systems Thinking Applied to Safety, MIT Press, 2012,

Lewes G.H., *Problems of Life and Mind*, volume 2 of 1st Ser. James R. Osgood and Co., Boston, MA, 1875

Lutz R.R., *Software Engineering for Safety: A Roadmap*, ICSE '00 Proceedings of the Conference on the Future of Software Engineering

Marais K., A *New Approach to Risk Analysis with a Focus on Organizational Risk Factors*, MIT PhD thesis, 2005

MD Anderson Proton Therapy Center at the University of Texas, *Why MD Anderson Proton Therapy Center?*, accessed August 7th, 2012 at http://www.mdanderson.org/patient-and-cancer-information/proton-therapy-center/why-choose-md-anderson/index.html

National Cancer Institute (NCI), *Radiation Therapy for Cancer* FAQ sheet, http://www.cancer.gov/cancertopics/factsheet/Therapy/radiation, accessed August 7th, 2012

Nichols R., Henderson R., Huh S., Flampouri S., Li Z., Bajwa A., D'Agostino H., Pham D., Mendenhall N., Hoppe B., *Proton Therapy for Lung Cancer*, Thoracic Cancer Vol. 3 Iss. 2, pp109-116, May 2012

Pereira S., Lee G., Howard J., *A System-Theoretic Hazard Analysis Methodology for a Non-advocate Safety Assessment of the Ballistic Missile Defense System,* Proceedings of the 2006 AIAA Missile Sciences Conference, Monterey, CA, November 2006.

Procertus Biopharm, http://www.procertus.com/markets.htm, accessed August 7th, 2012

PSI, 2012a, *Report on Proton Therapy Safety Measures for Gantry 2- Draft*

PSI, 2012b - *Evaluation of STPA in the Safety Analysis of the Gantry 2 Proton Radiation Therapy System* (authored by Antoine B., Rejzek M., Hilbes C.) - *Draft*

Spiegel Online - International Newsletter, *Doomed Flight AF 447: Questions Raised about Airbus Automated Control System*, Monday May 30th, 2011, accessed at http://www.spiegel.de/international/world/0,1518,765764,00.html

Sussman, J., *Complex Sociotechnical Systems: The Case for a New Field of Study*, 2012 Charles L. Miller Lecture, Wednesday April 25th, 2012, MIT

The Advisory Board Company, The Pipeline – Technology Insights, *Patient safety innovations take center stage at ASTRO 2011*, November 7, 2011 http://www.advisory.com/Research/Technology-Insights/The-Pipeline/2011/11/Patient-Safety-Innovations-Take-Center-Stage-at-ASTRO-2011, last accessed July 18, 2012

Thomas J., Leveson N., *Performing Hazard Analysis on Complex, Software- and Human-Intensive Systems*, submitted at the 2011 ISSC conference

Wikipedia, *Reliability*, accessed at http://en.wikipedia.org/wiki/Reliability on July 13th, 2011

PAGE INTENTIONALLY LEFT BLANK

2 Background: Hazard Analysis Techniques in the Medical Devices Industry

"One reason cited for the low incidence of errors occurring in the delivery of radiation therapy is the strict regulatory environment surrounding its practice."
Pennsylvania Patient Safety Authority, 2009

2.1 Introduction

Regulation of the production and trade of goods is aimed at correcting market deficiencies. Common deviations from a perfect market include the existence of oligopolies and monopsonies, the creation of positive and negative externalities from a given economic activity, and the imperfect information of at least one party to the market.

When it comes to healthcare products, the potential for these three factors to bring the market to an undesired state is large:

- Producers of drugs and medical devices are often in oligopolistic, if not monopolistic, situations in niche markets that they design and protect with the development, acquisition and defense of strong intellectual property rights. The power that they have over the end-customer is all the greater that their products are hoped to provide life and health benefits for which there is no alternative.
- The availability of healthcare products to society creates potentially large positive externalities (a healthy population and workforce), but also potentially tragic ones (harm and injury to people who were seeking a cure), neither of which would appear in the manufacturer's books.

- Finally, in the absence of regulatory constraints, the information available to manufacturers and to potential customers regarding benefits to be expected from the use of these products would be highly asymmetrical, especially as the time between a transaction occurs and the observation of its effects can be large. Under the maximizing profit paradigm, this information can fall prey to manipulation and distortion.

Regulation attempts to correct these potential problems by constraining manufacturing and sale behaviors, and deciding what products can be released to the market. This chapter explores the regulatory frameworks that apply to external beam therapy devices such as proton therapy machines. It specifically identifies the hazard analysis requirements imposed in the USA and Europe on device manufacturers, before describing the hazard analysis techniques currently used to meet these regulatory obligations and introducing the STAMP-based hazard analysis technique called STPA.

2.2 Regulation of radiotherapy devices

Manufactured products can be responsible for losses through improper design, inadequate manufacturing or improper use. Improper design can broadly be characterized as the system not being designed to meet foreseeable hazards, or creating situations that are hazardous to its users. Inadequate manufacturing results in products not meeting their design specifications. Improper use puts the system outside of the use assumptions envelope that was postulated by the design team. As a consequence of the detrimental effects potentially associated with each of these three pathways to losses, regulatory agencies concerned with healthcare safety attempt to only allow on the market systems that are considered safe for their intended use, control the manufacturing practices of drugs and device manufacturers, and work at ensuring that the people that will prescribe and use these drugs and devices know what they are doing.

The regulatory frameworks imposed in the United States of America (USA) and the European Economic Area (EEA) on the design, manufacture and use of radiotherapy devices are presented in the following paragraphs. Because the intent of this presentation is to identify whether hazard analysis is required by law and what techniques are accepted by the regulatory powers to perform these analyses, particular emphasis is put on the design phase.

2.2.1 Certification of medical devices in the USA

In the USA, government protection of public health and safety with regards to medical devices is primarily pursued through food and drug law[7]. Devices that make use of radiation produced either through use of sealed sources of nuclear materials (e.g. ^{60}Co to produce gamma radiation) or accelerators (e.g. X-ray machines) are additionally subject to regulation of nuclear matters (Title 10 of the Code of Federal Regulations). This later authority is shared among several government agencies at the federal, state and local levels (NRC, 2011):

Approval of device design and medical procedures: the Food and Drug Administration (FDA) oversees approval of drugs (including radiopharmaceuticals) and devices (including machines using nuclear radiation or material) for safety and efficacy but it does not regulate the use of the devices and pharmaceuticals that it approves (Food, Drug and Cosmetics Act, Chapter V).

Regulation of nuclear material production: the Nuclear Regulatory Agency (NRC) regulates the manufacture and distribution of sealed sources or devices containing byproduct material (10CFR32). Proton therapy machines do not use nuclear materials and are therefore not covered by these provisions.

Regulation of the medical use of nuclear material: the NRC, and relevant agencies of the States that have entered into an agreement with the NRC, regulate the use of radioactive material for medical use; they issue licenses to medical users such as university medical centers, hospitals, clinics and physicians in private practice. Proton therapy machines do not use nuclear materials and are therefore not covered by these provisions.

Regulation of the medical use of radiation-producing machines that do not produce radioactive material (such as X-ray machines or proton therapy machines) is provided by relevant health and environmental agencies of the States.

[7] Food and drug law regulates food, drugs, cosmetics, medical devices and biological products. By a rough estimate, 25 cents of every consumer dollar is spent for products that fall within the categories of products regulated by the FDA. (Hutt et al, 2007)

The focus of this thesis is on assessing hazards before designs are brought to market. As such, it is mostly concerned with the pre-market review of medical devices and will not provide detailed information about the regulatory framework applicable to the use of accelerator driven radiotherapy machines of which proton therapy are one instance.

2.2.1.1 Market clearance and approval of medical devices by the US Food and Drug Administration (FDA)

The US Food and Drug Administration clears medical devices after a pre-market review that includes the mandatory performance of a risk analysis

2.2.1.1.1 Origins and evolution of the FDA's mandate with respect to nuclear devices

Food and Drug Law is the oldest field of consumer protection legislation in the US with city, county and state food and drug laws originating in colonial times. The Federal Pure Food and Drugs Act of 1906 and the Federal Meat Inspection Act of the same year were Congress's first major efforts to protect consumers on a national level, long before the creation of the Federal Trade Commission (1914), the Environmental Protection Agency (1970) or the Consumer Product Safety Commission (1972). The Food and Drug Administration (FDA) was established in 1930 as part of the US Department of Health and Human Services (Hutt et al, 2007).

While the FDA's mission to protect the public health was originally limited to regulating adulteration and misbranding of food and drugs, the Food and Drug Law's later amendments expanded the realm of the law's purview to include cosmetics and medical devices (Hutt et al, 2007).

Although advances in bioengineering are likely to challenge it in the near future, the current distinction between drugs[8] and devices is based on the respective mechanisms through which they act on the body. While drugs and devices are both intended for curative purposes, drugs are basically exercising their effect through chemical action or by being metabolized by the body[9]

---

[8] which, until 1962, were regulated like food and did not require pre-market approval

[9] According to section 321(g) of the Federal Food Drug and Cosmetic Act, "the term “drug” means

- (A) articles recognized in the official United States Pharmacopoeia, official Homoeopathic Pharmacopoeia of the United States, or official National Formulary, or any supplement to any of them; and
- (B) articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals; and

while devices are intended to bring about changes in the structure or function of the body through nonchemical nor metabolitic means[10]. Drugs pre-market submissions are reviewed by the FDA Center for Drug Evaluation and Research while pre-market review of medical devices is done by the FDA Center for Devices and Radiological Health (CDRH) born from the 1982 merger of the Bureau of Radiological Health with the Bureau of Medical Devices (FDA, 2006).

The following pieces of legislation are the foundations of FDA's mandate to regulate the marketing of medical devices:

- The Federal Food, Drug and Cosmetic Act of 1938 established the FDA's regulatory structure and extended federal authority to include supervision of cosmetics and medical devices.
- The Medical Device Amendments of 1976 was the result of President Nixon's 1969 message to congress calling for minimum standards and premarket clearance for certain medical devices (FDA, 2006)[11]. It did not significantly enlarge FDA's jurisdiction, but transformed its approach to regulation of medical devices by introducing the concepts of safety and effectiveness of medical devices and classifying devices based on their risk. It significantly increased the FDA's regulatory powers by imposing a federal premarket clearance process.

- (C) articles (other than food) intended to affect the structure or any function of the body of man or other animals; and
- (D) articles intended for use as a component of any articles specified in clause (A), (B), or (C)."

Source: FD&C Act, Title 21 of the Code of Federal Regulation, Chapter 9

[10] According to section 321(h) of FD&C Act, a medical device is : "an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is:

- recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them,
- intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or
- intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes."

Syringes, gauze pads, condoms, hospital beds, in vitro reagents, X-ray machines are examples of devices.

[11] The then Department of Health, Education, and Welfare formed a committee that calculated that 10,000 injuries were attributable to “therapeutic devices” and more than 700 of these injuries were fatal.

- In the aftermath of the 1985-1987 Therac accident (Leveson, 1993) and following several studies showing that reporting of adverse events, although required since 1984 and necessary to improve FDA's information about actual experience with devices, was very limited[12], the Safe Medical Devices Act of 1990 confirmed FDA's way of processing medical device applications and strengthened its post-market authority (Samuel, 1991).
- Finally, the Food and Drug Administration Modernization Act of 1997 was intended to improve the efficiency of the review process and required, among many other provisions, that FDA recognize standards developed by national and international consensus organizations.

Radiation-emitting electronic devices only came under FDA oversight in 1971, through implementation of the Radiation Control for Health and Safety Act of 1968. GE's recall of 90,000 television sets that were emitting radiation beyond voluntary industry-set standards in 1967 had indeed led Congress to realize that the American public was exposed to vast amounts of unnecessary radiation. What was first established as the Radiological Health Unit in 1948 was transferred from the Public Health Service to the FDA, where it later was merged into the Center for Devices and Radiological Health. Its mission was to protect against unnecessary human exposure to radiation from electronic products in the home, industry and healing arts by setting federal radiation standards, monitoring compliance, and conducting research on the then poorly known harms of radiation exposure for home and occupational use (FDA, 2006)(Tran, 2006). The FDA has no legislative authority to regulate users of these devices, with the exception of facilities that perform mammography.

While the Nuclear Regulatory Commission regulates the manufacture and use of nuclear material and byproduct material emitting radiation, it does not play a role in the pre-market review of medical devices, even when they include nuclear material such as ^{60}Co sources of gamma rays.

[12] A 1986 General Accounting Office (GAO) study showed that less than one percent of device problems occurring in hospitals are reported to FDA, and the more serious the problem with a device, the less likely it was to be reported. A GAO follow-up study in 1989 concluded that despite full implementation of the Medical Device Reporting (MDR) regulation, serious shortcomings still existed (FDA, 2012a).

2.2.1.1.2 510(k) clearance and Pre-Market Approval (PMA)

FDA's main regulatory responsibilities with respect to medical devices consist in clearing medical devices for market and taking action to have devices withdrawn from the market when they are reported to be associated with accidents and injuries. While recognizing that post-market surveillance and action are essential to ensuring the safety of the public not only by withdrawing dangerous systems from the market but also by providing the regulator with information that can be used in later reviews, this thesis is only interested in the review mechanisms involved in the pre-market activities. They are described in the following paragraphs.

Within CDRH, the Office of Device Evaluation approves and clears medical devices through a pre-market review process. Clearance requires that the manufacturer provide reasonable evidence that the device is safe and effective, as defined in 21 CFR 860.7 (emphasis added):

> (d)(1) There is reasonable assurance that a device is safe when it can be determined, based upon valid scientific evidence, that the probable benefits to health from use of the device for its intended uses and conditions of use, when accompanied by adequate directions and warnings against unsafe use, outweigh any probable risks. The valid scientific evidence used to determine the safety of a device shall adequately demonstrate the absence of unreasonable risk of illness or injury associated with the use of the device for its intended uses and conditions of use.
>
> (e)(1) There is reasonable assurance that a device is effective when it can be determined, based upon valid scientific evidence, that in a significant portion of the target population, the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results.

No further indication is provided by the FDA as to what the criteria for safety and effectiveness stand for. It is specifically noteworthy that there is no suggestion that the safety claim need be substantiated by risk benefit studies that are quantitative.

Medical devices fall into one of three classes depending on their intended use, indications for use, and the risks that they pose to the patient and/or user. Class I are devices with the lower risk (e.g. scalpel, tongue depressors, stethoscopes...); Class II are moderate risk (e.g. X-ray systems, physiological monitors...); and Class III are the highest risk (they support or sustain human life,

are of substantial importance in preventing impairment of human health, or present a potential, unreasonable risk of illness or injury. Examples include pacemakers, artificial joint implants, ...).

A classification database is maintained by the FDA to help manufacturers identify the class that their devices fall in. For example, subpart F of the "Radiology" panel (21 CFR 892) deals with radiological "Therapeutic Devices". When applying for pre-market clearance of a proton therapy machine, one would need to receive clearance for at least four therapeutic devices, all of which are class II:

- 892.5050 Medical charged particle radiation therapy system (including treatment planning computer programs)
- 892.5710 Radiation therapy beam shaping block
- 892.5840 Radiation therapy simulation system
- 892.5770 Powered radiation therapy patient support assembly

All classes of devices are subject to General Controls, the baseline FDA requirements for device marketing. These General Controls consist in:

- Quality System Regulation (QS), also known as Good Manufacturing Practices (GMP). They require that design and manufacturing activities be subject to a quality assurance program. The GMP are described in 21CFR820,
- Labeling requirements,
- Establishment registration and Device listing with the FDA.

Class II and III are subject to special controls that include

- Special labeling requirements
- Mandatory performance standards
- Post-market surveillance
- FDA medical device specific guidance

Finally, depending on their class and as summarized in Table 2, medical devices will follow one of three possible market submission processes, listed below in order of increasing stringency:

- exemption from pre-market notification,
- pre-market notification via 510(k) clearance, or

- pre-market approval (PMA).

Table 2 - Device classification and FDA requirements

Device class	Requirements imposed by the FDA	Pre-market submission type	Examples
Class I	General Controls	Exemption[13] (most) or 510(k)	Scalpel, tongue depressor...
Class II	General Controls + Special Controls	Exemption (a few) or 510(k)	X-ray systems, physiological monitors...
Class III	General Controls + Special Controls + Pre-Market Approval including clinical data	510(k) if substantially equivalent to a pre-amendment device or is a pre-amendment device, or Pre-market approval (PMA)	Pacemakers, artificial joint implants, ...

A 510(k) is a premarket submission made to FDA to demonstrate that the device is "substantially equivalent" to a legally marketed device. Substantial Equivalence means that a device is as safe and as effective as a device that is already legally in the market. Since the Medical Device User Fee and Modernization Act of 2002, FDA levies a fee for reviewing applications. In exchange, it is required to provide an answer to 510(k) submissions within 90 days, a challenging timeline given the volumes of information to be reviewed. If the reviewer is not confident that the evidence presented ensures substantial equivalence, questions can be asked to the manufacturer. After a maximum of two rounds of questions, the substantial equivalence claim is rejected if still unconvincing[14].

A Premarket Approval (PMA) application is aimed at scientifically demonstrating to FDA the safety and effectiveness of the class III device for which approval is requested. It must include results of clinical and laboratory investigations. According to a consultant in device premarket submissions (Syring, 2003), the FDA often requires more time than the statutory 180 day review cycle to approve a PMA.

---

[13] Devices exempt from 510(k) are pre-1976 amendment devices not significantly changed or modified since then and for with a regulation requiring a pre-market approval has not been published by the FDA, or Class I and II devices specifically exempted by regulation. A list of these is available at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/315.cfm

[14] Interestingly, the CDRH has started to implement a Third Party Review Program. This program provides an option to manufacturers of certain devices of submitting their 510(k) to private parties (Recognized Third Parties) identified by FDA for review instead of submitting directly to CDRH. This arrangement echoes that in place in the European Economic Area, with private entities, the notified bodies, being in charge of checking that device manufacturers comply with regulatory requirements

In short, to be authorized by the FDA to market their device, manufacturers of non-exempt class II devices such as a proton therapy systems would have to follow the general controls regulations that include Good Manufacturing Practices, and to obtain pre-market clearance of their device by convincing the CDRH reviewers that the claim to substantial equivalence that they make in their 510(k) application is substantiated, especially with respect to software.

(Maisel, 2004) summarizes this information in a decision tree. Reproduced in Figure 1, it describes how regulatory constraints apply to medical devices depending on their risk, similitude with legally marketed devices, and exemption status.

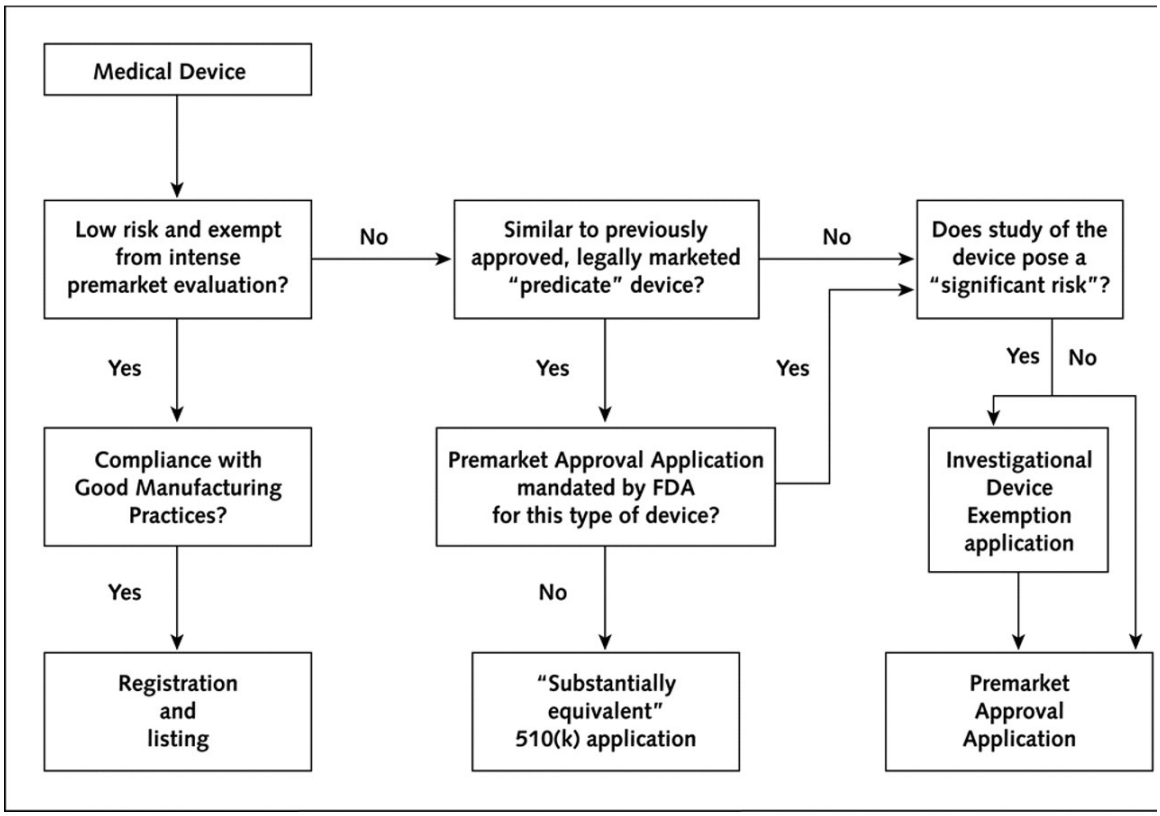


Figure 1 - Pathways to marketing of medical devices in the USA (Maisel, 2004)

2.2.1.1.3 Good Manufacturing Practices

The Safe Medical Devices Act of 1990 added design validation requirements to the Good Manufacturing Practices set forth in the Quality System Regulation (21 CFR 820). Put in effect in 1997 (FDA, 2006), these requirements are the first through which FDA requires that a risk analysis be performed as part of the device design activities (21 CFR 820.30):

> (g)*Design validation.* [...] **Design validation shall include** software validation and **risk analysis**, where appropriate. The results of the design validation, including identification of the design, method(s), the date, and the individual(s) performing the validation, shall be documented in the DHF.

Because these updated Good Manufacturing Practices do not specify technical details such as what risk analysis techniques the applicant is expected to use, the FDA issues non-mandatory guidance documents to help put together solid applications. The medical device industry being a highly diversified and fragmented one, with 24,000 manufacturers (Maisel, 2004) occupying numerous small niche markets with a few products only (Altenstetter, 2003) and not always having a vast amount of regulatory expertise available to them, these guidance documents, albeit no substitute for regulation and in no way mandatory, are a strong and commonly used industry reference.

The 1997 "Design Control Guidance For Medical Device Manufacturers" (FDA, 1997) mentions fault tree analysis (FTA) and failure modes and effects analysis (FMEA) as examples of verification methods and activities, thus suggesting that using them is deemed a valid way to meet the risk analysis requirement set forth in the 21CFR820.

Fifteen years later, the 2012 "Guidance for Industry and Food and Drug Administration Staff - Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approvals and De Novo Classifications" (FDA, 2012b) does not reference specific hazard analysis techniques but its appendix A points out ISO standard 14971 (ISO, 2007).

ISO 14971 (2007) outlines a process for risk management activities. While recommending that "*the manufacturer might find it convenient to use an as-low-as reasonably-practicable approach*" to establish its risk acceptability policy, it does not establish what level of risk can be considered acceptable ("*that decision is left to the manufacturer*"). It instead proposes a process

to assess risks and protect against them. The annexes, provided for informative purposes only, give more tactical guidance. More precisely:

- Annex C offers questions that a designer could ask to identify the devices' characteristics that could impact safety.
- Similarly helping with the hazard identification phase, annex E offers examples of hazards and of initiating circumstances for the manufacturer to consider.
- Annex D describes general frameworks for risk estimation. It suggests general categories of risk control options (designing for inherent safety, adding protective measures and providing information for safety). It finally proposes techniques (event tree analysis and fault tree analysis) for residual risk estimation, but notes that these are only "*possible techniques*", that (emphasis added) "*there is no preferred method for evaluating overall residual risk and [that] the manufacturer is responsible for determining an appropriate method*".
- Finally, Annex G offers information on a few techniques deemed to be useful to risk management programs. These techniques are all based on the chain of events accident model. They are described as complementary ("*it might be necessary to use more than one of them*"). They are: Preliminary Hazard Analysis (PHA), Fault Tree Analysis (FTA), Failure Mode and Effects Analysis (FMEA), Hazard and Operability Study (HAZOP) and Hazard Analysis and Critical Control Point (HACCP).

ISO 14971 (2007) is an FDA-recognized standard. As a consequence (emphasis added), "*assuring conformity with this standard may help device manufacturers meet the design validation requirements specified in the Design Controls section of Part 820 of FDA's regulations governing quality system*s" (FDA, 2012b), but conformity to the standard is neither necessary nor sufficient to establish safety in the eye of the regulator. Conversely, a manufacturer is legally allowed to use any hazard evaluation and risk assessment framework, even when not listed in ISO 14971 (2007). This means that STPA could technically be used to support a clearance application. As long as the case they make is convincing to the reviewer, the manufacturer's obligation to abide by the GMP requirements will be met, whatever risk assessment method they choose.

2.2.1.1.4 Legal implications of market clearance and approval: a (short) introduction to federal preemption of State tort law

Manufacturers create devices. But FDA authorizes their introduction to the market. Who should bear legal responsibility for tort claims brought to court as a consequence of a marketed device having caused harm to patients or users? This section will succinctly touch on product liability law as it pertains to medical devices.

Product liability claims are typically filed in State Courts as there is no federal product liability law. However, the 1976 Medical Devices Amendment to the FD&C Act contains an express preemption clause that enables federal regulation to trump state regulation of medical devices and offers manufacturers of medical devices some protection against defective design and defective marketing claims[15].

Without entering the debate, very active among law scholars (e.g. (Stewart, 2000), (Funk et al, 2007), (Jijon,2008), (Schwartz and Silverman, 2010)), about the respective merits of federal tort preemption (which is cited as encouraging efficient regulatory activities and industry innovation) and State authority for tort claims (which is claimed to create a stick necessary for adequate reporting of adverse events and to reduce the risks of regulatory capture by manufacturer lobbies), it is important to note that the Supreme Court has recently weighed in favor of tort preemption when the device has received a full safety review by the FDA through the Pre-Market Approval process. In Reigel v. Medtronic (2008), the Supreme Court ruled that individuals injured by a device that has received pre-market approval from the FDA may not sue the manufacturers for damages. This express preemption does however not apply to devices marketed following the less rigorous 510(k) clearance, as evidenced by Supreme Court ruling

[15] Section 2 of the *Restatement (Third) of Torts: Products Liability* distinguishes between three major types of product liability claims:

- defectively manufactured: the product does not conform to the designer's or manufacturer's own specifications.
- defective design (even though properly manufactured): some flaw in the intentional design of a product makes it unreasonably dangerous.
- defectively marketed: improper labeling or failure to warn of a product's hidden dangers

Metronic, Inc v. Lora Lohr (1996) that considered 510(k) to be an exemption from federal safety review[16] (Richards,1997).

According to (Maisel, 2004), the FDA annually receives approximately 4,000 510(k) applications compared with fewer than 100 Pre-Market Approval Applications. As a result, and if the Supreme Court's opposition to preemption revealed by Medtronic v. Lohr holds, less than 3% of non-exempt medical devices put on the market every year are subject to the Federal express preemption clauses. Others could lead to trial for tort liability in State courts, a perspective far from the liking of medical device manufacturers.

2.2.1.2 Ensuring safe use of medical devices

As this part of the device lifecycle is not the core interest of this thesis, it will not be dwelled upon in much detail. Further, this discussion will focus on radiotherapy machines that, like proton therapy machines, produce radiation not through the use of nuclear materials but with accelerators. Were radioactive materials or byproducts involved, then the responsibilities of the Nuclear Regulatory Commission (NRC) (or that of the States, when they have taken over regulatory authority for the possession and use of these materials under section 274 of the Atomic Energy Act of 1954) would have to be further emphasized.

Radiation therapy is regulated at both the federal and state level and is considered one of the most highly regulated medical practices. Basically, there are two parts to ensuring that medical devices are safely used: licensing facilities to allow them to use specific medical systems on one hand, and certifying personnel as having been trained to use the medical devices correctly.

Healthcare facilities are licensed by the States. Further certification can be voluntarily sought from professional organizations. For example, accreditation by the Joint Commission is meant to reflect the quality of the care that healthcare facilities provide.

---

[16] Two cases relative to drugs shed a complimentary light on the preemption situation for manufacturers of healthcare products.

- Wyeth v. Levine (2009) states that FDA approval of a drug (in this case, the drug's label) does not shield its maker from lawsuits brought by patients injured by use of the drug
- Pliva, Inc. v. Mensing (2011) validates implied preemption for equivalent drugs.

One of the reasons that radiotherapy is considered a high risk procedure is that it requires coordination of care by a team of several people. These include radiation oncologists, dosimetrists, medical radiation physicists, radiation therapists and radiation oncology nurses.

- Radiation oncologists develop and prescribe treatment plans and are certified by the American Board of Radiology.
- Dosimetrists create treatment plans and work with the oncologist and the medical therapist to select the treatment plan that optimally delivers dose to the patient's tumor. They are certified by the Medical Dosimetrist Certification Board.
- Medical physicists oversee the work of the dosimetrist and are responsible for making sure the equipment works properly. They are certified by the American Board of Radiology or the American Board of Medical Physics.
- Radiation therapists administer the daily radiation treatment to the patient. They are certified by the American Registry of Radiologic Technologists.

Authorization for these personnel to use the radiotherapy devices is granted by the NRC based on them following required training and having acquired certain degrees of experience with the machines they are to use.

Finally, although they do not bear any regulatory power, professional associations play a major role in disseminating best practices to their members through workshops, white papers and the issuance of specific guidelines. They also work with accreditation and certification agencies to develop educational curriculum. They include the American Association of Physicists in Medicine (AAPM), the American Society of Radiation Oncology (ASTRO), and the American Society of Radiologic Technologists (ASRT).

2.2.1.3 Criticism of the current regulatory framework for medical devices in the USA

The 510(k) review process has been heavily criticized by all parties to the regulatory process. Device manufacturers claim that it prevents valuable innovation to reach needy patients in a timely manner while patient advocacy groups criticize its leniency, especially as it allows devices to be marketed based on substantial equivalence to grandfathered, pre-1976 amendment, devices that never were evaluated for safety and effectiveness. In the heat of that debate, the FDA commissioned the Institute of Medicine (IOM) for a review of the 510(k) procedure, asking

if it optimally protected patients and promoted innovation in support of public health. The IOM 2011 report on the public health effectiveness of 510(k) was very critical, finding that "*510(k) clearance is not a determination that the cleared device is safe or effective*". The committee concluded instead that "*the 510(k) process lacks the legal basis to be a reliable premarket screen of the safety and effectiveness of moderate-risk devices and, furthermore, that it cannot be transformed into one*" (IOM, 2011). As a consequence, the IOM recommended that a new regulatory framework be designed that would integrate premarket and post market information to follow the device throughout its lifecycle, bringing to mind the adaptive licensing frameworks presented for example in (Eichler et al, 2012).

While the overhaul of the 510(k) process is likely to take several years, other areas of regulatory responsibilities may be capable of faster improvement. Such is the case of the post-market surveillance system. Consistent with the IOM's recommendation to enforce regulatory oversight throughout the lifetime of the medical devices rather than only reforming the gate of their entry into the market, many are the calls to strengthen post-market monitoring systems by creating a more accessible error and accident database from which knowledge could be extracted and disseminated.

Systems such as MedWatch and the Manufacturer and User Device Experience Database are in place to help the FDA conduct post-market surveillance (Maisel, 2004). The FDA receives 80,000 to 120,000 device related adverse events reports per year, but (Fraass, 2008) pleads for the development of improved reporting systems for errors in radiotherapy. Although medical devices manufacturers and users are under increased obligation by the Safe Medical Devices Act of 1990 to report adverse events, they only have to report to the FDA events suspected to have caused a death or, in certain cases, serious injuries (FDA, 2012b). Further, this information does not appear to be available to third parties making it impossible to conduct research on this data. Finally, "*study and analysis of radiotherapy planning and delivery errors are performed relatively infrequently*" by the medical community (Fraass,2008), resulting in lost opportunities to learn from near-misses and errors to improve therapy outcomes. According to (Fraass 2008), this situation likely derives from the "*current medico-legal climate in the United States [that] makes it quite difficult for anyone directly involved in an incident to publicly release any information about the incident (such as whether any particular device or procedure was directly*

responsible for the problem)", including the fact that an incident even occurred. Although it will be no easy task to change the mindsets of practitioners country-wide whose behavior has been shaped by this climate, the benefits that could be derived in terms of improving procedures and quality assurance practice are large. By performing a 10 year retrospective study of patient events from external beam radiotherapy at their hospital, Hunt et al (2012) illustrate how valuable incident and accident information is to identifying areas for improvement and convincingly advocates for improvement in reporting.

Not that the rate of these events is expected to be large. The published literature, although biased towards large and usually university-affiliated hospitals, whose error profiles can be expected to be very different from that of the more numerous smaller practices, offers a few reference points. (Macklin et al, 1998) reviewed 1925 patients who were treated at the Cleveland Clinic in 1995 with a total of 93,332 individual radiotherapy fields, revealing a crude radiation delivery error rate of 0.18%. These errors had negligible risk of adverse medical outcomes and their rate compared favorably with the error rates for drug administration in large tertiary care hospitals. (Fraass, 1998) found an error rate for all manually treated cases of 0.21% and an error rate of 0.085% for all computer-controlled cases. Yeung et al (2005) performed a study involving 13,385 patients treated during a 10-year period and found 624 incidents. 40% of the dose error accidents were discovered before any treatment occurred and 98% of the incidents had dose errors <5% (clinical relevance threshold). Hunt et al (2012) reported an average 0.93% event rate per course of treatment over 597,000 treatments delivered over 10 years. Errors were found in all steps of the process (treatment planning, treatment delivery, information transfer and communication) and, as demonstrated by the reports of international publications dedicated to disseminating information such as Shafiq et al (2009), ICRP (2000, 2009), and IAEA (2000), in all countries of the world

Despite their number being low and the frequency of serious incidents extremely rare, (Fraass, 2008) points out that many of these errors could be avoided through dissemination of error information to practitioners country-wide. In the aftermath of the 2000 Institute of Medicine Report on errors in the health care system (IOM, 2000), that reported the striking number of 90,000 yearly deaths from medical error in the USA, the light that incident reporting and analysis

could shed on what improvements ought to be brought to devices and treatment procedures should not be kept under a bushel.

2.2.2 Certification of medical devices in Europe

By certifying their conformity to essential requirements defined by law, private entities are responsible for authorizing the marketing of medical devices in Europe

2.2.2.1 The European regulatory landscape for medical devices

An overview of the regulatory landscape for medical devices in the European Economic Area (EEA[17]) as it pertains to safety assessment is provided here for comparison with the one in place in the USA. The interested reader is referred to the chapter written by C. Hodges in (O'Grady et al, 1999) and the publications of the European Commission's General Directorate for Health and Consumers (ECHC, 2012a) on the regulation of medical devices in Europe for more detailed information.

Regulatory constraints are defined in the EEA through directives that are provided by the European Commission and must be transposed into national legislation of the EEA member states to bear regulatory effect. There are three device directives that defines the constraints imposed on manufacturers for manufacturing and selling medical devices in the European common market:

- Directive 90/385 EEC on Active Implantable Medical Devices covers all powered implants or partial implants that are left in the human body
- Directive 93/42/EEC, amended by directive 2007/47/EC to include software, covers most medical devices. It became mandatory in 1998.
- Directive 98/79/EC covers In Vitro Medical Devices.

93/42/EEC is relevant for proton therapy machines. Article 1(2)a defines medical devices in a way similar to the FDA. They are:

---

[17] The EEA includes the 27 members of the European Union and 3 members of the European Free Trade Association: Iceland, Liechtenstein and Norway. EEA States are allowed to participate in the EU's internal market without a conventional EU membership, provided that they adopt all EU legislation related to the single market except laws on agriculture and fisheries (Wikipedia, 2012). Switzerland, which hosts the PROSCAN facility studied in this thesis, has a similar agreement with the EU.

"any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of:

- diagnosis, prevention, monitoring, treatment or alleviation of disease,
- diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap,
- investigation, replacement or modification of the anatomy or of a physiological process,
- control of conception,

and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted by such means."

The risk assessment principles behind European regulation of medical devices are similar to the American ones, but their implementation differs significantly. In the European Economic Area (EEA), marketing of a medical devices does not involve its assessment by a public agency, or the grant of a marketing authorization. Instead, "*the onus of ensuring and declaring that a product conforms to the legal essential requirements is placed on the manufacturer, but in many instances this is subject to approval by an independent technical organization (known as the notified body)*" (Hodges, 1999). Contrarily to the USA, where federal preemption protects certain manufacturers from certain categories of tort claims, the legal responsibility for the safe performance of the device unambiguously rests with the manufacturer.

The notified bodies are private, commercial testing houses that contract with the manufacturers to review the Declaration of Conformity in which manufacturers explain how they meet the Essential Requirements specified by the European Directive. When the review is favorable, notified bodies authorize the manufacturer to put the CE mark on their devices. Once this is done, they can start selling their devices (it is an offense to place a medical device on the market without CE marking). CE stands for "Conformité Européenne" (European conformity in French). In contrast to the US system where a device's clearance for market is granted once and for all

unless adverse effects are reported and the FDA takes the device off the market, the CE mark must be renewed every 5 years (MDC, 2009).

As in the USA, devices are classified according to the degree of risk they represent. Classification rules are set forth in the 93/42/EEC directive. Roughly, Class I covers devices that do not enter or interact with the body, Class IIa and IIb devices enter the body or interact with it, and Class III devices affect the functions of vital organs. Proton-therapy machines belong to Class IIb.

This classification is used to define options for conformity assessment methods. Once a manufacturer has selected the most appropriate conformity assessment module for its device, it selects a notified body to perform the third party conformity assessment tasks. The manufacturer prepares a Technical File and a Declaration of Conformity that document how the Essential Requirements defined in 93/42/EEC are met.

The essential requirements specify the aspects of safety and performance that must be satisfied at the time at which a relevant product is placed on the market. They are stated as principles, not as detailed technical requirements. As with the Good Manufacturing Practices of the US legislation, the detailed technical specifications are elaborated in voluntary harmonized standards[18]. While standard compliance is not sufficient to substantiating a safety claim before the FDA, compliance to harmonized standards, while remaining optional, does demonstrate conformity to European essential requirements.

One of the essential requirements is that manufacturers must eliminate, reduce, protect against and warn of risks as far as possible. This implies that a manufacturer must carry out a risk analysis. Standard ISO 14971 (2007), the same as referenced by FDA guidance, is harmonized with respect to 93/42/EEC and can thus be used as a guideline to perform the risk assessment that is required for the Declaration of Conformity.

[18] As defined by the European Commission, a harmonized standard is a European standard elaborated on the basis of a request from the European Commission to a recognized European Standards Organization to develop a European standard that provides solutions for compliance with a legal provision (ECEI, 2012)

As in the USA, where manufacturers have an obligation to list and register their products with the FDA, device manufacturers are required to provide certain information to competent authorities in the EEA member states. For example, notification of the first placement on the market of a medical device must be provided to the corresponding competent authorities and applications for approval of a clinical investigation of medical devices must be submitted to the national competent authority. These competent authorities are usually the member states' departments or agencies for public health (ECHC, 2012b). Finally, and like in the USA, device manufacturers are subject to vigilance and incident reporting requirements.

As for use regulation, each member state defines its own procedures to authorize health facilities and health practitioners to use medical devices. These will not be presented here. They are similar to those in place in the USA.

Finally, as in the USA, professional organizations play an essential role in disseminating best practices. ESTRO, for example, is the European Society for Radiotherapy and Oncology. Like its American counterpart ASTRO, it disseminates information to its members through education and professional development courses, publications of guidelines and research updates, and conferences. It also is active in promoting research.

2.2.2.2 Criticism of the European regulatory framework

The current system for approval and control of healthcare products is likely to evolve in Europe. Recent healthcare related affairs, such as the Mediator drug scandal where French pharmaceutical company Servier falsely marketed a product and manipulated the evidence it submitted to regulatory authorities (2010) and the PIP breast implant scandal in which breast implants were sold that had little resemblance to those for which CE mark had been obtained (2011), have indeed stimulated debate on the adequacy of the current regulatory structures.

As a response to the Mediator affair, French Minister for Work, Employment and Health Xavier Bertrand launched a large public dialogue to reform the way that patients and users are protected from the risks associated with health care products, drugs and medical devices included. Working Group 6 offered ideas to strengthen control and assessment of medical devices. Their recommendations were presented to the French government in May 2011.

The PIP scandal led the same Minister and other high-ranking authorities to call for more stringent European regulations on medical devices, including strengthened post-market oversight. It remains to be seen what evolutions will be decided to strengthen controls without stifling cross boundary trade of these products, the dual goal pursued by the European Commission (Altenstetter, 2003).

In the meantime, a new health agency was created in France in 2012 (ANSM) to replace the one previously in charge of approving the marketing of pharmaceutical drugs (AFSSAPS). Its authority includes the evaluation of medical devices and how their cost compares to the benefits that can be expected of them.

2.2.3 International bodies

Although non regulatory in nature, a few international organizations play a role in setting expectations for regulatory action with respect to safety of medical devices that use radiation. They include the following four institutions.

The Conference of Radiation Control Program Directors is a non-profit agency that serves as a common forum for many radiation protection agencies to communicate with each other and promote uniform radiation protection regulations and activities. Similarly, the Global Harmonization Task Force, created in 1992, aims to achieve greater uniformity between national medical device regulatory systems.

The International Commission on Radiological Protection (ICRP) is an independent, international organization staffed with some 200 volunteer members that represent leading scientists and policy makers in the field of radiological protection. It publishes reports on radiological protection and therefore participates in disease prevention and the protection of the environment.

The International Atomic Energy Agency publishes information about the safe use of ionizing radiation in medicine and organizes workshops around themes of common interest to the international community. IAEA TecDoc 1040 (IAEA, 1996) and the Radioprotection webpage (IAEA, 2012) are examples of the technical guidelines that they develop to guide national patient safety and radioprotection programs worldwide.

2.3 Hazard Analysis Techniques for Medical Devices

The rules defined by regulatory powers to authorize the sale and use of a medical device in respectively the USA and the EEA require that the manufacturer perform a risk assessment of its device. None of the legally binding requirements mandate the use of a specific risk assessment technique. Rather, both regulatory regimes point to ISO 14971 (2007) as one possible way of demonstrating conformity with the obligation of risk analysis. This section presents the hazard analysis techniques mentioned in ISO 14971 (2007) before describing their limitations and explaining how STAMP aims to address them.

2.3.1 Hazard analysis techniques recommended by ISO 14971

ISO 14971 (2007) does not impose the use of any specific risk analysis technique. It references and describes five of them for informative purposes: PHA, FTA, FMEA, HAZOP and HCCP. It recommends that more than one be used to take advantage of their respective strengths.

These methods are based on the linear chain of events accident model. This model assumes accidents result from a succession of several failure events, or the propagation of a specific failure throughout a system. The implication is that accidents can be prevented if the event chain can be interrupted. As a consequence, this model elevates component reliability, redundancy and barriers in depth as ultimate safety shields.

PHA, hardware FME(C)A, HAZOP and HCCP are bottom-up inductive methods: starting with a postulated fault or initiating condition, the analyst attempts to ascertain the effects of that fault or condition on system operations. Inductive methods are applied to determine what undesired system states are possible (Vesely et al, 1981).

On the contrary, FTA and the less common functional FMEAs follow a top-down deductive approach: starting with a postulated undesired system state, one attempts to find out what modes of system, subsystem, or component behavior contributes to this failure. Deductive methods are applied to determine how undesired system states can occur (Vesely et al, 1981).

The following paragraphs succinctly describe these five hazard analysis techniques. Their borrow from the US NRC's Fault-Tree Handbook (Vesely et al, 1981), the UK Health & Safety

Executives 2005 Review of Hazard Identification Techniques (Gould et al, 2005), Vincoli's Basic Guide to System Safety (Vincoli, 2006), FDA's 2004 revision of the 2001 HACCP Guidelines, IEC 61882 and ISO 14571. More detailed guidance on how to use these methods can be found in the following standards[19]:

- IEC 60300-3-1 (2003) *Dependability Management - Part 3: application guide - Section 1: Analysis techniques for Dependability - Guide on Methodology - Annex A: Brief Description of Analysis Techniques* (FTA, HAZOP, FMEA),
- IEC 60300-3-9 (1995) *Dependability Management - Part 3: application guide - Section 9: Risk Analysis of Technological Systems* (PHA),
- IEC 61025 (1990) *Fault Tree Analysis* (FTA),
- IEC 60812 (1985) *Analysis Techniques for System Reliability - Procedures for Failure Mode and Effect Analysis* (FMEA),
- IEC 61882 (2001) *Hazard and Operability Studies - Application Guide* (HAZOP),
- 2001 *HACCP Guidelines* by the US Food and Drug Administration (HACCP).

2.3.1.1 Preliminary Hazard Analysis (PHA)

Overview: As explained in standard EN 1050 (1997)[20] *Safety of Machinery - Principles for risk assessment*, "*PHA is an inductive method whose objective is to identify, for all phases of life of a specified system / sub-system / component the hazards, hazardous situations and hazardous events which could lead to an accident.*" Essential to this methodology is the determination of the criticality of potential accidents and the importance that should be given to their reduction by the designer.

Process: With the help of checklists such as the three pages of EN 1050 (1997)'s appendix A, hazard matrices, past engineering experiences and other available data, a preliminary list of hazards is created (hazards are defined as undesired events or conditions e.g. mechanical shock,

[19] For additional methodological references, the reader is referred to http://www.ntnu.no/ross/info/standards.php, a list of standards relevant for reliability, safety and security studies maintained by the ROSS network at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway

[20] EN 1050 (1997) was superseded in 2007 by ISO 14121 *Safety of machinery - Risk assessment - Part 1: Principles*, itself superseded in 2010 by ISO 12100 *Safety of machinery - General principles for design - Risk assessment and risk reduction*, which are expected to include the same kinds of checklists as EN 1050 (1997) but were not available to this dissertation's author at the time of writing.

corrosion, fire, explosion ...). The causes and the effects of these hazardous conditions are documented in a second step, along with a qualitative evaluation of the severity of the identified effects. In a third step, the probability of possibility of occurrence is evaluated. This leads to risks being ranked according to the attention that they should receive from the design team. Finally, recommendations are made to eliminate and control the hazards, and suggestions are offered for follow-on analyses. The description of the obtained results can be presented in different ways (e.g. tree, table such as Table 3).

Table 3 - Sample PHA worksheet (Vincoli, 2006)

Ref.	**Hazardous condition**	**Probable causes**	**Effects**	**Risk Assessment Code**	**Assessments**	**Recommendations**
Sequence number	Nature of the condition (see checklist if needed)	Describe what is causing the stated condition to exist	If allowed to go uncorrected, what will be the effect or effects of the hazardous condition?	Hazard level assigned, as guidance for designer attention	Probability or Possibility of occurrence (likelihood) and Severity (exposure, magnitude)	Recommended actions to eliminate or control the hazard

Purpose: PHAs are intended for use at the earliest stages of the design process. They are meant to be complemented by detailed hazard analysis such as FMEAs when more design information becomes available.

2.3.1.2 Failure Mode and Effect Analysis (FMEA) / Failure Modes, Effects and Criticality Analysis (FMECA)

Overview: Developed by the US Military right after World War II[21], FMECA was used by NASA's Apollo program (NASA, 1966). FMECA, and FMEA, later spread to civilian domains such as the automotive industry[22] and civil aviation[23].

[21] The first FMEA guideline was Military Procedure MIL-P-1629 "Procedures for performing a failure mode, effects and criticality analysis" dated November 9, 1949.
[22] 1967 publication by the Society of Automotive Engineers of "Design Analysis Procedure for Failure Modes, Effects and Criticality Analysis"
[23] inclusion of FMEA as a recommended method to assess the safety of a system in ARP4761, Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment published in 1996 by the Society of Automotive Engineers

FMEA is a bottom-up, inductive approach, evaluating common failure modes and examining their effects on the whole system (Vincoli, 2006). The C in FMECA indicates that FMEA is extended to consider the criticality of the failure modes' effects according to the combined influence of severity classification and probability of occurrence based on best available data (MIL-STD-1629A, 1980), and that assurance and controls are described for limiting the likelihood of such failures (Vesely, 1981).

MIL-STD-1629A (1980) presents two kinds of FME(C)As: functional FME(C)As and hardware FME(C)As (also called component (Vincoli, 2006) or piece-part (Wikipedia, 2012)). Functional FME(C)As consider the effects of failure at the functional block level, such as a power supply or an amplifier. Hardware FME(C)A consider the effects of individual component failures, such as resistors, transistors, microcircuits, or valves (Wikipedia, 2012).

Process: Per MIL-STD-1629A (1980) section 4.4.2, the FMEA process consists in 8 steps.

1. Define the system. This includes identification of internal and interface functions, expected performance at all indenture levels, system restraints and failure definitions.

2. Construct functional and reliability block diagrams for each item configuration involved in the system's use.

3. Identify all potential item and interface failure modes and define their effects on the immediate function or item, on the system, and on the mission to be performed.

4. Evaluate each failure mode in terms of the worst potential consequences that may result and assign a severity classification category[24].

5. Identify failure detection methods and compensating provisions for each failure mode.

6. Identify corrective design or other action required to eliminate the failure or control the risk.

[24] Severity classification categories are a qualitative measure of the worst potential consequences resulting from design error or item failure (e.g. minor, marginal, critical, catastrophic).

7. Identify effects of corrective actions or other system attributes, such as requirements for logistics support.

8. Document the analysis and summarize the problems which could not be corrected by design and identify the special controls which are necessary to reduce failure risk.

Once these eight steps are completed, a criticality analysis can be performed. Its purpose is to rank each potential failure mode identified by the FMEA according to the combined influence of severity classification and probability of occurrence. Depending on the availability of specific parts configuration data and failure rate data, the approach can be either qualitative or quantitative. The qualitative approach (section 3.1 of task 102 documented in MIL-STD-1629A) is appropriate when specific failure rate data are not available, e.g. at earlier stages of the design process where components have not yet been specified. Failure rate data is available from reliability handbooks such as MIL-HDBK-217 for electronic parts.

This information can be documented in a table such as Table 4.

Table 4 - Sample FMEA worksheet (MIL-STD-1629A, 1980, Figure 101.3)

SYSTEM _____ DATE _____
INDENTURE LEVEL _____ SHEET _____ _____
REFERENCE DRAWING _____ COMPILED BY _____
MISSION _____ APPROVED BY _____

IDENTIFICATION NUMBER	ITEM/FUNCTIONAL IDENTIFICATION (NOMENCLATURE)	FUNCTION	FAILURE MODES AND CAUSES	MISSION PHASE/ OPERATIONAL MODE	FAILURE EFFECTS			FAILURE DETECTION METHOD	COMPENSATING PROVISIONS	SEVERITY CLASS	REMARKS
					LOCAL EFFECTS	NEXT HIGHER LEVEL	END EFFECTS				

Purpose: As emphasized by the foreword of MIL-STD-1629A, while FME(C)As aim to identify all modes of failures within a system design, “their first purpose is the early identification of all catastrophic and critical failure possibilities so they can be eliminated or minimized through design correction at the earliest possible time” (MIL-STD-1629A, 1980). As a consequence, MIL-STD-1629A recommends that FMECA shall be initiated early in the design phase to aid in

the evaluation of the design and to provide a basis for establishing corrective action priorities (MIL-STD-1629A, 1980, section 4.2).

FMECA and FMEAs provide a basis for quantitative reliability and availability analysis, can assist in the selection of design alternatives with high reliability and high safety potential and can be used to guide maintenance planning. For these quantitative analyses to be accurately performed, detailed design data such as design drawings, system schematics and functional diagrams must be available, along with component reliability databases.

2.3.1.3 Fault Tree Analysis (FTA)

Overview: Conceived around 1961 by H. Watson of Bell Labs for evaluation of the Minuteman launch control system, Fault Tree Analysis (FTA) was further developed by the Boeing company and extensively used during the 1975 Reactor Safety Study (WASH-1400) (Ericson, 1999). It is a top-down, deductive approach to fault[25] analysis: starting with a potential undesirable event called a top event, it attempts to analyze the system in the context of its environment and operation to find all credible ways in which the undesired event can occur. It translates this failure behavior into a graphic model: the fault-tree (Vesely, 1981).

Process[26]: First, a top event is identified and displayed at the top of the fault tree. Then, the immediate, necessary and sufficient events and conditions causing the top event are identified. These contributors can be associated with component hardware failures, human errors, or any other pertinent events that can lead to the undesired event (Vesely, 1981). Contributors are connected with logic gates[27] to the top event. The first level contributors are in turn described as

---

[25] (Vesely, 1981) distinguishes between failures (basic abnormal occurrences such as relay failing to close when a voltage is impressed across its terminals) and faults ("higher order events" specifying what the undesirable component state is and when it occurs, e.g. a bridge that is supposed to open occasionally to allow the passage of marine traffic and that opens without warning because it - correctly - responded to an untimely command issued by the bridge attendant). It also suggests the existence of three categories of faults: primary (any fault of a subsystem/component that occurs in an environment for which the component is qualified), secondary (any fault of a subsystem/component that occurs in an environment for which it has not been qualified), command (proper operation of a subsystem/component but at the wrong time or in the wrong place, e.g. an arming device in a warhead train closes too soon because of a premature or otherwise erroneous signal origination from some upstream device).

[26] For the interested reader, the Fault Tree Analysis tutorial provided by P.L. Clemens and J. Sverdrup (Clemens, 1993) offers a very detailed yet accessible description of the FTA process, the assumptions it bears, its limitations, warnings of misuses and descriptions of common errors as well as several useful heuristics (e.g. symbol meaning. identification of top events, computing of event probabilities, handling of common cause failures) and examples.

[27] usually AND- or OR- gates, but others are possible such as Exclusive OR-, Ordered AND-, Inhibit- gates

resulting from the logic combination of second level events and conditions. The analysis proceeds in this way until the analyst reaches events (called basic events, leaves or initiators) that are independent and for which failure data is available.

Purpose: FTAs aim to identify hazardous scenarios that lead to the undesired top event, gauge system failure probability and support resource allocation to optimize control of risk by identifying the potential contributors to failure that are "critical" and guiding system reconfiguration to reduce vulnerability. The Fault Tree Handbook (Vesely, 1981) makes it clear that although fault trees "*can be evaluated quantitatively and often are*", they are, in essence, qualitative models that reflect the analyst's understanding of what credible faults can create the specified undesired top events. "*The fault tree is a graphic "model" of the pathways within a system that can lead to a foreseeable, undesirable, loss event. The pathways interconnect contributory events and conditions, using standard logic symbols. Numerical probabilities of occurrence can be entered and propagated through the model to evaluate probability of the foreseeable, undesirable event*" (Clemens, 1993 - original emphasis).

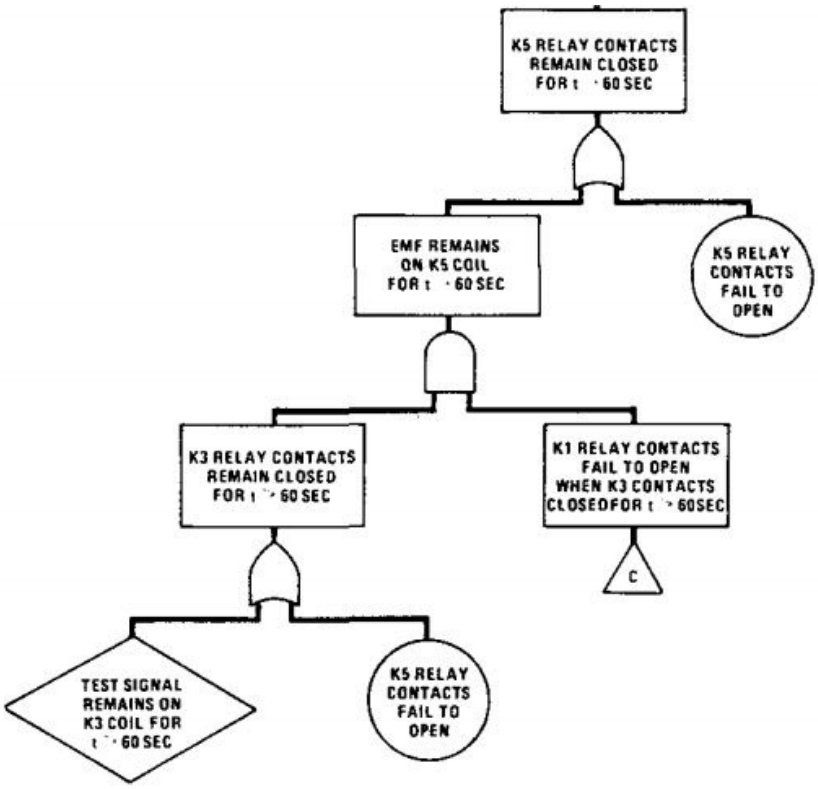


Figure 2 – Typical Fault Tree – extract (Vesely, 1981)

2.3.1.4 Hazard and Operability Study (HAZOP)

Overview: HAZOP was initially developed in the 1960s by the United Kingdom's chemical industry (Kletz, 2006).

The method is based on guidewords that a team of engineers uses to examine potential deviations of process and operations from design conditions. A characteristic feature of HAZOP studies is the "examination session" during which a multi-disciplinary team systematically examines all relevant parts of a design or a system, under the guidance of a trained, experienced study leader (IEC 61882:2001).

Table 5 – Basic guidewords (adapted from IEC 61882:2001)

Guideword	Meaning	Example
No (not, none)	Complete negation of the design intent	No data or control signal passed
More (more of, higher)	Quantitative increase in a parameter	Higher temperature than desired
Less (less of, lower)	Quantitative decrease in a parameter	Data passed at a lower rate than intended
As well as (more than)	Qualitative modification/increase (an additional activity occurs)	Some additional or spurious signal is present
Part of	Qualitative modification/decrease (only some of the design intention is achieved)	Only part of the system is shut down
Reverse	Logical opposite of the design intention occurs	Back-flow when the system shuts down
Other than (other)	Complete substitution (another activity takes place)	The data or control signals are incorrect
Early / Late	Relative to the clock time	Beam is turned off late
Before / After	Relative to the clock time	Reactor outflow valve is closed before reactor inflow valve is
Faster / Slower	The step is done/ not done with the right speed	Plane performs escalation maneuver slower than what air traffic controller expects
Where else	Applicable for flows. transfer, sources and destinations	Patient is not sent for the right procedure

Process: IEC 61882 describes the HAZOP study procedure in detail. The examination step consists in the following steps:

1. Divide the system into parts

2. Select a part and define the design intent

3. Identify deviations by using guidewords on each element

4. Identify consequences and causes

5. Identify whether a significant problem exists

6. Identify protection, detection, and indicating mechanisms

7. Identify possible remedial/mitigation measures (optional)

Then repeat steps 3 to 7 for each element, and 2 to 7 for each part of the system

The resulting analysis can be documented in a table such as Table 6.

Table 6 - Example HAZOP output sheet (IEC 61882:2001)

PART CONSIDERED:	ALARM SYSTEM								
DESIGN INTENT:	TO SOUND A GENERAL PURPOSE ALARM (GPA)								
ELEMENTS:	**INPUTS:**	INITIATION SIGNAL							
		ELECTRICAL ENERGY							
PERSONNEL:									
	SOURCES:	ALL ALARM GENERATORS							
	DESTINATIONS:	ALL PERSONNEL ON PLATFORM							
No.	**Element**	**Guide word**	**Deviation**	**Possible causes**	**Consequences**	**Safeguards**	**Comments**	**Actions required**	**Action by**
1	GPA Initiation signal and electrical energy	NO	No inputs	1) Instruments or personnel do not initiate GPA	Failure to alert personnel	None	Unlikely but possible	None	
				2) Personnel try to initiate GPA, but signal fails to reach alarm	As above	Duplicated connections and fail safe logic, i.e. "Current to open, spring to close"	Unlikely		
				3) No electrical energy	As above	Uninterruptible power supply	As above		
2		MORE	More inputs	1) False alarm	Personnel stressed unnecessarily	None	Possible	Should initiation require two buttons?	
				2) Mischief alarm	As above	Discipline and code of practice	Unlikely	None	
3	Inputs	MORE	More inputs	More electrical energy	Damage to alarm system	Dedicated protected power supply	Unlikely	None	
4		LESS	Less initiation	Initiation signal only reaches some alarms	Some personnel not alerted	Routine alarm checks		None	

Purpose: Although they should be carried out early in the design phase to have influence on the design, HAZOP studies require the availability of a design close to completion to be carried out effectively. As a result, and as for both FTA and FME(C)A, it is usually carried out when the detailed design has been completed or on an existing system to identify modifications that could reduce risks and operability problems.

2.3.1.5 Hazard Analysis Critical Control Point (HACCP)

Overview: Conceived by NASA with the assistance of food company Pilsbury as part of an effort to produce safe food for the astronauts in the 1960s, HACCP has been internationally adopted as a best safety management practice for the food industry (e.g. the EC Food Hygiene Directive 93/43/EEC requires that food businesses assess and control potential food hazards on the basis of principles used to develop HACCP). Its guidelines and application were codified by the Codex Alimentarius Commission, the United Nations body responsible for implementing the Joint FAO/WHO Food Standards Programme.

Process (FAO, 1997; FDA, 2001): HACCP is based on seven principles. The first requires that a hazard analysis be performed while the other ones are about risk management through the definition and control of Critical Control Points.

(FDA, 2001) recommends to use a flow diagram delineating the steps in the process from receipt of raw materials to sale. At each step of the flow diagram, hazards should be identified along with their severity and the preventive measures that exist or should be taken to control them.

The hazards that can cause food and related medical products to be unsafe are considered by HACCP to be of three kinds: microbiological (dangers associated with microorganisms), chemical (contamination from pesticides, oil residues, other process products) and physical (presence of wood, metal, glass... in the end product). (FDA, 2001) and other references provide lists of such hazards to facilitate their identification by the analyst as well as criteria to rank the severity of these hazards.

Purpose: HACCP stands out among the methods discussed so far by its specificity to the food industry. According to the FDA, "*HACCP is a management system in which food safety is addressed through the analysis and control of biological, chemical, and physical hazards from raw material production, procurement and handling, to manufacturing, distribution and consumption of the finished product*". It can valuably be applied to the safety management of the production of "food like" products such as pharmaceuticals but will not be discussed further as it has little relevance for non-biological medical devices.

2.3.2 Limitations of traditional hazard analysis techniques

Each of these techniques was developed in a specific context and for a specific task. As such, despite them being very well suited to their respective original purposes, the extension of their application to new settings and new types of systems necessarily faces limitations, especially in a world where systems are increasingly dependent on software for control and instrumentation, and where humans' interaction with controlled processes is increasingly remote. Further, the fundamental assumptions usually held by the users of these failure-based techniques, namely that accidents can be modeled as linear chains of events and that ensuring high levels of reliability is a guarantee for safety, do not hold in a world of complex systems with high degrees of interaction with their environment and between their components.

Many safety analysts have come to realize that factors other than failures have been, and increasingly are, found to be responsible for accidents. To illustrate this point, consider the following accident, presented by Hardy at the 6th IET International System Safety Conference. The replacement of a faulty breaker in a concrete blade mill by a maintenance crew led to death of an employee. While the breaker was being replaced, the mill technician was mending broken paddle tips inside the stopped mill. The breaker having been reset, the PLC that controlled it brought power back to the mill, lethally injuring the mill technician. The PLC controlling the breaker had indeed been modified to bring power to components after power losses, and understood the reset of the breaker as a power loss situation[28]. As illustrated by this example, harm does not require a failure to occur and perfectly operating elements can cause harm: their design can be flawed in the first place (e.g. sensors installed too far from the process they are to give a measure for, wrong computer algorithm, humans allocated tasks that they should not have been expected to perform well under conditions departing from normal ones etc.) and inadequate coordination between several stake-holders may allow the system to be moved towards unsafe states.

While traditional safety analysis techniques are based on the premise that losses result from failure events, with system components failing to perform the task they were designed for and mitigation measures failing to contain the resulting mishaps, (Marais, 2005), (Dulac, 2007),

[28] http://systemsafetyskeptic.com/yahoo_site_admin/assets/docs/THardy_IET2011.25940540.pdf

(Stringfellow, 2010) and (Thomas, 2012), among other followers of (Leveson, 2004), explain how accidents should instead be thought of as resulting from a more complex set of factors and safety understood to be a different system property than reliability. Their evaluation of traditional hazard analysis and risk evaluation methodologies concludes that hazard analysis techniques that take into account the

- specificities of a digital world,
- existence of hazards that are not associated with failure events such as inadequate requirement capture and design errors,
- organizational cultures,
- asynchronous evolution of sub-systems on process models held by controllers and operators
- and changing role of human and human cognition in system control

have so far been lacking, despite the growing evidence that these issues are major factors in system evolution towards unsafe states.

Increased reliance on computerized processes, characterized by fast information exchanges and automation of codified decision making, has enabled the creation of larger and more capable systems. It has also led to their materializations becoming harder to understand by individual designers and operators, with increased intellectual distance between the human controllers and the process they control, both at the design (an endeavor now greatly segmented across designer teams that may not be able to wholly communicate one with the other) and operational stages. The analytical challenges associated with the very nature of the flaws possibly introduced by increased reliance on software (design errors rather than reliability failures) and of those possibly created by humans or organizations are not well handled by failure-based accident models.

Further, since they are based on the assumption that accidents are caused by component failures, these traditional hazard assessment techniques need a mature design, complete with specific details about what components will be used to what end, in order to be effective. This approach has two major flaws, a practical one and a theoretical one.

- When hazards are identified at late stages of development, safety can only be achieved by the addition of costly extra redundancy and protection systems, or via costly and time-wasting rework.
- Because software errors and flawed human decision making do not involve random failures, traditional hazard analysis technique rarely provide information that can be used to eliminate or control them.

To address these issues, Leveson (1995, 2004, 2011, 2012) proposes a shift in approach. Rather than relying on linear accident models that assume that accidents result from a chain of events involving the failure of independent components, Leveson poses safety as a system control problem. Rather than taking a bottom-up approach that evaluates a system's freedom from harm by assessing the contribution of each component to preventing or mitigating a linear accident scenario, the Systems Theoretic Accident Model and Process (STAMP) builds upon both systems theory and control theory to propose a top-down approach to what is, by nature, a property that emerges[29] from component interactions and system behavior: safety (Leveson et al, 2004).

2.3.3 STAMP and STPA: a response to the limitations of traditional hazard analysis techniques

The STAMP model of accident causation treats safety as a hierarchical control problem rather than as a failure problem (Leveson, 2011). It thus offers a new perspective that can both facilitate the inclusion of safety consideration at early design stages and enhance learning derived from accident analysis by structuring the analysis around hazard definition, identification of safety constraints, consideration for controller process models and definition of control actions. This insight, however, does not negate the contributions made by traditional analysis methods to improving component reliability.

---

[29] According to Black (2009), the term emergent was first used by Lewes to characterize systems that behaved differently from those he called resultant (1895). According to Lewes, a resultant system could be expressed as a sum or product of its components' outputs, whereas the outputs of components in an emergent system are fundamentally different from each other and from the resulting system behavior.

"Add heat to heat and there is a measurable resultant; but add heat to different substances, and you get various effects, qualitatively unlike."

2.3.3.1 Understanding accidents as resulting from a lack of constraints on system behavior

Safety, as defined in the dictionary, is "*the condition of being safe from undergoing or causing hurt, injury, or loss*"[30]. In short, safety means freedom from loss. When systems are created whose operations include the control of large amounts of energy, of hazardous or valuable materials, information or lives, inadequate system design or improper system operations may result in losses that are important to one or several stakeholders. These losses may be human, spiritual, economic, or environmental. What they have in common is that someone, somewhere, would rather not incur them, and has the direct or indirect ability to make the system designer accountable for them, forcing him to consider their prevention as one of many design objectives. System designers therefore need tools to first identify the hazardous situations that their design might be associated with, and then assess the adequateness of the solutions that they propose to deploy in order to mitigate risks.

As a reaction to the limitations of traditional hazard analysis techniques in dealing with design, software and human contributions to losses, Leveson has proposed a change of perspective based on systems theory. Rather than consider safety as a by-product of component reliability, she suggests that safety is best thought of as an emergent system property resulting from the proper control of hazards through enforcement of safety constraints. "*The cause of an accident, instead of being understood in terms of a series of failure events, is viewed as the result of a lack of constraints imposed on system design and operations*."(Leveson, 2003). This postulate and the framing of safety as an emergent system property[31] form the foundations on which she built the Systems Theoretic Accident Model and Processes (STAMP) framework.

In this framework, the focus is not put on quantifying risks, but on identifying scenarios that lead to losses.

---

[30] www.merriam-webster.com

[31] Emergence characterizes a system property that belongs to the system, but cannot be observed in its parts. This implies that its analysis requires that the system be first viewed as a whole before the contribution of its parts to the emergent property can be investigated.

2.3.3.2 STPA: STAMP-based Hazard Analysis

In the STAMP model, the role of the system safety engineer is to "*identify the design constraints necessary to maintain safety and to ensure that the system design and operations enforce these constraints*" (Leveson, 2003). The STAMP based Hazard Analysis (STPA) is a process that intends to achieve the first of these two objectives.

A top down procedure that starts from the definition of system-level losses to be prevented, STPA helps the analyst to identify the discrete scenarios through which these losses may be realized. Its systemic approach contrasts with methods whose primary units of analyses are individual components whose failure modes are considered as the starting point of hazardous scenarios. STAMP's emphasis on functional identities rather than physical architecture allows STPA to be relevant at all design stages, opening the door to resource-effective safety-guided design rather than possibly less optimal post-design safety addenda.

Described in (Leveson, 2012), STPA proceeds in the following steps.

- After having been invited to define the losses that she wants to avoid and the potential accidents through which they would be realized, the analyst generates system level hazards and the system-level safety constraints that, when correctly enforced, will prevent accident occurrence.
- She then represents the system of her interest in the form of a hierarchical control structure, and identifies the commands that each (human or automated) decision-maker transfers to the lower-level so that an action is finally executed by the process actuators. The downward control and upward feedback channels that link the system's different functional elements together are made explicit in the control structure.
- STPA Step 1 consists in identifying how these commands and the resulting actions could lead to an unsafe state.
- Finally, for each unsafe control action identified in STPA Step 1, "STPA step 2" identifies scenarios through which they can be realized. This identification of hazardous causal factors is supported by the visualization of individual process loops.

Once unsafe control actions and hazardous causal factors are identified, so can be the safety constraints whose enforcement is necessary to keep the system away from harm. These

constraints can be used to formulate safety requirements that will ensure that the corresponding factors' ability to cause losses is limited. Once design changes are proposed, the STPA analysis must be repeated to not only verify that the hazardous situations previously identified have been attended to, but that no new hazard has been introduced in addressing the former ones. While STPA provides input most useful to system designers, it can also support post-design safety evaluation processes by identifying remaining safety gaps and evaluating the relevance of proposed safety ensuring devices and procedures.

2.4 Conclusion

This chapter described the regulatory frameworks applicable to the marketing of medical devices and, more specifically, external beam radiotherapy devices in the USA and the EEA. It showed that both systems are founded on the same principles and would theoretically be open to the use of new hazard analysis techniques to meet the manufacturer's obligation of performing a risk analysis.

It also described the hazard analysis techniques mentioned in the guidelines currently offered for informative purposes by the regulatory authorities on both markets. After having summarized the limitations of these traditional hazard analysis techniques, it presented the STAMP-based Hazard Analysis (STPA) technique as an alternative developed to meet the challenges faced by traditional failure-based hazard analysis techniques.

Most importantly, this chapter concludes that performing STPA can satisfy the legal requirements for regulatory mandated risk analyses in the US and EEA markets for medical devices. Its emphasis on informing design decisions rather than creating post-design safety cases resonates particularly with the regulators' intentions.

This chapter having established the regulatory relevance of STPA, the following one presents the application of STPA to the PROSCAN facility.

2.5 References

Altenstetter C., *EU and Member State Medical Devices Regulation*, International Journal of Technology Assessment in Health Care, Volume 19, Issue 01, January 2003, pp228 - 248

Clemens P.L, Sverdrup J., *Fault Tree Analysis tutorial*, 4th edition, May 1993, available at http://www.fault-tree.net/papers/clemens-fta-tutorial.pdf.

Dulac M., *A Framework for Dynamic Safety and Risk Management Modeling in Complex Engineering Systems*, MIT PhD thesis, 2007

Eichler HG, Oye K, Baird LG, Abadie E, Brown J, Drum CL, Ferguson J, Garner S, Honig P, Hukkelhoven M, Lim JC, Lim R, Lumpkin MM, Neil G, O'Rourke B, Pezalla E, Shoda D, Seyfert-Margolis V, Sigal EV, Sobotka J, Tan D, Unger TF, Hirsch G, *Adaptive licensing: taking the next step in the evolution of drug approval*, Clin Pharmacol Ther. 2012 Mar;91(3):426-37.

EN 1050:1997 *'Safety of machinery. Principles for risk assessment'*

Ericson C., *Fault Tree Analysis - A History*, Proceedings of the 17th International System Safety Conference, 1999. Available at http://www.fault-tree.net/papers/ericson-fta-history.pdf

European Commission General Directorate for Health and Consumers website, http://ec.europa.eu/health/medical-devices/index_en.htm, accessed August 21, 2012.

European Commission General Directorate for Health and Consumers website, http://ec.europa.eu/health/medical-devices/links/contact_points_en.htm, accessed August 21, 2012.

European Commission General Directorate for Enterprise and Industry website, http://ec.europa.eu/enterprise/policies/european-standards/harmonised-standards/, accessed August 21, 2012

FAO, *Hazard Analysis and Critical Control Point (HACCP) System and Guidelines for its Application,* Annex to CAC/RCP 1-1969 (General Principles of Food Hygiene), Rev. 3 (1997). Accessed August 28th, 2012 at http://www.fao.org/docrep/005/Y1579E/y1579e03.htm

FDA, *Design Control Guidance For Medical Device Manufacturers*, March 1997. Accessed August 20, 2012 at http://www.fda.gov

FDA, *Supplement to the 2001 Food Code - Annex 5: HACCP Guidelines*, rev. 2004, accessed August 28th, 2012 at http://www.fda.gov/Food/FoodSafety/RetailFoodProtection/FoodCode/FoodCode2001/ucm089302.htm

FDA, "About FDA / What we do / History / Milestones", 2006. Accessed August 21, 2012 at http://www.fda.gov/AboutFDA/WhatWeDo/History/FOrgsHistory/CDRH/ucm064949.htm

FDA, Medical Device Reporting (MDR) website, http://www.fda.gov/MedicalDevices /Safety/ReportaProblem/default.htm, 2012a. Accessed August 22, 2012

FDA, *Guidance for Industry and Food and Drug Administration Staff - Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approvals and De Novo Classifications*, 2012b. Accessed August 20, 2012 at http://www.fda.gov

Fraass B., Lash K., Matrone G., Volkman S., McShan D., Kessler M., Lichter A., M.D., *The impact of treatment complexity and computer-control delivery technology on treatment delivery errors*, International Journal of Radiation Oncology, Biology, Physics, Volume 42, Issue 3, 1 October 1998, Pages 651–659

Funk W., Shapiro W., Vladeck D., Sokol K., *The Truth about Torts: Using Agency Preemption to Undercut Consumer Health and Safety*, Center for Progressive Reform White Paper #704, 2007

Gould J., Glossopo M., Ioannides A., *Review of Hazard Identification Techniques*, HSL/2005/58, http://www.hse.gov.uk/research/hsl_pdf/2005/hsl0558.pdf

Hodges C., *European Regulation of Medical Devices*, Chapter 1 in O'Grady et al, 1999

Hunt M., Pastrana G., Amols H., Killen A., Alektiar K., *The Impact of New Technologies on Radiation Oncology Events and Trends in the Past Decade: An Institutional Experience*, Int J Radiat Oncol Biol Phys. 2012 Apr 10 (e-publication ahead of print)

Hutt P., Merrill R., Grossman L., Rose I.N., *Food and Drug Law - Cases and Materials*, 3d edition, 2007, Foundation Press, 1727 pages

International Standards Organisation, *ISO 14971: Medical devices — Application of risk management to medical devices*, 2nd edition, 2007

IAEA, *Design and implementation of a radiotherapy programme: Clinical, medical physics, radiation protection and safety aspects*, TecDoc-1040, 1996

IAEA, Safety Reports Series, *Lessons Learned from Accidents in Radiotherapy*, 2000

IAEA, *Radiation Safety in External Beam Radiotherapy*, 2012. Accessed August 22, 2012 at https://rpop.iaea.org/RPOP/RPoP/Content/InformationFor/HealthProfessionals/2_Radiotherapy/ RadSafetyExtBeamRadiotherapy.htm

IEC 61882, *Hazard and operability studies (HAZOP studies) – Application guide*, first edition, May 2001.

Institute of Medicine, *To Err is Human: Building a Safer Health System*, National Academy Press, 2000.

Institute of Medicine, *Medical Devices and the Public's Health: the FDA 510(k) Clearance Process at 35 years*, National Academy Press, July 2011

Jijon A.D., *Federal Preemption of State Tort Suits under the Medical Device Amendments of 1976*, HLS Student Papers, 2008, Accessed July 15, 2012 at http://nrs.harvard.edu/urn-3:HUL.InstRepos:7702808

Kletz, Trevor (2006). Hazop and Hazan (4th Edition ed.). Taylor & Francis

Leveson N. G., Turner C. S., *An investigation of the Therac-25 accidents,* Computer, volume 26, Issue 7, published by the IEEE computer society, 1993

Leveson N., *A New Accident Model for Engineering Safer Systems*, ESD-WP-2003-01.19, MIT Engineering Systems Division Internal Symposium, May 29-30 2002, accessed at http://esd.mit.edu/WPS/internal-symposium/esd-wp-2003-01.19.pdf

Macklin R., Meier T., Weinhous M.S., *Error rates in clinical radiotherapy*, J Clin Oncol, 16 (1998), pp. 551–556

Maisel W., *Medical Device Regulation: An Introduction for the Practicing Physician*, Ann. Intern. Med. 17 February 2004;140(4):296-302

Medical Devices Certification, GmbH; *Basic Information about the European Directive 93/42/EEC on Medical Devices*, 2009. Accessed August 20, 2012 at http://www.mdc-ce.de/downloads/040100_06_e.pdf

MIL-STD-1629A, Procedures for Performing a Failure Mode, Effects and Criticality Analysis, superseding MIL-STD-1629 (1984) and MIL-STD-2070 (1977), US Department of Defense, 1980, Accessed November 27th, 2012 at http://www.fmeainfocentre.com/handbooks/milstd1629.pdf

NASA, Procedure for Failure Mode, Effects and Criticality Analysis (FMECA), 1966, RA-006-013-1A, retrieved November 15th, 2012 at http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19700076494_1970076494.pdf

NRC, *Fact Sheet on Medical Use of Radioactive Materials*, 2011. Accessed August 22, 2012 at http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/med-use-radactiv-mat-fs.html

O'Grady J., Dobbs-Smith I., Walsh N., Spencer, M., *Medicines, Medical Devices and the Law*, Cambrige University Press, 1999

Pennsylvania Patient Safety Authority, *Errors in Radiation Therapy, Pennsylvania Patient Safety Advisory*, 2009

Richards E.P., *The Supreme Court rules on medical device liability--or does it?,* IEEE Eng Med Biol Mag. 1997 Jan-Feb;16(1):87-8,90.

Samuel F., *Safe Medical Devices Act of 1990*, Health Affairs, 10, no.1 pp 192-195, 1991

Shafiq J, Barton M, Noble D, Lemer C, Donaldson LJ., *An international review of patient safety measures in radiotherapy practice*, Radiother Oncol. 2009 Jul;92(1):15-21. Epub 2009 Apr 22.

Schwartz V., Silverman C., *Preemption of State Common Law by Federal Agency Action: Striking the Appropriate Balance that Protects Public Safety*, Tulane Law Review, Vol. 84:1203-1232, 2010

Stewart R., *Regulatory Compliance Preclusion of Tort Liability: Limiting the dual-track System*, Georgetown Law Journal; Jul 2000; 88, 7; pg. 2167

Stringfellow M., Accident *Analysis and Hazard Analysis for Human and Organizational Factors,* MIT PhD thesis, 2010

Syring G., *Overview: FDA Regulation of Medical Devices*, 2003. Accessed August 20, 2012 at http://www.qrasupport.com/FDA_MED_DEVICE.html

Thomas J.*, Extending and Automating a Systems-Theoretic Hazard Analysis for Requirements Generation and Analysis*, draft thesis, expected Fall 2012

Tran V., *The Radiation Control for Health and Safety Act of 1968: History, Accomplishments, and Future*, HLS Student Papers, 2006. Accessed August 22, 2012 at http://nrs.harvard.edu/urn-3:HUL.InstRepos:8846732

Vesely W., Goldberg F., Roberts N., Haasl D., *NUREG-0492: Fault Tree Handbook*, Office of Nuclear Regulatory Research of the US Nuclear Regulatory Commission, 1981. Accessed at http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0492/sr0492.pdf

Vincoli J., Basic Guide to System Safety, John Wiley & Sons, Inc. Hoboken, New Jersey, 2006

Yeung T.K., Bortolotto K., Cosby S., Hoar M., Lederer E., *Quality assurance in radiotherapy: Evaluation of errors and incidents recorded over a 10 year period*, Radiotherapy and Oncology, Volume 74, Issue 3, March 2005, Pages 283–291

Wikipedia, *European Economic Area*. Accessed August 22, 2012 at http://en.wikipedia.org/wiki/European_Economic_Area

Wikipedia, *Failure mode, effects, and criticality analysis*. Accessed November 15th, 2012 at http://en.wikipedia.org/wiki/Failure_mode,_effects,_and_criticality_analysis

PAGE INTENTIONALLY LEFT BLANK

3 Test Case: Proton-Therapy at the Paul Scherrer Institute

"The art of cancer treatment is in finding the right balance between tumor cure and injury to normal tissues."

Goitein, 2008

3.1 Introduction

The project presented in this chapter applies STPA to the safety review of PROSCAN's Gantry-2 treatment area. It is a case of mostly post-design analysis, where the great majority of architecture choices had been made before the STPA analysis was performed. Nonetheless, it appeared that the STPA framework, although originally meant for application at the design stage, was flexible enough to accommodate this post-design situation.

After providing some background information on cancer treatments and proton therapy, the introduction section describes the set-up of the PROSCAN STPA project. The following section describes the analysis that was performed, and the last one compares the STPA results with the Safety Report put together by the PROSCAN design team for the Swiss regulatory authorities tasked with authorizing the commissioning of Gantry-2.

3.1.1 Cancer Treatments

Cancer appears in one out of three human lives (Tubiana, 2009). It causes one out of every four deaths in the United States, a country where two out of five people will get cancer during their lifetime (Goitein, 2008). Although cancer occurrence increases with age, cancer is diagnosed at all stages of life and at increasingly earlier ones. As Michael Goitein, one of the pioneers of proton therapy at the Harvard Cyclotron Laboratory in the 1970s (AAPM, 2012; ASTRO, 2009), puts it: cancer is important to all of us (Goitein, 2008).

Cancer is a disease characterized by uncontrolled cell division, and the ability of the generated cells to migrate and spread throughout the human body (metastasis). This uncontrolled growth is the result of faulty cell-division commands, themselves a product of gene alteration in the parts of the genome responsible for controlling cell-division mechanism. Although the causes and mechanisms through which faulty mutations occur and survive in a cell's genetic material despite the vigilance of anti-oxidative barriers, phase II enzymes (capable of detoxifying carcinogenic elements), gene integrity control and repair processes are not yet fully understood, it appears that a majority of cancers are caused by changes in the cell's DNA imposed by environmental carcinogens or internal oxidative stress (such as resulting from inflammation, as is the case for asbestos induced cancer) and that a minority have a hereditary genetic basis (Cah. Nutri. Diet, 2001).

From lack of therapeutic means to directly restore the integrity of faulty DNA sequences, as well as from the necessity of removing abnormal tissue whose uncontrolled growth otherwise cannibalizes resources away from normal cells (blood flow, and thus oxygen, nutrients and waste evacuation capacity) and may exert damaging physical pressure on critical organs and structures (such as the optical nerves in the case of brain cancer), treatment consists in either removing (surgery) or destroying (chemotherapy, radiotherapy) unhealthy cells in the hope of eradicating the source of malignant growth.

Better understanding of cancer cells' biological specificities is key in improving treatment selectiveness for tumors that cannot be completely removed by surgery, either because they have grown too close to (or even within) critical body structures (e.g. the stem chord) or because they have spread so diffusely that surgery would not be practical (e.g. in the bones, the limbic and blood systems). Tumor cells have indeed been shown to have different metabolism and growth mechanisms than normal cells. For example, they grow more capillaries to access the blood supply system at higher flow, a mechanism that enables them to access the increased resources that their accelerated growth requires. Their metabolic activity is superior to that of normal cells. They divide at a faster rate. Their DNA repair mechanism is often deficient. They may exhibit different external cell markers. Treatment that is able to target these abnormal behaviors will reduce treatment induced damage to healthy cells.

3.1.2 Radiotherapy

By breaking chemical liaisons, radiation modifies molecules that are necessary to cell life and can cause lethal damage to cell. Highly reactive radicals can thus be formed in the intracellular material that can oxidize proteins, oxidize lipids in cell membranes, and break one, two, or the liaison between DNA strands, causing a cell to lose its ability to reproduce. The higher the dose, the greater the probability of sterilizing cells. Such damage is experienced both by the malignant cells one is trying to eradicate, and by the cells in the healthy tissues that receive radiation even though one would wish to spare them (Goitein, 2008; Association SOS Irradies 31, 2007).

Both chemotherapy and radiotherapy take advantage of the fact that cancer cells, whose focus is on fast rather than sustainable growth, are, in average, both more sensitive to pharmaceutical and radioactive attacks on their vital functions than healthy cells, and less able to recover from damage because of impaired repair functions. The reasons for this difference are complex, not completely understood and even controversial (Goitein, 2008). One such explanation is that a cell's genetic material is most vulnerable during cell-division (mitosis) since it is then completely exposed, away from the protection otherwise provided by histones. Since cancer cells divide at a much faster rate than normal ones, the probability of them being in mitosis at a given treatment time is all the larger, increasing the likelihood that treatment will affect them more than it will a normal cell.

Radiotherapy most commonly use high-energy photons (X-rays, gamma-rays). It can also administer radiation in the form of accelerated particle beams (pions, electrons, protons, heavy ions such as carbon). The range of these energetic particles in tissue depends on the mechanisms through which they lose energy by interacting with matter. These vary according to the particles' nature (photons penetrate further than charged particles; electrons themselves penetrate further than the much more massive protons) and energy, as well as on the electronic density of the penetrated material. As can illustrate common experience with X-ray imaging, the human body can, with respect to density, be characterized by bone (higher density, more energy absorption per distance traveled) vs. mostly water (lower density, less energy absorption).

The purpose of a radiation-therapy machine is to generate therapeutic particle beams and direct them to a patient's tumor with as little overlap as possible with healthy tissue. Radiation therapy

has been proven exceptionally effective in in-vitro experiments (Levin et al, 2005). However, in-vivo outcomes cannot reach these high theoretical limits because of such constraints as necessity to deliver sub-optimal doses when critical structures would otherwise receive higher than acceptable collateral radiation doses, and impossibility, for some tumor types, of delimiting a clean target boundary, as unhealthy cells spread out from the tumor's center into healthy tissue at lower and lower concentrations, always creating a risk that they will form the seed for future growth.

3.1.3 An Example of External Charged Particle Beam Therapy: Proton-Therapy

According to the Proton Therapy Cooperative Group, 96,500 patients had been treated with charged particles as of the end of 2011, 83,600 of whom with protons (PTCOG, 2012). As of the end of 2011, 49 centers were active worldwide, 10 of them in the USA.

The idea that energetic charged particles such as accelerated protons could be used for medical therapy was expressed as early as 1946 by Robert Wilson (Wilson, 1946). Proton treatment was introduced at the Berkeley Radiation Laboratory in 1954. Harvard University and the Massachusetts General Hospital began using protons for cancer treatment in 1961 at the Harvard Cyclotron Laboratory. The first hospital to offer treatment was the Loma Linda University Medical Center who opened the Conformal Proton Beam Treatment Center in 1990 (ASTRO, 2009).

While photo-electric, Compton scattering and pair-production interactions (NSE, 2012) cause a photon (X or g-ray) beam to lose energy near exponentially as it travels through matter (Goitein,2008), the more massive charged particles follow a very different interaction pattern. They are subject to Coulomb interactions with the electrons of the matter that they traverse, and are “invisible” to matter until they have slowed down enough to interact with electronic orbits, the point at which all of their remaining energy is released (OCW, 2004). As shown in Figure 3, this phenomenon allows energy deposition to be concentrated in a narrow “Bragg peak”. Energy is deposited into a smaller volume of the receiving tissue than achievable by conventional radiotherapy treatments, thus protecting healthy tissue around the cancer tumor from undesired irradiation, as confirmed by the imaging data presented in Figure 4.

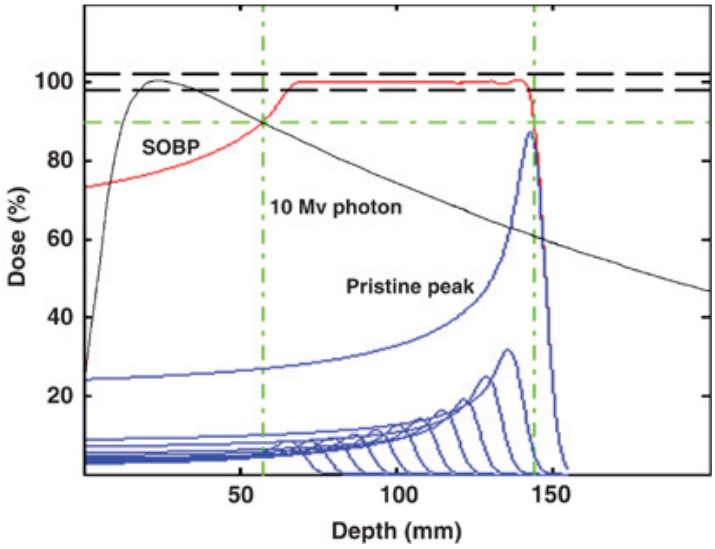


Figure 3 - Dose deposition for proton and photon beams as a function of tissue depth
Source: (Levin et al, 2005, originally published by Nature Publishing Group)

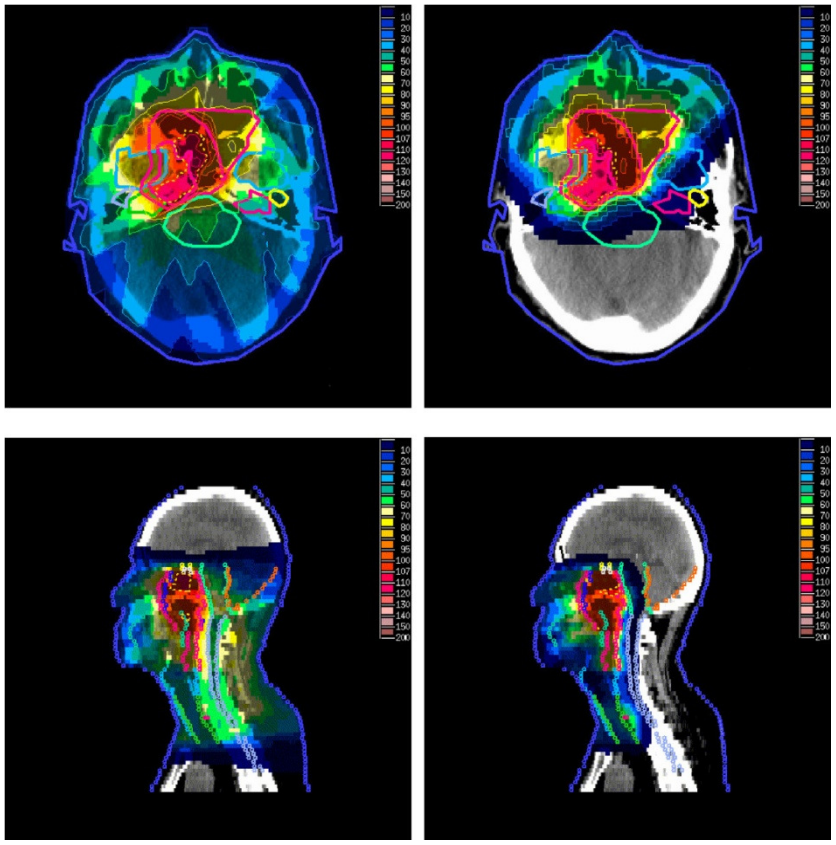


Figure 4 - Irradiation of nasopharyngeal carcinoma by photon (IMRT X-ray) therapy (left) and proton (IMPT) therapy (right)
Source: (Taheri-Kadkhoda et al, 2008, originally published by BioMed Central)

Although the unique deposition pattern of proton energy in matter makes protons a very attractive vehicle for tumor cell damaging radiation[32], the important costs associated with the building of necessarily large proton-delivery facilities that include complex control systems have thus far limited the deployment of this technology[33]. As a consequence and despite favorable therapeutic results[34] on specific indications and benefits expected from expanding the range of indications, especially for young and pediatric patients who benefit the most from radiation sparing of their healthy tissue, proton-therapy is under fire in the USA as driving healthcare costs to unnecessary heights. Albeit it is not the purpose of this dissertation to explore the US health-care landscape and the situation created by the interaction of several strong interest groups, recent press outcry comparing the creation of new proton-therapy facilities to a medical arms-race deserves a short detour (Emanuel and Pearson, 2012; Langreth, 2012).

Proton therapy will not be worth its financial effort if it cannot treat cancer with higher success rates and lower side effects than the currently available treatment options. Medicine, as a science, is constrained in its collection of evidence by strong ethical bounds and very practical concerns, such as the controllability of patient characteristics, the availability of experimental subjects at the times of trial, the variability of procedures and treatment protocols[35] from one study center to the next, or the long time to evaluate the effects of certain treatments on morbidity and mortality. As a result, while evidence for treatment efficacy is often available, evidence for treatment superiority that ought to be obtained from controlled comparison of several treatment options can be difficult to assemble, and will take time to accumulate.

---

[32] Proton energy is deposited at a constant rate until the proton has slowed enough that it can interact with matter, resulting in the release of beam energy in a spatially narrow Bragg peak. Not only are areas ahead of the target submitted to less radiation stress than in conventional photo-therapy (X and gamma ray), but areas behind the target volume are left free of all energy deposition. The depth at which the Bragg peak occurred depends on initial beam energy.

[33] As of early 2010, some 67,000 patients had received proton therapy in 30 facilities worldwide (PTCOG, 2010). Two years later, their number had risen to 83,700 (PTCOG, 2012)

[34] In the past 5 years, PROSCAN has achieved a 98% cure rate for OPTIS 2 patients and 81% for Gantry 1 patients (personal communication by T. Lomax, May 17th, 2011)

[35] For example, in the case of proton therapy, parameters such as whether the proton beam needs to be collimated (with more or less secondary neutron generation and correlated neutron irradiation of the patient) or not, the fractioning regimes, the total dose delivered to the tumor may vary from one center to the next, while treatment planning (field directions), contrary to drug treatment protocols, will necessarily be patient specific.

(Olsen et al, 2007), a review of 54 publications reporting clinical efficacy results for proton therapy, observed in 2007 that "*the evidence on clinical efficacy of proton therapy relied to a large extent on non-controlled studies*" and were thus associated with "*low level of evidence according to standard health technology assessment and evidence based medicine criteria*" for all tumor sites.

That same year, a Proton Task Force was assembled by ASTRO to evaluate the state of the art in proton beam therapy (PBT). Collecting data until November 2009, it agrees in (Allen et al, 2012) that "*current data do not provide sufficient evidence to recommend PBT in lung cancer, head and neck cancer, GI [gastro-interinal] malignancies. and pediatric non-CNS [central nervous system] malignancies*". In hepatocellular carcinoma and prostate cancer however, it found evidence for the efficacy of PBT "*but no suggestion that it [was] superior to [existing] photon based approaches*". Most encouragingly, PBT appeared superior to photon approaches for pediatric CNS malignancies (more data is needed) and evidence existed that PBT was superior to proton approaches for ocular melanoma and chordoma.

Later still, Bouyon-Monteau (2010) felt entitled by the accumulation of evidence and the best practices it observed to claim that [60-250 MeV] proton-beam therapy is now widely accepted as the "*gold standard*" for treatment of specific indications in adults - ocular melanoma, chordoma already mentioned in the ASTRO study, and chondrosarcoma of the base of skull - and is regarded as a highly promising treatment modality in the treatment of pediatric malignancies[36] (brain tumors, sarcomas...). However, these indications only represent a small fraction of all cancer diagnostics. Once a facility as expensive as a proton therapy center [37] has been built – and

---

[36] "*One of the most important advantages of proton therapy is its capacity to kill tumor cells while sparing healthy tissue nearby, such as in the brain or other vital organs, from the adverse effects of radiation exposure. Traditional radiation in pediatric brain tumors has been associated with long-term neurocognitive deficits including decreases in IQ, difficulties with attention, processing speed and other executive skills. Also, even low dose radiation to glands in the brain may have a life-long detrimental effect on hormone production and growth. Protons have the ability to target tumors with high precision and have no exit dose. The decreased radiation dose outside the tumor is especially critical for children since the risk of secondary, radiation-induced tumors may reach 25% in long term survivors treated with conventional radiotherapy*." (Pediatric Proton Foundation, 2011 press release)

[37] (Emanuel and Pearson, 2012) quote the two proton centers that the Mayo Clinic has announced it will be building cost $180M each. Five years earlier, (Huff, 2007) was quoting figures in the $125-207M range with $15-25M yearly operational costs.

why would it not be built, in an environment where healthcare centers compete for patients, hence for the best physicians, hence for state of the art equipment ?–, the temptation is high to maximize its uptime so that the initially large outlay of capital is made to turn a positive profit faster.

An easy way to do so is to extend the range of indications that are treated in the facility, even when half as affordable and satisfactory radiotherapy treatments (such as brachytherapy or photon therapy) are already available for these indications (Langreth, 2012).

For example, construction plans for 10 proton therapy centers have been recently announced in the USA on top of the 10 centers that currently exist. Several of these centers advertise protons for the treatment of prostate cancers, helped in their marketing campaign by such advocacy groups as the Brotherhood of the Red Balloon (ProtonBob, 2012), a group of patients who were successfully cured of their prostate cancer by proton therapy, or the author of the Affordable Proton Therapy blog on blogspot.com who even encourages fellow patients to go abroad (to Korea) for treatment. Is it only a coincidence that the most common type of cancer, according to the National Cancer Institute of the National Institute of Health, is prostate cancer, with more than 240,000 new cases expected in the United States in 2012 (NCI, 2012)?

Doubt unfortunately exists, as there does not seem to be sufficient evidence that proton therapy provides any distinctive therapeutic benefit for the treatment of this illness:

- (Samson et al, 2010) shows no evidence of benefit from a 2008 meta-analysis, from lack of data;
- ASTRO's review of the literature published until November 2009 and reporting about 2,000 cases of prostate cancer patients concludes that proton therapy is an option for prostate cancer, but that no clear benefit over the existing (and cheaper) intensity modulated photon therapy has been demonstrated (Allen et al, 2012).
- The US Agency for Healthcare Research and Quality follows (Olsen, 2007) to suggest in (Trikalinos et al, 2009) that there aren't enough comparative studies to conclude that proton therapy provides specific benefits as "*no trial reported significant differences in*

overall or cancer-specific survival or in total serious adverse events" among the 243 articles published on the therapy that were included in their meta-analysis.

- (Bouyon-Monteau, 2010) indicates that clinical validation of the use of proton therapy in prostate, lung and hepatocellular cancers is yet to be obtained, after results of on-going clinical studies in dose-escalation evaluations are made available.

Of course, progress in driving down proton therapy costs (e.g. by the creation of smaller accelerators that could be included in the machine's rotating Gantry such as commercialized by the Mevion company) and augmenting its capability (e.g. by allowing for intensity modulation and beam scanning rather than collimation) cannot be discounted. On the basis that technology improvements are driven by larger use (think Moore's law and other frameworks that show scale as being a driver of technological performance), an argument could thus be made that proton therapy, in that it is at least as effective as conventional radiotherapy, should be made available to patients despite its higher cost.

While this controversy introduces many avenues for research in the drivers of health care costs in the USA and the implications of healthcare sector structure on the efficacy and efficiency of patient treatment (e.g. (Hoffman, 2009)[38], (Vu, 2011)[39], (Grutters, 2011)[40]), it is time for us to close this parenthesis and return to our study of the PROSCAN facility at PSI.

---

[38] (Hoffman, 2009) examines the public debate about the cost effectiveness of proton technology from the point of view of ethics to point out how logic flaws (such as the "hopeful principle" and the "automatic escalator" protechnology arguments) can be introduced in debates about new technologies, and how they may impair their implementation even when they want to promote it.

[39] (Vu, 2011) offers a framework to evaluate the health and cost effectiveness of proton therapy (quoted as being 2 to 3 times more expensive to deliver than photon therapy) compared to intensity modulated radiation therapy in the case of radiation therapy of pediatric brain tumors. It comments that more data is needed to "*evaluate the advantages of proton therapy in reducing the dose delivered to relevant parts of the brain to lower the risks of adverse health effects, especially for IQ losses*".

[40] (Grutters, 2011) uses proton therapy as a case study for the application of real options analysis to weighing the costs associated with the introduction of new technology against its effects when the evidence for performance superiority is limited, as is often the case for newly introduced technologies.

3.1.4 Proton therapy at the Paul Scherrer Institute

3.1.4.1 History of charged particle therapy at the Paul Scherrer Institute

The Paul Scherrer Institute was born in 1988 from the merger of two prominent Swiss Federal Research centers: the Swiss Institute for Nuclear Research with expertise in particle physics and material science, and the Federal Institute for Reactor Research, devoted to topics of interest to the nuclear power industry. It currently employs close to 500 scientists supported by more than 700 technicians and engineers. It had a 2011 budget of CHF 365M ($372M), more than 80% of which are provided by the Swiss Confederation (PSI, 2012b). PSI is part of the Swiss Federal Institute of Technology (ETH) domain, along with three other research centers and universities ETH Zurich and EPFL Lausanne. Its research activities concentrate on three main subject areas: Matter and Material, Energy and the Environment, and Health (PSI, 2012c). It also hosts PROSCAN, Switzerland's sole proton-therapy facility for the treatment of specific malignant cancer tumors.

As shown in Table 7 and explained in more details in Appendix 2, the Paul Scherrer Institute has had experience with the use of charged particles beam to treat cancer patients since 1980, when the large on-site cyclotron was used to produce pi-mesons (pions) that were directed at cancer tumors in the Piotron facility. Twenty years ago, building on the success of OPTIS, a small facility dedicated to the treatment of ocular cancers, and using the same cyclotron, the Institute successfully designed Gantry-1, an experimental proton-therapy facility to treat cancer patients by irradiating deep-seated tumors while keeping secondary irradiation of healthy tissue and critical structures at a low level.

The success of this first experimental facility led to the launch in 2000 of project PROSCAN, an initiative to further expand PSI's activities in the field of proton-therapy. Existing user areas OPTIS and Gantry-1 benefitted in 2006 from the installation of a dedicated cyclotron, COMET. A new user area, Gantry-2, is soon to be commissioned (expected: 2014). Learning from the Gantry-1 experience and using faster actuation devices, it aims to make continuous scanning possible and thereby offer new therapeutic options, especially for cancers attached to mobile organs.

Table 7 - History of the use of charged particle beams for cancer therapy at PSI

Treatment Area	Piotron	OPTIS	Gantry 1	OPTIS 2	Gantry 2
First beam	1980	1983	1994/2006	2007	2008
First patient	1980	1984	1996	2010	-
Last patient	1993	2010	-	-	-
Patients until Dec. 31st, 2010	503	5458	655	47	-
Particle	Pions	Protons	Protons	Protons	Protons
Technique	60 converging beams, moving patient table	Fixed beam, single scattering	Rotating gantry (eccentric), spot scanning	Fixed beam, double scattering	Rotating gantry (isocentric), spot scanning and repainting
Magnetic scanning dimensions	0	0	1	0	2
Beam origin	PSI Ring cyclotron (590 Mev)	PSI Injector 1 (72 MeV)	PSI Ring cyclotron (590 MeV); later cyclotron "COMET" (250 MeV) both degraded to fixed energies	Cyclotron "COMET" (250 MeV degraded to 71 MeV)	Cyclotron "COMET" (250 MeV degraded to variable energies)

source: (PSI, 2012a)

Pedroni et al summarize the key features of new treatment area Gantry-2 as follows (2004, 2011a):

- 2D lateral scanning: fast sweeper magnets displace the proton beam laterally in two dimensions at the isocenter. Compared to prior versions/areas, where beam would move in only one direction, this limits the need to (slowly) move the patient table to cases of large tumors (field patching);
- Faster field application:
    - The complete beam line is constructed in such a way that fast energy changes are possible (i.e. fast changes in deposition depth)
    - Fast sweeper magnets are used by the advanced parallel beam scanning mechanism (i.e. fast changes in 2D spot position).
- Repainting: thanks to faster field application, the system will be able to apply the same spot sequence several times in a short amount of time. “Repainting” is a promising

strategy to cope with organ motion and will extend treatment possibilities to new medical indications;

- Smaller spot size: the beam energy modulation system is located right after the exit of the cyclotron, resulting in smaller spot sizes and, therefore, higher precision of the treatment.
- Easier patient set-up: the isocentric layout (no moving floor) allows for easier and safer patient handling and provides more comfort to the patient and medical staff;
- Target position control: state-of-the-art imaging devices (both CT and X-ray system) are available in the treatment room. In addition to in-room positioning, they will offer useful visualization in the context of moving organs.

3.1.4.2 Description of the PROSCAN facility

PROSCAN (see Figure 5 and Appendix 2) is an engineered facility that delivers an accelerated proton beam to four user areas, where the beam is used for either medical treatment of cancer tumors (OPTIS 2, Gantry-1, Gantry-2) or scientific experiments.

3.1.4.2.1 Overview of the facility's architecture

Featured in Figure 5, the PROSCAN facility consists in a set of hardware and software subsystems that create, steer, shape and distribute the beam to four distinct user areas: Gantry-1, Gantry-2, OPTIS-2 and an experimental area. While Appendix 2 offers a more detailed description of the facility, here is a succinct overview of its architecture.

The ion source produces protons by ionizing hydrogen atoms, stripping them of their sole electron by application of a high voltage electric field. The plasma thus created is channeled into the cyclotron, at the heart of which a negatively charged extractor will attract the positively charged hydrogen ions: protons.

Electric fields accelerate the charged and magnetic fields bend their trajectories. In the cyclotron, both effects are combined so that a particle will be accelerated by a series of electric field that she crosses at regular intervals when turning inside the cyclotron's magnetic fields. High speed coordination is requested so that the electric fields directions and bending potential are coordinated to speed up the particle by giving it successive and additional momentum boosts. It is achieved with a high frequency (HF) alternating power generator. Injected with quasi null speed at the cyclotron's center, the protons reach 250 MeV at the cyclotron's outer diameter.

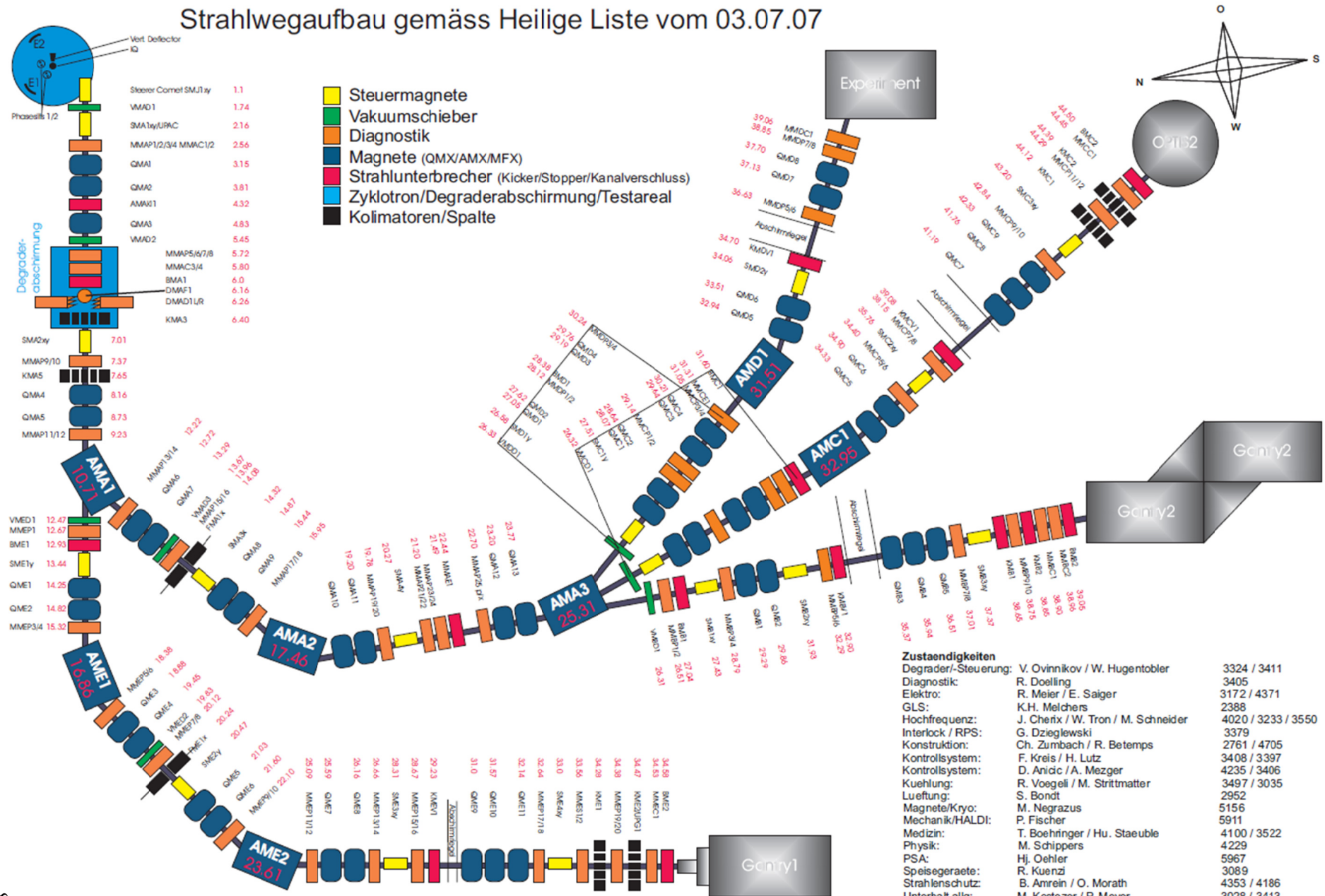


Figure 5 - PROSCAN facility overview

Upon acceleration, protons take wider trajectories in the cyclotron, until reaching the external extraction track. From there, they are directed into the facility's beamline, at a current defined by the ion source setting as well as by current adjustment systems embedded into the cyclotron (phase slits and deflector plate).

A beamline is a collection of devices (magnets, collimators, ionization chambers, dose counters, beam blockers...) whose purpose is to bring the high energy proton beam to the desired target, at the energy and focus that is requested by the facility user. The beamline has four functions:

- transport the beam to the required target,
- set the beam tune and select its momentum spread,
- diagnose the beam quality and
- stop the beam on demand, preventing its undesired entry into specific areas.

In PROSCAN, a shared beamline carries the proton beam away from the cyclotron and through the energy degraded. After the beamline splits in two branches, the beam is sent to each of the four user areas through dedicated local beamline. To make sure that only the user area that is receiving the beam can issue commands to the shared beamline actuators, a beam allocator (BALL) is tasked with controlling beam mastership and allowing only one area to send commands to the shared beamline at a time.

To enable communication of commands to the many elements featured in PROSCAN, a distributed computing environment was deemed necessary to overcome the otherwise complicated wiring issues that a centralized command and control system would have required. Communication between the different elements of the facility transit over an Ethernet network to which VME crates are connected. Their CPU run EPICS, an OpenSource toolkit that is used for distributed computing environments.

Finally, the design philosophy that was followed by the PROSCAN design team was to separate the protection and safety systems from the machine control system. In the words of one such designer, the separation of these duties was key to getting the facility licensed. These protection and safety systems are as follows:

- patient safety system: avoid overdose

- personnel safety system : avoid personnel exposure to radiation
- machine protection system : protect the facility's equipment
- user facility safety systems: protect the local beamlines.

Each of these systems can turn the beam off, and several redundant, escalating mechanisms can be used to do so:

- insert beam stoppers into the beamline or forbid their removal from the beamline
- kick the beam away (deflector/AMAKI magnet)
- reduce the power of the HF source to the cyclotron
- shut down the ion source: used as a last resort, since it can break the ion source by making a hole in the vacuum chamber.

3.1.4.2.2 Treatment essentials: Beam position and Dose application

Both Gantry-1 and Gantry-2 can deliver dynamic spot-scanning treatments while one of Gantry-2's novelties is the introduction of a continuous scanning mode. In both modes, two important treatment attributes must be under tight control: position of the dose deposition, and amount of dose being deposited (i.e. number of particles reaching the tumor).

Position: A tumor is a 3D element. The trajectory of the proton beam and its energy are controlled so that it is stopped inside the tumor. Given that human tissue can be modeled as a homogeneous medium whose interaction properties are very similar to those of water, depth is solely a function of incoming proton energy. Since protons are charged particles, their trajectory can be modified with the use of strong magnetic fields: their location in the 2D plane perpendicular to the beam's direction is defined by sweeping magnets. Since the intensity of the magnetic beam necessary to achieve a certain displacement in beam trajectory is a function of the particles' energy level, magnet setting to achieve a given displacement is a function of the beam's energy. Settings to achieve 2D position are therefore a function of desired depth and these parameters cannot be considered as independent.

Dose: The amount of dose that is deposited at the endpoint of the protons' trajectory is strictly proportional to the number of protons that are received in the end volume. In spot scanning mode, the beam 2D position is set, then packets of protons are sent, landing in the same spot. Dose in these conditions is simply the product of irradiation time and beam intensity at the

target. Current, time, proton count or any combination of these can thus be used as control variables for what level of dose is being locally delivered, as long as their counterparts can be accurately measured. In Gantry 1 (spot scanning), dose application is controlled by dynamically defining spot duration ("time") in such a way that the spot is interrupted when the number of protons calculated as being needed to achieve that dose in-situ has been counted by counters directly upstream from the patient ("proton count"); such a process is theoretically indifferent to current level. Note: practical considerations on the capacity of the monitors to accurately count protons at very low and very high currents are taken into account when defining the operationally acceptable range of current intensities.

In continuous scanning however, treatment speed will be increased by eliminating these beam interruptions: beam is swept in the 2D space at constant or varying speed, and beam intensity is modulated continuously. Dose then is the integral of sweeping speed and beam intensity over time. This relationship introduces the necessity of real-time coupling of sweeping speed and beam in the case of continuous scanning. Dose delivery control can be achieved by intervening on current level when dose application duration is known (e.g. if the sweeping speed was to be held constant, spatial dose variations would be achieved by varying current levels to create the desired dose map) or by modifying dose application duration ("time"), in this case defined by beam sweeping speed, when current is known (e.g. assuming beam current can be held constant at the Gantry, sweeping magnets speed would be varied to modulate dose deposition)[41].

[41] Ensuring either constant speed or current makes for simpler control loops: if either of these (current level or sweeping speed) is allowed to fluctuate, a feedback control loop must be created that would instantaneously correct commands affecting the dual control variables (resp. sweeping speed or current level). Further, setting one of these variables to a constant level can facilitate safety interventions: an irregular deviation from the equilibrium value target could promptly trigger treatment interruption through generation of an interlock command.
While fast intensity modulation is achievable through direct control over the vertical deflector plate, sweeping magnets have slow response times and can't sweep at all speeds. Choosing the constant speed option further makes it possible to keep the sweeping speed at its maximum value throughout treatment, thus minimizing treatment duration.

3.1.4.2.3 Treatment definition and delivery: a set of commands to be issued and implemented

During facility operations, numerous decisions have to be made to ensure that the beam is created and delivered as intended on one hand, and without inflicting damage to either the physical elements or the mission pursued by the beam user on the other.

First, based on the dose deposition map intended for either experiment or therapy, a set of commands must be defined; they correspond to the choice of a facility setting (therapy, diagnostic[42] or experimental modes that, in turn, will make certain commands and settings available or not to the user), and to the specification of a particular sequence of beam characteristics (energy, intensity, duration and trajectory definition of spots or beam continuous painting over time).

This set of commands, known as the treatment plan, must then be communicated to the facility so that the PROSCAN elements are properly configured to execute the operations plan (including the preliminary performance of several actions such as activation of safety systems, restrictions in command and messaging capability, setting of beam sensors...).

Finally, the commands must be executed so that the clinical intent is achieved.

More formally, the delivery of radiation therapy consists in the following steps:

- cancer diagnostic and choice of therapeutic option (surgery, chemotherapy, radiotherapy or a combination of the above)
- treatment planning: identification of the target volumes, identification of location specific dose constraints (e.g. limit dose received by critical structures to certain clinically defined levels), choice of the dose-map to be administered, definition of the actuators' settings schedule necessary to obtain the dose-map.
- treatment delivery: administration of the dose-map deemed clinically relevant
- follow-up surveillance and potential treatment reassessment.

[42] The diagnostic mode keeps the patient safety system in therapy mode but sets the run permit to experiment mode. This allows the operator to insert more sensors (while still having a "beam ready" condition from the Run Permit System) or control specific beamline elements from the otherwise unavailable user interfaces.

Errors in any of these steps can have disastrous consequences and each must be carefully evaluated to eliminate or, when elimination is not possible, reduce and mitigate hazards they can be associated with.

This study assumed that the diagnostic step has already been accurately performed, and focused on the activities that take place at PROSCAN: treatment planning and treatment delivery. While keeping in mind that there are three other beam client areas, it specifically studied the provision of radiation therapy to cancer patients in the Gantry 2 local area.

3.1.4.3 Motivation for the PROSCAN STPA project

High-energy radiation is a powerful tool to eliminate undesired cells, such as those growing in a cancerous tumor. These effects, proportional to the dose received, are however just as effectively devastating on healthy cells. A treatment error sending radiation dose to healthy body parts or resulting in the delivery of higher than needed doses to the tumor target can therefore have highly debilitating and irreversible consequences on radiotherapy patients as was unfortunately illustrated in the radiotherapy accidents described in the introductory section of this thesis. This radiation hazard highlights the importance of integrating safety considerations in the design of radiation equipment and operational procedures.

Safety analysts face the challenge of understanding both the technical features of the machine to be designed and how users will respond to these technical features if they are to design safeguards that will allow for continuously safe operations. In applications where performance requirements (such as fast time responses and high-intensity beams) dictate reliance on digital equipment whose processing mechanisms end-users have little control on or information about, and that are characterized by the necessity of multiple interfaces smoothly sharing information and being aligned in goal, STPA's contribution to evaluating system hazards through modeling of the system as a hierarchy of control issuing entities and information sharing channels appears as potentially very fruitful. A research project was put together to test this hypothesis, using PROSCAN's Gantry-2 treatment area as a test-bed on which to apply STPA.

3.2 The PROSCAN STPA project

3.2.1 Purpose and goals of the PROSCAN STPA project

Proposed by Dr. Christian Hilbes (School of Engineering of the Zurich University of Applied Sciences), the PROSCAN STPA project was supported, approved and funded by PSI management. Launched in April 2011, it proposed to use STPA to identify the hazards associated with the operation of the Gantry-2 proton-therapy unit designed at PSI as an example of how this hazard analysis technique, most potently used at early stages of system design, could also be applied to the hazard review of design proposals at late stages of the design development phase. Although this endeavor was distinct from the commissioning process, the project team included members of the Gantry-2 design team who were active in putting together the application for Gantry-2 licensing by the Swiss authorities, ensuring that potential findings from the STPA analysis would be available to the team putting together the risk evaluation for licensing purposes.

The project goals were formulated as follows:

- Perform a hazard analysis of the proposed proton-therapy facility, whose results could be compared to the existing safety report used by PSI to communicate with the Swiss commissioning authorities
- Further establish the applicability, feasibility, and relative efficacy of using STPA to perform hazard analysis of complex systems
- Transfer the STPA know-how to PSI, with the result of enabling independent use by PSI staff.

3.2.2 Project description

The project assumed that the facility designers were seeking market approval from a regulatory authority that accepts hazard analyses performed using STPA as a means to identify and document a risk analysis. Ideally, the design team would have followed the Safety Driven System Engineering Design Methodology described in (Leveson, 2012) and (Stringfellow et al, 2007). As a result, the STPA project followed a process similar to the initial steps of Safety Guided Design. It started with the identification of mission goals, high level requirements and constraints, moved on to model the facility as a control structure, and finally performed the STPA hazard analysis.

Intended as a demonstration project rather than as supporting the application of Gantry 2 for commissioning, this study put more effort into identifying and solving methodology issues than into providing a complete view of the risks associated with the design under consideration. The scope of the analysis was therefore defined over time as encompassing five sub-systems rather than the whole facility, and the depth with which each example was investigated depended on the lessons that the analysis team wanted to learn from them.

As described in the background chapter, STPA is grounded on the notion that system behavior must be controlled to avoid the realization of hazardous situations. Once the system of interest is described in control terms, STPA consists of 2 steps. First, STPA Step 1 identifies unsafe control actions whose realization would put the system into a hazardous situation. Then, STPA Step 2 identifies scenarios that could lead to these unsafe control actions. Mitigation measures can be discussed in view of these findings, either as preventing (or reducing, detecting and mitigating) unsafe control actions, or as preventing (or reducing, detecting and mitigating) the occurrence of scenarios through which they can occur, depending on how difficult each of these endeavors are.

After having performed a high-level analysis of the facility's control structure, the project team selected five Gantry-2 sub-systems to perform a detailed analysis on. For these, the team identified and documented the emerging safety issues that are created by the association of physical (e.g. proton beam lines), digital (e.g. computer control of guiding magnetic fields) and human components (e.g. the specialist entering the treatment work plan in a first user interface, the radiotherapy assistant delivering treatment to the patient through a second interface, the technicians building and maintaining the machine…). The results of these analyses are documented in (PSI, 2012a).

These five sub-systems were chosen to be representative of the typical natures that controller-controlled process interactions can take. The depth of analysis was selected so as to provide clear insight into the risks associated with operation of the facility at a fine-grained level of detail while making do with the limited time and human resources that were available to work on this project. Several ways of performing STPA Step 1 and 2 were experimented with (see Lessons Learned in Chapter 4). These choices are summarized in Table 8.

Table 8 - Gantry-2 sub-systems on which STPA was performed

Ex.	Example name	Relationship type	Step 1 "traditional"	Step 1 Thomas	Step 2 "process loop"	Step 2 "Step 2 Tree"	Workshop
1	Nurse	Human-Human		X			
2	Local Operator	Human-Machine		X		X	
3	Dose Controller	Machine-Machine spot mode		X		X	
4	Sweep controller	Machine-Machine continuous mode				X	
5	Mixed spot & continuous sweeping	Machine-Machine	X		X		X

3.2.3 Project results

The detailed STPA analysis that was performed in the PROSCAN STPA project is documented in (PSI, 2012d). This section presents how the PROSCAN facility and its users were modeled in a control structure, what high level system hazards were considered, and, for illustrative purposes, the detailed results for one of the five examples performed during the PROSCAN STPA project. Although Table 22 and Table 23 list all of the unsafe control actions and hazardous scenarios identified for the five examples, the reader is referred to (PSI, 2012d) for details about the other four examples.

3.2.3.1 Mission goals, requirements and constraints

The starting point Safety Guided Design is the definition of

- mission goals that the mission is designed to achieve,
- requirements that the mission must meet to achieve its mission goals, and
- constraints on the mission that must not be violated in the efforts to achieve the mission goals. These constraints are environmental constraints as well as customer design and programmatic constraints.

Starting from the postulate that a proton-therapy machine will be built (first design decision), the following goals and high level requirements were identified for the Gantry-2 system consisting in the personnel, procedures, facilities and equipment that handle the patient from the time that he is diagnosed with cancer to the time when he leaves the Gantry-2 treatment room after he has received full treatment.

Goal

G1. Treat localized cancer tumors

Rationale: although the set of targets considered includes moving ones (e.g. when in a moving organ such as the lungs or the prostate), the goal of this project is to treat localized, non-circulating tumors, not to treat moving metastases.

High Level Requirements

HLR1. The system shall identify the body volumes to which dose will be delivered. (←G1)

Rationale: dose delivery to healthy tissue may bring adverse outcome, depending on the intensity of the dose delivered and the radiological sensitivity of healthy tissue. On the other hand, while tumors do not always have well defined contours, delivering too little dose to tumor cells spread in healthy tissue may result in new tumor growth.

HLR2. The system shall calculate the amount of dose to be delivered to the identified dose delivery volumes. (←G1)

Rationale: same as HLR1.

HLR3. The system shall deliver the calculated dose to the identified dose delivery volumes. (←G1)

3.2.3.2 System accidents

The scope of the STPA analysis of the PROSCAN system was limited to radiation specific hazards, although it was acknowledged that other hazards could create additional losses.

The PROSCAN facility delivers high-energy radiation to patient tumors. As this radiation is harmful to living cells in the human body, it is intrinsically hazardous. STAMP proposes to eliminate sources of hazards from the initial design. However, once the decision has been made that therapy is to be provided using proton beams, the radiation hazard cannot be designed out of the system. It can only be reduced and mitigated.

In these conditions, four types of losses or accidents can occur during the facility's operations:

ACC1. Patient injury or death

ACC2. Ineffective treatment

ACC3. Loss to non-patient quality of life (esp. personnel)

ACC4. Facility or equipment damage

3.2.3.3 System-level hazards

Next, the engineer defines the high-level hazards that could lead to system accidents, prevent mission goals and requirements from being met, or violate the mission constraints.

Radiation-related hazards

H-R1. Patient tissues receive more dose than clinically desirable

Rationale: Patient reception of too much dose may induce negative health effects in both healthy (risk of secondary tumor - see (Tubiana 2009)) and unhealthy tissue (risk of tissue necrosis), as well as prevent delivery of the required amount of dose to unhealthy tissue.

H-R2. Patient tumor receives less dose than clinically desirable

Rationale: A treatment could be ineffective because it does not deliver enough dose to unhealthy cells..

H-R3. Patient treatment is improperly fractioned

Rationale: If fractions are not planned adequately, dose delivery may not be appropriately timed, for example because of longer than desired treatment interruption without treatment redefinition

H-R4. Non-patient (esp. personnel) is unnecessarily exposed to radiation

Rationale: In addition to the negative physiological effects induced by accidental exposure to high radiation doses, the toll that knowledge of one's exposure to lower levels of radiation may take on one's emotional well-being is acknowledged.

H-R5. Equipment is subject to unnecessary stress

Note: not radiation specific. Unnecessary stress was defined as one that is not required for the system to achieve its performance goals. *Such examples of stress include inadequate rotational control of the Gantry around the patient table leading to potential collision and equipment damage, extra wear resulting from cycling (e.g. of the ion source), mechanical damage to sensitive equipment (e.g. vacuum windows) during maintenance, dose deposition in equipment that were not meant and therefore not designed to receive these levels of dose, leading to potential material activation (consequences: higher disposal and remediation costs, possibility to go beyond maximum authorized dose levels in the facility with resulting risk of having to shut down the facility) or loss of equipment functionality (with negative consequences including increased replacement and downtime costs, in addition to potential personnel exposure to facility hazards).*

Other potential hazards

H-O1. Person is hit by moving mechanical equipment

H-O2. Person is submitted to intense magnetic field

H-O3. Person is in proximity to high energy source

H-O4. ...etc.

3.2.3.4 High-level safety constraints (SC-R)

The high-level hazards will guide the identification of lower-level hazardous situations throughout the analysis[43]. They also lead to the specification of high-level safety constraints whose enforcement will prevent the hazards from being realized.

Safety constraints are requirements that eliminate or mitigate the hazards. If the hazard is of the form "Hazardous state occurs", generating the safety constraint involves a simple but important translation from the hazard into an engineering goal.

H-R1. Patient tissues receive more dose than clinically desirable

SC-R1. The system must be able to prevent delivery of higher than clinically desirable dose

H-R2. Patient tumor receives less dose than clinically desirable

SC-R2. The system must be able to deliver sufficient dose to treat the tumor.

H-R3. Patient treatment is improperly fractioned

SC-R3. Each fraction must not exceed more than TBD Gy and must not be delivered TBD' hours after the previous one without treatment plan being reevaluated.

H-R4. Non-patient (esp. personnel) is unnecessarily exposed to radiation

SC-R4. The system must be able to prevent unnecessary exposure of personnel and non-patients to radiation

H-R5. Equipment is subject to unnecessary stress

SC-R5. The system must prevent excessive equipment exposure to radiation.

Similar safety constraints would need to be derived with the non-radiation related hazards.

3.2.3.5 Environment and customer constraints

The next step in the methodology is to identify:

- environmental constraints (C-E) and environmental assumptions (C-A)
- customer-derived system design constraints (C-C), including customer programmatic constraints (e.g. budgets etc.)

Environmental description, constraints and assumptions describe and constrain the environment of the system. They are design independent. *Here would be documented assumptions and constraints that, for example, would convey information on the fact that the facility is to be used by hospital staff... (hence must not require advanced engineering knowledge to be operated and fixed on a daily basis).*

A-E1. The system will be staffed by trained medical doctors, medical physicists, and operators including experienced maintenance staff.

[43] The reverse (that lower level hazards could be higher level hazards) is not necessarily true.

A-E2. The system will be hosted at a physics research center, the Paul Scherrer Institute. Continuous upgrades and fixes will be designed by a dedicated research team.

Customer derived system constraints are constraints on the design of the system that are technical in nature. Typically, they involve how the system must interact with existing resources, engineering mandates, or initiatives the customer wishes to implement. They are critical to document as conflicts between safety and customer constraints must be investigated.

Note: given the intended health application of the facility, safety is also part of what performance is about. Hence the inclusion of dose limits in the customer constraints highlighted below.

C-C1. Whenever the Gantry-2 treatment facility is operational, it must do so without interfering with the successful completion of processes in the other PROSCAN areas.

Rationale: Gantry-2 treatment facility receives beam from a cyclotron whose output is shared with three other user areas (Gantry-1, OPTIS-2 and an experimental area).

C-C2. Healthy body tissue and structures must not receive more than TBD Gy of radiation per treatment and must not receive more than TBD' Gy of radiation per treatment fraction, where TBD and TBD' depend on the radio-sensitivity of given body structures.

Rationale: Patient reception of too much dose may induce negative health effects in both healthy (risk of secondary tumor - see Tubiana 2009) and unhealthy tissue (risk of tissue necrosis), as well as prevent delivery of the required amount of dose to unhealthy tissue.

C-C3. Unhealthy tissue must not receive more than TBD" Gy of radiation per treatment and must not receive more than TBD"' Gy of radiation per treatment fraction, where TBD" and TBD"' depend on the radio-sensitivity of given body structures.

Rationale: Same as C-C2

C-C4. Treatment of a TBD liter tumor volume must be performed in less than TBD min.

Rationale: 1. reduce patient discomfort, 2. enable treatment of tumors associated with mobile organs.

3.2.3.6 High-level functional decomposition

This functional analysis would record not only the functions necessary to meet the system's requirements and constraints, but also the physical and informational interactions between these functions.

The following functions are identified as being necessary to meeting the goal of the facility:

1. Patient selection for treatment
2. Treatment plan definition
3. Treatment plan transfer to executing facility
4. Treatment plan execution
    a. Proton beam creation and delivery, which requires control of four beam attributes:
        i. Proton beam on/off

ii. Proton beam tuning (energy)
iii. Proton beam intensity setting
iv. Proton beam positioning

b. Patient identification and positioning

5. Treatment verification

When performing a hazard review, the system to be analyzed is one in existence. In this case, the review would consist in verifying that all the functions identified as necessary for effective treatment by the verification team have indeed been considered by the design team.

3.2.3.7 Control structures

Following the STAMP framework, the PROSCAN facility was modeled using a functional control structure (see Leveson, 2012 for an introduction to STAMP control structures). It is important to note that a control structure is not a description of the software or hardware architecture, but a representation of the functions that the system must perform and how these functions are related to each other.

The PROSCAN system includes all the people and equipment that contribute to defining and delivering the radiation treatment to cancer patients. the analysis team chose to exclude PSI management, Swiss regulatory authorities, Swiss government (esp. in its funding capacity of such large research projects as the PROSCAN facility), deontological bodies and other such external stake-holders from our system representations on the grounds that the system designers do not have the power to modify them in the course of their design work. Their influence on system evolution and operation is nonetheless acknowledged.

- Controls for the medical profession (such as definition of best practices by professional bodies including dose limits, deontological rules, certification of physicians…) provide guidance, commands and rules to the "Treatment Definition" group,
- Constraints on the design team include funding considerations by the Swiss government,
- Regulation and certification authorities, including institutions responsible for the emission of technical standards, are at work to control the design of the "Treatment Planning" and "Treatment Delivery" sub-system. Audits and inspections constrain "Treatment Delivery" operations.

In the figures below, each box represents a system element; downward arrows represent controls and constraints imposed by upper levels on lower ones; upward arrows represent feedback channels through which information on the controlled processes is transferred to each controller.

↓: control in the form of directive(s) or command(s)

↑: control feedback in the form of state information or sensor measurements

← or →: physical and informational transfer other than control and feedback

□: functional element with the controller of its internal interactions (i.e. the functional element-level interactions)

At the highest level (called D0), the facility is composed of two elements, one defining the treatment which the other has to deliver. At the lower level (here D2, but could as well be D10), it is made up of the smallest elements, functional or physical, that must be analyzed for the system's behavior to be understood.

A full STPA analysis would start by performing STPA Step 1and 2 using the D0 description of the system. It would then zoom into the D0 elements identified to be prone to hazardous behavior by describing them at a higher degree of detail (D1) and performing STPA Step 1 and 2 on this more detailed control structure. This process is iterated until further refinement is not expected to provide benefit in terms of hazardous scenario identification.

D0 - High level description of the PROSCAN facility

Figure 6 (D0) gives a very high level description of the system defined above. It presents the two primary functions that must be built into the system for it to provide adequate therapy to the patient: (1) define the treatment plan characteristics and (2) implement the treatment plan, thus controlling the state of tumor cells in the patient body and, more generally, the patient's health.

The Treatment Definition group provides the Treatment Delivery group with instructions on what energy field is to be deposited in what location inside the patient's body; it may also request new capabilities to be developed and implemented by the Treatment Delivery group. The Treatment Delivery group is in charge of preparing the patient for treatment, and of creating and delivering the beam so that energy is deposited in the patient body following the patterns demanded by the Treatment Definition group. Both of them can adjust their actions as well as, over longer periods of time, their process models and control algorithms based on information that they receive from the patient. The Treatment Definition group also receives feedback from the Treatment Delivery group that it provides commands to.

In this representation, the patient is described as being directly controlled by the Treatment Delivery group: the Treatment Delivery group's function is to perform activities that will act upon the patient's body, ideally delivering an energetic proton beam that will kill the tumor cells and improve the patient's health.

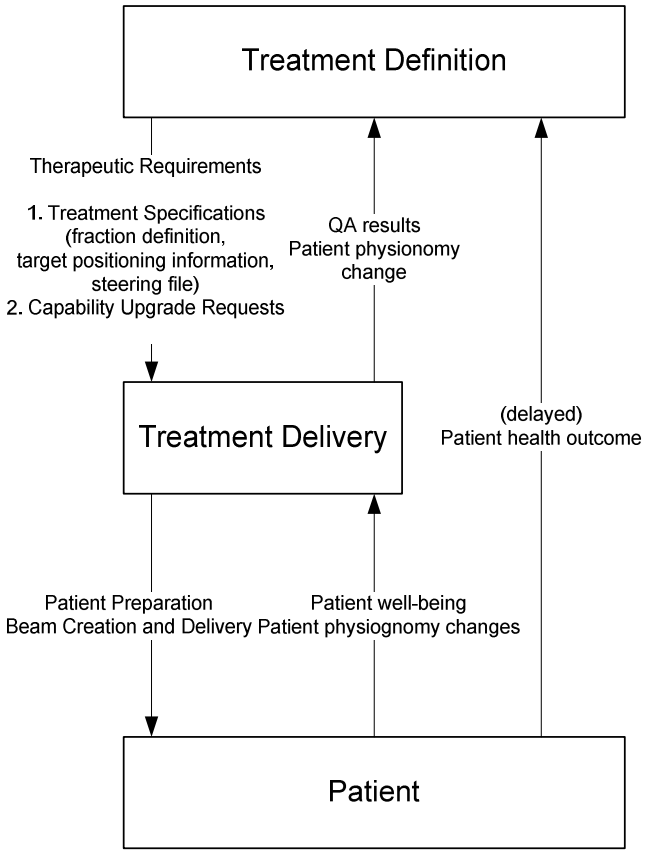


Figure 6 - High-level functional description of the PROSCAN facility (D0)

A full STPA analysis of the PROSCAN/G2 facility would start by an analysis at D0. After unsafe control actions (e.g. "transmission to Treatment Delivery of an incorrect set of therapeutic requirements"), high-level safety constraints would be defined (e.g. "Treatment plan provided to Treatment Delivery must define a combination of radiation fields that deliver TBD Gy to the control volume, and no more than TBD' Gy to the critical structures", where TBD and TBD' are two sets of radiation limits that correspond to different body parts and are defined by scientific consensus) that would have to be enforced by controllers at all "depths" of detail (D1, D2...).

D1 - Treatment Definition

Figure 7 zooms into the Treatment Definition box and describes the control and feedback channels that link its human and automated elements together. The combination of their action results in a patient being approved for treatment, and a steering file (which codes the treatment plan into commands that can be read by the facility's automated controllers) being generated that will include all the information that is needed to create and deliver the desired radiation field to the patient.

Given the slow kinetics of tumor evolution after irradiation, most of the feedback to the Treatment Definition elements about the effectiveness and relevance of the commands they provided is delayed, meaning that it cannot be used for dynamic steering of the lower-level controllers via such means as updates to the steering file. This intrinsic limitation of the feedback channel highlights the importance of creating a delivery system with predictable outcomes, and the need to regularly verify that the Treatment Definition elements' process models are correct, i.e. that they match the reality of treatment delivery. This is achieved at PROSCAN by using regular QA tests.

Detailed explanation of the D1/Treatment definition control structure:

Once a patient has been diagnosed with cancer, his physician (Medical Doctor) may decide to refer him to the PROSCAN facility for treatment. Based on his age, physical condition, treatment history, tumor type and tumor extent, the PROSCAN Tumor Board, composed of physicians and medical physicists, will reserve the limited treatment spots to those patients for whom treatment by proton-therapy at PROSCAN is expected to be most beneficial.

Upon receiving slot approval, the Medical Doctor provides the Medical Physicist with the doses that will be delivered to the treatment volumes that he has identified (tissue volumes that will be irradiated), along with constraints on dose tolerance levels, a list of critical structures and the maximum dose that they should be given.

With assistance from the Treatment Planning Software and information provided by an imaging facility, the Medical Physicist devises a treatment plan that matches the Medical Doctor's prescription and its constraints. Such a treatment plan consists in the identification of several fields whose superposition will create, in situ, the dose distribution requested. A field is characterized by one Gantry position with respect to the patient's body. It consists of the application of a proton beam whose energy and current varies over time.

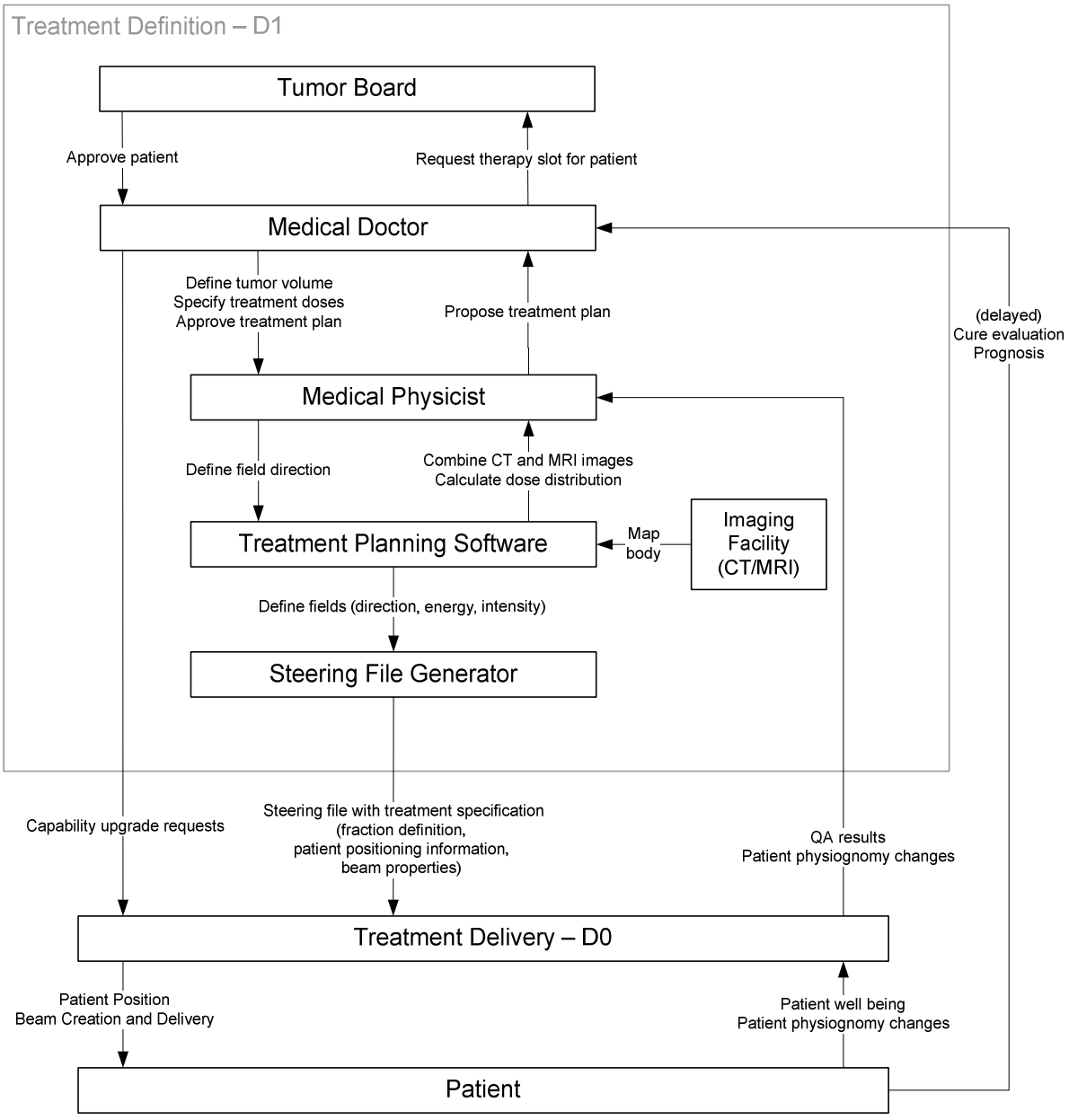


Figure 7 - Refined description of the Treatment Definition group (D1)

The constraints imposed by the Medical Doctor, such as requesting a maximum level of dose to neighboring critical structure while demanding a high level of dose to the tumor's center, may define an impossible set given the precision margins of the Treatment Delivery system and the physics of proton interaction with matter. In that case, the Medical Physicist will request trade-offs to be made. After a few iterations, the Medical Doctor approves a treatment plan.

The Treatment Planning Software is captured as an independent element in this control structure, rather than thought of as a simple tool used by the Medical Physicist, since its logic (=control algorithm) is not only controlled by a different set of actors (design team) but may also be responsible for the definition of an ill-appropriate treatment plan. The Treatment Planning Software translates the desired dose distribution into a list of fields, resulting in a desired sequence of gantry and table positions and beam characteristics to be modulated over time. These fields are applied in different daily fractions to maximize treatment effectiveness.

These commands are sent to the Steering File Generator that transforms them into commands understandable by the facility and its operators. Based on the Generator's designers' model of how the PROSCAN facility creates a beam of certain characteristics and is able to move both the gantry and the patient table in the room referential, the Treatment Planning Software's field and position descriptions are translated into device configurations and settings (e.g. power and voltage to be supplied to a magnet at a given point in time, presets to be configured into dose detectors, information to be observed by potentiometers...). Summarized in the Steering File, this sequence of configurations and settings is then transferred to the Treatment Delivery Group.

D1 - Treatment Delivery

Figure 8 provides more details about the Treatment Delivery group, describing the control and information interactions between its functional elements. Since PROSCAN is an experimental facility, hosted by a research institution and continuously tuned by researchers and scientists who work at improving its capability, the analysis team decided to include the PROSCAN Design Team in this operational control structure. Our description of a mature system whose architecture would have been essentially set (although the possibility of future fixes and revisions would remain) would not have given the Design Team such a role in operations.

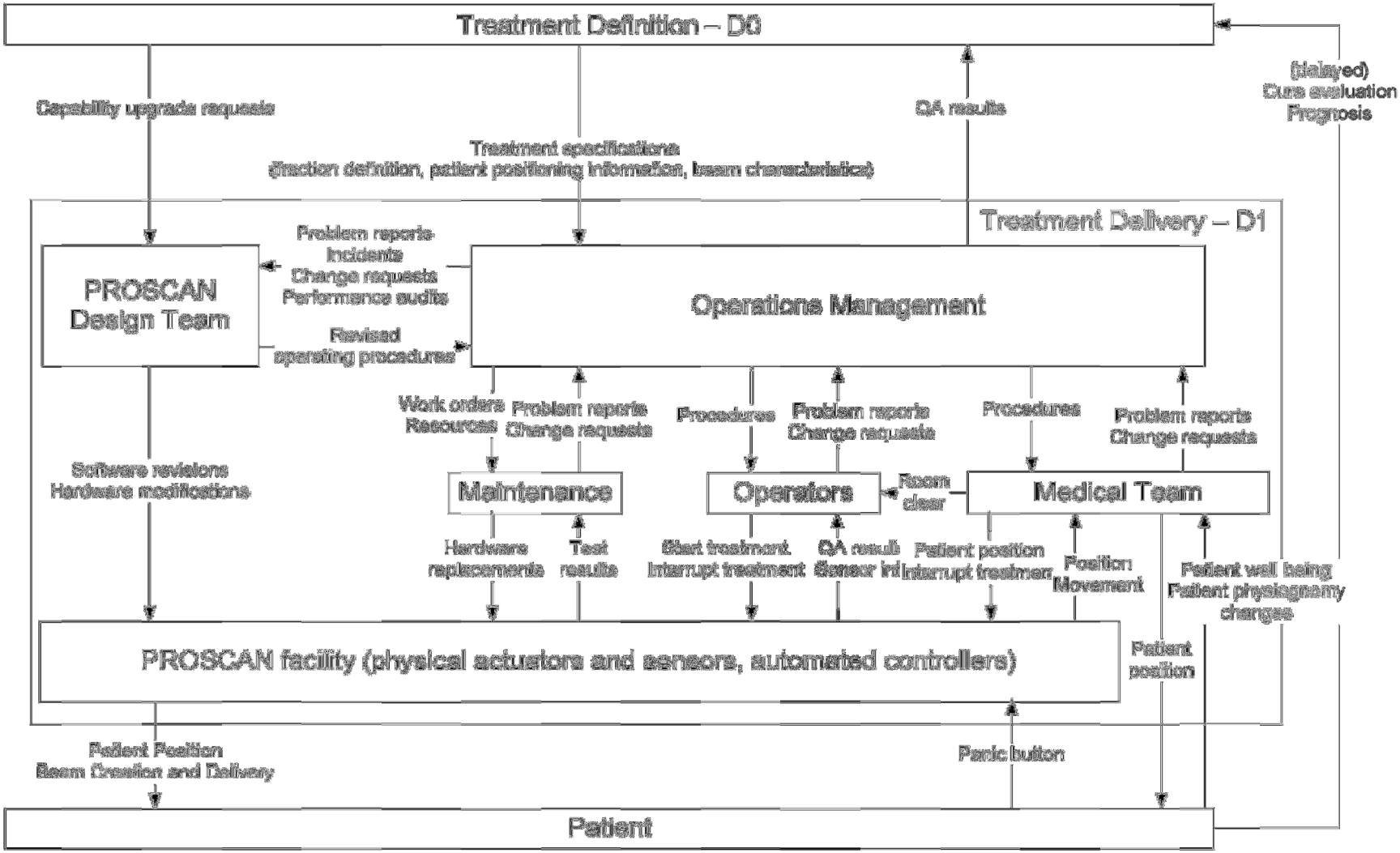


Figure 8 - Refined description of the Treatment Delivery group (D1)

Note: the human activated safety features (panic button from patient to PROSCAN and capability of operators and medical team to interrupt treatment) were retained in this representation. The automated safety features were not.

Previous STPA analyses by academic and industry researchers in various sectors of activity usually aimed at identifying constraints that should apply to the technical system in its entirety, or to its more detailed functional groups. In contrast, one of this project's ambitions was to dive deeper into the physical realization of these functions and assess what contribution STPA can make to their risk analysis. It therefore appeared necessary to describe the PROSCAN facility (bottom element of the Treatment Delivery box of Figure 8) in further detail.

The PROSCAN facility keeps the process of patient irradiation in check by controlling several beam attributes. As the definition of the hazards in 3.2.3.3 shows, negative outcomes to the patient and the system's personnel are the result of either the wrong dose being delivered (too high or too low), or dose being delivered at the wrong place (including in non-patient bodies). Providing safe process outcomes therefore requires that two specific characteristics of the dose delivery process must be well controlled:

- amount of dose delivered
- location at which the dose is delivered.

In PROSCAN, the dose amount is controlled by controlling both the intensity of the beam and its on/off status. The location of dose deposition is controlled by controlling the beam tune (dose deposition depth is a function of beam energy) and the alignment of the beam with the tumor location.

Figure 9 summarizes how irradiation control is therefore obtained by simultaneous control of four basic process attributes: beam on/off status, beam intensity, beam tune (energy and shape), beam to target alignment[44].

Given our objective to apply STPA to lower degrees of abstraction, down to the physical elements through which the functions described above are embodied, we zoomed in further into the loops that control these four basic process attributes. Figure 10 presents the results of this process for the "Beam to Patient alignment" attribute.

[44] These characteristics are not independent one from the other (e.g. beam current is set differently depending on what energy the beam is at; sweeping speed affects both dose deposition and beam alignment with target). However it was helpful to describe the PROSCAN architecture by how it controlled each of these beam attributes separately.

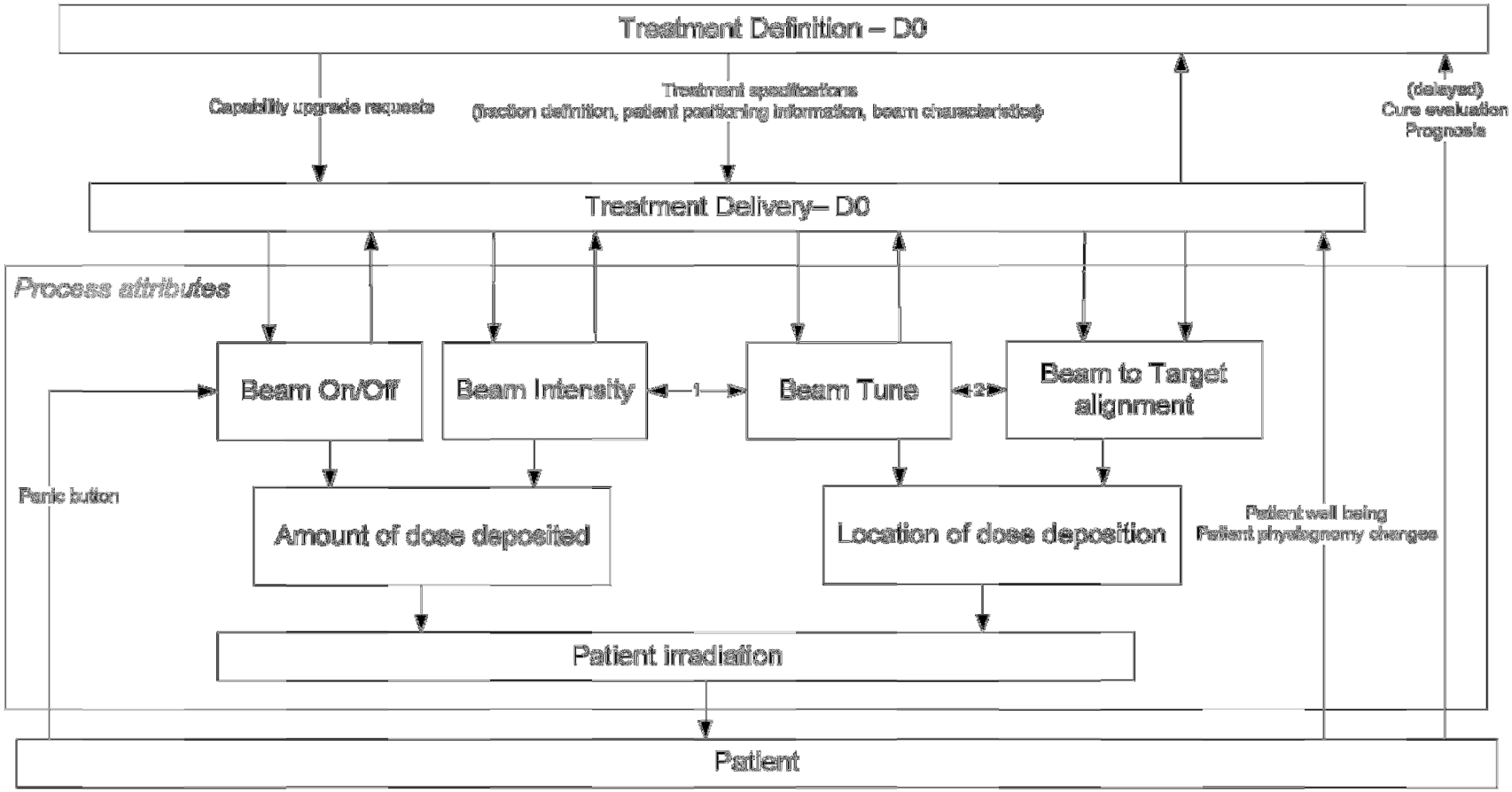


Figure 9 - Patient irradiation process attributes

1: setting of the mechanisms that regulate the intensity of the beam over the beamline (esp. defocusing magnets and collimators) depend on the energy of the protons.

2: setting of the bending and sweeping magnets to achieve a given beam trajectory similarly depend on beam energy.

This representation of the beam attributes includes controllers (e.g. the Treatment Delivery group) and different controlled process, some of which feed into the others. For example, the Beam intensity attribute does not "control" the amount of dose delivered in the sense that it is not a controller that issues commands that will affect the amount of dose that will be delivered, but, along with the Beam on/off status, defines the amount of dose being delivered. This extension of the control structure framework to describe relationships between beam attributes provided the means to abstract the physical realization of the PROSCAN facility into the functions that its various physical and software elements were meant to embody.

"Beam to Target alignment" is achieved by controlling "Beam to Patient alignment" and ensuring that knowledge of the tumor position in the patient body ("Target to Patient alignment") is accurate and well communicated to all the elements that need to be aware of it to take safe control actions. "Beam to Patient alignment" is controlled when the following contributing process attributes are controlled:

- position of the Beam in the Gantry referential
- position of the Gantry + Table system in the room referential
- position of the Patient on the Table.

Several actuators (sweeper magnets, gantry & table motors, the medical team and the patient itself) and sensors (strip chambers, encoders and potentiometers, CT-imaging) are at work in the process loops tasked with ensuring that these attributes take the values that the higher-level controllers (TCS and other Treatment Delivery group controllers) request and expect.

Through this "zoom-in" process, an adequately detailed functional representation of the PROSCAN facility was obtained. It organized PROSCAN's physical elements (human controllers, software logic, actuators and sensors) in a control structure whose process loops (such as the four loops labeled in Figure 10) were then examined to identify unsafe control actions (STPA Step 1) and the scenarios through which they could be realized (STPA Step 2).

3.2.3.8 Perform STPA Step 1 and Step 2 on the Control Structure's Process Loops

While the result of our STPA analysis on the five examples chosen for this study are listed later in this chapter when they are compared to the hazards documented in the PROSCAN Safety Report, this sub-section will provide one of these five examples for illustrative purposes. The control actions that the local operator is responsible for will be analyzed as an example of STPA Step 1 and Step 2 being applied to human interaction with automated systems. This and the other examples can be found in (PSI, 2012d).

3.2.3.8.1 Human controller interacting with automated system: the local operator

This example covers the procedure through which the operator starts a patient treatment. STPA Step 1 was performed using the "Thomas process", defined in (Thomas, 2011). STPA Step 2 was performed on a subset of the potentially unsafe control actions found during Step 1 using the Step 2 Tree described in Chapter 4.

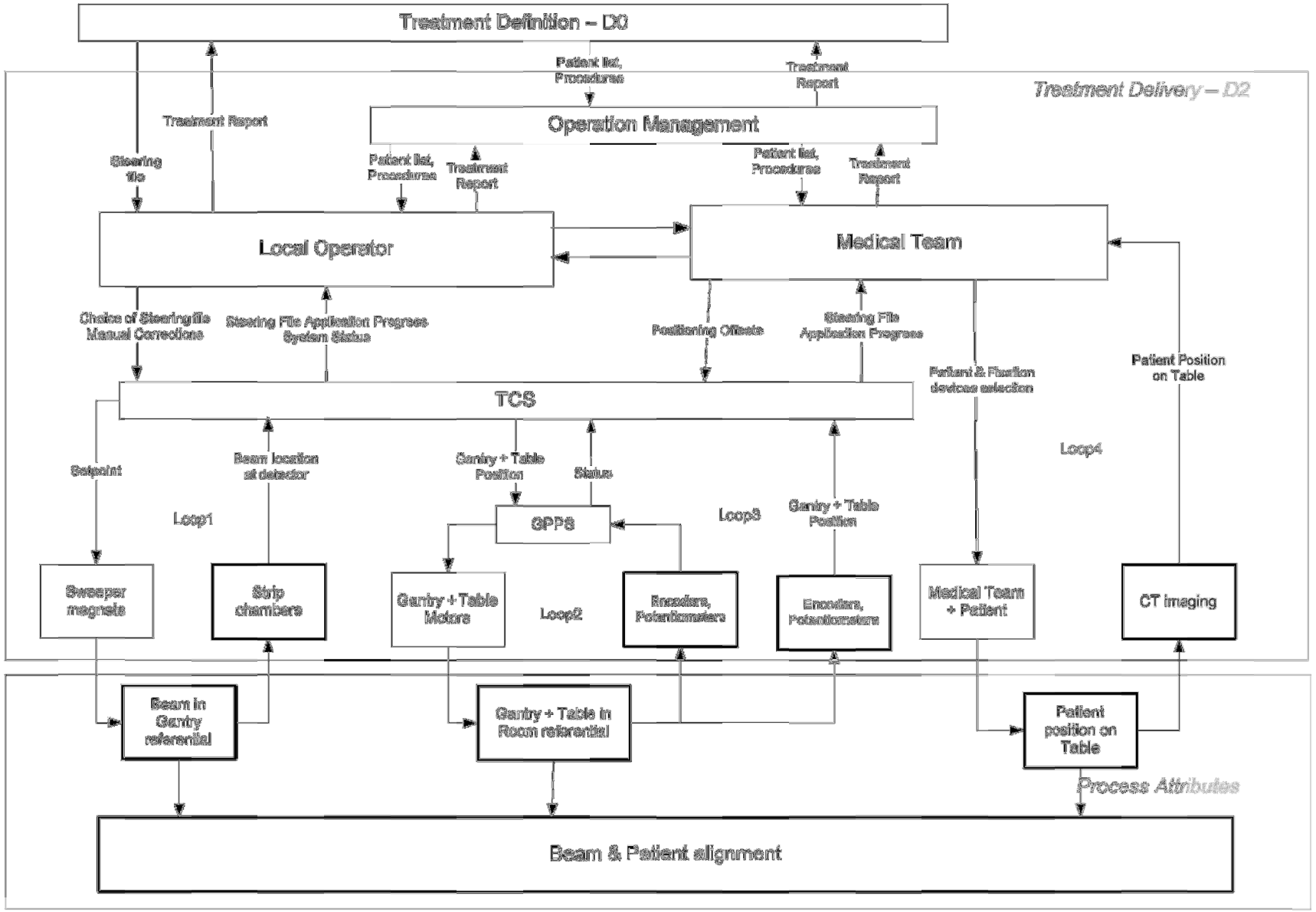


Figure 10 - Control loops for the Beam & Patient alignment attribute (D2).

TCS refers to the Therapy Control System and GPPS is the Gantry & Patient Positioning System. Both consist of software logic that is implemented by different circuit boards hosted on the PROSCAN premises.

3.2.3.8.1.1 Role of the operator in view of radiation related hazards

The operator is in charge of preparing the machine control system for each new patient treatment, and of starting their treatment. As such, he is tasked with controlling that the applied steering file matches that corresponding to the patient-specific treatment plan, and that the treatment progresses without complication. He contributes to preventing the delivery of dose to the wrong person or to the wrong location, and to preventing the delivery of the wrong dose (H-R1, H-R2, H-R3).

As shown in Figure 11, the operator does not exercise direct steering of the treatment process. The actuation speed indeed requires a cognitive ability far beyond human capability. Nonetheless, the operator is empowered with the capacity to interrupt treatment were he to witness unsafe changes to the system state (e.g. patient moving).

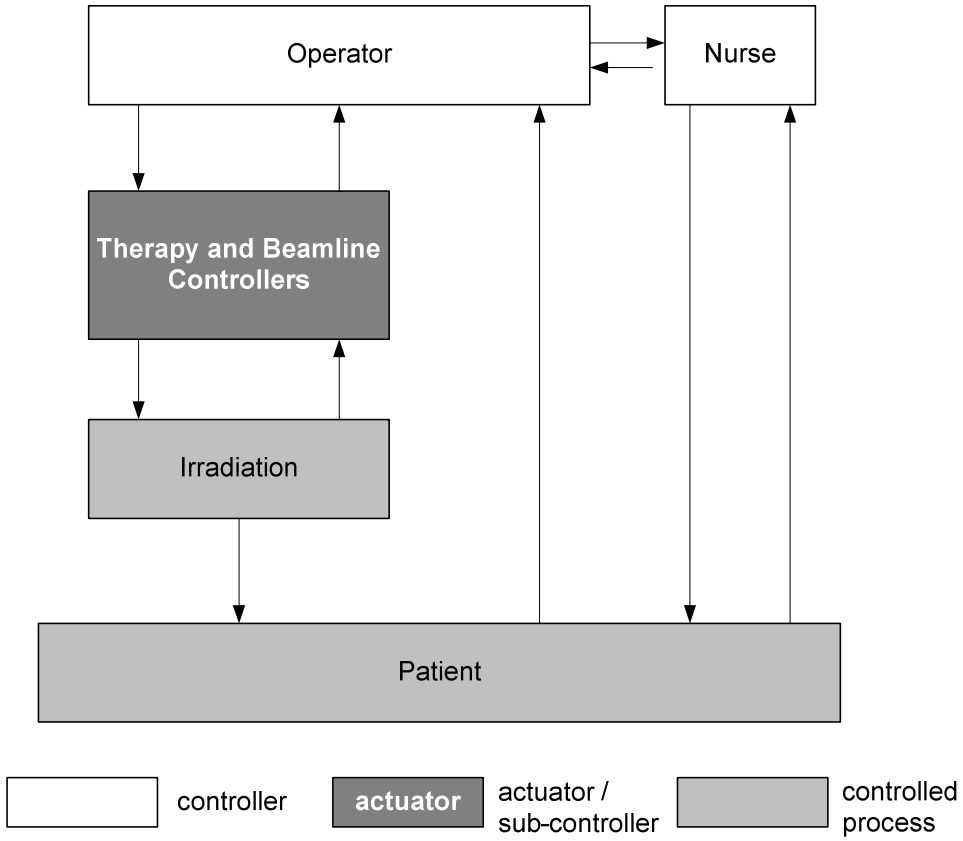


Figure 11: Control structure showing the operator as human controller acting on the beam line controllers and elements by providing them commands through the GUI of his workstation (Source: PSI, 2012d)

Operator, nurse and patient use an audio intercom for communication between them. Additionally the operator and nurse can observe the patient during treatment by video.

3.2.3.8.1.2 Detailed process loops

In the STPA framework, the control actions that the operator is responsible for, the actuators through which he implements them and the feedback channels that provide him with information on the state of the system that he/she is tasked with controlling are identified. These elements and their interactions are visualized in detailed process loops.

As explained in (PSI, 2012d), after the facility is turned on and the treatment mode has been selected for the beamline, operator can control the beam by taking the following control actions:

- CA 1.1 Request mastership
- CA 1.2 Load steering file
- CA 1.3 Start treatment

CA 1.1) Request mastership

Prior to treating a patient, the treatment area must be granted beam mastership. The operator requests mastership through the graphical user interface (GUI) on his workstation. That same GUI reports back whether mastership has been granted or not. This information is summarized in the process loop drawn in Figure 12.

CA 1.2) Load steering file

Once mastership has been granted, the operator identifies the patient to be treated on the daily-plan and loads the appropriate steering file again using the GUI. Figure 13 shows the associated process loop.

CA 1.3) Start treatment

Finally when the nurse clears the treatment area, the operator starts treatment through the GUI. He constantly gets feedback via the audio and video intercom from the patient and the nurse, as well as treatment progress information from the GUI. See Figure 13.

As explained in (PSI, 2012d), after the facility is turned on and the treatment mode has been selected for the beamline, operator can control the beam by taking the following control actions:

- CA 1.1 Request mastership
- CA 1.2 Load steering file
- CA 1.3 Start treatment

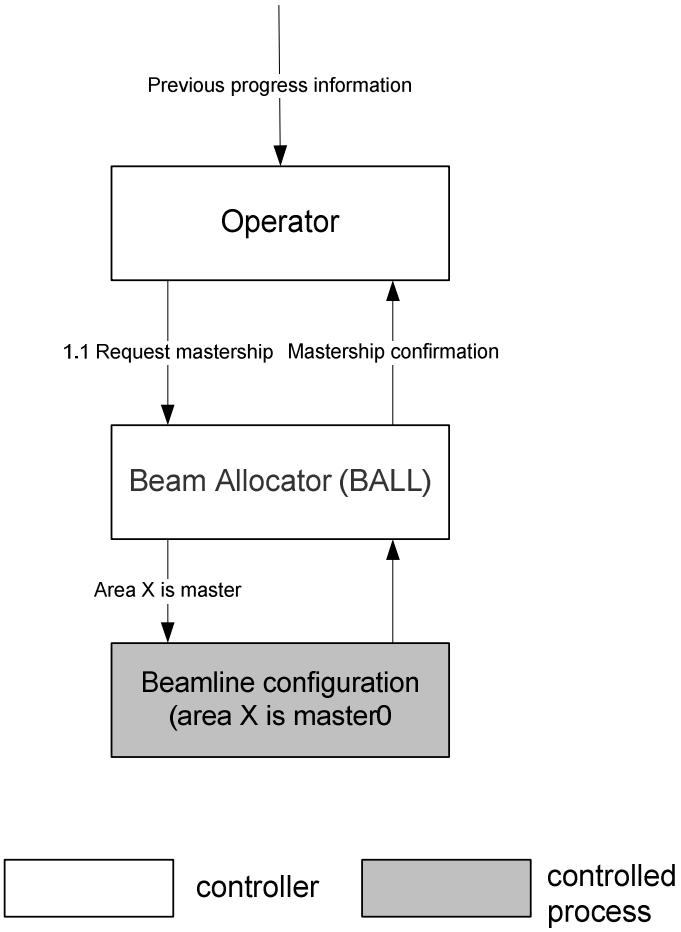


Figure 12: Process loop of the operator requesting mastership.

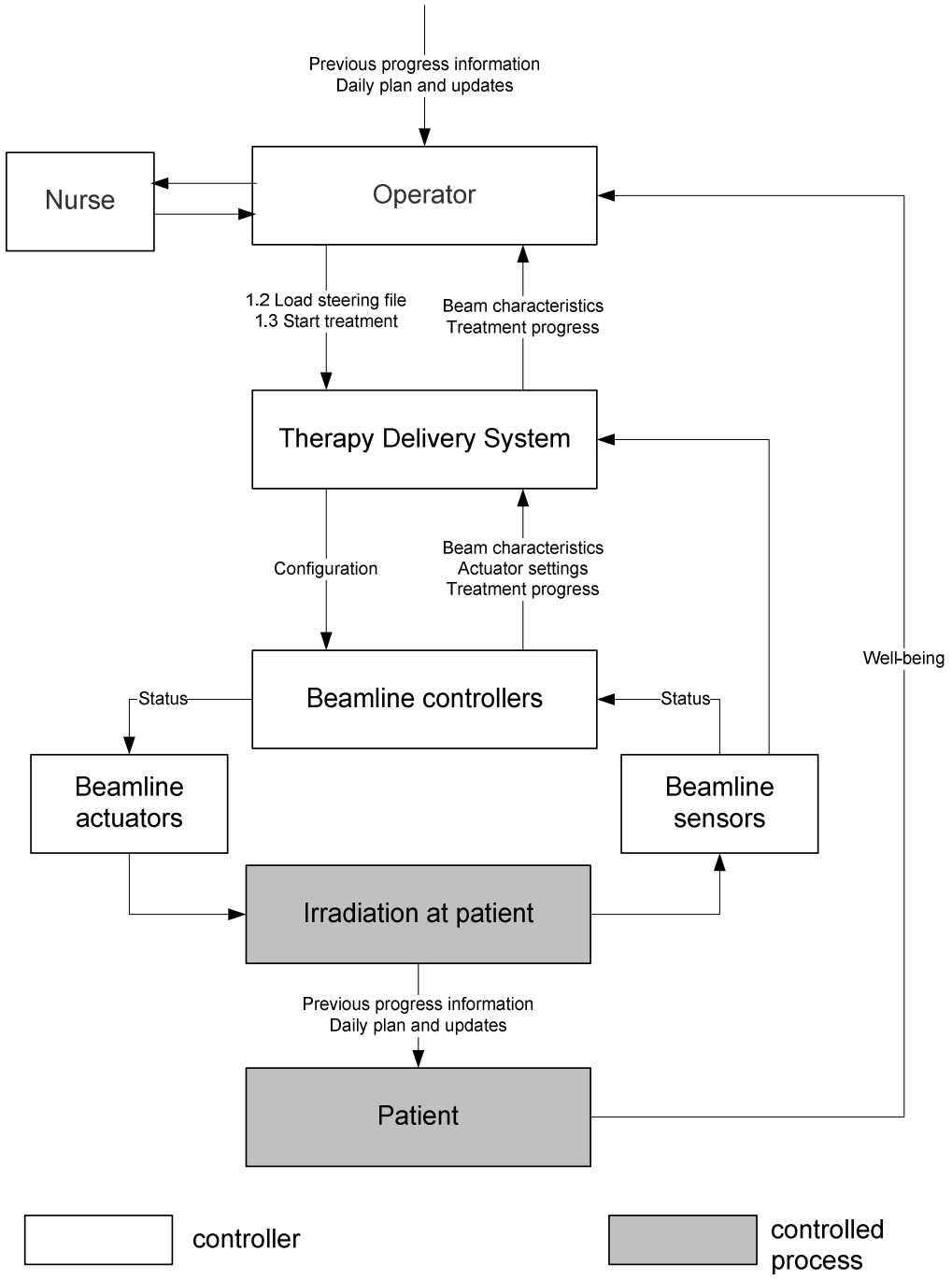


Figure 13: Process loop of the operator loading the steering file and starting treatment.
Commands from and feedback to the operator are transferred via the graphical user interface of his/her workstation.

3.2.3.8.1.3 STPA Step 1 on the operator

STPA Step 1 consists in identifying unsafe control actions that the controllers can take. This analysis was performed for only one of the operator's control actions: CA 1.3) Start treatment.

Relevant process variables

To initiate and supervise the safe delivery of the radiotherapy treatment, the operator must have information on the following:

- presence of non-patients in the treatment room (called "personnel" in Table 9)
- patient readiness for treatment (esp. position on the table)
- match of treatment plan ID with patient ID
- equipment readiness for treatment (incl. facility mode choice, table & Gantry positions as well as beamline actuators, sensors and interlock settings)
- grant of area mastership
- facility mode
- treatment status.

Table 9 - Process variables associated with the operator's control actions.

Process variable	Possible values	Comment
Personnel	• close • not close	"close" is to be understood as "potentially leading to detrimental radiation exposure". "personnel" is to be understood as "non-patient" (can include visitors, family members...) "close" = close to beamline or inside the treatment room "not close" = not close to beamline and outside of the treatment room
Patient readiness (esp. position)	• no patient • ready • not ready	“Ready”: patient is in treatment room, at treatment point, in the correct position and ready for dose delivery "Not ready": patient is in treatment room, but not ready for dose delivery (e.g. incorrect position)
Treatment plan ID	• right • wrong • none	"Right"/"Wrong" refer to whether the correct treatment plan has been selected and loaded for the patient awaiting treatment. "None": no treatment plan has been loaded
Equipment readiness	• ready • not ready	"Ready"/"Not ready": with respect to treatment start
Mastership status	• master • not master	"Not master": other areas have the power to control beamline elements
Facility mode	• therapy • non therapy	"Therapy": facility is configured for patient treatment application. All the patient safety and machine interlocks are enabled and remote operator control is disabled. "Non-therapy": facility is configured to allow more flexibility for experimental purposes. Some patient safety and machine protection interlocks are disabled by default. The interlocks can be remotely disabled/enabled by an operator.
Treatment status	• no treatment • in progress • interrupted	"Interrupted": treatment was previously in progress, but was stopped before completion and is expected to resume for the dose delivery to be complete

These are the minimal information requirements that shall be provided to the controller as they condition the potential impact of his actions on the safety of the system. Therefore, the area operator, when issuing the "start treatment" command, must care about the process variables documented in Table 9.

Identifying unsafe control actions

The Thomas process was used to identify the unsafe control actions associated with CA 1.3. Combining the values that the different process variables can take, a complete set of contexts was identified. It was then evaluated whether performing (Thomas process Part 1) or not performing (Thomas process Part 2) control action CA 1.3 was hazardous or not. If it was hazardous, then an unsafe control action (UCA) was identified. This proscess is documented in Table 10 and Table 11.

As a result of this process, eleven unsafe control actions were identified as being associated with CA 1.3.

UCA1. Treatment is started while personnel is in room (↑H-R4)
UCA2. Treatment is started while patient is not ready to receive treatment (↑H-R1, H-R2
Note: This includes "wrong patient position", "patient feeling unwell", etc.
UCA3. Treatment is started when there is no patient at the treatment point (↑H-R2, H-R3)
UCA4. Treatment is started with the wrong treatment plan (↑H-R1,H-R2)
UCA5. Treatment is started without a treatment plan having been loaded (↑H-R1,H-R2)
UCA6. Treatment is started while the beamline is not ready to receive the beam (↑H-R1, H-R5)
UCA7. Treatment is started while not having mastership (↑H-R1, H-R2, H-R4)
UCA8. Treatment is started while facility is in non-treatment mode (e.g. experiment or trouble shooting mode) (↑H-R1, H-R2)
UCA9. Treatment start command is issued after treatment has already started (↑H-R1, H-R2)
UCA10. Treatment start command is issued after treatment has been interrupted and without the interruption having adequately been recorded or accounted for (↑H-R1, H-R2)
UCA11. Treatment does not start while everything else is otherwise ready (↑H-R1, H-R2)

Part 1: Control actions provided in a state where action is hazardous

Table 10 – First step of the Thomas process applied to CA 1.3.

Index	Personnel	Patient readiness	Treatment plan ID	Equipment readiness	Area master-ship status	Facility mode	Treatment status	Is this Hazardous? consider whether this can be hazardous if performed too early/too late or if applied too long/stopped too soon.	Hazard	UCA
1.3-p.1	Not close	Ready	Right	Ready	Master	Therapy	No treatment	Non hazardous		
1.3-p.2	Close	*	*	*	*	*	*	Hazardous since personnel could be exposed to beam	H-R4	UCA1
1.3-p.3	*	Not ready	*	*	*	*	*	Inadequate patient position (possibly from patient feeling unwell) leading to wrong dose at wrong place	H-R1, H-R2	UCA2
1.3-p.4	*	No patient	*	*	*	*	*	Hazardous if possibility of messing up patient treatment history exists (e.g. system records the fraction as having been given when it has not been), thus leading to potential underdose	H-R2, H-R3	UCA3
1.3-p.5	*	*	Wrong	*	*	*	*	Hazardous because wrong treatment is applied to patient.	H-R1, H-R2	UCA4
1.3-p.6	*	*	None	*	*	*	*	Hazardous if beam is nonetheless delivered upon treatment start, using no guidance or following the guidelines of the treatment plan that was last used, leading to wrong dose at the wrong place.	H-R1, H-R2	UCA5
1.3-p.7	*	*	*	Not ready	*	*	*	Could lead to equipment damage	H-R5	UCA6
								Could be hazardous for patient if results in beam being switched “on” too early or with incorrect characteristics	H-R1	
1.3-p.8	*	*	*	*	No	*	*	Hazardous 1. to processes in the other user areas if the start treatment command is transferred to the beamline despite the mastership block and affects the beam used in other areas 2. to local area if the start treatment command provided when area is not master is interpreted by the local beamline as being receivable and puts the system in an ndesirable state (e.g. disable beam blocking devices in the	radiation hazards to users and patients in all (other and local) areas (H-R1, H-R2, H-R4)	UCA7

								local beamline before the shared beamline has been configured to beam readiness, which can only be done after mastership has been granted)		
1.3-p.9	*	*	*	*	*	Experiment		Hazardous when safety checks are disabled, especially if operator is not aware that treatment is proceeding without these safety nets. Could lead to patient over/under dose.	H-R1, H-R2	UCA8
1.3-p.10	*	*	*	*	*	*	In progress	Hazardous if treatment has already started, and issuing a start command could lead to reload of treatment file, resulting in overdose, or abort of fraction delivery (resulting in underdose if not adequately documented)	H-R1, H-R2	UCA9
1.3-p.11	*	*	*	*	*	*	Inter-rupted	Hazardous if extent of delivery was not recorded and delivery starts from the beginning of the treatment plan again.	H-R1, H-R2	UCA10

Part 2: Control actions not provided or inadequately executed in a state where inaction is hazardous

Table 11 – First step of the Thomas process applied to CA 1.3.

Index	Personnel	Patient readiness	Treatment plan ID	Equipment readiness	Area mastership status	Facility mode	Treatment status	Is this Hazardous? consider whether this can be hazardous if performed too early/too late or if applied too long/stopped too soon.	Hazard	UCA
1.3-np.1	Not close	Ready	Right	Ready	Master	Therapy	No treatment	Hazardous as could lead to long wait time or traffic jam in patient preparation area, both potentially resulting in patient moving.	H-R1, H-R2	UCA11

Treatment not started in situation 1.3-np.1 is hazardous in the same way that "treatment started too late" would be.

There does not seem to be a context where "incorrectly starting treatment" would have detrimental safety consequences. This probably has to do with the fact that starting treatment is a binary command that does not involve any parameter and is not performed through a sequence of steps that could be taken in an inappropriate order.

Note: "incorrectly" performing a control action is meant to refer to an action not being performed as it should be, i.e. steps being taken in the order they should, or the right parameters being used. It is not meant to refer to performing a control action at the wrong time or in the wrong context.

3.2.3.8.1.4 STPA Step 2: identifying hazardous scenarios associated with the operator's actions

Step 2 was applied to two of the above described unsafe control actions:

UCA3. Treatment is started when there is no patient at the treatment point. Patient treatment history is messed up leading to consecutive treatment errors (↑HR-H-R21, HR-H-R32) (1.3-p.4)

UCA4. Treatment is started with the wrong treatment plan (↑H-R1, HR-H-R24) (1.3-p.5)

Step 2 aims to understand how the treatment could be started while there is no patient at the treatment point. Once these causal scenarios are elucidated, elimination measures can be provided to prevent their occurrence. If they cannot be eliminated, attempts must be made to reduce the likelihood of their occurrence, detect them and mitigate their consequences.

The search for these causal scenarios was performed using the Step 2 Tree.

UCA3. Treatment is started when there is no patient at the treatment point [45]

-----reduce: require confirmation from operator about the desire to start treatment (e.g. start reading the steering file) to avoid the facility starting treatment without the operator intending treatment to start; require confirmation of nurse that treatment can proceed to avoid treatment being started by operator while patient is not at treatment point.

-----detect: provide feedback to operator about treatment being in progress (i.e. steering file being read) so he can stop it if finds out that it is proceeding while there is no patient (i.e. provide operator with an ability to detect and catch the UCA)

-----mitigate: when UCA is detected, make it possible for treatment team to re-initiate treatment at the point where it was left off after the error is documented in the (i.e. override the dose history so that it only reflects dose that was actually delivered to the patient). Note that this overwriting capacity must be strictly framed so as not to be abused (necessity to devise a procedure that provides for checks by several people, and is only to be used in very special circumstances[46]).

P1: inadequate command generation, i.e. the "treatment start command is generated despite there being no patient to treat"

P1T1: goal/input is missing, delayed or wrong

P1T1F1 and P1T1F2 (missing or delayed goal/input)

---

[45] this could, among others, lead to inadequate documentation of dose received by patient and is therefore hazardous

[46] making it cumbersome (e.g. by requiring signature of several people including medical team) should help in this regard ...

- Scenario 1 - operator was expecting patient to have been positioned, but table positioning was delayed compared to plan (e.g. because of delays in patient preparation or patient transfer to treatment area; because of unexpected delays in beam availability or technical issues being processed by other personnel without proper communication with the operator).

----**reduce:** provide operator with direct visual feedback to the gantry coupling point, and ask him to check that patient has been positioned before starting treatment (M1).

P1T1F3 (wrong goal/input)

- Scenario 2 - operator is asked to turn the beam on outside of a treatment sequence (e.g. because the design team wants to troubleshoot a problem) but inadvertently starts treatment and does not realize that the facility proceeds with reading the treatment plan.

----- **reduce**: reduce the likelihood that non-treatment activities have access to treatment related input by creating a non-treatment mode to be used for QA and experiments, during which facility does not read treatment plans that may have been previously been loaded (M2); make procedures (including button design if pushing a button is what starts treatment) to start treatment sufficiently different from non-treatment beam on procedures that the confusion is unlikely.

<u>P1T2: flawed control algorithm</u>

P1T2F1 and P1T2F2 (missing or wrong rules):

- Scenario 3 - operator's control algorithm is not equipped with a rule to deal with absence of patient at the treatment point (e.g. if time of treatment start is defined along with the daily treatment plan and operator asked to start treatment at exactly that time without checking for patient presence at the gantry, or with such a rule as "start treatment 5 minutes after trolley has entered the treatment room", without providing accurate feedback about possible delays) or is equipped with a wrong rule, leading him to start treatment before patient is actually brought to treatment point.

----**eliminate**: check that the procedure that the controller has to follow includes a rule such that "if there is no patient at treatment point, do not start treatment."

P1T2F3 (wrong clock):

- Scenario 4 - operator is asked to start treatment at a precise absolute time that is meant to coincide with time of patient positioning inside the treatment room by the nurse, but reference time for the operator differs from that used by the nurse, or nurse took longer than expected to position patient, resulting in treatment being started.

----**eliminate**: coordination between operator and nurse is not made by reference to a common treatment schedule, but by visual and oral coordination through the audio com (M3) and camera (M1) channel.

<u>P1T3: flawed process model</u>

design flaw

P1T3F1 (missing PV):

- Scenario 5 - operator not required to check for patient positioning before starting treatment or before resuming treatment after it was interrupted

----**reduce**: create procedure that asks for operator to confirm (e.g. via audio channel (M3)) with local area nurse that treatment is ready to proceed

inadequate updating

P1T3F2: wrong interpretation of information by controller

- Scenario 6 - patient had been positioned correctly, operator got ready to start treatment, but patient was removed from the treatment point before operator actually started treatment (e.g. because felt unwell/uncomfortable; because nurse detected an error in positioning; because table fell from the gantry....) and operator did not update his process model in time to account for this change in patient presence.

----**eliminate**: real-time visual feedback given to operator about patient presence at the treatment point via the video channel (M1) combined with procedural requirement that operator shall monitor patient well-being via that video channel (check that video does point to treatment location).

P1T3F3: missing/inadequately timed source of information (feedback or external)

- Scenario 7 - operator is not provided with real-time information about patient being installed at the treatment point

----**reduce**: visual feedback given to operator about patient presence at the treatment point via the video channel (M1). Note that this measure unfortunately does not eliminate this causal factor as the video processor could freeze, displaying a static picture of the last recorded frame.

P1T3F4: inadequate information transmission (distorted, lost, inadequate timing) –

- Scenario 8 - patient is removed from treatment point but camera image freezes so that operator sees patient as being still at the treatment point, when he is not.

----**detect:** none known of. It would be possible to include, at treatment point close to patient position, a clock so that even when patient and couch are immobilized (and the image therefore should not be expected to change), something (the numbers of the clock) keeps moving in the image, proving that it is not frozen.

- Scenario 9 - patient is removed from treatment point but there is a delay in the image transmission to the operator workstation, so that operator sees patient as being still at the treatment point and ready for treatment, when he isn't anymore.

---- **mitigate**: For this purpose it might be enough to require the use of video systems with high enough frame rate capability. This would be especially critical in case of ethernet based cameras sharing the same network as the computers.

- Scenario 10 - Patient has not arrived at treatment point, but image from prior treatment is still on the screen - operator leaves his workstation between two consecutive treatments (e.g. to go to the

restroom) and, upon returning, believes the new patient has been brought to treatment point and is ready for treatment.

----**eliminate:** a procedure could be used for operators to confirm patient identity before treatment start.

P1T3F5: inadequate sensor operation

- Scenario 11 - Camera is not working, leading either to one of the issues identified above (e.g. image freeze or transmission delay) or to an absence of image; but team decides to proceed with treatment anyway without providing enough information to the operator with respect to patient readiness for treatment. It could be that patient removal from treatment point was caused by a mechanical problem (e.g. table would not properly coupled with gantry) that also impacted camera functioning (e.g. mechanical shock that impacted both the coupling point and the camera).

----**reduce**: camera maintenance and audit program

----**mitigate**: in the absence of properly functioning video channel and when decision is made to nonetheless proceed with treatment, provide a procedure that substitutes audio communication between treatment floor nurse and local control room operator to the video channel (M1), with explicit confirmation by nurse to operator that patient is at treatment point and ready for treatment (M3). This type of solution can be seen as a provision to ensure a safe level of performance under degraded system conditions, a contribution to maintaining system functionality despite some elements not being available anymore.

P2: safe command not followed

<u>P2T1: flawed transmission</u>

P2T1F1: command not issued – *not applicable: could not find a scenario where the non-emission of an expected command led to the treatment being started when it should not have been started.*

P2T1F2: command inadequately timed

- Scenario 12 – Operator expected patient to have been positioned but was not provided with on-time feedback about the treatment floor's situation (see scenarios 5, 6, 7) and started treatment despite patient absence from the treatment point.

P2T1F3: command distorted

- Scenario 13 – Operator issued another command or performed a set of keystrokes on his workstation that, in this context, were interpreted by TDS as requesting treatment start.

---- **eliminate**: command sequence to trigger the start command should be unique and should only be available through a single input device.

<u>P2T2: flawed execution</u>

not executed: not applicable (P2T2F1: missing input; P2T2F2: command not received; P2T2F3: actuator failure; P2T2F4: inadequately timed execution)

wrongly executed

P2T2F5: wrong input

- Scenario 14 although operator did not issue the start treatment command, a signal (e.g. command from another user area; other command issued by operator) was interpreted by the TDS / beamline actuators as necessitating treatment start.

---- **reduce**: concept of mastership is meant to prevent other areas from sending input to shared beamline actuators
----------some form of confirmation requested by TDS before proceeding to avoid spurious signals in the command channel to be wrongly interpreted as starting treatment? (role of TVS?)

---- **eliminate**: TDS should only listen to/accept/receive commands from local area

P2T2F6: actuator failure

- Scenario 15 TDS starts to read through treatment files spuriously, e.g. because is stuck in a recursive loop following some unusual set of commands….

Note: Some mitigation measures have the ability to address several hazards: M1 (direct visual link between operator and patient through video channel) can help mitigate several scenarios; M3 (audio communication between nurse and operator + requirement that nurse confirm that patient is positioned and ready for treatment) is probably helpful for several UCAs.

UCA4. Treatment is started with the wrong treatment plan

-----**reduce**: require verification of nurse and operator data. In the PROSCAN facility: requirement that operator check plan identifying number against daily plan when uploading it to work station; redundancy provided in treatment plan ID verification by requiring nurse to enter patient ID information according to daily plan into treatment room workstation and have PatBase verify that nurse and operator input match.

-----**detect**: require post-treatment verification of dose delivery, and verification that it matches expected treatment plan

-----**mitigate**: fractioning reduces the impact of inadequate treatment being applied; following fractions could for example be adjusted to take into account this initial error

P1: inadequate command generation, i.e. the "treatment start command is generated despite the treatment plan not being the right one"

P1T1: goal/input is missing, delayed or wrong

P1T1F1 and P1T1F2 (missing or delayed goal/input)

- Scenario 1 – the proper steering file failed to load (either because operator did not load it, or previous plan was not erased from system memory and overwriting is not possible) and the system uses a previously loaded one by default.

----reduce: when fraction delivery is completed, used steering file could for example be automatically dumped out of the system's memory (M4)

P1T1F3 (wrong goal/input)

- Scenario 2 – operator is provided with a steering file produced with a treatment plan that is wrong (error in treatment planning).
- Scenario 3 - Operator is provided with steering file that corresponds to the right patient and the right treatment, but wrong fraction (all the more likely that steering file names are not coded for fraction, but only for patient ID), or loads a steering file from previously delivered fraction.

----- **eliminate**: fractions that have already been delivered could for example be dumped out of the workstation memory, and future fractions' steering files should not be loaded in advance

- Scenario 4 – daily plan used by the operator does not match the actual sequence of patients (e.g. because scheduling changes were made that were not communicated to the operator).
- Scenario 5 – steering files' name matches the patient ID but does not match the patient ID that is recorded inside the file (e.g. because someone – including the operator – with writing powers on the file name changed it – either purposefully or through an involuntary mistake)

----- **reduce**: restrict file naming powers. Note: in Gantry 1, file names are not locked. At this stage and unless otherwise motivated and circumscribed, this should be considered a safety gap.

- Scenario 6 - nurse provides wrong patient ID when asked to confirm ID by operator.

----- **detect & mitigate**: if the nurse's input is not matching his own daily plan, operator can detect an error and procedure can be restarted.

P1T2: flawed control algorithm

P1T2F1 and P1T2F2 (missing or wrong rules):

- Scenario 7 – operator not aware that he has to load the patient specific steering file, thinking that operations' planning pre-load all files corresponding to daily plan onto local area computer platform, possibly resulting in previous file being used

----**eliminate**: clear procedures and training; M4

P1T2F3 (wrong clock): not applicable

P1T3: flawed process model

design flaw

P1T3F1 (missing PV) -

- Scenario 8 - operator, not required to check for patient ID before starting treatment or before resuming treatment after it was interrupted, follows the treatment plan, and possibly skips a name, repeats a name or is not made aware of potential changes

----**reduce**: create procedure to compare patient ID as confirmed by nurse and treatment plan ID (done by PatBase in PROSCAN). Note: choice to name treatment plan with same name as patient ID is a useful measure in that it makes this comparison straightforward.

P1T3F2: wrong interpretation of information by controller

- Scenario 9 – operator misunderstands possible change made to the daily plan and does not realize that patient order may have changed
- Scenario 10 – Several steering files have names so similar that they are easily confused; order of files in GUI Interface repository which operator selects a file from differs from that on daily plan, leading to potential confusion.

P1T3F3: missing/inadequately timed source of information (feedback or external)

- Scenario 11 - operator is not provided with real-time information about patient ID of person being installed at the treatment point, or about fraction number

----**reduce**: audio confirmation of patient ID from nurse; process for updates to daily plan to be communicated to operator

inadequate updating

P1T3F4: inadequate information transmission (distorted, lost, inadequate timing)

- Scenario 12 – changes to daily plan or daily plan data are imperfectly communicated to the operator

----**reduce**: unknown. Currently, such communication is written on a sheet of paper. If changes were to be made without the operator having acknowledged them, UCA3 can occur (e.g. him not realizing that someone put a new piece of paper on this desk and him thereby using the old one without knowing it is out of date).

P1T3F5: inadequate sensor operation – not applicable (sensing is performed by operator's eyes)

P2: safe command (here: start treatment with the proper steering file in the right context) not followed leads to UCA 4 (treatment is started with the wrong steering file)

P2T1: flawed transmission

P2T1F1: command not issued – *not applicable: could not find a scenario where the non-emission of a command led to UCA4.*

P2T1F2: command inadequately timed – *not applicable*

P2T1F3: command distorted

- Scenario 13 – operator selects the right steering file for loading on his workstation, but a different one is sent to TDS

P2T2: flawed execution

not executed: not applicable, since this is a case of execution in the wrong conditions rather than non execution

(P2T2F1: missing input; P2T2F2: command not received; P2T2F3: actuator failure; P2T2F4: inadequately timed execution)

wrongly executed

P2T2F5: wrong input – *not applicable: could not find such a scenario leading to UCA 4*

P2T2F6: actuator failure

- Scenario 14 – GUI fails to transmit the steering file selected by operator to TDS which, by default, uses one that had previously been loaded.

3.2.3.8.1.5 Assessment of safety constraint enforcement in current system

The PROSCAN STPA study dealt with a system in existence, not one that is being designed from scratch. Protective measures thus already exist to prevent hazardous situations from occurring. The conclusion of our STPA study consisted in verifying, for each of the examples whose analysis was performed, whether the unsafe control action that were identified are met with appropriate prevention, detection and mitigation provisions.

The results of this evaluation for the two unsafe control actions associated with CA 1.3"Start Treatment" is summarized in Table 12.

Table 12 – List of protective measures for UCAs uncovered in STPA Step 1.

Unsafe Control Action		**Elimination/Detection[47]/Reduction / Mitigation Measures**
UCA1	Treatment is started while personnel is in room	Prevent: Personnel radiation protection system blocks beam from entering treatment area as long as area not cleared.
UCA2	Treatment is started while patient is not ready to receive treatment	Prevent (re: making sure that the patient is in the correct position): • Prior to bringing the patient into the Gantry room the positioning of the patient is verified with CT-scans (operative measure). • X-ray images are taken at random of the patient installed on the couch and attached to the gantry. (See UCA11) • The therapy Verification System verifies that the table and the gantry are correctly positioned before allowing that the beam be switched on. This applies also to the first dose element. Detect: The patient is supervised by the nurse and operator through audio and video by redundant cameras and from different angles.
UCA3	**Treatment is started when there is no patient at the treatment point**	**See the detailed evaluation of how the Step 2 scenarios can be addressed in the previous section**
UCA4	**Treatment is started with the wrong treatment plan**	**See the detailed evaluation of how the Step 2 scenarios can be addressed in the previous section**
UCA5	Treatment is started without a treatment plan having been loaded	Prevent: Therapy Control System is based on state machine which does not allow to start a treatment when no treatment plan has been loaded.
UCA6	Treatment is started while the beamline is not ready to receive the beam	The machine interlock system allows beam only when all beamline devices are ready. However, **the requirement** "Treatment must not be started without a treatment plan", **formulated as SC-r5, is not satisfied.** The treatment start can indeed be sent despite the beamlines not being ready. As a result, the system would just wait until the beamline is ready and then proceed. It is crucial to check at what step the system waits. If it waits after having started a dose element application, the system will accumulate background noise as dose, which could result in a wrong dose delivery.
UCA7	Treatment is started while not having mastership	Prevent:. • The treatment delivery system is aware of which area is granted "mastership". This means that starting a treatment from an area which does not hold mastership is disabled. • The patient safety system is also aware on mastership and will ignore all "beam on" requests from an area that is not master.

[47] Detection measures are insufficient if not associated with corresponding reduction and mitigation measures.

UCA8	Treatment is started while facility is in non-treatment mode (e.g. experiment or trouble shooting mode)	There is no technical measure to prevent this from happening. Operational measures require the facility to be in therapy mode for patient treatments. Rules applying to patient treatments done in a degraded mode have been defined. The degree of degradation is in principle limited by the Therapy Control System, but this limitation function is itself dependent on the mode setting. It is disabled in experiment mode.
UCA9	Treatment start command is issued after treatment has already started	The treatment delivery system uses a state machine to handle the treatment status transitions. While a treatment is in progress only a "treatment abort" is available. This abort consequently generate all relevant log-files and status information
UCA10	Treatment start command is issued after treatment has been interrupted and without the interruption having adequately been recorded or accounted for	See UCA5. The Therapy Control System logs every treatment interruption with the necessary information to continue the treatment at a later time.
UCA 11	Treatment does not start while everything else is otherwise ready	**Recognition that this can be an issue:** No special prevention measures in place except close coordination efforts by treatment delivery team. Mitigation of expected consequence of this UCA (change in patient position):X-rays of the patient on the gantry could be used to verify correct patient positioning.

Similar results were obtained for the four other examples. Although they did not cover the whole facility and its subsystems, they were deemed sufficiently complete to be the object of a comparison with the Safety Report where the PROSCAN design team documented the hazards faced by PROSCAN and the protective measures that were taken to address them. This comparison is reported in the following section.

3.2.3.8.2 Example 5: using STPA to facilitate brainstorming

Example 5 (mixed mode with both continuous and spot scanning) was performed by a team of PROSCAN designers assembled during one afternoon for a workshop moderated by Dr. Christian Hilbes and Martin Rejzek at PSI. The example focused on a functionality of the Gantry-2 facility that was not yet fully defined and still the object of much discussion and design effort. As explained in (PSI, 2012a), the workshop participants were briefed on STPA Step 1 and Step 2 by the moderators, and then asked to identify unsafe behavior using the STPA framework as a stimulator for discussion. In less than four hours, the workshop identified hazardous behavior that had not previously been thought of by the design team. It was deemed a success as it showed that designers that are not safety experts can promptly be made familiar enough with the methodology that they can contribute effectively to the analysis. It also showed that STPA processes can effectively be used to stimulate brainstorming and discussion, and that results could be obtained within reasonable amounts of time (PSI, 2012a).

3.3 Comparison to the Gantry-2 Safety Report

To evaluate whether STPA allows for more effective (i.e. identification of more relevant hazardous scenarios, provision of more detailed insight) and/or more efficient (i.e. less resources spent performing the analysis) evaluation of the hazards associated with a given device requires that its outputs can be compared against that of a study performed using a different technique on the same system, controlling for availability of data and experience of the analysts with the system under consideration.

Unfortunately, such a baseline study was not available and could not be performed. An indication of what scenarios the design team evaluated as being relevant is however offered by the Gantry-2 Safety Report, a report written by the PROSCAN design team that summarizes the hazard scenarios (called "errors") that were considered during the design of Gantry-2.

The draft PSI Safety Report is being put together for communication to the Swiss regulatory authorities of the hazards associated with the commissioning of Gantry-2. Its format was developed over time to be a communication tool between the PROSCAN designers and the Swiss regulatory authorities. It is meant as a summary, not a full blown description, of the main risks identified by the design team and the solutions that were proposed to them. It does not follow any specific hazard analysis framework, such as those that were described in the Background Chapter of this thesis.

Given the ad hoc documentation process used in the draft Safety Report on one hand, and the limited scope of the analysis that was performed for the PROSCAN STPA project on the other, a formal comparison is out of order. No positive conclusion could validly be made from that data, especially regarding the ability of different hazard analysis techniques to identify hazardous scenarios or the resources that must be allocated to that effort.

While a rigorous effectiveness and efficiency assessment is impossible to make, it remains valuable to estimate whether the scenarios uncovered by the PROSCAN STPA study are any close to those that the PROSCAN design team built into their design: are all scenarios identified by one also found by the other? Does one identify scenarios that the other one did not include?

The draft Safety Report indeed represents a smaller set of scenarios than those that were considered by the designers. If the STPA project managed to identify those scenarios (and possibly others), then nothing could be concluded. However, if, within the scope that they share, the PROSCAN STPA project failed to identify at least the scenarios listed in the draft Safety Report, then legitimate questions could be asked regarding the ability of STPA to uncover as rich a set of scenarios as necessary to ensure safe design and operations of a system.

After a brief comparison of the type of information that they present, this section summarizes and contrasts the results of both hazard analyses.

3.3.1 Comparing the analytical set-up of the STPA study and Gantry-2 draft Safety Report

Table 13 compares the analytical set-up of the STPA study and of the draft Safety Report in order to evaluate the scope normalization that is required for the scenarios documented in the draft Safety Report to be compared to those uncovered by the PROSCAN STPA project. To that end, the motivation, focus, scope, methodological references of the documentation endeavors and composition of the assessment teams are described.

These differences in motivation, scope, focus, assessment teams and methodologies are significant enough that the output of the two analyses cannot be compared directly[48]. However, checking whether, in the scope they share, the STPA scenarios include the 53 "errors" of the draft Safety Report will provide a preliminary response to the question of whether STPA does a good job of identifying hazardous scenarios that are considered important enough to be reported to a commissioning authority.

The following section proposes a semantic analysis to evaluate how much overlap there is between the results of the two studies. Its goal is to discover whether one study found causes of potential accidents that were missed by the other despite being included in their mutual scope.

[48] For example, using the fact that 53 "errors" are documented in the draft safety report as possibly leading to accidents while the STPA analysis that was performed on only a subset of PROSCAN controllers reports 101 unsafe control actions (UCA) with an average of 16.8 causal scenarios for the 5 UCAs on which STPA Step 2 was performed (thereby estimating that performing Step 2 on the 101 UCAs identified would yield some 1700 hazardous scenarios) to claim that STPA is superior at discovering hazardous behavior is not valid.

Table 13 - Motivation, scope, focus, methodology and assessment teams respectively used in the Gantry-2 draft Safety Report and the PSI STPA study.

Commonalities between the approaches and their scopes are highlighted in grey.

	Draft Safety Report (PSI, 2012a)	**STPA study (PSI, 2012d)**
Motivation: different	Document protective measures for purpose of communicating with the licensing authorities	Identify hazardous scenarios for purpose of method development and checking that protective measures are adequate
Scope: same	"Risk analysis"	"Hazard analysis"
Scope *(losses considered):* different with overlap	Patient radiological safety	Patient radiological safety Non-patient (personnel & visitors) radiological safety Equipment safety (in the context of radiation creation and delivery)
Scope *(sub-systems studied)*: different with overlap	Treatment Delivery: analysis • Gantry 2 - facility and operations design (spot scanning)	Treatment Definition: introductory discussion Treatment Delivery: analysis • Gantry 2 - beam sweeping function • Gantry 2 - therapy delivery system in mixed spot and continuous scanning modes • Gantry 1 & 2 - spot scanning dose control • Gantry 1& 2 - nurse • Gantry 1& 2 - local operator
Scope *(facility modes):* different with overlap	Spot scanning	Spot scanning and Continuous scanning
Focus: different	Protective measures	Hazardous scenarios
Methodology: different	No specific technique but top-down analysis starting with undesired system behavior). Reference to IEC 61508 to evaluate adequacy of protective measures provided. Provisions to include consideration for severity and likelihood of risk occurrence, but data not included in the version of the document that was made available for this comparison.	STPA (top-down analysis, starting with undesired system states). No attempt to evaluate adequacy of protective measures provided beyond verifying that they exist.
Assessment teams: *different with overlap*	Current PROSCAN designers including Martin Rejzek	Former PROSCAN team member + MIT graduate student + current PROSCAN designer Martin Rejzek

3.3.2 Comparing the analytical results of the Gantry-2 draft Safety Report and the PROSCAN STPA Study: a semantic approach

Starting with the definition of losses that the designers must attempt to prevent, an STPA analysis provides different outputs to its users:

- high level safety constraints
- hazardous states
- hazardous scenarios (unsafe control actions and the causal factors that can lead to them being taken).

Although it uses a different vocabulary, the Gantry-2 draft Safety Report is organized in a similar manner. After translating the terminology used in both reports to a common standard, this section proceeds with comparing the sets of scenarios reported respectively in the STPA Project and the draft Safety Report.

3.3.2.1 Identification of high level safety constraints

Although trivial, this comparison is included for the sake of completeness.

Safety report data: The draft Safety Report starts with a general constraint formulated as follows:

> "*The aim of the irradiation is to deliver a predetermined sequence of dose spots. Each dose spot should be delivered:*
>
> - *to the correct position;*
> - *in the correct quantity."*

Translation into PROSCAN STPA terminology: attributes "dose position" and "dose quantity" must be controlled.

PROSCAN STPA data: "location of dose deposition" (i.e. "beam alignment with target" and "beam tune") and "amount of dose deposited" (i.e. "beam on/off" and "beam intensity") must be controlled

Conclusion: IDENTICAL - the high-level safety constraints for both analyses are the same and are meant to ensure that each dose element delivered (spot for spot-scanning which is the only mode discussed by the draft Safety Report, or line for continuous scanning which the PROSCAN STPA discusses in addition to spot-scanning) is sent to the clinically correct location in the clinically correct amount.

3.3.2.2 Definition of situations to be avoided (i.e. hazardous states in STAMP terminology)

Safety Report Data: the draft Safety report identifies 4 safety goals that must be achieved so that the high level safety constraints are met:

> SG 1: NO RADIATION ACCIDENT: No serious overdose should be delivered to the patient.
>
> SG 2: NO ERROR IN THE DELIVERED DOSE: No incorrect dose should be delivered.
>
> SG 3: NO ERROR IN DOSE POSITION: The dose must be applied at the correct position
>
> SG 4: DELIVERED DOSE AND DOSE POSITION MUST BE KNOWN AT ALL TIMES: If the irradiation is interrupted at any time, the dose already deposited and the beam position must be known.

PROSCAN STPA Data: The STPA analysis identifies 5 system level hazards that must be protected against, either by their elimination or their reduction and mitigation:

> H1. OVERDOSE: Dose delivered to patient tissues (healthy tissue and tumor) is higher than clinically desirable.
> H2. UNDERDOSE: Dose delivered to tumor is lower than clinically desirable.
> H3. WRONG FRACTIONS: Radiation delivery is improperly fractioned.
> H4. NON-PATIENT IS UNNECESSARILY EXPOSED TO RADIATION (esp. personnel and visitors)
> H5. EQUIPMENT IS SUBJECT TO UNNECESSARY STRESS.

Comparison: The two analyses adopt different viewpoints as to what behavior should be considered hazardous. The draft Safety Reports describes situations to be avoided in terms of specific operational *output* or *behavior* (e.g. "No error in delivered dose", "Know delivered dose at all times"). As such, this description of desired behavior is close to those used to generate specifications and requirements or, in STPA terminology, safety requirements. On the other hand, STPA hazards are defined as system (where the system includes not only its operations team and facility, but also its designers, medical users and patients) *states* that one wishes to avoid. These states are in turn understood to be created through certain (unsafe) system behavior: the unsafe control actions (UCA). Following the definition of high level hazards, high-level system constraints will be defined whose enforcement will prevent the hazards from being realized.

Building on this framing of specific hazardous states (STPA's concern) as resulting from specific and hazardous system behavior (draft Safety Report's concern), understanding that hazards can be organized in causal hierarchies, and making sure that the two analyses are compared within the same scope, it is straight-forward that the 4 Safety Goals of the draft Safety Report are bijective images of the 2 main patient related STPA hazards:

H1 (overdose) is caused by either an error in the delivered dose (i.e. "too high dose at right position") or an error in the dose position, possibly associated with an error in delivered dose (i.e. " dose at wrong position"). It is therefore caused by SG1, SG 2 and/or SG 3 being breached.

H2 (underdose) is caused by either an error in the delivered dose (i.e. "too low dose at right position") or an error in the dose position, possibly associated with an error in delivered dose (i.e. " dose at wrong position" that may lead to dose thresholds for critical structures and non target volumes to be exceeded, a situation requiring that treatment be interrupted or modified, with the result that less dose is delivered to the tumor than clinically desired). It is therefore caused by SG 2 and/or SG 3 being breached.

SG 4 being breached may contribute to H1 and H2, especially in cases of beam interruption, a situation identified during the STPA analysis.

As explained above, H4 and H5 are outside of the scope defined by the draft Safety Report as relevant for its investigation. STPA scenarios that will have been identified to cause these hazards shall therefore be excluded from our assessment of whether STPA identified more causes than the draft Safety Report did.

H3 is a more delicate matter. Although the draft Safety Report acknowledges that failing to apply the correct fractioning regime (scenario 8.f) will cause an error in the delivered dose or beam position (SG 2 and 3), it only considers issues associated with the application of the fraction regime, not with this definition since it only concerns itself with the operation of the facility and not the design of the treatment plan. On the other hand, the PROSCAN STPA study does include consideration for treatment plan preparation. So, within the scope of the draft Safety Report, H3 can be considered to be included in both analyses although the scope of how the STPA study discusses it is larger.

Conclusion: The definition of situations to be avoided is similar within the scope shared by both analyses, but STPA adopts a larger scope (that includes a broader set of hazards, treatment delivery and also treatment definition). More specifically, for the comparison to be valid, STPA scenarios uniquely associated with H4 or H5, and the treatment definition phase are taken out of this comparison effort. The reader should note that the STPA study looked at a larger set of hazards and operational issues.

3.3.2.3 Organization of the information

"What's in a name? That which we call a rose
By any other name would smell as sweet" (Shakespeare,1597)

Both studies organize the results of their investigation in a top-down hierarchy of causal relationships. The draft Safety Report describes 13 high level *failure*[49] *scenarios* (labeled 7a to 7d, 8a to 8f, 9a to 9c), grouped in 3 families of failure scenarios (7: overdose scenarios, 8: error in delivered dose or beam position, 9: errors in dose position and dose recording). Each of the 13 failure scenarios is then understood to be caused by various *errors* (numbered E 7.a.1 to E 9.c.1), 53 of which are reported in the draft Safety Report (see Table 15). The STPA project on the other hand records 101 *unsafe control actions* (numbered UCA1.1.1 to UCA5.3.4) that are understood to be caused by *hazardous scenarios* identified as referring to certain *causal factors.* There are 84 hazardous scenarios recorded for the 5 unsafe control actions on which STPA Step 2 was performed (see Table 16 for an illustrative excerpt and Table 23 for the full list).

While the naming conventions used by both reports are different, looking into the specific scenarios that these names refer to help realize that they aim to describe identical concepts. This similarity is made obvious by using a neutral expression as a semantic pivot point in Table 14.

Table 14 - Causal structure and naming conventions in the Gantry-2 draft Safety Report and the PROSCAN STPA study.
Both reports share a similar top-down causal structure but their naming conventions differ. Indents in this table's cells imply causal relationship: a lower level indent is a cause to the higher level indent.

Draft Safety Report	Neutralization wording	STPA study
Safety Goals	Situations to be avoided	High Level Hazards
	Refined description of the situation to be avoided	Refined Hazards
Failure Scenarios	Unsafe behavior, one that will result in the the situations to be avoided being realized	Unsafe Control Actions
Errors	Detailed explanation of how the unsafe behavior could occur.	Causal Factors / Scenarios
Technical Measures *Operational Measures*	*Design choices aimed at eliminating the unsafe behavior, reducing it or mitigating it.*	*Protective Measures*

[49] Semantic clarification: the term failure is used to describe situations where a hazard is realized despite no component having failed per se, but errors happened—such as wrong configuration of certain logic modules—that led the system to "fail to behave safely"; while in STPA salons failures is understood to refer specifically to individual components failing to function as intended.

Table 15 – Draft Safety Report Results: failures and errors leading to the safety goals being breached

SG 1: NO RADIATION ACCIDENT: No serious overdose should be delivered to the patient.
SG 2: NO ERROR IN THE DELIVERED DOSE: No incorrect dose should be delivered.
SG 3: NO ERROR IN DOSE POSITION: The dose must be applied at the correct position
SG 4: DELIVERED DOSE AND DOSE POSITION MUST BE KNOWN AT ALL TIMES : If the irradiation is interrupted at any time, the dose already deposited and the beam position must be known.

7. Overdose scenarios (SG 1)
7a Error during spot application
E 7a.1 Dose delivery not correctly terminated by dose monitoring system 1.
E 7a.2 Undetected fault in the dose monitoring system 2.
E 7a.3 Dose delivery not correctly terminated by dose monitoring system 1 AND 2.
E 7a.4 Local partial power failure of the dose monitoring.
E 7a.5 Asynchronous reset of the watchdogs.
E 7a.6 Failure at beam switch-off due to an error in the PaSS.
E 7a.7 Failure to limit the maximal spot dose and duration due to an error in the watchdogs.
E 7a.8 Failing of the beam switch-off function due to an error in one or several final elements.
E 7a.9 Failing of the beam switch-off safety function due to an error in the central PaSS.

7b Error before or after the spot application
E 7b.1 The patient safety and control systems are not correctly set up.
E 7b.2 Miss-configuration or errors in the MPSSC modules
E 7b.3 Miss-configuration or error in the ETOT modules
E 7b.4 Interferences from the RPS or MCS on the beam line and beam blockers.

7c Unplanned beam switch-on
E 7c.1 Unplanned beam switch-on during patient setup and treatment.
E 7c.2 External interference during the execution of the therapy.

7d Excessive beam intensity
E 7d.1 COMET delivers an extraordinary beam current greater than 1000 nA.
E 7d.2 The failing of the deflector plate or the use of an inadequate characteristic lead to high beam currents.
E 7d.3 The variable intensity correction is wrong and causes therefore a too high current in the area of Gantry 2.
E 7d.4 An excessive beam current is delivered to the area due to wrong beam tunes.

8. Error in delivered dose or beam position (SG 2 and 3)
8a Dose measurement error
E 8a.1 Dose measurement error in dose monitoring system 1.
E 8a.2 Dose measurement error in dose monitoring system 2.
E 8a.3 Simultaneous failure of the high-voltage supply to the Dose Monitors.
E 8a.4 Saturation of the beam monitoring systems.
E 8a.5 Excessive fluctuation in beam intensity.

8b Error in the beam
E 8b.1 Error in the setting of beam line to Gantry 2.
E 8b.2 Undesired changes of the beam line element parameters during the field application.
E 8b.3 Errors in the beam set-up not detected before first spot application.
E 8b.4 Incorrect shape, position and range of the beam due to faults in the scanning elements (sweeper, pre-absorber, nozzle cover, table).
E 8b.5 Error in the sweeper magnets.
E 8b.6 Error in the pre-absorber position or wrong pre-absorber.
E 8b.7 No or wrong nozzle cover is installed.
E 8b.8 Error in the position of the mechanical axes.
E 8b.9 Incorrect position of the beam due to error in the computer tomograph (CT) calibration.

8c Error in the patient position
E 8c.1 Incorrect position of the patient.
E 8c.2 Error in the patient position offset correction.
E 8c.3 Selection of the wrong patient.

8d Error in the therapy systems.
E 8d.1 An incorrect spot dose is assumed by the TDS or the TVS.
E 8d.2 Simultaneous faults in both TDS and TVS.
E 8d.3 Incorrect dose requested by the therapy planning.
E 8d.4 Selecting the wrong steering file.
E 8d.5 Incorrect selection of the treatment field.

8e Incorrect state of the treatment facility.
E 8e.1 The patient treatment facility is not correctly set up
E 8e.2 Error in Mastership allocation or in the central PaSS area reservation control.
E 8e.3 Error in facility mode User Therapy for patient treatment.
E 8e.4 Wrong calibration of the dose defining monitor 1.
E 8e.5 The cyclotron delivers too much beam to the gantry 2 beam line.
E 8e.6 The settings of any of the beam intensity defining elements, KMA5/3, DMAD1, DMAF1, AMAKI and the slits FMA1x are wrong.
E 8e.7 Error in the beam tune verification system (BTVS) and data links.

8f Failure to apply the correct fractionation regime.

9. Errors in dose and position recording
9a Power failure or computer crash.
E 9a.1 Power failure or Therapy Control System crash.

9b Interruption of the treatment
E 9b.1 Errors following an intentional interruption of the treatment.
E 9b.2 Errors following an intentional interruption of the treatment.

9c Procedures for restarting an interrupted treatment.
E 9c.1 Errors at a restart following an interruption of the treatment.

Table 16 - PROSCAN STPA Results (excerpts – see Table 23 for full data)

Unsafe control actions and hazardous scenarios for a subset of the system's controllers and control actions

H1. OVERDOSE: Dose delivered to patient tissues (healthy tissue and tumor) is higher than clinically desirable.
H2. UNDERDOSE: Dose delivered to tumor is lower than clinically desirable.
H3. WRONG FRACTIONS: Radiation delivery is improperly fractioned.
H4. NON-PATIENT IS UNNECESSARILY EXPOSED TO RADIATION (esp. personnel and visitors)
H5. EQUIPMENT IS SUBJECT TO UNNECESSARY STRESS.

UCA 1.1.1 Patient is positioned and immobilized on table imminent to treatment (↑H1, H2)
UCA 1.1.2 Wrong patient or patient of unknown identity is positioned on the table (↑H1, H2)
UCA 1.1.3 Patient is positioned on table too early. (↑H1, H2)
UCA 1.2.1 Wrong patient or patient of unknown identity is brought to treatment point (↑H1, H2)
UCA 1.2.2 Table is positioned at treatment point too early or too far from time of treatment start. (↑H1, H2)
UCA 2.3.1 Treatment is started while personnel is in room (↑H4)
UCA 2.3.2 Treatment is started while patient is not ready to receive treatment (↑H1, H2)
UCA 2.3.3 Treatment is started when there is no patient at the treatment point (↑H2, H3)

S 2.3.3.1 (P1T1F1 & P1T1F2) operator was expecting patient to have been positioned, but table positioning was delayed compared to plan (e.g. because of delays in patient preparation or patient transfer to treatment area; because of unexpected delays in beam availability or technical issues being processed by other personnel without proper communication with the operator).

S 2.3.3.2 (P1T1F3) operator is asked to turn the beam on outside of a treatment sequence (e.g. because the design team wants to troubleshoot a problem) but inadvertently starts treatment and does not realize that the facility proceeds with reading the treatment plan.

S 2.3.3.3 (P1T2F1 & P1T2F2) operator's control algorithm is not equipped with a rule to deal with absence of patient at the treatment point (e.g. if time of treatment start is defined along with the daily treatment plan and operator asked to start treatment at exactly that time without checking for patient presence at the gantry, or with such a rule as "start treatment 5 minutes after trolley has entered the treatment room", without providing accurate feedback about possible delays) or is equipped with a wrong rule, leading him to start treatment before patient is actually brought to treatment point.

S 2.3.3.9 (P1T3F4) patient is removed from treatment point but there is a delay in the image transmission to the operator workstation, so that operator sees patient as being still at the treatment point and ready for treatment, when he isn't anymore.

S 2.3.3.10 (P1T3F4) Patient has not arrived at treatment point, but image from prior treatment is still on the screen - operator leaves his workstation between two consecutive treatments (e.g. to go to the restroom) and, upon returning, believes the new patient has been brought to treatment point and is ready for treatment.

UCA 2.3.4 Treatment is started with the wrong treatment plan (↑H1,H2) (1.3-p.5)

S 2.3.4.11 (P1T3F4) changes to daily plan or daily plan data are imperfectly communicated to the operator

S 2.3.4.12 (P2T1F3) operator selects the right steering file for loading on his workstation, but a different one is sent to TDS (is that possible at all?)

.....

3.3.2.4 *Identifying intersect and disjoint sets within the results*

Having established that the results in both studies are comparable when considered within the same scope despite different naming conventions, this paragraph explains how they intersect and whether one set is larger than the other.

To achieve this comparison, STPA unsafe control actions and scenarios are matched against those found in the draft Safety Report, as illustrated in Table 17 below.

Table 17 – Matching Safety Report scenarios with STPA unsafe control actions and causal scenarios: an illustrative example.

Safety Report (Failures and Errors)	**STPA Project (Unsafe Control Actions and Hazardous Scenarios)**
8. Error in delivered dose or beam position (SG 2 and 3)	UCA 3.2.13- Beam is not turned off when dose limit has been reached or exceeded UCA 3.2.8 - Beam is turned off too late (incl. UCA 3.2.5 : after dose limit has been reached) UCA 3.2.14 Beam not turned off when beamline settings result in beam attributes not being according to treatment plan requirements

Three types of adjustments had to be made to account for the differences in scope of the two reports. They are briefly described before the results of the comparison are provided in Table 22.

1. Excluding STPA data that corresponds to high level hazards not considered by the draft Safety Report: Because the draft Safety Report defined personnel protection and the treatment mode of continuous beam scanning as out of its scope, the UCAs listed in Table 18 were excluded from the comparison.

Table 18 - UCAs found while performing the PROSCAN STPA study but considered outside of the draft Safety Report's scope

UCA Ref.	**UCA Text**
UCA 2.3.1	Treatment is started while personnel is in room
UCA 3.1.1	Beam is turned on when personnel is close to the beamline (and causal scenarios identified via Step 2)
UCA 3.2.1	Beam is turned off too late after personnel has entered area deemed close to the beamline
UCA 3.2.17	Beam not turned off fast enough to avoid equipment damage
UCA 3.1.9	Beam is turned on when equipment is not ready to receive beam
UCA 4.1.1	The sweeper magnets bring beam to position n+1 before the dose goal for position n has been reached (and causal scenarios identified via Step 2)
UCA 5.3.1	Mode is not switched after a line was applied, so current would be 0 and dose detector accumulating noise and leading to underdose
UCA 5.3.2	Mode is not switched after a spot was applied, so current would be constant and set at an obsolete value
UCA 5.3.3	Mode is switched to dynamic during application of a spot
UCA 5.3.4	Mode would switch to static during application of a line

2. Accounting for the fact that STPA did not analyze the safety dedicated systems and highlighting differences in the degree of detail retained by both reports when documenting hazardous behavior.

The scope of the STPA project did not include evaluating how the dedicated safety systems (such as interlocks) could create hazardous situations. Further, STPA step 2 was only applied to a few of the UCAs identified by performing STPA Step 1. On the other hand, the draft Safety Report does both report on how errors in the dedicated safety systems can be hazardous and refer to very specific, STPA Step 2 like, causal factors (such as inadequate operation of specific elements). When this was the case, judgment was used to evaluate whether STPA could have identified those factors had it been performed on the dedicated safety systems and had all the unsafe control actions identified during STPA Step 1 been analyzed using STPA Step 2. Occasions where such judgment was used are respectively highlighted in light grey and italicized, as illustrated in Table 19 below.

Table 19 - Evaluating whether performing STPA Step 2 would have identified the scenarios reported in the draft Safety report: an illustrative example.

Draft Safety Report	**STPA Project**
7. Overdose scenarios (SG 1)	UCA 3.2.13- Beam is not turned off when dose limit has been reached or exceeded UCA 3.2.8 - Beam is turned off too late (inc UCA 3.2.5 : after dose limit has been reache UCA 3.1.5 Beam is turned on when dose lin has been reached or exceeded (e.g. because meant to turn beam off but ended up turning beam on)
7a Error during spot application	
E 7a.1 Dose delivery not correctly terminated by dose monitoring system 1. E 7a.2 Undetected fault in the dose monitoring system 2. E 7a.3 Dose delivery not correctly terminated by dose monitoring system 1 AND 2. E 7a.4 Local partial power failure of the dose monitoring.	Causal factors to above UCAs - faults in the feedback channel and the command generation (control algorithm and process model) (Step 2)
E 7a.5 Asynchronous reset of the watchdogs. E 7a.7 Failure to limit the maximal spot dose and duration due to an error in the watchdogs.	*Would be revealed when performing STPA on the protective measures (watchdogs are meant to protect again failures of the dose control system)*
E 7a.6 Failure at beam switch-off due to an error in the PaSS. E 7a.9 Failing of the beam switch-off safety function due to an error in the central PaSS.	Causal factors to above UCAs - faults in command generation (control algorithms and process model update) (Step 2)

It is plausible that these factors would have been found by performing STPA Step 2 on the UCAs identified by STPA Step 1

It is plausible that these factors would have been found by performing STPA on the dedicated safety system

3. Excluding semantic ambiguities when their inclusion would be to advantageous to STPA. There were a few occasions where it was ambiguous to decide whether the UCAs uncovered by the STPA analysis were included in the draft Safety Report or not. Nineteen of the 101 STPA UCAs are in this case. They are documented in Table 20 below and conservatively excluded from the comparison.

Table 20 - Cases where deciding whether the STPA results are included in the draft Safety Report is ambiguous

1. Issues associated with the On/Off beam function associated with the Setting and Resetting of the Preset in the GeCo:

UCA 3.1.1.1	Preset is set when beam is off and the dose condition for this sequence has been reached or exceeded
UCA 3.1.1.2	Preset is set after beam has been turned on and treatment in on-going
UCA 3.1.1.3	Preset is set before the previous sequence is complete
UCA 3.1.2.1	Reset is set while sequence is in progress (beam can be on or off, but dose deposition is not finalized)
UCA 3.1.2.2	Reset is set before documentation of the previous sequence is complete (hazardous unless documentation processes receive the monitor value information from another source than the GeCo)
UCA 3.1.2.3	Reset is set long before the beam is turned on (hazardous unless there is no background noise to be picked up by the monitors that will be understood as proton counts)
UCA 3.1.1.5	Preset for sequence N not set or set at a wrong value in GeCo and patient irradiation proceeds with no or wrong new preset
UCA 3.1.1.6	Preset for sequence N not set or set at a wrong value in GeCo and patient irradiation proceeds using preset values from previous dose sequences
UCA 3.1.2.4	GeCo monitor values not reset or reset to a wrong value for sequence N and irradiation proceeds (note: right value is 0; wrong value could be negative or positive)

2. Issues associated with the beam being on unbeknownst to the facility operators and while procedures are being performed in the treatment room (could fall under "8.e incorrect state of the treatment facility" but doesn't appear to be covered by the scenarios discussed in this section of the draft safety report).

UCA 1.2.5	Patient is brought to treatment point while beam is on.
UCA 1.2.9	Table is positioned incorrectly in room referential and beam is on
UCA 1.2.14	Trolley is brought to gantry coupling position while beam is on
UCA 1.2.20	Trolley is configured for table coupling to the gantry while beam is on
UCA 1.2.25	Trolley is moved to parking position while beam is on
UCA 3.2.15	Beam not turned off when beam is on, patient is in the room, but no treatment is being applied
UCA 3.2.16	Beam not turned off when beam is on, patient is in the room, treatment is in progress, and beam is being reported as being turned off

3. Possibility of havoc introduced by inadequate software provisions for repeat commands

UCA 2.3.9	Treatment start command is issued after treatment has already started
UCA 3.1.7	Beam on command is given when beam is already on
UCA 3.2.7	Beam off command is given when beam is already off

Once these adjustments were made, scenarios that were found by one of the study but not by the other were brought to the attention of the reader by highlighting them in grey and surrounding them with a box as illustrated in Table 21.

Table 21 - Highlighting scenarios uniquely documented by only one of the studies

Draft Safety Report	Correspondence in STPA
8. Error in delivered dose or beam position (SG 2 and 3)	UCA 3.2.13- Beam is not turned off when dose limit has been reached or exceeded UCA 3.2.8 - Beam is turned off too late (incl. UCA 3.2.5 : after dose limit has been reached) UCA 3.2.14 Beam not turned off when beamline settings result in beam attributes not being according to treatment plan requirements
8d Error in the therapy systems.	
E 8d.4 Selecting the wrong steering file. Other causes of wrong steering file being loaded are not discussed	Causal factor (wrong goal/input) to UCA 3.1.2 Beam is turned on when patient is in treatment room and ID does not match treatment ID UCA 2.3.4 Treatment is started with the wrong treatment plan (see the 13 scenarios identified as leading to UCA 2.3.4 that cover more pathways to the wrong steering file being read than error in file selection)

As a result of this reclassification work, Table 22 is obtained. Thanks to the highlighting scheme that was used, three conclusions stand out:

1. No text is grey highlighted and boxed in the right hand side column: all unsafe behavior documented in the draft Safety Report was found by STPA
2. Three grey boxed highlights in the left hand side column: the following three unsafe behaviors are documented by STPA and are not present in the draft Safety Report
   - Hazardous situation created by the repetition of a command when the receiving controller is not aware that he should ignore such repeat commands.
   - Possibility of patient swap in the operations area when several treatment rooms share access to one or more patient preparation rooms is ignored by the draft Safety Report.
   - Scenarios that could lead to the wrong steering file being loaded onto the Treatment Delivery Platform are not considered.
3. Twenty-five grey highlights in the right hand side: slightly less than half of the errors documented in the draft Safety Report are not described with the same level of detail by the STPA report. Since they can be nested under UCAs identified by the STPA project, it is however estimated that they would have been found had STPA Step 2 been performed.

The STPA project succeeded in identifying all the unsafe behavior described in the draft Safety Report. Albeit this is not a sufficient condition to judge STPA as being good enough for certification purposes, it was a necessary one for it to possibly be considered a candidate for this intent.

Table 22 - Hazardous scenarios uncovered respectively by the draft Safety Report and the PROSCAN STPA Study

Draft Safety Report	**Correspondence in STPA**
7. Overdose scenarios (SG 1)	UCA 3.2.13- Beam is not turned off when dose limit has been reached or exceeded UCA 3.2.8 - Beam is turned off too late (incl. UCA 3.2.5 : after dose limit has been reached) UCA 3.1.5 Beam is turned on when dose limit has been reached or exceeded (e.g. because meant to turn beam off but ended up turning beam on)
7a Error during spot application	
E 7a.1 Dose delivery not correctly terminated by dose monitoring system 1. E 7a.2 Undetected fault in the dose monitoring system 2. E 7a.3 Dose delivery not correctly terminated by dose monitoring system 1 AND 2. E 7a.4 Local partial power failure of the dose monitoring.	Causal factors to above UCAs - faults in the feedback channel and the command generation (control algorithm and process model) (Step 2)
E 7a.5 Asynchronous reset of the watchdogs. E 7a.7 Failure to limit the maximal spot dose and duration due to an error in the watchdogs.	*Would be revealed when performing STPA on the protective measures (watchdogs are meant to protect against failures of the dose control system)*
E 7a.6 Failure at beam switch-off due to an error in the PaSS. E 7a.9 Failing of the beam switch-off safety function due to an error in the central PaSS.	Causal factors to above UCAs - faults in command generation (control algorithms and process model update) (Step 2)
E 7a.8 Failing of the beam switch-off function due to an error in one or several final elements.	Causal factor to above UCAs - faults in command actuation (Step 2)
7b Error before or after the spot application E 7b.1 The patient safety and control systems are not correctly set up. E 7b.2 Miss-configuration or errors in the MPSSC modules E 7b.4 Interferences from the RPS or MCS on the beam line and beam blockers.	Causal factors to above UCAs -faults in command generation (command algorithm and process model update) (Step 2)
E 7b.3 Miss-configuration or error in the ETOT modules	*Would be revealed when performing STPA on the protective measures (ETOT is meant to protect against failure of the dose control system)*
Apparently not considered	UCA 3.1.5 Beam is turned on when dose limit has been reached or exceeded

Draft Safety Report	Correspondence in STPA
7c Unplanned beam switch-on E 7c.1 Unplanned beam switch-on during patient setup and treatment.	UCA 2.3.2 Treatment is started while patient is not ready to receive treatment UCA 3.1.8 Beam is turned on when patient is in treatment room but outside of treatment sequence
E 7c.2 External interference during the execution of the therapy.	Causal factor to above UCA

7d Excessive beam intensity	Out of Scope chosen for STPA demonstration but related causes include the following:	
	UCA 5.1.1	Beam intensity is not configured, leading to random, possibly too high, too low or extremely high current being applied
	UCA 5.1.2	Constant beam intensity is requested during static spot application
	UCA 5.1.3	High current intensity is requested before the area is cleared
	UCA 5.2.4	Beam intensity table download is interrupted before download is complete
	UCA 5.2.3	A table previously written to “odd” is subsequently overwritten by a table with destination “even”
	UCA 5.2.1	Beam intensity table is not downloaded and random or obsolete values stored in memory are processed instead
	UCA 5.2.2	Beam intensity tables are downloaded when previous table is being processed and could as a result be overwritten
	UCA 5.1.5	Beam intensity is systematically configured too late (problem when sequence contains spots with high dose/current and spots with low dose/current
	UCA 5.1.4	Beam intensity is configured before the previous element is finished
E 7d.1 COMET delivers an extraordinary beam current greater than 1000 nA. E 7d.2 The failing of the deflector plate or the use of an inadequate characteristic lead to high beam currents. E 7d.3 The variable intensity correction is wrong and causes therefore a too high current in the area of Gantry 2.		Out of Scope chosen for STPA demonstration
E 7d.4 An excessive beam current is delivered to the area due to wrong beam tunes.	Out of Scope chosen for STPA demonstration but explanatory factors include: UCA 2.3.6 Treatment is started while the beamline is not ready to receive the beam UCA 3.1.9 Beam is turned on when equipment is not ready to receive beam (hence wrong tune)	

Draft Safety Report	Correspondence in STPA
8. Error in delivered dose or beam position (SG 2 and 3)	UCA 3.2.13- Beam is not turned off when dose limit has been reached or exceeded UCA 3.2.8 - Beam is turned off too late (incl. UCA 3.2.5 : after dose limit has been reached) UCA 3.2.14 Beam not turned off when beamline settings result in beam attributes not being according to treatment plan requirements

8a Dose measurement error	
E 8a.1 Dose measurement error in dose monitoring system 1. E 8a.2 Dose measurement error in dose monitoring system 2. E 8a.3 Simultaneous failure of the high-voltage supply to the Dose Monitors. E 8a.4 Saturation of the beam monitoring systems.	Causal factors to dose related UCAs (mentioned above + others, both low and high doses) - faults in the feedback channel
E 8a.5 Excessive fluctuation in beam intensity.	Causal factors to dose related UCAs (mentioned above + others, both low and high doses) - controlled process variations

8b Error in the beam		
E 8b.1 Error in the setting of beam line to Gantry 2.	UCA 2.3.6 Treatment is started while the beamline is not ready to receive the beam UCA 3.1.6 Beam is turned on when beamline settings result in beam attributes not being according to treatment plan requirements UCA 3.2.6 Beam is turned off too late after beamline settings have changed to the point of creating a beam with attributes differing from treatment plan	
E 8b.2 Undesired changes of the beam line element parameters during the field application. E 8b.3 Errors in the beam set-up not detected before first spot application. E 8b.4 Incorrect shape, position and range of the beam due to faults in the scanning elements (sweeper, pre-absorber, nozzle cover, table). E 8b.5 Error in the sweeper magnets. E 8b.6 Error in the pre-absorber position or wrong pre-absorber. E 8b.7 No or wrong nozzle cover is installed. E 8b.8 Error in the position of the mechanical axes.		Causal factor to UCA 3.1.6 and UCA 3.2.6 above - actuators
E 8b.9 Incorrect position of the beam due to error in the computer tomograph (CT) calibration	*Would be revealed when performing STPA on the protective measures (CT is meant to protect against "wrong patient position")*	

E 8c.2 Error in the patient position offset correction.	*Would be revealed when performing STPA on the protective measures (CT is meant to protect against "wrong patient position")*
E 8c.3 Selection of the wrong patient.	UCA 1.1.2 Wrong patient or patient of unknown identity is positioned on the table
Missing - allows consideration for cases where patient "swap" is possible (e.g. if multiple patient preparation rooms can be used for the same treatment room)	UCA 1.2.1 Wrong patient or patient of unknown identity is brought to treatment point
	UCA 3.2.3 Beam is turned off too late after wrong patient has been set on treatment table UCA 3.2.11 Beam not turned off when wrong patient is in the room

8f Failure to apply the correct fractionation regime.	Out of Scope chosen for STPA demonstration - cause of Hazard H3

Draft Safety Report	Correspondence in STPA
8c Error in the patient position	UCA 3.1.4 Beam is turned on when patient position is wrong UCA 3.2.12 Beam not turned off when patient position is wrong
E 8c.1 Incorrect position of the patient	
Incorrect position of tumor vs. body (out of scope: has to do with treatment planning)	
Incorrect position of body vs. table:	
UCA 1.2.4 Patient is brought to treatment point but patient position with respect to table is wrong	
Incorrect initial position	
UCA 1.2.6 Patient is not correctly positioned on the table	
Moved during transfer to gantry or at Gantry...	
UCA 1.2.10 Trolley is not stopped when arrives at gantry coupling position ---> also can cause incorrect position of table vs. gantry	
... possibly because was made to wait for too long	
UCA 1.1.1 Patient is positioned and immobilized on table imminent to treatment	
UCA 1.1.3 Patient is positioned on table too early.	
UCA 1.2.2 Table is positioned at treatment point too early or too far from time of treatment start	
UCA 1.2.3 Table is positioned at treatment point too close to time of treatment start	
UCA 1.2.7 Table is not brought to the treatment point despite patient having been positioned and treatment room being ready to receive patient.	
UCA 1.2.26 Trolley is not brought to gantry coupling position	
UCA 2.3.11 Treatment does not start while everything else is otherwise ready	
Incorrect position of table vs. gantry	
UCA 1.2.8 Table is not correctly positioned in room referential prior to treatment start	
UCA 1.2.11 (repeat command) Trolley is commanded to move to gantry coupling position when already at gantry coupling position.	
UCA 1.2.12 Trolley is brought to gantry coupling position while GPPS is moving or is away from coupling position.	
UCA 1.2.13 Trolley is brought to gantry coupling position while coupling procedure is running.	
UCA 1.2.15 Trolley is configured for table coupling to the gantry while trolley is moving	
UCA 1.2.16 (repeat command) Trolley is configured for table coupling to the gantry while coupling procedure is already running	
UCA 1.2.17 Trolley is configured for table coupling to the gantry while GPPS is moving or is away from coupling position.	
UCA 1.2.18 Trolley is configured for table coupling to the gantry while GPPS is in coupling position for the wrong type of couch.	
UCA 1.2.19 Trolley is configured for table coupling to the gantry, but GPPS is not running the coupling procedure.	
UCA 1.2.21 Trolley is commanded to park while moving or not at the right starting position	
UCA 1.2.22 Trolley is moved to parking position while trolley coupling procedure is running	
UCA 1.2.23 Trolley is moved to parking position while GPPS is moving or is at a position other than the coupling one	
UCA 1.2.24 Trolley is moved to parking position while GPPS coupling procedure is running	
UCA 1.2.27 Trolley is not brought to correct position for coupling	
UCA 1.2.28 Table is not coupled to the gantry	
UCA 1.2.29 Table is not correctly coupled to the gantry	

Draft Safety Report	Correspondence in STPA
8d Error in the therapy systems	
E 8d.1 An incorrect spot dose is assumed by the TDS or the TVS.	Can be caused by several factors (case of "Wrong process model" for controllers TDS and TVS), e.g. UCA2.3.5 Treatment is started without a treatment plan having been loaded (and TDS assumes that obsolete data are to be used) Causal factor to all UCA involving "beam not turned off when should have been" (UCA 3.2.x) and "beam turned on when should not have been" (UCA 3.1.y) - incorrect process model
E 8d.2 Simultaneous faults in both TDS and TVS.	Causal factor to all UCA involving "beam not turned off when should have been" (UCA 3.2.x) and "beam turned on when should not have been" (UCA 3.1.y) - hardware fault, inadequate control algorithm, incorrect process model
E 8d.3 Incorrect dose requested by the therapy planning.	Causal factor to all UCA involving "beam not turned off when should have been" (UCA 3.2.x) and "beam turned on when should not have been" (UCA 3.1.y) - wrong goal/input
E 8d.4 Selecting the wrong steering file. Other causes of wrong steering file being loaded are not discussed	Causal factor (wrong goal/input) to UCA 3.1.2 Beam is turned on with patient in treatment room and ID different from treatment ID UCA 2.3.4 Treatment is started with the wrong treatment plan (see the 13 scenarios identified as leading to UCA 2.3.4 that cover more pathways to the wrong steering file being read than error in file selection)
E 8d.5 Incorrect selection of the treatment field.	Causal factor to all UCA involving "beam not turned off when should have been" (UCA 3.2.x) and "beam turned on when should not have been" (UCA 3.1.y) - incorrect control algorithm of the controllers reading the steering file (a similar example can be found at S3.1.1.3 " Spot dose controller timing algorithm or hardware platform fails to read the steering file correctly: reads it "too fast", skips instructions, gets stuck in a recursive reading of the same command.")

8e Incorrect state of the treatment facility.	UCA 3.1.6 Beam is turned on when beamline settings result in beam attributes not being according to treatment plan requirements UCA 2.3.6 Treatment is started while beamline is not ready to receive beam
E 8e.1 The patient treatment facility is not correctly set up	Cause to UCA 3.1.6
E 8e.2 Error in Mastership allocation or in the central PaSS area reservation control.	UCA 2.3.7 Treatment is started while other areas are using beam
E 8e.3 Error in facility mode User Therapy for patient treatment	UCA 2.3.8 Treatment is started while facility is in non-treatment mode
E 8e.4 Wrong calibration of the dose defining monitor 1.	Causal factor to all UCA involving "beam not turned off when should have been" (UCA 3.2.x) and "beam turned on when should not have been" (UCA 3.1.y) - inadequate feedback channel (sensor)
E 8e.5 The cyclotron delivers too much beam to the gantry 2 beam line.	Out of Scope chosen for STPA demonstration - but causal factors associated with analysis of beam intensity controller's UCAs would bring this up as either "wrong process input" (process = intensity at patient) or "inadequate actuator operation" (actuator = cyclotron)
E 8e.6 The settings of any of the beam intensity defining elements, KMA5/3, DMAD1, DMAF1, AMAKI and FMA1x are wrong.	Out of Scope chosen for STPA demonstration - but causal factors associated with analysis of beam intensity controller's UCAs would bring this up as "inadequate actuator operation" (where actuator are the intensity shaping elements in the beamline)
E 8e.7 Error in the beam tune verification system (BTVS) and data links.	Out of Scope chosen for STPA demonstration- but causal factors associated with analysis of beam intensity controller's UCAs would bring this up as "inadequate feedback" (from sensors, transmission links, and controller's process model)

Draft Safety Report	Correspondence in STPA
9. Errors in dose and position recording	UCA 2.3.10 Treatment start command is issued after treatment has been interrupted and without the interruption having adequately been recorded or accounted for
	UCA 3.1.3 Beam is turned on when treatment is in progress but no patient is at the treatment point and dose is counted towards treatment
	UCA 3.2.2 Beam is turned off without the actual dose delivery being documented (e.g. when treatment is in progress and dose condition has not been reached)
	UCA 3.2.4 Beam is turned off after dose has been delivered while no patient is in the room and with dose deposition being documented as having been delivered inside the patient's body
	UCA 3.1.10 Beam is not turned on while treatment is in progressed, but the corresponding sequence is documented as having been applied
	UCA 3.2.18 Beam not turned off when there is no patient in the room but treatment is being applied
	UCA 3.1.1.4 Preset is set before documentation of the previous sequence is complete (hazardous unless documentation processes receive the preset information from another source than the GeCo)
	UCA 2.3.3 Treatment is started when there is no patient at the treatment point (see 15 scenarios describing how this could happen. Likely consequence would be inadequate documentation of dose delivery, leading to underdose and wrong fractioning)
9a Power failure or computer crash. E 9a.1 Power failure or Therapy Control System crash.	
9b Interruption of the treatment E 9b.1 Errors following intentional interruption of the treatment. E 9b.2 Errors following unintentional interruption of the treatment.	Out of Scope chosen for STPA demonstration - *Would be revealed when performing STPA on the protective measures that protect the system against hazardous behavior by interrupting the treatment*
9c Procedures for restarting an interrupted treatment. E 9c.1 Errors at a restart following an interruption of the treatment	Out of Scope chosen for STPA demonstration - *Would be revealed when performing STPA on the protective measures that protect the system against hazardous behavior by interrupting the treatment*

3.4 Conclusion

The main conclusion derived from the PROSCAN STPA project was that using STPA to perform the hazard review of a close to final design appeared practical. Few resources were spent performing the analysis, design team members who were not safety experts could be easily taught about the STPA process and moderated to perform this analysis with recognized benefit, and the facility could be analyzed to a very low level of detail without becoming embarrassingly cumbersome.

Further, the STPA project succeeded in identifying all the unsafe behaviors documented in the draft Safety report. Three unsafe behaviors were identified that were not documented in the draft Safety Report and would warrant further attention.

In the process of performing the STPA project, several methodological questions were raised that are likely to arise in the analysis of other safety critical systems. Solutions are documented in Chapter 4, including a scheme to organize the STPA results and help make sense of the wealth of information that it brings to light.

3.5 References

AAPM (American Association of Physicists in Medicine), *Biography of Michael Goitein*, Accessed August 30th, 2012 at http://www.aapm.org/org/history/bio/1767/

Allen A.M., Pawlicki T., Dong L., Fourkal E., Buyyounouski M., Cengel K., Plastaras J., Bucci M.K., Yock T.I., Bonilla L., Price R., Harris E.E., Konski A.A., *An evidence based review of proton beam therapy: the report of ASTRO's emerging technology committee*, Radiother Oncol. 2012 Apr;103(1):8-11. Epub 2012 Mar 9.

Association SOS Irradies 31, *Plainte contre X des chefs de mise en danger de la vie d'autrui, coups et blessures non intentionnels et homicide non intentionnel,* déposée par Maitre Leguevaques, 2008

(Cah. Nutr. Diet.) Anonymous, *Alimentation et cancer*, Cah. Nutr. Diét., 36, hors série 1, 2001

Emanuel E., Pearson S., *It Costs More, but Is It Worth More?*, The Opinion Pages of the New York Times, January 2, 2012, accessed at http://opinionator.blogs.nytimes.com/2012/01/02/it-costs-more-but-is-it-worth-more/

EPICS, *About EPICS*, http://www.aps.anl.gov/epics/about.php accessed July 15, 2012

Goitein M., Radiation Oncology: A Physicist's-Eye View, Springer Science, 2008, ISBN 978-0-387-72644-1

Grutters J.P., Abrams K.R., de Ruysscher D., Pijls-Johannesma M., Peters H.J., Beutner E., Lambin P., Joore M.A.. *When to wait for more evidence? Real options analysis in proton therapy*, Oncologist. 2011;16(12):1752-61. Epub 2011 Dec 6.

Hofmann B., *Fallacies in the arguments for new technology: The case of proton therapy*, Journal of Medical Ethics: Journal of the Institute of Medical Ethics 35. 11 (Nov 2009): 684-687.

Huff C., *Catching the Proton Wave*, Hospital and Health Network, 2007, accessed August 7th, 2012 at http://www.hhnmag.com/hhnmag/jsp/articledisplay.jsp?dcrpath=HHNMAG/Article/data/03MAR2007/0703HHN_FEA_Protonwave&domain=HHNMAG

Langreth R., *Prostate Cancer Therapy Too Good to Be True Explodes Health Cost*, March 26, 2012, Bloomberg, http://www.bloomberg.com/news/2012-03-26/prostate-cancer-therapy-too-good-to-be-true-explodes-health-cost.html

Leveson N., Engineering a Safer World: Systems Thinking Applied to Safety, MIT Press, 2012, ISBN 978-0-262-01662-9

Levin W.P., Kooy H., Loeffler J.S. and DeLaney T.F., *Proton Beam Therapy*, British Journal of Cancer (2005) **93**, 849–854. doi:10.1038/sj.bjc.6602754, originally published by Nature Publishing Group

Meer D., *Developments towards advanced scanning in the Gantry 2 test area at PSI*, presentation at PTCOG 46, May 18-23, 2007, Wanjie Proton Therapy Center, China

National Cancer Institute, http://www.cancer.gov/cancertopics/types/commoncancers, last accessed July 15, 2012, quoting American Cancer Society: Cancer Facts and Figures, 2012, Atlanta, Ga: American Cancer Society, 2012.

NSE, NIT Nuclear Science and Engineering Department's virtual reading room, http://mightylib.mit.edu/Course%20Materials/22.01/Fall%202001/photons%20part%201.pdf, accessed July 15, 2012

Open Course Ware, *Principles of Radiation interactions with Matter*, Lecture notes from a Fall 2004 class in the MIT Nuclear Science and Engineering Department, http://ocw.mit.edu/courses/nuclear-engineering/22-55j-principles-of-radiation-interactions-fall-2004/lecture-notes/energy_depos_hcp.pdf, accessed July 15, 2012

Pediatric Proton Foundation, *National Association for Proton Therapy and Pediatric Proton Foundation Release First Snapshot of Treatment Data on Children at U.S. Proton Centers*, September 18, 2011 press release, http://www.prweb.com/releases/pediatric/proton_therapy/prweb8803383.htm, last accessed July 15, 2012

Pedroni E., Bearpark R., Böhringer T., Coray A., Duppich J., Forss S., George D., Grossmann M., Goitein G., Hilbes C., Jermann M., Lin S., Lomax A., Negrazus M., Schippers M., Kotrle G., *The PSI Gantry 2: A Second generation proton scanning gantry*, Z. Med. Physik, 14 (1), 25-34, 2004.

Pedroni E., D. Meer, C. Bula, S. Safai, S. Zenklusen, *Pencil beam characteristics of the next-generation proton scanning gantry of PSI: design issues and initial commissioning results*, The European Physical Journal Plus 126:66, 2011

ProtonBob, 2012a, website of the Brotherhood of the Red Balloon http://www.protonbob.com/faqs/faq05.asp, accessed July 15, 2012

ProtonBob, 2012b, http://www.protonbob.com/proton-treatment-homepage.asp, accessed July 15, 2012

PSI, 2011a – PSI website, http://radmed.web.psi.ch/asm/gantry/gantry_master.html, accessed June 2011

PSI, 2011b – PSI website, http://p-therapie.web.psi.ch/e/gantry2.html, accessed June 2011

PSI, 2012a, *Report on Proton Therapy Safety Measures for Gantry 2- Draft*

PSI, 2012b – PSI website, Facts and Figures, http://www.psi.ch/facts-and-figures accessed July 15, 2012

PSI, 2012c – PSI website, About PSI, http://www.psi.ch/about-psi accessed July 15, 2012

PSI, 2012d - *Evaluation of STPA in the Safety Analysis of the Gantry 2 Proton Radiation Therapy System* (authored by Antoine B., Rejzek M., Hilbes C.) - Draft

PTCOG, *Hadron Therapy Patient Statistics 2009*, March 2010 http://ptcog.web.psi.ch/Archive/Patientstatistics-updateMar2010.pdf accessed August 10, 2012

PTCOG, Hadron Therapy *Patient Statistics 2011*, March 2012 http://ptcog.web.psi.ch/Archive/Patientstatistics-updateMar2012.pdf accessed July 15, 2012

Samson D J, Ratko TA, Rothenberg BM, Brown HM, Bonnell CJ, Ziegler KM, Aronson N., *Comparative Effectiveness and Safety of Radiotherapy Treatments for Head and Neck Cancer*, Comparative Effectiveness Review No. 20. (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-02-0026.) Rockville, MD: Agency for Healthcare Research and Quality, May 2010. Available at:www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Stringfellow M., Owens B., Leveson N., Ingham M., Weiss K., *A Safety-Driven, Model-Based System Engineering Methodology Part I*, MIT Technical Report, December 2007.

Taheri-Kadkhoda Z., Björk-Eriksson T., Nill S., Wilkens J., Oelfke, U., Johansson K.A., Huber P. and Münter M., "*Intensity-modulated radiotherapy of nasopharyngeal carcinoma: a comparative treatment planning study of photons and protons*", Radiation Oncology 2008 3:4, doi:10.1186/1748-717X-3-4 http://www.ro-journal.com/content/3/1/4, originally published by BioMed Central

Trikalinos TA, Terasawa T, Ip S, Raman G, Lau J., *Particle Beam Radiation Therapies for Cancer. Technical Brief No. 1.* (Prepared by Tufts Medical Center Evidence-based Practice Center under Contract No. HHSA-290-07-10055.) Rockville, MD: Agency for Healthcare Research and Quality. September 2009. available at http://www.effectivehealthcare.ahrq.gov/ index.cfm/search-for-guides-reviews-and-reports/ ?pageaction=displayproduct&productid=174

Tubiana M., *La prévention du cancer et la relation dose-effet: l'effet cancérogène des rayonnements ionisants*, Cancer/Radiotherapie 13 (2009) 238-258.

Vu A.T., *Radiation Therapy of Pediatric Brain Tumors: a Comparison of Long-Term Health Effects and Costs between Proton Therapy and IMRT*, S.M. Thesis MIT/ESD and MIT/NSE, 2011. Accessed at http://dspace.mit.edu/handle/1721.1/65511

Wilson R., *Radiological Use of Fast Protons*, Radiology, November 1946 47:5 487-491

3.6 Appendix: full list of STPA study hazardous scenarios

Table 23 - PROSCAN STPA Results

Unsafe control actions and hazardous scenarios for a subset of the system's controllers and control actions

H1. OVERDOSE: Dose delivered to patient tissues (healthy tissue and tumor) is higher than clinically desirable.
H2. UNDERDOSE: Dose delivered to tumor is lower than clinically desirable.
H3. WRONG FRACTIONS: Radiation delivery is improperly fractioned.
H4. NON-PATIENT IS UNNECESSARILY EXPOSED TO RADIATION (esp. personnel and visitors)
H5. EQUIPMENT IS SUBJECT TO UNNECESSARY STRESS.

UCA 1.1.1	Patient is positioned and immobilized on table imminent to treatment (↑H-R1, H-R2) (I-p.2)
UCA 1.1.2	Wrong patient or patient of unknown identity is positioned on the table (↑H-R1, H-R2) (I-p.3)
UCA 1.1.3	Patient is positioned on table too early. (↑H-R1, H-R2) (I-p.4)
UCA 1.2.1	Wrong patient or patient of unknown identity is brought to treatment point (↑H-R1, H-R2) (II-p.2)
UCA 1.2.2	Table is positioned at treatment point too early or too far from time of treatment start. (↑H-R1, H-R2) (II-p.3)
UCA 1.2.3	Table is positioned at treatment point too close to time of treatment start. (↑H-R1, H-R2) (II-p.4)
UCA 1.2.4	Patient is brought to treatment point but patient position with respect to table is wrong. (↑H-R1, H-R2) (II-p.5)
UCA 1.2.5	Patient is brought to treatment point while beam is on. (↑H-R1, H-R2, H-R4) (II-p.6)
UCA 1.2.6	Patient is not correctly positioned on the table (↑H-R1, H-R2)(I-ie.1)
UCA 1.2.7	Table is not brought to the treatment point despite patient having been positioned and treatment room being ready to receive patient. (↑H-R1, H-R2) (II-np.1)
UCA 1.2.8	Table is not correctly positioned in room referential prior to treatment start (↑H-R1, H-R2, H-R5) (II-ie.1)
UCA 1.2.9	Table is positioned incorrectly in room referential and beam is on (↑H-R1, H-R2, H-R5) (II-ie.1)
UCA 1.2.10	Trolley is not stopped when arrives at gantry coupling position (↑H-R1, H-R2, H-R5, patient injury) (IIa.1-p.1)
UCA 1.2.11	(repeat command) Trolley is commanded to move to gantry coupling position when already at gantry coupling position. (↑H-R1, H-R2, H-R5) (IIa.1-p.2)
UCA 1.2.12	Trolley is brought to gantry coupling position while GPPS is moving or is away from coupling position. (↑H-R1, H-R2, H-R5, patient injury) (IIa.1-p.3)
UCA 1.2.13	Trolley is brought to gantry coupling position while coupling procedure is running. (↑H-R1, H-R2, H-R5, patient injury) (IIa.1-p.4)
UCA 1.2.14	Trolley is brought to gantry coupling position while beam is on. (↑H-R1, H-R2, H-R4) (IIa.1-p.5)
UCA 1.2.15	Trolley is configured for table coupling to the gantry while trolley is moving (↑H-R5, patient injury) (IIa.2-p.1)
UCA 1.2.16	(repeat command) Trolley is configured for table coupling to the gantry while coupling procedure is already running (↑H-R1, H-R2, H-R5) (IIa.2-p.3)
UCA 1.2.17	Trolley is configured for table coupling to the gantry while GPPS is moving or is away from coupling position. (↑H-R1, H-R2, H-R5, patient injury) (IIa.2-p.4)
UCA 1.2.18	Trolley is configured for table coupling to the gantry while GPPS is in coupling position for the wrong type of couch. (↑H-R1, H-R2, H-R5, patient injury) (IIa.2-p.5)
UCA 1.2.19	Trolley is configured for table coupling to the gantry, but GPPS is not running the coupling procedure. (↑H-R1, H-R2, H-R5) (IIa.2-p.6)
UCA 1.2.20	Trolley is configured for table coupling to the gantry while beam is on(↑H-R1, H-R2, H-R4) (IIa.2-p.7)
UCA 1.2.21	Trolley is commanded to park while moving or not at the right starting position (↑H-R1, H-R2, H-R5) (IIa.3-p.2)
UCA 1.2.22	Trolley is moved to parking position while trolley coupling procedure is running (↑H-R1, H-R2) (IIa.3-p.3)
UCA 1.2.23	Trolley is moved to parking position while GPPS is moving or at a different position than the coupling

one (↑H-R1, H-R2, H-R5) (IIa.3-p.4)

UCA 1.2.24 Trolley is moved to parking position while GPPS coupling procedure is running (↑H-R1, H-R2, H-R5) (IIa.3-p.5)

UCA 1.2.25 Trolley is moved to parking position while beam is on (↑H-R1, H-R2, H-R4) (IIa.3-p.6)

UCA 1.2.26 Trolley is not brought to gantry coupling position (↑H-R1, H-R2, H-R3) (IIa.1-np.1)

UCA 1.2.27 Trolley is not brought to correct position for coupling (↑H-R1, H-R2, patient injury, H-R5) (IIa.1-ip.1)

UCA 1.2.28 Table is not coupled to the gantry (↑H-R1, H-R2, H-R3) (IIa.2-np.1)

UCA 1.2.29 Table is not correctly coupled to the gantry (↑H-R1, H-R2, patient injury, H-R5) (IIa.2

UCA 2.3.1 Treatment is started while personnel is in room (↑H-R4) (1.3-p.2)

UCA 2.3.2 Treatment is started while patient is not ready to receive treatment (↑H-R1, H-R2) (1.3-p.3)

UCA 2.3.3 Treatment is started when there is no patient at the treatment point (↑H-R2, H-R3) (1.3-p.4)

- S 2.3.3.1 (P1T1F1 & P1T1F2) operator was expecting patient to have been positioned, but table positioning was delayed compared to plan (e.g. because of delays in patient preparation or patient transfer to treatment area; because of unexpected delays in beam availability or technical issues being processed by other personnel without proper communication with the operator).
- S 2.3.3.2 (P1T1F3) operator is asked to turn the beam on outside of a treatment sequence (e.g. because the design team wants to troubleshoot a problem) but inadvertently starts treatment and does not realize that the facility proceeds with reading the treatment plan.
- S 2.3.3.3 (P1T2F1 & P1T2F2) operator's control algorithm is not equipped with a rule to deal with absence of patient at the treatment point (e.g. if time of treatment start is defined along with the daily treatment plan and operator asked to start treatment at exactly that time without checking for patient presence at the gantry, or with such a rule as "start treatment 5 minutes after trolley has entered the treatment room", without providing accurate feedback about possible delays) or is equipped with a wrong rule, leading him to start treatment before patient is actually brought to treatment point.
- S 2.3.3.4 (P1T2F3) operator is asked to start treatment at a precise absolute time that is meant to coincide with time of patient positioning inside the treatment room by the nurse, but reference time for the operator differs from that used by the nurse, or nurse took longer than expected to position patient, resulting in treatment being started.
- S 2.3.3.5 (P1T3F1) operator not required to check for patient positioning before starting treatment or before resuming treatment after it was interrupted
- S 2.3.3.6 (P1T3F2) operator is not provided with real-time information about patient being installed at the treatment point
- S 2.3.3.7 (P1T3F2) patient had been positioned correctly, operator got ready to start treatment, but patient was removed from the treatment point before operator actually started treatment (e.g. because felt unwell/uncomfortable; because nurse detected an error in positioning; because table fell from the gantry....) and operator did not update his process model in time to account for this change in patient presence.
- S 2.3.3.8 (P1T3F4) patient is removed from treatment point but camera image freezes so that operator sees patient as being still at the treatment point, when he is not.
- S 2.3.3.9 (P1T3F4) patient is removed from treatment point but there is a delay in the image transmission to the operator workstation, so that operator sees patient as being still at the treatment point and ready for treatment, when he isn't anymore.
- S 2.3.3.10 (P1T3F4) Patient has not arrived at treatment point, but image from prior treatment is still on the screen - operator leaves his workstation between two consecutive treatments (e.g. to go to the restroom) and, upon returning, believes the new patient has been brought to treatment point and is ready for treatment.
- S 2.3.3.11 (P1T3F5) Camera is not working, leading either to one of the issues identified above (e.g. image freeze or transmission delay) or to an absence of image; but team decides to proceed with treatment anyway without providing enough information to the operator with respect to patient readiness for treatment. It could be that patient removal from treatment point was caused by a mechanical problem (e.g. table would

not properly coupled with gantry) that also impacted camera functioning (e.g. mechanical shock that impacted both the coupling point and the camera

S 2.3.3.12 (P2T1F2) Operator expected patient to have been positioned but was not provided with on-time feedback about the treatment floor's situation (see scenarios 5, 6, 7) and started treatment despite patient absence from the treatment point.

S 2.3.3.13 (P2T1F3) Operator issued another command or performed a set of keystrokes on his workstation that, in this context, were interpreted by TDS as requesting treatment start.

S 2.3.3.14 (P2T2F5) although operator did not issue the start treatment command, a signal (e.g. command from another user workstation area; other command issued by operator) was interpreted by the TDS / beamline actuators as necessitating treatment start

S 2.3.3.15 (P2T2F6) TDS starts to read through treatment files spuriously, e.g. because is stuck in a recursive loop following some unusual set of commands….

UCA 2.3.4 Treatment is started with the wrong treatment plan (↑H-R1,H-R2) (1.3-p.5)

S 2.3.4.1 (P1T1F1 & P1T1F2) the proper steering file failed to load (either because operator did not load it, or previous plan was not erased from system memory and overwriting is not possible) and the system uses a previously loaded one by default.

S 2.3.4.2 (P1T1F3) operator is provided with a steering file produced with a treatment plan that is wrong (error in treatment planning).

S 2.3.4.3 (P1T1F3) Operator is provided with steering file that corresponds to the right patient and the right treatment, but wrong fraction (all the more likely that steering file names are not coded for fraction, but only for patient ID), or loads a steering file from previously delivered fraction.

S 2.3.4.4 (P1T1F3) daily plan used by the operator does not match the actual sequence of patients (e.g. because scheduling changes were made that were not communicated to the operator).

S 2.3.4.5 (P1T1F3) steering files' name matches the patient ID but does not match the patient ID that is recorded inside the file (e.g. because someone – including the operator – with writing powers on the file name changed it – either purposefully or through an involuntary mistake)

S 2.3.4.6 (P1T2F1 & P1T2F2) operator not aware that he has to load the patient specific steering file, thinking that operations' planning pre-load all files corresponding to daily plan onto local area computer platform, possibly resulting in previous file being used

S 2.3.4.7 (P1T3F1) operator not required to check for patient ID before starting treatment or before resuming treatment after it was interrupted follows the treatment plan, and possibly skips a name, repeats a name or is not made aware of potential changes

S 2.3.4.8 (P1T3F2) operator is not provided with real-time information about patient ID of person being installed at the treatment point, or about fraction number

S 2.3.4.9 (P1T3F3) operator misunderstands possible change made to the daily plan and does not realize that patient order may have changed

S 2.3.4.10 (P1T3F3) Several steering files have names so similar that they are easily confused; order of files in GUI Interface repository which operator selects a file from differs from that on daily plan, leading to potential confusion.

S 2.3.4.11 (P1T3F4) changes to daily plan or daily plan data are imperfectly communicated to the operator

S 2.3.4.12 (P2T1F3) operator selects the right steering file for loading on his workstation, but a different one is sent to TDS

S 2.3.4.13 (P2T2F6) GUI fails to transmit the steering file selected by operator to TDS which, by default, uses one that had previously been loaded.

UCA 2.3.5 Treatment is started without a treatment plan having been loaded (↑H-R1,H-R2) (1.3-p.6)

UCA 2.3.6 Treatment is started while the beamline is not ready to receive the beam (↑H-R1, H-R5) (1.3-p.7)

UCA 2.3.7 Treatment is started while other areas are using beam (↑H-R1, H-R2, H-R4) (1.3-p.8)

UCA 2.3.8 Treatment is started while facility is in non-treatment mode (e.g. experiment or trouble shooting mode) (↑H-R1, H-R2) (1.3-p.9)

UCA 2.3.9 Treatment start command is issued after treatment has already started (↑H-R1, H-R2) (1.3-p.10)

UCA 2.3.10 Treatment start command is issued after treatment has been interrupted and without the

interruption having adequately been recorded or accounted for (↑H-R1, H-R2) (1.3-p.11)

UCA 2.3.11 Treatment does not start while everything else is otherwise ready (↑H-R1, H-R2) (1.3-np.1)

UCA 3.1.1 Beam is turned on when personnel is close to the beamline

- S 3.1.1.1 (P1T1F1) The Spot Dose Controller does not receive any input from the above controller and proceeds with an irradiation program that was left in its memory and not cleared, possibly while personnel is in the treatment room or close to the beamline.
- S 3.1.1.2 (P1T1F2) Local area operator request beam on (starts treatment or proceeds with some other activity requiring beam on) while personnel is close to the beamline (hasn't left the treatment room or other sections in close contact to the beamline) and Spot Dose Controller abides to the request.
- S 3.1.1.3 (P1T1F2) Spot dose controller timing algorithm or hardware platform fails to read the steering file correctly: reads it "too fast" or skips several instructions resulting in early read of a later command sequence, gets stuck in a recursive reading of the same command.
- S 3.1.1.4 (P1T1F3) Spot dose controller interprets a signal (spurious signal, other command) from higher level controllers as a command to start treatment or download the steering file while personnel is close to the beamline
- S 3.1.1.5 (P1T2F1) The Spot Dose controller is not programmed to check that personnel is away from the beamline and the treatment room before turning beam on (both at treatment start and when starting a new spot in an on-going treatment).
- S 3.1.1.6 (P1T2F2) The Spot Dose Controller includes rule that permit beam on when personnel is known to be close to the beamline.
- S 3.1.1.7 (P1T2F3) The Spot Dose Controller's control algorithm does not allow for enough time for personnel to exit beam-exposed areas after they have cleared their presence out. For example, it does not include the time variable at all, includes a wrong parameter for the time variable because of inadequate knowledge of the time it takes people to exit the beamline environment or larger than warranted expectations of the magnet's actuation time, the clock is too fast…
- S 3.1.1.8 (P1T3F1) Process Variable “personnel presence” is absent from the Spot Dose Controller’s control algorithm.
- S 3.1.1.9 (P1T3F1) There are no channel to inform the spot dose controller of the presence of personnel or non-patients (e.g. family members) in beam-exposed areas
- S 3.1.1.10 (P1T3F2) Although a direct feedback channel exists, the controller's processing unit does not update its process model at regular enough intervals, possibly leading to a situation where the room was adequately considered as cleared when personnel left areas close to the beam, but not revised as not cleared when they eventually returned to these areas faster than the process model was updated.
- S 3.1.1.11 (P1T3F3) Although the process variable if in interest is included in the Spot Dose Controller’s control algorithm and is provided with adequate feedback channel, the controller’s processing of that information is inadequate and leads to erroneous interpretation of the value of that process variable.
- S 3.1.1.12 (P1T3F4) Information about the value of the process variable of interest is distorted or lost during transmission to the controller.
- S 3.1.1.13 (P2T1F1) Hardware failure prevented emission of the otherwise safe command by the Spot Dose Controller
- S 3.1.1.14 (P2T1F3) The kicker magnet interpreted a different (another command, or spurious noise) signal as prompting the opening of the beam and does so while personnel is in radiation-exposed areas.
- S 3.1.1.15 (P2T2F2) The Beam On command was generated at a time when no personnel was close to the beamline but, before it was executed and without it having been cancelled, personnel re-entered the beam exposed areas.
- S 3.1.1.16 (P2T2F3) Power source to the kicker magnet suffers from voltage discontinuities that result in kicker magnet being turned on, turning beam on when it should not.
- S 3.1.1.17 (P2T2F3) Beamline design is such that when breaking down (from wear, from missing

power or cooling input, from unanticipated behavior such as extreme hysteresis), the kicker magnet opens the beamline lets beam through without having been prompted to do so

S 3.1.1.18 (P2T2F3) Sensitivity to environmental factors (e.g. temperature?) or stage of treatment (e.g. does kicker magnet's reliability depend on number of cycles it was engaged in previously ....?) is not adequately factored in and lead to unanticipated opening of the beamline.

S 3.1.1.19 (P2T2F3) Kicker magnet cannot withstand keeping the "beamline off" setting for too long and automatically resets after a certain period of time, that might be smaller than the desired interval between two consecutive updates to the process variable of interest.

S 3.1.1.20 (P2T2F3) Kicker magnet support systems (e.g. power source, cooling system) cannot withstand keeping the "off" setting for too long and automatically resets after a certain period of time.

S 3.1.1.21 (P2T2F3) External disturbances stuck the beamline in an open position; beam cannot be turned off, remote diagnostic means are not available, and personnel has to be sent close to the beamline to take remedial action

S 3.1.1.22 (P2T3) Another controller issues a beam on command while personnel is in the room

S 3.1.1.23 (P2T3) Spot Dose controller and another controller with access to AMAKI issue simultaneous commands that result in actuator misinterpreting the command or in an information jam that results in opening of the beamline if a "confused" AMAKI local controller puts it in a "open by default" mode. AMAKI letting beam through due to a “confused” AMAKI local controller

UCA 3.1.2 Beam is turned on when patient is in treatment room and ID does not match treatment ID

UCA 3.1.3 Beam is turned on when treatment is in progress but no patient is at the treatment point and dose is counted towards treatment

UCA 3.1.4 Beam is turned on when patient position is wrong

UCA 3.1.5 Beam is turned on when dose limit has been reached or exceeded

UCA 3.1.6 Beam is turned on when beamline settings result in beam attributes not being according to treatment plan requirements

UCA 3.1.7 Beam on command is given when beam is already on

UCA 3.1.8 Beam is turned on when patient is in treatment room but outside of treatment sequence

UCA 3.1.9 Beam is turned on when equipment is not ready to receive beam

UCA 3.2.1 Beam is turned off too late after personnel has entered area deemed close to the beamline

UCA 3.2.2 Beam is turned off without the actual dose delivery being documented (e.g. when treatment is in progress and dose condition has not been reached)

UCA 3.2.3 Beam is turned off too late after wrong patient has been set on treatment table

UCA 3.2.4 Beam is turned off after dose has been delivered while no patient is in the room and with dose deposition being documented as having been delivered inside the patient's body

UCA 3.2.5 Beam is turned off too late after dose limit has been reached

UCA 3.2.6 Beam is turned off too late after beamline settings have changed to the point of creating a beam with attributes differing from treatment plan

UCA 3.2.7 Beam off command is given when beam is already off

UCA 3.2.8 Beam is turned off too late

UCA 3.1.10 Beam is not turned on while treatment is in progressed, but the corresponding sequence is documented as having been applied

UCA 3.1.11 Beam is only partially turned on while patient is at the treatment position and treatment is in progress

UCA 3.1.12 Beam is only partially turned on outside of treatment sequence while patient is at the treatment position

UCA 3.1.13 Beam is only partially turned on while treatment is in progress but no patient is at the treatment position and dose is counted as having been applied delivered

UCA 3.1.14 Beam is only partially turned on

UCA 3.2.10 Beam not turned off when personnel becomes close to the beamline

UCA 3.2.11 Beam not turned off when wrong patient is in the room

UCA 3.2.12 Beam not turned off when patient position is wrong

UCA 3.2.13 Beam not turned off when dose limit has been reached or exceeded

UCA 3.2.14 Beam not turned off when beamline settings result in beam attributes not being according to treatment plan requirements

UCA 3.2.15 Beam not turned off when beam is on, patient is in the room, but no treatment is being applied

UCA 3.2.16 Beam not turned off when beam is on, patient is in the room, treatment is in progress, and beam is being reported as being turned off

UCA 3.2.17 Beam not turned off fast enough to avoid equipment damage

UCA 3.2.18 Beam not turned off when there is no patient in the room but treatment is being applied

UCA 3.1.1.1 Preset is set when beam is off and the dose condition for this sequence has been reached or exceeded

UCA 3.1.1.2 Preset is set after beam has been turned on and treatment in on-going

UCA 3.1.1.3 Preset is set before the previous sequence is complete

UCA 3.1.1.4 Preset is set before documentation of the previous sequence is complete (hazardous unless documentation processes receive the preset information from another source than the GeCo)

UCA 3.1.2.1 Reset is set while sequence is in progress (beam can be on or off, but dose deposition is not finalized)

UCA 3.1.2.2 Reset is set before documentation of the previous sequence is complete (hazardous unless documentation processes receive the monitor value information from another source than the GeCo)

UCA 3.1.2.3 Reset is set long before the beam is turned on (hazardous unless there is no background noise to be picked up by the monitors that will be understood as proton counts)

UCA 3.1.1.5 Preset for sequence N not set or set at a wrong value in GeCo and patient irradiation proceeds with no or wrong new preset

UCA 3.1.1.6 Preset for sequence N not set or set at a wrong value in GeCo and patient irradiation proceeds using preset values from previous dose sequences

UCA 3.1.2.4 GeCo monitor values not reset or reset to a wrong value for sequence N and irradiation proceeds (note: right value is 0; wrong value could be negative or positive)

UCA 4.1.1 The sweeper magnets bring beam to position n+1 before the dose goal for position n has been reached

- S 4.1.1.1 (P1T1F2) Push commands from the external timing reference are received at a higher than 100 kHz rate.
- S 4.1.1.2 (P1T1F2) SDS receives a trigger signal from TDS before a valid delivery table for {U,T} has been downloaded and proceeds with application of an invalid delivery table
- S 4.1.1.3 (P1T1F2) (transmission) SDS interpreted a spurious signal between TDS and SDS as trigger and would start with the application of the next delivery table element. [1]
- S 4.1.1.4 (P1T1F2) (transmission) SDS interpreted a spurious signal as push command from the external timing reference and would move to the next position N+1.
- S 4.1.1.5 (P1T1F2) TDS believes current element has been fully completed when it hasn't, resulting in TDS trigger command being issued too early.
- S 4.1.1.6 (P1T1F2) TDS downloads a delivery table too early resulting in overwriting a valid delivery table (not yet applied) with another valid delivery table.
- S 4.1.1.7 (P1T1F3) Steering file read by TDS and, therefore, delivery tables downloaded by SDS code for faster than clinically desirable beam movement (e.g. because time step assumed between sequential settings is larger than that actually used by SDS)
- S 4.1.1.8 (P1T1F3) Steering file ready by TDS and, therefore, delivery tables downloaded by SDS code for beam movement that is faster than dynamics of intensity modulation
- S 4.1.1.9 (P1T1F3) SDS downloads a corrupt delivery table from TDS. (This includes downloading only a part of the delivery table)
- S 4.1.1.10 (P1T1F3) SDS downloads the delivery table from TDS to an “odd” table memory space while the next to be applied table is supposed to be an “even” table
- S 4.1.1.11 (P1T2F1) SDS is configured with a delivery table which is not conformal with the restriction that the x value of the last triplet must be <2^21-1 (the table holds 23 bits so theoretically a value of 2^23-1) is possible. The interpolation algorithm will fail unless

it is provided with a rule to handle that situation.

S 4.1.1.12 (P1T2F1) SDS is configured with a delivery table that does not define a triplet for t0 (meaning the x value of the first triplet is >0). The interpolation algorithm will wrongly use the last values of the previous element unless it is coded with a rule to better handle that situation.

S 4.1.1.13 (P1T2F1) SDS does not check for completion of dose delivery, or conformity of intensity levels with expectations, in location N before moving to location N+1.

S 4.1.1.14 (P1T2F3) Time steps used by SDS (parameter xStep) do not match those assumed in the delivery table (especially, xStep is shorter than assumed in the delivery tables).

S 4.1.1.15 (P1T2F3) SDS skips entries in the delivery tables.

S 4.1.1.16 (P1T3F1) The designers have not included means for SDS to know whether the dose condition has been reached and whether the beam can thus be safely swept. Hence the process relies heavily on the intensity modulation of the proton beam being performed on par with expectations, and in synch with sweeping movement.

S 4.1.1.17 (P2T1F1) A defect of the POF-transceivers (plastic-fiber-optic-transceiver) at either SDS on the Sweeper Magnet side or POF cable brake prevents magnet current controller from receiving SDS commands AND magnet current controller is programmed to modify magnet current in such a way that beam is swept at a speed non-correlated with steering file expectations, and possibly faster than desired.

S 4.1.1.18 (P2T1F2) Signal from SDS to magnet controller travels faster than expected by treatment planning team, resulting in earlier beam sweep than clinically desirable.

S 4.1.1.19 (P2T1F3) Signal integrity is not ensured (e.g. because of a bad POF cable interconnection or a blurred cable) and parameters from an SDS command are distorted, leading to magnet moving farther than desired.

S 4.1.1.20 (P2T1F3) Since the link to the sweeper magnet is not only used to request a new position but also for general data transfer (like reading out status information or warnings and errors) overloading the link can result in signal distortions when information fluxes are not prioritized correctly by the firmware: another command or information request (e.g. status information) might be distorted to resemble a magnet current setting command, resulting in earlier than desired magnet configuration (magnet moves beam to right position earlier than desired), or wrong configuration (magnet moves beam farther than desired).

S 4.1.1.21 (P2T2F2) Magnet current controller fails to update settings requested by SDS command, and stays stuck in an autonomous and uncontrolled setting variation procedure (e.g. maintains constant beam sweep speed when intensity modulation had been coded assuming variable beam sweep speed).

S 4.1.1.22 (P2T2F3) A miss configuration or power loss of the sweeper magnet power supply controller could result in the command not being followed. losing power and consequently loosing the magnetic field of the sweeper magnets result in the beam position moving back to the iso-center (no beam displacement) which can be interpreted as “undesired movement” too

S 4.1.1.23 (P2T2F4) A large difference between the last and current set point can result in a delayed actuator behavior when the power supply or magnet is not possible to increase/decrease the magnetic field appropriately.

S 4.1.1.24 (P2T2F4) Magnet actuation speed is faster than was anticipated by the treatment planning team (example: sensitivity to environmental factors (e.g. temperature?) or stage of treatment (e.g. does actuation speed depend on the number of cycles previously performed; could voltage ramp-up speed be slower at treatment start than later ....?) is not adequately factored in and leads to faster than expected response).

S 4.1.1.25 (P2T2F5) Because of environmental disturbances, current source yields larger/smaller current than desired, resulting in different magnetic field than desired, resulting in beam being swept further than desired (i.e. preventing complete dose delivery in desired location n and n+1)

S 4.1.1.26 (P2T2F5) Magnet support systems (e.g. power source, cooling system) cannot withstand keeping the same setting for too long and automatically resets after a certain period of time, that might be smaller than the desired interval between two consecutive settings.

S 4.1.1.27 (P2T2F6) Magnet cannot withstand keeping the same setting for too long and automatically resets after a certain period of time, that might be smaller than the desired interval between two consecutive settings.

S 4.1.1.28 (P2T2F6) Possibly early or farther than desired actuation because of hysteresis effects

UCA 5.1.1 Beam intensity is not configured, leading to random, possibly too high, too low or extremely high current being applied

UCA 5.1.2 Constant beam intensity is requested during static spot application

UCA 5.1.3 High current intensity is requested before the area is cleared

UCA 5.1.4 Beam intensity is configured before the previous element is finished

UCA 5.1.5 Beam intensity is systematically configured too late (problem when sequence contains spots with high dose/current and spots with low dose/current

UCA 5.2.1 Beam intensity table is not downloaded and random or obsolete values stored in memory are processed instead

UCA 5.2.2 Beam intensity tables are downloaded when previous table is being processed and could as a result be overwritten

UCA 5.2.3 A table previously written to “odd” is subsequently overwritten by a table with destination “even”

UCA 5.2.4 Beam intensity table download is interrupted before download is complete

UCA 5.3.1 Mode is not switched after a line was applied, so current would be 0 and dose detector accumulating noise and leading to underdose

UCA 5.3.2 Mode is not switched after a spot was applied, so current would be constant and set at an obsolete value

UCA 5.3.3 Mode is switched to dynamic during application of a spot

UCA 5.3.4 Mode would switch to static during application of a line

4 Lessons Learned: Processes for Carrying out the Analysis and Organizing the Results

Look deep, deep, deep into nature, and then you will understand everything.
Albert Einstein (1879-1955)

Il ne faut rien laisser au hasard.
French idiom *(Leave nothing to chance)*

L'arbre cache souvent la forêt.
French proverb *(One can't see the wood for the tree)*

4.1 Introduction

The PROSCAN-STPA project was undertaken to evaluate the applicability of STPA to the safety review of a complex system whose behavior is mostly under software control and allow the PROSCAN team to develop in-house expertise in the use of STPA. It demonstrated that STPA was indeed applicable, showed that it provides an adequately flexible framework that can be tailored to the needs of the user and the degree of detail she wants to use in her investigation of safety issues, and included a PROSCAN team member who became a key performer of the analysis. In addition, this project surfaced several methodological questions of interest to the study of other complex systems.

Some of these questions had to do with the analysts' initial lack of familiarity with the STAMP framework. Others were due to the fact that they were not using STPA to identify hazards and create design requirements, but were instead dealing with a system already in existence. For example, putting together the first control structures proved challenging as the analysts knew the system's physical architecture so well that it was difficult for them to abstract it back to its functional representation. Others were more fundamental and are likely to arise in other projects.

Processes to help analysts carry out an STPA analysis and organize its results were developed as a response to these issues.

4.2 Processes for carrying out an STPA analysis

Starting with a specification by the system stakeholders of the losses that they want to prevent, STPA is a top-down method that aims to provide safety designers with a systematic approach to identify hazardous scenarios and create safety requirements (and thus design features or operational procedures) that will prevent their occurrence.

As explained in the Background Chapter, an STPA analysis flows as follows:

- Define the system goal
- Specify the losses to be prevented
- Identify the high-level hazards that will lead to these losses and specify high-level safety constraints that will prevent their realization
- Define the system boundaries and create a control structure representation of the functions performed by the system elements
    - For each control loop, identify the actions that must be performed in this loop
        - STPA Step 1: For each of these control actions, identify the associated unsafe control actions that would lead to safety constraints being breached
        - STPA Step 2: For each unsafe control action, identify the scenarios through which it could be realized.

Once these scenarios have been identified, they can be used to either generate safety requirements that system designers will be tasked with addressing if the system is still in design stage or to check that the system design effectively prevents their occurrence, reduces their likelihood or mitigates their consequences if the system is undergoing a safety review.

Questions were raised and answered about the creation of functional control structures, the definition of control actions, and the causal analysis of STPA Step 2.

4.2.1 Questions about the creation of the PROSCAN/G2 control structures

Once system boundaries are defined, a control structure can be drawn to model the system of interest. It describes the control and feedback relationships that link the system's functional elements together and serve as the backbone used in the rest of the analysis.

Mapping the existing architecture of the project subject (PROSCAN's Gantry2) into the functional framework on which STAMP is based proved to be less trivial than expected. The following questions were raised in that process:

1 - What level of detail should one adopt to ensure the thoroughness of the analysis without unnecessarily expanding the analysis effort?

2 - Should the control structure include safety-dedicated elements or should it be a representation of the system's elements reduced to their operational functions?

3 - How should one model a controller that takes the form of allowing or forbidding (through the use of "veto") another controller from providing commands to a system element?

In the context of the PROSCAN STPA study, where the physical realization of the system was mostly complete, the questions were answered as described below.

4.2.1.1 What level of detail should be chosen to perform the analysis?

It proved challenging for the PROSCAN technical experts brought to work on the STPA project to step back from the component-oriented descriptions of the facility's physical architecture that they knew in great detail and adopt the functional representation adopted by STAMP and STPA.

To facilitate that process, multiple control structures were drawn, describing the system with increasing degrees of detail, or increasing depth levels. D0 provides the higher level, most abstract, more superficial representation of the system's functional relationships. D1, D2 etc. zoom into how these functions are divided among higher and lower level controllers. This refinement process of the control structures led to another question: at what degree of detail should the analysis be performed to be most useful?

The plurality of views allowed by the multi-degrees of detail description of the facility naturally supports a multi-level STPA analysis where safety requirements are derived from a top-down analysis. First, high-level constraints are identified when applying STPA to the more abstract representation of the system, such as describing the PROSCAN facility into two sub-systems: the Treatment Definition group and the Treatment Delivery group. Then, applying STPA on more detailed control structures, lower-level requirements are identified that must contribute to enforcing the higher level safety constraints.

What matters more than choosing the degree of detail at which the analysis will be performed is that both the representation and the analysis be performed top-down. Delving into a deeper level of detail will only refine the findings obtained at the higher level of detail. The added precision is valuable as it leads to the enumeration of specific requirements. It is proposed that the highest degree of detail to be analyzed be the last at which design decisions need to be made.

To further facilitate the conversion of component-focused professionals to the functional lens adopted by STPA, this dissertation suggests focusing on process attributes to start a description of the overall system. When describing in detail how the system's elements influence the controlled process (irradiation of patient in the PROSCAN case) and with respect to the radiation-related hazards listed at the beginning of the analysis, it was indeed useful to distinguish between several process attributes whose simultaneous control is needed to ensure that the treatment is delivered as intended:

- beam tune (energy and shape)
- beam intensity
- beam on/off state
- beam to target alignment.

In a way, not only were the control structures refined, but so was the controlled process, understood to be the combination of four distinct although mutually dependent process attributes.

4.2.1.2 How should "veto" controllers be modeled?

Rather than asking a vigilant system to shut the beamline when conditions for safe “beam on” are achieved, PROSCAN instead requires that all controllers capable of blocking the beam be aware

that the conditions for safe “beam on” are achieved and individually allow the beam to be turned on for that command to be executed. This is akin to these controllers having a veto power that prevents the beam on command to be implemented by the beamline elements in situations that could lead to an unsafe state.

As on one hand the command flows through this controller to lower level actuators, but on the other they are more powerful than the command issuing controller because they can prevent its command from being executed, it is unclear what their position should be in the control structure. Do they control the command issuing controller (since they can force the facility to ignore its commands) or are they controlled by it (since they only act as “command filters "or command “gates”, and not command issuers)?

STAMP detailed process loops make provisions for the existence of external input into the control loop. They can be input to the controller (e.g. command from a higher level controller, information from out of loop sensors), to the actuator (e.g. power necessary for activating the actuator) or to the controlled process (e.g. output of a related controlled process). One solution is to add another category of external interference: that of other controllers whose activity may affect the controlled process or the actuator. This extension benefits from the work of Ishimatsu et al. (2011) on multiple controllers STPA analysis.

Another option consists in considering them as a lower-level controller, positioned in the process loop between the command issuing controller and the lower level actuator. The "veto controller's" rebuttal to the command transmission request produced by the higher level controller can indeed be interpreted as feedback informing that controller that the facility is not ready for the command to be implemented.

4.2.1.3 Should the model of the system under consideration include safety dedicated elements?

The PROSCAN facility includes several protective systems whose role is to shut the beam off in the event that a hazardous state is detected. The goal of the hazard review being to evaluate their adequacy, it was decided to model and analyze the system without these protective features. Doing so allowed us to identify safety issues that had to be eliminated from the design or, when elimination strategies were not available, protected and mitigated against. The existing protective

measures could then be evaluated with respect to the extent to which they addressed the issues identified by the hazard analysis. The final analytical step consists in verifying whether the safety dedicated elements are capable of causing hazardous behavior themselves.

4.2.2 Questions asked about STPA Step 1 & Step 2

4.2.2.1 Should all controller actions be considered as control actions?

The nurse brings the patient to the CT-scanner before accompanying him into the treatment room. There, she is in charge of positioning the patient within the CT-scanner so that its position relative to the table is measured, and position offsets are sent to software tasked with incorporating them in the steering file. Should the nurse's action in obtaining CT-scan position offsets be considered as a control action on which STPA step 1 and 2 ought to be performed?

This example illustrates the fact that controllers, especially humans, can participate in several distinct control loops in each of which their roles can be different. When positioning the patient on his table, the nurse participates in a first control loop where she uses fixation devices (actuators) to immobilize the patient on the couch (controlled process = patient position). When obtaining CT confirmation of the patient position, she belongs to a second control loop, being part of the feedback channel to the treatment delivery system whose steering file will be updated to account for possible changes in patient position. However, when zooming into that second loop, the "feedback" nurse becomes a controller in a third control loop: she is in charge of setting up the patient's table into the CT-scan to ensure adequate position evaluation.

Depending on the degree of details, not all actions should be considered as control actions. At the more abstract degree of detail where the nurse controls patient position, her responsibility in setting up the CT-scan feedback channel is not to be scrutinized as that feedback could well be provided by other means. However, when looking into her role with respect to operating the CT-scanner, then this very same set of actions must be studied in full detail.

4.2.2.2 How can we make STPA Step 2 easier?

4.2.2.2.1 Motivation for providing additional heuristics to perform STPA Step 2

(Thomas, 2011) proposed a very detailed process that defines the steps of STPA Step 1 with such precision that they could be performed by a computer with limited human input. It also offers an attractive format for reporting the outcome of STPA Step 1 that is likely to be adopted as a

standard by STPA users. In contrast, the practice of STPA Step 2 would still benefit from the introduction of additional heuristics and templates. Leveson (2003) created a list of generic causal factors whose use is expected to guide the analyst in identifying design limitations that must be addressed to ensure that the safety constraints are enforced. (Leveson, 2012) maps these control flaws on a process loop. They can be used as guidewords for the Step 2 search of hazardous scenarios that lead to unsafe control actions (see Figure 14).

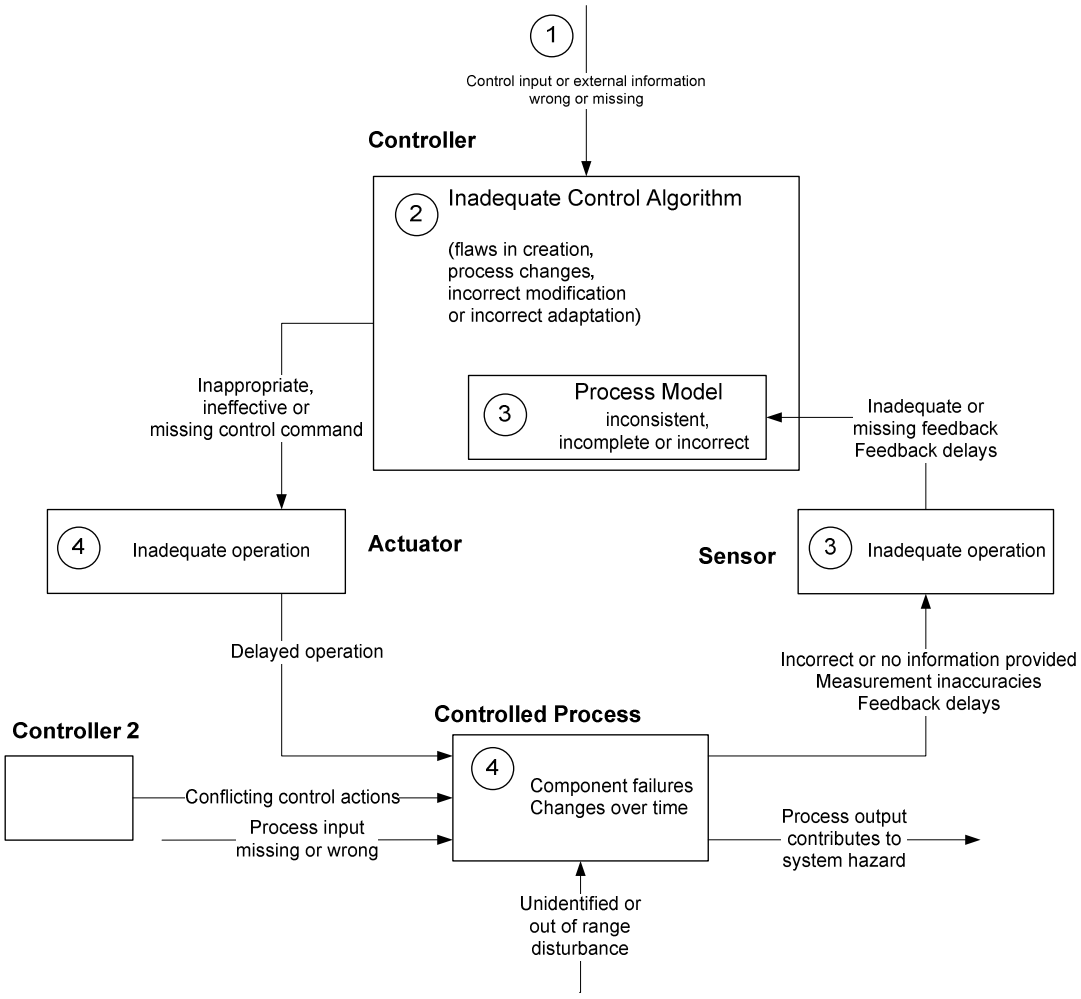


Figure 14 - A classification of control flaws leading to hazards (Leveson, 2012)

(Leveson, 2012) mentions that this process loop can also be edited to document the specific causal factors found in a given analysis. This visual representation is attractive for how simple it makes telling the stories of how the unsafe scenarios unfold. However, it is not practical when large numbers of scenarios can be found to explain a single unsafe control action (e.g. the average 16.8 scenarios per unsafe control action found in the PROSCAN study) and does not

allow for a detailed narration of the scenarios involved. It also fails to show the path through which certain causal factors contribute to the unsafe control action. For example, inadequate sensor operation and process model inconsistency are not disjoint explanations of how an unsafe control action can occur: inadequate sensor operation is in fact unsafe because if makes the process model inconsistent. Finally, because it would entail the duplication of the same control structure and repetition of similar information, it is not very efficient when several unsafe control actions are associated with the same control loop and share common causal factors.

In an attempt to address these concerns and provide additional structure to the analyst performing STPA Step 2, the information presented in Figure 14 was reorganized in a tree meant to highlight the common pathways through which causal factors lead to unsafe control actions. This "Step 2 Tree" was successfully used to perform STPA Step 2 in three of the five examples of the PROSCAN STPA project.

4.2.2.2.2 The Step 2 Tree: proposal for exhaustive listing of hazardous scenarios

The Step 2 Tree approach starts from the premise that control actions are the result of control commands being successively provided by the controller, transmitted to the actuator, processed and finally implemented by the actuator. Unsafe control actions can therefore be thought of as belonging to one of two categories:

- the safety constraints were not enforced by the controller and the controller generated an unsafe control command, or
- appropriate control commands were provided but not followed.

As a consequence, they are understood to be caused by two separate categories of hazardous scenarios: those that affect command generation, and those that affect command implementation. These general categories can, in turn, be described as consisting of several sub-categories of scenarios, which, in turn, contain several sub-sub-categories of scenarios etc. This “Russian doll” structure (or hierarchical and top-down refinement), where a given sub-category is one of several ways that the higher level category is realized and potentially creates the unsafe control action of interest, is well organized using a tree.

The Step 2 Tree starts with two nodes (P1, P2). Each of these nodes is attached to lower level nodes (T_i, then F_j), causal factors through which the high level nodes can be caused[50]. This hierarchy of causal factors categories is depicted in Figure 15. The first nodes of the tree are connected as follows:

- P1: controller generated an unsafe control command,
  - P1/T0: unidentified hazard
  - P1/T1: missing rule[51], wrong or inadequately timed control input
  - P1/T2: inadequate control algorithm
  - P1/T3: inadequate process model
- P2: appropriate control commands were provided but not followed
  - P2/T1: inadequate transmission[52] of a safe control command,
  - P2/T2: inadequate execution of a safe control command.
  - P2/T3: conflicting control actions are issued by different controllers[53].

[50] P, T and F are legacy notations. The first time this tree was used, it was described with an analogy to human population studies where populations (P) are understood to be made up of tribes (T) that are themselves made of families (F).

[51] missing a rule for unknown situations (e.g. unanticipated signal in the command or feedback lines)

[52] following an analogy with the semantics used in communication studies, we will understand transmission of a command to include: the issuance by the controller of a command that its internal logic generated (issuing), the transport of the message to its recipient (transport), and the processing/understanding of the message received by the recipient (reception). We acknowledge that unintended loss, delay and distortion of information can happen at each of the three points of this transmission chain.

[53] Inadequate coordination between different controllers (see (c) and (f) in

Figure 16) can explain wrong goal input to the controller (P1T1F1), and perturb the command communication channel such that the command is lost (P2T2F2), distorted (P2T1F3) or delayed (P2T1F2). We could choose to leave this cause of unsafe control actions implicit and advise that it be categorized with P1T1F1, P2T2F2, P2T1F3 or P2T1F2. However, we prefer to categorize these issues as an independent 2nd level node (P2/T3) to bring the attention of the analyst and the designer to the interaction and coordination issues that arise from allowing controllers to share control of the same actuators or controlled processes, especially in situations where we anticipate that coordination between multiple controllers may require special attention (e.g. when multiple controllers command the same actuator).

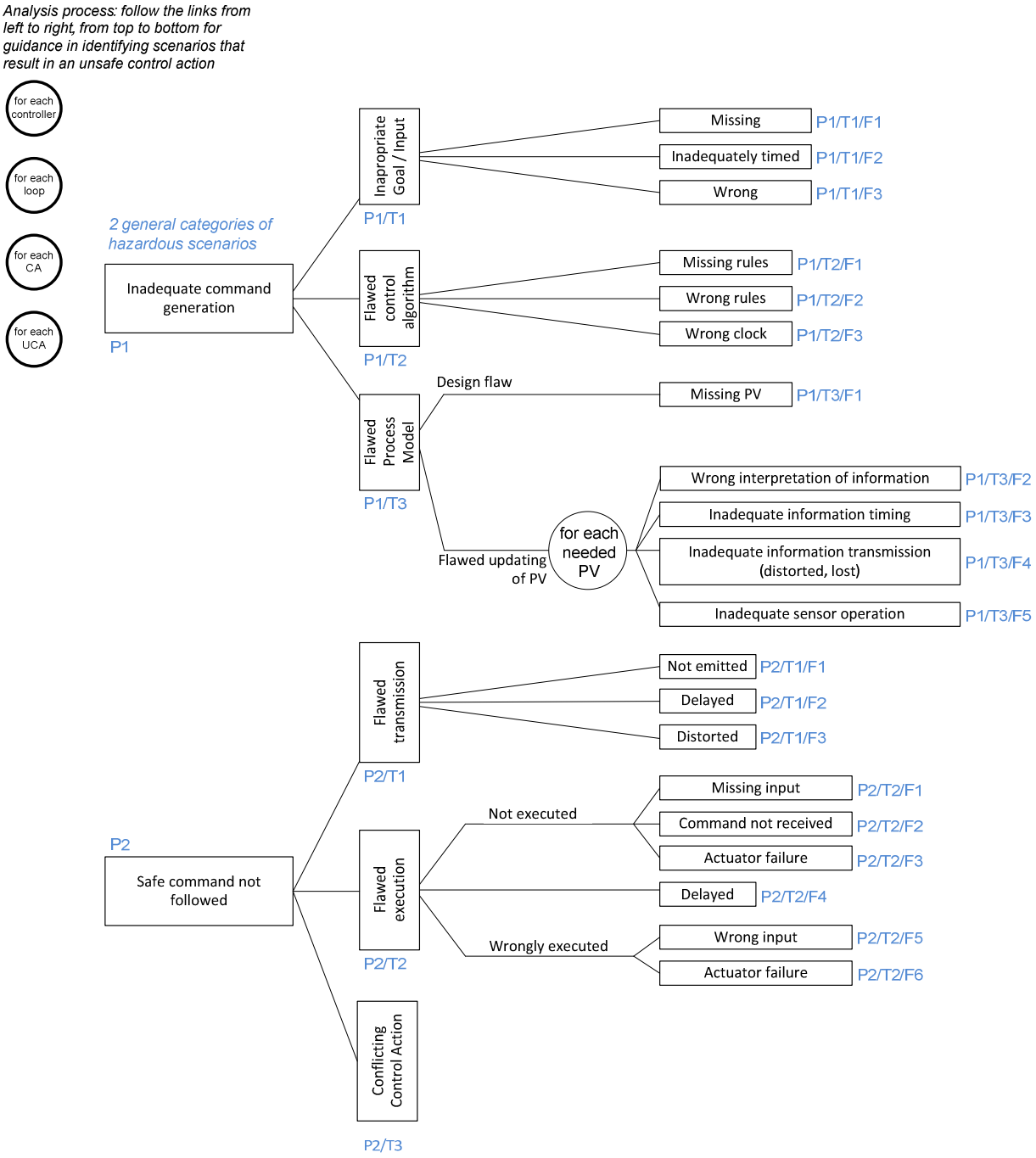


Figure 15 - The Step 2 Tree

Figure 16 maps the Step 2 Tree categories to the causal factors listed in Figure 15.

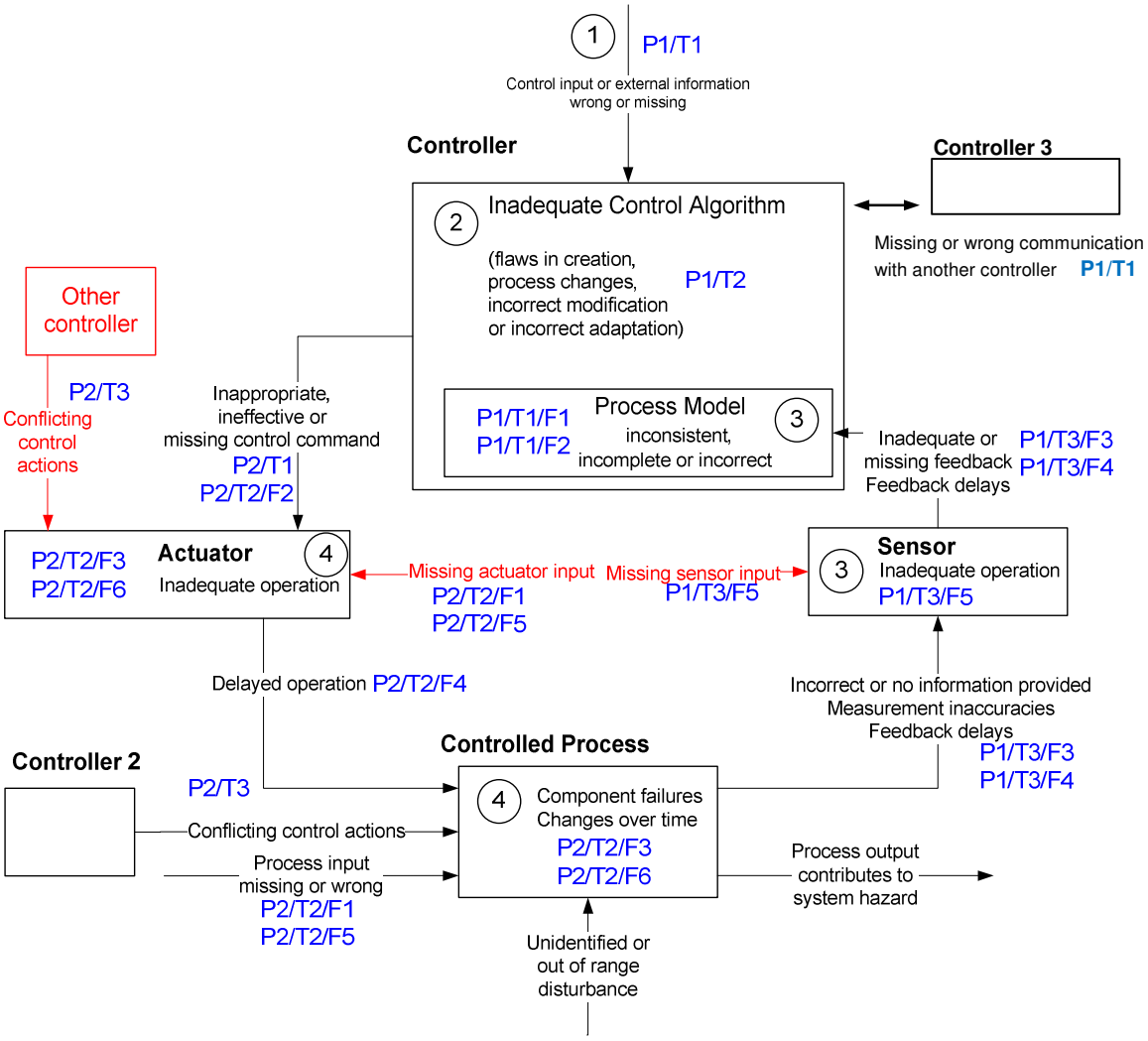


Figure 16 – Mapping Step 2 Tree categories onto the Control Flaws Process Loop[54]

Red elements correspond to additions to (Leveson, 2012 – updated in personal communication by adding Controller 3).

This nomenclature is meant to provide guidance to the analyst; it can in no way substitute for solid understanding of how the system works under a diverse set of conditions. The analyst would pick one unsafe control action identified by STPA Step 1 at a time and work his way through the Step 2 Tree to generate causal scenarios with the help of these guidewords. One family of scenarios after the other, he would basically answer the question "is there a scenario in this family of theoretical scenarios that would contribute to the unsafe control action that I want

[54] The red arrows referring to sensor and actuator inputs represent physical inputs (e.g. power) or logic inputs (e.g. calibration information) that are necessary to the sensor and actuator's operations and are provided by elements outside of the process loop under consideration.

to prevent?" by writing out the "story" based on this family's guidewords that would explain how the unsafe control action of interest is realized.

The Step 2 Tree brings structure to the Step 2 process by making the causal connections between the causal factors listed in Figure 14 explicit. Further, it opens the door to more efficient documentation of the scenarios as it provides an opportunity to avoid duplicating parts of the analysis effort when investigating several UCAs associated with the same CA (e.g. "turn beam on" leading to "turn beam on too early", "turn beam on too late", "not turn beam on when needed"). Indeed, the control algorithm at work in all these scenarios would be the same, the feedback channels will be identical, and so will the actuation devices. The same set of scenarios may therefore be responsible for several UCAs, and the analysis will gain in communicability by not including them more than once while making clear that they are associated with several UCAs.

4.2.2.2.3 Causal factors for unsafe control actions – additional guidance for using the Step 2 Tree

The following paragraphs provide additional guidance with respect to what the guidewords offered for the causal factors that are proposed on the process loop and in the Step 2 Tree mean.

P1: Inadequate command generation. What scenarios could result in a command being inadequately generated by the controller's logic?

A controller usually needs external information such as environmental/contextual information and the imposition of an objective/goal to be accomplished in order to perform its function correctly or, in other words, to achieve the said goal. P1/T1 ("inappropriate goal/input") groups hazardous scenarios that are caused by flaws related to these inputs and goals. The term input is to be understood as any input external to the detailed process loop, physical flows (e.g. energy) or data that is necessary for the controller to operate safely. In the case of automated controllers, this input would include external power for the controller itself (e.g. its circuit boards must be powered for their logic to be operational) or control commands coming from higher levels of the control structure. In the case of human controllers this usually refers to information that is communicated to the controller by other controllers.

With this in mind, an inadequate command generation is possible when such goals and input are:

P1/T1/F1: missing

P1/T1/F2: inadequately timed, i.e. is received too early or too late or

P1/T1/F3: wrong.

The next node, P1/T2, groups scenarios that result from flaws in the control algorithm. Even when all goals are adequately provided and all input to the controller is correct, a flaw in the decision algorithm of the controller could result in an inadequate command generation. This could be due to:

P1/T2/F1: missing rules

P1/T2/F2: wrong rules

P1/T2/F3: wrong clock, which includes timing consideration (e.g. is the controller making decision fast enough and at a rhythm that match the tempo of control needs) as well as the possibility that the rules it was provided with are not presented in the correct sequence or priority.

Finally, inadequate command generation can also be the result of process model flaws. The controller (whether it is an automated controller or a human being) has a model of the process in his "mind". This process model must be accurate (esp. in its understanding of the consequences to be expected from the actions that the controller may require to be taken on the state of the controlled process) and must be updated often enough via the analysis of information obtained from loop sensors and external sources of information regarding the state of the system (commonly described as being conveyed via "feedback channels") to serve as a valid representation of the current state of the process such that the controller, which uses this model as the basis for decision-making, can act according to what the actual situation requires. P1/T3 is the heading under which all these scenarios can be found. Reasons for lack of coherence between the controller's process and the real world can be either traced to design flaws, or to flawed updating of the representation that the controller has of the real situation (that is, flawed updating of process variables, including spontaneous change of information due to hardware issues such as bit errors in memory).

Design flaws associated with the lack of adequate feedback are the result of:

P1/T3/F1: existence of process variables whose values the controller should be aware of but is not. This also covers the possibility that the wrong sensor (one that senses another state variable than the one of interest) be chosen.

Flawed updating of process variables could be the result of:

P1/T3/F2: wrong interpretation of the feedback provided by the sensor and the external sources of information by the controller. For example, a controller is programmed to take action on temperature levels based on the Fahrenheit scale – but fails to convert information about the current temperature that is obtained in Celsius degrees.

P1/T3/F3: Inadequate timing of information (e.g. information comes too late or too early, information comes at too fast a pace, etc.)

P1/T3/F4: Inadequate information transmission, when information is lost, distorted, corrupted, changed, etc. while it is being transmitted to the entity that intends to use it.

P1/T3/F5: Inadequate sensor operation (e.g. offsets, stuck-at behavior, excessive noise, wrong input etc.)

The second branch of the Step 2 Tree is about scenarios that cause a safe command to not be followed.

P2: Safe command not followed. What causal factors could result in a safe (and correctly provided) command not being followed?

Each command issued by the controller must be transmitted to the actuator. A flawed transmission of this command could result in the command not being followed. These scenarios are labeled as P2/T1. Reasons that can explain their occurrence include:

P2/T1/F1: The command is not issued

P2/T1/F2: The command is transmitted but delayed

P2/T1/F3: The command is distorted (e.g. lost, changed, corrupted, etc.)

The second of the P2 nodes aims at finding causal factors related to flawed execution. First, the command could fail to be executed at all as a result of:

P2/T2/F1: Missing input to the actuators (e.g. the actuator is not powered) or controlled process

P2/T2/F2: The command was not received

P2/T2/F3: The actuator failed to react to the command (e.g. is blocked, broken, etc.)

The command could be executed, but too late, i.e. is

P2/T2/F4: Delayed

… or the command could be executed wrongly due to:

P2/T2/F6: Wrong input to the actuators or controlled processes (e.g. the actuator is powered but the power is too low)

P2/T2/F7: The actuator failing to behave as requested by the command it received despite having received the proper inputs.

Finally, P2/T3 refers to coordination problems with the actuator or controlled process receiving commands or actuation from several controllers or actuators at the same time, with possible conflicting content.

4.2.2.2.4 Causal factors for unsafe control actions – another categorizing scheme

The list of causal factor presented in both (Leveson, 2012) and the Step 2 Tree described in the previous section includes two kinds of causal factor: some that are related to a logic error due to either requirements/design error or inadequate flow and processing of information, and others that result from the failure of a system to perform the part of its function that is realized in the physical world. Colloquially speaking, these two generic categories could be referred to as logic errors on one hand and hardware errors on the other.

STPA controllers send commands to actuators through transmission links, and receive information from how well the action they intended to be performed succeeded in modifying the system's attribute that they wished to impact. Several logic and physical elements are involved in making this loop function. Going around the process loop, the following generic sequences happen:

Table 24 - Organizing STPA Step 2 causal factors long the control loop (1): from generation to actuation

Circled in red are considerations not explicitly present as guidewords to STPA Step 2 in (Leveson, 2012) and the Step 2 Tree

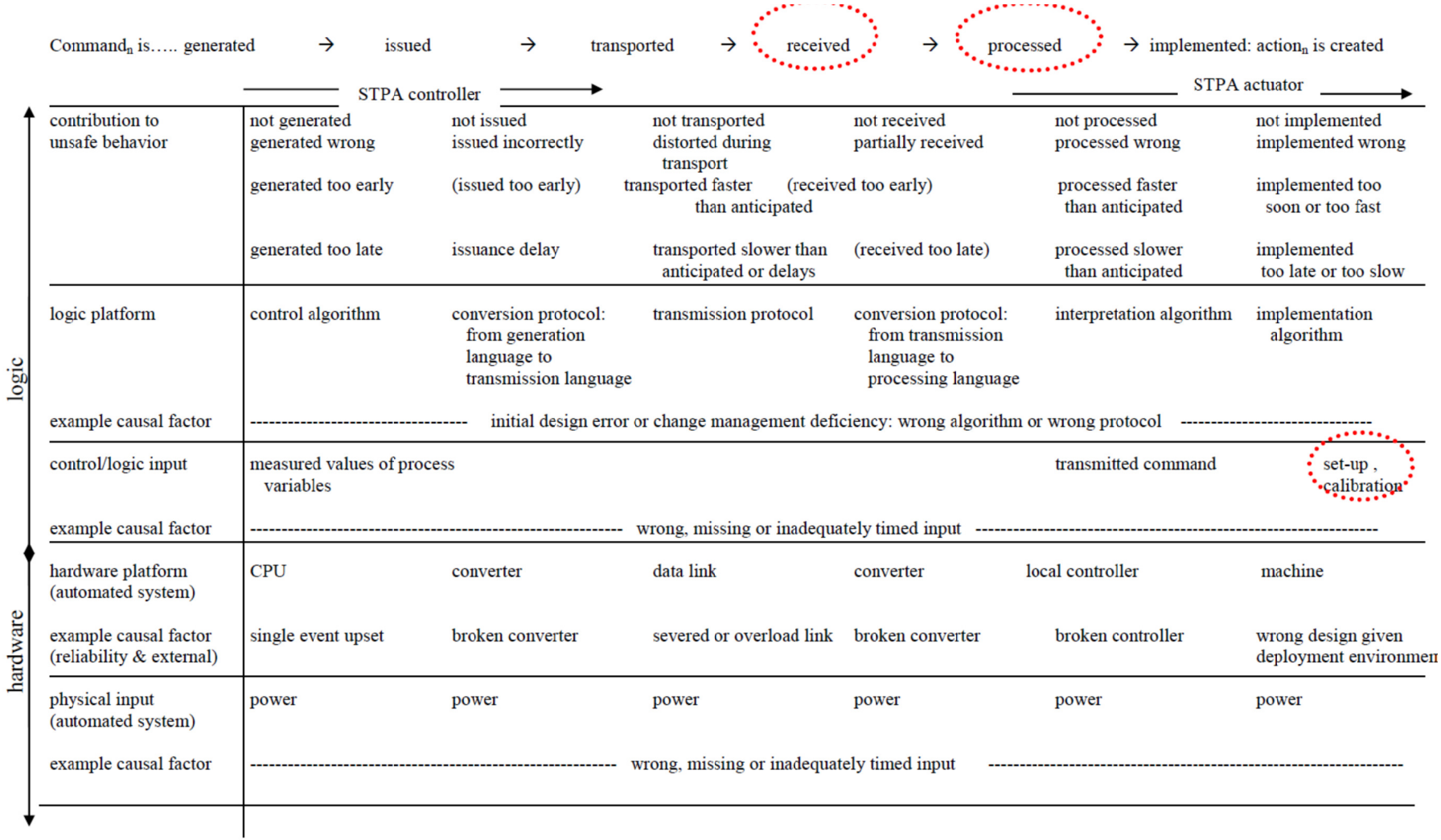
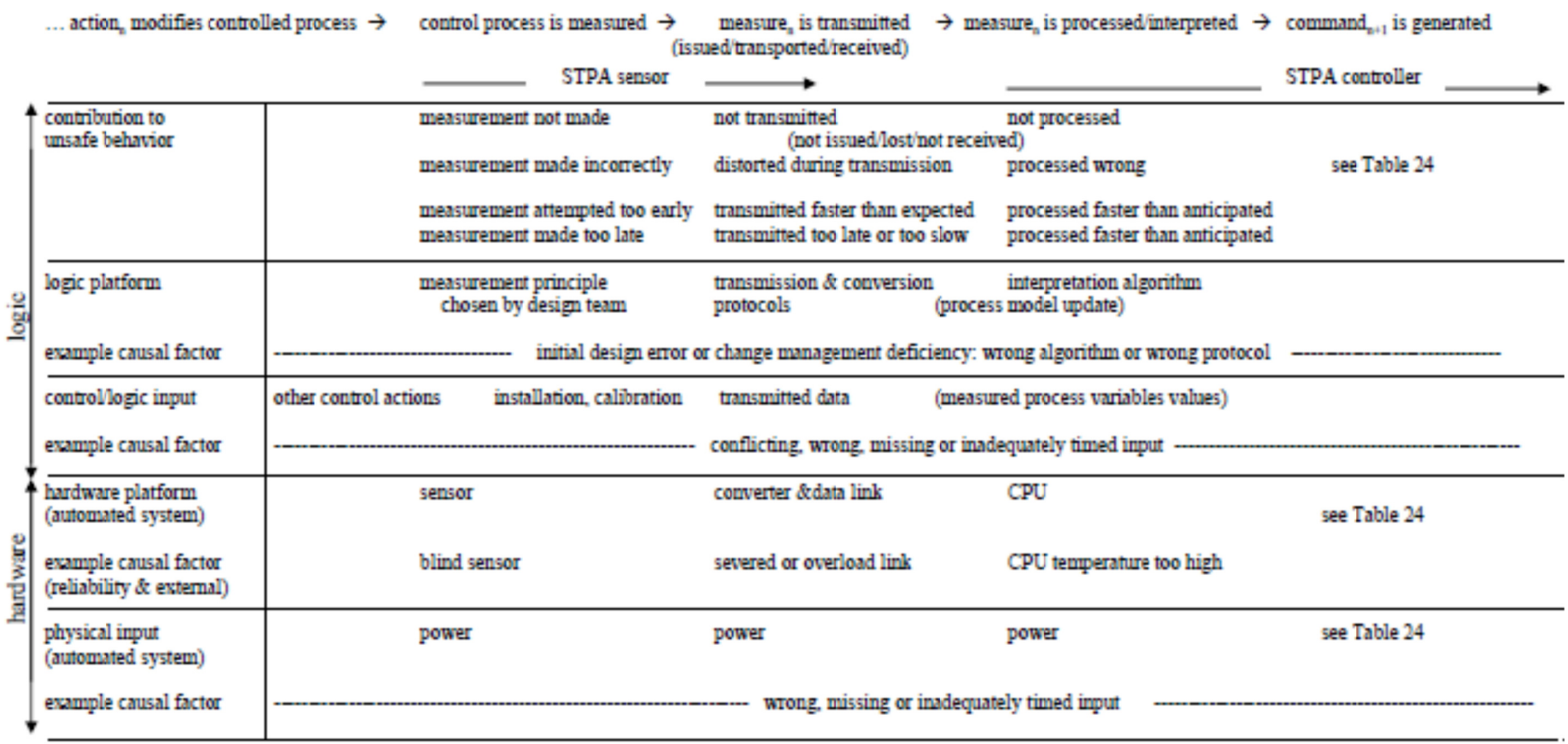


Table 25 - Organizing STPA Step 2 causal factors along the control loop (2): from actuation to updated process variables



command$_n$ is…..

generated → issued → transported → received → processed → before it is implemented,

This results in action$_n$ being taken. Then:

… action$_n$ modifies controlled process → control process is measured →

measure$_n$ is transmitted → measure$_n$ is processed/interpreted → commandn$_{+1}$ is generated

where transmitted refers to issued + transported + received .

Each of these action steps can be performed in such a way that it causes an unsafe control action to be taken. As Table 24 and Table 25 show, these causes can have their roots both in the logic or in the hardware platforms through which the command issuance and control action steps are performed.

These tables do not bring new causal factors to light. They however propose a systematic way of categorizing them that allows for simple recreation of the list of factors presented in (Leveson, 2012) and the Step 2 Tree above. As such, and given that human store information in categories (Collins & Quillian, 1969), they can be useful teaching tools for STPA Step 2.

4.2.2.3 Would it make sense to use STPA in combination with traditional hazard analysis techniques? Of the complementary use of STPA and fault trees.

Contrary to other hazard analysis techniques that are grounded in failure analysis, STPA does not focus on hardware failures. It is rather born from the understanding that not all failures are hazardous and not all hazards originate in component failure. STPA's distinctive contribution to hazard analysis has instead to do with focusing on functional behavior and evaluating the adequacy of logic and requirement design. To STPA, safety is a functional problem, not a failure problem. Safety can be endangered when a sub-system does not perform its function. This can arise as a consequence of component failure but not necessarily so: reliability and safety are understood to be two separate system characteristics.

This is not to say that hardware failure analysis is not valuable nor that STPA analyses should ignore them. Rather, hardware failure analysis should only be undertaken when high level functional analysis has helped identify what functional behavior is required of hardware and

what hardware failure mode is known to be hazardous. For example, after STPA Step 2 identifies that a sensor not providing the controller with information about the state of the controlled process is hazardous, one would need to make the system robust to this type of event, for example by designing the controller so that not receiving information from the sensor after a certain time has elapsed instructs the controller to put the system in a safe state rather than let it assume that the process variable that the sensor provides information on has kept the same value as that stored in the controller's memory. It would not be strictly necessary to investigate the failure modes of the sensor that make it go offline as long as the controller is designed to deal with this situation. However, one could still choose to design the sensor to be less likely to go offline, either by modifying its design or by preventing or mitigating the external disturbances that could affect its operation. Improving the sensor's reliability and/or robustness would reduce the diagnosis and start-up burden on the system, and increase its uptime. In the case of many systems, especially health-care related, such improvement does contribute to safety considered in a broader sense, as it allows for more patients to be treated.

FMEA (Failure Modes and Effects Analysis) and traditional hazard analysis techniques such as HAZOP (HAZard and OPerability analysis) and FTA (Fault Tree Analysis) offer frameworks to evaluate hardware reliability, robustness and susceptibility to external events that are now part of an established corpus of engineering techniques. They are well suited to the evaluation of hardware choices and could be used by the STPA analyst as he would a magnifying glass, after STPA has helped him identify the hardware issues that would benefit from further study, places where understanding of what caused a certain undesired behavior.

In this context and when this understanding is needed, it seems that, of these three methods, FTA would be the most valuable candidate for integration with STPA. Indeed, FMEA and HAZOP start from individual components. They do not start from hazardous situations, the hazardous consequence of a component having failed in a certain way (e.g., sensor not sending information leading to controller issuing an unsafe command). Plugging FMEA or HAZOP on the hardware problem identified by STPA would result in the analyst generating the list of all failure scenarios that could be associated with the component being studied, only a few of which would be relevant to the situation investigated. Of course, these results could be trimmed so that they match the STPA request, for example by only detailing the FMEA failure modes that can result

in the hazardous situation being studied. However, the waste born from the fact that these methods are not hazard but component centric makes them less attractive complements to STPA than Fault Tree Analyses.

FTAs are a better match to STPA because, contrary to the other two methods, they follow a top-down approach. Putting the events to be avoided at the top of the failure tree, FTAs are used to record how these events could occur by using a logic tree that links lower level events that contribute to the top-level failure using AND and OR gates. After STPA has identified that the sensor not sending information is an important problem, "sensor not sending information in x, y, z conditions" would be put at the top of a fault tree dedicated to the sensor, where x, y and z define the context in which this scenario has been evaluated to be hazardous (see Thomas process). This situation could be caused by the "sensor not generating a signal" OR "sensor signal not having been issued", both of which would be displayed as lower level nodes in the fault-tree. These lower level nodes would be further analyzed, resulting in a new level being added at the bottom of the fault tree being built: the "sensor not generating a signal" could be caused by "sensor being off" OR "sensor not able to sense controlled process" OR "sensor not able to process sensory input". Causes leading to these causal nodes would in turn be identified as lower level causal nodes: "sensor not able to sense controlled process" could be the result of "sensor disconnected from controlled process" OR "sensor overload" OR "sensor calibration wrong" etc. Once the "bottom" causes are identified specifically for the design being studied, design alternatives can be generated that eliminate some of the fault tree's leaves or reduce the likelihood of the chain of events that the fault trees document.

As such, it highlights the complementary of traditional hazard analysis, very well suited to the evaluation of hardware choices in the context of safety assessments, and STPA. Further, the top-down nature of STPA allows the analysis for both hardware and logic parts of each element to only evaluate hazardous scenarios and not the full array of possible, both hazardous and non-hazardous, scenarios.

4.3 Organizing and Displaying the Results of an STPA Analysis

Despite the analysis results being generated hierarchically, the output of an STPA analysis performed at a high degree of detail is a long and loosely ordered list of unsafe behaviors and scenarios through which they could be realized. This section addresses this problem. Its goal is to propose an organization structure that improves understanding of the system's behavior and facilitates design decisions related to safety issues.

4.3.1 Introduction

It is the beauty of a thorough hazard analysis to dig so deep into a system's structure and behavior that it can tell myriad stories about what unsafe paths this system could possibly take and what measures are or must be taken to cut these trajectories short. However, the amount of information hence created can be so difficult to handle that little use ends up being made of that data.

Organizing the results in a manner that renders their inclusion into the design space effortless is a necessary condition for hazard analysis results to be incorporated in design requirements and, therefore, safety to really become a design attribute. This is achievable thanks to the nature of the elements that are in need of ordering: hazardous scenarios, associated with specific system elements, causing a set of unsafe control actions, that well identified controllers can end up taking. All these dimensions can be thought of as categories according to which the hazardous scenarios revealed by the STPA analysis can be grouped: let us, for example, find all scenarios associated with this HF generator, with this specific command, with this desired function.

The value of defining dimensions along which similar scenarios can be grouped lies in making the wood more apparent to who might otherwise be blinded by the large number of its trees, and of doing so without sacrificing important details. Indeed, because categorizing will not eliminate but only reorganize data, the benefit of STPA generating an extensive list of unsafe system trajectories is maintained, and as little as possible of the system's behavior is left to chance. Further, bringing the hierarchy of these hazardous scenarios to light will facilitate the design of smart prevention, reduction and mitigation solutions that provide effective and efficient coverage against the scenarios that will have been uncovered.

4.3.2 Data Structure

4.3.2.1 Data structure and notations

Let $\mathcal{S}$ be the output of our STPA analysis. $\mathcal{S}$ is a set of several scenarios $\mathcal{S} = \{S_i , i = 1 \text{ to } s\}$. To identify what characterizes each S_i and how they could be organized into meaningful groups, it is useful to remember where they come from, how they were created.

First came system-level hazard definition: $\mathcal{H} = \{H_i, i = 1 \text{ to } h\}$. As proposed in the analysis of the PROSCAN facility presented in chapter 2, the higher level hazards can be understood to be the result of lower-level ones.

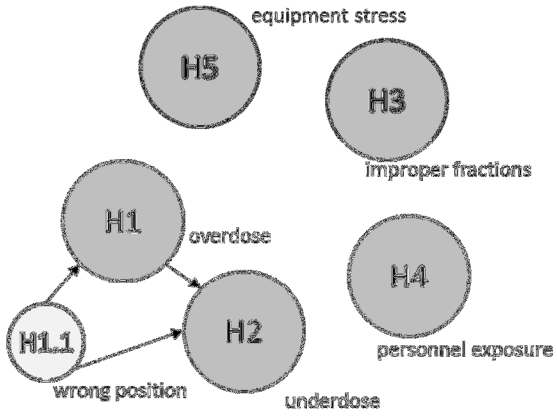


Figure 17 - System level hazards

H1.1 "wrong position" is a hazardous situation that can result in higher level hazards H1 "overdose" and H2 "underdose"

Control structures were then drawn at varying degrees of detail. These control structures form a set containing as many items as they are degrees of detail used by the analyst: $\mathcal{G} = \{G_g , g = 1 \text{ to } d\}$.

In each of these control structures, a set of functional elements was identified $\mathcal{E}_g = f_1(G_g) = \{E_{g,i} , i = 1 \text{ to } e_g\}$, a subset of which form the control structure's set of controllers $\mathcal{C}_g = f_1'(G_g) = \{C_{g,i} , i = 1 \text{ to } c_g\}$, with $\mathcal{C}_g \subseteq \mathcal{E}_g$.

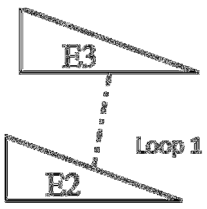


Figure 18 - Control loop elements.

These include actuators (e.g. magnets), sensors (e.g. irradiation chamber), transmission links (e.g. EPICS channel), controllers (e.g. Treatment Delivery System).

Each loop of the control structure includes at least one controller. To each controller $C_{g,i} \in \mathcal{C}_g$, a set of control actions $\mathcal{CA}_{g,i} = f_2(C_{g,i}) = \{CA_{g,i,j}\ ,\ j = 1 \text{ to } ca_{g,i}\}$ was attributed.

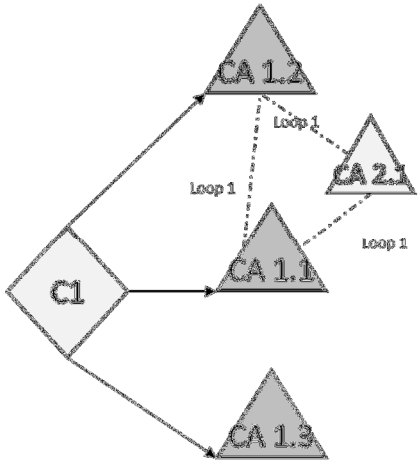


Figure 19 - Control actions

These are the control actions associated with loop 1, assuming loop 1 includes two controllers, C1 and C2

Each control action can be unsafe in two ways: unsafe because providing it, possibly too early or too late, leads to a hazard (p) or unsafe because not providing it leads to a hazard (n). The PROSCAN STPA project team singularized (p) and (n) UCAs that shared the characteristic of being provided with a wrong parameter. They are referred to as incorrectly provided (i) in the current discussion.

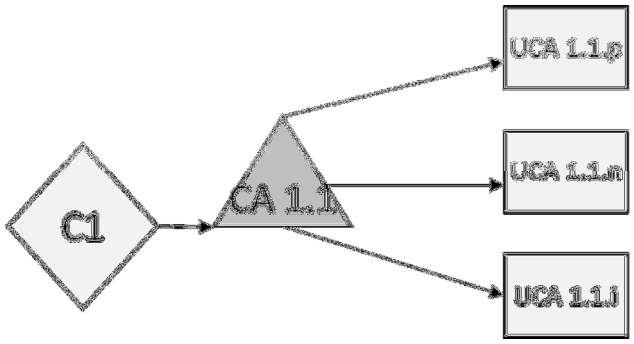


Figure 20 - Types of unsafe control actions

A control action can be unsafe in two ways: if it is provided and leads to a hazard (i.e. provided in the wrong context), or if not providing it leads to a hazard (i.e. not providing it in a certain set of contexts). In the PROSCAN study, UCAs that were born from a CA being provided with the wrong parameter (and thereby not being provided with the right one) were categorized independently as being incorrectly provided (in yet another set of contexts).

Let T be the unsafe control action type: $T \in \mathcal{T} = \{p, n, i\}$.

For each control action and with guidance from the high level hazards defined at the onset of the analysis, process variables $\mathcal{PV}_{g,i,j} = f_3(CA_{g,i,j}, \mathcal{H}) = \{PV_k\ ,\ k = 1 \text{ to } v \text{ where } v = f_4(CA_{g,i,j})\}$ were identified that were deemed relevant to evaluating the hazardous nature of specific system behavior in given system states. For example, because it is necessary to know whether personnel is present in the treatment room to evaluate whether turning the beam on might lead to personnel exposure to radiation (hazard H-R4), then process variable "personnel presence in treatment room" is deemed relevant to evaluating the safety of the "turn beam on" control action.

The combinations of the values taken by these process variables define the contexts in which the control actions can be provided. These contexts can therefore be described as vectors whose coordinates are the values taken by the process variables in the system state described by that context: $\mathbf{CT}_j = [PV1_j, PV2_j, ..., PVm_j]$. For example, the control action "turn beam on" can be provided when personnel are or are not in the treatment room (PV1 = "personnel presence", with possible values {yes, no}), patient is or is not at the right position (PV2 = "patient position", with possible values {none, yes, no}) and equipment is or is not ready to receive the beam (PV3 = "equipment readiness" with possible values {yes, no}). CT_1 = [no, yes, yes] and CT_2 = [yes, yes, yes] are two examples of possible contexts in which the control action "turn the beam on" could be provided. In the first case, personnel is out of the treatment room, patient is at the right position and equipment is ready to receive the beam. In the second case, patient and equipment are in the proper configurations, but personnel are still present in the treatment room.

It results from the above that an unsafe control action ($\mathbf{UCA}_{g,i,j,k}$) is defined as a three element tensor whose coordinates are

1. the control action that it stems from, $CA_{g,i,j}$,
2. the type of unsafe control action that it belongs to, $T_k \in \mathcal{T} = \{p, n, i\}$
3. the context in which it unfolds, $\mathbf{CT}_k$.

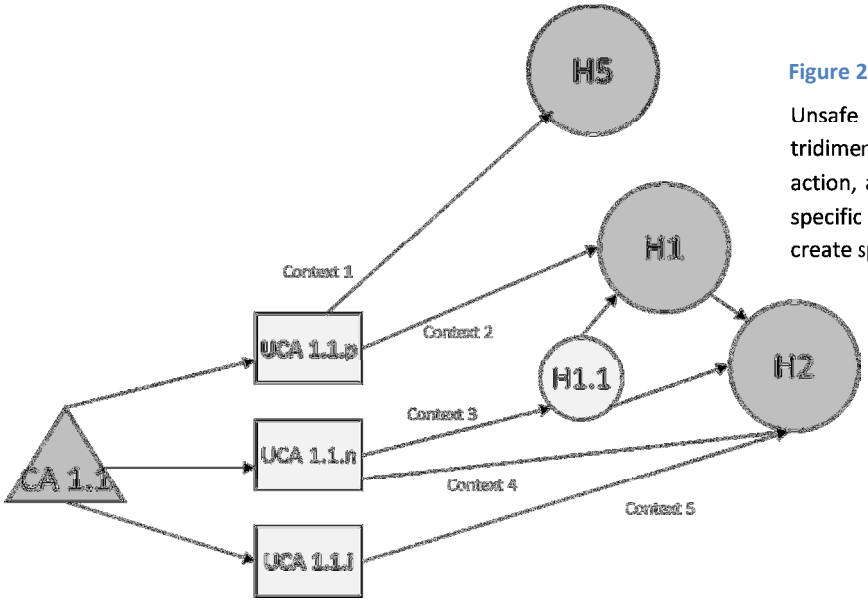


Figure 21 - From control actions to hazards

Unsafe control actions are defined as a tridimensional tensor referring to a control action, an unsafe control action type and a specific context. They are unsafe in that they create specific hazards.

$\mathcal{uca}_{g,i,j}$ is the set of unsafe control actions associated with control action $CA_{g,i,j}$. It is defined as: $\mathcal{uca}_{g,i,j} = \{\mathbf{UCA}_{g,i,j,k}$, k = 1 to $uca_{g,i,j}\}$ where $\mathbf{UCA}_{g,i,j,k} = [CA_{g,i,j}, T_k, \mathbf{CT}_k]$

Finally, detailed process loops were drawn that describe the command, actuation, feedback and information channels that allow the controller to learn about the system's state and take action to move the system towards a more desirable state. Then, exploring this loop using STPA Step 2, causal factors are identified for each of the unsafe control actions that are supported by that loop, leading to the creation of as many scenarios as causes can be found for each $\mathbf{UCA}_{g,i,j,k}$. The scenarios associated with UCA_{gijk} are $\boldsymbol{S}_{g,i,j,k} = \{S_i$, i =1 to s_{gijk}, S_i being a cause of $\mathbf{UCA}_{g,i,j,k}\}$.

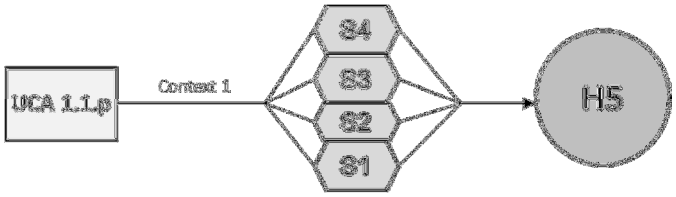


Figure 22 - Hazardous scenarios cause unsafe control actions

Scenarios describe how the unsafe control actions could happen

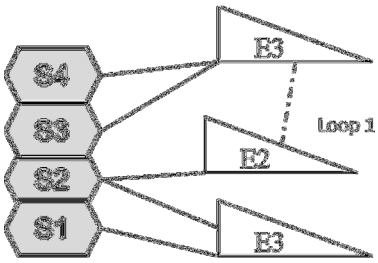


Figure 23 - Scenarios are associated with system elements

Each of these scenarios refers to one or more elements of the detailed process loop L_l, explaining how this element's behavior can lead, by itself or in conjunction with other factors, to the unsafe control action being taken in the context that defines UCA_{gijk}.

As summarized in Figure 24, it follows from the above that each scenario S_i is by nature associated with a specific controller $C_{g,i}$, one of its control actions $CA_{g,i,j}$, the process loop in which that control action is implemented L_l, an unsafe control action type T_k, a given context CT_k, one or more hazards $H' \subseteq H$ and a well-defined set of system elements $E'_g = \{E_{g,i}$, i = 1 to $e_g'\}$, with $E'_g \subseteq E_g$.

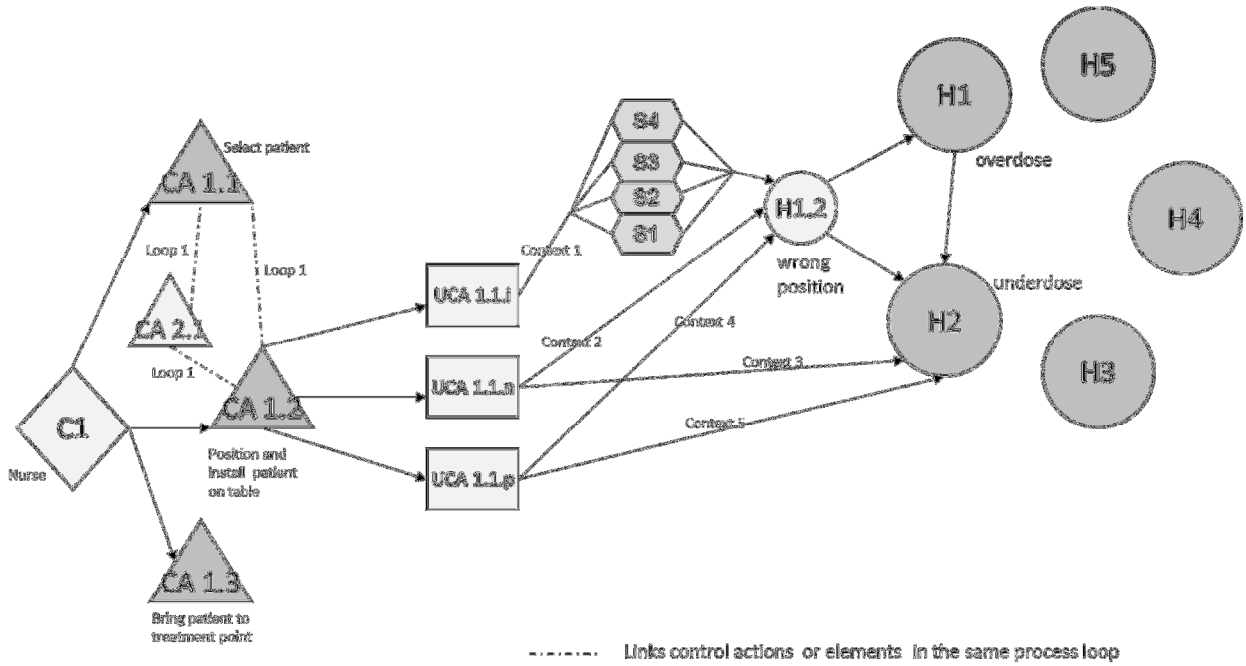


Figure 24 - Overall structure of the data generated when performing STPA

(C_i) are controllers, ($CA_{i.j}$) are control actions taken by controller C_i, ($UCA_{i.j.k}$) are unsafe control actions of type T_k that are derived from control action $CA_{i.j}$, (S_l) are scenarios through which $UCA_{i.j.k}$ can occur, (H_n) are system level hazards and (E_l) the elements associated with scenario S_l. The same paths as those that link CA1.1 to H5 via UCA 1.1.p and S1 to S4 could be drawn to link CA1.1 to H1 or H2, as well as CA1.2 and CA1.3 to the set of high level hazards.

4.3.2.2 Observations on the structure of data generated by STPA

4.3.2.2.1 Systemic factors

Systemic factors are causal factors that affect the system as a whole. Their existence makes it impossible to consider occurrences of inadequate behavior by different system elements as independent one from the other. For example, the likelihood of a back-up pump not coming online is higher if the reason why the pump it replaces failed to provide the expected flow rate was inadequate maintenance. If the inadequate maintenance schedule was caused by a reduction of maintenance resources due to reduced attention to safety issues under profit-making pressures possibly encouraged by a good safety record and the associated management and staff complacency, then it is likely that the back-up pump will receive even less maintenance scrutiny. A similar concern would be raised if the plant came offline because of material damage to the plant's structure such as resulting from an external shock, earthquakes and the like: unless it was specifically designed to withstand this kind of blow, the replacement system will also have

suffered from it. These systemic factors include external shocks, organizational factors (e.g. culture, resource allocation), and design flaws, especially with respect to sub-system interactions.

By emphasizing the role that context plays in turning given actions into hazardous ones, STPA brings appropriate focus to causal roots that may have effects throughout the plant via their role in modifying the context in which the control actions are taken. Let us consider for example a boiling water reactor nuclear plant that is faced with a large loss of coolant (LOCA) situation. Mitigation systems include a pump that re-circulates water inside the core and is powered by a turbine driven by reactor steam. Under LOCA conditions, the chances that the pump will not be able to provide the required flow are increased as re-circulating water is contaminated with debris. These debris can block the hydraulic lines and prevent the pump from providing flow because of beyond design hydraulic head. Since these debris were created by the LOCA conditions, the pump failing on demand is not independent from that of the LOCA happening. By prompting the analyst to consider the LOCA condition as context to controllers' actions, STPA forces verification of how that systemic factor will affect all control actions.

Further, the contextual factors help identify cases of hazardous interactions between elements not functionally or physically related one to the other. For example, if treatment area 1 is allowed to stir the beam while treatment area 2 is using it to treat a patient, then improper dose deposition will result in the patient in area 2. Such a situation is identified when the process variable "beam mastership" is used to define contexts in which control action "stir the beam" by controller of area 1 is unsafe and the analyst is prompted to imagine situations in which providing this action can be unsafe.

Finally, by maintaining a strong top-down focus (identifying UCAs by looking into means that may lead to a specific set of hazards) throughout an analysis dealing with increasingly detailed descriptions of the system ("zooming in" the control structure), the STPA framework encourages the analyst to creatively explore hazardous situations. The analyst is really asked "what can cause this hazard?" and, although guidance is provided in STPA Step 1 and Step 2, given as much free-dom as he would wish to propose answers to that question. For example, when UCA "thrust is too low" is identified for a plane and causal factor "engine inadequate operation" is listed,

external disturbance from neighboring equipment (e.g. a hydraulic fluid leak) should come to mind and highlight the need for adequate isolation of the fluid lines from the hot engine.

4.3.2.2.2 Many to Many and One to Many relationships

While multiple scenarios (S) can cause the same unsafe control action (UCA), a given scenario, especially if worded in such a way that it emphasizes a system elements' behavior rather than the context in which it operates, could also cause several different unsafe control actions. Further, in cases where components belong to several loops, each loop being controlled by a different controller, a single scenario could be linked to unsafe control actions taken by different controllers (e.g. a given sensor not correctly reporting on the actual state of the system may cause several controllers to each issue inappropriate commands).

Each scenario involves a possibly unique but never null set of structural system elements (E), while it is possible that each element is associated with several scenarios and several process loops (e.g. the same sensor may report to different controllers, each operating in a different process loop and responsible for different control actions).

Different unsafe control actions (UCA) can lead to the same hazard (H), while a single unsafe control action can lead to several hazards (e.g. selecting the wrong patient fixation devices can cause the patient to be ill-positioned, leading to both under and overdose hazards).

Each unsafe control action is defined as the combination of a control action (CA) with an unsafe control action type (p, n or i). Each control action is under the responsibility of a unique controller[55], while a given controller can be responsible for several control actions.

The fact that the relationships between the different data elements introduced in section 4.3.2.1 of this chapter are not bijective complicates attempts to record and communicate STPA results as they require the introduction of fastidious duplication of the data being recorded. For example, when an unsafe control action can cause two hazards, it must be recorded twice in a hazard centric representation of the results: once for each hazard. Paper and word documents are a poor recording medium, especially as they do not support dynamic reorganization of data to match the

[55] Cases where several controllers have the authority to issue the same control action are treated as several controllers issue different control actions.

reader's immediate interest. A computer database is much better suited to this task. Nonetheless, the many to many relationships identified above require that particular attention be given to the set-up of this data repository.

4.3.2.2.3 Use oriented exploration of STPA result database

4.3.2.2.3.1 Existence of several possible entry points

The dimensional complexity mentioned above is the easiest type of complexity to handle. As a consequence of the inherent structure of the data generated when performing STPA, it is indeed possible to query the list of scenarios according to any of the data categories presented in 4.3.2.1. This will allow the user to access the data from different points of views depending on the goals that she is pursuing. It is expected that this flexibility will facilitate access to the data and thereby promote its integration into the design roadmap.

Each of the data categories introduced in 4.3.2.1 offers a unique vantage point into the STPA results:

- Find all scenarios leading to a given hazard: if a hierarchy of hazards is established, this will allow the user to prioritize the prevention, reduction or mitigation of scenarios when conflicts happen or resources are insufficient to address them all.

- Find all scenarios associated with a specific control action , i.e. find $\boldsymbol{S}_{g,i,j}$, the set of all scenarios $\boldsymbol{S}_t$ associated with control action $CA_{g,i,j}$: useful when considering preventive measures that modify a control action or the associated process loop and when desiring to evaluate how well these measures prevent, reduce or mitigate the scenarios that would otherwise be possible causes of the unsafe control actions associated with this UCA.

- Find all scenarios $\boldsymbol{S}_{g,i,j,k}$ associated with a given control action $CA_{g,i,j}$ and a specific type of unsafe control action T_k: this selection helps make exploring means to detect that a given type of unsafe control action is being performed by the system, prerequisite to designing "healing" and "recovery" mechanisms, more efficient. For example, it is likely that a system that can detect that beam is turned on (control action: beam on; unsafe control action type: provided) when personnel is in the treatment room can also help detect that beam is turned on (control action: beam on; unsafe control action type:

provided) when patient position is wrong. This is the case as in both instances, there is a need to detect "beam on". It is expected that the more elegant solution that makes uses of the same subsystems to detect both situations (i.e. case when "beam is on" when it should not) offers a more efficient and possibly more robust[56] means to attain this detection capacity objective than a solution that provides separate "beam on" detection capacity for as many situations as there are expected contexts.

- Find all scenarios associated with a given context: since a large number of scenarios are the consequence of a context not being adequately evaluated by the controller (flaws in the feedback channel or the process model), measures that allow for more accurate and robust contextual information to the controller(s) will contribute to mitigate many scenarios related to the same contextual situations.

- Finally, finding all scenarios associated with a given element appears very valuable in a world where design is split across many teams whose expertise is focused on a few functional or physical elements. When retrieving these scenarios, they will be made aware of not only behavior that is directly caused by inadequate behavior of the element that they are tasked with working on, but also behavior that results from its interaction(s) with other system elements.

4.3.2.2.3.2 User scenarios

Hazard analysis, albeit sometimes performed as an after-thought on a design to be proven safe for its operation or marketing to be authorized, is at its best when accompanying the design process and helping steer design decisions towards eliminating hazardous situations and creating a safe system. After high level hazards and safety constraints are identified at the onset of the design phase, design decisions can first be checked against them as and when they are made, then analyzed to verify that their creation does not introduce new hazardous scenarios to the system's set of behaviors.

---

[56] Using less resources to achieve a given objective is, by essence, more efficient. The robustness case is less direct and has to do with the idea that adding more elements than is necessary to the design will create hardware clutter that will be harder to provide maintenance to, as well as signal clutter and unnecessary interactions that can cause hazardous situations on top of increasing the review burden.

It is however recognized that in large, complex systems, design responsibilities are split across several teams, each possibly comprised of many individual designers and experts. While some of these teams have integration responsibility and are therefore in charge of spotting and solving the interface and coordination issues that are often the cause of inadequate system behavior, few of the sub-system designers own a holistic vision of the system and of its behavior. One of the system integrators' challenges therefore consists in communicating to these myopic subsystem designers the constraints that their design must enforce so that its behavior adequately contribute to higher level goals without creating higher level safety issues. This is best done by integrating these constraints into the design agenda in the form of requirements to be attributed to different system functions, sub-systems and components. These requirements would be identified as contributing to the safety of the system and known to be prioritized accordingly when conflicts exist with requirements contributing to other system goals.

User scenario 1: assigning design constraints to a sub-system or individual element

The purpose here is to ensure that a sub-system or individual element's designer's choices will be adequately constrained by him being made aware of the implications of the features he chooses for his design on the overall system's behavior. Allowing him to visualize the interactions that the sub-system he is tasked with creating has with other sub-systems and presenting him with the scenarios that his design activities can contribute to eliminate (e.g. choosing a beamline kicker magnet that is closed when off and open when powered rather than one that is open when unpowered and closed when powered), reduce or prevent, are necessary conditions for him to make safety-guided decisions.

User scenario 2: evaluate whether a design change will create new hazards

A power plant utility engineer is responsible for integrating new technology in an aging plant, such as replacing analog electro-technical controllers with digital software controls. Along comes an equipment vendor that offers a technical solution meant to address the obsolescence and reliability issues faced by the utility engineer. The utility engineer is tasked with examining the hazards that would be introduced by replacing the old equipment with the new one.

In such a context, it appears that focusing on the loops and control actions that are affected by the design change would allow the operator to systematically evaluate the safety of the proposed

change, considering both the direct effects of this substitution (such as differences in hardware reliability and failure modes) and those that arise from interaction with distant sub-systems on the behavior of the whole system.

1. What process loops are affected by the proposed change? Those are loops that feature the new equipment (e.g. as a feedback source, controller or actuator) as well as loops that the new equipment provides input to. How are they changed (e.g. a process variable is introduced in the new version of the controller's process model that previously was not considered, creating the need for a new feedback channel)? Are new loops created?

2. What control actions are modified by the proposed change? Are new ones added?

3. What scenarios are affected by the proposed change? Are new ones added?

Identifying the commonalities and differences in behavior between the original and the newly proposed design by asking these questions will help evaluate whether the existing hazard prevention and mitigation measures appropriately address the hazardous scenarios associated with the design change, and whether new ones should be considered.

User scenario 3: evaluate whether an existing design contains adequate hazard prevention and mitigation measures

As emphasized by previous authors (Stringfellow et al, 2007; Leveson 2011) hazard analysis is at its best when guiding design. In cases where the documentation of design requirements or design choices motivated by safety considerations is lacking, STPA can provide a solid framework for performing an "after the fact" safety review of the close to final design, as demonstrated in our application of its premises to the review of the PROSCAN facility. Providing tools to facilitate the final verification that all safety concerns are adequately addressed by the design will be valuable in both situations.

An STPA-based safety review starts by associating each safety concern to the solution that the design proposes for it. This could be performed in several ways, depending on the documentation culture of the group in which the hazard assessment was performed:

1 - First, validation (as in "is what we want to build the right thing to build?": do we have the right plans?): map the safety constraints identified in STPA Step 1 to the requirements that were used to create the design, these requirements being the place where design choices are captured. Some of these requirements will be contributors to the potential hazard (e.g. creation of proton beam to treat cancer patient is what explains why there is a dose hazard in the first place) and others will have been intended to address these safety constraints (e.g. requirement that the beam must not be allowed on unless personnel are out of the treatment room). Safety constraints that have no image in the requirement space must be investigated as a hint that the proposed design may have safety holes, along with requirements that have no image in the safety constraint space as the hint that the STPA analysis may have overlooked system behaviors that could very well lead to undesirable states.

Then, verification (as in "is the thing we built what we wanted to build?": did we build it according to plans?): assess how well the design lives up to its requirements' expectations, including how robust the safety features meant to prevent, reduce and mitigate the system's hazardous behaviors are to the environment that they will be deployed within.

2- In cases where requirement documentation is less rigorous or has been lost over time, the safety concerns identified by performing STPA (esp. unsafe control actions and hazardous scenarios) on a design stripped of its safety dedicated features can be mapped against the technical and operational measures that contribute to alleviating them. These solutions can then be questioned as to whether they are expected to achieve their intended goal "well enough" through an evaluation of their expected behavior in the hazardous contexts identified by the STPA analysis.

Whichever process is followed, questioning the relevance of the features identified as being relevant to safety can be done by applying STPA to their behavior. Solutions to their expected deficiencies can then be proposed. These solutions can in turn be scrutinized for the possibility that they may introduce new causes of system hazards, opening the door to an endless iteration of analysis and a possibly paralyzing mise en abîme, one safety measure's potential inadequacies leading to the installment of the next safety feature, suffering in turn from new potential hazardous modes themselves to be addressed by a third... When should the addition of reduction

and mitigation measures stop? Ultimately, it is up to the designer to recognize that there are diminishing returns, increasing costs and increasing burdens due to the creation of a more complicated system by adding protection features to protection features. Hazard analysis techniques can provide information about what residual hazards are left with the system, but it is the designer's responsibility to assess whether these residual risks are acceptable given his design mandate.

4.3.3 Organizing STPA results for visual display

Although STPA is intended to be integrated into the various phases of the design process, with more elaboration and detail occurring as the design process proceeds, much of the current STPA literature has to do with helping generate initial high level requirements that will be fed to system designers. In contrast, the PROSCAN safety evaluation project was not dealing with a white sheet of paper but to a system very close to its final architecture. As such, it was able to look at the system in much more detail and take a finer grain approach to the hazard analysis than have thus far been published on STPA, down to the level of individual components,.

This finer grain approach came with a challenge: that of capturing in a meaningful way 1. the large amount of data that was generated in the form of myriads of hazardous scenarios and 2. the system's characteristics that were meant to address these hazardous scenarios. More generally, given the large amount of data that the final STPA round will include for a design close to completion, text editors and readers, although appropriate tools for capturing, reporting and safe-keeping these results, are limited in their ability to inform the data users of the multiple links that tie that information together without resorting to inefficient data duplication. Can this multitude be made sense of?

4.3.3.1 A hierarchical taxonomy of hazardous behaviors

As illustrated in Figure 24, STPA data is by the existence of several data categories (controllers, control actions, unsafe control actions, elements, scenarios, hazards) that can be thought of as many "tagging" dimensions. These tags lend themselves well to the establishment of a computer database query system that will allow users to promptly identify the unsafe behaviors and causal scenarios associated with their subsystem of interest.

The possibility of taking advantage of this causal hierarchy to display the information obtained by the hazard analysis and the resulting benefits for improved readability and communication is already recognized in intent specification (Leveson, 2000, 2012) where hyperlinks are used to make connections between different sections of the requirement documentation. A proposal is made here to enhance this referencing scheme by displaying related hazards and causal scenarios closer one to the other, and allow for visual representation of the protective measures in place to address them. Instead of taking safety constraints as input data points, this exercise will be concerned with classifying hazards, unsafe control actions, hazardous scenarios and protective measures.

Taxonomy, the academic discipline whose goal, according to the online Merriam-Webster dictionary, is the "*orderly classification of plants and animals according to their presumed natural relationships*" can be called upon for some useful tips. "*A classification should be truly stable in that it is not disturbed by the addition of further information; it should be robust in that alterations to a small number of items of information should not produce major changes in the classification; and it should be predictive in that a property known to exist in most members of a group will be expected to occur in those members which have not yet been examined for it*" (Williams; 1967). Taking advantage of the fact that STPA generates information in an orderly manner and following a simple causal hierarchy, a taxonomic approach is proposed for documenting the hazardous scenarios that STPA helps identify.

Hazards, unsafe control actions and causal scenarios follow the following causal relationships:

System Level Hazards ← Unsafe Control Actions ← Scenarios

Based on the observation made in the PROSCAN example described in Chapter 3 that several unsafe control actions (UCAs) can be grouped per the unsafe behavior that they engender (e.g. bringing the patient to a wrong position is a contributor to the patient receiving dose in the wrong place, which is in turn a contributor to the system level hazards of dose and overdose), it appears useful to refine the high level system hazards into contributing hazards for purpose of categorizing UCAs into semantically close sets.

System Level Hazards ← Refined Hazards (first level) ←Refined Hazards (second level) ← ... Refined Hazards (n-th level) ... ← Unsafe Control Actions ← Scenarios

The number of relevant refinement levels will vary from one study to the next, depending how diverse the system's behavior and how rich the paths leading to system level hazards are. These levels and the qualitative information they carry can be displayed in a causal tree, where lower levels branches are understood to causally contribute to higher levels ones.

In the following paragraphs, the PROSCAN study will be used as an example of how one can proceed with this refinement. Greater indentation with respect to the left hand side margin will be used to visualize belonging to a more detailed hazard category.

4.3.3.2 The PROSCAN example

The STPA study of the PROSCAN Gantry-2 user area started by identifying five "system level hazards" labeled H.1 to H.5.

H.1: patient overdose

H.2: patient underdose

H.3: improper treatment fractioning

H.4: personnel exposure to radiation

H.5: equipment damage

These system level hazards can be refined as having originated in one or several more detailed hazardous situations. For example, H.1 “Overdose” can result from either one, or both, of the more detailed hazards H.1.1 and H.1.2 being realized:

H1.1: the facility delivers the wrong dose, or

H1.2: the alignment between patient and beam is incorrect.

Meaning: H1 happens when H1.1 or H1.2 happen.

H1.1 and H1.2 can in turn be refined at a higher level of detail. H1.2, "incorrect alignment between patient and beam", is the result of one or more of the following hazardous situations:

H1.2.1: wrong patient is positioned at treatment point, or

H1.2.2: patient position is incorrect with respect to the table at treatment point, or

H1.2.3: table position is incorrect with respect to the room referential, or

H1.2.4: gantry nozzle position is incorrect with respect to the room referential, or

H1.2.5: beam position is incorrect within gantry nozzle.

Meaning: H1.2 happens when H1.2.1, H1.2.2, H1.2.3, H1.2.4 or H1.2.5 happen.

In this example, it makes sense to provide yet other levels of refinement. H1.2.1 and H1.2.2 can indeed, and for example, result from one or several of the following situations:

H1.2.1: wrong patient is positioned at treatment point because:

H1.2.1.1: wrong patient is positioned at the treatment point, or

H1.2.1.2: patient identity changed during transfer to treatment point

H1.2.2: patient position is incorrect with respect to the table at treatment point, because:

H1.2.2.1: patient position was incorrect in the preparation room, or

H1.2.2.2: patient position changed during transfer to the treatment point, because:

H1.2.2.2.1: patient waiting time is long during preparation, transfer or at treatment point, or

H1.2.2.2.2: uncomfortable environment (e.g. heat, distracting noises, fear), or

H1.2.2.2.3: patient is medically unwell, or

H1.2.2.2.4: motion disturbances during transfer

H1.2.2.3: patient position changed when at the treatment point (including after treatment has started).

Meaning: H1.2 happens when H1.2.1, H1.2.2 or H1.2.3 are realized.

H1.2.1, in turn, happens when H1.2.1.1 or H1.2.1.2 are realized.

H1.2.2 happens when H1.2.2.1, H1.2.2.2 or H1.2.2.3 are realized
And H1.2.2.2 happens when H1.2.2.2.1, H1.2.2.2.2, H1.2.2.2.3, or H1.2.2.2.4 are realized.

This level of refinement appears appropriate for the purposes of the PROSCAN study. The formulation of these more refined hazards indeed corresponds to semantically and causally relevant categories in which the unsafe control actions identified when performing STPA Step 1 can be organized. Unsafe control actions (UCAs) classified in the same category cause the hazard associated with this category. Consequently, they will cause the higher level hazards that are caused by this refined hazard.

Table 26 – Hierarchical classification of causes leading to unsafe behavior of the Gantry-2 system

H.1: patient overdose

H1.1: facility delivers the wrong dose,

H1.1.1: .....

H1.2: alignment between patient and beam is not correct.

H1.2.1: wrong patient is positioned at the treatment point (UCA4 in nurse example)

H.1.2.1.1 patient identity was wrong in the preparation room,

C1. Nurse

CA1.1. select patient

T3: incorrectly provided

UCA 1.1.3.1: Wrong patient or patient of unknownidentity installed on the table (UCA2 in nurse ex.)

S1: because patient was called in too early, left the room to go the rest room and the following patient was called in in between;

S2: foreign patient or patient who cannot hear or speak not able to confirm his identity;

S3: patient is mistaken with a homonym;

S4: wrong daily plan used because updates were not communicated to the treatment floor;

.....

H1.2.1.2 patient identity changed during transfer to treatment point

C1. Nurse

CA1.3. bring patient to gantry coupling position

T3: incorrectly provided

UCA 1.3.3.1: Wrong patient or patient of unknown dentity brought to treatment point (UCA4 in ex.)

S1: several patients had been installed on trolleys in a setting where the same preparation room feeds several treatment areas, and one went was sent to the wrong user area.

S2:...

C2:…

H1.2.2: patient position is incorrect with respect to the table at treatment point,

H1.2.2.1: patient position was incorrect in the preparation room,

C1. Nurse

CA1.2. position and immobilize patient on table

T1: provided

UCA 1.2.1.1: .... imminent to treatment (i.e. "too late"), leading to rush in treatment preparation and possibility of fixation devices not being correctly installed (UCA 1 in nurse example)

CF1 - inadequate process model

S1 - belief by nurse that prior treatment was still far from completion from lack of timely update by the operator room

S2 - ....

CF2 - ....

T3: incorrectly provided

UCA 1.2.3.1: Patient is not correctly positioned on the table (UCA9 in nurse example)

CF1:....

H1.2.2.2: patient position changed during transfer to the treatment point,

H1.2.2.2.1: patient waiting time is long during preparation, transfer or at treatment point

C1. Nurse

CA1.2. position and immobilize patient on table

T1: provided

UCA 1.2.1.3: .... too early, leading to long wait time and patient unrest on the table and possibility of the fixation devices becoming untied (UCA 3 in nurse example)

H1.2.2.2.2: uncomfortable environment (e.g. heat, distracting noises, fear)

H1.2.2.2.3: patient is medically unwell

H1.2.2.2.4: motion disturbances during transfer

H1.2.2.3: patient position changed when at the treatment point (including after treatment has started).

H1.2.3: table position is incorrect with respect to the room referential,

H1.2.4: gantry nozzle position is incorrect with respect \to the room referential,

H1.2.5: beam position is incorrect within gantry nozzle

H.2: patient underdose

…

H3: .....

As explained previously, UCAs are defined as a triptych including a control action associated to a specific controller, an unsafe control action type, and a context: $\mathbf{UCA}_{g,i,j,k}= [CA_{g,i,j}, T_k, \mathbf{CT}_k]$. The three dimensions of the UCA tensors can be used to further organize the control actions within each of the semantic categories defined above. Although they do not carry causal significance, they uniquely define each control action. An illustration is provided with "**H1.2.1.1: patient position was incorrect in the preparation room**".

H.1.2.1.1 " patient position was incorrect in the preparation room " can be caused by the controller **C1. Nurse** taking the following control actions in a way that leads to a hazard[57]

CA1.1. select patient or

CA1.2. position and immobilize patient on table

More specifically, control actions **CA1.1 select patient** and **CA1.2 position and immobilize patient on table** can be unsafe when being provided. This corresponds to UCA type **T1: providing.**

Actions of type T1 that are associated with CA 1.2 were identified in (PSI, 2012) as hazardous when they were taken in the following contexts:

CT1: .... too early, leading to patient unrest on the table and possibility of the fixation devices becoming untied (UCA3 in the nurse example)

CT2: .... imminent to treatment (i.e. "too late"), leading to rush in treatment preparation and possibility of fixation devices not being correctly installed (UCA1 in nurse example)

This association of CA1.1, T1, CT1 and CT2 define unsafe control actions UCA 1.1.1.1 and UCA 1.1.1.2:

UCA 1.1.1.1: Nurse selects patient too early, possibly leading to patient unrest on the table and resulting in the fixation devices becoming untied (UCA3 in the nurse example) or

UCA 1.1.1.2: Nurse selects patient imminent to treatment start time (i.e. "too late"), possibly causing treatment preparation to be rushed and fixation devices to not be correctly installed (UCA1 in nurse example)

Performing the same association with the other unsafe control action types yields:

[57] these control actions are referenced respectively as CAI.1 and CAI.2 in example 1 "the Nurse" of (PSI, 2012)

T2: not providing it -- not deemed hazardous

T3: incorrectly providing it

UCA 1.1.3.1: Wrong patient or patient of unknown identity is installed on the table (UCA2 in nurse example) or

UCA 1.1.3.2: Patient is not correctly positioned on the table (UCA9 in nurse example)

Once the unsafe control actions are thus organized per their type, the control action they stem from, their controller and the refined hazards that they cause, a similar process can be used to nest the scenarios describing how control flaws cause unsafe control actions under them. Table 26 and Figure 27 present the results of applying this taxonomic endeavor to the positioning hazard in the PROSCAN example.

4.3.3.3 Evaluating the breadth of the protection offered by hazard protection measures

One of the biggest values derived from logically organizing the results of a hazard analysis is the possibility this opens to visualize and therefore promptly grasp the extent of the protection that given elimination or mitigation strategies can offer against undesired system behavior.

Once the hazardous behaviors have been identified, they can be dealt with through different means. Ideally, they would be eliminated: perhaps a different technology or architecture choice would not be prone to the same behavior while equally meeting the performance objectives of the design. For example, designing an alarm such that it will sound when a detection sensor's circuit intensity becomes null eliminates scenarios where sensor losing power cause personnel presence to not be detected when the beam is to be turned on. Similarly, ensuring that the safety constraints identified by a STPA analysis are indeed translated into actual software code would eliminate the possibility that software creates the hazards that the analysis will have identified and studied.

When hazardous behaviors cannot be eliminated, attempts can be made to reduce their likelihood and mitigate their consequences. Reduction and mitigation endeavors often require that a diagnosis be made of the system's state. It is therefore useful to add detection measures to the arsenal of technical and operational prevention measures that can be proposed to address the safety weaknesses that have been identified in a design but haven't been found to be possible to eliminate.

Verifying patient position prior to treatment by performing a CT-scan comparison of actual position with expected position at the point of treatment, OR performing it in the patient waiting area before patient is brought to treatment points cover different parts of the unsafe control action map that was presented in Table 26. The extent at which reduction, mitigation and detection measures address the safety issues that the STPA review will have identified is related to how high they can target the hierarchy presented in the previous section. The more powerful ones target higher levels of the hazard taxonomic tree (Table 26).

Hazard prevention measures can be thought of as attempts to eliminate and, when elimination is out of reach, reduce and mitigate the hazards and the unsafe control actions that cause them. They are all the more valuable that

1. they eliminate rather than simply reduce or mitigate hazards,
2. they target hazards of greater severity and
3. they target higher levels of hazards, cutting the taxonomic tree through a large branch rather than refined bough after refined bough.

Figure 27 illustrates a proposal to visualize the "coverage" brought by different hazard prevention measures (labeled M1 to M8) to the hazardous landscape identified in the PROSCAN study:

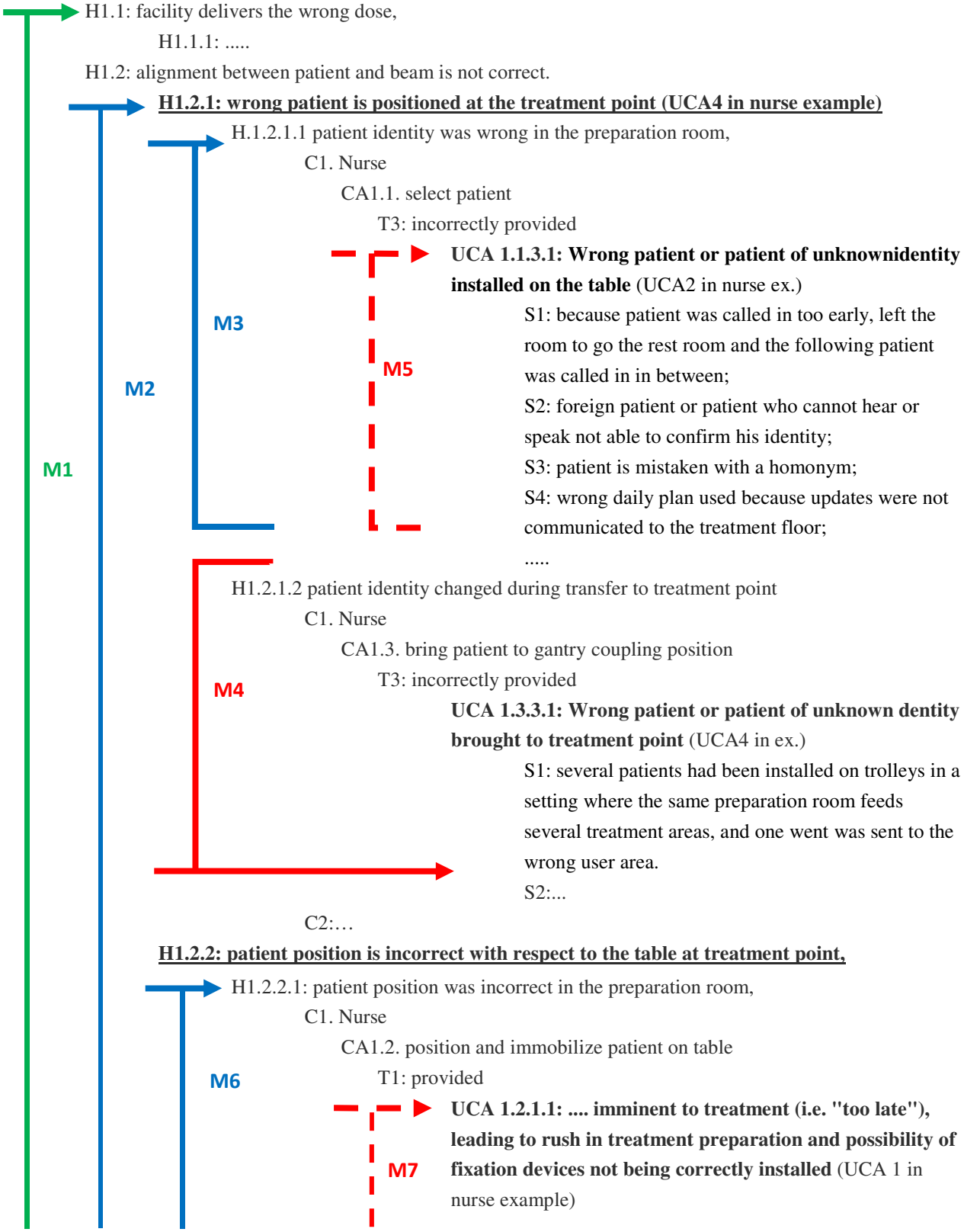
- M1: total treatment dose delivered is fractionated so that only a portion of it is delivered every day.
- M2: has two components. 1. Nurse verifies gross alignment of couch with table at treatment point for all irradiation sequences. 2. X-ray or CT-scan verification is sometimes performed in the treatment room on patient coupled to the gantry.
- M3: Patient scheduling software checks that patient ID entered when defining CT-offsets matches the loaded steering file. Checks "patient anatomy" as seen by medical operator on CT image against selected steering file
- M4: use only one treatment room, associated with a single preparation room. Only two trolleys are present on the treatment floor, and they never are in the same room when carrying a patient. One is used for patient treatment, the other for the next patient's preparation
- M5: use of patient specific molded couch and fixation devices, all labeled with patient ID
- M6: CT-images are made to verify that the patient is correctly positioned before being transferred to the treatment room

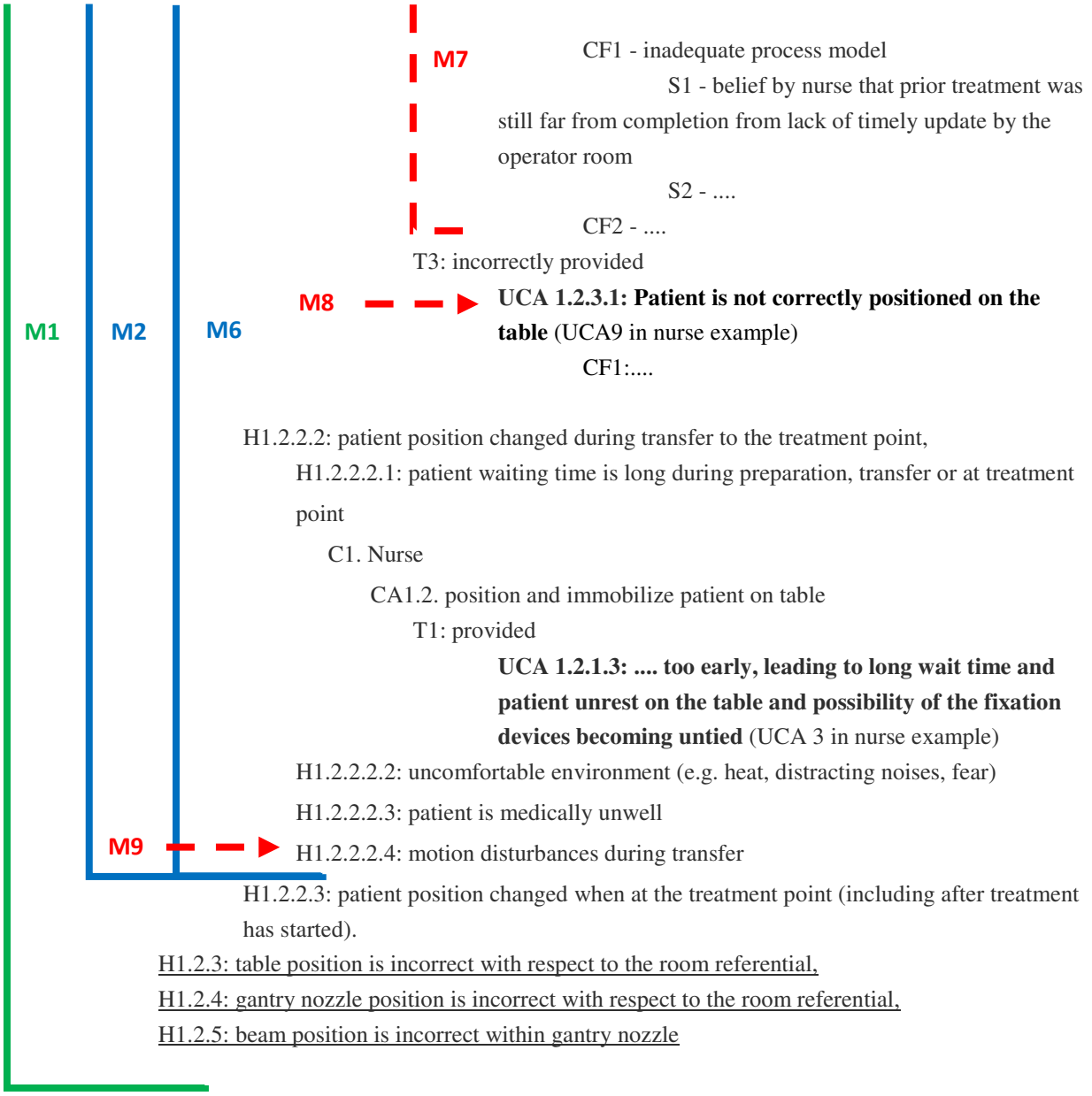
- M7: close coordination efforts by treatment delivery teams[58].
- M8: use of fixation devices that match the patient's anatomy reduces potential for nurse to be off when setting them in place.
- M9: measures to minimize motion disturbances during transfer from the CT-room to the treatment point such as creating a dedicated transfer path on an even plane with optically guided trolley driven at very low speed, mechanical system that prevents couch from sliding out of the table rails when clamps that hold it to the table are released (couch needs to be lifted out to be set free) and elimination in Gantry 2 of the Gantry 1 feature requiring that floor be lowered before table is installed at treatment point.

Each prevention measure targets one or more levels of the hazard taxonomic tree. When targeting a higher level (i.e. a less indented one), broader hazard, it covers all the refined hazards that are causally related to that higher level hazard. The targeted level is pointed at with an arrow; the lower level hazards covered by that arrow are included in the horizontal bracket. A given measure could target different levels of the classification. For example, including a sensor to detect whether beam is on could have an effect on unsafe control actions that depend on different controllers and would thus be grouped under different branches of the taxonomic tree.

[58] Is more coordination needed? The question needs to be asked, especially if the facility expands to include several treatment rooms, making beam possibly less available for each treatment rooms, or areas start sharing spaces such as the patient preparation rooms.

Table 27 – Hierarchical classification of causes leading to unsafe behavior of the Gantry-2 system





CF1 - inadequate process model

S1 - belief by nurse that prior treatment was still far from completion from lack of timely update by the operator room

S2 - ....

CF2 - ....

T3: incorrectly provided

UCA 1.2.3.1: Patient is not correctly positioned on the table (UCA9 in nurse example)

CF1:....

H1.2.2.2: patient position changed during transfer to the treatment point,

H1.2.2.2.1: patient waiting time is long during preparation, transfer or at treatment point

C1. Nurse

CA1.2. position and immobilize patient on table

T1: provided

UCA 1.2.1.3: .... too early, leading to long wait time and patient unrest on the table and possibility of the fixation devices becoming untied (UCA 3 in nurse example)

H1.2.2.2.2: uncomfortable environment (e.g. heat, distracting noises, fear)

H1.2.2.2.3: patient is medically unwell

H1.2.2.2.4: motion disturbances during transfer

H1.2.2.3: patient position changed when at the treatment point (including after treatment has started).

H1.2.3: table position is incorrect with respect to the room referential,

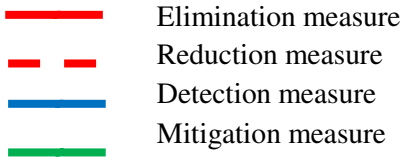
H1.2.4: gantry nozzle position is incorrect with respect to the room referential,

H1.2.5: beam position is incorrect within gantry nozzle

H.2: patient underdose

….

H3: .....



Another way to visualize the extent of the coverage offered by the prevention measures consists in remembering the data structure presented in the earlier pages of this chapter and summarized in Figure 24. The mitigation measures can indeed be represented on the links that tie data elements together, as is done in Figure 25 below.

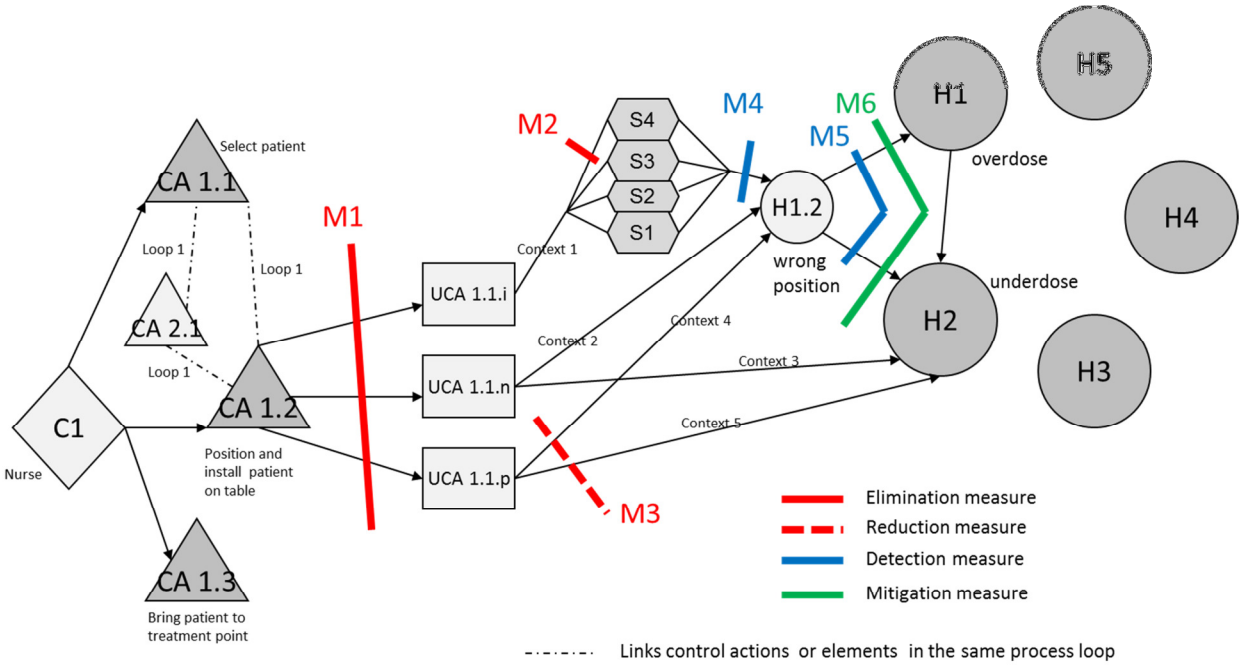


Figure 25 - Visualization of system protection measures (lines)

In this figure, used for illustration purposes only, it was assumed that prevention and mitigation measures existed such that:

- M1 eliminates all unsafe control actions from CA 1.2 "Position and install patient on table"
- M2 eliminates scenario S4 that could cause UCA1.1.3.2 (of type "incorrectly provided")
- M3 could reduce the likelihood of all UCA 1.1.1.x (of the type "provided when should not have been")
- M4 detects that UCA 1.1.3.2 ("Patient is not correctly positioned on the table") has happened (e.g. by performing a CT-scan position verification in the treatment preparation area), a first step towards being able to correct the inadequate position
- M5 detects that hazard H.1.2.2 ("Patient position is incorrect at treatment point") is realized (e.g. by performing a CT-scan position verification at the treatment point), a first step towards being able to correct the inadequate position
- M6 mitigates the consequences of hazard H.1.2.2 (e.g. by fractioning the dose so that it be delivered over a large number of sequences rather than in one large amount).

Figure 25 helps visualize the effect that the prevention and mitigation measures hope to have in preventing the system level hazards. It shows that:

- M5 is more valuable than M4 as a detection means. If one had to choose between the two, because resources are not sufficient to implement both or because they are negative side-effects to implementing both (e.g. multiplying radiologic diagnostics augments radiation exposure to patients and personnel), then it would be better to implement M5 (verification of the patient position at the treatment point) than M4 (verification of the patient position before transfer to the treatment point) as M5 will "catch" more hazardous situations than can the more upstream M4.
- M2 eliminates S4 but fails to address other scenarios that carry potential for the same hazards as are associated with S4 to be realized. It is therefore to be considered a weak protective measure.
- M3 ought to be complemented by mitigation measures.
- M1 is the most valuable safety measure proposed and should therefore be given a lot of attention. Were it to be absent, safety gaps could be identified and protective measures must be found for S1, S2, S3, and UCA 1.1.n, as well as mitigation measures to reduce the severity of the events associated with H2 being realized.

This first visualization attempt suffers from its dependency on how the drawing is laid out. Figure 26 tackles this issue by representing hazard prevention measures not as lines, but as dots. Full red dots (“elimination measure”) can be thought of as severing all causes downstream and therefore eliminating the hazardous pathway. Red circles are prevention measures that reduce the likelihood of that causal pathway unfolding; they include such measures as the design of protection systems dedicated to moving the system to a safe state after a hazardous condition has been detected. They often require the presence of detection measures and include interlocks, protection systems and “safety functions” in the sense of IEC 61508. The green dots refer to mitigation measures that make the consequence of the hazardous scenario unfolding less dire than without the mitigation measure.

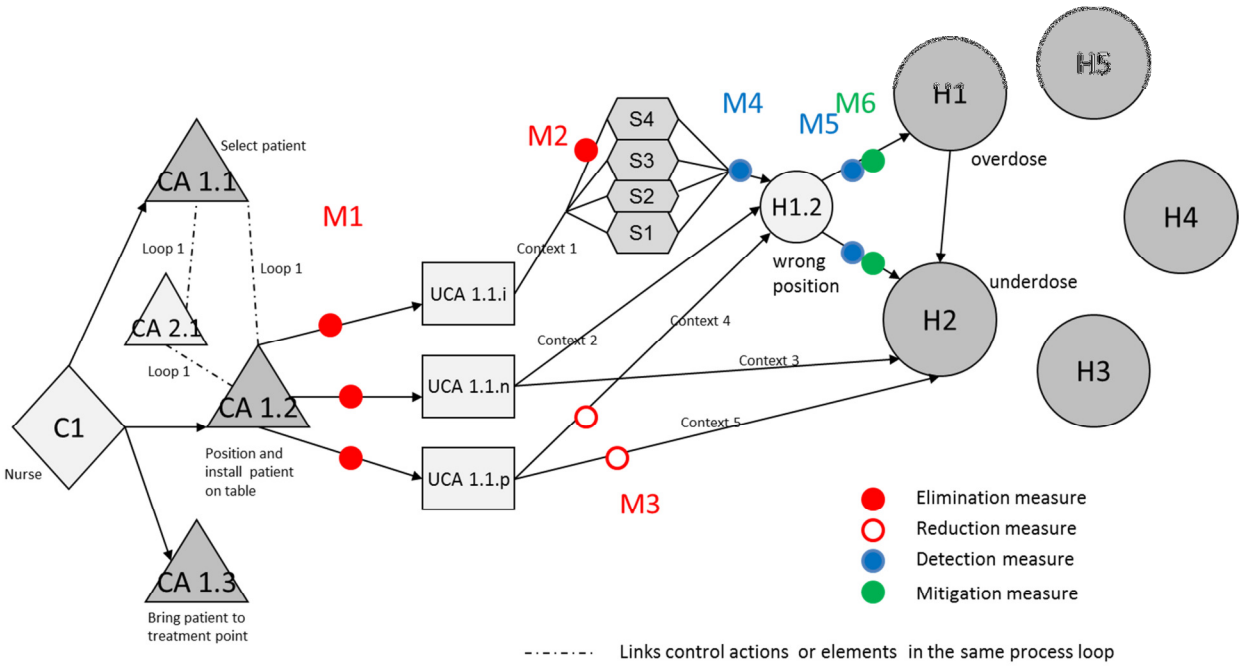


Figure 26 - Visualization of system protection measures (dots)

The purpose of visualization is to foster communication by making sets of data easier to grasp by people and to generate new understandings through eloquent display of information. By demonstrating how they can map hazardous behaviors in a consistent hierarchical manner, display proposed protection measures and identify possibly existing safety gaps in a design, the tabular and graphical organization of hazardous behaviors proposed in this section offer a solid foundation to build tools that will support the work of system designers and system reviewers towards designing systems less prone to hazardous behavior[59].

[59] Just like there is more than one correct way of classifying species and coming up with the corresponding taxonomies, we suspect that there is more than one correct scheme to organize that information. Users of the STPA methodology and of its results should feel free to experiment with several proposals and retain that which is most useful to them in that it allows for the most effective communication between requirements and design teams, and between design teams and safety reviewers.

4.3.4 Software assisted scenario browsing: a basis for further discussion

The STAMP framework is built around two types of hierarchies:

- Hierarchies of hazards (see taxonomic discussion above as well as the hierarchy in hazard severity that allows one to prioritize possibly conflicting constraints),
- Hierarchies of controllers (see the concept of control structure).

The visualization proposals that have been discussed thus far (see Figure 24, Figure 25 and Figure 26) are based on hazard hierarchies. However, many users would find it beneficial to visualize the analysis' results directly on the control structures. Further, if a software tool were available to store STPA data and display its causal information, an extension could be developed so that it can be used as support for performing the hazard analysis: users would not only be allowed to display STPA results, but could also, upstream, input information and perform the analysis with the help of that software tool. Such an extension would be especially helpful in collaborative settings, where the visual display of the control structure has been experienced to be a strong stimulator of expert discussion (c.f. the STPA workshop held at PSI during the PROSCAN STPA project). It could also allow data to be entered in such a way that it would lend itself to semi-automatic generation of STPA results through such procedures as are being developed by CSRL researcher John Thomas (Thomas, 2011, 2012). It could finally be made to support trouble-shooting sessions: users could enter issues as they arise into the supporting database via the data capturing interface; the software could suggest investigation questions that the STPA methodology could help generate – e.g. based on the list of Step 2 control flaws.

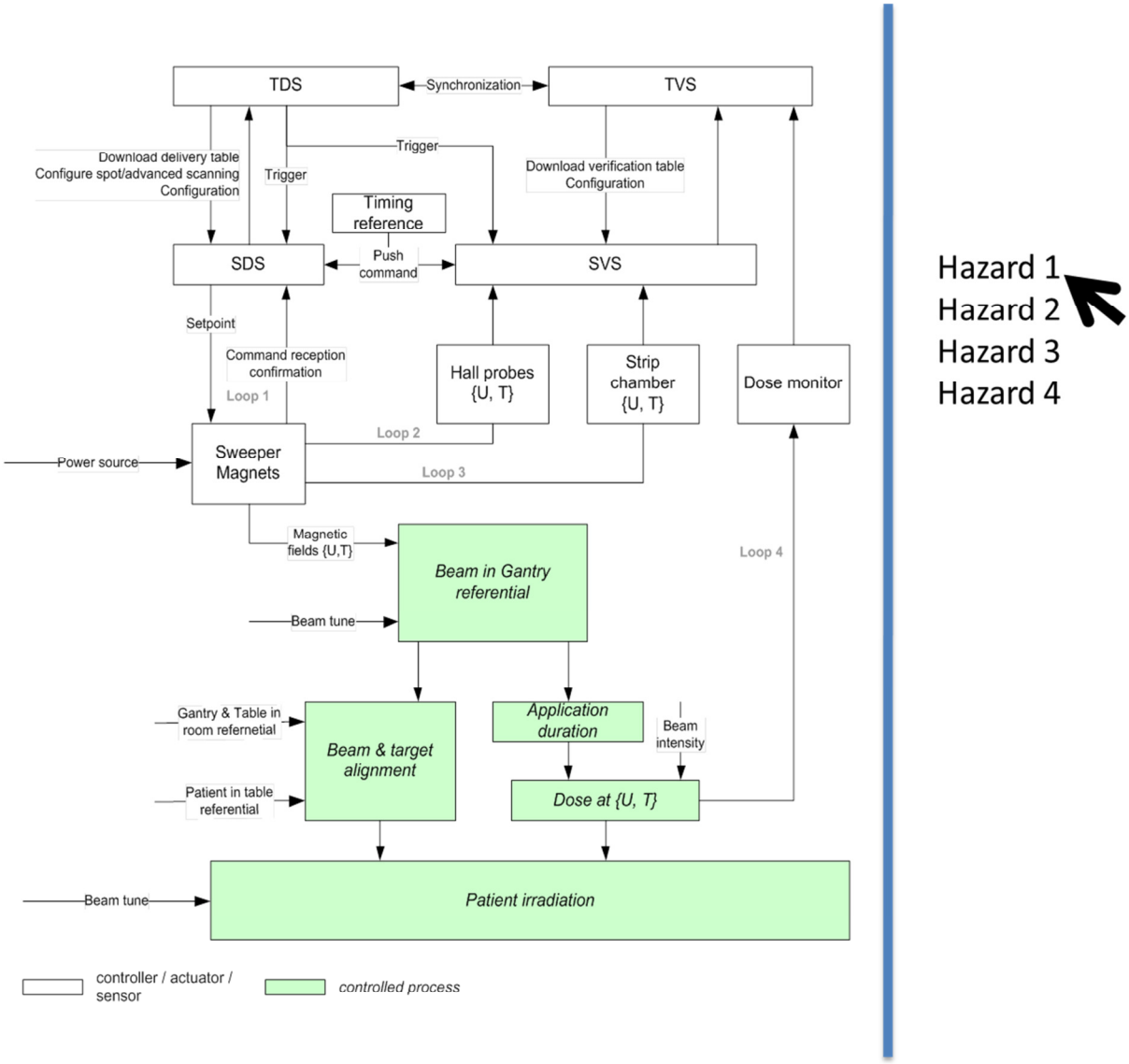
Figure 27 illustrates how such user navigation in the forest of hazards could be guided by an interactive software tool that would allow for the user cases discussed in the first pages of this chapter to be performed with no more than a few clicks. The structure of the backstage database that would be sampled by the user through such an interface could be built to mirror the graph representation of the data that was proposed in Figure 24[60].

[60] Note: as of July 2012, a prototype software implementing some of these ideas was under development in the complex systems research lab at MIT.

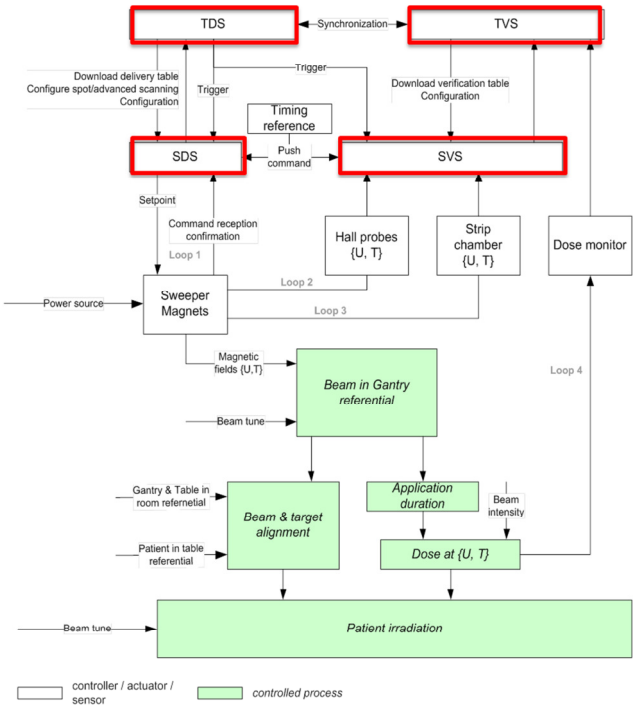
Figure 27 - Proposal for screenshots from an STPA display tool

The black arrow represents the user's mouse's shadow on the screen.

1. User selects the hazard he wants to focus on.

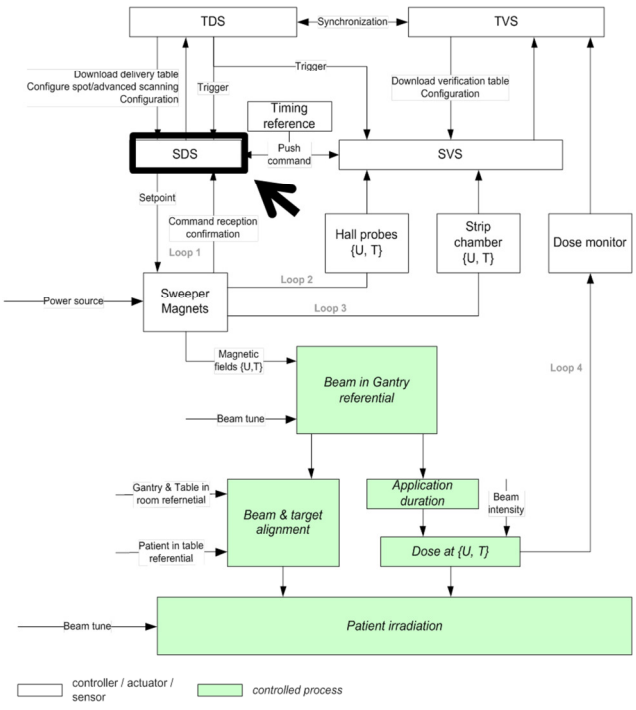


2. Software displays the controllers whose actions can cause the hazard that user selected.



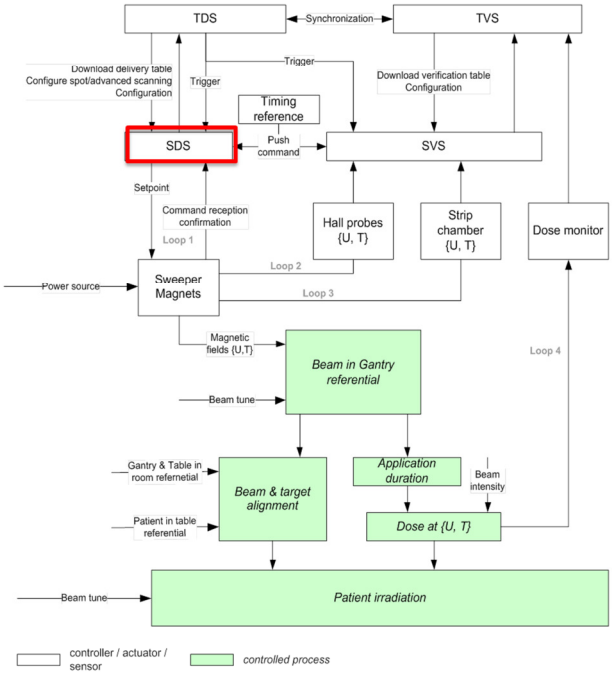
Hazard 1
Hazard 2
Hazard 3
Hazard 4

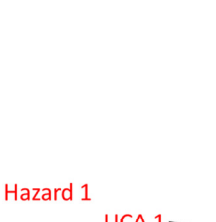
3. User selects the hazards he is interested in for that controller



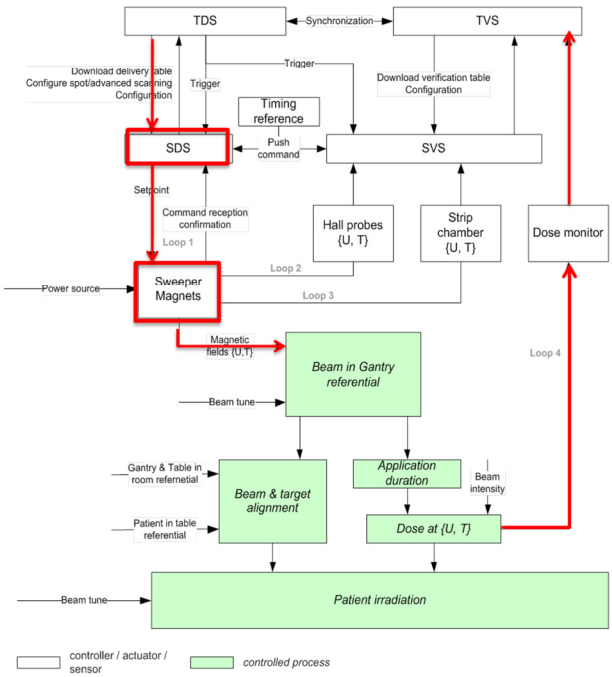
Hazard 1
Hazard 2
Hazard 3
Hazard 4

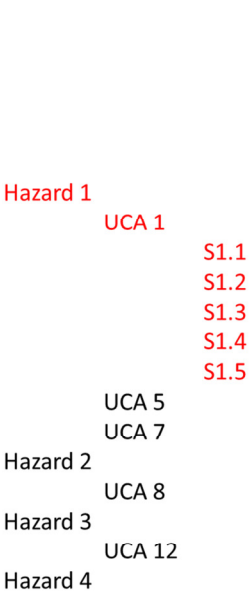
4. Software lists the unsafe control actions that the controller of interest could take that would lead to the hazards.





5. User selects a UCA whose causes she wants to study. Software lists the scenarios that could cause it and highlights the scenario path on the control structure





4.4 References

Fleming C., Ishimatsu T., Miyamoto Y., Nakao H., Katahira M., Hoshino N., Thomas J., Leveson N., *Safety Guided Spacecraft Design Using Model-Based Specifications*, in Proceedings of the 5th Conference of the International Association for the Advancement of Space Safety (IAASS), 2011

Leveson N., *TCAS qualitative fault trees*, 1981, as presented in http://www.safeware-eng.com/software%20safety%20products/Specifications.htm accessed August 10th, 2012

Leveson N., *Intent Specifications: An Approach to Building Human-Centered Specifications*, IEEE transactions on software engineering, Vol. 26, No. 1, January 2000

Leveson N., *A New Accident Model for Engineering Safer Systems*, ESD-WP-2003-01.19, MIT Engineering Systems Division Internal Symposium, May 29-30 2002, accessed at http://esd.mit.edu/WPS/internal-symposium/esd-wp-2003-01.19.pdf

Leveson N., Engineering a Safer World: Systems Thinking Applied to Safety, MIT Press, 2012a, ISBN 978-0-262-01662-9

Leveson N., *The Use of Safety Cases in Certification and Regulation*, Journal of System Safety, Vol 47, No 6, 2011. Accessed at http://www.system-safety.org/ejss/past/novdec2011ejss/spotlight1_p1.php. As well as Leveson N., *White Paper on The Use of Safety Cases in Certification and Regulation,* accessed September 6th, 2012 at http://sunnyday.mit.edu/SafetyCases.pdf.

PSI, 2012 - *Evaluation of STPA in the Safety Analysis of the Gantry 2 Proton Radiation Therapy System* (authored by Antoine B., Rejzek M., Hilbes C.) - *Draft*

Stringfellow M., Owens B.D., Leveson N.G., Ingham M., Weiss K.A., *Safety-Driven Model Based System Engineering Part I: Methodology Description*, MIT Technical Report for JPL, 2007.

Thomas J., Leveson N.G., *Performing Hazard Analysis on Complex, Software- and Human-Intensive Systems,* 29th International System Safety Conference, 2011. Accessed on August 10th, 2012 at http://www.system-safety.org/conferences/2011/

Thomas J., *Extending and Automating STPA for Requirements Generation and Analysis*, presentation given at the First STAMP/STPA Workshop, MIT, April 18th, 2012, available at http://csrl.scripts.mit.edu/home/get_pdf.php?name=2-6-Thomas-Extending-and-Automating-STPA-for-Requirements-Generation-and-Analysis.pdf

Williams W.T., *Numbers, Taxonomy, and Judgment,* The Botanical Review, Volume 33, Number 4 (1967), 379-386, DOI: 10.1007/BF02858741

5 Conclusions and Future Work

This dissertation investigated the applicability of STAMP-based Hazard Analysis STPA to the identification of unsafe behavior in complex systems on the basis of a case study in the medical device industry. This included:

- Describing the regulatory frameworks that apply in the USA and Europe to the market authorization of accelerator driven radiotherapy machines and evaluating whether STPA could be used as a method to perform the mandatory hazard analysis,
- Evaluating the applicability of STPA to the risk review of a complex system based on a case application to the PROSCAN proton therapy facility,
- Proposing new methodological processes to obtain, organize and display STPA results.

Several conclusions can be drawn as a result of this effort.

1. This research confirmed that STPA would fit into the regulatory structure that currently exists for authorizing medical devices on the US and European markets.

2. The PROSCAN STPA project added to the existing body of projects that have applied STPA to analyzing hazards in complex systems, further demonstrating the applicability and feasibility of applying STPA to an advanced technology system.

Although the evaluation potential was limited and more formal comparison is needed, some information about potential efficacy was derived from a comparison with the summary of hazardous behavior reported by the PROSCAN design team in its draft Safety Report to the Swiss regulatory agencies. Although more formal evaluation is needed, preliminary information about the benefit and ease of using STPA as a brainstorming support tool in a workshop for engineers was obtained from a PSI report on how the methodology allowed identification of scenarios that had previously not been identified.

STPA provides step by step guidance to facilitate performance of structured hazard analysis on complex systems. Its top-down nature and focus on functional architecture render its use

worthwhile at any design stage. Beyond providing assistance to system designers, it can be a powerful tool for hazard reviews and certification activities.

3. This research advanced the state of the art in nuclear medicine safety, and complex medical devices and systems in general, by introducing a systems perspective to hazard analysis in this field.

4. In the process of performing the PROSCAN STPA study, this research found that the amount of information generated required notations, ways of organizing and ways of presenting the results so that they could be located and, more important, understood by those who will be reviewing or using the analysis. The following methodological developments were proposed as solutions to these problems.

First, the importance of describing the hierarchical relationships of the system with increasing depth of detail was emphasized.

Second, notations and a process for specifying STPA scenarios (Step 2 Tree), presenting their causal relationships (hazard taxonomic trees), displaying hazardous scenarios (STPA data graphs and control structure displays) and evaluating the coverage provided by protective measures against unsafe behavior and displaying (STPA data graphs) were created.

Third, it was proposed that a software tool could store STPA results in a database coded with the data dimensions described in the STPA data graphs, query them according to several user scenarios and display them. Visual display would facilitate communication of hazardous behavior to design stakeholders and understanding of the extent to which protective measures address identified unsafe behavior of the system.

5. Finally, STPA was put into context in terms of how it can interact with other, more traditional hazard analysis methods, and especially Fault Tree Analysis.

These steps towards improving the design of medical devices and further developing the hazard analysis technique STPA open new investigation avenues. Three streams are identified for future research:

1. Further evaluating the contribution that STPA can make to safety in healthcare settings

Further evaluating the contribution that STPA can make to safety in healthcare settings includes investigating not only the unsafe behaviors that a medical device could assume, but also analyzing the workflow of the patient caring team that define and deliver care as well as looking into the relationships between device manufacturers, regulatory authorities and health-care providers. The PROSCAN facility could be studied from this angle.

Moreover, a retrospective analysis of past radiotherapy accident analyses could be undertaken to map the causal factors that were identified as causing the accident to the list of control flaws provided to perform STPA Step 2. Finding that past accident causes do fit into STPA Step 2 categories would not mean that a prior STPA analysis would have identified these unsafe scenarios before the accident happened and thus potentially contributing to avoiding the accident. However, finding accident causes that do not belong to any STPA control flow category would indicate that STPA would need to be improved to offer more thorough identification of hazardous behavior. The information provided by the retrospective analysis could be used to develop specific procedures for applying Step 2 in this particular application.

2. Developing tools to support the analysis and use of STPA results

STPA generates large amounts of data whose management can benefit from the power of computer databases and searches. Software tools that would both assist the analyst in performing the STPA analysis (listing hazards, creating the control structures, identifying unsafe control actions, identifying hazardous scenarios) and documenting protective measures, but also enhance the readability of the analysis by facilitating the selection of data subsets and visualize findings at different levels of abstraction.

Further, STPA understands that structure dictates behavior. Developing tests that would allow the analyst to point out aspects or missing pieces of a control structure that are likely to create unsafe behavior would be particularly helpful. After all, a control structure is a directed graph. It can therefore probably be analyzed using network analysis techniques. Similarly, tests can probably be designed to characterize pathways from controllers to hazards in the STPA data graphs presented in Chapter 4 and evaluate the relevance and strengths of protective measures that can be inserted between STPA data nodes.

Finally, it would be useful to develop heuristics and guidelines to help the STPA data user make decisions as to what criteria they should retain for deciding whether residual hazards are acceptable.

Ideally, one would be able to design hazards out of a system. It is however not always possible: as long as radiation is to be used for patient irradiation, keeping the radiation source shut is not an option. Design choices can reduce the likelihood of a hazard being realized, such as designing the transfer path from preparation room to treatment room to be short and smooth in order to prevent patient position from changing, and protective measures can be taken to reduce both the likelihood and the consequence of hazardous events, such as creating devices that immobilize the patient. However, as long as hazards, although made less likely, are not eliminated, one cannot rule out the possibility that they may occur.

When hazards cannot be eliminated and protective measures have to be put in place, they must in turn be evaluated in terms of their potential for hazard creation and chosen or designed to minimize this possibility. But should protection systems needed to prevent the hazards associated with these first protection systems also be included in the analysis? If so, when should this iteration stop?

STPA does not answer this question. It is a hazard analysis tool and, as such, aims to inform design decision-making, but it is not a decision-making tool. It leaves it to the designer to decide whether he deems the residual risk to be acceptable by the system's users. Recommendations on how to make this decision would be welcome in contexts where resources available for design and for system construction are limited.

3. Formally evaluating the resources needed to perform a complete STPA.

STPA has been applied to systems as diverse as missile systems (Pereira et al., 2006), space missions (Ishimatsu et al., 2010), weather satellites (NASA-JAXA research project, MIT CSRL 2012), road tunnels (Kazaras et al., 2012), civil aviation (NASA research project on ITP, MIT CSRL 2011-2012), radiotherapy (PROSCAN STPA study, PSI & MIT CSRL 2011-2012), nuclear power (NRC research grant, MIT CSRL 2012), truck engine design (Stefanie Goerges' MIT SDM thesis) and others. This experience points to the ease of its use and feasibility by small teams in reasonable amounts of time. However, a more formal assessment of the resources needed to perform a full analysis would be welcome as potential users are always concerned about their ability to adopt resource intensive processes. Ideally, this assessment would be compared to an evaluation of the resources needed to perform a full FMEA and FTA analyses, the leading methodological incumbents for hazard analysis in most industries.

□ □ □

This dissertation would not be complete without a grateful note to you, tenacious reader who generously spent much time in the company of this work.

The domain in which this research project was completed is one of growing importance, especially as exposure to harmful environmental factors increases, thus increasing the likelihood of harm to human health. The topic that was addressed, safety, is at the heart of the social contract that tacitly binds all system designers to users of their creations, all engineers to the societies that they live in. Although this work did not demonstrate the optimality of using STPA to create safe systems and assess how dangerous systems can be, it did bring a unique and convincing perspective to the hazard analysis of medical devices. Ultimately, the goal is to help make these products and the facilities that use them safer for both patients and personnel. By proposing avenues for the creation of STPA tools, this dissertation is also relevant to other domains.

I hope that reading this dissertation will have kindled interesting thoughts in your mind. May you go on to generate powerful ideas that will make this world a safer place!

PAGE INTENTIONALLY LEFT BLANK

Appendix 1 - Glossary

With assistance from the Merriam-Webster[61] dictionary, this dissertation proposes the following definitions for a few concepts used in this report. The reader's attention is brought to how the definitions and use of the words "actuator", "control action" and "controlled process", while still following (Leveson, 2012), have been extended compared to STAMP practice to date.

ACCIDENT: undesired or unplanned event that results in a loss, including loss or injury to human life, property damage, environmental pollution, mission loss etc.

ACTUATOR: a human operator or mechanical device tasked with directly acting upon a process and changing its physical state. Valve systems (valve + the motor associated to it), doors, magnets (their electronic controller and power source included) or a nurse are actuators that respectively implement control on the following processes: "fluid flow", "egress availability", "beam position", "patient position". Actuators, like sensors, can be smart in that they can be programmable; they may therefore need to be studied with the same concepts as the controllers are.

CAUSAL FACTOR: cause of a hazardous scenario (STPA Step 2).

COMMAND: a signal providing a set of instructions (goals, set points, order) issued by a controller with the intent of acting upon a process by activation of a device or implementation of a procedure. Communication and Control, along with Hierarchy and Emergence, are fundamental systems theory concepts at the foundation of STAMP. Commands are issued by Controllers, with the intent that they be implemented by Actuators to act on the Controlled Process[62].

[61] www.merriam-webster.com

[62] For information, one of the Merriam-Webster's definition of a command is: "a. an order, b. a signal that actuates a device (as a control mechanism in a spacecraft or one step in a computer)"

CONTEXT: the circumstances that form the setting for the issuance of a control command or an event. Is often defined in terms of interrelated conditions in which a control action occurs (STPA Step 1).

CONTROL ACTION: the bringing about of an alteration in the system's state through activation of a device or implementation of a procedure with the intent of regulating or guiding the operation of a human being, machine, apparatus, or system. They are the result of an Actuator implementing a control Command issued by a Controller, and aim at controlling the state of the Controlled Process[63].

CONTROL ALGORITHM: a step-by-step procedure used by a computer or human controller for deciding what control action must be taken to solve a problem or accomplish some end.

CONTROL STRUCTURE: hierarchy of process loops created to steer a system's operations and control its states. In the context of a hazard analysis, we are most concerned with the control of hazardous states aimed at eliminating, reducing or mitigating them.

CONTROLLED PROCESS: although at times reducible to the state of a physical element (e.g. framing a “door” as a controlled process whose values can be “open” or “shut”), it appears fruitful to rather consider the controlled process identified in STAMP process loops to be the system’s attribute or state variable that the controller aims to control (e.g. thinking of the door not as the controlled process but, together with its motor, as an actuator that implements control on the possibility of egress).

CONTROLLER: a human or automated system that is responsible for controlling the system's processes by issuing commands to be implemented by system actuators.

ELEMENTS: design units that, together, compose the system. Although they most often refer to physical or logical entities that can be found in very detailed system descriptions (e.g. a specific relay switch, a motor, a CPU and the software that runs on it), elements can also be conceptual.

[63] For information, one of the definitions that the Merriam-Webster dictionary provides for the noun "control" is: "a device or mechanism used to regulate or guide the operation of a machine, apparatus, or system". One of its definition for the noun "action" is: "the bringing about of an alteration by force or through a natural agency".

For example, the “spot dose controller” does not exist per se. However, conceptualizing the collaboration of several architectural elements that, together, aim at controlling the dose given in each spot as the responsibility of this conceptual controller brings valuable abstraction to the first steps of the analysis, ensuring that it does not depend on the precise architectural choices made and is therefore able to uncover valuable alternatives to these choices.

FEEDBACK: evaluative or corrective information about an action, event, or process that is transmitted to the original or controlling source..

HAZARD: system state of set or conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident.

LOSS: decrease in amount, magnitude or degree including destruction or ruin.

SAFETY: freedom from loss.

SAFETY CONSTRAINT: bound set on system design options and operations to restrict, compel to avoid or forbid the performance of actions that would lead to a hazard.

SAFETY REQUIREMENT: design requirement formulated to include the enforcement of safety constraints as a design objective.

(HAZARDOUS) SCENARIO: an account or synopsis of a possible course of action or events resulting in a hazard. See Causal Factor.

SENSOR: human or mechanical device tasked with measuring a process variable by responding to a physical stimulus (as heat, light, sound, pressure, magnetism, or a particular motion) and transmit a resulting impulse (as for measurement or operating a control)..

UNSAFE CONTROL ACTION: control action that leads to a hazard (STPA Step 1).

PAGE INTENTIONALLY LEFT BLANK

Appendix 2 - PROSCAN Architecture

The purpose of radiotherapy is to deliver a certain amount of energy to tumor cells to impair their replication mechanisms and kill them. In charged particle therapy, this translates into delivering a certain number of charged particles (pions, protons, heavy ions…) to the patient's tumor. The PROSCAN facility hosted at the Paul Scherrer Institute in Switzerland accelerates protons to treat ocular and deep seated tumors, especially in pediatric patients. This appendix presents additional information about its architecture.

The beamline's first function is to transport the beam up to the required target. When the target is not set directly at the exit of the cyclotron and the path to the target is not a straight line, beam steering is achieved by using magnetic fields that bend the trajectory of beam particles with an angle. This angle depends on both the particles' energy and the magnetic field that they go through.

The beamline's second function is to select the beam tune and momentum spread at the target according to beam users' expectations. If the proton beam coming out of the cyclotron is a well-defined one in terms of tune (=energy) and momentum, several mechanisms lead to degradation of the beam's quality as it travels through the beamline. Since protons are charged particles, electrostatic repulsion will lead to spatial diffraction of the beam, resulting in non optimal spread of the pencil beam at the target. Quality control of the beam is therefore actively ensured by a set of engineered systems:

- To reduce this effect, the beam is refocused in magnetic lenses distributed along the beamline: quadrupole and sextupole magnets, which can be thought of as exerting lateral pressure on the beam to maintain its focus.
- While the cyclotron is designed to produce 250 MeV protons, experimenters and treatment designers require a broad range of proton energies to perform their activities. Means must therefore be provided to degrade the beam's energy before it reaches its target. This is

achieved by having the beam interact with devices through which proton energy is dissipated. In PROSCAN, these consist in the shared degrader's graphite wedges, Gantry 1's range-shifter plates, Optis 2 degrading and modulation wheels, and Gantry 2's energy degrader used to obtain beams of very low energy. A well-designed degrader can very quickly change the energy of the beam in small steps to provide a spread-out-Bragg peak (SOBP).

However, these degrading devices spoil the quality of the beam through scattering and energy straggling (Pedroni et al 2004): as beam energy loss and momentum modification through interaction with matter follow a statistical distribution, not all 250 MeV protons will lose the same amount of energy. At the exit of the degrading device, the beam will therefore not only have diffracted some, but it will also exhibit momentum and tune spread, a statistical distribution of energy vectors that, if left unchanged, would reduce treatment and experimental precision. In order to maintain the quality of the scanning pencil beam, it is "beam-optically analyzed in the beam line section(s) following the degrader" (Pedroni et al, 2004):

- Collimators and slits are used to shave out the external fractions of the beam's spatial signature[64]. Then, advantage is taken of the fact that magnetic fields will bend a particle's trajectory with an angle that will vary depending on its incoming energy, making it possible to further use collimators and slits to perform energy selection by skimming away particles based on their post-bending trajectory.
- It is important to note that these filtering effects take protons out of the beam, and therefore reduce the current that will be delivered to the target. These quality-control induced losses

[64] Gantry-2's continuous scanning aptitude relies on fast energy and current modifications. Since inserting mechanical collimators is slow, an alternative current setting system is being explored: instead of relying on cyclotron current actuators, beamline collimators and slits, Gantry 2 could use quadrupole magnets to electrically defocus the beam on purpose onto a collimator so as to reduce current at will from a high current baseline. Thus eliminating the protons that would be in surplus at higher energies where losses through the degrader are smaller would make it possible to keep the 100% aperture settings for the ion source and phase slits at all energies. By compensating the lower degrader losses observed at high energies with the addition of controlled losses through the quadruple/collimator filter, a beam current constant across all energies could be delivered to the user area for a given deflector plate setting, meaning that current range would be accessible through direct setting of the deflector plate, without any change to either the phase slits or the ion source.

must therefore be accounted for when deciding upon the cyclotron's current actuators' settings. Further, they are highly energy dependent: for example, transmission[65] of a 100 MeV beam across the degrader is only 1% (Meer, 2007). This energy dependency makes it necessary to modify current levels each time the beam at target's energy is changed to get the desired current output: when producing beam at low energy, beam current losses in the degrader and the energy-selection devices downstream must be compensated with higher beam current intensities at the source.

The third function performed by the beamline is to provide diagnostic means for the beam. Both direct and indirect measurements are used to infer the beam's status. Direct measurements such as obtained from ionization chambers and strip chambers result from an instrument's interaction with the beam and can give information on its current (proton count per amount of time) and spatial spread. Indirect measurements do not come from instrument interaction with the beam, but consist of information obtained from the status of beamline elements that do interact with the beam. They, for example, include magnetic field measurement by Hall probes in bending magnets, electric current read-outs from actuators powered by electricity, information on insertion and position of the degrader wedges or the OPTIS 2 diffracting lenses obtained from potentiometers etc.

The beamline's fourth and final function is to stop the beam on demand and prevent it from entering undesired areas. A fast kicker magnet is used as a switch to turn the beam on and off, and beam stoppers are inserted into the beamline as a cork that would prevent beam entry at undesired times and place. There are three kinds of beam stoppers used in PROSCAN: slow and large graphite absorbers, fast copper inserts, and slow "Kanalverschlussen", rotating devices that, when open, align a hollow tube with the beam direction and, when closed, offer their side as a wall to the beam.

These functions of the beamline being understood, PROSCAN and its physical infrastructure can be described as follows:

[65] ratio of beam current intensity exiting the degrader over beam current intensity entering the degrader.

- **Beam creation**: cyclotron COMET (COmpact MEdical Therapy cyclotron) and its ion source, including their power supplies. Their joint purpose is to produce a 250 MeV proton beam at the current requested by the client area.

  Compared to synchrotrons, cyclotrons have the drawback of only being able to produce beams of a single energy level. However, they have the advantage of producing continuous beams whose intensity can be dynamically controlled rather than pulses.

  Current adjustment is achieved by acting upon the ion source, the cyclotron's phase slits, the cyclotron's deflector plate, and collimators and slits associated with beam quality control after spoilage through the degrader. In Gantry 2, these current modulation functions will be sped up by the inclusion of a defocusing quadrupole paired with a collimator. The speed benefit is realized by the fact that the slow-to be changed ion source and phase slits settings can then be kept constant at levels ensuring maximum cyclotron output current even at high energy levels, where losses through the degrading mechanisms are smaller and can now be augmented by appropriate defocusing of the beam.
- **Beam transport**: shared beamline, from COMET to bending magnet AMA1 (see Figure 5). The beamline consists in a collection of actuators, complete with their power supplies, controllers and diagnostic tools, whose role is to set the beam on or off (kicker magnet), define the energy at which it will be delivered to the area coupling point (degrader), and keep it focused in momentum and space (quadrupole and sextupole magnets, collimator, slits) until it is delivered to the local user area. The collimators, slits and, in the case of Gantry-2, defocusing magnet, change the current of the beam.
- **Beam transport to local areas**: four separate local beamlines. These consist in a collection of actuators (magnets, collimators, beam stoppers) and sensors whose goal is to keep the beam focused until it is delivered to the client area. The beam stoppers ensure that no beam is delivered to the local area unless it has been requested, unless the beamline settings allow for safe beam delivery, and unless the local area is ready to receive it. From AMA1 to AMA3, actuators including three bending magnets, two stirring magnets, one collimator, several quadrupoles, one beam stopper and several diagnostic tools are shared by Gantry 2, OPTIS 2 and the experimental area. After AMA 3 however, each area is sole master of its local beamline actuators including one (experiment), three (OPTIS 2) and five (Gantry 2) beam stoppers each.

- **Beam use**: four separate client areas. Upon reaching the end of each local beamline, the beam goes through the coupling point and enters the client area. There, the beam goes through a final beamline section. Depending on which client area it is, it will be further bent (bending magnets), tuned (local degradation systems, such as OPTIS 2's modulator wheel or Gantry 1's degrader plates), focused (quadrupole and sextupole magnets), steered (steering magnets) and swept (magnets that achieve directional control of the beam at target in Gantry 1 and 2) before being delivered to its target.
- **Communication backbone**: To enable communication of commands to the many elements featured in PROSCAN, a distributed computing environment was deemed necessary to overcome the otherwise complicated wiring issues that a centralized command and control system would have required. Further, properly abstracting the communication and computing hardware away from the logic that it implements was considered essential to provide the facility with the flexible adaptability that it would need, over its long lifetime[66], to keep up with changes in its operating environment, such as discontinuity in hardware parts supply when the market ends up moving away from the once state-of-the-art solution that would have been chosen at PROSCAN. This multiple controller environment provides a high degree of flexibility to the architecture, but requires enforcement of a good hierarchy of commands to be effective.
  - In this distributed architecture, centrally issued commands are sent over an Ethernet network to which VME crates are connected. These crates host computers and logic boards that are locally responsible for stirring and monitoring individual elements to match therapy expectations. Each local element is thus controlled by a local controller who receives setting commands from higher level controllers, implements a local control loop to ensure that actuator state will converge towards the desired setting, and provides higher level controllers with status information on the actuator that it is in charge of.
  - CPU card (a.k.a IOC) in the VME crate communicates with the network on one hand and with other cards in the crate through the crate's bus. These other cards can be controllers for individual actuators (e.g. high voltage power system controller card used to control voltage supply on a dose

[66] In the words of Simon Rees, in charge of the EPICS system at PSI, "we need to design a control system that is flexible, because change is inevitable". Personal communication, August 2011.

detector). Within PROSCAN, there are 28 IOCs, each of which are in charge of controlling different devices.

- These CPU cards host VxWorks, the standard operating system at PSI, which eliminates the need to have to code software from scratch. These run EPICS, an OpenSource toolkit to write applications for distributed environment (written by Argonne National Lab and Los Alamos National Lab, it stands for "Experimental Physics Industrial Control System" (EPICS, 2012)). Complete with the ability to easily design user interfaces, it is a standardized platform to manage control and monitoring problems in a distributed system.
- EPICS defines several control points, each known as a process variable (PV). Each PV is a point you can write to or read from. As an illustration, a magnet controller could use the following three PVs: Magnet1.on, Magnet1.desired_field_strength, Magnet1.actuat_field_strength. There are some 28,000 such PVs used in PROSCAN.
- Most of these cards were bought from external manufacturers. A few had to be designed in-house, when unique performance (high speed, high precision) was sought.
- VMEs are used to provide a highly reliable, low-risk backbone to control the beamline elements. Another option could have been to use cheaper Programmable Logic Controllers (PLC). However, not being meant to support real programming language, they are less versatile and therefore offer fewer functionality to easily design complex systems.
- The strategy chosen to ensure high availability of the communication backbone in the event of hardware failure was to stock replacement parts and ensure prompt maintenance rather than create a redundant and more complicated architecture.

■ **Coordination master:** All user areas have the ability to issue commands over the network, not only to their local beamlines but also to the shared beamline elements. To prevent the reception of conflicting commands by beamline elements and operator mode confusion (an area user believing that he is controlling the beam when he is not), it is critical that only one user area be known to be using the beam and allowed to issue commands to the beamline elements. The PROSCAN design team therefore made commands over local beamline elements only issuable by the corresponding local areas, and defined the concept of beam mastership through which only one area will be authorized to transmit commands over the network to the shared beamline elements. Instead of forbidding the transmission of commands to local beamline elements by the non-master areas, local areas maintain the

capability to issue executable commands to their local actuators at all times, resulting in efficiency gains. Further, the condition of "sub-mastership", which does not allow beam to enter the local area but permits the activation of local elements, was defined to provide control by the sub-master to the elements that Gantry 2, OPTIS 2 and the experiment area share between AMA1 and AMA3 when Gantry 1 has beam mastership.

- **Safety dedicated systems**: the design philosophy that was followed by the PROSCAN design team was to separate the protection and safety systems from the machine control system. In the words of one such designer, the separation of these duties was key to getting the facility licensed. These protection and safety systems are as follows:
  - patient safety system: avoid overdose
  - patient safety system : avoid personnel exposure to radiation
  - machine protection system : protect the facility's equipment
  - user facility safety systems: protect the local beamlines

  Each of these systems can turn the beam off, and several redundant, escalating mechanisms can be used to do so:

  - insert beam stoppers into or remove beam stoppers from the beamline
  - kick the beam away (deflector/Amaki)
  - reduce the power of the HF source to the cyclotron
  - shut down the ion source: used as a last resort, since it can break the ion source by making a hole in the vacuum chamber.

Pions: To ensure conformal distribution to the target volume, pions were sent to the target from 60 beams. 140 MeV negatively charged pions, upon absorption by atomic nuclei, lead to a tiny atomic explosion that damages the cells in which that interaction took place. However, these light particles tend to scatter in tissue, leading to non-optimal dose distribution in the target. Further, it was difficult to produce, collect and focus the particles, whose spot's trace was 5 cm wide. The pion project was stopped in 1992 after some 500 patients had been treated for pelvic tumors and large sarcomas.

Protons: in the early 1980s, it was decided to explore the use of protons instead, first for optical tumors in the OPTIS facility, then, around 1990, for head and body tumors in the Gantry-1 user area. PSI was already equipped with a proton-accelerating cyclotron that was used to perform physics experiments and materials research. As the maximum current rating of this cyclotron and the energy at which its beam was delivered (590 MeV) were much higher than what was required for proton therapy, protons were skimmed off from the cyclotron's output to enter the treatment beamline, where their energy was severely reduced by a degrading mechanism consisting in inserting graphite into the beamline before the beam was refocused and delivered to the treatment area. The treatment center was then simply one of many beam users, moreover not a priority one.

The current produced by this large cyclotron was too unstable to be relied upon as an input to the dose calculation that would define how long the treatment should last. The PSI team worked around this constraint by inventing the spot-scanning technique. In dynamic spot-scanning, dose is delivered as a series of short pulses (i.e. "spots"), whose number of protons is calculated by dose counters; the spot is stopped when the required number of protons has been counted as having been delivered and the beamline actuators' settings are changed to position the beam to its new location. This solution made the short-time current fluctuations irrelevant, since dose was calculated as the integration over time of the current that was actually observed.

Unfortunately, the continuity of the beam used for therapy was negatively impacted by the priority that was given to experiments run with the cyclotron. Trying to get the most out of the machine, experimenters often pushed the beam generation capacity to its limits, leading to frequent occurrences of short "no-beam" phases that impacted the smoothness of medical treatment operations. Given the success that was nonetheless achieved by the proton therapy project[67], PSI decided in 2000 to launch PROSCAN, an initiative to further expand its activities in this field.

PROSCAN's first step was the installation of a dedicated commercial superconducting cyclotron. Designed to match the needs of the therapeutic client areas, this smaller cyclotron was co-

[67] from 2004 Pedroni et al, "PSI is still the only location in which proton therapy is applied using a dynamic beam scanning technique on a very compact gantry".

engineered with ACCEL Instruments GmbH, a German firm since then bought by American manufacturer of radiation therapy devices Varian. COMET started operations in 2006. It produces a 250 MeV proton beam characterized by more stable particle currents.

Gantry-1 was linked to the new proton source. Using dynamic proton spot-scanning, it has successfully treated more than 650 infants, children and adults of various ages for a set of brain and body tumors since 1996. In order to expand the range of treatment indications to include mobile tumors (e.g. those that are attached to deformable and mobile body structures such as the lungs or even the prostate), and as explained in (Pedroni et al, 2004) a faster scanning instrument is needed. The second step of the PROSCAN project is meant to tackle this challenge through the development of fast (spot or continuous) scanning in a new gantry, Gantry 2. Through the availability of a faster scanning system enabled by a new set of design choices including dynamic control of beam intensity by the deflector plate at the ion source, 2D magnetic sweeping of the beam that limits the need to move the patient table to the application of large fields and faster dynamic variation of beam energy than Gantry-1's mechanical range shifters could achieve, it will be possible to treat the target volume repeatedly in the same session.

The key features of new treatment area Gantry-2 can be summarized as follows (Pedroni et al, 2004, 2011a):

- 2D lateral scanning: fast sweeper magnets displace the proton beam laterally in two dimensions at the isocenter. Compared to prior versions/areas, where beam would move in only one direction, this limits the need to (slowly) move the patient table to cases of large tumors (field patching);
- Faster field application:
    - The complete beam line is constructed in such a way that fast energy changes are possible (i.e. fast changes in deposition depth)
    - Fast sweeper magnets are used by the advanced parallel beam scanning mechanism (i.e. fast changes in 2D spot position).
- Repainting: thanks to faster field application, the system will be able to apply the same spot sequence several times in a short amount of time. "Repainting" is a promising

strategy to cope with organ motion and will extend treatment possibilities to new medical indications;

- Smaller spot size: the beam energy modulation system is located right after the exit of the cyclotron, resulting in smaller spot sizes and, therefore, higher precision of the treatment.
- Easier patient set-up: the isocentric layout (no moving floor) allows for easier and safer patient handling and provides more comfort to the patient and medical staff;
- Target position control: state-of-the-art imaging devices (both CT and X-ray system) are available in the treatment room. In addition to in-room positioning, they will offer useful visualization in the context of moving organs.

Appendix 3 – Notes on Quantitative Risk/Safety Assessment

This short discussion of quantitative risk assessment is meant as an introduction to how it is used in different industry settings and to some of the issues raised by this use.

A frequent treatment of safety as a design characteristic follows the paradigm that, to put it bluntly, a system is safe when it is acceptably[68] risky. The definition of "reasonable" in the ALARP concept of the UK occupational safety law as well as the related ALARA[69] concept used in radio-protection settings, the existence of Safety Integrity Levels (SIL) thresholds in the Functional Safety standard ISO 61508, the inclusion of probabilistic safety objectives in nuclear regulatory matters, the establishment of quantitative safety risk criteria in the chemical industry or the computation of safety performance targets to define acceptable levels of safety (ALoS) in civil aviation all point to the idea that not only can safety be quantitatively measured, but it should be so in terms of susceptibility to risk, where risk is defined as a function of both an accident's consequences and its probability of occurrence[70], and relatively to the cost of eliminating or mitigating that risk. Despite the fact that ample evidence has been collected to prove that "*"riskiness" means more to people than "expected number of fatalities"*" (Slovic,

[68] or "tolerably risky". As observed by the RCGuidelines (Frank et Jones, 2010), risk tolerance is not the same as risk acceptance. While o accident harming people should be considered acceptable, "*based upon the fact that there is no such thing as zero risk, society tolerates some level of risk in return for the benefits derived from the activity posing the risk*".

[69] respectively As Low As Reasonably Practical and As Low As Reasonably Achievable

[70] For example, ISO 14971 defines risk as the "*combination of the probability of occurrence of harm and the severity of that harm*" and RCGuidelines (quoted by (Frank et Jones, 2010)) as "*a measure of human injury, environmental damage, or economic loss in terms of both the incident likelihood and the magnitude of the loss or injury*". This understanding has remained astonishingly stable for three centuries (Bernstein, 1996), since French mathematician Abraham de Moivre formulated it in his letters to the Spring of 1711 edition of the Philosophical Transactions as "*The Risk of losing any sum is the reverse of Expectation, and the true measure of it is, the product of the Sum adventured multiplied by the Probability of the Loss*" (Hald, 1984).

1987), the resulting metric are usually then compared to acceptance criteria meant to reflect society's tolerance of risk and define the safety constraints of the design tradespace[71].

1 Using quantitative risk metrics to guide design choices and regulatory decisions

ALARP (As Low as Reasonably Practical, UK HSE): The UK Health and Safety Executive (UK HSE, 2012), for which ALARP and SFAIRP (So Far as Reasonably Practical), as encoded in the general duties of the Health and Safety at Work etc. Act of 1974[72] and reinforced following Lord Cullen's 1989's investigation report on the Piper Alpha oil platform fire, is word of law when it comes to making provisions to avoid risks, defines these concept by referring to the Court of Appeal in its judgment in Edwards v. National Coal Board73, [1949] 1 All ER 743 (emphasis added). In the words of Lord Justice Asquith (Jones-Lee, 2009): "'*Reasonably practicable' is a narrower term than 'physically possible' … a computation must be made by the owner in which the quantum of risk is placed on one scale and the sacrifice involved in the measures necessary for averting the risk (whether in money, time or trouble) is placed in the other, and that, if it be shown that there is a gross disproportion between them – the risk being insignificant in relation to the sacrifice – the defendants discharge the onus on them*."

[71] Risk tolerance, as defined by the United Kingdom Health and Safety Executive quoted by the RCGuidelines (Frank et Jones, 2010): "*refers. . .to a willingness by society as a whole to live with a risk so as to secure certain benefits in the confidence that the risk is one that is worth taking and that it is being properly controlled. However, it does not imply that . . . everyone would agree without reservation to take this risk or have it imposed on them*".

[72] whose long title is:" An Act to make further provision for securing the health, safety and welfare of persons at work, for protecting others against risks to health or safety in connection with the activities of persons at work, for controlling the keeping and use and preventing the unlawful acquisition, possession and use of dangerous substances, and for controlling certain emissions into the atmosphere; to make further provision with respect to the employment medical advisory service; to amend the law relating to building regulations, and the Building (Scotland) Act 1959; and for connected purposes."

[73] The case revolved around whether it was reasonably practical to prevent even the smallest possibility of a rock fall in a coal mine (Jolliffe, 2008)

Use of the ALARP concept to define design acceptability criteria: In practice, ALARP/SFAIRP is said to be achieved when good practices[74] are followed or, in complex situations that may involve new technology for which no good practice is available, when a formal cost-benefit analysis identifies further risk mitigation measures to be (grossly) disproportionate to the benefits they would achieve in cases where the risk is neither broadly acceptable (case where no mitigation is required) or unacceptable (Jolliffe, 2008). The ALARP concept allows the regulator to prescribe a performance goal to the operator rather than follow a prescriptive approach that would mandate detailed requirements and procedures to be enforced[75], thus addressing the conclusion that "*the single most important cause of accidents was apathy on part of all concerned in industry, and second, that a major cause of this was that there was simply too much law*", not encouraging industry to take ownership of its own safety (Rimington et al., 2003, describing the 1972 Robens Report on Safety and Health at Work, whose recommendations were put into effect in the Health and Safety at Work etc. Act of 1974). It puts the responsibility for designing safe systems on the operator's shoulders. It shall be noted that albeit (Rimington et al, 2003) emphasizes the necessity for ALARP regulators to be "*technically competent to conduct the necessary dialogue*", goal-oriented prescription require less resources from the regulator since its staff is not expected to draft detailed and must therefore be on top of all technological developments.

SIL (Safety Integrity Levels, IEC 61508): IEC Standard 61508 deals with the contribution of electrical and electronics systems to functional safety defined as "the ability of a safety instrumented system or other means of risk reduction to carry out the actions necessary to achieve or to maintain a safe state for the process and its associated equipment" (ISA, 2002).

[74] These are not defined as the higher standards of risk control but as those recognized by consensus among several stakeholders to be satisfying the law when applied to a relevant particular case .

[75] For more information on the difference between prescriptive and performance-based regulatory approaches to safety, see (Leveson, 2012b)

In this context, Safety Integrity Levels (SIL) are designed to measure a relative level of risk-reduction provided by a safety function76, or to specify a target level of risk reduction. In short, SIL are a measurement of the probability of failure on demand required for a Safety Instrumented Function within a Safety Instrumented System77 based on the ANSI/ISA 84, IEC 61508, and IEC 61511 standards, complemented by prescriptions about the design process of software (Bell, 2005; Net Safety Inc, 2012).

Table 28 - SIL Levels Defined in IEC 61508

SIL Level	Probability of Failure on Demand (per year)	Risk Reduction Factor
1	0.1-0.01	10-100
2	0.01-0.001	100-1,000
3	0.001-0.0001	1,000-10,000
4	0.0001-0.00001	10,000-100,000

As explained in (Net Safety Inc., 2012), standard IEC 61508 defines SIL using requirements grouped into two broad categories: 1. hardware safety integrity and 2. systematic safety integrity.

1. The SIL requirements for hardware safety integrity are based on a probabilistic analysis of the device. To achieve a given SIL, the device must have less than the specified probability of dangerous failure and have greater than the specified safe failure

---

[76] IEC 61508 (whose scope is defined as follows: " *This International Standard covers those aspects to be considered when electrical/electronic/programmable electronic (E/E/PE) systems are used to carry out safety functions ")* defines a safety function in part 3.5.1 as a "*function to be implemented by an E/E/PE safety-related system or other risk reduction measures, that is intended to achieve or maintain a safe state for the EUC [equipment under control], in respect of a specific hazardous event. Examples of safety functions include:*

– functions that are required to be carried out as positive actions to avoid hazardous situations (for example switching off a motor); and

– functions that prevent actions being taken (for example preventing a motor starting)."

[77] A set of components arranged for the purpose of taking the process to a safe state when predetermined conditions are violated. It is composed of sensors, logic solvers, final elements and is also known as Instrumented Protective Systems, Safety Interlock or Emergency Isolation/Shutdown System.

fraction. These failure probabilities are calculated by performing a Failure Modes and Effects Analysis (FMEA). The actual targets required vary depending on the likelihood of a demand, the complexity of the device(s), and types of redundancy used.

2. The SIL requirements for systematic safety integrity define a set of techniques and measures required to prevent systematic failures (bugs) from being designed into the device or system. These requirements can either be met by establishing a rigorous development process, or by establishing that the device has sufficient operating history to argue that it has been proven in use.

While implicitly acknowledging the fact that software reliability is a moot concept when prescribing the use of a rigorous development process to reduce the likelihood of software contributing to the creation of hazardous situations rather than suggesting the same probabilistic approach as recommended for hardware, it does philosophically equate safety with the achievement of a required reliability performance by safety dedicated systems: the larger the risk, the more reliable risk reduction measures must be implemented and the greater the reliability[78] the components used must exhibit.

CDF (Core Damage Frequency, US NRC): As presented in (NEA, 1994), almost every member country of the Nuclear Energy Agency of the Organization for Economic Cooperation and Development used Probabilistic Safety Criteria for the safety assessment of nuclear power plants as of 1994. These might be dose limits for anticipated occupational exposure and design basis accidents with implicit or explicit frequency considerations, or calculations related to the probability of loss of core integrity. At the time this report was written, probabilistic criteria for mortality risk had only been formulated in three countries: the Netherlands, the United Kingdom and the USA.

The US nuclear industry is famous for the developments that it brought to probabilistic risk analysis/assessments (PRA) in the wake of the Reactor Safety Study published as WASH-1400, NUREG/75.014 and also known as the Rassmussen report. (Kadak and

[78] probability that a system can perform a defined function under stated conditions for a given period of time.

Matsuo, 2007) provide a detailed account of how the risk-informed approach based on PRA was progressively introduced in US regulation of the nuclear industry, presenting the culture change that it required and assessing its effects on the operational and safety performance of plants that adopted it as positive.

The approach developed in 1975 in WASH-1400 indicated that it was possible to create quantitative, probabilistic measures of plant safety, helping determine that risk was not dominated by design basis accidents but by much smaller in consequence but much more likely to occur accidents. Given the shortcomings identified by the Lewis Committee especially with respect to the treatment of uncertainties, the Nuclear Regulatory Commission's staff was instructed in 1978 that the regulatory decision could not be solely based on PRAs and must have a deterministic basis. However, NRC found PRAs to be useful for insights into risks and problem identification, requiring it be used for identification and analysis of plant safety vulnerabilities in 1988, then allowing plants to develop risk-informed maintenance programs in 1991 and finally agreeing to amend plant licenses based on risk information in 1996, as part of the PRA Implementation Plan of 1994. From the Risk Informed Regulation Implementation Plan that followed in 2000, to the Risk-Informed, Performance-Based Plan started in 2007, the transition to a risk-informed regulatory system continues, with the goals of reinforcing the identification of plant vulnerability and helping prioritize maintenance efforts according to equipment safety significance (Kadak and Matsuo, 2007). It is interesting to note that the NRC Strategic Plan to move toward risk-informed performance-based regulation was in part motivated by the 1993 "Government Performance and Results Act" passed by Congress to "improve Federal program effectiveness [...] by promoting a new focus on results, service quality, and customer satisfaction". (NRC, 2012)

As summarized on its website, the NRC uses event trees and fault trees to perform PRA that estimate nuclear power plants' risk by identifying what can go wrong, how likely these events are, and what are their consequences. A Level 1 PRA's output is the calculation of the frequency of accidents that cause damage to the nuclear reactor core, commonly called core damage frequency (CDF). A Level 2 PRA then estimates the frequency of accidents that release radioactivity from the nuclear power plant. A Level 3

PRA finally estimates the consequences in terms of expected injury to the public and damage to the environment.

Important note: Albeit NRC increasingly considers probabilistic risk information in its oversight of nuclear power plants design and operations, the legal basis for certification and regulatory approval remains the traditional deterministic approach which requires the inclusion of safety systems capable of preventing and/or mitigating the consequence of severe accidents, following the defense-in-depth principle and including relevant safety margins. This is quite clear in the 1995 PRA Policy Statement (60 FR 42622 of August 16th, 1995 - emphasis added): "*The use of PRA technology should be increased in all regulatory matters to the extent supported by the state of the art in PRA methods and data, and in a manner that complements the NRC's deterministic approach and supports the NRC's traditional defense-in-depth philosophy."*

Chemical industry: Guidelines for Developing Quantitative Safety Risk Criteria (RCGuidelines) are provided by the American Institute of Chemical Engineers's Center for Chemical Process Safety (CCPS, 2009) to promote "more effective risk decision making" and provide "improved risk assessment tools", especially with respect to determining the risks associated with infrequent incident events, whose probability of occurrence is said to be challenging for even the most knowledgeable teams to assess with qualitative methods such as hazard and operability studies (HAZOP) and What If studies.

ALoS (Acceptable Level of Safety, ICAO, 2009): Arguing that the elimination of accidents and/or serious incidents are unachievable goals in open and dynamic operational contexts and that hazards are integral to aviation environments in spite of the best efforts to prevent them, ICAO stresses safety as a relative concept rather than an absolute one[79], defining it as (emphasis added) "The state in which the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management." (Ch. 2)

[79] "Safety is therefore a concept that must encompass relatives rather than absolutes, whereby safety risks arising from the consequences of hazards in operational contexts must be acceptable in an inherently safe system" (ICAO, 2009, Chapter 2, paragraph 2.2.3)

Following "the basic management axiom that one cannot manage what one cannot measure 80", it goes on to introduce the notion of acceptable level of safety (ALoS) as "the way of expressing the minimum degree of safety that has been established by the State and must be assured by an SSP", and the notion of safety performance as "the way of measuring the safety performance of a service provider and its SMS" in the context of a perceived need to "complement the historical approach to the management of safety based upon regulatory compliance, with a performance based approach" (Ch. 6).

These measures are claimed to help guide attention towards places where possibly corrective "action may be required to bring operational performance of the system to the level of design expectations" and allow for the assessment of whether safety critical activities are performed in a way that safety risks "can be maintained ALARP" (Ch. 6).

While these few examples demonstrate the prevalence of the quantitative "risk-based" safety measure in current risk assessment and risk management practices, the clarity and appeal of this construct must not mask the difficulties it carries with it, and important drawbacks associated to its use. These can be summarized as follows: 1. subjectivity is inherent to the process but quantification trumps the decision-makers with the disguise of objective truth carried by the numerical representation, 2. data is often missing for a truly representative probabilistic risk evaluation to be performed adequately, 3. the definition of acceptability levels pose ethical and practical issues and 4. the existence of acceptability levels may lead designers to satisfice[81] design by meeting these thresholds rather than focus on eliminating the hazards.

[80] "In any system, it is necessary to define a set of measurable performance outcomes in order to determine whether the system is truly operating in accordance with design expectations, as opposed to simply meeting regulatory requirements." (ICAO, 2009, Chapter 6, paragraph 6.4.3)

[81] While in a different context, the term is here used as in (Gigerenzer, 2010): "*A heuristic is a mental process that ignores part of the available information and does not optimize, meaning that it does not involve the computation of a maximum or minimum. Relying on heuristics in place of optimizing is called satisficing*."

2 Difficulties and drawbacks associated with the risk measurement of safety

2.1 A subjective process in the guise of an objective metric

As (Jonkman et al, 2003) points out when summarizing the literature on quantitative risk assessment, there are four distinguishable phases in quantitative risk management:

• Qualitative analysis: Definition of the system and the analysis scope; identification and description of the hazards, failure modes and scenarios.
• Quantitative analysis: Determination of the probabilities and consequences of the defined events. Quantification of the risk in a risk number or a graph as a function of probabilities and consequences.
• Risk evaluation: Evaluation of the risk on grounds of the results of the former analyses. In this phase the decision is made whether or not the risk is tolerable, often with a reference to acceptability thresholds.
• Risk control and risk reduction measures: Depending on the outcome of the risk evaluation, measures may have to be taken to reduce the risk. It should also be determined how the risks can be controlled (for example by inspection, maintenance or warning systems).

Risk assessment, be it quantitative or qualitative, is concerned with the first two phases, both the identification of the hazardous scenarios and the subsequent determination of their probabilities and consequences. As emphasized by (Woolridge, 2008), both these phases are extremely subjective, even when a specified risk question has been defined and the search for numerical input into either qualitative or quantitative assessment is very thorough:

> *"Within any risk assessment, of either [qualitative or quantitative] type, judgments will be used through-out. These may be the risk assessors' judgments, or expert opinion, or both, and these will always be subjective. This will apply when selecting (and rejecting) data[82], delineating the risk pathways, applying weightings to data or model pathways,*

[8282] We would like to point out as a side-note that this selection bias unfortunately plays a large role in another activity concerned with safety: that of investigating accidents.

selecting the distributions used in a stochastic model, as well as selecting a description of high, low and so on, in a qualitative assessment."

Unfortunately, this subjectivity may not be recognized by the consumers of the hazard analysis as numbers bear a magic that most closely resembles truth. "*From the earliest times, man has endowed numbers with magical properties*" writes computer scientist W.T. Williams when addressing the use of numerical methods for taxonomy (Williams, 1967). He probably had in the back of his mind the symbolism associated with numbers in religious practices as diverse as the Egyptian mythology, the Eleusinian Mysteries, the Maya ceremonies and the Jewish Kabbalah, not to mention the sect organized around mathematician Pythagoras or the continued success of numerology in pretending to reveal their personality to people in search of self-knowledge.

That they have come to characterize scientific endeavors, in essence the pursuit of knowledge about the physical truth(s) of our universe, adds to the common perception that numbers are truthful representations of reality. Yet, warns Williams, although "*to most of us they are entities to be viewed with apprehension and awe*", "*numbers are contingent properties of the landscape*". Even the numerical classifications churned by his era's computers to deliver new "objective" (as in without bias in following the rules that are given to them) classifications for the use of taxonomists are subjective at heart: the requirements that guided the creation of the computer

---

While accident investigation are prime data for system designers and system users to learn from in order to improve their design, management and operations, they are often performed in the context of litigation (victims insisting that their grief be recognized), power struggles (powerful forces having stakes in certain organizations and technologies not being questioned), and the shared beliefs that human error is strongly associated with incompetency or willful complacency and that accidents are caused by "weak links" or "root causes".

As such, their account of an accident not only may omit consideration of factors that may lead to interpretations inconsistent with these constraints (including confirmation bias when a hypothesis is strongly aligned with the investigators' beliefs) but may focus on meeting a social purpose: that of finding a scapegoat (term coined from the Old Testament, where the sins of Israel are symbolically placed upon the head of a goat then sent into the desert to roam, while another goat is sacrificed to God,) to alleviate the victims' griefs, a process which has, from immemorial times, allowed archaïc societies to survive the violence that they generate (Girard, 1982).

(Dekker, 2003, 2007, 2011) convincingly presents the toll that the common understanding of accountably turned into an assigning blame practice puts on safety, notably by preventing the reporting of incidents that the system could usefully learn from. He offers in contrast the concept of "just culture", one foundation for which consists in providing much social care and attention to victims, giving the example of Nepalese sherpas and Scandinavian societies, where laying blame accurately is considered much less important than generous treatment of the victims.

programs that generate them are born of assumptions made by the programmers about how species should be organized, i.e. guided by the model they had of species' organizations. As such, and despite both our cultural tendency to grant numbers more signifying power than they hold and that of number crunchers to suggest that their objectivity is larger than that offered by subjective methods, numbers and numerical models should not be endowed with more powers than they have. "*In some numerical writing [Williams] think[s] [he] ha[s] detected a hint of authoritarianism - a suggestion that the methods are objective and absolute, free from human error, enshrining some form of revealed truth*." They are not. They offer one representation of the world, one that is as weak or as strong as the subjective models which they come from.

This issue could certainly be addressed by ensuring that the assumptions that led to the establishment of the risk estimates here discussed, such as the uncertainty associated with extrapolation of reliability data obtained from one plant to a whole fleet of nuclear power plants whose age, maintenance policies and other operational factors may be very different one from the other, that due to the small sample of elements tested or to the large variation in expert response to a probability distribution function elicitation question, are presented to the decision-maker[83]. But these are non trivial matters. "*With a quantitative risk assessment, many people may not have the knowledge base to directly understand the computations involved. They will need to rely on the explanations and opinions of the risk assessor to explain how the result was reached and what were the underlying assumptions, judgments and uncertainties*" whereas "*most people should be able to understand and follow the arguments*" of a logically written qualitative assessment (Woolridge, 2008)

Finally, when the simplicity of a number is available, how to be sure that, despite its known deficiencies, it will not end up being the sole basis for decision? Put otherwise - why should it,

[83] *" Therefore, any risk manager, policymaker, or other stakeholder who needs to use, or wishes to understand, a given risk assessment should not look only at the final "result". They should have some understanding of how that result was reached."* (Woolridge, 2008)

when other propositions can be made to describe the hazards associated with given architecture choices?[84]

2.2 Issues related to QRA data availability

Here comes a profound methodological difficulty: although estimating the size of losses possibly associated with specified undesired system states is, when staying short of converting them to a common unit (see the discussion below on the value of a statistical human life), straightforward, estimating their frequency or likelihood is a much harder and at times impossible task. Rare is the case where probability data is available, unambiguous and certain:

Historical data offers a solid reference point for identifying expected external disturbances (e.g. a site's susceptibility to natural catastrophes) and physical component behavior under normal, expected conditions and specific stress instances (e.g. reliability data, test results, probability distributions for the frequency of given component failures). It can provide a first approximation of their frequency, under the weak assumption that the drivers behind these phenomena will remain unchanged (e.g. assume the frequency of strong storms will not change despite surface temperature of the oceans increasing as a consequence of global warming) or the slightly stronger one that it is the relationship between their drivers that will remain unchanged (e.g. assume that the relationship between the frequency of strong storms and ocean surface temperature remains constant, and account for the change in surface temperature due to global warming). Nonetheless, even in cases where these assumptions hold, it is lacking on several grounds:

- the frequency of very rare events will not be properly estimated for lack of observable data and may require the use of subjective expert judgment elicitation, themselves subject to uncertainty that is difficult to characterize in quantitative terms beyond the variability that can be observed within a sample of specialists when it should also include an

[84] While being skeptical about the use of numerical estimates of risk as measures for system safety, the author believes that numerical risk assessment methods, used in conjunction with other design evaluation techniques and within the limits to their validity (i.e. not for humans or software) addressed in the following paragraph, have a useful role to play in guiding system design towards systems that effectively mitigate hazardous behavior. Understanding a probabilistic risk assessment as one possible albeit necessarily imperfect model of system behavior can lead to valuable identification of design aspects that should be given more attention for the benefit of the whole. This search for "high potential" nodes can be performed with importance metrics and sensitivity analyses on key uncertainty parameters.

evaluation of how certain the epistemic foundations on which their knowledge is founded is[85],

- component and human behaviors are influenced by the environment they are deployed in, meaning that data gathered in one setting may not be validly applied to another (e.g. calculating erosion rates for structures exposed to different external environments based on data gathered in a given - possibly laboratory - setting, estimating the probability of specific operator action during an accident based on slip measurements in a controlled human factors experiment or from incident databases, applying industry wide averages to individual, possibly best or worst of their class, organizations),
- it may not be available for essential system parts such as software (an abstraction for whom "reliability"[86] is a moot concept as repeated by Leveson (2012 and earlier)) and technologies so recently introduced that they don't have so large behavioral records that statistical inference would make sense[87],
- its use requires that very detailed knowledge of the system's architecture be known, making quantitative assessment an unlikely guide for early design decisions.

These fundamental deficiencies have to do with the existence of both epistemic and aleatory uncertainty in the data being assessed and the generation of this assessment. (Paté-Cornell, 1996) discusses them in a straightforward manner, along with six possible levels for treating them that depend on the alternatives of the decision, on the management rule that one intends to apply, on

---

[85] For example, pre-Copernic astrophysians knew that their calculations of star trajectories, based on a variety of mostly geocentric models (Luminet, 2008 presents for example Tycho Brahé's obsession for coming up and imposing his own), had to be adjusted to match their observations. They would admit to the errors in their projection - but how could an observer of their time have quantified the probability of error in the geocentric understanding of the universe that was the basis of their calculations? Although such extreme examples are probably hard to be found today, similar epistemic uncertainty still exists about certain technologies (e.g. should consequences of low level radiation exposure be calculated using a linear or a threshold assumption?) or natural phenomena (e.g. genesis of large earthquakes, climate phenomena including role of the oceans in climate change models).

[86] Probability that a system will perform its intended function over a given period of time. Computed based on the observed frequency of failure events.

[87] They are technical means to make do with sparse data and provide avenues for updating probability distribution function as knowledge of a component's behavior is augmented over time, such as the use of Bayesian statistics. However, the uncertainty about the data used to guide decision-making remains high in these cases, and must be acknowledged adequately.

the magnitude of the outcomes, and on the probabilities of the outcomes. Epistemic, reducible uncertainties, reflecting our incomplete knowledge of fundamental phenomena (e.g. (NEA, 1992) resistance of certain components under accident conditions, poorly understood physical phenomena such as climate change) and uncertainties associated with modeling assumptions (e.g. treatment of human actions or software), are sometimes ignored, tend to be under-reported and are hard to integrate in mathematical models. They are sometimes dealt with by eliciting expert judgment or assuming conservative hypotheses that, in effect and when the risk metric is then used for decision-making, distort the assessment of the merits of possible intervention (Viscusi, 2005, p38). Aleatory, irreducible uncertainties (i.e. randomness, data variability) can be addressed by the use of sophisticated probabilistic model that propagate individual sub-system uncertainties through-out the system based on the modeler's understanding of their interaction with the other sub-systems[88].

Of course, these uncertainties are not specific to probabilistic safety assessments (PSA), and one can legitimately argue that one of the benefits of conducting a PSA is that they can identify areas about which more needs to be known (Apostolakis, 2004). Nonetheless, once PSAs/QRAs/PRAs are done, resisting the temptation to forget about these limitation is difficult when the numbers they allow the analyst to come up with make attractive safety sales pitch in performance oriented evaluation frameworks.

[88] (NEA, 1992) proposes a complementary presentation of the limitations of probabilistic risk assessments due to uncertainties, listing three main sources of uncertainty:

- uncertainties due to a lack of comprehensive data regarding the area under consideration. It is impossible to demonstrate the exhaustiveness of a PSA, even when the scope of the analysis has been extended to as large a number of situations as possible --notably in terms of various reactor operating states and potential initiating events.
- uncertainties regarding data. Such uncertainties concern the reliability data for plant components, the frequency of initiating events, common-mode failures and failures resulting from human actions. The main uncertainties are those relating to the frequency of rare initiating events (for example, the combination of a steam piping break and a steam-generator tube break), as well as data relating to human factors.
- uncertainties associated with modeling assumptions that cannot easily be quantified, such as the resistance of certain components under accident conditions, poorly understood physical phenomena or human actions.

2.3 Issues associated with the definition of acceptability criteria[89]

The belief that safety can only be defined as a relative construct and the persisting paradigm that freedom from harm either cannot be achieved (e.g. Perrow, ICAO quoted below), or would be too costly to achieve and should therefore not be pursued in a world of limited resources[90], justify the creation and use of safety target levels as design goals and their growing embracement by regulatory agencies to whom the appeal of performance based regulation, in contrast with the prescriptive approach that was until recently the norm91, can be at least hypothetically attributed to pressure for them to design stable regulatory frames that are unambiguous, do not stifle innovation and allow for efficient licensing as well as the growing discrepancy between the resources at their disposal and that of the industries that they are tasked with regulating. The attractiveness of quantitative risk estimates, be them derived from probabilistic data or qualitative assessment, allow just that. Their attractiveness lies in the ranking that they allow between initiatives as well as with respect to a reference point meant to represent a "safe" condition (which would more validly be described as "believed to be safe enough").

---

89 (Johansen, 2010) offers a detailed review of risk acceptability criteria and methods used to establish them. She offers three principles for judging risk acceptability and, thereby, creating relevant risk acceptability criteria: equity (all risks must be kept below an upper limit), utility (risk acceptability is the balancing of costs and benefits) and technology (risk must be as low as that of a reference system).

[90] For example, (Jones-Lee, 2009) starts its discussion of frameworks to evaluate whether safety improvements are worth undertaking in a world of finite resources that are also needed for "*other beneficial uses, such as education, environmental protection, crime prevention and so on*" by taking two points as axiomatic. It first assumes that every activity that humans engage in carries some risk of death or injury. Second, "*it will be taken that in most situations safety can be improved, but typically only at a cost*".

This implicitly assumes that the situation one starts with is already optimal in terms of safety-cost trade-off, i.e. that the initial situation corresponds to a point on the Pareto frontier, when any improvement in one dimension (higher safety or lower cost) can only be achieved to the detriment of the other property (lower safety or higher cost): how can that be the case, especially when no methodology exists to guide the designer's choices in trading these attributes one for the other?

Similarly, the RCGuildeines described by (Frank et Jones, 2010) reinforce the notion that "*significant step changes in risk reduction require increasingly greater expenditures of resources*".

[91] For example, (Paté-Cornell, 1996) indicates that bill HR 1022 passed in 1995 by the US House of Representatives called for a balancing of costs and benefits (without specifying how this should be done, in particular without specifying any "value of life") after US federal laws had for many years stipulated that economic considerations were not to be taken into account when setting health and safety regulations (e.g. zero-risk approach of the Delaney amendment to the FD&C Act).

A fundamental concern with acceptability criteria is that the entity performing the trade-off analysis between system performance and safety, and the stakeholder who will suffer from unsafe design are different, creating a strong agency bias.

Moreover, acceptability criteria are so difficult to define that the values they take vary widely across industries and countries. Given that the unit most generally used to compare the costs of providing safety to the benefits this brings is monetary, an important concept in any discussion of the worth of a safety related measure is the value of a statistical human life (VSL): while all efforts should be made to avoid the loss of any given human life, how much is society willing to pay to avert the loss of a statistical human life is measured by the VSL, for whose elicitation several methods are available[92]. If it comes as no surprise that these estimates, by definition dependent on cultural constructs and the social contract that ties a society to the performers of economic activities that it benefits from, differ broadly across geographies[93], the difficulty in coming up with meaningful risk tolerance/acceptance criteria is made obvious by the discrepancies observed in any given country between VSL estimates in different sectors (Is the loss of a life at work really worth 5 times the loss of a life from inadequate health measures - see data for Australia in the figures below? That lost at work or from environmental hazards 3 times that from transportation - see data for the UK in the figures below?), and in any "country/sector" pair between the observed estimates (see Table 4-17 of (ASCC, 2008), as well as figures 4-8, 4-1 and 4-3 reproduced below as Figure 28, Figure 29 and Figure 30)

That risk perception be different from one sector to the next (e.g. dying in a car crash is less dreaded than dying from a plane crash, albeit the consequence is the same - death - and the likelihood of the former is larger than that of the later[94]) is a well-known issue, one much

---

[92] (Abelson, 2008) references to this end three types of methodology: wage-risk studies, revealed preferences studies and stated preferences studies.

[93] (ASCC, 2008), an Australian meta-study of 244 VSL studies, reveals that "countries such as the UK and Japan had average VSL estimates that were three times as high as that for Australia".

[94] (Slovic, 1987) presents a variety of quantitative and qualitative characteristics that have been shown to influence people's perceptions and attitudes towards risk. These are proven to be not only determined by the sort of uni-dimensional statistics generated by QRA, but also by such features as controllability, potential for catastrophic consequences, observability, equitability, reduction potential.

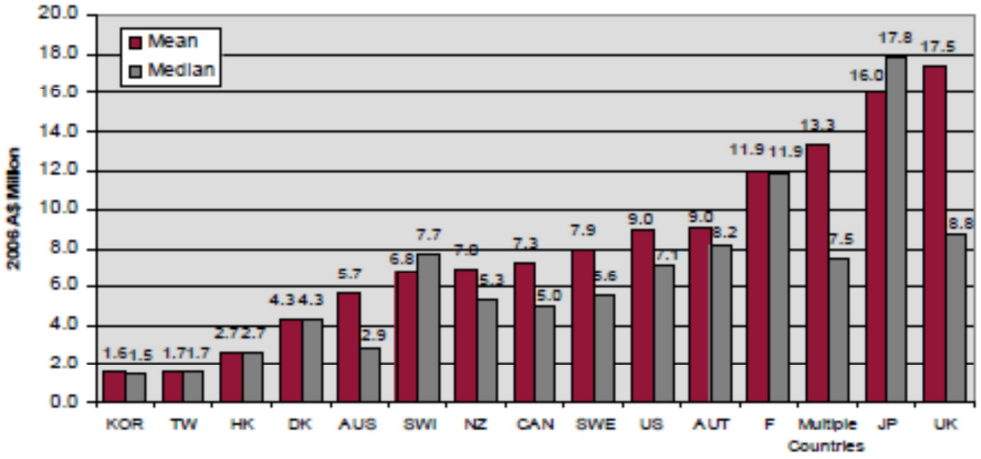


Figure 28 - VSL estimates by country in 2006 A$ millions (ASCC, 2008)

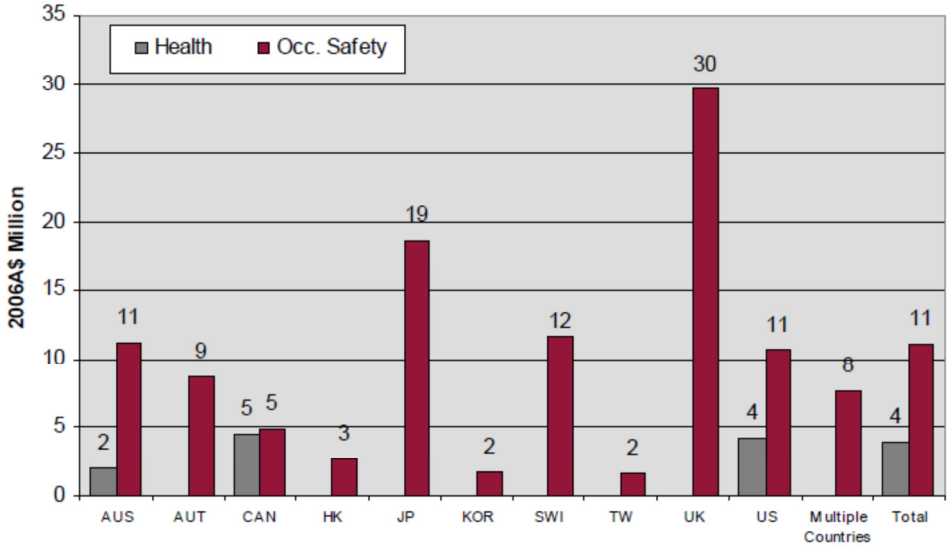


Figure 29 - Range of VSL estimates (means) by country - health and occupational safety in 2006 A$ millions (ASSC, 2008)

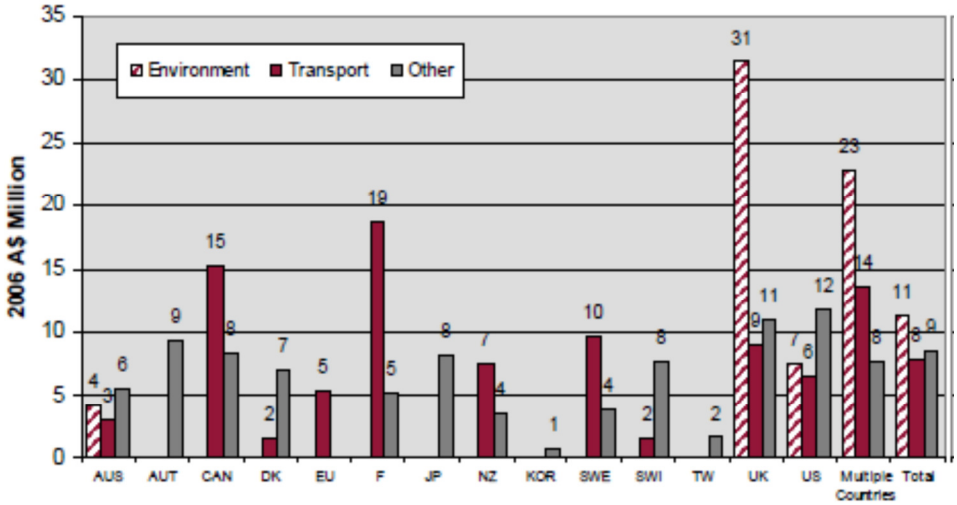


Figure 30 - Range of VSL estimates (means) by study and country – other sectors in 2006 A$ millions (ASSC, 2008)

investigated by psychologists and social scientists and is not the topic of the work here presented albeit still in need of a definitive explanation (Sjöberg, 2000; Nordgren, 2007). Nonetheless, should, following (Slovic, 1987)[95], this perception affect the valuation that ultimately goes into estimating what safety measures should or should not be encouraged by society? In other words, although road transportation risks are less dreaded by individuals than, say, occupational safety risks, should society value a life saved from reducing factory accidents more than one saved from a car accident, accepting to spend more per accrued benefit on air safety than on road safety when both, in the end, are lives that are equally valuable to society itself? Should not society be expected to be more rational in choosing the allocation of its resources than the individual elements it is comprised of?

Finally, (Frank et Jones, 2010) point out that even when valuation is not an issue, criteria that aim for the same safety level (defined as thresholds for the accepted frequency of accidents of given outcomes) may be formulated very differently depending on the practice that is expected they will live within. For example, both the UK and the Netherlands apply the ALARP principle. However, the Dutch acceptance criteria are several orders of magnitude more stringent than the British one (see Figure 31). According to these authors, this situation stems from differences in application of that principle, the premise behind the UK approach being that risks are always being driven down to the extent practicable, with less emphasis on continuous risk reduction in the application of the Netherlands criteria. This example points out that the practice of using these decision criteria does influence their definition - and should be the object of as much scrutiny as the setting of the criteria themselves, as pointed out by (Rimington et al, 2003).

In addition to the possibility that their existence may bias system proponents into constructing cases that a-posteriori justify their designs based on them meeting the acceptance criteria rather than a-priori aiming to create them free from hazards, risk acceptance criteria also run the risk of becoming design targets rather than design constraints. As (Johansen, 2010) puts it, "*interpreting risk and probability as subjective constructs are not seen to threaten the validity of risk acceptance criteria. What may cause a problem, are regulators and practitioners understanding risk acceptance criteria as objective cut-off limits".*

[95] "Attempts to characterize, compare, and regulate risks must be sensitive to this broader conception of risk."

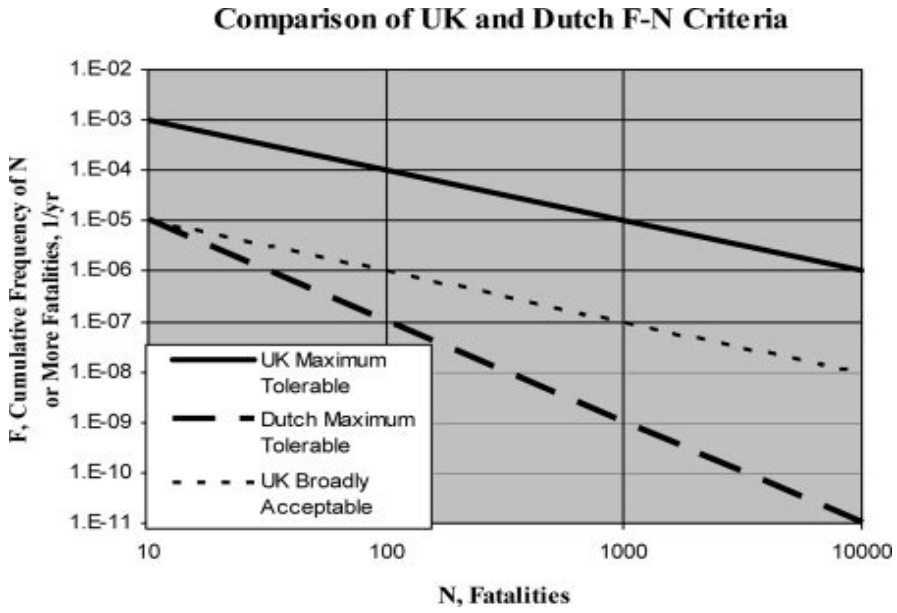


Figure 31 - F-N (frequency vs. consequence) acceptability criteria imposed by the UK and the Netherands regulatory authorities (Frank et Jones, 2010)

Take-away: The "risk-based" safety measure is a quantitative approach to evaluating safety benefits, one that, when complemented with benefit valuation methods, meets the measure and unit conditions necessary for comparison of design alternatives. Its limitations include methodological aspects such as incompleteness (inability to handle several important contributors to safety or lack of thereof such as human behavior during accident conditions, digital software faults, safety culture, design and manufacturing errors (Apostolakis, 2010)), and practical difficulties (esp. in the sourcing and elicitation of uncertainty data and risk acceptability criteria) as well as the inadequate uses it is bound to suffer from.

This is not to say that quantitative risk assessment does not have a role to play in risk-related decision-making[96]. However, just as conformance to a safety related standard should not be understood as evidence that a system is safe but as reducing the likelihood that a system is unsafe

[96] which, guided by quantitative assessments, would consist in the following activities: establish the probability and magnitude of the hazards respecting the inherent scientific uncertainties (a technical process), evaluate the benefits and costs (a social process), and set priorities in such a way that the greatest social benefits are achieved at the lowest cost. (Amendola, 2002) notes an evolution towards making the risk assessment and risk management parts of that process be less strictly separated, with more importance given to participatory processes that allow for increased public participation in the characterization of risks.

(Fowler, 2000, criticizing the idea that conformance to IEC 64508 should be used as proof that a system is safe), QRA must not form the sole basis on which a decision is made, as several of the authors previously quoted insist:

> "*Similarly, the tremendous power of numerical classifactory methods will be dissipated if one insists on claiming for them powers they do not possess and using them for purposes for which they are not intended*". (Williams, 1967 - albeit discussing not QRA but the use of numerical methods in taxonomy, his statement does echo the point here made)
>
> "*The overall conclusion is that acceptance criteria offer sound decision support, but only if authors and users understand the assumptions and limitations of the applied metrics and approaches*" (Johansen, 2010)
>
> "*However, it is easy in pursuing ALARP to overstate the part that numerical estimates can play in decision-making, and therefore they must always be understood as "contributors" rather than as "deciders" in a final judgement and must not in particular be allowed to override considerations of "good" or "best" engineering practice and satisfactory systems of work.* " (Rimington et al, 2003)
>
> "*A risk assessment is intended to provide one of the sources of information for a risk manager or policymaker to use in deciding whether a risk is acceptable. The numerical result from a quantitative risk assessment, per se, does not necessarily make it easier to decide whether a risk is acceptable*" (Apostolakis, 2010, emphasizing the benefits associated with using QRA to inform design and licensing decisions and rejecting the idea of using them as basis for safety decisions)

The UK's participatory approach of the ALARP principle as described in (Rimington et al., 2003) and (HSE, 2012) seems in that regard to have achieved an equilibrium position, one that takes of QRA the insights it can help generate by providing a formal framework in which knowledge about the system's possibly many parts can be aggregated and challenged, but does not take their outcome as sound basis for high-level design decisions (emphasis added):

"The UK approach [...] is to regard QRA outcomes as expressing mainly an artefact – the outcome of applying a particular model, methodology and set of assumptions. These help to

achieve consistency, to rank risks and priorities, and to show where changes on an installation could produce significant risk reduction. However the outcome is in itself no more than an aid to judgment. Partly for that reason[…], tolerability limits are not used as instruments of precise control; the ALARP dynamics are relied on to bring down the risk." (Rimington et al., 2003)

3 References

Abelson P., *Establishing a Monetary Value for Lives Saved: Issues and Controversies,* paper prepared for the conference 'Delivering better quality regulatory proposals through better cost-benefit analysis' hosted by the Australian Office of Best Practice Regulation on 21 November 2007, accessed June 26th, 2012 at http://www.finance.gov.au/obpr/docs/Working-paper-2-Peter-Abelson.pdf

Allen T., Moses J., Hastings D., Lloyd S., Little J., McGowan D., Magee C., Moavenzadeh F., Nightingale D., Roos D., Whitney D., *ESD Terms and Definitions, version 12*, ESD Working Paper, October 2001, available at http://esd.mit.edu/wps/esd-wp-2002-01.pdf

Amendoal A., *Recent paradigms for risk informed decision making*, Safety Science, Vol. 40, Issues 1–4, February–June 2002, Pages 17–30

Apostolakis G., *How Useful Is Quantitative Risk Assessment?*, Risk Analysis, Vol.24, No. 3, 2004

ASCC (Australian Safety and Compensation Council), *The Health of Nations: the Value of a Statistical Life*, July 2008, available at http://www.safeworkaustralia.gov.au/sites/SWA/AboutSafeWorkAustralia/WhatWeDo/Publications/Documents/330/TheHealthOfNations_Value_StatisticalLife_2008_PDF.pdf

Bell R., *Introduction to IEC 61508*, SCS '05, Proceedings of the 10th Australian workshop on Safety critical systems and software - Volume 55, p 3-12, accessed June 25th, 2012 at http://delivery.acm.org/10.1145/1160000/1151817/p3-bell.pdf?ip=18.34.2.17&acc=PUBLIC&CFID=118147094&CFTOKEN=52882792&__acm__=1340656180_0a932bbb934db612d860d8f1505a173b

Bernstein P.L., Against the Gods: the Remarkable Story of Risk, John Viley and Sons, New York, 1996

CCPS, *Guidelines for Developing Quantitative Safety Risk Criteria*, Wiley, ISBN: 978-0-470-26140-8, August 2009

Dekker S., *When human error becomes a crime*, Human Factors and Aerospace Safety, 3(1), 83-92, 2003

Dekker S., *Just Culture: Balancing Safety and Accountability*, Ashgate Publishing Limited, England, 2007

Dekker S., *The criminalization of human error in aviation and healthcare: A review*, Safety Science , Volume 49, Issue 2, February 2011, Pages 121–127

Frank W., Jones D., *Choosing Appropriate Quantitative Safety Risk Criteria: Applications from the New CCPS Guidelines*, AIChE - Process Safety Progress, Volume 29, Issue 4, 2010

Fowler D., Bennett P., *IEC 61508: A Suitable Basis for the Certification of Safety-Critical Transport-Infrastructure Systems ??* , Computer Safety, Reliability and Security: Lecture Notes in Computer Science, 2000, Volume 1943/2000, 250-263, DOI: 10.1007/3-540-40891-6_22

Gigerenzer G., *Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality*, Topics in Cognitive Science 2 (2010) 528-554

Girard R., Le Bouc Emissaire, éd. Grasset, 1982105p, 2010

Hald O., *A. de Moivre: 'De Mensura Sortis" or 'On the Measurement of Chance'*, International Statistical Review 52(3):229-262, 1984

ISA (the Instrumentation Systems, and Automation Society), *Safety Instrumented Functions (SIF) -- Safety Integrity Level (SIL) Evaluation Techniques. Part 1: Introduction*, 2002, ISA-TR84.00.02-2002-Part1, accessed June 26th 2012 at http://www.isa.org/Content/Microsites195/SP5_2,_Binary_Control_Logic_Diagrams_for_Process_Operations/Home193/Committee_Archives126/TR_8402p1.pdf

Johansen I.L., *Foundations and Fallacies of Risk Acceptance Criteria*, ROSS (NTNU) 201001,

Jolliffe G.E., *Considerations of ALARP for Complex Safety Related Systems*, System Safety, 2008 3rd IET International Conference on, 20-22 Oct. 2008

Jones-Lee M., *Safety Expenditure: where should we draw the line?,* in "Safety-Critical Systems: Problems, Process and Practice", edited by Dale C. and Anderson T., Proceedings of the Seventeenth Safety-Critical Systems Symposium, Brighton, UK, 3-5 February 2009

Jonkman S.N., van Gelder P.H.A.J.M., Vrijling J.K, *An Overview of Quantitative Risk Measures for Loss of Life and Economic Damage*, Journal of Hazardous Materials, A99 (2003) 1-30

Kadak A.C., Matsuo T., *The Nuclear Industry's Transition to Risk-Informed Regulation and Operation in the United States*, Reliability Engineering and System Safety 92 (2007) 609–618

Luminet J.P., La discorde céleste - Kepler et le trésor de Tycho Brahé. Les bâtisseurs du ciel, vol II., éd. JC Lattès, Paris, 501pages, Février 2008

Matott L. S., Babendreier J. E., and Purucker S. T., *Evaluating uncertainty in integrated environmental models: A review of concepts and tools*, Water Resources Research, Vol. 45, W06421, 14 pp., 2009, doi:10.1029/2008WR007301.

Net Safety Monitoring Inc., *Safety Integrity Levels (SIC) - IEC 61508/61511*, accessed June 25th, 2012 at http://www.net-safety.com/about/whitepaper/wpt0015.pdf

Nordgren L. Van der Pligt J., van Harreveld F., *Unpacking Perceived Control in Risk Perception: The Mediating Role of Anticipated Regret*, Journal of Behavioral Decision Making, 20: 533–544 (2007)

NRC, *History of NRC's Risk-Informed Regulatory Programs*, 2012, accessed June 26th, 2012 at http://www.nrc.gov/about-nrc/regulatory/risk-informed/history.html

OECD/NEA/Committee on the Safety of Nuclear Installations, *The Use of Quantitative Safety Guidelines in Member Countries, Addendum to CSNI Report N°177 Consideration of Quantitative Safety Guidelines in Member Countries*, NEA/CSNI/R(94)15, June 1994

OECD/NEA, NEA Issue Brief, *An analysis of principal nuclear issues*, N°8, January 1992, accessed June 25th, 2012 at http://www.oecd-nea.org/brief/brief-08.html

Rimington J., McQuaid J., Trbojevic V., *Application of Risk Based Strategies to Workers Health and Safety Protection - UK experience*, May 2003, ISBN 90-5901-275-5, accessed on June 26th, 2012 at http://www.risk-support.co.uk/SZW-published_report.pdf

Perrow C., Normal Accidents: Living with High-Risk Technologies, Basic Books, NY, 1984

Sjöberg L., *Factors in Risk Perception*, Risk Analysis, Vol. 20, No. 1, 2000

Slovic P., *Perception of Risk*, Science, Vol.236, pp280-285, April 17th, 1987

Sussman, J., *Ideas in Complexity in Systems - Twenty Views*, accessed June 22nd 2012 at http://web.mit.edu/esd.83/www/notebook/20ViewsComplexity.PDF.

UK HKE website, *ALARP at a glance*, accessed June 25th, 2012 at http://www.hse.gov.uk/risk/theory/alarpglance.htm

Viscusi W.K., Harrington J.E., Vernon J.M., Economics of Regulation and Antitrust, 4th edition, the MIT Press, 2005

de Weck O., Ross A., Rhodes D., *Investigating Relationships and Semantic Sets amongst System Lifecycle Properties (Ilities)*, Third International Engineering Systems Symposium, CESUN 2022, Delft University of Technology, 18-20 June 2012, available at esd.mit.edu/wps/2012/esd-wp-2012-12.pdf

Woolridge M., *Qualitative Risk Assessment*" in Microbial Risk Analysis of Foods edited by D.W. Schaffner, ASM Press, Washington D.C., 2008