Probabilistic Incorporation of Uncertainties in Radiation Therapy

by

Peter Ferjančič

A dissertation submitted in partial fulfillment
of the requirements for the degree of:

Doctor of Philosophy
(Medical Physics)

At the

University of Wisconsin – Madison

2021

Date of final oral examination: 11/22/2021

The dissertation is approved by the following members of the Final Oral Committee:

John Bayouth – Professor, Human Oncology

Bryan P. Bednarz – Associate Professor, Medical Physics

Robert Jeraj – Professor, Medical Physics

Randall Kimple – Associate Professor, Human Oncology

Jennifer B. Smilowitz – Clinical Professor, Medical Physics

Table of contents

List of Abbreviations and Definitions

- CI: Confidence interval
- CT: Computed tomography
- CTV: Clinical target volume
- DP: Dose painting
- DVH: Dose-volume histogram
- FDG: [^{18}F]Fluoro-2-deoxy-2-d-glucose – measure of glucose metabolism
- FET: O-(2-[^{18}F]Fluoroethyl)-L-tyrosine – measure of amino acid uptake
- GTV: Gross tumor volume
- ICRU: International Commission on Radiation Units
- IM: Internal margin
- ITV: Internal target volume
- IMPT: Intensity modulated proton therapy
- IMRT: Intensity modulated radiotherapy
- MLC: Multi leaf collimator
- NTCP: Normal tissue complication probability
- OAR: Organ at risk
- PET: Positron emission tomography
- PTP: Probabilistic target definition and planning
- PRV: Planning organ at risk volume
- PTV: Planning target volume
- RO: Robust optimization
- RT: Radiotherapy
- SD: Standard deviation

- SM: Setup margin
- SUV: Standardized uptake value (unit: g/mL), the concentration of radioactive decays in a region of interest normalized to injected activity and patient body mass.
- TCP: Tumor control probability
- TLM: Tumor likelihood map
- TM: Target map
- gTLM: Gross tumor likelihood map
- miTLM: Microscopic tumor likelihood map
- VMAT: Volumetric-modulated arc therapy
- WiscPlan: In house TomoTherapy treatment planning software

List of Figures

List of Tables

Acknowledgements

This dissertation would not have been possible without the help and support of the many great people I've had the privilege and pleasure of working with over my many years.

I would first like to express my gratitude the many teachers, professors and mentors who helped me become who I am today, especially all the faculty and staff at the University of Wisconsin, Madison Department of Medical Physics, who always put great effort into creating a friendly and welcoming environment. When I left my home behind, you made me feel welcome and like I belong.

I would like to thank all the members of the Jeraj group who preceded me, some of who are: Stephanie Harmon, Tim Perk, Matt Scarpelli, Christie Lin, Allison Roth, Amy Weisman, Mauro Namías, Surendra Prajapati, Tyler Bradshaw, Damijan Valentinuzzi, and Song Chen. You not only helped me with my work, but you taught me how to *be* a scientist.

I would also like to thank the lab colleagues still in the group: Daniel Huff, Victor Santoro Fernandez, Alison Deatsch and Brayden Schott, and the international members, including: Urban Simončič, Andrej Studen, Maruša Turk, Ana Marin, Jošt Stergar, Matija Milanič, Rok Dolenec, Nina Verdel, Žan Klaneček, Jošt Stergar, Luka Rogelj, Doroteja Novak, Gašper Razdevšek, Matej Perovnik and Luciano Rivetti. You provided advice, help and constructive criticism with this work, and helped make it what it is today.

I would also like to extend heartfelt thanks to the physicians and scientists who have patiently collaborated and provided guidance, some of who are: Glenn Liu, Scott Perlman, Guy Starbuck, Vincent Streif, Jens Eickhoff, Martin Ebert, Ros Francis and Stefan ten Eiklelder. Special thanks goes to the members of the ICRU committee on probabilistic planning and dose painting: Vincent Grégoire, Rock Mackie, Edmond Sterpin, Uulke van der Heide, Daniela Thorwarth, Cynthia

Ménard, Brian O'Sullivan, and Thomas Bortfeld, who provided expert guidance like few graduate students can ever hope for.

None of this would be possible without my advisor and mentor Robert Jeraj, who, ten years ago, first introduced medical physics to me. You provided guidance and support at every step of the way, not just in my thesis work, but in life.

I especially thank my friends and family, but there are no words that can capture the eternal gratitude that I feel for my parents. Zares iskrena hvala za vso ljubezen, skrb, podporo in dobre nasvete, ki sta mi jih nudila ko sem prosil zanje, in ko sem jih potreboval. Za vedno vaju imam rad, in nikoli mi ne bi uspelo ustvariti kar sem brez vaju.

Last, but not least, I am forever thankful to my wife Silvia Villalobos. Your love and support have made the most difficult parts of my life bearable, and the best parts of my life blissful.

Abstract

The main challenge of radiation therapy is delivering sufficient radiation dose to target, while minimizing irradiation of healthy tissues in the presence of uncertainties. Current radiotherapy guidelines, based on ICRU reports, recommend accounting for these uncertainties using target volume expansion by margins. While this approach is historically sensible, it oversimplifies much of the complexity exhibited by tumors and limits more complex dose planning made possible by the advancements in technology over the last decades.

In this thesis we present an alternative approach – a probabilistic incorporation of uncertainties in radiotherapy. In this approach that we call “probabilistic target definition and planning” (PTP), uncertainties are not accounted for using target expansion by volume, as is done in the classical GTV-CTV-PTV approach. Rather, probabilistic target volumes are used to capture microscopic disease presence uncertainties, and robust optimization is used to account for treatment delivery uncertainties.

To evaluate the viability of this approach, we implemented PTP in the in-house TomoTherapy treatment planning system WiscPlan and applied it on realistic examples to evaluate its feasibility. We performed example quantification of imaging uncertainties using test-retest repeatability analysis on a dataset of patients with glioblastoma imaged with FET PET/CT, used it to derive voxel-level target maps and directly implement them in the PTP treatment optimization.

To evaluate the ability of the PTP approach to optimize non-uniform dose prescriptions, we expanded the analysis to a dataset of patients with prostate cancer imaged with multiparametric MRI, where we performed a quantitative comparison between classical and PTP planning. Finally, we applied the method to different disease sites, each faced with its own challenges, and evaluated the performance of their dose plans.

PTP achieved comparable performance to the classical approach, but with significant improvements observed for some treatment sites. PTP also allows incorporation of probabilistic normal tissue descriptions and dose-painting-by-numbers optimizations under uncertainty.

The work presented in this thesis presents a novel treatment planning approach, which endeavors to be an alternative, more flexible method to optimize external beam photon dose plans by removing the limitation of margins and by directly implementing patient's disease information in treatment plan optimization.

1 Summary

The current clinical paradigm of radiation therapy (RT) planning was first defined by the International Commission on Radiation Units (ICRU) Report 50 [1], which established precise terminology to describe different descriptions of tumor presence with binary tumor volumes while accounting for uncertainties using Gross Tumor Volume (GTV), Clinical Target Volume (CTV), and Planning Target Volume (PTV). Newer ICRU reports 62 and 83 expand on these concepts, but ultimately maintain the same central paradigm of accounting for uncertainties using target expansion by margins [2], [3]. The first volume, GTV, describes the area where tumor presence is palpable or visible, for example through Positron Emission Tomography (PET) imaging. It can include just the primary tumor, or it can also include distant metastases or lymph nodes. Expanding the GTV by some margin to account for suspected microscopic infiltration of the tumor cells into nearby tissues gives the CTV. Expanding the CTV by another margin to account for uncertainties in dose delivery results in the PTV.

Although expanding volumes by margins is a pragmatic and historically sensible approach, it is an oversimplification that poorly captures the underlying uncertainties. First, delineating the volumes can vary considerably depending on the method used or the person segmenting [4], [5]. Secondly, the infiltration of malignant cells into nearby tissues has been shown to be a stochastic process describable with a continuous probabilistic function, rather than a binary delineation [6]. Thirdly, whether the CTV receives the prescribed dose depends on more than just geometric margins - in reality, the dose distributions are not equally conformal on all sides of the CTV, especially in cases where tumor coverage needs to be balanced against organ at risk (OAR) sparing [7]. Lastly, the current approach largely assumes uniform dose distribution to target – a concept that is intrinsically incompatible with methods of non-uniform dose prescription such as dose painting [8], [9].

The main objective of this work was to **develop, implement, and evaluate a probabilistic target definition and planning (PTP) approach, which accounts for uncertainties in radiation therapy planning using non-binary tumor maps and robust optimization.**

In this thesis we addressed three specific aims, summarized as follows:

Specific Aim 1: To assess feasibility of probabilistic target definition and planning (PTP)

- SA 1a: To incorporate imaging uncertainties in probabilistic tumor descriptions
- SA 1b: To establish clinical feasibility of dose plans created using PTP approach
- SA 1c: To evaluate back-compatibility between PTP and classical treatment planning

Specific Aim 2: To evaluate performance of the new treatment planning approach

- SA 2a: To create example PTP dose plans and compare them to classical dose plans.
- SA 2b: To establish dose reporting metrics for probabilistic planning

Specific Aim 3: To explore opportunities of PTP

- SA 3a: To evaluate PTP performance in highly complex geometries
- SA 3b: To demonstrate PTP for probabilistic OAR sparing
- SA 3c: To demonstrate PTP for dose painting

The first specific aim was to assess the feasibility of the proposed PTP approach. To demonstrate the process starting from the imaging data, rather than literature derived uncertainties, we first performed a quantification of uncertainties using test-retest repeatability analysis (SA 1a). We employed an example dataset of eight patients with glioblastoma imaged with O-(2-[^{18}F]Fluoroethyl)-l-tyrosine (FET) PET/CT. We created maps of gross and microscopic disease presence, which were then directly incorporated into the optimization and dose plan calculation (SA1b). As the final sub-aim, we evaluated the feasibility and back-compatibility of the PTP approach (SA 1c) by comparing classically and PTP optimized dose plans.

The second specific aim was evaluation of performance of the PTP approach on a more complex disease site with higher anatomic variability and quantitatively compare the performance of the

PTP and classical planning approaches. To achieve this aim, we used a cohort of patients with prostate cancer who received multiparametric magnetic resonance imaging (MP-MRI) imaging to generate tumor likelihood maps. We calculated PTP and classically optimized dose plans for both homogeneous dose prescriptions and plans with integrated dose boosts, and we quantitatively compared them using a variety of dose metrics.

The third specific aim was to explore the opportunity of the proposed PTP approach and its non-reliance on the margin-based accounting for uncertainties. To do so, we implemented and validated the PTP approach in various sites, each facing their own unique challenges. The cases explored in this work were highly complex geometries in cases of patients with head and neck cancers (SA 3a), patients with whole-brain irradiation with hippocampus sparing (SA 3b), and patients with PET imaging-based dose-painting-by-numbers prescription (SA 3c). While this list of cases is by no means exhaustive of the possible applications of the PTP approach, it does present some examples of the different implementations of the approach.

2 Background

In this section we briefly review some key topics relevant to this work. In section 2.1 we discuss the uncertainties in radiation therapy. In section 2.2 we talk about treatment planning, and in 2.3, we review existing ICRU reports and their recommendations relating to managing uncertainties.

2.1 Uncertainties in radiation therapy

The treatment process of external beam radiation therapy (EBRT) inherently introduces uncertainties that arise during treatment preparation and execution and limit the accuracy of the delivery. If unaccounted for, they can lead to severe degradation of the quality of the delivered treatment as compared to the planned [7].

The main sources of uncertainties are:

1. Imaging uncertainties
2. Microscopic infiltration
3. Biological uncertainties
4. Residual positioning uncertainties
5. Anatomical changes uncertainties
6. Physics uncertainties

Uncertainties are usually classified into systematic and random. Systematic uncertainties are those that are persistent throughout multiple separate deliveries (fractions) of a treatment. Random errors are those that can change between different deliveries, such as setup errors, organ motion and anatomical changes.

2.1.1 Imaging uncertainties

Target volume definition depends on the interpretation of obtained medical imaging. Previous works have shown that even when segmenting identical images, significant variation is observed in volume delineation of GTV. This variability can be caused by intraobserver variation or "observer noise" i.e. when the same observer segments the same image and comes to slightly different results [10], [11], or it can be caused by interobserver variability, i.e. variability between different people [4], [12]. Automated segmentation methods, where a computer algorithm segments the disease rather than a trained human, can remove the observer variability, but they are subject to their own segmentation uncertainties [13], [14]. A thorough evaluation of the latter is performed in the AAPM TG report 211 [15].

This variation is caused by current medical imaging modalities not imaging cancer cells directly, but instead visualizing disease surrogates [16]. These surrogates can be anatomical anomalies, such as changes in tissue density or shape that can be imaged with computed tomography (CT) and magnetic resonance imaging (MRI), or they can be functional changes, such as increased metabolism or proliferation that can be imaged with positron emission tomography (PET). Medical imaging is also limited by spatial resolution – this is usually a fraction of a millimeter for CT imaging, and 1-4 mm for MRI and PET imaging. In addition to spatial limitations, tumors have also been shown to exhibit temporal variability, meaning that the same disease imaged at two different time points can present differently [17]. The exact variation of this temporal variability depends on the disease and the type of medical imaging used, but can be evaluated in each scenario by test-retest imaging analysis that provide limits of repeatability [18].

2.1.2 Microscopic infiltration uncertainties

Target definition in RT is not affected only by the imaging uncertainties, but also by limited sensitivity within tissues where the concentration of diseased cells is too low to cause noticeable

anatomical or functional changes. Malignant, or cancerous, tumors are by definition made of cancer cells that can invade nearby tissues. This invasion can be local in the neighboring tissues or nearby lymph nodes, or it can be distant, to other parts of the body in a process called metastasis.

While metastatic growths represent a whole different challenge in treatment of cancer, local spread and invasion of cancer cells into neighboring tissues can be accounted for in the planning of radiation therapy. The ICRU reports recommend expanding the visible GTV by a margin, to obtain the CTV and account for microscopic infiltration. The exact size of the margin, with consideration of nearby anatomical structures, depends on the cancer type and location. While there are some guidelines, the delineation of CTV boundaries currently lacks a clear “gold standard” [19], although there are published models that try to address this issue for some disease sites [20], [21].

Knowledge of the exact presence of microscopic disease is challenging, but can be estimated with populational histological analysis, where tissue samples are resected from patients and examined under a microscope to identify cancer infiltration on a cellular level. Several different studies have been performed in various disease sites, quantifying probabilities of cancer cell infiltration in tissue based on a range from visible disease [22], [23]. Three-dimensional histological studies have been made possible over the last decade, where an entire volume of tissue were analyzed, rather than two-dimensional slices [24]. Such histological studies provide the most accurate ground truth but are very labor and time consuming. They can provide the most accurate data on microscopic infiltration due to their ability to identify individual cancer cells in a large volume, but their need for a resected tissue sample makes them viable for studies investigating population statistics and less useful for RT treatment planning.

2.1.3 Residual positioning uncertainties

After the planning volumes are identified, the next major category of uncertainties is positioning uncertainties. These describe the accuracy of the patient position compared to the planning position. They can be split into a systematic component, which is constant over all treatment fractions, and a random component which varies between fractions. Systematic errors are typically introduced during treatment planning, for example if the patient in the planning CT is positioned differently from their normal treatment position. Random errors are related to daily patient anatomy, described in greater detail in the next section, and positioning variations between fractions.

In practice, good clinical protocols and clinical adherence to them, the availability of daily, real-time imaging and positioning equipment allow good accounting for most of these uncertainties. However, even in the best scenarios such protocols and equipment cannot completely account for all positioning uncertainties – the part that remains is referred to as *residual positioning uncertainties*.

These residual uncertainties depend strongly on the disease site and positioning equipment available. For example, a typical prostate treatment can have positioning uncertainties of 2-6 mm [25], [26], but other disease sites, e.g. brain tumors, which are well encased in the rigid skull, can measure less than 1 mm for some Gamma Knife treatments [27]. Due to their dependence on availability of positioning equipment and clinical protocols, these uncertainties can vary considerably between different treatment sites. On the other hand, due to their relative ease of measurement, there is a large body of literature reporting these uncertainties in various scenarios that can provide an estimate of typical uncertainties [28]–[31].

2.1.4 Uncertainties of anatomical changes

Uncertainties of anatomical changes, unlike residual positioning uncertainties, are not caused by inaccuracies in patient positioning, but rather changes within patients themselves that would be present even if patients were placed on the treatment table perfectly. These uncertainties can be caused by organ motion, such as breathing, heartbeat or peristalsis, by filling of cavities at or near the disease sites such as bladder and rectum, or by changes in body weight that patients often undergo during treatment.

These uncertainties depend on the anatomy and disease site and can be considerable. For example, in patients with lung cancer, tumors can move during the breathing cycle upwards of 1 cm within a breathing cycle [32]. Motion management techniques such as breath holds or respiratory gating help with reducing these uncertainties, but only to some extent. For patients with prostate cancer, bladder and rectum filling can move volumes by 2 mm or more, resulting in significantly different dose profiles to those organs [33].

2.1.5 Biological response uncertainties

In cancer treatment, response is often sought through radiation induced cell death of the cancerous tissues. Cell survival and cell death can mean different things in different contexts, but in oncology, this is often defined with the loss of capacity for sustained proliferation [34]. The radiation dose required to achieve this loss of proliferative capacity varies on a range of factors, such as cell cycle stage, clonal density, intrinsic radiosensitivity of different tissue types and immune response [35]–[39]. As a result, the response of any cell group to radiation – even cells within a single tumor – is highly variable.

A frequently used way of evaluating the response of cells under different conditions to radiation is using the in-vitro survival curves. In these studies, a specimen is taken from a tumor or normal tissues and prepared in a culture dish. Cells that retain their proliferative abilities will grow into

colonies visible with the naked eye, which allows for easy comparison between different cells or conditions. Exposing such specimens to different radiation doses allows for an estimate of radiosensitivity of different tissues and the creation of cell survival curves, some examples of which for various tumors of human origin (and one murine tumor) can be seen in figure 1.

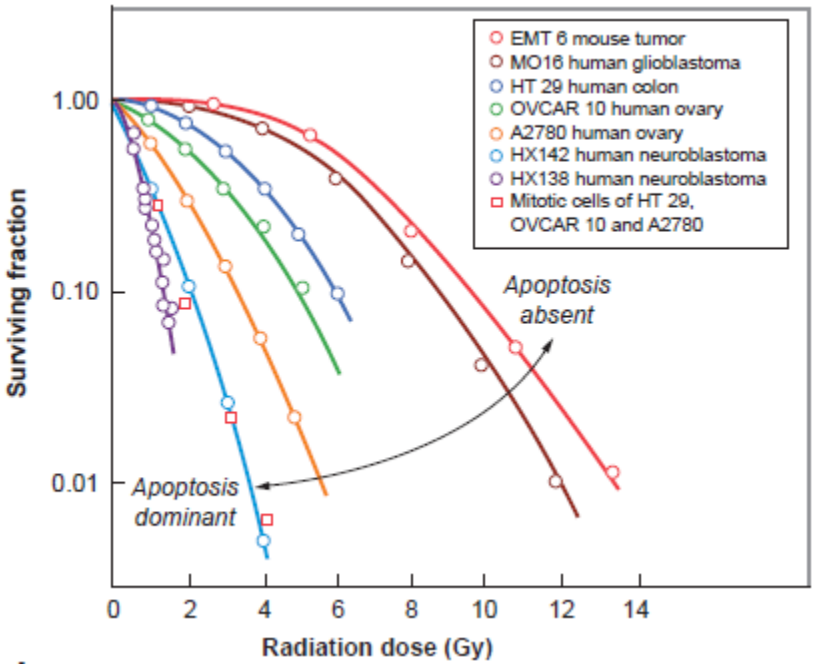


Figure 1: Compilation of survival curves of several cell lines of human (and one murine) cancers. Note the wide range of radiosensitivity exhibited by different cancer cell types. Adapted from [34].

Due to the general shape of the survival curves, the “linear-quadratic model” is often used to describe the sensitivity of different tissues to radiation. However, when evaluating response in-vivo, the effect of radiation on cells becomes even more complicated. Multiple studies have shown the existence of inter- and intra-tumoral heterogeneity caused by different levels of hypoxia, perfusion and even clonogenic differences, which can affect the efficacy of radiation therapy [40]–[46]. Some of these principles are captured in the so-called “four R’s of radiation therapy”: *repair*

of sub-lethal injury in normal and diseased cells, *reoxygenation* of the tumor, *redistribution* of cells throughout the division cycle and *regeneration* of surviving normal and diseased cells.

Additional effects, such as the structure of the intracellular matrix and the immunocompetence of the host, also affect the overall tumor response. Of particular interest are the bystander/abscopal effects [47]. The bystander effect refers to the induction of biologic effects in cells that are not directly traversed by charged particles but are located close to cells that are. Studies using micro-beam irradiation have shown that nearby cells not directly targeted by radiation can be killed, presumably due to the release of cytotoxic molecules into the medium [34], [48], [49]. While these effects have not yet been ubiquitously embraced, their mounting evidence provides ample motivation to consider response models outside the traditional standard.

At this point it should be apparent that tumor environments are highly complex and warrant a more detailed examination. Some imaging techniques have been developed, such as various functional MRI sequences and PET imaging tracers that allow exploration of intra-tumoral heterogeneity [50]. Some such as [^{18}F]Fluoro-2-deoxy-2-d-glucose (FDG) or O-(2-[^{18}F]Fluoroethyl)-L-tyrosine (FET) allow for differentiation of different levels of tumor activity, while others, for example ^{61}Cu(II)-diacetyl-bis(N4-methylthiosemicarbazone) (Cu-ATSM), allow imaging of tumor hypoxia.

The ability to visualize different sub-volumes of tumors has in turn led to studies exploring heterogeneous dose prescriptions such as dose painting, where these sub-volumes are preferentially targeted with higher doses [51], [52]. Such approaches are not yet widely established, in part due to the lack of validation in clinical studies and uncertainties regarding appropriate prescription [53]. Consequently, the ideal dose distribution that RT should aim to achieve has not yet been sufficiently explored in clinical trials.

2.1.6 Physics uncertainties

Finally, there is uncertainty in the delivered dose distribution that arises from approximations and inaccuracies used in treatment planning and beam modeling, such as using depth-independent kernels [46], [54]. These uncertainties were considerable in the past, when computational power was limited and more simplifications were necessary to ensure tractability, but modern computers and commercial planning systems can achieve very high accuracies. Monte Carlo simulations in particular are the widely accepted gold standard, and they can achieve highly accurate dose calculations. Improvements in algorithms resulted in much faster calculations while maintaining very accurate doses. For example, Acuros dose calculation algorithm (Varian Medical Systems, Palo Alto, CA) achieves dose accuracy within 2% even in tissues with varying density such as lung [55]. Proton and particle therapies are subject to somewhat different uncertainties, including particle range uncertainties and a different formulation of positioning or depth uncertainties, but they are still modeled to a high precision [56]. For today's treatment planning, AAPM TG 142 specifies that the delivered dose should have at least a 95% probability to be within 2% of the planned dose for all dose voxels [57], and studies have shown that dose inaccuracy of <1% is often achieved [58]. As these uncertainties are comparatively small when compared to previous categories, an in-depth analysis of their effects was not considered for the work we present in this thesis.

2.2 Treatment planning

In the early days of radiation therapy planning, dose plans were created using the so-called "forward planning", where planning starts with the desired beam number, direction and dose that should be delivered to a volume and then calculates forward what the final dose within the geometry will be and whether it meets all the requirements. With the development of intensity modulated radiation therapy (IMRT), a process of "inverse treatment planning" became prevalent.

In inverse planning, the desired solution is known, and then the algorithm works backward to find the beam parameters needed to achieve it, or a solution sufficiently close to it. The inverse-planning process works iteratively to determine optimal beam shapes and fluence patterns to achieve a desired dose distribution. A schematic representation of this process, as shown in ICRU report 83 can be seen in figure 2.

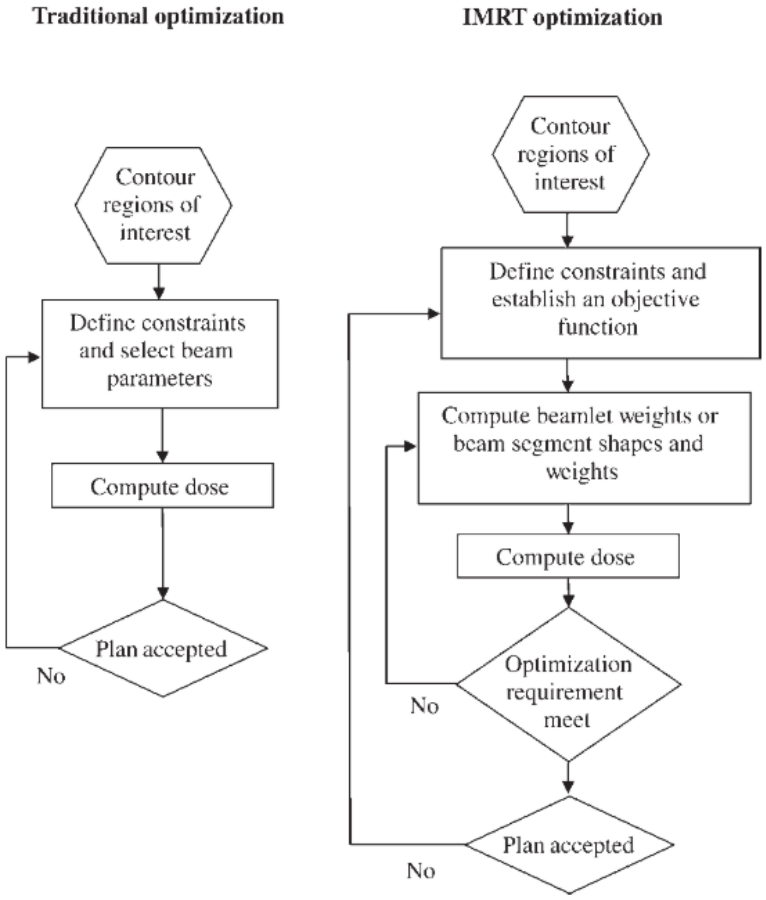


Figure 2: Comparison between traditional (left) and IMRT (right) optimization processes. Adapted from ICRU report 83 [3].

Example optimization goals could be minimum or maximum doses to the target volumes, maximum doses to organs at risk (OAR), or various dose-volume specifications (e.g. no more

than 20% of the total lung volume should receive more than 20 Gy). These can be combined into a mathematical objective function that calculates a single value representing the “goodness” of the dose plan. The optimization procedure iteratively searches for the solution minimizing the cost or maximizing the “goodness” of the dose plan. Due to the iterative nature of this optimization approach, there is no guarantee that a globally optimal solution is obtained. The person performing the dose planning is expected to adjust the weights – the values assigned to different competing objectives - to achieve a desired compromise between different goals.

As it has been shown that the optimal radiation pattern from any single direction is typically non-uniform [40], [59]–[61], the beam is often split into smaller sub-units called “beamlets”, which can be independently weighted. This is the defining characteristic of IMRT. This allows better conformity while still maintaining target volume dose uniformity. This also opened the way for non-uniform dose distributions, such as simultaneous integrated dose boost techniques [62], [63].

2.2.1 Robust Optimization (RO)

Optimization of a problem with uncertainty in input parameters can be performed in multiple ways. In Classical RT optimization, a dose distribution is sought where the therapeutic dose is delivered to a volume that is larger than the expected target volume, an approach that tends to be over-conservative and necessarily increases the radiation exposure of healthy tissue and organs-at-risk [64]. An alternative approach is robust optimization (RO), a field of optimization theory that deals with optimizing problems in the presence of input parameter variability. RO searches for solutions over a range of “scenarios” – realizations of the problem with different starting parameters and evaluating the performance of a solution over all scenarios. While this can provide a solution that is less-than-optimal in the idealized scenario, such solutions perform consistently well even when subject to uncertainties [65].

The origins of RO date back to the 1950s and the use of worst-case analysis and Wald's *maximin* theorem, where decisions are ranked based on their worst-case outcomes. In other words, the optimal solution is the one with "the best worst-case-outcome". In mathematical terms, this is described as

$$f(x) = \max\left\{\sum_{j=1}^{N}\left(w_j \cdot f_j\big(q, d(x,s)\big)\right)\right\} \tag{2.1}$$

For:

$$v \coloneqq \max(\min f(d,s)) \tag{2.1}$$

Where $f(x)$ is the objective function for a set of values x to be optimized, w_j is the weight assigned to optimization objective j and f_j is the penalty associated with said objective.

Robust optimization became a discipline of its own in the 1970s, with developments in other scientific and technological fields. Most notably, RO is being used extensively in intensity modulated proton therapy (IMPT), where the dose cloud approximation does not perform well due to the Bragg peak shape of dose deposition [65]. As small variations in initial parameters, such as stopping power, can lead to much larger changes in dose deposition when compared to photon beams, robust and probabilistic optimization methods have been developed that directly incorporate motion and uncertainty into treatment plan optimization. Some commercial planning systems have already implemented robust planning methods for IMPT, making them available for clinical use.

2.2.2 WiscPlan

While some commercial planning systems (TPS) have started implementing RO functionality, open-source TPS offer better access to their low-level algorithms and better opportunities for modification and expansion. For this reason, we selected our in-house developed TomoTherapy TPS WiscPlan as the framework for our probabilistic target definition and planning [66].

WiscPlan can use various formats of patient's CT imaging and contours to generate patient geometries. The user can specify the target region of interest (ROI) and perform the x-ray dose calculation using a convolution-superposition (CS) algorithm. TomoTherapy is known for its helical delivery of radiation, where the x-ray source rotates around the patient, which allows for delivery of highly complex plans that would be difficult to achieve on a different modality [67], [68]. For optimization purposes, the doses are calculated for 51 discrete angles around the central axis per gantry rotation, corresponding to about 7° of movement of the source. In addition, dose is calculated independently for each beamlet – a part of the beam corresponding to one of the 64 binary multileaf collimators (MLC) that shape the beam profile.

As the MLCs are binary and can open and close very quickly, it is possible to modulate the beamlet doses independently. For practical purposes, some limits on this modulation are usually set, in terms of a *modulation factor*, which is the ratio between the maximum and the average dose in all beamlets.

2.3 Handling of planning uncertainties today

Delivering radiation therapy in the presence of uncertainties is a fundamental challenge in clinical RT medical physics, and it is no surprise that hundreds of publications exist on the topic. To get a summarized understanding of the history and current guidelines, reports of the International

Commission of Radiation Units (ICRU) serve as an excellent review and backbone for most such publications, and we will start our review with these reports.

2.3.1 ICRU 29

The first ICRU report to tackle the topic of target definition and reporting was *ICRU report 29: Dose specification for reporting external beam therapy in photons and electrons* [69]. While it was primarily concerned with good reporting practices, it did include a definition of Target Volume, which was a volume containing those tissues that are to be irradiated to a specified dose. It defined that target volume "*… consists of the demonstrated tumor(s), if present, and any other tissue with presumed tumor*". On top of that, the Target Volume had to account for expected movement, variations in shape and size and inaccuracies or variation in treatment set-up during treatment, effectively defining what we know today as the PTV.

ICRU report 29 also defined "Treatment Volume", which is the entire volume receiving at least the same absorbed dose as any part of the target volume, and Irradiated Volume, which receives a "significant" dose in relation to tissue tolerance, such as 50% of the specified target dose. It also introduced "Organs at Risk" as radiosensitive organs in or near the target volume, whose presence influence treatment planning or prescribed dose.

As the majority of the 29-page report was dedicated to reporting the therapy, many questions remained unaddressed. Most relevantly, it failed to provide any guidelines or definition on the margins, or other methods of accounting for uncertainties, and left the question open to the reader. Despite this, it still emphasized the importance of treating invasive disease and the need to deliver sufficient radiation to the target under the presence of uncertainties, which paved the way for the more specific target definition volumes presented in ICRU report 50 which are still in use today.

2.3.2 ICRU 50

An update to ICRU 29 came when *ICRU Report 50: Prescribing, Recording and Reporting Photon Beam Therapy* was published. It primarily aimed to create guidelines that would allow radiation oncologists and physicists to create reports that would be consistent and contain sufficient information to be clearly understood by their colleagues in other departments or institutions. It stressed the importance of consistent parametrization, particularly in target volumes, to be used for three purposes: prescription, recording and reporting. This way, consistent treatment policies and improvisation could be done in the light of experience and facilitate communication and comparisons between institutions.

Most notably, ICRU report 50 introduced several distinct steps and volumes that may be defined, and it introduced the GTV, CTV and PTV volumes. A simple summary of these volumes can be seen in figure 3.

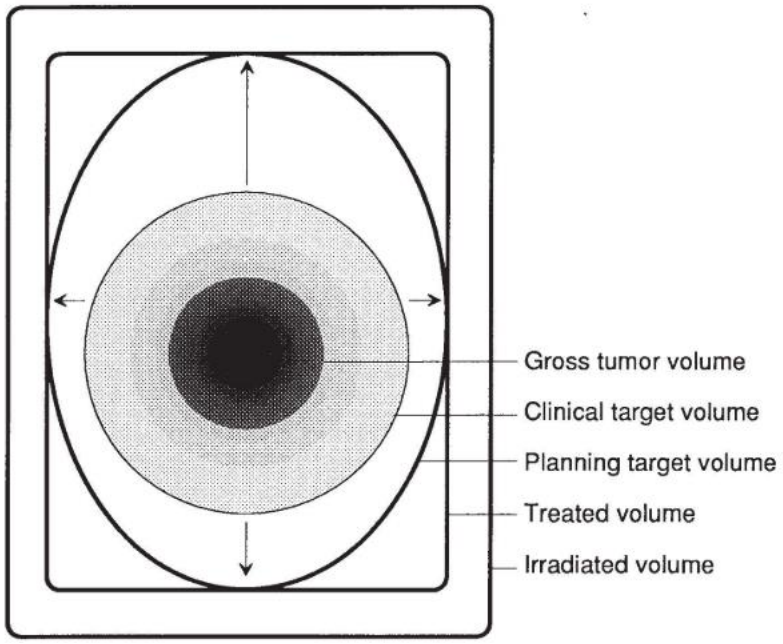


Figure 3: Schematic illustration of the different volumes, as per ICRU Report 50. Adapted from [1].

ICRU report 50 defined the GTV as the palpable, or visible, extent of the malignant disease, where tumor cell density is largest. It is also the volume where adequate dose must be delivered to achieve local tumor control. The GTV may be determined using different methods, such as clinical examination or various imaging techniques. The report recognized that this may lead to significantly different sizes and shapes, so therapists should indicate which methods have been used for definition and evaluation of the GTV, but the report did not elaborate on the consequences of such differences.

Next, the report introduced the CTV, which is defined as the volume that includes subclinical involvement of the disease, such as individual malignant cells, small clusters or microinfiltrations which cannot be detected by medical imaging or clinical examination. The CTV is defined as the GTV with an added margin, or in cases where the GTV has been removed prior to therapy, it could be a standalone volume. The CTV could also include volumes that do not contain demonstrable tumor but with suspected malignant cells such as regional lymph nodes, potentially resulting in multiple different CTV volumes. While margins were recommended as the way to account for microscopic infiltration, the report acknowledged that the volume surrounding the GTV often has a high tumor cell density close to the visible disease with decreasing density towards the periphery. The report mentioned the possibility of using “boost” therapy which allows treating the same volume with two different doses but did not otherwise expand on the consequences of variant cell density on dose planning or optimization.

ICRU Report 50 set to define the PTV as “*a geometrical concept defined to select appropriate beam sizes and arrangements, taking into consideration the net effect of all the possible geometrical variations in order to ensure that the prescribed dose is actually absorbed in the CTV*”. The PTV is used for dose planning and specification and can be related to the machine

coordinate system. It is also worth noting that the definition of the PTV is identical to the previous (ICRU 29) definition of “Target Volume”. Another key difference between CTV and PTV is that the PTV is more of a utility structure that is helpful in planning, but does not exactly represent defined tissues or tissue borders and may not receive the actually planned dose distribution – particularly around the borders. The optimal PTV may depend on beam selection, anatomical location, patient immobilization devices, target motion effects etc. It is also possible that a single CTV is assigned multiple PTVs, such as for cases with very large CTVs where multiple fields have to be combined to ensure coverage.

ICRU report 50 also maintained the definitions of “Treated Volume” as volume enclosed by an isodose surface, appropriate to achieve the purpose of treatment, and “Irradiated Volume”, which is the volume receiving dose considered significant in relation to normal tissue tolerance.

A new addition in ICRU Report 50 was also the guideline for homogeneous dose prescription, which should be kept within +7% and -5% of prescribed dose, or else confirmation by the radiation oncologist is recommended. However, the authors also recognized that this level of homogeneity cannot always be achieved, and that sometimes higher doses can be found within the GTV where they might even be advantageous. Higher heterogeneity is also accepted for palliative treatments and for subclinical disease. While the use of maximum dose reporting was suggested, the report also expanded on the concept that a clinically meaningful volume must be dosed as defined if its minimum diameter exceeds 15 mm, except in smaller organs such as eyes or optical nerves.

ICRU report 50 introduced several key concepts that still today form the basis of any radiation therapy treatment. While some aspects of the report are somewhat dated now, it is still a remarkably relevant document, with the many concepts that it lays out still forming the backbone of radiation therapy treatments around the world today. This is in part due to the commendable foresight of the authors who, where the technology of 1993 was not yet fully developed,

commented on the potential the future might bring, and kept the research paths open. One crucial path was the publication of the ICRU report 62 that further expanded on the introduced concepts.

2.3.3 ICRU 62

Six years after the previous report, *ICRU Report 62: Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU Report 50)* was published. The aim of this publication was to more accurately formulate the definitions and concepts, as well as update the report for the considerable technical and clinical advancements. While more detail was provided across many levels, the overall intention of guiding physicists and radiation oncologists to complete comprehensive reports of radiation therapy treatments remained. It updateed and expanded the target and OAR volume definitions, added the benefit versus complication question, split sources of uncertainty into random and systemic, and introduced the conformity index. It also expanded on reference points, coordinate systems, graphics and other elements, which are less relevant to the work we describe in this thesis and will not be discussed in further detail.

The definitions of GTV *"… the gross demonstrable extent and location of the malignant growth*", and CTV *"… volume that contains a demonstrable GTV and/or subclinical malignant disease that must be eliminated."* remains similar to definitions in ICRU report 50.

While this report further detailed the probabilistic nature of microscopic infiltration, such as shown in figure 4, and admitted considerable interest in the public for a more probabilistic approach, it ultimately used these descriptions as steppingstones to obtain appropriate CTV margins and failed to expand on truly probabilistic approaches to therapy.

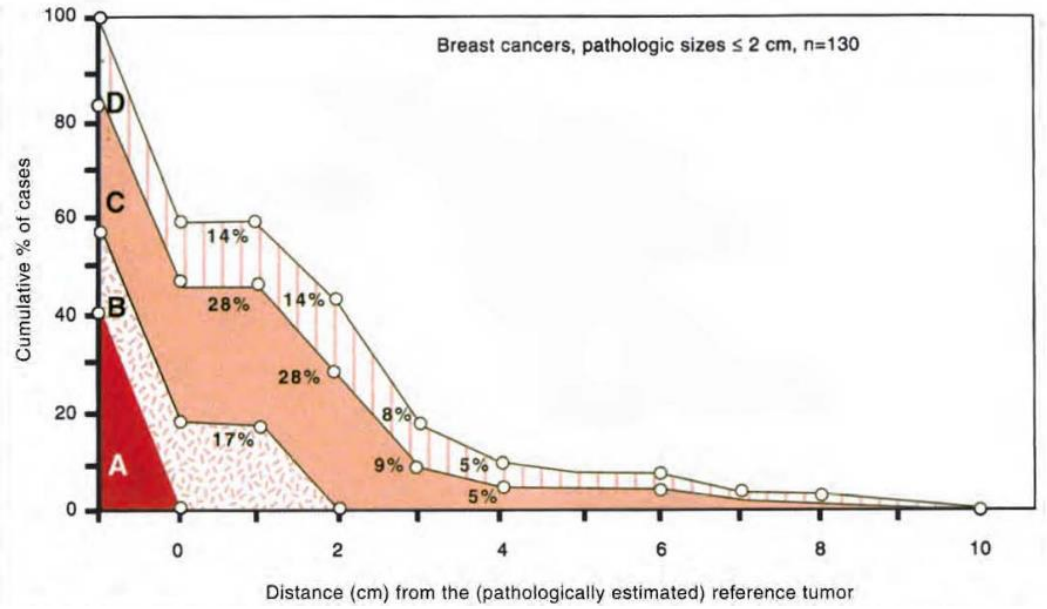


Figure 4: Example of the distribution of tumor foci at different distances from the border of the GTV. This figure shows the remaining proportion of cases with tumor foci at or beyond the specified distance from the reference tumor.

A: No Tumor foci outside the reference tumor (41% of all cases)

B: Tumor foci within w cm of the reference tumor (17% of cases)

C: Noninvasive tumor foci extending farther than 2 cm from the reference tumor (28% of all cases)

D: Invasive tumor foci extending farther than 2 cm from the reference tumor (14% of all cases) Adapted from [2].

ICRU report 62 also further broke down the expansion of the CTV volume with two, instead of one, margins and volumes. The first pair are the Internal Margin (IM) and Internal Target Volume (ITV), which are intended to compensate for all variations in site, size and shape of the organs and tissues contained in or adjacent to the CTV. Basically, this volume compensates for the physiological changes in site or size, which are not easily controlled, and would be somewhat

analogous to the proposed uncontrolled motion uncertainties. In some geometries, internal margin is a very low value, such as in brain tumors that are encased in a rigid structure. The second margin expansion would be the set-up margin (SM), that is intended to account specifically for uncertainties in patient positioning and alignment of the beams during treatment planning and delivery. As such, a set-up margin depends on each beam used, as well as the setup and positioning equipment available.

When both margins are combined, the PTV is obtained. Since margins are used to compensate for both random and systematic uncertainties, a quadratic combination approach is often employed for this combination, however, this report expanded on the “theoretical concept” as introduced in ICRU 50, and stateed that delineation of the PTV is a matter of compromise. As such it is therefore the judgment and responsibility of the radiation oncologist and radiation physicist, based on experience and judgment drawn from observation and evaluation of the risk of failure, especially in cases where PTV and Planning Organ at Risk Volume (PRV) partially overlap. A schematic representation of these combinations can be seen in figure 5.

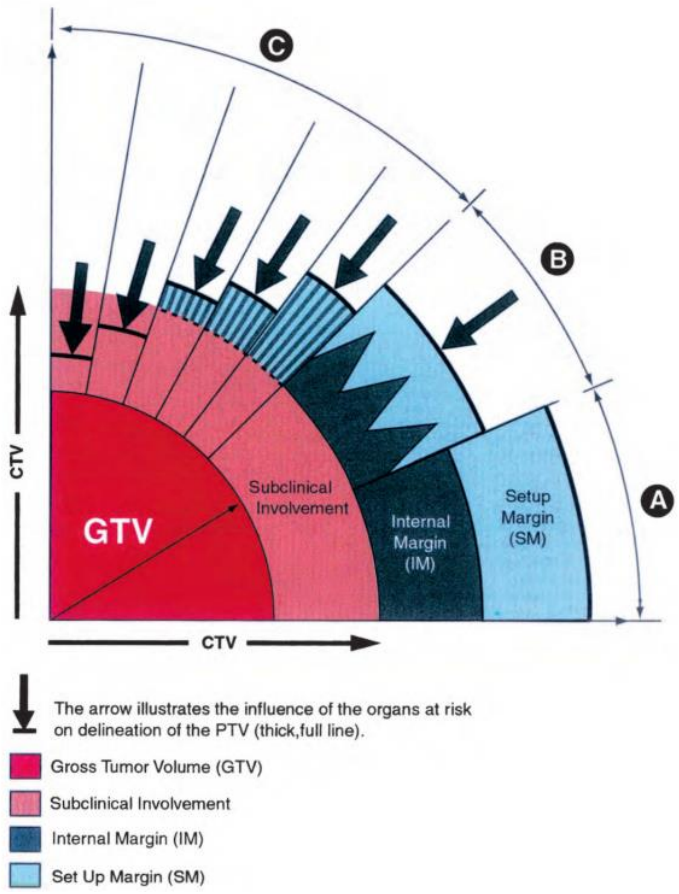


Figure 5: Schematic representations of the relations between the different volumes (GTV, CTV, PTV and PRV) in different clinical scenarios. Scenario A represents a conservative approach with linear addition of IM and SM. As scenario A often leads to an excessively large PTV, a smaller margin must be accepted. A quantitative approach (e.g. using the $\sqrt{\sum \sigma^2}$) is only relevant in optimizing the margin size, if all uncertainties and their σ are known, i.e. such as in a few sophisticated protocols. Scenario C represents a case where a nearby OAR dramatically reduces the width of the acceptable safety margin, and it might be preferable to save the patient's quality of life, even with a somewhat diminished chance of cure. Adapted from [2].

While this updated PTV definition hints at it, ICRU report 62 also explicitly described cases with a tradeoff between target underdosage and OAR overdosage, which, it recommended, should be

balanced considering the patient as a whole, rather than as a purely radiotherapy driven decision. For example, one must avoid overtreating patients, who will fail either locally, or at a distance, or who have good prognosis due to successful treatment of the GTV. Therefore, it is a matter of clinical judgement whether treatment of border tissues is justified. While the report acknowledged the intrinsic trade-off between target underdosage and OAR overdosage, it left the decision on how to balance the two up to the clinical experience.

2.3.4 ICRU 83

As technology kept advancing, and intensity-modulated radiation therapy (IMRT) became an established concept, an update to the guidelines was needed. This led to *ICRU Report 83: Prescribing, Recording and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT)* being published in 2010.

ICRU Report 83 primarily updated the guidelines to reflect the unique challenges associated with IMRT treatments, such as the new opportunities made available by this technique, and with proper reporting of these treatments. The report sought consistency with previous reports, keeping the GTV-CTV-PTV paradigm and advising against introducing new volumes, such as biological target volume, proliferative target volume and hypoxic target volume, since such concepts can already be captured less ambiguously in the existing paradigm.

ICRU Report 83 primarily updated the reporting guidelines from previous reports. The main shift was from dose reporting in only specific reference points, to dose-volume reporting, such as with dose-volume histograms (DVH) and dose-volume descriptors such as D_V, or highest dose delivered to at least V of volume. It also explored concepts such as the clinical and biological metrics of tumor control probability (TCP) and normal-tissue complication probability (NTCP). It also recommended using a remaining volume at risk (RVR) as all the volume within a patient not captured by other regions of interest.

While ICRU Report 83 did not give specific recommendations on PTV margins, it did provide an overview of published methods to combine random and systematic uncertainties, which can be seen in figure 6 and is explored further in the next section.

Author	Region	Recipe	Comments
Bel *et al.* (1996)	PTV	0.7σ	Statistical uncertainties only (linear approximation)—Monte Carlo.
Antolak and Rosen (1999)	PTV	1.65σ	Statistical uncertainties only, block margin?
Stroom *et al.* (1999a)	PTV	$2\,\Sigma + 0.7\sigma$	95 % absorbed dose to on average 99 % of CTV tested in realistic plans.
van Herk *et al.* (2000)	PTV	$2.5\,\Sigma + 0.7\,\sigma$ (or more correctly): $2.5\Sigma + 1.64$ $(\sigma - \sigma_e)$	Minimum absorbed dose to CTV is 95 % for 90% of patients. Analytical solution for perfect conformation.
McKenzie (2000)	PTV	$2.5\,\Sigma + \beta + (\sigma - \sigma_e)$	Extension of van Herk *et al.* (2000) for fringe dose due to limited number of beams. The factor β depends on the beam organization.
Parker *et al.* (2002)	PTV	$\Sigma + \sqrt{(\sigma^2 + \Sigma^2)}$	95 % minimum absorbed dose and 100 % absorbed dose for 95 % of volume. Probability levels not specified.
van Herk *et al.* (2002)	PTV	$2.5 + \Sigma + 0.7\sigma + 3$ mm (or more correctly): $\sqrt{2.7^2\Sigma^2 + 1.6^2\sigma^2} - 2.8$ mm	Monte Carlo based test of 1 % TCP loss due to geometrical errors for prostate patients, fitted for various σ and Σ.
Ten Haken *et al.* (1997), Engelsman *et al.* (2001a, 2001b)	PRV (liver and lung)	0	No margin for respiration, but compensation by absorbed-dose escalation to iso-NTCP, reducing target-dose homogeneity constraints.
McKenzie *et al.* (2000)	PRV	A	Margin for respiration on top of other margins when respiration dominates other uncertainties.
van Herk *et al.* (2003)	PRV (lung)	$0.25\,A$ (caudally); $0.45\,A$ (cranially)	Margin for (random) respiration combined with random setup error of 3 mm SD, when respiration dominates other uncertainties ($A > 1$ cm).
McKenzie *et al.* (2002)	PRV	$1.3\,\Sigma \pm 0.5\,\sigma$	Margins for small and/or serial organs at risk in low (+) or high (−) absorbed-dose region.

Symbols: Σ, standard deviation of systematic uncertainties; σ, standard deviation of statistical (random) uncertainties; σ_e, describes width of beam penumbra fitted with a Gaussian function; A, peak-to-peak amplitude of respiration.

Figure 6: Summary of various published recommendations for margins around target volumes (CTV) and OAR. Modified from [3])

2.3.5 Combining uncertainties

For clinical treatment planning, simplicity in workflow is desirable, and different sources of uncertainty are typically combined into a single target volume that should ideally capture them.

While GTV to CTV expansion depends considerably on previous experience and knowledge of the disease, as well as various clinical parameters, the CTV-to-PTV margin requirements have been extensively explored. An excellent summary of different ways of performing this combination is presented by Marcel van Herk in his 2004 publication [70].

Consistent with the ICRU reports 62 and 83, uncertainties are classified as random (type A) and systematic (type B). Of the two, systematic errors have been consistently found to be much more important in affecting the final dose, since random errors cause a relatively predictable blur in the dose distribution, whereas systematic errors cause its shift relative to the target. The best way of combining uncertainties is not a straightforward question, since they are typically not correlated, and a linear addition of their standard deviations is not correct.

Several margin recipes have been published: Bel et al. showed that a margin for random deviations of 0.7 times their standard deviation (SD) is adequate to keep a 95% dose coverage [71]. Aaltonen et al, derived that a factor of 0.5 to 0.7 times the SD is a sufficient margin, based on biological modeling [72]. Killoran et al. introduced the concept of probability of prescription dose, which can be described for each ROI [73].

Some groups expanded on the probability approach and explored the use of coverage probability matrices to derive margins. Stroom et al. in 1999 provided a margin recipe based on coverage probability, using 2 times the total SD of systematic uncertainties (Σ) plus 0.7 times the total SD of random uncertainties (σ) to ensure 99% of target volume receives at least 95% of the prescribed dose, statistically [74]. A variation on this approach was published by van Herk et al. in 2000, which used dose population histograms to derive a margin recipe to guarantee that 90% of patients in the population receive a minimum cumulative dose of at least 95% of the prescribed dose [75]. This very popular formulation can be described as $2.5\,\Sigma + 0.7\,\sigma$ and is commonly known as “the Van Herk equation”. An updated variation was published by van Herk et al. two

years later, based on Monte Carlo simulation that was based on 1% TCP loss due to geometrical errors for patients with prostate cancer. Many other variations of the concept followed, with overall agreement that adding different random errors should be done quadratically, same as for systematic errors. Combining the two resulted in many different recipes, but overall systematic errors were assigned 3-4 times larger contribution to the margin than random errors.

2.3.6 Probabilistic planning

With the continued improvements in imaging data and planning capability, more complex approaches to incorporating uncertainties in radiation therapy started to appear. In 2005 Chu et al. presented a margin-less approach to planning using robust optimization through different scenarios of positioning errors rather than margins [76]. In 2006, Chan et al. presented a different margin-less approach to create dose plans for tumors in the thorax and abdomen robust to breathing motion [77] and presented dose distributions featuring “horns” - increases in dose near the edges that help ensure adequate coverage of those edges in the presence of uncertainties. In 2007, Unkelbach et al. expanded probabilistic planning to intensity-modulated proton therapy (IMPT) [78]. There, the strongly nonlinear dose deposition presents challenges that are not well addressed using the margin approach, and probabilistic planning quickly became the preferred planning approach [65], [79]–[81]. Various implementations of probabilistic incorporation of uncertainties were also developed for IMRT, such as [82]–[84], accounting for uncertainties using various probability-weighted optimization functions. An excellent review of these approaches was published in 2018 by several of the key authors on this topic [7], where they summarize key studies over various disease sites and implementations.

2.3.7 Probabilistic target definition

While the GTV-CTV-PTV paradigm introduced by ICRU reports allowed for simple and consistent planning, it also oversimplified the heterogeneity and complexity exhibited by tumors [17], [40],

[42], [85]–[87]. In 2001, studies using “slow” CT scans to capture tumor motion during radiotherapy presented an early “probabilistic” description of tumor [88]. Studies investigating tumor growth via modeling embraced the probabilistic descriptions as well, since they were not directly limited by the requirements of radiation therapy, such as works by Swanson et al in patients with glioblastoma [20], [21]. In recent years, several publications from Massachusetts General Hospital explicitly proposed using a probabilistic target definition over the classical target volumes [89]–[91].

2.4 Current shortcomings

While handling of uncertainties in radiation therapy is a popular, and often researched subject, there are still considerable uncertainties and inconsistencies present. The ICRU reports established a well-defined set of volumes which standardized dose prescription and reporting and greatly facilitated radiation therapy comparisons between different research centers. However, they also refrained from defining a single recipe for calculating margins. This led to several publications with different recommendations on how exactly to calculate margins in different disease sites. Additionally, most of these recommendations do not intrinsically account for cases where these margins would overlap with OAR and often defer to the clinical experience of the radiation oncologist. This typically involves defining smaller target margins close to OAR, with the justification that cancer cells are less likely to be present further away from the tumor.

In addition, whether the CTV receives the prescribed dose depends on the dose distribution, rather than just geometric margins. In reality, dose distributions are neither perfectly conformal nor equally conformal on all sides of the CTV, and the optimal margin may be anisotropic. Depending on the dose distribution, a margin may not be needed in some regions, while the same margin may be insufficient in other regions. Margins also fail to balance tumor coverage and OAR sparing, which may be especially crucial for target close to sensitive serial organs (e.g. the spine).

While prescribing multiple dose levels to the same target is possible using the classical ICRU approach, it is cumbersome, as it involves defining multiple CTVs. Using the classical planning approach for integrated boosts may also lead to overly conservative plans that would deliver excessive doses to the target and healthy tissues alike.

Probabilistic target definition may include the answer to some of these challenges, and sufficient work has been performed in disease modeling to justify it as a method [89]–[91]. However, very few works have so far attempted to combine probabilistic target definition with a probabilistic approach to treatment planning. One exception is a publication by Sterpin et al., where they explore robustness evaluation strategies for statistically consistent reporting in a population of patients with head-and-neck cancers receiving homogeneous dose prescriptions [92].

In this thesis we seek to address some of these gaps, by presenting a probabilistic target definition and planning approach in multiple disease sites and including heterogeneous disease sites. To present the method, we first apply it to patients with glioblastoma that received test-retest imaging, allowing us to perform our own evaluation of imaging uncertainties, presented in greater detail in the next chapter.

3 Probabilistic Target Definition and Planning (PTP) Methodology

In this chapter we present the theoretical framework of the methodology used in this thesis. In Section 3.1 we perform a high-level introduction of the approach. In Section 3.2, we examine in further detail the probabilistic tumor volumes, and in section 3.3 we discuss the implementation of the PTP optimization. Section 3.4 is dedicated to a review of probabilistic dose reporting metrics.

3.1 PTP Introduction

Probabilistic target definition and planning (PTP) is a novel treatment planning approach incorporating probabilistic, non-binary target definition and robust optimization to create radiation therapy dose plans. The general workflow of this proposed planning approach can be seen in figure 7.

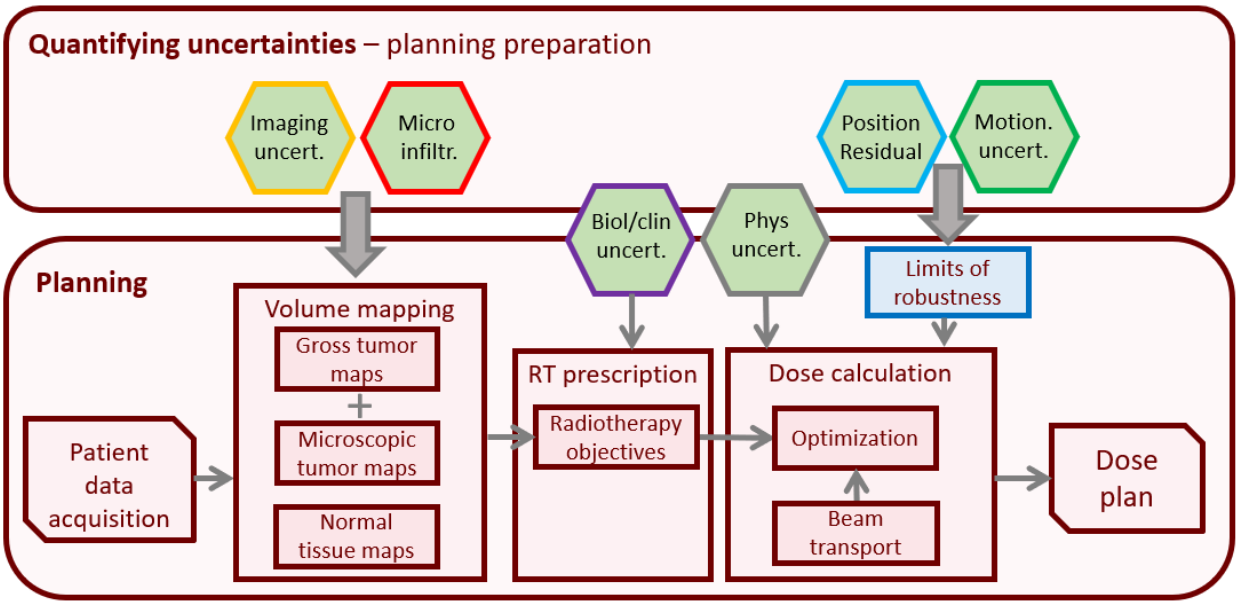


Figure 7: Schematic outline of the main sections of the new RT workflow, based on proposed ICRU guidelines. The key difference from previous guidelines is the use of probabilistic target definition and probabilistic planning. First, all relevant uncertainties need to be quantified in the

pre-planning phase. After patient data acquisition, tumor and normal tissue probabilistic maps are created and used to optimize the treatments.

In probabilistic planning, rather than using binary target definitions, we use “tumor likelihood maps” or TLMs. These are 3D matrices containing a non-binary description of disease presence that allow for a continuous description of disease likelihood, where any value in the range [0,1] can be assigned to a voxel. TLM is composed of two components: gross tumor likelihood map (gTLM) and microscopic tumor likelihood map (miTLM).

Equivalent to classical planning, the PTP approach requires a quantification of all relevant uncertainties to be performed. These can be sourced from literature or evaluated specifically for the considered implementation. Rather than use binary target volumes to describe disease presence, probabilistic volumes are used to describe gross and microscopic disease presence. These tumor likelihood maps can either be used in treatment planning directly, or first converted into Target Maps, where the values do not represent the likelihood of disease presence directly, but are adjusted to better achieve a desired dose distribution.

To account for motion and residual positioning uncertainties such as accurate patient placement, a robust optimization (RO) approach is used, with scenario selection using multiple reality realizations. This allows for a plan optimization with non-uniform dose prescriptions while simultaneously considering the trade-offs between target and OAR coverage based on user determined optimization weights.

This approach represents a paradigm shift which allows us to move away from simplistic volumetric expansions and traditional optimization. It also allows for creation of non-uniform dose plans and an intrinsic way of defining trade-offs in cases of nearby organs at risk.

3.2 Probabilistic volume mapping

3.2.1 Gross tumor likelihood maps (gTLM)

Gross tumor likelihood maps (gTLM) are 3D matrices containing a non-binary description of disease presence. Unlike the before-mentioned GTV/CTV/PTV volumes that are binary by nature, gTLMs allow for a continuous description of disease likelihood, where any value in the range [0,1] can be assigned to a voxel based on medical imaging provided source information. This step is analogous to defining GTV in the classical planning workflow; however, with the continuous values allowing for more information than the binary definition used for classical planning.

gTLM mapping can be performed manually, using multiple different probability levels, or as a computer-based numerical analysis. In the latter case, a range of functions could be used, as long as the function converts the input values to a range not exceeding [0,1]. For example, in Chapter 4 we use a cumulative distribution function of the normal distribution (i.e. the Gauss error function). Other linear, exponential or polynomial functions can also be used, as well as multiparametric functions combining information of multiple different imaging modalities or other sources of information.

In this thesis we mainly focus on using continuous TLM values, however this mapping could also be performed using a few discrete probability levels. As a limiting case of binary probabilities, just 0 or 1 values could be used, equivalent to the traditional GTV segmentations.

3.2.2 Microscopic tumor likelihood maps (miTLM).

Similar to gTLM for visible disease, microscopic tumor likelihood maps (miTLM) are probabilistic representations of expected extent of microscopic infiltration. By definition, microscopic infiltration is a process on a microscopic level where malignant cells infiltrate nearby tissues in a concentration insufficient to be visible on medical imaging. Previous studies in various disease

sites have evaluated this process and found that the infiltrations follow an approximately exponential shape [6], [93]. Additional guiding principles for the algorithm have been identified, as summarized below:

1. The method needs to be probabilistic. It should be able to accept non-binary disease maps as input and similarly output non-binary maps showing probability of disease presence.
2. The method should not decrease the probability of disease in gTLMs or increase their values above 1.
3. The method should not allow tumor infiltration into anatomical constraints like air cavities or bone for nearby soft tissue tumors. The method needs to be able to take as input the information about these invalid zones and calculate miTLM accordingly.
4. The method should not allow the tumors to “tunnel” through obstacles. For example, we would not expect a tumor contained within the skull to be able to penetrate bone, unless a gap in the bone is present.

While these four guidelines represent a reasonable example model, some diseases might require exceptions, such as disease in boney sites.

Some of the common shortcomings of miTLM derivation have been identified. For example, using simple convolution methods modifies the values within the tumor (figure *8*a). Using a more generalized volume expansion approach generates tumor presence in volumes where we would not expect the tumor to be present, such as bone or nearby air cavities (figure *8*b). Correcting only the values within the mask of illegal expansion generates cases where a tumor might seem to be penetrating through anatomical constraints, and generate tumor presence on the other side, where it is not expected (figure *8*c).

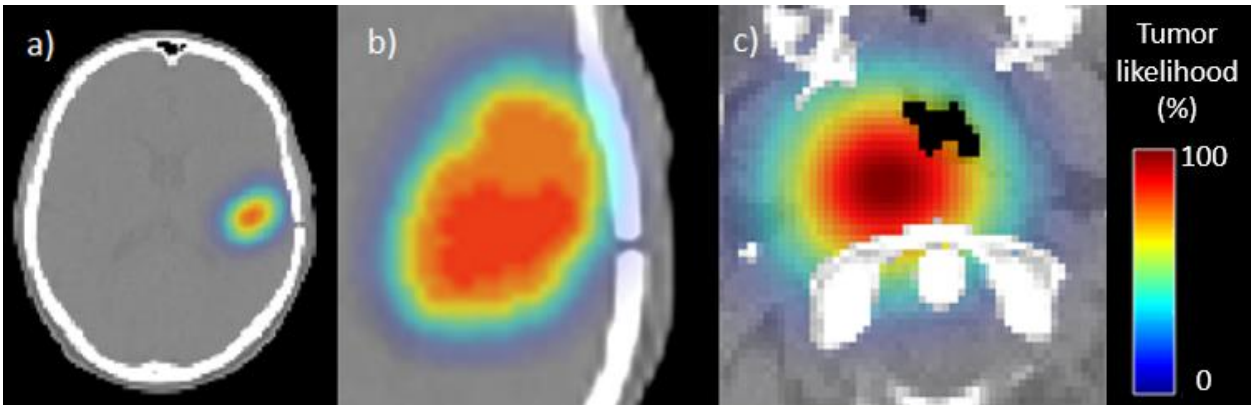


Figure 8*: Example shortcomings of insufficient methods of calculating miTLM. Example a) shows the infiltration model using convolution where values within gTLM have been reduced below those of the initially defined disease in comparison with figure 22. Example b) shows miTLM expansion using simple exponential expansion that expanded into skull. Image c) shows a case where illegal volumes (bone and air) have been corrected for, but the method allowed disease "tunneling" though a bone obstacle.*

One of the methods developed in this work that meets the guidelines described above is based on iterative region growing. This expansion accepts a gTLM, a mask of invalid anatomical volumes generated by thresholding the CT image, and a range of microscopic infiltration as identified from literature [20]–[22] as inputs, and calculates miTLM.

In the first step, the gTLM values are adjusted to account for likelihood of nearby infiltration. The likelihood of disease presence in a voxel is set to the maximum of either the initial gTLM value in that voxel, or the likelihood of infiltration from any of the nearby voxels with higher disease value, adjusted for distance. This is performed in all voxels of nonzero gTLM and their immediate neighbors. This smooths out the gTLM values, especially in areas where voxels of low likelihood of tumor presence exist close to voxels of high likelihood of tumor presence. An example of this procedure can be seen in figure 9.

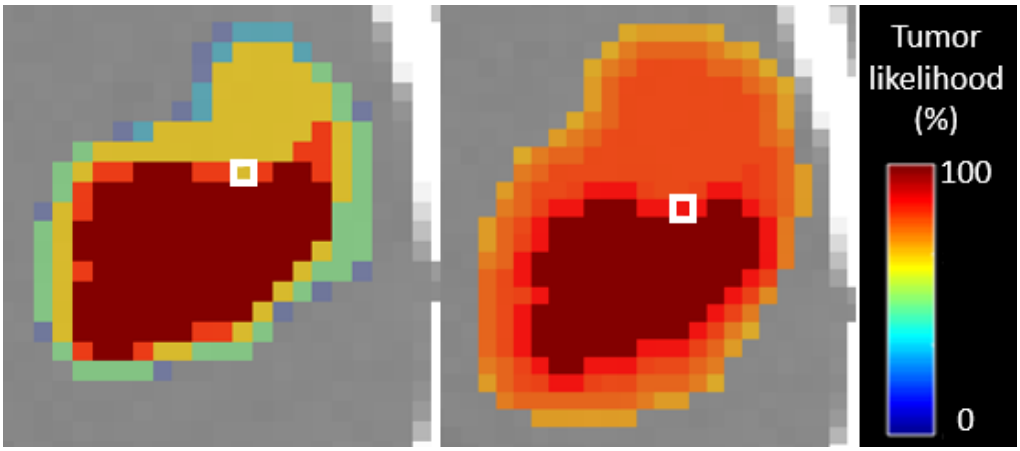


Figure 9: Initial gTLM of an example glioblastoma lesion (left) and the 1st iteration of calculating the miTLM (right). The same example voxel is shown with a white frame on both images. Despite starting with a lower TLM value than its neighbors, the map is smoothed out due to the proximity to the highest probability region (dark red).

Next, an outer layer of new nonzero miTLM voxels that are not defined as anatomical constraint obstacles by the CT mask is identified. For every voxel in this new outer layer, the new miTLM value is calculated based on the previous outer layer miTLM values and distances from every nearby voxel, as shown in figure 10. Layers are added until the desired precision is achieved. As layers are added, each new layer is based on the existing outermost layer, allowing only infiltration from immediately adjacent volumes. This prevents “tunneling” of the tumor through different tissues.

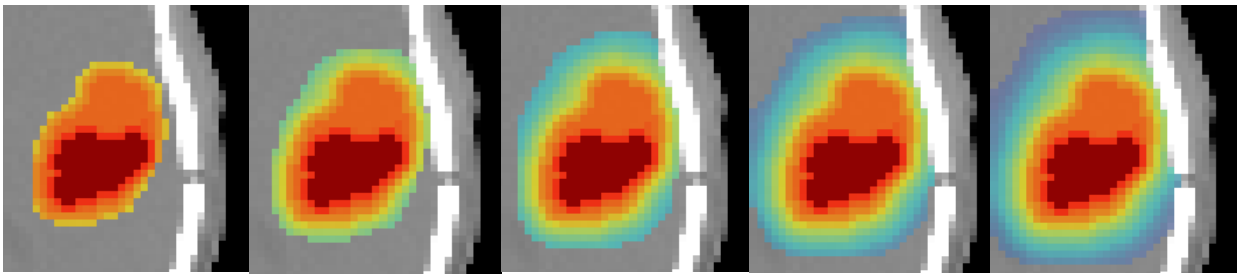


Figure 10: Five iterations of calculating a gTLM+miTLM map shown on an axial CT slice of a patient with glioblastoma. In each iteration, only miTLM values in the immediately neighboring voxels are calculated. The expansion is performed in 3D, allowing for infiltration in superior/inferior directions.

3.3 PTP optimization and objective functions

For classical optimization, a uniform dose is prescribed to the entire planning volume and optimized using an objective function. While different functions may be used, a common implementation is using linear quadratic penalty functions for different objectives, mathematically formulated in the general form as:

$$f(x) = \sum_{j=1}^{N} \left(w_j \cdot f_j\left(q, d(x)\right) \right) \tag{3.1}$$

for:

$$0 \leq x \leq mf \cdot \bar{x} \tag{3.2}$$

where j is an optimization objective, w_j is the optimization weight for objective j, f_j is the objective penalty function, q is the prescription dose, $d(x)$ is the dose calculated for beamlet weights x, and mf is the modulation factor.

For probabilistic optimization, we used a minimax RO formulation to calculate the objective function value [80], [94]. We implemented a static dose cloud and a supervoxel volume grouping with ROI-variant density to reduce the computation burden and help with the tractability of the optimization. Mathematically, the minimax RO objective function is of the form:

$$f(x) = \sum_{j=1}^{N} \left(w_j \cdot \max \left\{ f'_j \left(q, d(x,s) \right) \right\} \right) \tag{3.3}$$

while still adhering to the modulation constraint of equation 3.2. The variables are the same as in equation 3.1, and the newly added variable s represents scenarios considered. For the scenario selection, we used a "good practice scenario selection" approach [95], using 2 mm shifts in the six cardinal directions to account for positioning uncertainties, analogous to the 2 mm CTV-PTV expansion in the classical case.

For PTP approach we developed an expansion of the RO, where we used the voxel-level TM values to adjust the objective function within the target volume. Where the penalty function for the optimization goal *j0*, setting the minimal dose to target, is of the form:

$$f'_{j0}\left(q, d(x,s)\right) = \sum_{i=1}^{N_i} \left(TM(i)\right) \cdot \left(\min\left(D_i, D_{ref,j0}\right) - D_{ref,j0}\right)^2 \tag{3.4}$$

Here TM represents the value of the target map in voxel i, based on the likelihood of disease presence, D_i is the achieved dose to that voxel, and D_{ref} is the desired dose in that voxel. While we used a homogeneous dose prescription in this case, it is worth pointing out, that this formulation is intrinsically compatible with heterogeneous dose prescriptions, such as dose

painting. From the equations 3.3 and 3.4 it is also apparent that if a uniform TM is used, PTP optimization is simplified to standard RO.

For all optimization, we set a maximum modulation factor of 5 for all plans and objectives. We used the same relative optimization weights for plans optimized with the classical and RO approaches. We iterated the optimization process until a threshold step size was reached, or for a maximum of 100 iterations, and we observed convergence to a stationary point in all optimizations where the optimization did not converge within this frame.

It is important to note that penalty functions other than linear quadratic can be used in PTP. We presented the methodology using these penalty functions for ease of understanding, but other functions, such as linear, exponential, and step-quadratic have also been implemented and tested for different implementations.

3.4 Plan Reporting

Currently established evaluation procedures feature a limited exploration of plan robustness and generally do not provide confidence levels on their dose metrics. In the presence of uncertainties, doses to different volumes can change considerably. Reporting not only the nominally expected dose, but also the range of such values over the range of expected realizations of uncertainties, would provide insight on the robustness of a plan to such uncertainties.

Some existing studies have introduced different ways of incorporating uncertainty information into dose plan reporting. In this section we summarize different existing ways of probabilistic dose reporting, and introduce a new dose metric – the QV maps.

In this section we first discuss evaluation scenario selection. Evaluation scenarios, in general, can be different from planning scenarios, since evaluation presents less of a computational burden than optimization. Regardless of which scenario selection method is used, it should be reported clearly, as it may affect the report metrics.

In the second section, we review several existing dose reporting metrics: Dose benchmarks and ranges, DVH clouds and QV histograms. We also introduce an additional metric: QV maps, which represent adherence to desired dose optimization goals.

3.4.1 Scenario selection

The first question is proper evaluation scenario selection – what realizations of uncertainties will the plan be evaluated for [54], [96]. An excellent summary can be found in Sterpin et al. [92]. A summary of the different scenario selection approaches can also be seen in figure 11.

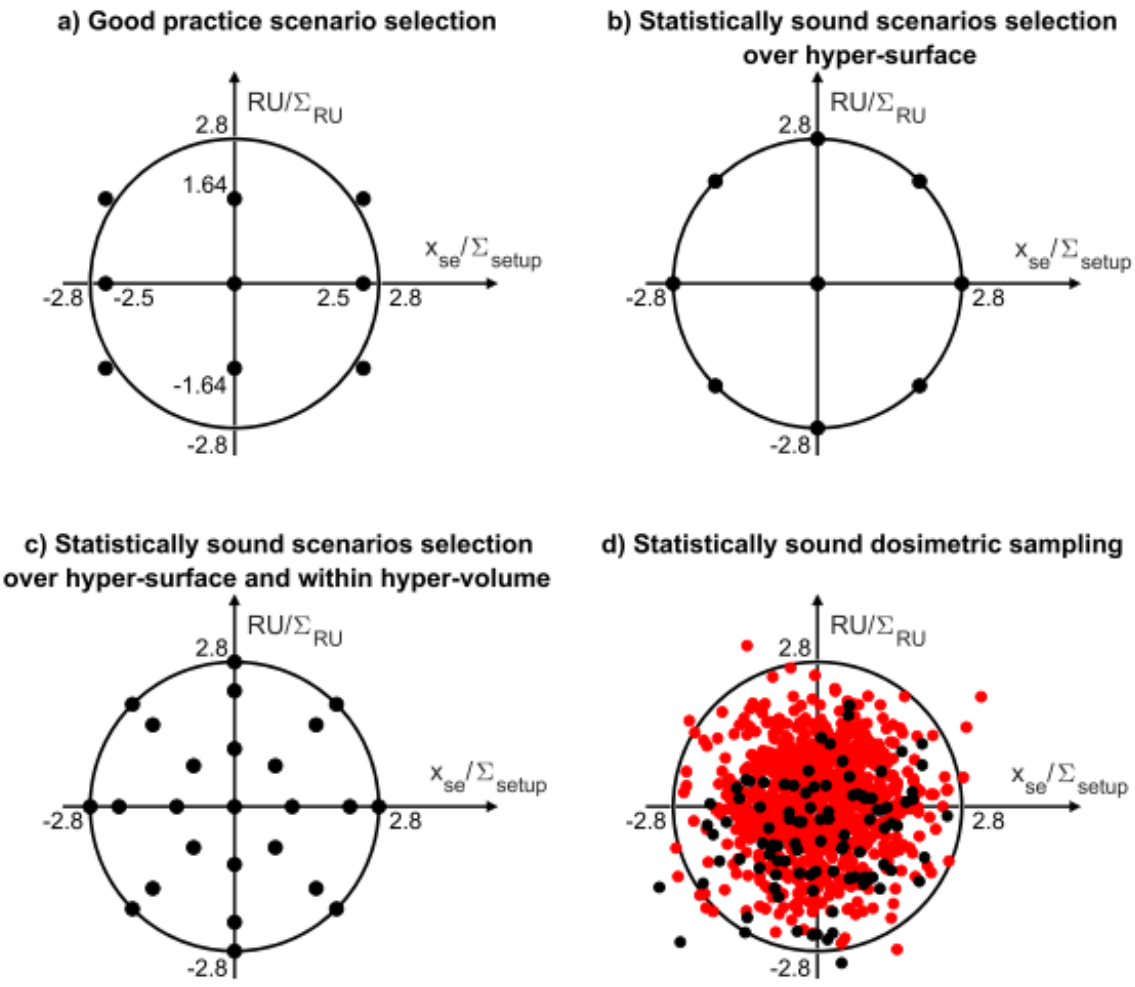


Figure 11: Illustration of the scenario selection approaches in a 2D slice (a)-(c), or projection (d), with each dot representing a simulated scenario. Adapted from [92]*. As the article considered proton plan evaluation, RU represents range uncertainties, but the principles remain the same. In (d) red dots represent selected dose distributions, black dots the discarded distributions (for when the total sampled uncertainty was greater than the predefined limit).*

3.4.1.1 Nominal scenario evaluation

Nominal scenario evaluation refers to the results obtained when evaluating a dose plan on the original, un-modified geometry. This returns a single value, and as such it does not capture any variation in results, but it is the most intuitive reporting and is easily understood due to its similarity

to current practice. It should be noted that such evaluations would typically still be performed on the CTV, rather than the PTV volume, which could lead to slightly different results.

3.4.1.2 Good practice scenario selection

Good practice scenario selection refers to a setup where the CTV to PTV margin is replaced with a systematic setup error of comparable magnitude. This results in a few scenarios being evaluated, depending on uncertainties considered. For example, for positioning uncertainties, shifts in the cardinal directions may be considered, resulting in 7 scenarios, including the nominal case. In proton planning, range uncertainties may be added, resulting in up to 21 scenarios.

The advantage of this approach is that due to relatively low number of scenarios considered it is comparatively easy to calculate. It also offers clear parallels that can be drawn to classical planning using margins. Its main disadvantage is that is represents a coarse sampling of uncertainty space, and consequently the confidence intervals are not as accurate as when obtained with more advanced methods.

3.4.1.3 Statistically sound scenario selection

In statistically sound scenario selection, scenarios are sampled uniformly on the hyper-surface of an isoprobability sphere. In other words, scenarios are sampled so that they represent the border surface for a confidence level α in any of the parameters or a combination of multiple parameters. Due to the added combination of multiple uncertainties, this approach requires a larger number of scenarios to be considered in order to adequately sample the uncertainty space. Additionally, hyper-spheres with multiple (reduced) different radii can also be considered to populate not only the surface of the largest sphere, but also some internal points, representing scenarios that are better than the worst-considered case. The advantage of this method is a more accurate evaluation of possible scenarios but at the downside of higher computational cost.

Statistically sound scenario selection and *statistically sound dosimetric selection* also require the user to correctly adjust for additional dimensions of uncertainty (such as radiation range, rotations, tissue density, etc.). The isoprobability surface of the hyper-sphere defining the hyper-volume that needs to be considered depends on the number of degrees of freedom, assuming a constant confidence interval α. For example, in 3-dimensional error space, a sphere that encompasses 90% of all scenarios would have a unit radius of 2.5, as is also used in the famous van Herk equation [75]. In 4-dimensional errors space, the radius of the hyper-sphere increases to about 2.8 to maintain encompassing 90% of all scenarios. A visual representation of this dependency on the number of dimensions can be seen in figure 12.

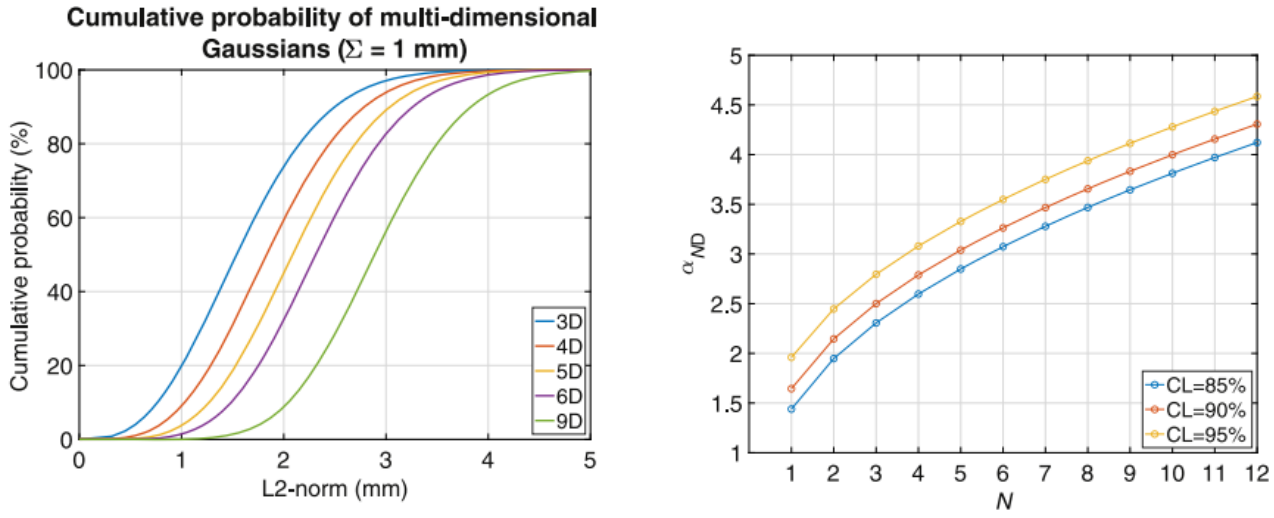


Figure 12: Left: examples of cumulative probabilities for isotropic multi-dimensional independent normal (Gaussian) distribution. Decrease of the cumulative probability using the same L2-norm can be seen as the number of dimensions increases. Right: the radius dependency on the number of dimension in order to ensure a confidence level (CL) of desired size. Adapted from [92].

3.4.1.4 Statistically sound dosimetric selection

In statistically sound dosimetric selection, scenarios are randomly sampled according to the pdfs of the uncertainties. This can be done through an analytical tool, or Monte Carlo sampling. This results in a method that most accurately models real behavior, as well as enabling an evaluation of the confidence interval in the dosimetric space. This, however, comes at a price of simulating a large number of scenarios in order to get the necessary statistics. For example, in the work of Sterpin et al. 1000 scenarios were simulated for each evaluation, resulting in slightly more accurate dose reporting but requiring the use of powerful hardware that may not be readily available.

3.4.2 Dose reporting metrics

3.4.2.1 Dose benchmarks and ranges

Once the desired scenario selection method is chosen, reporting the uncertainty information in dose planning becomes possible. One way of reporting is using established dose metrics and adding ranges of values evaluated over the scenarios. An example of that would be using the "nominal [min-max]" notation, where the dose metric evaluated on the nominal scenario is reported first, and then the minimum and maximum values over the scenarios. Reports of average and mean values can also be added or used in place of the nominal value.

3.4.2.2 DVH clouds

Dose-volume histograms have been introduced by Shipley et al. in 1979, and have been an invaluable tool for dose reporting ever since [97]. An expanded version of the DVH are the DVH clouds, where multiple DVH are overlaid on top of each other, resulting in either a group of lines, each representing a single scenario, or a band representing a range of values [54], [76]. While a band width equal to the standard deviation of the data at that point could be used, dose distribution

across scenarios does not necessarily follow a Gaussian function, and therefore standard deviations may not be accurate representations of uncertainties.

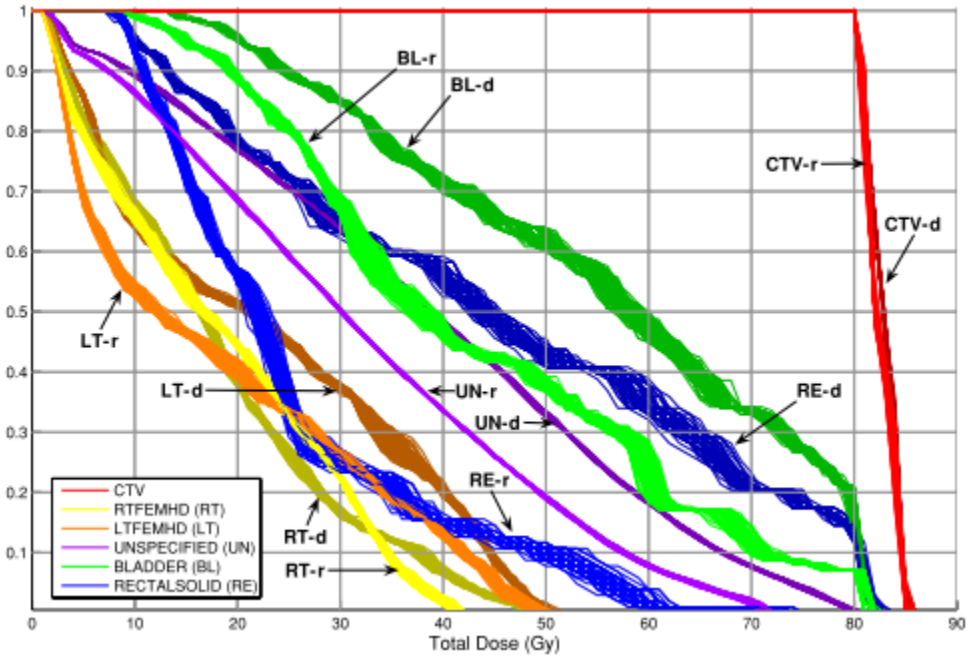


Figure 13: Dose-volume histogram cloud for 100 simulated treatments for a patient with prostate cancer. Here, the DVH cloud is presented with an overlay of different lines, rather than a min-max band. Adapted from [76].

3.4.2.3 QV histograms

While DVH help presenti the relationship between the absolute dose and volume, such comparisons are less useful in cases of heterogeneous dose prescriptions, as it is not clear what part of the volume receives a higher or lower dose. One way of addressing that is with Quality-Volume histograms (QVH) [98]. This method defines “quality” as the ratio between the planned and prescribed dose:

$$Q_p = \frac{D_p}{D_{presc}} \tag{8.1}$$

This way, each voxel is normalized to the prescription before any evaluation is being made, allowing for easy-to-use and intuitive representation of results.

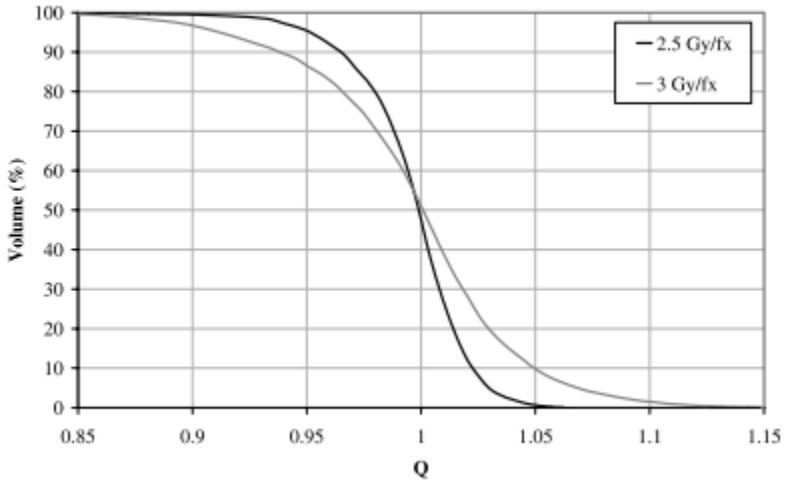


Figure 14: QV histogram of the PTV for example plan. Adapted from [98]*.*

One can also summarize the overall adherence of a volume to a prescription using the *quality factor (QF)*. This is defined as the average absolute deviation of the quality value from 1. Mathematically this is formulated as:

$$QF = \frac{1}{n}\sum_{p}\left|Q_p - 1\right| \tag{8.2}$$

Where *p* is each voxel within the VOI, and *n* is the total number of voxels in the volume.

3.4.2.4 QV maps

Extending the same principle to images of dose distributions, it is also possible to visualize planned dose, normalized to the prescription, or “QV maps”. This way it is possible to retain the spatial information, as well as the dose relative to the prescription. Examples of QV maps can be seen in figure 15.

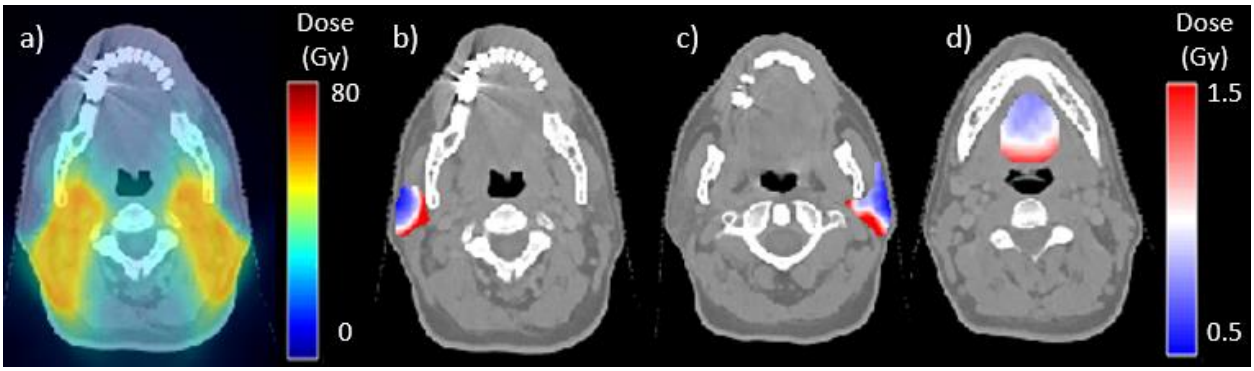


Figure 15: Example dose map (a), and QV maps for left and right parotid glands (c, d), and oral cavity for a patient with head and neck cancer.

Using this visualization, it is easy to see the areas where the delivered dose is higher than the norm – such as in the interior and posterior edges of the parotid glands that are closest to the target volume, or the posterior volume of the oral cavity, as indicated by the red volume. However, it is important to keep in mind that the QV maps will strongly depend on the reference dose value used, which in this case was the maximum mean dose to the volume allowed. As such, these ROI still met the planning objectives despite the red colored areas.

4 PTP Implementation – Step 1 - Uncertainties

The first step in the proposed treatment planning approach, as seen in figure 7, is quantification of uncertainties, initially as an input to defining tumor likelihood maps, typically based on imaging. These uncertainties have been evaluated for several imaging modalities and could be used as a starting point for the proposed probabilistic target definition and planning (PTP) [18], [99]–[101].

However, as an example for the purposes of this thesis chapter, we sought to apply PTP in an anatomical site with limited uncertainties to facilitate validation of the approach. We additionally wanted a dataset imaged with an imaging modality where imaging uncertainties have not yet been evaluated, to demonstrate how the magnitude of such uncertainties can be obtained from imaging data. We chose to implement it in a population of patients with glioblastoma, who were imaged with a less common radiotracer ^{18}F-fluoro-ethyl-tyrosine (FET), whose test-retest characteristics have not yet been sufficiently analyzed. Brain anatomy is also comparatively simple and does not suffer from large positioning or organ motion uncertainties.

The purpose of the work we present in this chapter is to test our methodology in a case with limited number of uncertainties. In brain, motion uncertainties can be neglected, and residual positioning uncertainties can be acquired from existing literature. In this case imaging uncertainties were unknown, but could be measured. As such, we sought to quantify imaging uncertainties on a new tracer and evaluate the first of the uncertainties presented in figure 7, to be used in probabilistic target definition. While different methods of uncertainty quantification than ones presented in this work may be more appropriate for other imaging modalities, the concepts presented can still be used as general guidelines for performing imaging uncertainty quantification.

4.1 Motivation

Gliomas are the most common malignant primary brain tumor, of which glioblastomas (GBM) constitute about 70% [102]. They are also the most aggressive type of brain tumor, with median survival under 12 months [103], [104]. The current gold standard treatment for GBM consists of surgery followed by radiation therapy (RT) with concurrent and adjuvant chemotherapy [19]. While maximum resection of the tumor is usually attempted, functional impairment and quality of life have to be considered when determining the aggressiveness of resection [105]. Thus for both surgical and RT planning, accurate delineation of disease volumes is highly important. Accurate quantification of a physiologically active tumor may allow high dose RT to be appropriately targeted while minimizing risk to healthy tissue [106].

RT planning target volumes are derived using functional [107], [108] or CT-MRI fusion images [19]. A promising tracer for imaging GBM is ^{18}F-fluoro-ethyl-tyrosine (FET). It is an artificial amino acid PET tracer developed in the late 1990s that has preferential uptake in malignant cells with increased expression of amino acid transporters, but is not incorporated into proteins [109]–[112]. FET uses a similar biological mechanism as the longer established [^{11}C]-methyl-L-methionine (MET), with the advantage of the longer lived [^{18}F] isotope [113]. FET has been shown to have excellent performance for diagnosing primary brain tumors, including volumes not visible on MRI, as well as for survival prediction [106], [110], [114]. In recent years, it is becoming increasingly recommended as a tracer for guiding treatment target definition [106], [115].

While multiple publications exist exploring the use of FET [104], [107], [116], [117], little analysis has been done regarding the uncertainties and repeatability of its uptake. Evaluation of a biomarker's repeatability, defined as the variation in measurements when an experiment is repeated under the same conditions [118], [119], is necessary for accurate assessment of tumor presence and response, as demonstrated in many ways for other tracers [99], [101], [118], [120]–[124]. However, repeatability of FET remains largely unexplored. Healthy tissue normalization

techniques attempt to partially address this innate variability in FET; however, no studies so far reported the impact of such normalization over varying volume sizes.

The goal of this chapter is to demonstrate a quantification of imaging uncertainties in a novel example – in this case FET PET/CT imaging of patients with glioblastoma. Evaluating the repeatability of a PET tracer is an essential step in establishing the imaging uncertainties. Knowing the magnitude of these uncertainties will allow us to calculate probabilistic tumor likelihood maps, which will in turn be used in probabilistic radiotherapy planning.

4.2 Materials and methods

4.2.1 Patient population and Region of Interest (ROI) definition

A cohort of patients with histologically confirmed GBM received test-retest FET PET/CT imaging at the Sir Charles Gairdner Hospital (SCGH, Perth, Western Australia) as part of a prospective study to evaluate FET for radiotherapy planning after surgical removal of gross tumor. The study was approved by the institutional review board under SCGH study number 2014-004, and all subjects signed an informed consent form. The study was registered at the Australian New Zealand Clinical trial registry under number ACTRN12614001114639).

The first FET PET/CT scan (test) was scheduled within 4 weeks before the start of RT, and the second scan (retest) was planned to be acquired 7 days after the first scan. Patients received intra-venous administration of 200 MBq (±10%) of FET and were subsequently imaged over the cranial region on a Siemens Biograph 16 PET/CT (Siemens Medical Solutions, Malvern, PA). A 10-minute single frame scan was acquired at an average of 28 minutes (range 23-37 minutes) post tracer injection, with an average difference in scanning time of 1.9 minutes between test and retest scans. Retest images were rigidly registered to test images using monomodal intensity-

based registration (*imregconfig*) implemented in Matlab R2017b (Mathworks Inc, CA). Example axial slices of FET PET/CT data and their voxel-wise differences can be seen in figure 16.

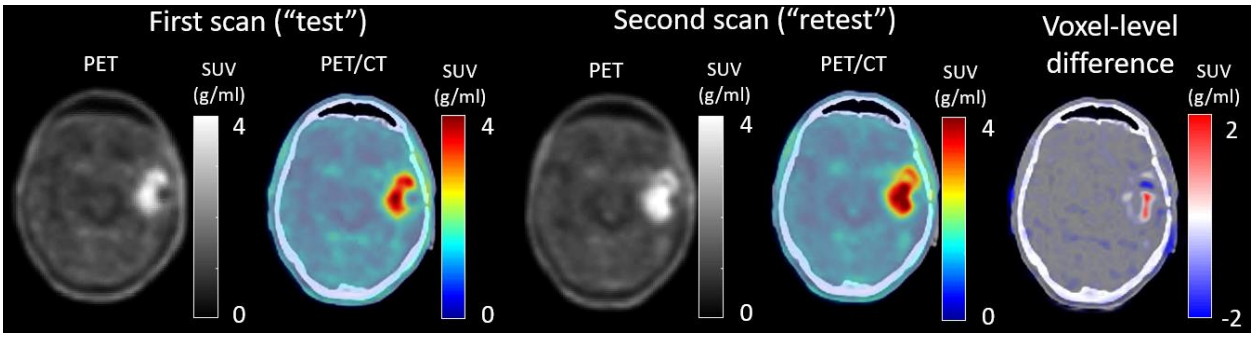


Figure 16: Example comparison of first baseline scan (“test”, left), second baseline scan 7 days later (“retest”, middle) and their voxel-level difference (right). Administered FET activity was 200 MBq and the image was acquired for 10 minutes at 20 minutes post injection. Disease delineations are indicated by a white contour for test and retest images. While the general shape of the lesion persists, considerable changes in uptake are present between the two scans.

We segmented the intracranial volume on baseline CT images. Using a threshold of 200 Hounsfield Units, we acquired bone masks followed by morphological closing using a spherical structure with a 10 mm radius. We performed morphological closing on the largest connected region enclosed in this mask and visually inspected the results to ensure quality of segmentation. For the purpose of this work we considered only uptake within the intracranial volume.

To segment diseased volumes, we implemented an adaptive thresholding method, as it was previously shown that signal-to-background adaptive thresholding performs better in segmenting GBM than fixed threshold metrics [110], [124]. First, we selected three spherical regions with a

radius of 5 voxels (6.4 cm^3) in the contralateral healthy brain as ROIs for normal uptake reference. We calculated SUV_{mean} values of these reference regions and took 1.6x SUV_{mean} as the threshold for lesion segmentation. We performed this on both test and retest images and compared both segmentations. We considered volumes greater than 1.1 cm^3 usable for analysis, as specified in the protocol of the SCGH study number 2014-004, introduced at the beginning of this section.

We used original test and retest contours for whole-lesion comparisons and a union of both volumes for the voxel-wise comparisons to have equal volumes on both time points. We defined healthy brain volume as all intracranial volume, excepting the union of the test and retest disease contours dilated by a five-voxel margin. A schematic representation of this segmentation approach can be seen in figure 17.

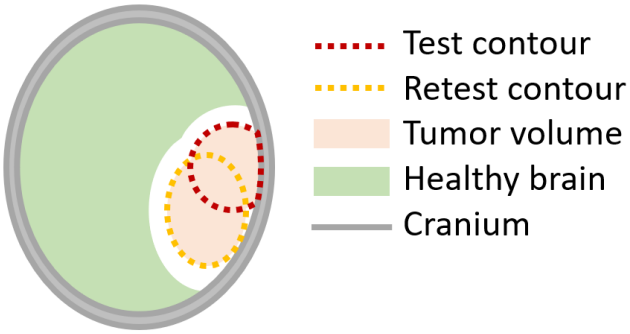


Figure 17: Schematic representation of segmentations. All volumes are constrained by the cranium. Healthy brain segmentation consists of entire intracranial volume, without the expanded union of test and retest tumor volumes.

*4.2.2 **FET repeatability***

We used contours on both test and retest imaging to extract tracer uptake in both tumor and healthy brain. We then calculated basic matching metrics: volume, Dice coefficient, and Hausdorff distance. We extracted SUV_{mean} SUV_{max}, SUV_{total} and $SUV_{98\%}$ for both volumes and both test and

retest imaging. $SUV_{98\%}$ is the 98th quantile of SUV values in the ROI and was investigated as a measure of highest uptake robust to noise, a drawback of SUV_{max} [125].

We calculated Bland-Altman 95% Limits of Agreement (LoA) of SUV for log transformed values, as a way of accounting for increase in absolute differences that occurs with higher SUV [119], [120]. First, we calculated the differences d of log transformed SUV measurements on test (M_A) and retest (M_B), as well as their bias B, standard deviation σ and repeatability coefficient RC.

$$d = \log(M_B) - \log(M_A) = \log\left(\frac{M_B}{M_A}\right) \tag{3.1}$$

$$B = mean(d) \tag{3.2}$$

$$\sigma = std(d) \tag{3.3}$$

$$RC = 1.96 \times \sigma \tag{3.4}$$

Finally, we calculated LoA as

$$LoA = \left[e^{(B-RC)}, e^{(B+RC)}\right] \tag{3.5}$$

This provided a range in which a measurement could be found with a probability of 95%; e.g. LoA of [0.85, 1.15] would indicate that the ratio M_B/M_A is expected to fall between 0.85 and 1.15 with 95% frequency.

4.2.3 SUV repeatability dependence on volume

We evaluated the repeatability of SUV values on the healthy brain volume, and on the tumor volume, defined as the union of segmentations on test and retest imaging, and using groups of nearby voxels or *neighborhoods* of varying sizes.

For each voxel (v_i) we calculated the mean voxel SUV of a neighborhood ($SUV_i^{NH}{}_{mean}$) within healthy or diseased brain. We calculated Bland–Altman 95% Limits of Agreement (LoA) for the test and retest $SUV_i^{NH}{}_{mean}$ on a voxel level. To account for potential global SUV shift, we set bias B to 0 when calculating LoA for varying volumes. We calculated the neighborhoods of every size in the determined range for every voxel within their respective volumes

These neighborhood sizes ranged from 0.01 cm^3 (single voxel) to 3.15 cm^3 (335 voxels), which was the largest neighborhood size which could be fully contained within the larger lesion segmentations and was selected as the cutoff for this analysis. While larger neighborhoods can be used, $SUV_i^{NH}{}_{mean}$ would be impacted by values outside the diseased volumes, obscuring the differences between healthy and diseased tissues.

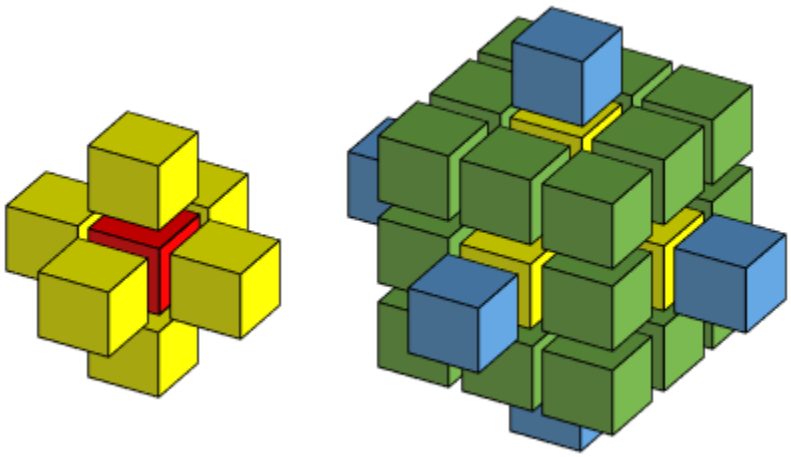


Figure 18: Example neighborhoods around the central (red) voxel of interest. Left side shows a 3/1 neighborhood. All voxels are contained in a cube with side length of 3 voxels, and with voxels no more than 1 unit away from center. E.g. a 3/1 neighborhood returns the 6 closest neighbors. Right side shows a larger 5/2 neighborhood, or all voxels contained within a cube 5 voxels wide, and voxels that are no more than 2 voxels away from center. All voxels shown are

considered for $SUV_{iNHmean}$ calculation. Different colors are shown solely for easier visual interpretation.

We present a full list of neighborhood sizes we considered in table 1. The X/Y label indicates that neighborhood voxels need to be within a cube of width X voxels, centered on the voxel of interest, and no more than Y voxels away from the voxel of interest. Examples of 3/1 and 5/2 neighborhoods can be seen in figure 18. We implemented this way of defining neighborhoods in order to decrease the difference in the number of voxels between neighborhood sizes, compared to simply taking all voxels within distance Y.

Table 1: List of neighborhood sizes considered in the analysis.

Label (X/Y)	Volume (cm^3)	Num. of voxels
1	0.01	1
3/1	0.09	7
3/2	0.33	27
5/2	0.41	33
5/2.5	1.00	81
7/3	1.52	123
7/4	3.11	251
9/4	3.18	257
9/5	4.15	335

4.2.4 Healthy brain normalization

To assess the impact of healthy brain normalization on uptake repeatability, we compared the repeatability of non-normalized lesion SUV_{mean} to the repeatability of lesion SUV_{mean}, normalized by mean uptake in a healthy brain reference region. We used a range of different reference region volumes following in contralateral normal brain, employing the same neighborhood approach as in the previous section (range: 0.01 cm^3 to 22.7 cm^3). Since healthy brain encompassed a much

larger volume, we were able to use larger reference regions than within tumor contours. Therefore, we investigated additional crescent-shaped reference volumes, allowing for considerably larger reference volumes [126], some examples of which can be seen in figure 19.

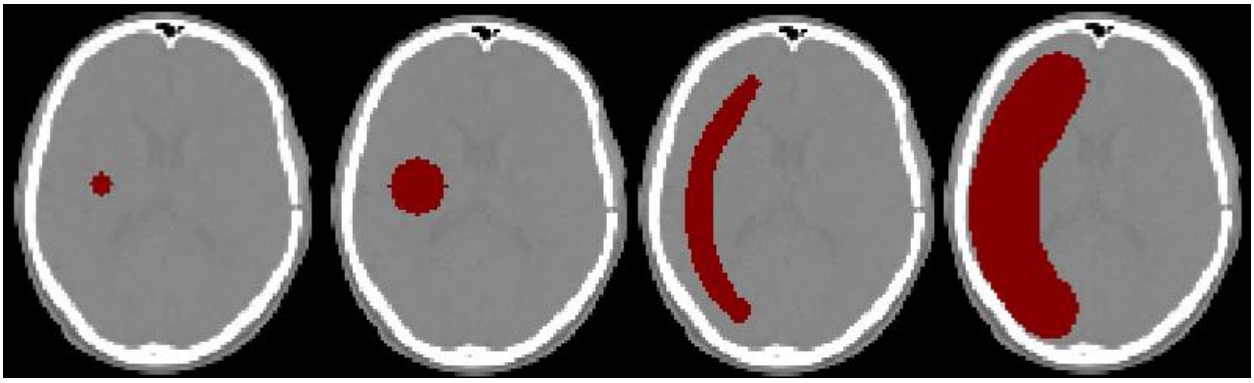


Figure 19: Examples of spherical and crescent-shaped volumes used in healthy tissue normalization.

4.3 Results

4.3.1 Patient population and ROI definition

Twenty-four patients were enrolled in the study, of whom nine consented to receive pre-treatment test-retest FET PET/CT scans, all of which were obtained within 6-9 days of each other. One patient had no visible increased uptake of FET in either test or retest. We selected the other eight for further analysis. Mean patient age was 53 years (range: 36-61) and there were 5 male and 3 female patients. We reconstructed all PET images using PSF reconstruction with 3 iterations and 24 subsets (3i24s) and smoothed each using a 4 mm Gaussian filter, as per Australasian Radiopharmaceutical Trials Network (ARTnet) recommendations to ensure semi-quantitative consistency in images.

We segmented the lesions with a background uptake-based thresholding method (1.6 x background uptake). The average tumor segmentation threshold was 1.62 g/ml [range: 1.08-2.66

g/ml]. Using these thresholds, we generated lesion contours with an average volume of 33.2 cm^3 (range 2.6 – 149.1 cm^3). The volumes, differences in volumes, dice coefficient and Hausdorff distances of these contours are shown in Table 2. Lesion volume varied considerably between test and retest. While average lesion volume changed by only 6% and was found not significant using a paired sample t-test, individual lesion volume change ranged from -57% to +87%. Average absolute volume difference was 32%, and average Dice coefficient between test and retest lesion volumes was relatively low at 0.66, with poorest matching observed in smallest volumes of slightly increased uptake surrounding areas of surgical resection. For larger disease volumes, matching between test and retest was better, reaching Dice coefficient values of up to 0.90.

Table 2: Tumor volumes for both test (1) and retest (2), volume differences, Dice coefficients, and Hausdorff distance for data used in the analysis.

Patient	Vol.[1] (cm^3)	Vol.[2] (cm^3)	Vol. diff (cm^3)	Vol. diff (%)	Dice coefficient	Haus. Dist. (cm)
1	47.82	56.53	8.70	18	0.90	0.14
2	2.96	5.43	2.48	84	0.39	0.37
3	117.66	149.06	31.41	27	0.62	0.36
4	32.67	16.21	-16.47	-50	0.66	0.39
5	29.08	29.34	0.26	1	0.83	0.17
6	2.59	1.10	-1.49	-57	0.28	0.43
7	5.29	5.30	0.01	0	0.83	0.10
8	13.74	17.20	3.45	25	0.81	0.15
Mean	31.48	35.02	3.55	6	0.66	0.26

4.3.2 FET repeatability

We evaluated the repeatability of FET SUV metrics by comparing values between test and retest scans. The summary of this analysis for the tumor volumes can be found in Table 3. We performed a similar analysis for the volumes of the healthy brain as shown in Table 4. We defined healthy volume outside joint contours of both test and retest imaging, therefore we used only one segmentation for both images. Because of that, SUV_{total} is equal to SUV_{mean} multiplied by the healthy brain volume and is not shown in the table below. For lesion-level SUV_{mean}, SUV_{max}, and $SUV_{98\%}$ metrics, we calculate LoA using test and retest contours across all patients, the results of which are shown in Table 5. Basic FET SUV metrics had limits of agreement around [0.80, 1.25] for both healthy and diseased intracranial volumes and for most SUV metrics.

Table 3: SUV metrics for tumor volumes: SUV mean, max, total, and 98th percentile. Secondary column labels indicate whether data refers to test (T) or retest (R) image.

Patient	SUV_{mean} (g/ml)		SUV_{max} (g/ml)		SUV_{total} (g/ml)		$SUV_{98\%}$ (g/ml)	
	T	R	T	R	T	R	T	R

1	2.72	2.91	4.78	5.40	129.99	164.38	4.31	4.48
2	1.25	1.15	1.56	1.34	3.69	6.23	1.46	1.28
3	1.87	1.46	4.06	3.80	219.57	217.79	3.07	2.70
4	2.60	3.65	5.48	5.33	84.89	59.15	4.98	5.08
5	2.01	2.07	4.07	4.02	58.44	60.74	3.35	3.13
6	2.60	2.71	3.23	3.00	6.72	2.99	3.13	2.98
7	1.84	1.64	2.66	2.43	9.73	8.71	2.46	2.32
8	2.32	2.77	3.35	4.04	31.83	47.58	3.18	3.75

Table 4: Volume and SUV metrics for healthy volumes: SUV mean, max, and 98th percentile. Secondary column labels indicate whether data refers to test (T) or retest (R) image.

Patient	**Volume (cm^3)**	**SUV_{mean} (g/ml)**		**SUV_{max} (g/ml)**		**$SUV_{98\%}$ (g/ml)**	
	T and R	**T**	**R**	**T**	**R**	**T**	**R**
1	1512	1.05	1.06	2.15	2.30	1.57	1.59
2	1390	0.77	0.72	1.99	2.10	1.14	1.11
3	1063	0.83	0.70	1.81	2.05	1.19	1.00
4	1375	0.89	0.92	2.34	2.10	1.35	1.35
5	1069	0.96	0.98	2.08	2.32	1.38	1.42
6	1265	1.35	1.52	2.46	2.49	1.91	2.09
7	1267	0.94	0.85	1.94	1.76	1.33	1.19
8	1358	1.12	1.34	2.52	2.26	1.59	1.89

Table 5: Limits of Agreement for several metrics for both lesion and healthy brain, calculated over all their respective voxel elements.

	Tumor			**Healthy brain**		
	Mean ratio	**CV**	**LoA**	**Mean ratio**	**CV**	**LoA**
SUV_{mean}	1.042	0.185	[0.72, 1.46]	1.007	0.114	[0.80,1.25]
SUV_{max}	0.991	0.118	[0.79,1.23]	1.011	0.097	[0.80,1.25]
$SUV_{98\%}$	0.978	0.102	[0.80,1.18]	1.006	0.106	[0.81,1.23]

4.3.3 SUV repeatability dependence on volume

We calculated values of $SUV_i^{NH}{}_{mean}$ over a range of neighborhood volume sizes, and calculated LoA for each patient for each of the neighborhoods considered. Resulting LoA are shown as the dashed gray lines in Figure 20. We calculated the mean LoA for $SUV_i^{NH}{}_{mean}$ across all patients for each neighborhood and we present the results in table 6 and by the solid blue line in Figure 4. We found single voxel SUV in tumor to have LoA of [0.76, 1.32], but for the largest neighborhoods considered, the LoA narrowed to [0.90,1.12].

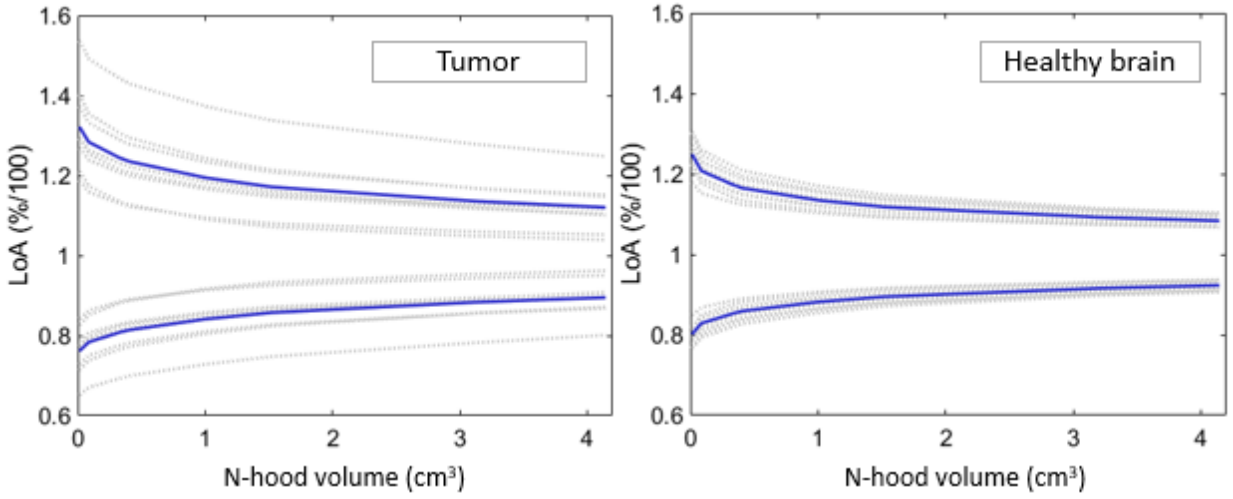


Figure 20: LoA with bias correction of SUV_{mean} of neighborhoods of varying volumes for both tumor volumes (left) and healthy brain (right).

Table 6: LoA of $SUV_i^{NH}{}_{mean}$ dependency on neighborhood (Nhood) size.

Nhood size (cm^3)	Nhood size (vox)	Mean LoA (tumor)	Mean LoA (healthy)
0.01	1	[0.76, 1.32]	[0.80, 1.25]
0.09	7	[0.78,1.28]	[0.83,1.21]
0.33	27	[0.81,1.24]	[0.85,1.17]
0.41	33	[0.81,1.23]	[0.86,1.16]
1.00	81	[0.84,1.19]	[0.88,1.13]

1.52	123	[0.86,1.17]	[0.89,1.12]
3.11	251	[0.88,1.14]	[0.91,1.09]
3.18	257	[0.88,1.13]	[0.92,1.09]
4.15	335	[0.90,1.12]	[0.92,1.08]

4.3.4 Healthy brain normalization

To explore the effect of healthy tissue uptake normalization on repeatability of lesion metrics, we calculated the LoA for lesion SUV_{mean}, SUV_{max}, and $SUV_{98\%}$ under two conditions: (1) calculating metrics from the original lesion SUV values in the PET image and (2) calculating metrics using the ratio of lesion SUV values to the mean SUV of healthy brain reference region volumes of varying sizes. We considered two kinds of reference volumes: smaller neighborhoods and larger crescent-shaped regions, as presented in Figure 19. The resulting LoA of normalized values are shown in figure 21.

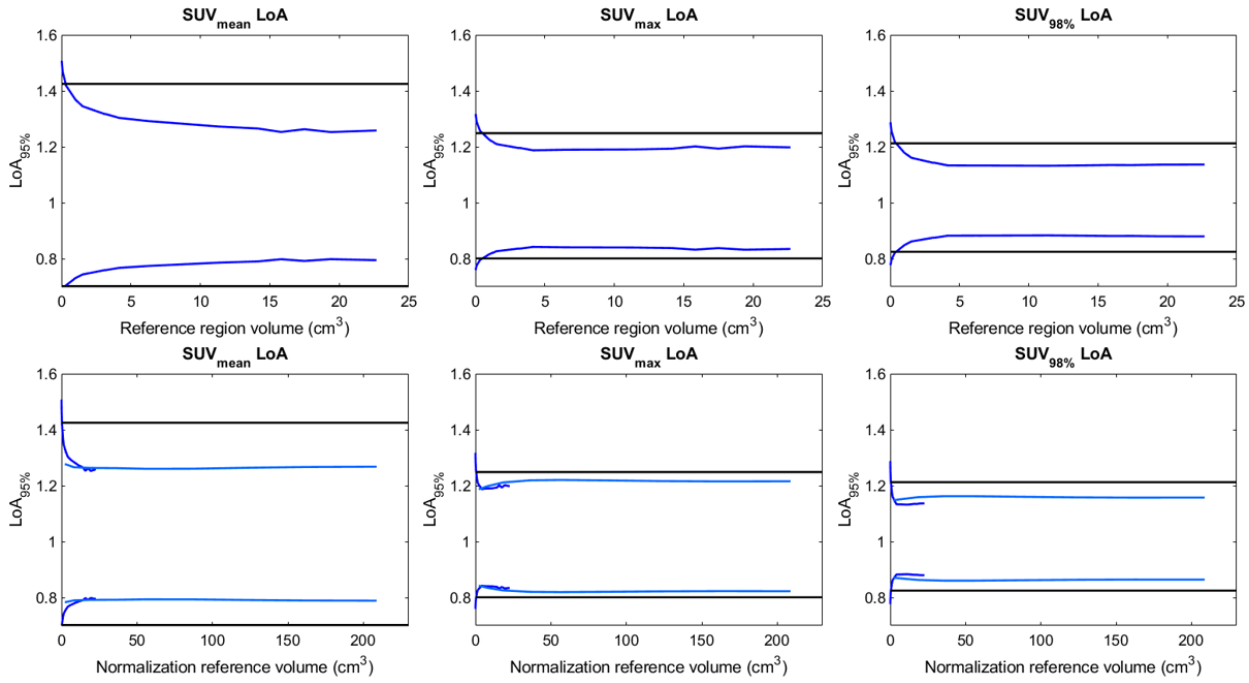


Figure 21: SUV LoA width dependence on reference region volume size. Dark blue lines represent values using smaller spherical reference regions, light blue lines represent values

normalized using larger crescent-shaped reference regions and black the reference non-normalized values, also shown in Table 5. For all three SUV metrics, normalization improved repeatability, demonstrated by the fact that the vertical distance between the two blue lines is narrower than the reference black lines. As seen on the bottom three plots with expanded X-axis, using much larger reference volumes did not seem to improve repeatability of lesion SUV metrics.

Normalization using single voxels or very small reference volumes resulted in poorer repeatability than using non-normalized values across all metrics. Normalization did improve repeatability for the tumor metrics starting at reference volumes of about 1 cm^3 and continued to improve it for all metrics up to about 5 cm^3. For SUV_{max} and SUV_{95}, we observed no improvement in repeatability for larger volumes. For SUV_{mean}, repeatability continued to improve up to reference volumes of about 15 cm^3, but no improvement was observed for volumes larger than that. We observed no improvement even when using much larger crescent shaped reference volumes of up to 200 cm^3. Results for SUV_{total} are just SUV_{mean} results scaled by healthy brain volume and are therefore not explicitly shown.

4.4 Discussion

4.4.1 Patient population and basic repeatability

In this chapter we present a quantification of imaging uncertainties in a previously unevaluated example – patients with glioblastoma imaged with FET PET/CT. Evaluating the repeatability of a PET tracer allows for the accurate differentiation of random uptake fluctuations from statistically significant changes in uptake and is therefore a critical step in tracer validation for response assessment. Knowing the magnitude of this variability is especially important in quantitative disease assessment of treatment, as it allows differentiation of significant changes from random

fluctuations. Additionally, it provides an insight on the intrinsic imaging uncertainties, necessary for probabilistic target descriptions.

Segmented lesion volumes varied considerably between test and retest, with an average absolute change of 32%. For reference, MRI-based 3D volume repeatability has found variation in tumor volumes of approximately 10% [127] when comparing gross tumor volumes before surgical resection. This large variation in volumes can be partially explained with the fact that thresholding as a segmentation method is sensitive to gradual changes in uptake. Even so, this is an important consideration for any FET quantitative analysis as it highlights considerable variability in FET uptake. This variability could also affect FET-based radiation therapy planning, although target expansion to PTV is expected to somewhat reduce these effects.

4.4.2 Basic FET repeatability

We found healthy brain measurements to be slightly more repeatable than those of tumor volumes. We found SUV_{mean} of tumor volumes to be least repeatable, although that could be caused by the segmentation method used. This result of [0.80, 1.25] is comparable to previous studies in other tracers in brain that found SUV_{mean} repeatability of 20%-30% for FDG [122] and 18-24% for FLT PET uptake variability [101]. Furthermore, we use this result in Chapter 4 to generate probabilistic tumor likelihood maps.

4.4.3 SUV repeatability dependence on volume

We evaluated the dependence of SUV repeatability on volume size. Averaging SUV over larger neighborhoods improved repeatability, however it was only for the largest neighborhoods when $SUV_i^{NH}{}_{mean}$ was calculated over hundreds of voxels. This result indicates that when evaluating whether a significant change of SUV occurred within a volume of interest, the size of the volume should be considered as well.

4.4.4 Healthy brain normalization

We also compared LoA between non-normalized whole-volume SUV metrics and metrics where uptake was normalized using healthy brain reference region uptake. Normalization using a single voxel did not improve repeatability for any metric compared to using non-normalized values, but repeatability did improve with increasing normalization volumes, up to a point. For SUV_{max} and $SUV_{95\%}$, repeatability continued to improve up to about 5 cm^3, after which the LoA plateaued with no noticeable improvement. For SUV_{mean}, the repeatability kept improving up to about 15 cm^3 and plateaued after that. Using much larger, crescent shaped reference volumes ranging up to 200 cm^3 did not improve repeatability compared to these volumes. This is an important consideration for future studies using FET PET/CT.

4.4.5 Limitations

We performed this analysis on eight patients with data suitable for analysis, a number that could not be increased, due to cessation of patient accrual. The patient scans were also acquired in the weeks post-surgery, which could introduce undesired effects in FET uptake and may not be representative of repeatability in other clinical scenarios such as pre-operatively or in assessment of recurrence. Most lesions appeared larger on retest images seven days after test, although these changes were found not to be statistically significant in the current population. Finally, we registered the images to each other using direct rigid registration, which we deemed appropriate for the constrained intracranial volumes, but it could still affect the exact results.

4.5 Conclusions

In this chapter we demonstrated a quantification of imaging uncertainties in an example dataset where these uncertainties were previously unknown. As seen in figure 7, quantification of these uncertainties is necessary to obtain non-binary, probabilistic gross tumor likelihood maps, which we can in turn use in probabilistic radiotherapy planning. We found FET uptake limits of

agreement (LoA) to be around [0.80, 1.25] for whole-volume metrics. We quantified SUV LoA dependence on volume size for both healthy brain and tumor volumes and we found that uptake normalization by healthy tissue uptake should be performed using a reference volume of at least 15 cm^3.

5 PTP Implementation – Step 2 – Treatment Planning

The work we showed in previous chapter is an example of direct quantification of uncertainties arising in medical imaging, the first and crucial step in incorporating uncertainties in a probabilistic treatment planning approach. While the results are specific to imaging with FET PET, it is possible to perform such quantification of uncertainties for any imaging modality or find it in the already published literature. We also refer to literature for evaluating the magnitude of other uncertainties in the “Quantifying uncertainties” section of figure 7, since while they each represent an interesting scientific question, the quantification of all categories of uncertainties is outside the scope of this thesis.

In this chapter we continue the analysis of the dataset shown in the previous chapter and build on the results of imaging uncertainties obtained there. But where Chapter 4 focused primarily on the quantifying uncertainties part of figure 7, Chapter 5 is concerned with those uncertainties can be used for probabilistic radiotherapy planning, or “planning” section of figure 7. While these two chapters could be considered part of the same unit, they were split into two parts due to their different foci and research questions.

This chapter focuses in greater detail on the “Planning” section of figure 7. We use the imaging uncertainties obtained in the previous chapter to define gross tumor likelihood maps and explore how they can be used to calculate probabilistic maps of microscopic disease. We use these maps to define dose prescriptions for our example cases and introduce the probabilistic optimization engine that we implemented into the in-house treatment planning system WiscPlan. We create dose plans using the classical planning approach to establish a baseline for comparisons, RO methodology to validate the RO implementation, and PTP using the probabilistic tumor likelihood maps to evaluate the proposed approach. We compare dose plans obtained using RO and PTP to classically optimized dose plans obtained using the existing treatment planning paradigm.

To summarize, in this chapter we aim to validate whether the proposed PTP workflow can produce clinically feasible dose plans and to evaluate how such plans compare to the classical treatment planning approaches. It is important to note that the goal of this work was not to create identical dose plans with the different planning approaches, as the optimizations are fundamentally different, and differences in results is to be expected. Rather, general agreement and performance was sought.

5.1 Motivation

The purpose of this chapter was to develop and evaluate the feasibility of a probabilistic planning approach for RT treatment optimization. We aimed to develop a new treatment paradigm which directly uses patient's disease information in treatment plan optimization. The results obtained in Chapter 3 allow us to incorporate the imaging uncertainty information directly in treatment planning. Unlike in dose painting, in this approach we modulate the optimization weights using disease information, rather than change the value of prescribed dose.

5.2 Materials and methods

5.2.1 Volume management in GBM

5.2.1.1 Classical treatment planning approach

To allow dose planning using classical, RO and PTP approaches, we created multiple treatment volumes in the GBM patients. For the classical treatment planning approach, we segmented the baseline FET PET image using a thresholding segmentation approach using 1.6x healthy brain uptake as the threshold to obtain the GTV. We then expanded the GTV 10 mm intracranially to account for microscopic infiltration and obtain the CTV, as per relevant literature [20], [21]. Lastly, we expanded the CTV isometrically by 2 mm to account for positioning and delivery uncertainties and obtain the PTV, which was used in dose prescription and optimization for the classical

approach. For the second approach, we used the CTV for dose prescription and optimization since RO accounts for positioning uncertainties directly.

5.2.1.2 New treatment planning approach

Using the FET imaging repeatability results introduced in the previous chapter, we first converted the PET/CT image into gTLMs, using the normal cumulative distribution function (*norm*).

$$gTLM_i = norm\left(\frac{SUV_i - T}{\sigma \cdot T}\right) \quad (5.1)$$

Where $gTLM_i$ represents the value of gTLM in voxel *i*, *SUVi* is the standardized uptake value for voxel *i*, *T* is the threshold, calculated as 1.6x average healthy brain uptake in each patient, and σ is the standard deviation of uptake calculated from the limits of agreement we estimated in the previous chapter and shown in table 5. The cumulative distribution function, by definition, returns the probability of a value (SUV_i, in this case) being higher than a reference (T in this case) with standard deviation σ. As such it provides a direct translation of uptake to likelihood of disease presence.

For this gTLM generation, we used only intracranial volumes within 1 cm of the initially defined disease volume, as described in Chapter 3, to avoid TLM marking in other areas which demonstrated increased uptake. Most commonly this occurred close to the skin near the disease site, presumably due to healing after surgery. An example of increased uptake and its handling can be seen in figure 22.

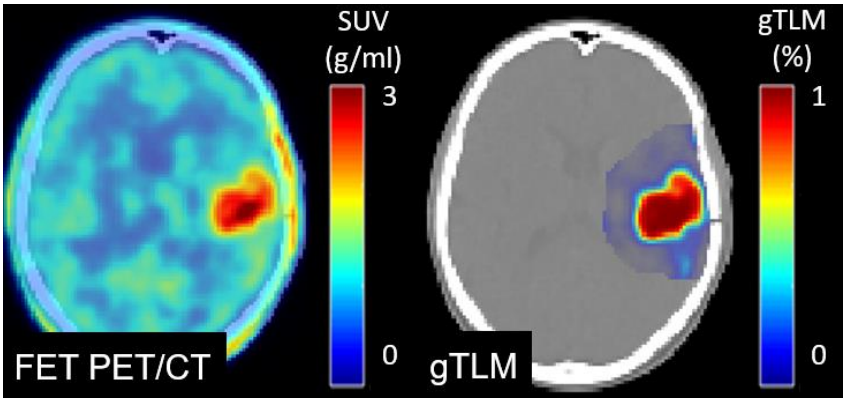


Figure 22: Example conversion of FET PET imaging to gTLM using equation 5.1. Note that only intracranial volume within 1 cm of the initially defined tumor volume was used. While a large part of the tumor is mapped with likelihood close to 100%, an anterior (top of the image) part of the disease receives less confidence and is colored with a more yellow-orange color in the gTLM.

While some artefacts and sharp boundaries can be seen on the gTLM image in figure 22, these effects were additionally smoothed out during calculation of microscopic infiltration and did not end up affecting the treatment significantly.

5.2.1.3 Microscopic tumor likelihood maps (miTLM)

For our work, we assumed an infiltration constant of 15 mm, based on previous results and margin recommendations in gliomas [19]–[21], [105], [128]. Using the approach described in section 3.2.2, we observed better conformity of gTLM+miTLM to anatomical constraints to anatomical constraints than other applied methods, as shown in figure 23. For implementation purposes, we generated a “Target Map” (TM) by applying thresholding to the TLM at p=0.05 value. We used this TM to prescribe the voxel-weighted dose prescription. While this does mean that volumes under 5% disease likelihood were ignored in optimization, this value is considerably lower than the levels where dose increases were observed in PTP optimization.

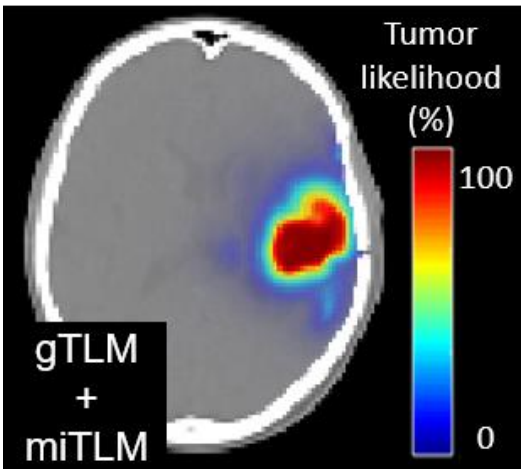


Figure 23: Axial slice of miTLM as calculated for the example patient shown in figure 22. A large part of the tumor is mapped with a likelihood of disease presence close to 100%, reflecting high FET uptake in those volumes. While all edges exhibit a gradual descent reflecting the likelihood of infiltration, the anterior (top of the tumor in figure) part of the tumor maintains the comparatively lower likelihood values. Note that for this case, microscopic infiltration was limited to intracranial volumes to ensure better comparability between planning approaches, so no infiltration in the gap in the skull is permitted.

5.2.2 Dose prescription

We evaluated three different dose planning approaches for each patient to validate the RO implementation and compare PTP to the classical planning approach. We set the optimization goal to 60 Gy to target while minimizing dose to healthy brain volumes. For the classical planning, we prescribed the dose to the PTV volume. For the RO approach, we assigned the dose to the CTV directly and positioning uncertainties were accounted for using RO parameters. For PTP, we assigned the dose to the entire TM volume and weighted each voxel by its respective TM value.

5.2.3 Dose optimization

We calculated intensity modulated radiation therapy (IMRT) dose plans using an in-house treatment planning software WiscPlan [66], [94] for TomoTherapy delivery systems. We calculated the dose for a helical delivery plan using 64 binary MLCs, 51 angles for the X-ray source position per rotation, a pitch of 0.86 and a 2.5 cm jaw width. We excluded dose calculation for beamlets that did not intersect the CTV expanded by 4 mm (the classical PTV), and the weights corresponding to those beamlets were omitted in calculations to improve tractability.

We performed the classical planning using the existing WiscPlan optimization engine [66], and for RO and PTP planning, we expanded the treatment planning software with probabilistic robust optimization functionality for setup uncertainties, using an objective-level minimax optimization [80], [94]. To find the function minimum, we used the constrained nonlinear multivariable function minimum search (*fmincon*), implemented in Matlab 2018b (The MathWorks, Inc., Natick, MA, United States). We performed all optimizations on a system with a 4-core Intel i5-7500 CPU (3.40 GHz) and 16 GB of RAM.

5.2.4 Dose plan comparison

We compared dose plan performance between different planning approaches using different dose metrics. Due to the probabilistic nature of RO and PTP, we modified some traditionally used dose metrics to better capture plan performance under uncertainty as presented in Chapter 3.

Aside from the commonly used dose benchmarks for various volumes of interest, we measured a range of expected values across all scenarios considered, to provide the range of possible results within the considered uncertainty range. This provided an estimate to the expected range of results, as well as some insight on whether the distribution of the metric tends to be approximately one-sided or symmetrical. We evaluated the performance of dose plans using eight dose benchmarks: average dose (D_{mean}) for both CTV and GTV, the 95^{th} and 5^{th} dose percentile

(D_{95} and D_{05}) for both CTV and GTV, and average dose and volume receiving at least 55 Gy (V_{55}). We quantified the differences between dose metrics of classical and RO approaches and between classical and PTP approaches and tested them for statistical significance using one-sided student's T-test. The tests were one-sided, testing for "plan worsening" in comparison to the baseline classical planning – that is decreases in target dose metrics or increases in healthy brain dose metrics. While we performed multiple tests, we did not use p-value correction for multiple hypothesis testing. This results in more conservative test results that are not dependent on the number of metrics being evaluated.

We used dose-volume histogram clouds (DVH clouds), where a range of values was visualized using an additional band rather than a single line [79], [129]. The main line represents plan performance in the nominal case, and the band represents the range of DVH values across all scenarios considered, as introduced in Chapter 3.

5.3 Results

To validate the implementation of the RO algorithm and test the validity of PTP planning, we performed quantitative comparisons of dose plans generated using the three optimization approaches. Dose plans of the three planning approaches for a representative case can be seen in figure 24.

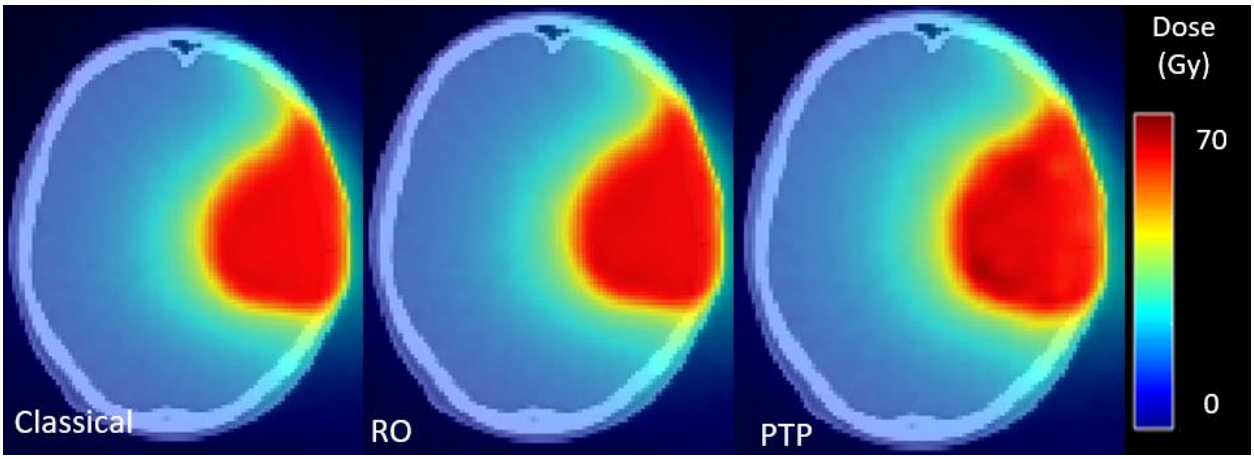


Figure 24: Axial dose plan slices of a representative patient with glioblastoma. RO and classical optimization both produce similar plans, with PTP resulting in somewhat more heterogeneous results.

We observed very similar performance between the classical and RO approaches, with comparatively higher target dose heterogeneity observed in PTP. This is also reflected in the dose-volume histogram (DVH) clouds, as shown in figure 25 and figure 26.

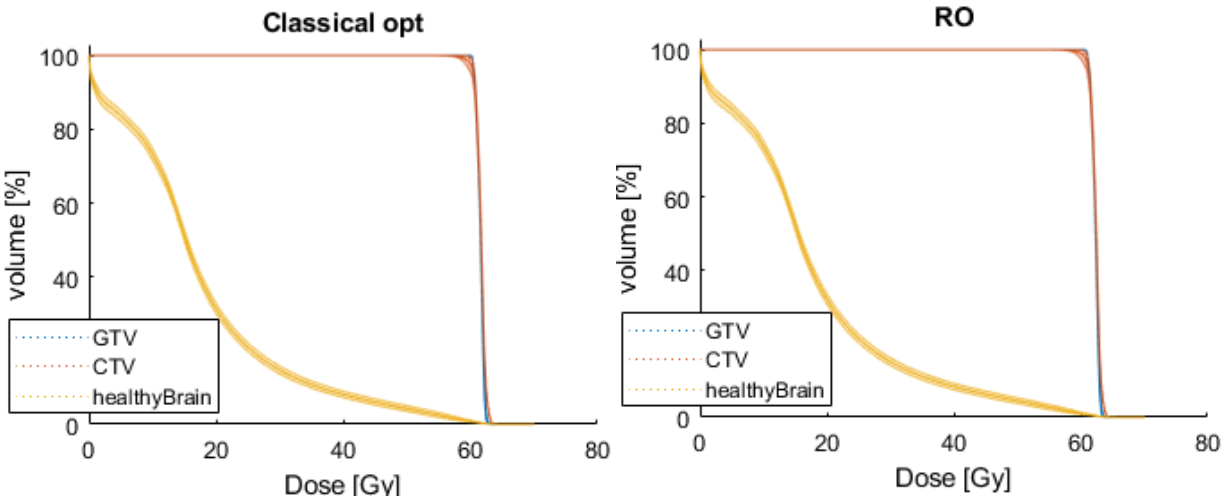


Figure 25: DVH cloud comparison between classical and RO plans for a single example patient. Very similar performance is observed between these two approaches, with only minutely increased target and OAR doses.

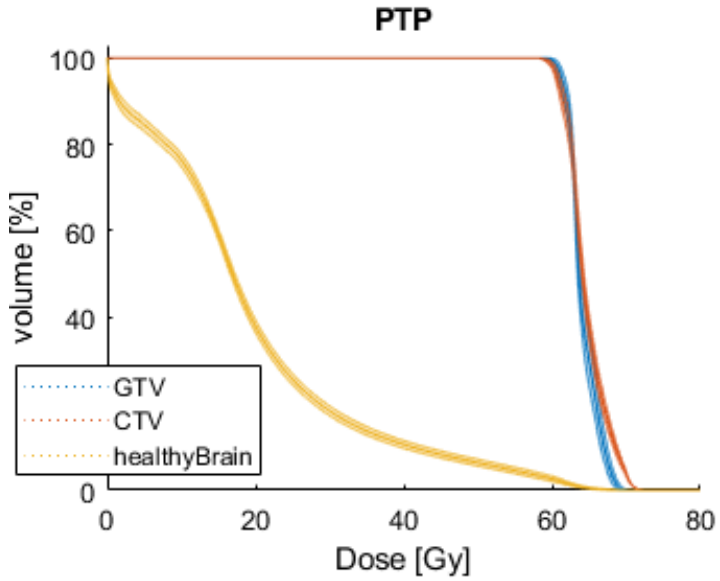


Figure 26: DVH cloud for the PTP dose plan for the example patient. Higher dose heterogeneity, in particular higher maximum dose is observed, but dose to healthy brain remains similar to classical planning.

While DVH plots of individual cases illustrate the behavior of optimization results, they do not capture overall trends. For this purpose, we aggregated the planning results in table 7.

Table 7: Summary of average dose metrics for the three planning approaches in the format: (average [range]).

Metric \ Plan	**Classical**	**RO**	**PTP**
CTV D_{95} (Gy)	60.1 [59.0-61.3]	60.1 [58.8-62.1]	57.7 [38.2-67.0]

CTV D_{mean} (Gy)	61.1 [60.5-61.8]	62.1 [60.7-66.6]	62.9 [52.1-69.8]
CTV D_{05} (Gy)	62.1 [61.6-63.1]	64.3 [61.9-73.4]	67.9 [62.6-74.6]
GTV D_{95} (Gy)	60.5 [59.8-61.3]	61.0 [59.2-63.7]	61.6 [54.1-67.3]
GTV D_{mean} (Gy)	61.1 [60.4-61.8]	62.1 [59.9-66.7]	64.1 [58.7-68.5]
GTV D_{05} (Gy)	61.9 [61.0-62.9]	64.0 [61.3-72.7]	67.0 [63.5-70.5]
Brain V_{55} (%)	18.1 [11.3-28.4]	18.5 [11.1-31.4]	19.5 [6.75-32.3]
Brain D_{mean} (Gy)	3.45 [2.26-6.16]	4.21 [1.99-11.8]	5.44 [0.138-12.9]

We also evaluated the differences in dose metrics between classical and RO plans using one-sided T-tests against the null hypothesis that RO and PTP perform worse, either by decreasing target dose, or increasing dose to healthy brain. As shown in table 8, no statistically significant differences were observed.

Table 8: Evaluation of statistical differences between different planning approaches. One-sided T-tests were used for these evaluations, testing for "plan worsening" in RO and PTP approaches – increases in OAR doses or decreases in target doses. These p-values were not adjusted for multiple test hypothesis, even so, no significant differences are observed.

Metric \ Plan	**Classical v RO (p-value)**	**Classical v PTP (p-value)**
CTV D_{95}	0.47	0.25
CTV D_{mean}	0.87	0.80
CTV D_{05}	0.92	1.00
GTV D_{95}	0.79	0.75
GTV D_{mean}	0.90	0.98
GTV D_{05}	0.93	1.00

Brain V_{55}	0.15	0.15
Brain D_{mean}	0.16	0.08

5.4 Discussion

In this chapter, we implemented and validated a new planning optimization approach which can directly incorporate probabilistic imaging disease likelihood information. To evaluate whether it achieves reasonable results, we aimed to develop and deploy a prototype PTP implementation and validate it on a dataset with simple geometry.

As our practical example, we chose a cohort of patients with glioblastoma imaged with FET PET as an example to perform this evaluation. We used previously established limits of repeatability of FET PET imaging to include imaging uncertainty information into target definition using TM. We then implemented these TMs directly in treatment planning using a voxel-weighted optimization function that used TM voxel values in its objective function. Optimizations using RO methodology with classically-determined volumes performed similarly to optimizations using the classical approach. Achieving equivalent results between these two approaches shows that our implementation of the RO algorithm behaves as expected and can produce feasible dose plans.

Comparison between classical and PTP plans shows no significant decreases in plan quality, as measured using one-sided t-tests to evaluate worsening (i.e., decrease in target dose or increase in OAR dose). While we sought general agreement, we should not expect identical dose plans with these planning approaches, since the objective functions used are fundamentally different.

Generation of gTLM is not limited to the normal cumulative distribution function, as shown in this example. We chose this approach since the cumulative normal distribution is the direct translation of known baseline variability to a probabilistic tumor map capturing this uncertainty. Other ways

that we could generate gTLMs include fuzzy logic merging of multiple segmentation methods or machine learning approaches.

This chapter aims to demonstrate the general idea and feasibility of the PTP approach, and as such several assumptions were made. First, the parameter of infiltration, while taken from previously published works, is subject to uncertainties. In cases of glioblastoma infiltration, the actual speed and range of infiltration can vary considerably [20]. Moreover, infiltration is not equal in all directions, as was the assumption for our work, but tends to follow the neural pathways and anatomical section of the brain. However, increased knowledge of disease likelihood and infiltration would not fundamentally change the planning approach, as it can be easily and directly incorporated into the workflow.

Secondly, scenario selection is an important aspect of probabilistic treatment planning, and multiple different methods can be used. In our case we chose a “good practice scenario selection” approach, where the geometry is uniformly shifted in the six cardinal directions: along the craniocaudal, anteroposterior and frontal axes. We chose this approach due to its ease of interpretation and straightforward equivalence to the classical approach of expanding the target volume by margins. Other possible scenario selection methods include alternate approaches such as “statistically sound scenario selection” or “statistically sound dosimetric selection” [95]. While one can argue that such scenario selection does not cover all possible, or even plausible, realizations of positioning uncertainties, it is important to remember that using the commonly used van Herk formula for PTV margins also results in coverage that is sufficient for 90% of all patients, not all of them.

Finally, while the examples on patients with glioblastoma presented here serve as a simple demonstration of the proposed treatment planning approach, they are a comparatively simple geometry to other disease sites. As the brain is encased in a hard, well visible structure,

positioning uncertainties are small. Here, 2 mm was used, which is typically narrower than the dose falloff at the edge of target and does not play as important of a role as microscopic infiltration. Additionally, only two target volumes (miTLM/CTV and gTLM/GTV) and a single OAR (healthy brain) were used. Lastly, applying PTP to a different brain tumor type may affect the microscopic disease presence, but the approach of microscopic disease calculation would remain conceptually the same.

These limitations, while worth considering, do not seriously affect the proposed treatment planning approach itself. However, PTP should be applied to and evaluated on other disease sites to verify its consistent performance under different planning conditions.

5.5 Conclusions

In this work we introduced a PTP optimization approach where tumor likelihood information based on imaging uncertainties quantified in Chapter 3 is used directly in optimization to create dose plans. We have demonstrated that this probabilistic target definition in combination with known probabilistic optimization produces feasible clinical plans in patients with glioblastoma, based on baseline FET PET imaging. While PTP produced plans with somewhat wider dose spread in target volumes, the dose metrics were not significantly different, and plans were overall comparable. Target volumes equivalent to the ones used in classical planning can be used, resulting in plans comparable to the classic planning approach, thus ensuring back-compatibility. Due to its non-reliance on expansion by margins, PTP provides a framework for more complex dose planning and better use of imaging information that could potentially improve dose plans in more complex geometries. We explore one such geometry in greater detail in the next chapter.

6 PTP Applications - Prostate

In the previous chapters, we established probabilistic target definition and planning on a dataset of patients with glioblastoma. While performing treatment planning starting at the quantification of imaging uncertainties is helpful for complete understanding of main steps in probabilistic treatment planning, always performing such an analysis is impractical, time consuming, and in most cases simply not possible, unless the patients receive test-retest imaging. For these reasons, it is crucial that the proposed PTP approach be independent of the availability of such data. Additionally, there are unique challenges associated with different anatomical sites in RT. Treatment planning for patients with glioblastoma simplifies several aspects of the process due to its simple geometry and low positioning uncertainties, and while planning for brain tumors serves as a proof-of-concept, we also want to evaluate the performance of any new planning approach in other disease sites.

For this reason, we sought to expand our analysis to a larger dataset with more anatomical variability and a different source of tumor mapping. We selected a cohort of patients with prostate cancer and with tumor likelihood maps generated with multiparametric magnetic resonance imaging. Not only did this dataset present an opportunity for treatment planning in a different anatomy, but also prescription and planning of integrated dose boost plans, due to the specific nature of the disease and existing clinical studies.

To summarize, the purpose of the work we present in this chapter is to expand the PTP approach to more complex clinical sites and to evaluate the opportunity in dose distributions obtained with the PTP approach.

6.1 Motivation

The purpose of this study was to expand the evaluation of the probabilistic target definition and planning (PTP) introduced in the previous chapter. We sought to evaluate it on a more complex geometry of patients with prostate cancer and physician contours of several volumes of interest. While brain lesions are encased in a hard skull and rather homogeneous tissue, the prostate is located in a more varied anatomy and in close proximity to bladder and rectum – organs at risk that should be preferentially spared over other nearby tissues.

Multiparametric magnetic resonance imaging (MP-MRI) derived tumor likelihood maps were used for probabilistic target definition. In addition, performance was evaluated for two different dose prescriptions: (1) for cases with uniform dose prescription and (2) for dose plans containing an integrated dose boost to the prostate subvolumes with higher disease presence.

While the effects of heterogeneous dose prescriptions in different disease sites are still being investigated, previous works in prostate cancer, such as the FLAME trial, have shown the benefits of focal lesion ablative micro boost, where specific tumor sub-volumes are treated with higher doses than the rest of the target volume [130]. The results of the FLAME trial, conducted on 100 patients, have shown that such boosts can be achieved without an increase in dose toxicity [131] and with an improvement in disease free survival [132]. These findings demonstrate the potential benefits of non-uniform dose planning – a well-established concept, but one that is still cumbersome to use when dealing with different prescription dose levels [7], [52].

6.2 Materials and methods

We applied the PTP approach, on a dataset of patients with prostate cancer. We calculated dose plans using both the classical ICRU approach (using the GTV-CTV-PTV volumes), and the proposed PTP approach using probabilistic tumor likelihood maps (TLM). For both planning approaches, a uniform dose plan and a plan containing a dose boost to the volumes of high

disease presence were calculated, resulting in four dose plans per patient. We evaluated comparative performance in both uniform and boost plans between the two planning approaches using dose benchmarks, presented in more detail below.

6.2.1 Patient cohort

Patients with biopsy proven T1-T2 prostate cancer (median age 67, range 54-71 years) were considered for this analysis. The image data for these patients was acquired as part of a previously published investigation [133]. Patients underwent MP-MRI examinations on a 3.0 T Philips Achieva MR scanner (Philips Healthcare, Best, the Netherlands) at the Netherlands Cancer Institute (NKI). An mDIXON sequence was scanned, from which a pseudo-CT scan was derived using the MRCAT algorithm (Phillips Medical Systems MR Finland Vantaa, Finland). A total of 30 image features from T2w, ADC, and K^{trans} were combined to derive a voxel-level gross tumor likelihood map (gTLM) for the prostate using a logistic regression model [134], [135]. The study was approved by the ethics committee of the Netherlands Cancer Institute, and patients provided written informed consent.

6.2.2 Planning volumes

We created planning volumes based on patient data, and an experienced radiation therapist manually segmented relevant volumes (prostate, OAR) . To account for microscopic disease infiltration, we expanded MP-MRI derived gTLMs using a sequential probabilistic region growing algorithm [136] to obtain microscopic tumor likelihood maps (miTLM). We set the range of infiltration parameter to 3 mm, based on existing histological studies evaluating range of extracapsular infiltration [6]. To compensate for the low sensitivity of MP-MRI [137]–[139], we created a Target Map (TM) by additionally setting the TM value within the prostate volume to 0.7 to ensure sufficient coverage within the prostate gland. This TM creation process can be seen for an example patient in figure 27.

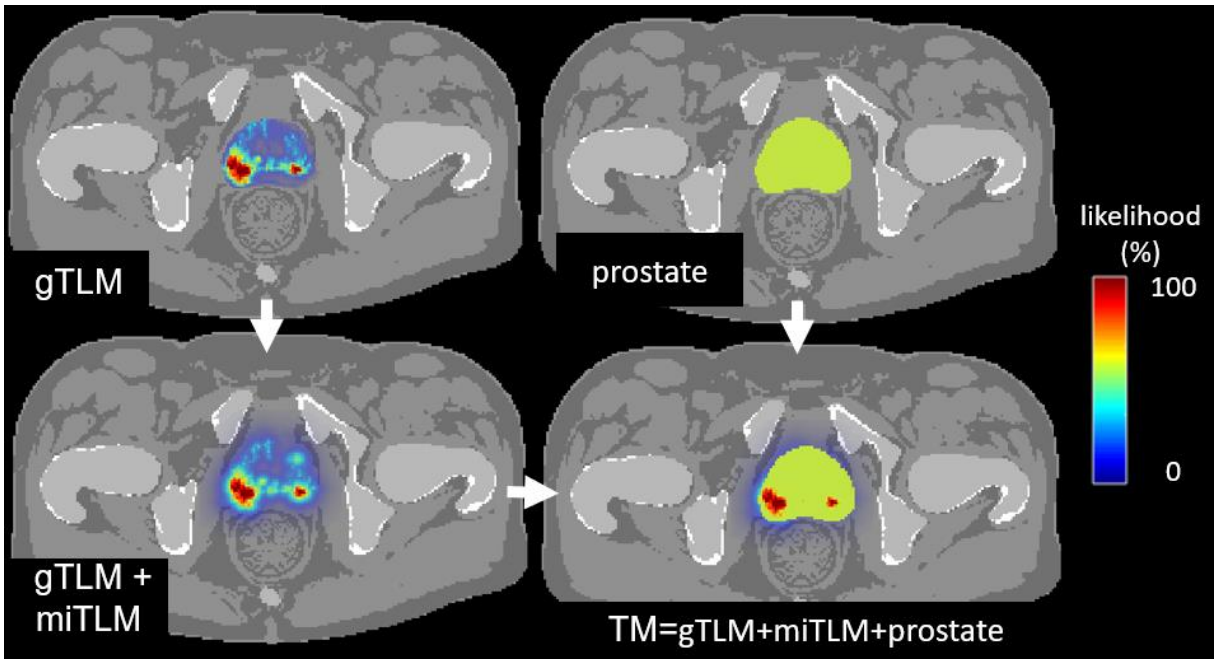


Figure 27: Axial slices of MR-derived CT overlaid with different disease mappings. Top left: initial MP-MRI derived tumor likelihood map (gTLM). Bottom left: gTLM with expected microscopic infiltration (miTLM). Note that unlike Gaussian smoothing, this algorithm does not decrease tumor likelihood values within the TLM. Top right: prostate volume. Bottom right: Target Map (TM) that combines the information of visible and microscopic disease as well as baseline prostate volume and is used for planning. Areas of high disease presence and some suspected microscopic extension outside the prostate can be seen.

We used these TM directly in the optimization for PTP optimization as described in the previous chapters. For the classical approach, we defined the CTV empirically as the volume where TM probabilities were greater than p=0.30 and the PTV as a 4 mm isometric expansion around the CTV [140].

For integrated dose boost plans, we first smoothed the gTLM using a 3 mm Gaussian filter to obtain "TM_{boost}", as seen in figure 28. As each TomoTherapy multileaf collimator corresponds to

a beam width of just over 6 mm at isocenter, this smoothing reduced the spatial variation observed in the MRI to a level comparable to the resolution of a treatment delivery system, while still being fine enough to retain target sub-volume heterogeneity. For the PTP boost plans, we used TM_{boost} directly as described in Chapter 3.

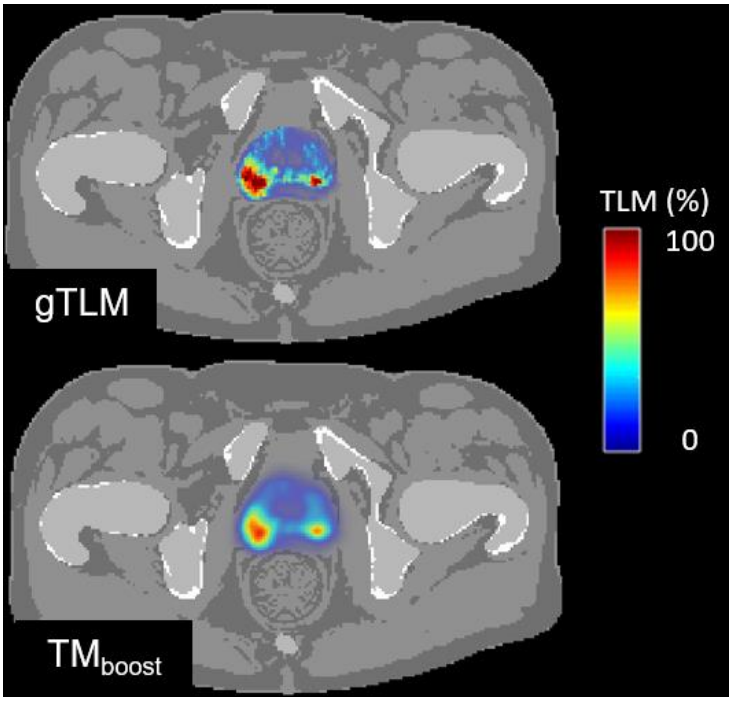


Figure 28: Axial slices of target map used for dose boosting (TM_{boost}). In this case, target volume was smoothed using a Gaussian filter to reduce the spatial variation observed in the MRI to a level comparable to the resolution of a treatment delivery system, while still being fine enough to retain target sub-volume heterogeneity.

For the classical boost planning, we created gross tumor volumes (GTV) using a TM_{boost} threshold of $p>0.30$, followed by a 4 mm expansion to account for positioning uncertainties. An example representation of all described volumes can be viewed in figure 26.

6.2.3 Dose prescription and parametrization

We calculated four dose plans for each patient, to compare classical and PTP planning approaches.

The *Classical uniform* approach was a plan where we prescribed a dose of 77 Gy to the PTV. We aimed at limiting the dose to the rectum and bladder so that the maximum dose to both organs was below 77 Gy. We added a secondary objective to try and minimize the dose to these organs with a comparatively small optimization weight compared to the other objectives.

The second optimization plan was a *PTP uniform* approach, where we prescribed the same dose of 77 Gy to the entire volume of the prostate and the seminal vesicles, however in this case the prescriptions were voxel weighted by the TM values as shown in equations 3.1 - 3.4. We used an objective-level penalty function, meaning that each optimization goal was evaluated separately in all scenarios, and the worst performing one (the one with the largest penalty) was selected before summing the penalties from different optimization goals. This approach results in a more intuitive balancing of competing optimization objectives than summing the penalties from different optimization objectives first. We accounted for the uncertainties in patient positioning and dose delivery using a minimax approach in optimization over scenarios of patient misplacement of 4 mm, equivalent to the PTV margins used in the first plan. We kept the dose constraints to the OAR equivalent to the Classical uniform planning approach.

The third optimization plan was a *Classical boost* approach. This plan was identical to the first (Classical uniform) approach, except that the maximum dose constraints to the CTV were lowered, and we set the desired dose within the GTV to 95 Gy, as per the FLAME trial [141].

The fourth optimization plan was a *PTP boost* dose plan. Here, we use the 77 Gy baseline from PTP Uniform case and add a prescription for 95 Gy dose boost to the entire target volume, using

a voxel-level penalty based on the TM_{boost} in each voxel. We handled the OAR prescriptions and uncertainties in patient positioning and dose delivery in the same way as for PTP uniform plan. Example volumes for the prescriptions can be seen in figure 29.

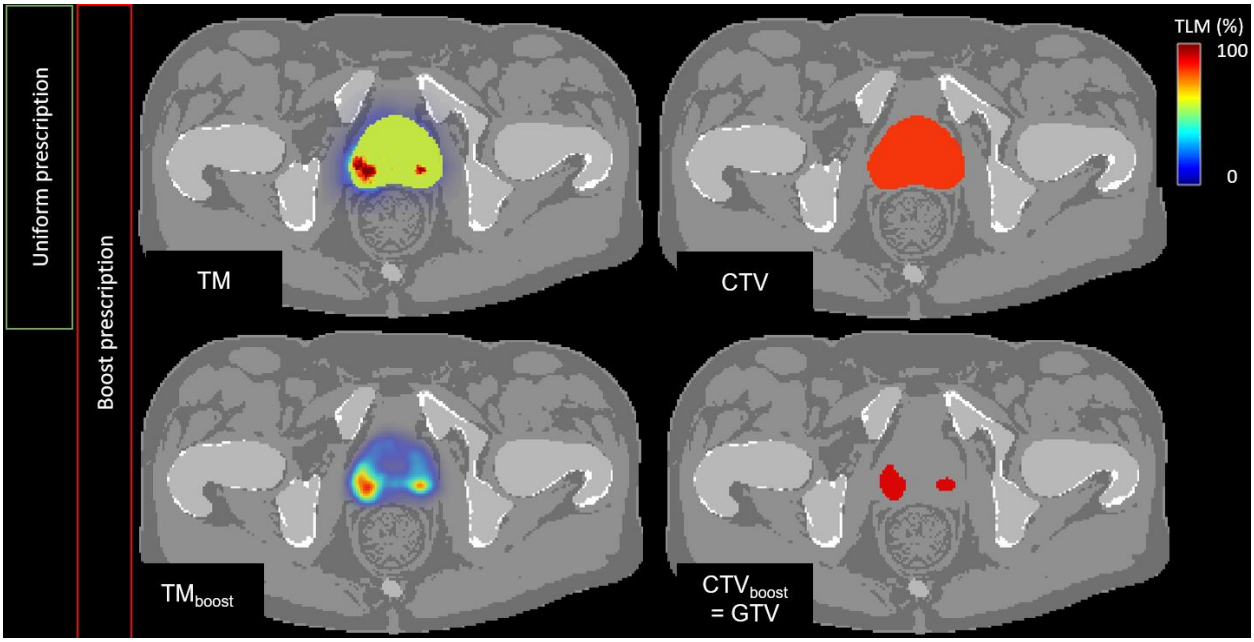


Figure 29: Comparison of probabilistic target volume (TM and TM_{boost}) (left) and their analogous classical volumes (CTV and CTV_{boost}/GTV) for the example patient (right).

6.2.4 Implementation and dose optimization

We calculated intensity modulated radiation therapy (IMRT) dose plans using an in-house treatment planning software WiscPlan [66], [94] for TomoTherapy delivery systems. We calculated dose for a helical delivery plan using 64 binary multileaf collimators, 51 angles for the X-ray source position per rotation, a pitch of 0.86 and a 2.5 cm jaw width. We excluded dose calculation for beamlets that did not intersect the CTV expanded by 4 mm (the classical PTV) and omitted the weights corresponding to those beamlets in calculations to keep computational burden within machine capabilities.

In the same manner as described in the previous chapter, we performed the classical planning using the existing WiscPlan treatment planning software[66]. For PTP, we expanded the treatment planning software with probabilistic robust optimization functionality for setup uncertainties, using an objective-level minimax optimization [80], [94]. We used a “good practice scenario selection” [95] approach, where six scenarios of positioning shifts along the main axes are used, along the “nominal” scenario using unchanged geometry. While the use of higher number of scenarios selected via statistically sound methods may improve robustness of results [95], a simpler method had to be used in our work due to computational resource limitations. To find the function minimum, we used the constrained nonlinear multivariable function minimum search (*fmincon*), in Matlab 2018b (The MathWorks, Inc., Natick, MA, United States). We performed all optimizations on a system with a 4-core Intel i5-7500 CPU (3.40 GHz) and 16 GB of RAM.

We set a maximum modulation factor of 5 for all optimizations. We used a static dose cloud approximation to reduce the computational requirements and to avoid introducing additional variability between the two planning approaches.

We used equivalent optimization weights for plans optimized with the classical and with the PTP approaches. The optimization process ran until step size decreased below a preset threshold, or for a maximum of 200 iterations. We observed optimization convergence to a local minimum in all optimizations where the computation did not converge within this frame.

6.2.5 Dose plan comparison

We first evaluated dose plans qualitatively using dose-volume histogram clouds (DVHs) and quantitatively using dose benchmarks for four volumes of interest. For the CTV and GTV, we evaluated average dose, and 95^{th} and 5^{th} dose percentile (D_{mean}, D_{95} and D_{05}). For bladder and rectum, we evaluated average dose (D_{mean}) and fraction of volume receiving at least 77 Gy (V_{77Gy}).

To compare planning approaches, we calculated the dose benchmarks for each planning approach in the nominal scenario and compared in pairs: classical uniform and PTP uniform plans; and classical boost and PTP boost plans. Two-sided paired t-tests were used to test the PTP approach for significant differences. We used bonferroni correction to account for multiple hypothesis being tested, and considered *p*-values of less than 0.05 to be significant.

6.3 Results

We selected twelve patients for planning, resulting in 48 total dose plans. For each patient, target volume and GTV volume are shown in table 1.

Table 9: Number of beamlets, size of combined volumes of prostate and suspected microscopic infiltration (Target volume), and volume with high disease likelihood (GTV).

Patient	Target volume (cm^3)	GTV volume (cm^3)
1	67.3	3.6
2	37.2	1.7
3	104.2	5.4
4	27.2	7.2
5	53.4	2.8
6	39.8	5.1
7	52.6	1.3
8	27.1	7.5
9	91.5	5.0
10	113.0	29.6
11	45.9	11.9
12	61.7	11.2

An axial slice of a representative dose plan can be viewed in figure 30, and their dose-volume histogram clouds (DVHs) can be viewed in figure 31. We performed the DVH cloud analysis with

shifted geometries on classically optimized plans to maintain comparability, even though target expansion by margins was used in optimizations, rather than scenarios of shifts.

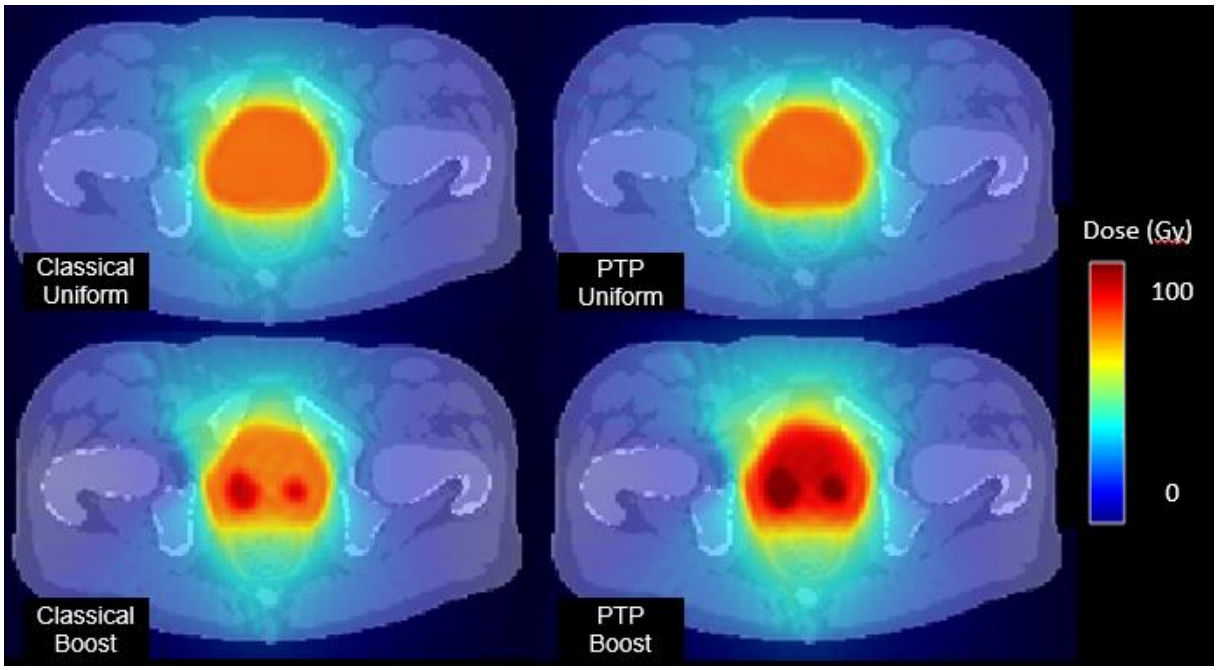


Figure 30: Axial slices of the four different dose plans for the example patient shown in figure 27. Top row shows the classical and PTP uniform plans. The two plans are comparable with very minute differences. Bottom row shows dose plans with integrated dose boosts using the classical and PTP approaches. Here, higher target doses can be seen in the PTP approach both within and outside of the GTV.

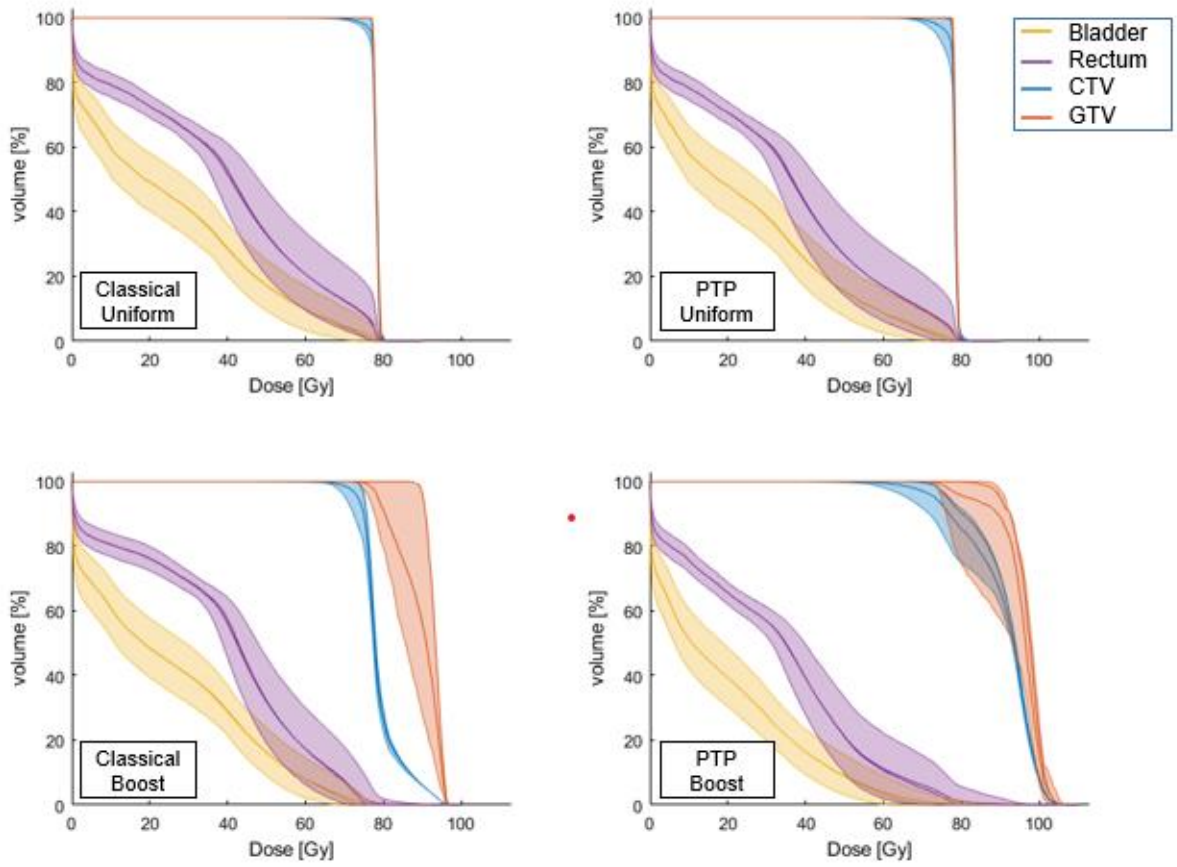


Figure 31: Dose-volume histogram clouds of the four planning approaches for the example patient. Histogram bands represent the range of dose performance among all the scenarios considered. For integrated dose boost plans, higher target mean and maximum doses can be observed, as well as lower doses to the OAR.

A quantitative overview of different planning approaches can be seen in table 10, and the p-values for statistically significant changes between the different approaches can be seen in table 11.

Table 10: Average values and their range across scenarios of different realizations of positioning uncertainties for dose benchmarks considered in the comparison between different dose planning approaches.

	Classical uniform	**PTP uniform**	**Classical boost**	**PTP boost**
CTV D_{95} (Gy)	77 [75.0-78.3]	76.3 [64.4-77.8]	80.3 [75.0-83.4]	79.8 [73.2-84.9]
CTV D_{mean} (Gy)	78.4 [76.1-81.4]	78.3 [74.2-81.0]	87 [79.5-91.9]	91.4 [86.7-94.4]
CTV D_{05} (Gy)	80 [77.4-86.0]	80.2 [79.2-84.8]	94.6 [90.9-95.8]	101 [94.0-107.0]
GTV D_{95} (Gy)	77.1 [75.2-78.4]	77.5 [71.5-80.1]	91.7 [79.2-94.8]	92.4 [87.1-102.0]
GTV D_{mean} (Gy)	78.2 [76.0-81.3]	78.5 [76.5-82.1]	93.9 [89.8-95.3]	97.3 [93.5-105.0]
GTV D_{05} (Gy)	79.4 [76.9-86.0]	79.8 [78.7-84.9]	95.7 [94.6-96.7]	103 [96.0-110.0]
Bladder D_{mean} (Gy)	19.1 [3.51-38.8]	18.4 [3.4-36.6]	18.8 [3.33-35.7]	17.3 [3.34-34.6]
Bladder V_{77Gy} (%)	0.66 [0.00-2.71]	0.60 [0.014-2.64]	0.941 [0.00-3.79]	0.278 [0.00-1.51]
Rectum D_{mean} (Gy)	31.3 [22.6-41.0]	28.9 [21.3-36.6]	31.5 [22.6-41.7]	29.8 [21.2-40.8]
Rectum V_{77Gy} (%)	1.82 [0.00-5.28]	1.26 [0.04-3.66]	2.15 [0.134-4.74]	1.77 [0.00-5.85]

Table 11: Table of p-values comparing dose metrics between different plans. P-values below the threshold of 0.05 are bolded. In uniform dose plans, lower average rectum doses are achieved. For dose boost plans, lower bladder doses and higher CTV D_{mean} and D_{05}, as well as higher GTV D_{05} are observed. P-values were corrected using the Bonferroni method and a factor of 20. Corrected values were capped at a maximum value of 1 for clearer presentation.

Metric \ Plan	**Uniform Classic v PTP**	**Boost Classic v PTP**
Rectum D_{mean}	**0.017**	0.191
Rectum V_{77Gy}	1.000	1.000
Bladder D_{mean}	0.660	**0.047**
Bladder V_{77Gy}	1.000	1.000

CTV D_{95}	1.000	1.000
CTV D_{mean}	1.000	**0.041**
CTV D_{05}	1.000	**0.004**
GTV D_{95}	1.000	1.000
GTV D_{mean}	1.000	0.097
GTV D_{05}	1.000	**0.005**

As seen in table 10, for uniform plans, dose differences between the classical and the PTP approach were typically less than 1 Gy across all metrics, with few differences observed, although Rectum D_{mean} was lower in PTP planning. For Classical Boost and PTP boost plans, the differences were greater.

PTP planning achieved lower D_{mean} to bladder by 1.5 Gy, which was a significant difference, as shown in table 11. The dose to rectum was decreased by 1.7 Gy, although that did not achieve threshold of significance. While CTV D_{95} did not change significantly, CTV D_{mean} and D_{05} were significantly increased by 4.4 and 6.4 Gy, respectively. All GTV doses were higher in PTP planning, although only GTV D_{05} achieved significance threshold, with an increase of 7.3 Gy.

A visual representation of differences in various dose metrics across patients can be seen in figure 32. Consistent increases in GTV and CTV D_{mean} and D_{05}, as well as consistent decreases in dose for both bladder and rectum can be observed.

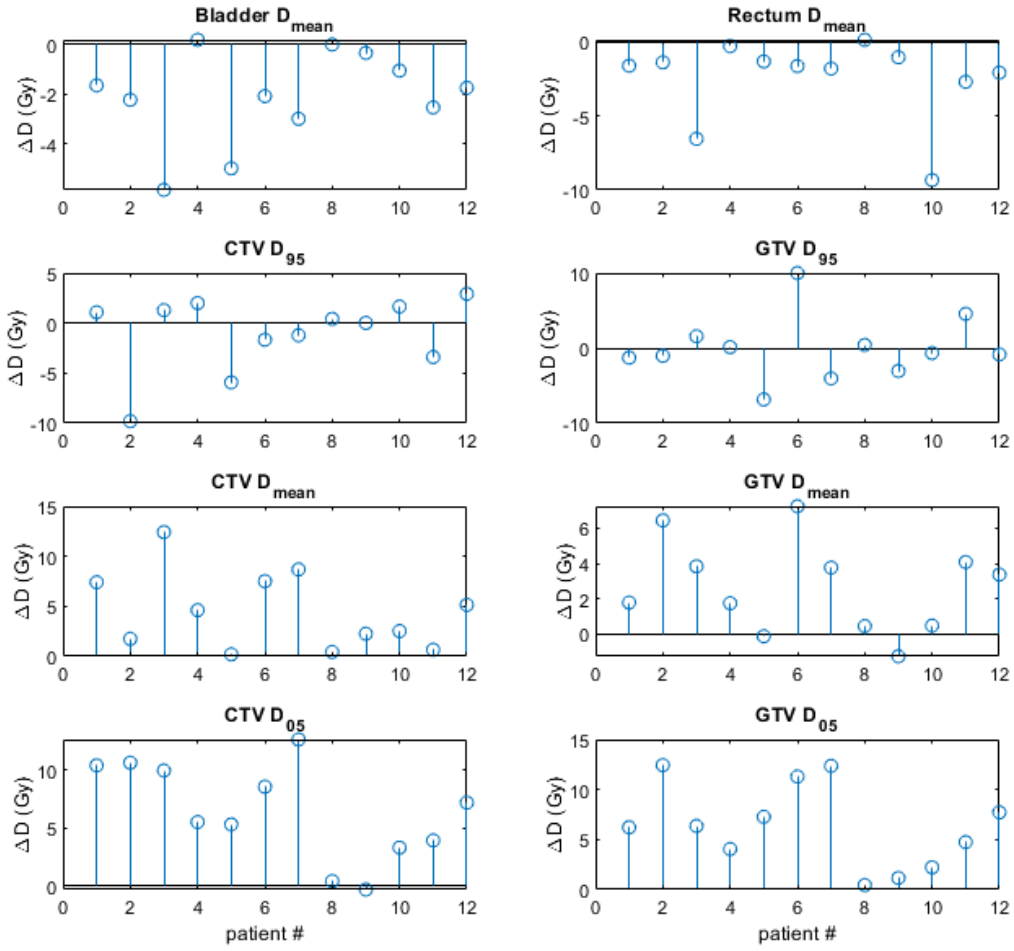


Figure 32: Differences of between PTP and classical planning for all patients and dose benchmarks in plans with integrated dose boost. Values above 0 represent increased, and values below 0 represent decreased doses in PTP, when compared to classical planning. Vertical lines to y=0 are shown to facilitate visual interpretation.

6.4 Discussion

In this chapter we presented an application of the PTP approach to a dataset of patients with prostate cancer. We showed that with the PTP approach, both uniform and non-uniform dose

distributions can be prescribed and optimized in the presence of uncertainty. For uniform dose prescriptions, both classical and PTP approaches performed similarly, with PTP achieving lower rectum doses, but otherwise performing comparably. Identical results were not expected since the two planning approaches are fundamentally different, with different objective functions and optimization algorithms.

For plans with integrated dose boosts, PTP resulted in plans with more heterogeneous target dose with higher average and maximum doses, while at the same time reducing doses to bladder and rectum. These results show that PTP can create dose plans with more favorable target/OAR ratios when compared to the classical approach.

Broadening of the target dose range is consistent with other probabilistic planning implementing shift scenarios, as target optimization goals near OARs tend to compete against OAR limits. In other words, the PTP approach intrinsically balances target dose optimization objectives for OAR objectives in the presence of uncertainties.

This trade-off can also happen within one volume in the presence of multiple competing optimization objectives, such as in the case of dose boost prescribed to the entire CTV volume and weighted using TM_{boost}. Since in the PTP approach the dose boost is prescribed to the entire prostate volume weighted with the TLM values, the optimizer will be penalized less when delivering higher doses to most of the prostate.. This allows the optimal dose to breach the non-boosted dose maximum if it permits for an overall better result, such as target coverage or OAR sparing. This approach can take advantage of previous findings that show that allowing some dose heterogeneity within the target volume can lead to sharper dose falloffs and overall better plans [142].

PTP is generalizable beyond the application shown in this work. We chose prostate as our example for this work as existing studies have shown the benefits of focal lesion ablative boost in large-scale clinical trials, but there is no aspect in this approach that prevents its generalization to other disease sites.

While the approach and main concepts of this work are generalizable, we made several assumptions that limit its scope. First, the initial gTLM maps were subject to considerable uncertainty. The microscopic infiltration algorithm was based on parameters from a single histological study and did not include information regarding anatomic barriers or extracellular matrix that might cause anisotropic cell motility. Additionally, the baseline value of 0.7 for TM within the prostate, and the threshold of 0.3 for the GTV were determined empirically, but as TM values were used as part of overall objective weighting, small variations in this baseline did not substantially affect optimization results.

While some uncertainty exists in all these input parameters, changing them would not invalidate the general approach introduced in this work. Any additional knowledge of disease presence and uncertainties can still be simply and directly integrated into the PTP approach.

6.5 Conclusions

In this chapter we presented an implementation of the probabilistic target definition and direct incorporation into treatment planning on patients with prostate cancer. We have demonstrated that this PTP approach in combination with known probabilistic optimization produces feasible clinical plans in patients with prostate cancer for both plans of uniform prescription and plans with integrated dose boosts. Due to its non-reliance on expansion by margins to account for uncertainty, PTP provides great opportunity for more complex dose planning and better use of imaging information to preferentially target subvolumes of increased disease presence.

7 PTP Applications – Complex Cases

So far, we explored probabilistic planning on patients with glioblastoma, where we started the process at image analysis and quantification of uncertainties before proceeding to RT planning and prostate, where we also examined probabilistic planning based on MP-MRI derived TLMs for integrated dose boosts. However, probabilistic targeting and planning is by no means limited to just those examples. To demonstrate PTP ability to deal with complex RT situations, we will examine a few other applications, each presenting its own challenge.

The first group of examples involve patients with cancers in head and neck. Due to the small volume and highly diverse anatomy, target volumes in head and neck tend to be strongly non-spherical, surrounded by multiple radiosensitive organs at risk, and in their very close proximity. This requires highly conformal dose planning, that requires compromises between target coverage and sparing of essential OAR. The intrinsically probabilistic nature of PTP offers an opportunity to sensibly take these trade-offs into account and achieve maximum coverage wherever possible, while simultaneously remaining conscious of sensitive normal tissues.

The second group involves patients with brain cancers that require whole-brain irradiation with hippocampus sparing. As the organ at risk (OAR) – the hippocampus – does not have well defined boundaries in all directions, a probabilistic description of the OAR can be utilized to describe with greater degree of freedom the volumes that we wish to avoid with intermediate priority. This means that the optimizer can spare the tissues where it is easier to achieve, and compromise sparing on the edges when necessary to achieve lesion coverage.

Lastly, we explored dose prescription and optimization for a patient case with a molecular imaging-based dose painting prescription. While several studies on dose painting prescriptions have already been published, many existing works rely on accurate patient positioning without

accounting for uncertainties. Due to its non-reliance on margins, the PTP approach would allow continuous dose painting prescriptions while accounting for uncertainties, potentially increasing robustness of such plans and expanding possible applications of dose painting.

To summarize, the purpose of the work we present in this chapter is to demonstrate the use of the PTP approach for different applications that represent different special challenges of treatment planning.

7.1 PTP in complex anatomies

7.1.1 Motivation

In Chapter 5 we evaluated probabilistic dose planning on patients with prostate cancer. While that allowed us to perform a sound demonstration of PTP on a realistic geometry, prostate anatomy is not as complex as some other disease sites. Conversely, radiotherapy of head and neck cancers presents one of the most complex sites in terms of geometry, organs at risk and conformal dose plans. For this reason, we expanded the PTP analysis to a set of three patients with head and neck cancers.

Multiple target volumes are often used to capture both the primary tumor and suspected nodal involvement. Sites of suspected subclinical spread are typically prescribed lower doses than the visible disease, reflecting a lower likelihood and cancer cell density .As the anatomy of the area is highly diverse, target volumes can be complexly shaped. Additionally, multiple nearby OAR such as parotid glands, spinal cord, optical nerves, etc. must be considered to ensure acceptable quality of life for the patient. These OARs are frequently located close to the target, sometimes even overlapping with their target volumes.

Due to the highly diverse anatomy, manifestations of head and neck cancers can vary considerably between different patients, based on the exact location and extension of the disease.

For this reason, we organized this section into a series of case studies where comparisons on individual level are performed, rather than as a population.

7.1.2 Materials and methods

Three patients with head and neck cancers were accepted for treatment at Centre Léon Bérard, in Lyon, France. Patients received planning CT scans as part of their standard-of-care treatment, and their CTVs and OARs were segmented by an experienced radiation oncologist.

7.1.2.1 Patient 1

The first patient was a 56-year-old male presenting with odynophagia and hemoptysis. Nasoendoscopic examination revealed multiple left and right neck nodes with a nasopharyngeal mass infiltrating the posterior pharyngeal wall and the top of the left tonsil fossa. A biopsy sample was taken and revealed undifferentiated carcinoma. MR imaging using volumetric interpolated breath-hold examination that achieved fat suppression via chemically selective fat saturation (FS-VIBE) and a gadolinium contrast agent were used to confirm a large nasopharyngeal tumor infiltrating the L&R tonsil fossae, and a bony infiltration in the clivus and petrous bone. Multiple bilateral lymph node involvement was also observed. PET/CT imaging did not reveal any distant disease. A coronal and a sagittal slice of the patient's MRI can be seen in figure 33.

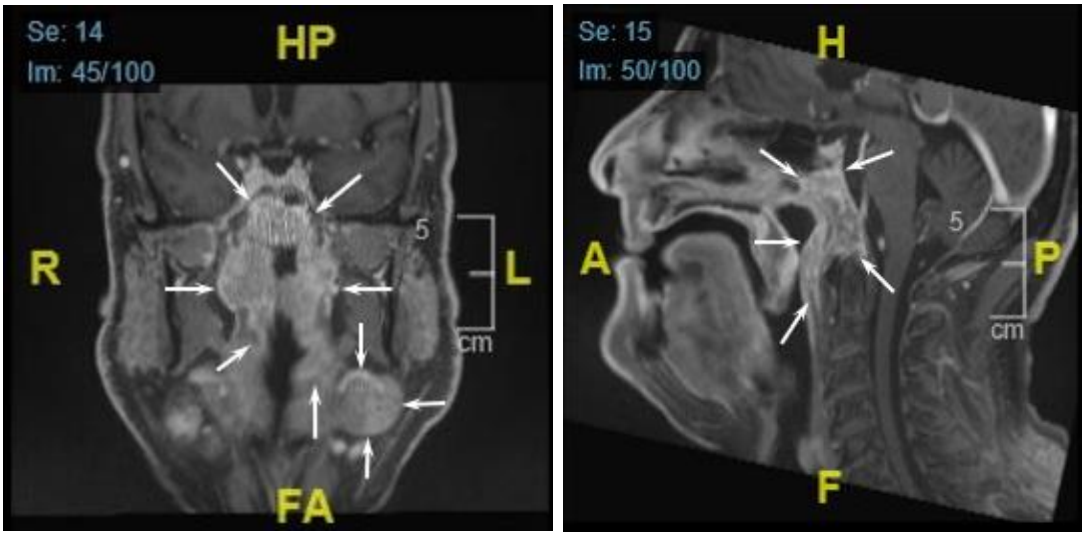


Figure 33: Coronal and sagittal slices of the patient's MRI. Arrows indicate location of disease.

We identified four target volumes. CTV-P1 included the target in the nasopharynx, clivus, apex of the right and left petrous bone and tonsil fossae. CTV-P2 encompassed sphenoid sinus, posterior 1 cm of the nasal fossae, right pterygoid muscle and left parapharyngeal space. We also identified four additional nodal CTVn in the left and right retropharyngeal and retro styloid lymph nodes, for a total of six target volumes.

We prescribed a dose of 70 Gy to the CTV-P1 and therapeutic CTVn, and a dose of 54.25 Gy to CTV-P2 and to remaining two CTVn prophylactically. We set target dose benchmarks as $D_{99\%}$ = 90%, $D_{95\%}$ = 95%, and $D_{05\%}$ = 107%. In addition to that, we prescribed doses for the nearby OARs: spinal cord, brain stem, parotid glands, oral cavity, eyes, optic nerves, larynx, trachea, esophagus, and mandible. These target and OAR volumes can be seen in figure 34.

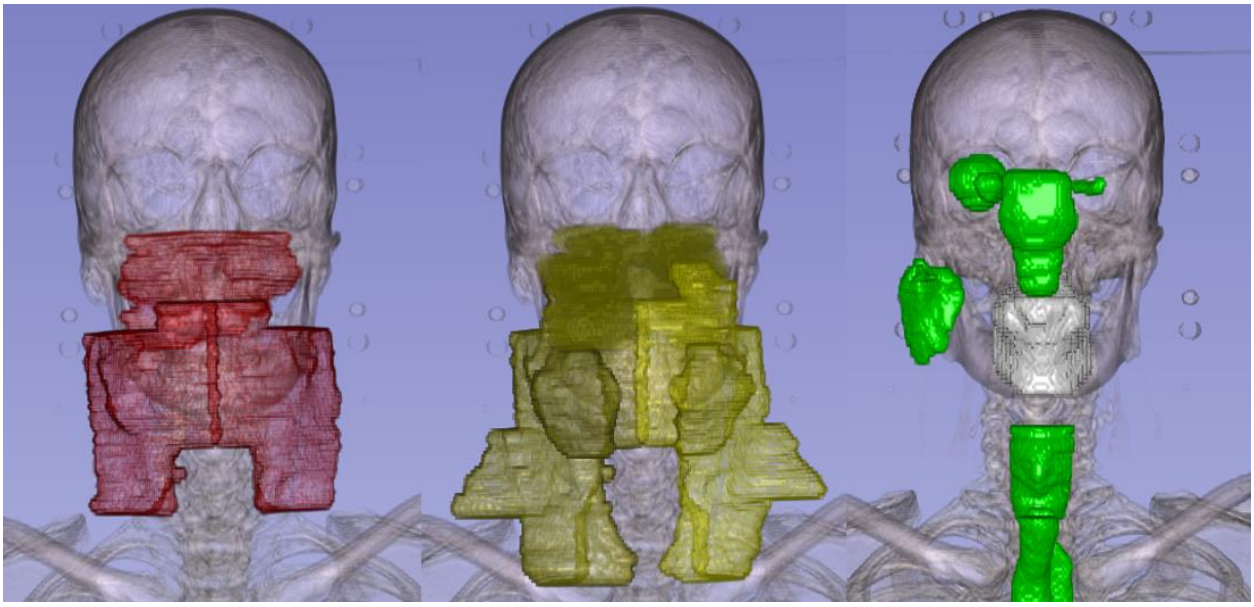


Figure 34: Different VOI for patient 1, split by category. Target volumes with 70 Gy prescription (red), target volumes with a 56.25 Gy prescription (yellow) and OAR (green).

We created dose plans using both classical and PTP approach. A CTV-PTV margin of 4 mm was used to account for positioning uncertainties in the classical approach. For the PTP approach, positioning uncertainties were accounted for using scenarios of 4 mm shifts in the six cardinal directions. We qualitatively compared dose plans using dose-volume histograms, and quantitatively using dose benchmarks and their ranges. We evaluated all target metrics for their respective CTVs, for both planning approaches and under the 4 mm shift scenarios, as used for the PTP planning.

7.1.2.2 Patient 2

The second patient was a 57-year-old male presenting with mild pain and ulcer on the left lateral border of mobile tongue. Examination confirmed a 1 cm ulceration of the posterior left lateral border of the mobile tongue with no palpable node. A biopsy revealed a well differentiated

squamous cell carcinoma, and MRI and CT images revealed no visible lymph node or distant metastasis. A coronal slice of the patient's MRI can be seen in figure 35.

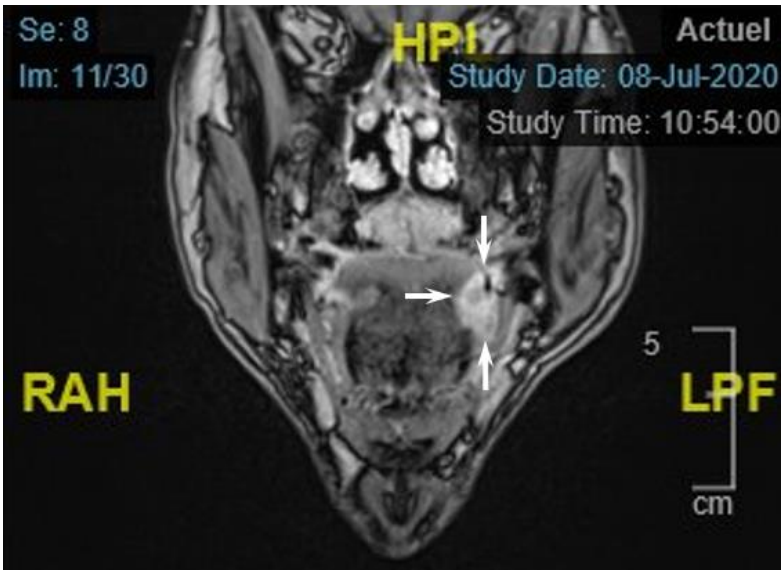


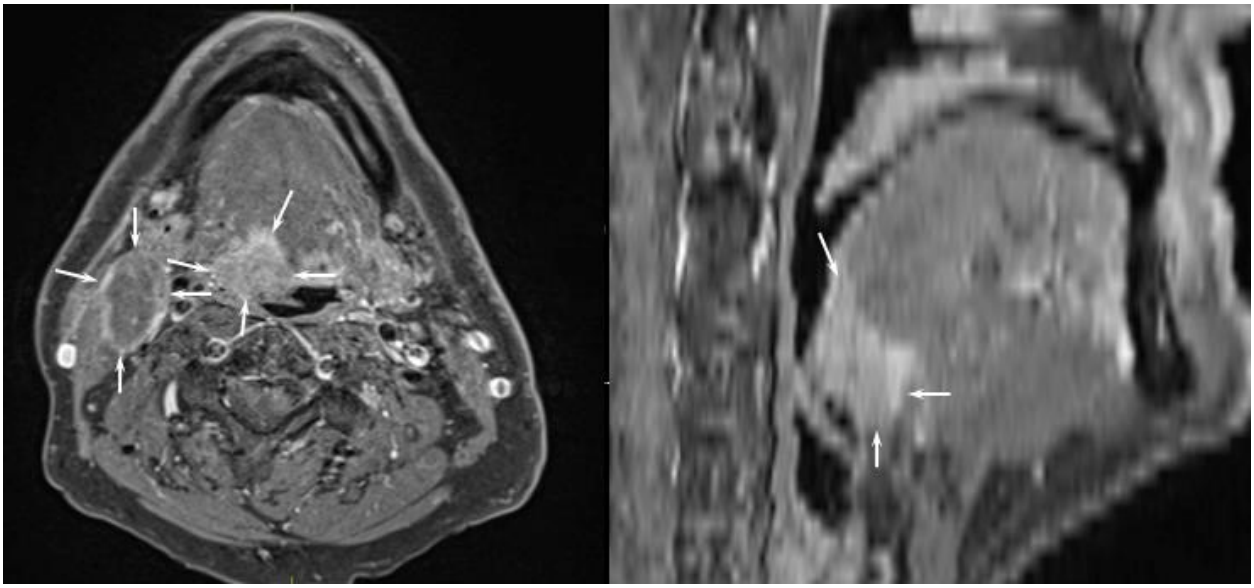
Figure 35: Coronal slice of the patient's MRI image, with the white arrows indicating the location of the disease.

For this patient, we identified one target volume and prescribed a dose of 64 Gy. We set target dose benchmarks as $D_{99\%}$ = 90%, $D_{95\%}$ = 95%, and $D_{05\%}$ = 107%. In addition to that, we prescribed doses for the nearby OARs: spinal cord, parotid glands, right sub-mandibular gland, oral cavity, pharyngeal constrictor muscle, larynx, trachea and mandible.

We calculated dose plans using both classical and PTP approaches, using a 4 mm margin in the classical planning and scenarios of 4 mm shifts for the PTP planning approach. We performed dose plan comparisons qualitatively using dose-volume histograms, and quantitatively using dose benchmarks and their ranges. We evaluated all target metrics for their respective CTVs, for both planning approaches and under the 4 mm shift scenarios, as used for the PTP planning.

7.1.2.3 Patient 3

The third patient was a 58-year-old male presenting with a lump in his right neck and soreness swallowing. An examination revealed a 3 cm node in his right side of the neck, arising from the base of the tongue, and with no involvement of lingual surface of epiglottis. Biopsy revealed moderately differentiated squamous cell carcinoma, and MRI imaging confirmed the 3 cm mass at the right base of the tongue, and necrotic right lymph nodes. PET/CT imaging revealed no distant disease presence.



We defined six target volumes for this patient. We prescribed a dose of 70 Gy to the primary target volume CTV-P1 and nodal volume CTVn_70 Gy, and prescribed the remaining four target volumes a dose of 54.25 Gy, with the same dose benchmarks as in patient 1. In addition to that, we prescribed doses for the nearby OARs: spinal cord, brainstem, parotid glands, oral cavity, larynx, trachea, esophagus and mandible. After planning, we compared dose plans qualitatively using dose-volume histograms, and quantitatively using dose benchmarks and their ranges. We evaluated all target metrics for their respective CTVs, for both planning approaches and under the 4 mm shift scenarios, as used for the PTP planning.

7.1.3 Results

7.1.3.1 Patient 1

For both classical planning and PTP, we obtained generally feasible dose plans that met most optimization objectives. We achieved all optimization objectives for target volumes, in classical planning for the nominal scenario evaluation, and failed to achieve one objective (left nodal CTV) when evaluating all scenarios. PTP target coverage also achieved all dose objectives for nominal scenarios, and several objectives were not met for all scenarios. A summary of these results can be seen in table 12.

*Table 12: Dose benchmarks for Classically and PTP optimized plan of head and neck patient 1. The smaller column in each section represents the objective that the dose plan should ideally achieve. Metrics are reported as "mean [min-max]" as evaluated over all scenarios. Orange text means the goal is not achieved. Dose requirements are marked with * and colored red if they overlap with a higher dose prescription, which overrides the prescription.*

	Volume	**$D_{99\%}$**		**$D_{95\%}$**		**$D_{05\%}$**	
Classical	CTV_T70	>63.0	67.4 [63.6-68.5]	>66.5	69.5 [67.5-69.5]	<74.9	74.6 [74.5-74.6]
	$CTV_{NR}70$	>63.0	69.3 [63.7-69.3]	>66.5	69.7 [69.4-69.7]	<74.9	73.8 [73.6-73.8]
	$CTV_{NL}70$	>63.0	69.2 [62.2-69.2]	>66.5	69.7 [69.3-69.7]	<74.9	73.8 [73.5-73.8]
	$CTV_T54.25$	>48.82	60.5 [56.1-61.8]	>51.5	65.3 [61.6-65.9]	<58.0*	74.3* [74.2-74.3]
	$CTV_{NR}54.25$	>48.82	55.7 [52.2-55.8]	>51.5	56.8 [56.3-57.0]	<58.0*	73.2* [73.0-73.2]
	$CTV_{NL}54.25$	>48.82	55.2 [51.6-55.2]	>51.5	56.5 [55.9-56.5]	<58.0*	73.2* [73.0-73.3]
PTP	CTV_T70	>63.0	64.5 [58.3-64.9]	>66.5	68.7 [65.0-68.8]	<74.9	74.9 [74.7-75.3]

	$CTV_{NR}70$	>63.0	68.2 [63.5-68.2]	>66.5	69.7 [68.8-69.7]	<74.9	74.3 [74.3-74.6]
	$CTV_{NL}70$	>63.0	66.8 [57.8-68.1]	>66.5	69.5 [68.7-69.6]	<74.9	74.9 [74.3-75.3]
	$CTV_{T}54.25$	>48.82	58.6 [49.1-58.6]	>51.5	64.5 [60.3-64.5]	<58.0*	74.9* [74.7-75.1]
	$CTV_{NR}54.25$	>48.82	52.5 [45.4-52.8]	>51.5	55.7 [54.2-55.7]	<58.0*	74.1* [74.1-74.3]
	$CTV_{NL}54.25$	>48.82	49.3 [40.3-51.4]	>51.5	54.9 [53.4-55.6]	<58.0*	74.5* [74.4-74.7]

In OAR volumes, classical planning failed to meet baseline requirements in the nominal scenario for two of the OARs: left parotid and oral cavity. In contrast, PTP achieved all dose requirements for OARs in the nominal case. Both planning approaches failed to achieve OAR dose benchmarks for all scenario evaluations. Detailed results for OAR dose benchmarks can be seen in table 13.

Table 13: OAR dose benchmarks for classical and PTP plans. Metrics are reported as "mean [min-max]" as evaluated over all scenarios. Orange text means the goal is not achieved.

Volume (#)	**Prescription**	**Classical result**	**PTP result**
L parotid (10)	D_{mean}<30 Gy	D_{mean} = 31.6 [27.6-37.6] Gy	D_{mean} = 27.0 [24.3-32.0] Gy
R parotid (11)	D_{mean}<35 Gy	D_{mean} = 34.6 [30.5-41.1] Gy	D_{mean} = 31.2 [27.6-37.8] Gy
Oral Cavity (12)	D_{mean}<35 Gy	D_{mean} = 36.3 [35.8-39.6] Gy	D_{mean} = 33.7 [32.8-35.6] Gy
Larynx (17)	D_5<55 Gy, D_{mean}<35 Gy	D_5 = 45.8 [40.8-55.8] Gy D_{mean} = 31.8 [30.8-34.1] Gy	D_5 = 51.5 [44.8-59.6] Gy D_{mean} = 33.9 [32.3-36.3] Gy
Brainstem (11)	D_2<60 Gy	D_2 = 59.4 [58.9-61.6] Gy	D_2 = 51.5 [46.2-57.4] Gy
PRV Brainstem	D_2<62 Gy	D_2 = 61.5 [59.8-65.4] Gy	D_2 = 54.1 [49.0-61.1] Gy
PRV Spinal Cord	D_2<40 Gy	D_2 = 40.0 [39.5-50.2] Gy	D_2 = 30.8 [27.8-41.1] Gy
Eye L (15)	D_5<40 Gy	D_5 = 22.9 [20.6-26.6] Gy	D_5 = 24.4 [22.4-26.5] Gy

Eye R (16)	D_5<40 Gy	D_5 = 22.3 [17.0-28.7] Gy	D_5 = 23.3 [17.8-28.3] Gy
PRV Occ Nrv R (17)	D_2<50 Gy	D_5 = 29.6 [22.2-39.1] Gy	D_5 = 32.3 [24.2-43.1] Gy
PRV Occ Nrv L (17)	D_2<50 Gy	D_5 = 32.0 [25.4-39.1] Gy	D_5 = 31.3 [25.5-38.0] Gy
Trachea	D_5<55 Gy, D_{mean}<35 Gy	D_5 = 30.8 [29.7-33.2] Gy D_{mean} = 13.04 [12.1-14.0] Gy	D_5 = 29.5 [28.3-32.5] Gy D_{mean} = 13.0 [11.9-14.0] Gy
Esophagus	D_5<45 Gy, D_{mean}<30 Gy	D_5 = 26.5 [26.4-29.0] Gy D_{mean} = 6.3 [5.5-7.4] Gy	D_5 = 27.0 [26.0-28.3] Gy D_{mean} = 6.7 [5.9-7.8] Gy

For a qualitative visual comparison, we examined dose-volume histogram clouds for classical and for PTP. Results can be seen in Table 12 for target volumes and Table 13 for OAR.

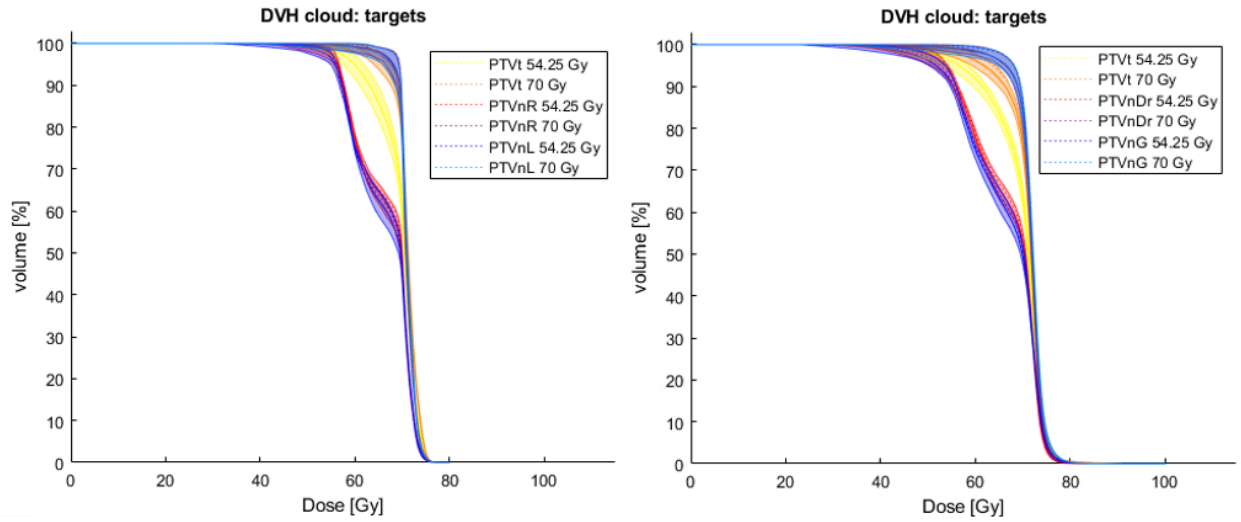


Figure 36: DVH cloud plots for all six target volumes for classical planning (left) and PTP (right) for patient 1.

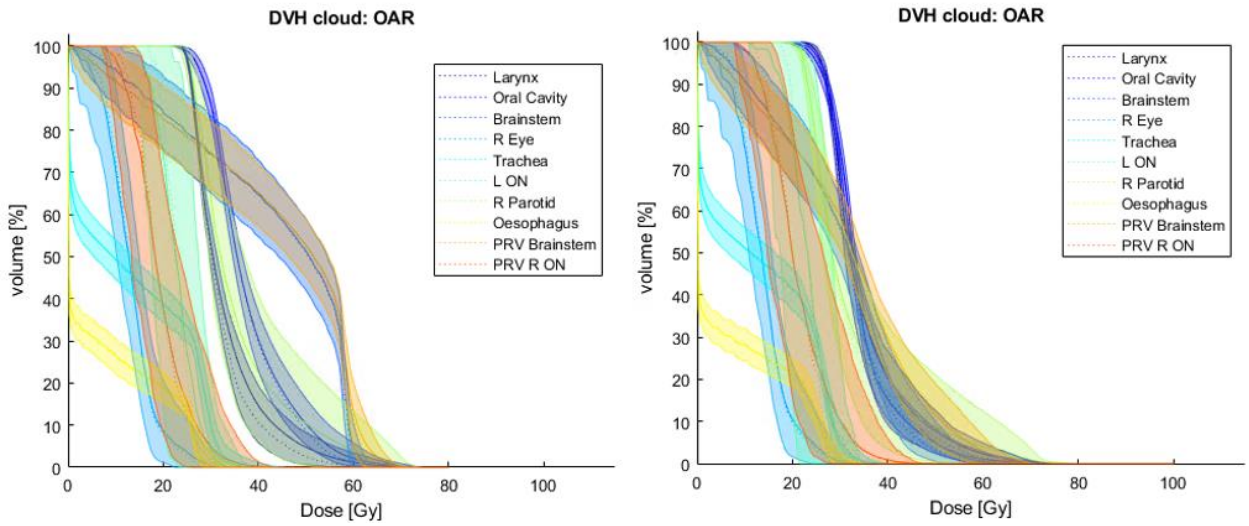


Figure 37: DVH cloud plots for all OAR volumes for classical planning (left) and PTP (right) for patient 1.

We observed some differences in OAR, where PTP generated somewhat reduced doses, especially in PTV Brainstem and PRV PRV.

7.1.3.2 Patient 2

We observed some differences in achieving dose requirements for this patient between the classical and PTP approaches. In classically optimized dose plan, the target failed to achieve all the target dose constraints, while also failing to meet the OAR constraints in the worst-performing scenario considered. PTP failed to achieve the minimum dose requirements for the target, however it achieved higher doses than when compared to classical planning. In addition, PTP failed the dose requirements only for the larynx in the worst performing scenario considered. A summary of these results can be seen in table 14.

Table 14: Target and OAR dose benchmarks for classical and PTP plans. Metrics are reported as "mean [min-max]" as evaluated over all scenarios. Text color is adjusted to orange if the goal is not achieved.

Volume (#)	Prescription	Classical result	PTP result
CTVnG_64	$D_{99\%}$>60.8 Gy $D_{95\%}$> 62.7 Gy $D_{05\%}$< 68.5 Gy	47.4 [31.8-49.3] 57.5 [45.5-57.5] 71.2 [71.2-71.2]	58.6 [46.1-58.6] 62.1 [53.0-62.1] 68.3 [68.2-68.3]
L parotid	D_{mean}<25 Gy	23.1 [18.9-29.9]	20.5 [17.3-24.9]
R parotid	D_{mean}<5 Gy	3.99 [3.53-4.58]	3.29 [2.96-3.58]
Oral Cavity	D_{mean}<30 Gy	20.1 [17.5-22.8]	17.5 [15.3-19.9]
Sub-mandibular Gl.	D_{mean}<7 Gy	6.81 [6.20-7.91]	4.99 [4.63-5.52]
Larynx	D_5<50 Gy, D_{mean}<35 Gy	39.4 [31.2-57.3] 25.9 [22.9-30.9]	40.6 [32.2-55.6] 24.6 [21.4-29.3]
PRV SC	D_2<30 Gy	25.7 [23.9-30.3]	24.8 [23.2-28.2]
Trachea	D_5<55 Gy, D_{mean}<35 Gy	19.7 [16.3-22.7] 2.92 [2.13-3.8]	19.5 [16.2-22.6] 2.9 [2.13-3.76]
Mandible	D_5<64 Gy	59.4 [53.2-64.4]	51.5 [47-55.9]
M Const. pharynx	D_{mean}<30 Gy	23 [19.9-28.8]	22.4 [19.5-26.7]

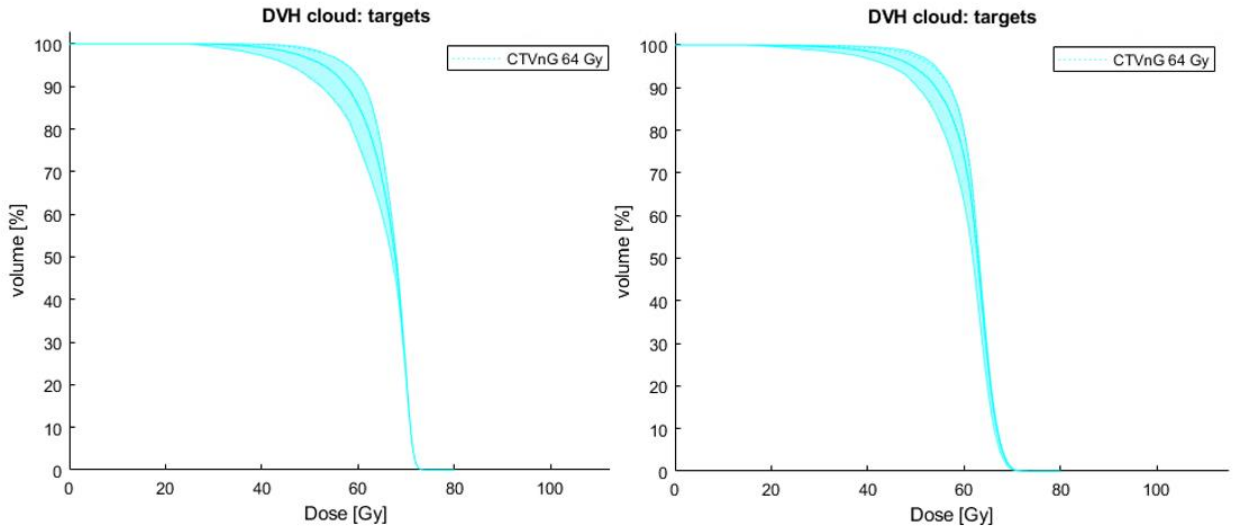


Figure 38: DVH cloud plots for all six target volumes for classical planning (left) and PTP (right) for patient 2.

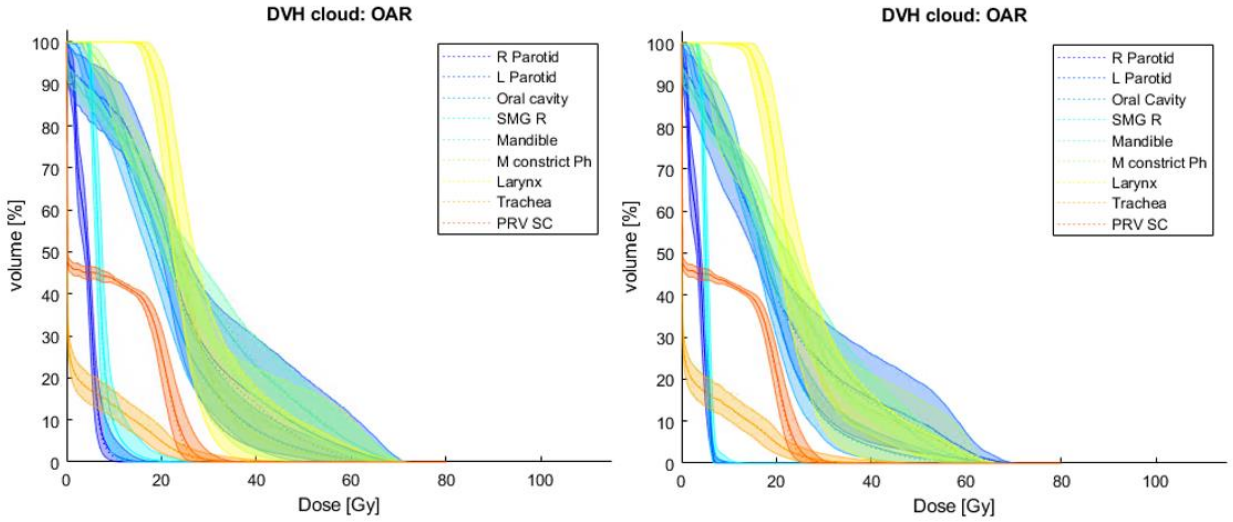


Figure 39: DVH cloud plots for all OAR volumes for classical planning (left) and PTP (right), for patient 2.

7.1.3.3 Patient 3

In the third patient, PTP target dose distributions were wider, and less consistent in achieving the minimum target doses when compared to classical planning, as seen in table 15.

*Table 15: Dose benchmarks for Classically and PTP optimized plan of head and neck patient 3. The smaller column in each section represents the objective that the dose plan should ideally achieve. Metrics are reported as "mean [min-max]" as evaluated over all scenarios. Orange text means the goal is not achieved. Dose requirements that overlap with a higher dose prescription are marked with * and colored red.*

	Volume	**$D_{99\%}$**		**$D_{95\%}$**		**$D_{05\%}$**	
Classical	CTV-P1	>63.0	69.4 [69.2-69.5]	>66.5	69.8 [69.7-69.8]	<74.9	73.1 [72.6-73.6]
	CTVn_70	>63.0	69.5 [68.7-69.6]	>66.5	69.8 [69.6-69.8]	<74.9	72.6 [72.2-73.2]
	CTV-P2	>48.8	69.0 [62.3-69.0]	>51.5	69.7 [66.9-69.7]	<58.0*	72.9 [72.5-73.3]
	Level II-IVa L	>48.8	52.9 [52.7-53.0]	>51.5	53.4 [53.3-53.4]	<58.0	55.5 [55.5-55.7]
	Level Ib/II/RPh/R	>48.8	56.0 [55.1-56.5]	>51.5	59.6 [59.0-60.0]	<58.0*	72.4 [72.2-72.7]
	Level III,IVa,V R	>48.8	52.5 [48.3-52.5]	>51.5	53.2 [52.5-53.2]	<58.0	57.5 [56.2-63.5]
PTP	CTV-P1	>63.0	62.7 [57.1-64.6]	>66.5	66.5 [62.0-66.6]	<74.9	72.0 [71.7-72.2]
	CTVn_70	>63.0	65.2 [60.2-65.2]	>66.5	66.8 [64.9-66.8]	<74.9	71.1 [71.0-71.2]
	CTV-P2	>48.8	52.0 [40.0-56.6]	>51.5	58.3 [48.4-60.6]	<58.0*	71.6 [71.2-71.8]
	Level II-IVa L	>48.8	45.4 [37.2-49.3]	>51.5	51.7 [48.2-52.2]	<58.0	58.4 [58.3-58.5]
	Level Ib/II/RPh/R	>48.8	47.7 [43.8-49.7]	>51.5	52.3 [51.0-52.6]	<58.0*	70.6 [70.5-70.6]

	Level III,IVa,V R	>48.8	50.6 [45.2-51.2]	>51.5	52.3 [49.6-52.4]	<58.0	58.3 [57.3-59.6]

PTP target doses were typically below the desired value for D_{99} metrics, and also did not deliver the desired dose in all scenarios for target D_{95}. On the other hand, PTP planning was much more successful in adhering to the prescribed limitations, as seen in table 16.

Table 16: OAR dose benchmarks for classical and PTP plans. Metrics are reported as "mean [min-max]" as evaluated over all scenarios. Text color is adjusted to orange if the goal is not achieved.

Volume (#)	Prescription	Classical result	PTP result
PRV Spinal cord	D_{02}<35 Gy	33.6 [33.5-34.4]	27.7 [27.6-30.2]
Brain stem	D_{02}<30 Gy	19.7 [18.1-22.2]	16.6 [15.0-18.0]
L parotid	D_{mean} <20 Gy	20.5 [17.0-25.3]	14.2 [12.7-17.9]
R parotid	D_{mean} <25 Gy	28.5 [23.8-33.3]	23.4 [18.8-29.5]
Oral Cavity	D_{mean} <30 Gy	39.0 [35.5-42.9]	29.4 [26.6-33.1]
Larynx	D_5<55 Gy D_{mean} <35 Gy	59.1 [49.3-66.2] 33.2 [30.9-36.0]	52.8 [39.2-63.1] 27.7 [25.8-30.7]
Trachea	D_5<55 Gy D_{mean} <35 Gy	35.8 [34.2-39.4] 15.5 [14.7-16.5]	35.6 [34.6-41.6] 15.5 [14.8-16.7]
Esophagus	D_5<50 Gy D_{mean} <30 Gy	28.6 [28.2-32.1] 12.0 [11.4-13.0]	32.4 [31.2-35.4] 12.5 [11.6-13.5]
Mandible	D_5<70 Gy D_{20}<50 Gy	59.8 [54.7-65.0] 49.4 [45.7-52.4]	55.3 [53.3-58.4] 48.4 [44.7-51.3]

Most notably, in classical planning the nominal dose to oral cavity is over the desired mean by 9 Gy, whereas in PTP all the nominal OAR dose objectives are successfully achieved.

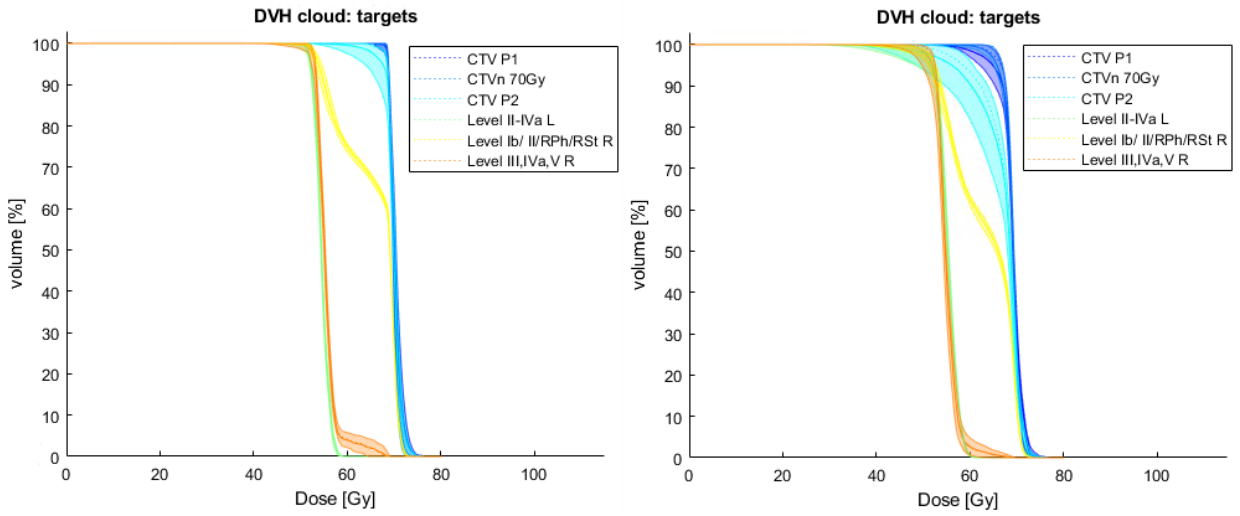


Figure 40: DVH cloud plots for all six target volumes for classical planning (left) and PTP (right) for patient 2.

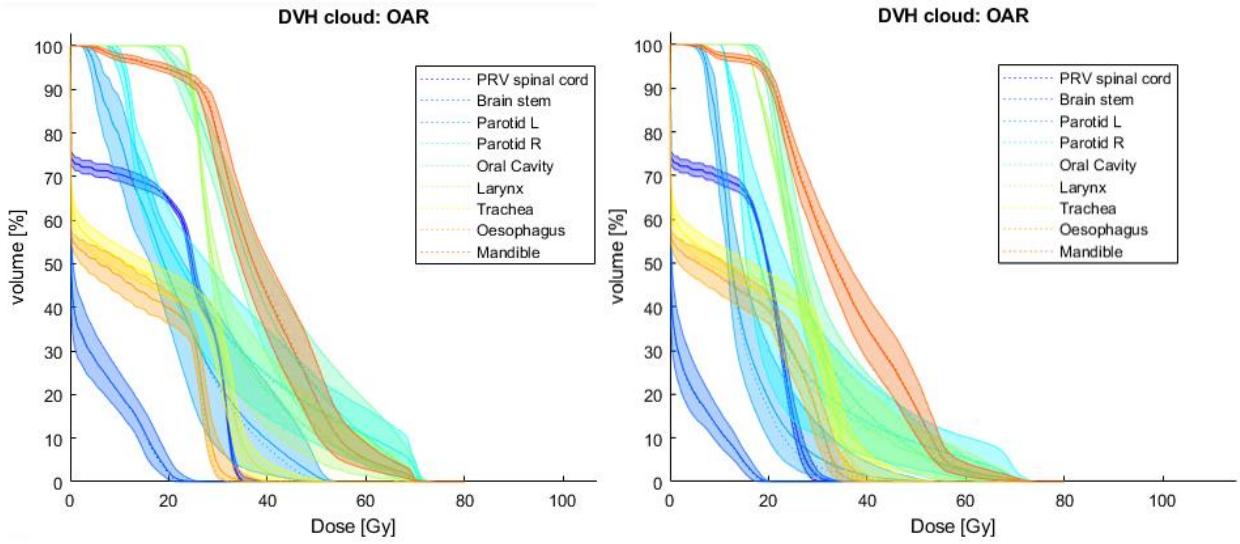


Figure 41: DVH cloud plots for all OAR volumes for classical planning (left) and PTP (right), for patient 3.

7.1.4 Discussion

In this section we compared the dose plans created with classical and PTP methods in three patients with H&N cancers. Due to the complex geometry and conflicting optimization goals, it was not possible to achieve all optimization objectives for all cases, but both planning approaches were still able to create generally feasible dose plans. We elected not to group the patients for comparison between planning approaches due to their small number and greatly different disease presentations, but we instead evaluated them as separate case studies.

In patient 1, in the classically optimized plan utilizing PTVs, we observed good target coverage for almost all target volumes and scenarios of uncertainty realizations. The plan failed to achieve desired dose coverage for the left nodal target volume under some scenarios, due to the proximity to the left parotid gland. Classically optimized plans also failed to achieve dose requirements in two of the OARs: the left parotid and the oral cavity. Both volumes were in close proximity to target volumes and were therefore directly opposed with competing dose objectives.

In PTP optimized dose plans, all target and all OAR requirements were met for the nominal scenario. While PTP results did meet all target dose requirements for the nominal scenario, target dose coverage was less consistent in PTP than in classical, with several objectives breaching their desired prescriptions in at least some scenarios. However, we observed consistently lower OAR doses in PTP plans when compared to classical, such as 10 Gy lower dose in the spinal cord.

In patient 2, the classically optimized plan struggled to achieve sufficient dose coverage in the target, presumably due to its elongated shape and proximity to several OAR that limited dose deposition. While PTP planning also failed to meet target D_{95} and D_{50} values, it achieved higher doses to both metrics, as well as meeting all other dose constraints other than worst scenario D_5 in larynx – where it still performed better than classically optimized plans.

In patient 3, the PTP dose plan resulted in somewhat lower D_{99} metrics in comparison to classical planning, with most values being slightly below the desired thresholds. However, in classical planning, the dose to oral cavity was a full 9 Gy higher than the prescription, where PTP planning achieved all OAR limitations.

While the small number of patients and different anatomies prevented any statistical comparisons as populations, these cases still provided valuable insight on the comparative performance of PTP planning and demonstrated the usability of PTP in some of the most complex geometries available. Our results also demonstrated that the PTP approach balanced target and OAR objectives where required, resulting in dose plans with good adherence to desired optimization objectives.

7.2 PTP for probabilistic organ sparing

7.2.1 Motivation

We have shown the usefulness of the PTP approach for cases with probabilistic targets. However, the PTP approach can also be used for probabilistic normal tissue definitions. One such example is whole-brain irradiation with hippocampal sparing and dose escalation on metastases. Whole-brain irradiation therapy is the standard treatment modality for patients with multiple brain metastases, as it was shown to significantly improve tumor control and reduce the rate of neurological death compared to radiosurgery alone [143]. However, whole-brain irradiation is associated with long-term side effects, with multiple patients reporting significantly poorer quality of life than patients receiving local therapy. The main side effects are neurocognitive deficiencies, in particular worsening of memory abilities. The hippocampus is the human brain structure that is most critically involved in episodic memory processing, and damage to the hippocampus is associated with memory dysfunction. Using hippocampus sparing whole-brain irradiation approaches has been shown to preserve memory and quality of life, compared to standard whole-brain irradiation.

While hippocampi can be well identified on anatomical imaging, the exact extension of the nerves involved in hippocampi function is uncertain. As the PTP formalism in theory allows probabilistic handling of not just targets, but also OAR, the goal of this section is to evaluate whether probabilistic handling of OAR is feasible using a probabilistic description of hippocampus sparing.

7.2.2 Materials and methods

We performed this proof-of-concept analysis on a dataset of two patients with brain cancer intended for hippocampus sparing whole brain irradiation. The patients were imaged with a Brilliance Big Bore CT scanner at the Universitaetsklinikum Tuebingen, Germany, in 2017 using a O-MAR reconstruction algorithm. An experience physician segmented the images.We obtained

volumes of whole brain, left and right hippocampi, and expanded hippocampus volumes for both sides to be used in classical planning, eyes and any existing cancer nodes. We additionally calculated normal tissue likelihood maps (NTLM) for hippocampi for the PTP approach. Example patient images can be seen in figure 42.

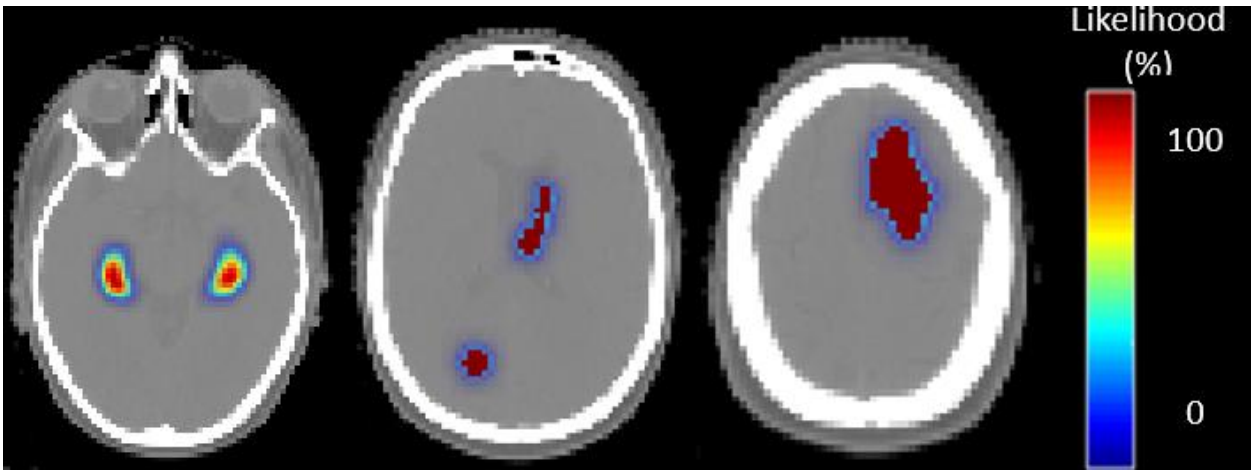


Figure 42: Axial CT slice with hippocampi NTLM (left), and two axial slices for TLM for brain metastases (middle, right). The hippocampi were prescribed a lower, and the metastases a higher dose compared to the rest of the brain volume.

For radiation therapy planning, we resampled CT volumes and all corresponding segmentations to a 2 mm cubical grid. We calculated TomoTherapy beamlets for 51 angular origins per gantry rotation using a pitch of 0.86 and 64 beamlets per source origin. For PTP planning, we expanded hippocampus volume using a region growing approach and planned for using both classical and PTP approaches.

We evaluated the results using dose benchmarks in the brain, the hippocampi, and for the second patient, in metastases. While the first patient did have a metastasis labeled, it was very small and partially overlapping with the hippocampus volume, and was therefore excluded from the planning and analysis. Dose benchmarks were compared between classical and PTP approaches.

7.2.3 Results

7.2.3.1 Patient 1

We created dose plans conforming to the overall expectations using both classical and PTP approaches. The different dose benchmarks obtained with classical planning for the brain CTV and the hippocampi, can be seen in table 17 and table 18. In all the tables in this section, entries are written by first stating the mean scenario dose and then the min and max doses in the brackets.

Table 17: Classically optimized plan dose benchmarks for patient 1.

	$D_{95\%}$	$D_{50\%}$	$D_{05\%}$
Brain CTV	9.34 [9.34-9.34]	30.1 [30.0-30.1]	32.4 [32.4-32.5]
Hippocampus L	4.03 [4.03-4.04]	7.16 [7.13-7.22]	18.4 [14.5-22.8]
Hippocampus R	6.27 [6.27-6.27]	7.72 [7.71-7.73]	10.4 [9.79-11.1]

Similarly, the results of PTP planning for the same patient can be seen in table 18. We observed generally comparable performance for both volumes, with PTP planning achieving 1 Gy lower doses in the hippocampi for most dose volumes and 9 Gy for the D_{05} of the left hippocampus. PTP also achieved higher D_{50} dose in the brain volume and 1.8 Gy lower dose in brain D_{95}.

Table 18: PTP optimized plan dose benchmarks for patient 1.

	$D_{95\%}$ (Gy)	$D_{50\%}$ (Gy)	$D_{05\%}$ (Gy)
Brain CTV	7.56 [7.55-7.57]	32.7 [32.6-32.7]	35.2 [35.1-35.2]

Hippocampus L	5.46 [5.46-5.46]	6.52 [6.5-6.61]	9.54 [8.45-10.8]
Hippocampus R	5.39 [5.39-5.39]	6.4 [6.38-6.47]	8.92 [8.14-9.62]

7.2.3.2 Patient 2

The second patient presented with three separate metastases with prescriptions of 51, 51 and 42 Gy, respectively. Dose benchmarks for the classical approach can be seen in table 19 and table 20 for the PTP approach.

Table 19: Classically optimized plan dose benchmarks for patient 2.

	$D_{95\%}$ (Gy)	**$D_{50\%}$ (Gy)**	**$D_{05\%}$ (Gy)**
Brain CTV	15.3 [15.1-15.5]	29.6 [29.6-29.6]	35.8 [35.7-35.8]
Hippocampus L	4.75 [4.75-4.75]	6.66 [6.66-6.67]	9.40 [8.88-9.92]
Hippocampus R	4.91 [4.91-4.91]	7.06 [7.05-7.09]	9.37 [8.87-10.0]
CTV 1 (51 Gy)	48.4 [46.5-49.6]	50.9 [50.7-50.9]	52.1 [52.0-52.1]
CTV 2 (51 Gy)	39.5 [36.0-43.8]	47.5 [45.8-48.6]	51.9 [51.7-52.0]
CTV 3 (42 Gy)	40.5 [40.1-40.9]	41.9 [41.9-41.9]	42.7 [42.7-42.8]

Again, PTP planning for patient 2 produced dose plans where Brain D95% was 1.7 Gy lower than in classical planning. PTP plans also resulted in considerably reduced doses to both hippocampi volumes. Comparable dose results are observed for all three brain metastases receiving the integrated dose boost.

Table 20: PTP optimized plan dose benchmarks for patient 2.

	$D_{95\%}$ (Gy)	$D_{50\%}$ (Gy)	$D_{05\%}$ (Gy)
Brain CTV	13.6 [13.4-13.9]	30.2 [30.2-30.2]	36.1 [36-36.1]
Hippocampus L	2.74 [2.74-2.74]	4.86 [4.82-4.98]	8.03 [7.65-8.71]
Hippocampus R	2.98 [2.98-2.98]	5.27 [5.24-5.34]	7.94 [7.51-8.7]
CTV 1 (51 Gy)	48.3 [46.2-49.9]	50.9 [50.8-51.1]	52.0 [51.9-52.0]
CTV 2 (51 Gy)	38.1 [34.9-42.9]	46.1 [44.0-47.7]	51.1 [50.7-51.3]
CTV 3 (42 Gy)	41.9 [41.6-42.4]	43.4 [43.3-43.4]	44.1 [44.1-44.2]

7.2.4 Discussion

In these cases of probabilistic hippocampus sparing, both treatment planning approaches achieved the desired effect of creating homogenous baseline dose with dose reduction in the hippocampi. For the second patient, the integrated dose boosts effectively targeted visible metastases. PTP dose planning slightly favored hippocampus sparing over brain dosage in both patients, as is observed by brain $D_{95\%}$, but achieved overall higher $D_{50\%}$ and $D_{05\%}$ for brain, as well as lower doses to the hippocampi. Doses to the metastases for the second patient were comparable between the two approaches.

These results show the viability of PTP use for whole brain irradiation with hippocampus sparing, where the OAR is defined in a probabilistic manner, rather than with sharply contoured boundaries. This probabilistic OAR description allows for optimization where volumes of intermediate priority are spared where possible, but not when conflicting with other, higher priority optimization objectives.

7.3 PTP for dose painting

7.3.1 Motivation

Many tumors show significant spatial variation in biologic characteristics such as proliferation, hypoxia, cell density and perfusion. As such, striving for a homogeneous dose distribution to target volume may not be the most rational way to deliver the radiation dose [144]. Therefore, the concept of dose painting has been introduced, where a higher dose is delivered to the more resistant tumor areas while reducing the dose toward the more sensitive areas. First introduced by Ling et al., it can also be viewed as reading biological, rather than physical conformity of the radiation to the tumor [145]–[147]. Dose painting could allow target dose escalation without an increase in toxicity, potentially improving treatment efficacy. Current dose painting approaches, poorly account for uncertainties in planning and delivery, often simply including a PTV volume in dose prescription [148]. This section seeks to evaluate the feasibility of PTP approach for dose painting, which could allow for a more realistic incorporation of uncertainties into treatment planning.

7.3.2 Materials and methods

For this proof-of-concept demonstration of dose painting planning, we used a single patient from the Avastin dataset. This dataset is a 10-patient phase I study designed to test the safety of administering Avastin (bevacizumab) to patients with head and neck squamous cell carcinoma. The Avastin protocol involves a rich set of imaging data (FDG, FLT, CuATSM, and DCE CT imaging) as well as target contours. For the purposes of this analysis, we used baseline FDG PET/CT to generate a variable dose prescription and to create separate target maps for PTP planning.

First, we resampled the CT and PET images to a 2 mm cubic grid. For the largest contoured target volume, we created a continuous dose prescription map using the conversion function of:

$$D_i = 50 + 15 * \sqrt{SUV_i} \tag{7.1}$$

Where D_i is the prescribed dose in voxel *i* in units of Gy, and SUV_i is the measured FDG uptake in that voxel. While some dose painting studies use linear equations to convert PET uptake to dose prescription, the square root conversion allows better targeting of multiple different tumor subvolumes [148]. To reduce the effect of image noise on optimization, we smoothed the prescription maps using a 6 mm Gaussian filter. This resulted in a prescription map that ranged from 50 to 110 Gy. In addition, we defined separate probabilistic target maps for top and bottom prescriptions. The minimum dose target prescription was defined simply as shown in equation 7.2

$$TM_{min,i} = D_i/Gy - 50 \tag{7.2}$$

Where $TM_{min,i}$ is the target map value in voxel *i*, calculated by offsetting the dose map prescription from the original baseline of 50 Gy. This meant that high-prescription volumes were preferentially weighted to receive a minimum dose, compared to the low-prescription volumes. For the maximum dose prescription, an inverse target map was defined as the difference from the maximum TM_{min} value to the $TM_{min,i}$ value in voxel *i*. This relation is also shown mathematically in equation 7.3.

$$TM_{max,i} = \max\left(TM_{min}\right) - TM_{min,i} \tag{7.3}$$

As most of the patient's heterogeneity affecting dose planning was captured in these maps, only three optimization objectives were used for this patient. These three were the minimum and maximum dose prescription, each using the dose map for dose prescription and target map for voxel-level weights and a ring structure surrounding the volume to avoid dose hot spots outside of the target. Target maps were defined in this way because subvolumes receiving the increased dose are small when compared to the larger baseline and therefore get less dose to satisfy maximum dose limitations in the larger surrounding volumes with a lower prescription. An example representation for all these volumes can be seen in figure 43.

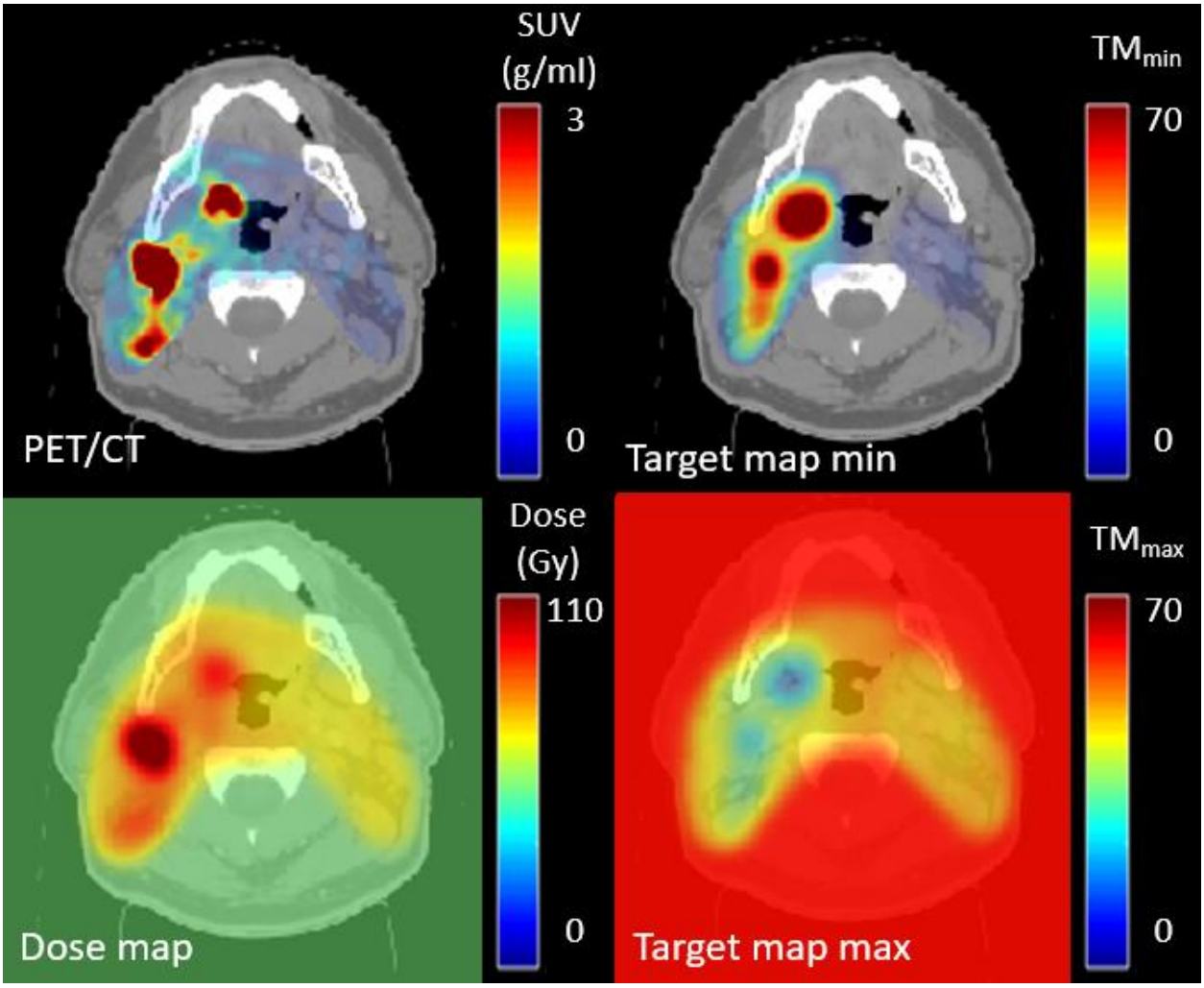


Figure 43: Example representation of input and planning volumes for dose painting. Input FDG PET/CT can be seen in top left, and its transcription into the dose map in bottom left. The voxel weights used for minimum and maximum prescriptions can be seen in the top right and bottom right, respectively.

We calculated the dose for a TomoTherapy helical delivery plan using 64 binary multileaf collimators, 51 angles for the X-ray source position per rotation, a pitch of 0.86 and a 2.5 cm jaw width. We excluded dose calculation for beamlets that did not intersect the CTV expanded by 4

mm (the classical PTV), and we omitted the weights corresponding to those beamlets in calculations to keep computational burden within machine capabilities. We used a supervoxel grouping approach to further ease the computational constraints, using 2000 supervoxels within the target volume. We performed all optimizations on a system with two 16-core Intel Xeon E5-2620 CPU (2.10 GHz) and 32 GB of RAM.

As direct optimization of dose painting prescriptions is not easily performed using the classical planning approach, we evaluated only the results of dose painting plans with PTP for this section. The classical planning approach does not lend itself to direct optimization of dose painting unless substantial modifications are performed. As these modifications are outside the scope of this work, we only evaluated results of dose painting plans created using PTP approach. We evaluated the plans using dose-volume histograms, where we additionally segmented the target volume We discretized the target volume into several subvolumes with 10 Gy increments ranging from 60-100 Gy (i.e. 60-70 Gy, 70-80 Gy, etc.). This was done to evaluate dose adherence to prescription in the different prescription ranges independently. Example discretized volumes can be seen in figure 44.

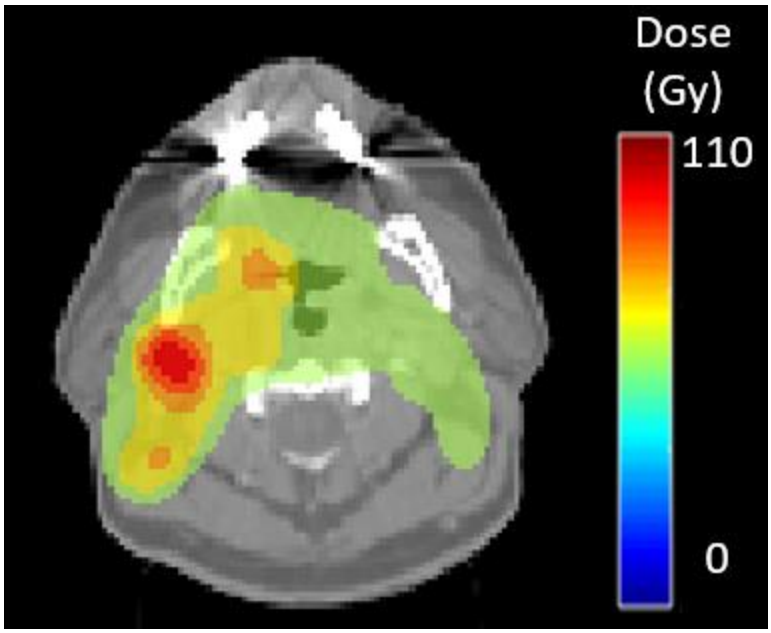


Figure 44: Example axial slice of discretized target volume

We successfully calculated the dose painting dose plan for the example patient and we demonstrated that it generated preferentially higher doses in expected areas. An example axial slice of the dose plan can be seen in figure 45.

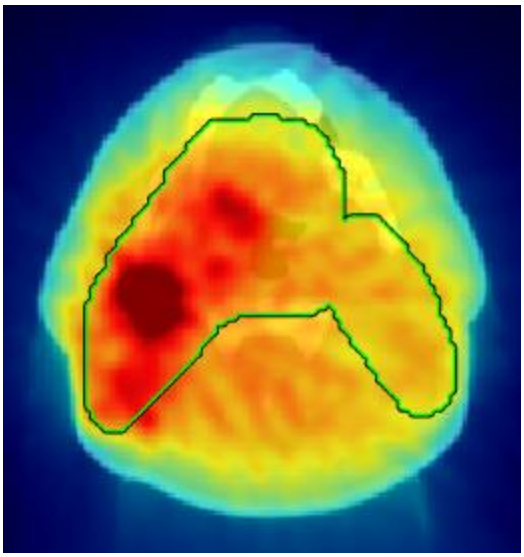


Figure 45: Axial slice of the example dose painting dose plan. The joint target volume is contoured in green. Two nodes of higher dose are visible on patient's right side (left in image). Good spatial co-occurrence with dose prescription in figure 43 can be seen.

DVH and QVH cloud plots for different target subvolumes can be seen in figure 46. Satisfactory conformity to planned doses can be observed in the QVH plot, with all subvolumes achieving doses close to prescribed, as seen by QVH values close to 1.

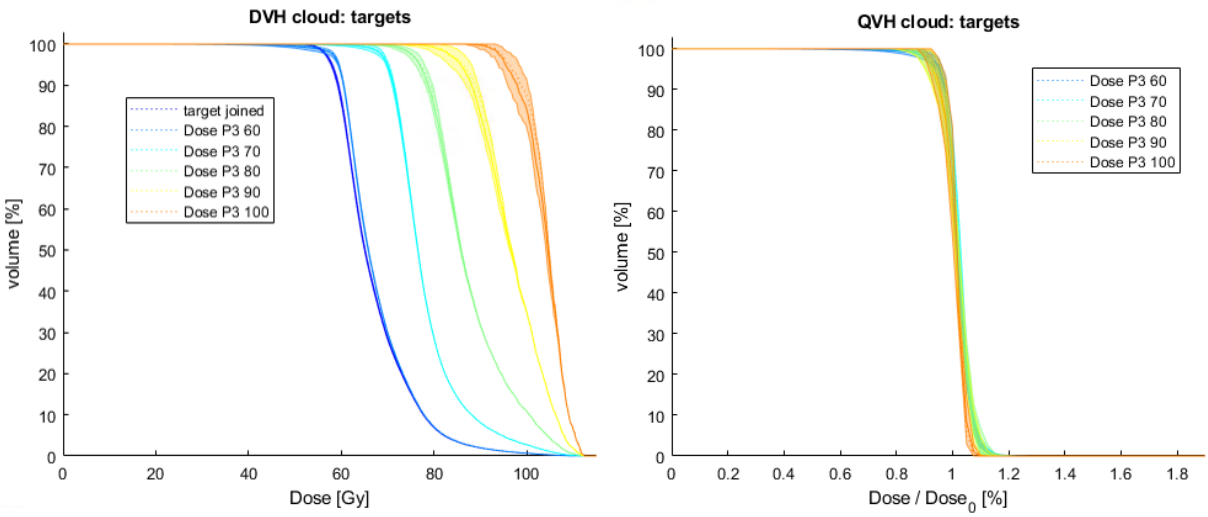


Figure 46: DVH cloud plot (left) and QVH cloud plot (right) for different target subvolumes. Dose values close to planned were observed for all subvolumes.

Examining the histogram of the delivered to normalized dose, shown in figure 47, we see that the majority of dose is close to the prescription, with 90% of the volume receiving between 95% and 106% of prescribed dose.

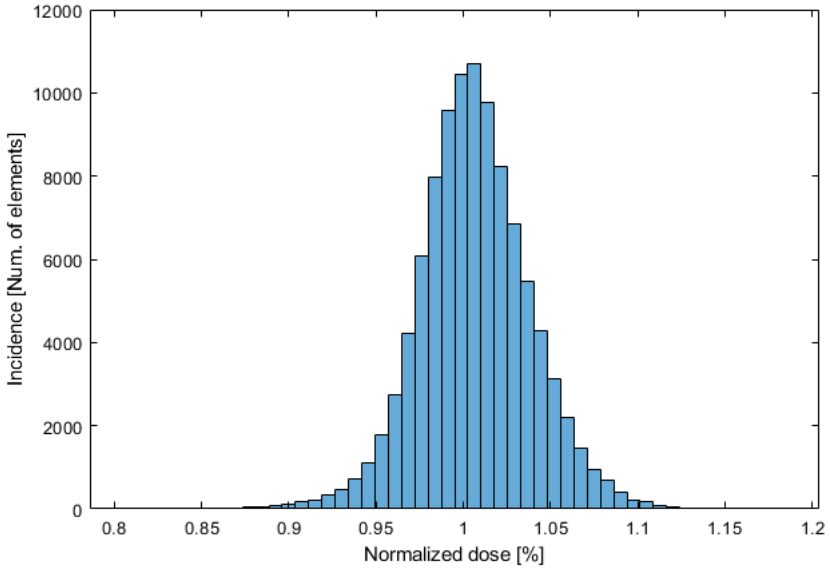


Figure 47: Histogram of planned and prescribed dose for each voxel within the joined target volume.

7.3.3 Discussion

The purpose of this sub-section was to perform a demonstration example on a patient with dose painting dose prescription. PTP approach managed to create dose plans that achieved good dose conformance (within 0.95-1.07 range) of the prescribed dose. This result was consistent between subvolumes of different dose prescription, as shown by the QVH plot. While these results do not suffice to draw any conclusion or comparison between different planning approaches, it nonetheless serves as a validation that PTP can be used to create dose painting dose plans, while still accounting for uncertainties.

7.4 Conclusions

This exploratory chapter was aimed at evaluating the opportunity provided by the PTP approach. For this goal, we looked at different examples of dose plan optimization using the PTP approach in various disease sites, each faced with a unique challenge.

In subsection 7.1, we evaluated patients with head and neck cancers, which present highly complex geometries with several targets of varying dose prescription and close proximity to multiple organs at risk. As these targets are located next to sensitive OAR, mutually competing optimization objectives are encountered, often making it impossible to achieve all specified goals simultaneously, presenting a unique challenge for any optimization approach. In our results we observed that PTP was able to create dose plans that sensibly balanced target and OAR objectives and achieved a higher number of optimization objectives.

In subsection 7.2, we evaluated patients with whole brain irradiation and hippocampus sparing. For these cases, we used a probabilistic expansion of the sensitive organ, to define the desired probabilistic dose limits. The PTP approach achieved desired dose goals for both cases, including for integrated boosts to multiple brain metastases, demonstrating the possibility of probabilistic OAR incorporation into plan optimization.

Finally, in subsection 7.3, we applied the PTP approach on a patient with a continuous dose prescription, or a dose painting case. Here, we used the patient's FDG PET image to define both the desired dose prescription and corresponding target maps for minimum and maximum doses. While these prescriptions resulted in dose plans that appear "noisier" on visual inspection, they nonetheless achieved a dose distribution that is within acceptable deviation of prescribed dose.

The examples presented in this chapter of the thesis are meant as an exploratory evaluation of the PTP approach in different applications, rather than a critical comparison of this approach to classical planning. Because of this, we were able to use datasets with fewer patients than would be normally required for a full comparison. Such a comprehensive analysis would also require an even deeper evaluation of the specifics of each individual disease site and is outside the scope of this thesis but present an avenue for future work that could build on the proof-of-concept demonstrated in this work.

8 Summary and future directions

The main purpose of this thesis is to present and evaluate a novel probabilistic target definition and planning (PTP) approach. We assessed the feasibility of PTP by developing probabilistic tumor maps combined with a new treatment planning optimization engine that could directly intake these probabilistic maps into its treatment optimization. First, we illustrated the PTP concept on an example dataset of patients with glioblastoma. Second, we evaluated the performance of this treatment planning approach and compared it to traditional treatment planning on a prostate cancer example. Finally, we explored the opportunities provided by this treatment planning approach in three other disease sites, each faced with additional treatment planning challenges.

We demonstrated the workflow, including quantification of imaging uncertainties, as well as the performance of the PTP in different anatomical sites over a wide range of challenges. Our results showed that PTP is a viable treatment planning approach that offers opportunity to better and more flexibly incorporate available patient specific information into treatment planning.

8.1 Research summary

8.1.1 Specific Aim 1: To assess feasibility of probabilistic target definition and planning (PTP)

The first goal of assessing PTP feasibility was separated into two parts. The first part was quantifying FET PET repeatability in patients with glioblastoma, as described in Chapter 3. Since any medical imaging modality, including PET, is subject to intrinsic uncertainties, this is a crucial step that allows the differentiation of random uptake fluctuations from statistically significant changes in uptake. Evaluating several different metrics, such as SUV_{mean}, SUV_{max} and $SUV_{98\%}$, for healthy and diseased volumes separately, we found that no statistically significant differences existed between the population averages of test and retest imaging. The metrics for each patient

did vary considerably, with uptake limits of agreement for whole-volume metrics around [0.80, 1.25].

Once we established these limits of repeatability, we used them in treatment planning as described in Chapter 4. We first included the uncertainty information by creating probabilistic gross tumor likelihood maps (gTLM). We then used the TLM directly in treatment planning, using a voxel-weighted optimization function that uses voxel-level TLM values in the calculation of objective function. To account for uncertainties, we used robust optimization with scenario selection of uncertainty realization along the six cardinal directions of target shifts. We compared this planning approach to standard robust optimization and to classical planning approaches, following the GTV-CTV-PTV paradigm prescribed by the previous ICRU reports. Dose plans obtained using RO methodology and classical volumes performed similarly to optimizations using the classical approach, allowing us to validate our implementation of the PTP. Dose plans created using the PTP approach resulted in plans with somewhat wider dose distributions, i.e., higher maximums and lower minimums, than the classically optimized plans, however no significant differences in plan quality were observed using one-sided t-tests evaluating for plan differences. It is also worth noting that the goal of this work was not to create identical dose plans with the different planning approaches, as the optimizations are fundamentally different, and differences in results is to be expected. Rather, general agreement and performance was being sought.

The analyses in these glioblastoma cases served as a demonstration of the general feasibility of the PTP approach. Working on this dataset provided the opportunity to demonstrate how imaging uncertainty analysis can be performed, but also highlighted the often underestimated magnitude of uncertainties that medical imaging is subject to. These cases were also uniquely suited for the initial part of our PTP development, as the skull drastically reduces expected motion, leaving remaining uncertainties mainly affected by microscopic infiltration, and providing only one

generally applicable organ at risk – the healthy brain. While this facilitated the initial planning and analysis, it also raised the question of how the PTP approach performs in more anatomically complex regions, which was the focus of the next specific aim.

8.1.2 Specific Aim 2: To evaluate the performance of the new treatment planning approach

The second aim of this thesis was to evaluate in greater depth how PTP performs in a more complex scenario and how the results in those cases qualitatively compare to the classical planning approach. To find an answer to this question, in Chapter 5 we applied the PTP approach to a cohort of patients with prostate cancer. This selection offered several advantages in evaluation. First, radiotherapy is a very common treatment modality for patients with prostate cancer, making the evaluation of PTP in these cases highly relevant. Second, the prostate is subject to greater positioning uncertainties than brain tumors, allowing us to validate the performance of the planning approach under these uncertainties. Third, while each patient is subject to greater individual uncertainties, planning is more comparable between different patients than in patients with brain tumors, as the exact location of the disease does not impact the dose distribution as much. The dataset we used in our analysis also allowed a different source of tumor mapping and imaging uncertainties, which were obtained by other researchers in a previous analysis, allowing a demonstration of a different initial imaging approach. Finally, recent studies have shown the benefit of integrated dose boost plans as evaluated with patients' disease free survival and side effects. These results justify the exploration of the planning of non-uniform dose prescriptions based on visible disease.

In chapter 5, dose plans obtained using the PTP approach had higher target maximum and mean doses and simultaneously achievied lower doses to bladder and rectum. While both plans are delivered on the same treatment delivery system, PTP is able to intrinsically evaluate trade-offs

between different competing objectives, resulting in plans that favor dose boost magnitude over target dose homogeneity, which is unreliable and optimizationally expensive under uncertainty. With these results we have demonstrated that PTP produces feasible clinical plans in patients with prostate cancer for both plans of uniform prescription. For plans with integrated dose boosts, PTP achieved a better target/OAR ratio, which would likely result in a favorable patient performance based on the results of current clinical trials. Due to its non-reliance on expansion by margins to account for uncertainty, PTP can achieve intrinsic balancing of competing optimization objectives. Additionally, it provides the opportunity for more complex dose planning, which was the research focus more explored in Specific Aim 3.

8.1.3 Specific Aim 3: To explore the opportunity of PTP

In the final section of this thesis, we aimed to explore the opportunities of using PTP for different disease sites and challenges, as shown in Chapter 7.

The first such evaluation was done in patients with head and neck tumors, which present highly complex geometries with several targets of varying dose prescription and close proximity to multiple organs at risk. With many planning objectives often competing against each other and tightly defined optimization objectives to the point where it is impossible to achieve all simultaneously, these geometries presented a unique computational and optimization challenge for the PTP approach. In these cases, PTP achieved comparable performance to classical planning and in some cases achieved better OAR objective adherence, while still delivering sufficient dose to the target. In other cases, no differences were observed, as the classical and PTP approaches achieved comparatively similar results.

The next application of the PTP approach was to cases of patients receiving whole brain irradiation with hippocampus sparing. For these cases, rather than applying a probabilistic target description, we implemented a normal tissue likelihood map (NTLM) for the organ at risk – the

hippocampi. There we showed that the PTP approach was able to optimize dose plans for probabilistically described organs at risk, sparing volumes where possible, but still optimizing for adequate dose to higher priority optimization objectives.

Finally, we used PTP to optimize dose plans for patients with head and neck tumors, where the dose prescription was adjusted based on the patient's FDG PET scan – a dose painting case that cannot be optimized while accounting for uncertainties using the classical dose planning approach. While the populations in these examples were not sufficient to make definitive comparisons, we demonstrated the opportunities of how PTP can be used to capture and parametrize different dose prescriptions.

8.2 Future directions

In this thesis we presented a novel radiation therapy planning and optimization approach that allows for a fundamentally probabilistic incorporation of uncertainties. While we demonstrated the application in a few select cases, there are still many questions that can be explored in further detail with this approach. Additionally, the ability of PTP to capture a broader category of uncertainties makes it a promising candidate for the application to different treatment modalities. In this section we will discuss some of these potential applications in greater detail.

The first avenue of exploration would be the application of PTP to additional disease sites or anatomical considerations. One such example would be patients with lung cancer – either in free breathing or with deep inspiration breath hold technique. Lung targets present an additional challenge not yet evaluated in PTP, as they can exhibit very large motion during treatment. With measured reproducibility or defined limits of breath hold repeatability, PTP would allow for incorporation of multiple inspiration scenarios in planning and dose plan evaluation, potentially

improving plan robustness. We would expect similar results of comparable and potentially more robust plans in application of PTP to other treatment planning sites.

A second, more challenging avenue, is whether probabilistic planning results in a clinically observable benefit to the patients. While the dose plans created with PTP were overall comparable to the plans created with the classical planning approach, validating the non-inferiority or any potential benefits of this approach is still a necessary step.

Another expansion of this approach would be the application of PTP to different treatment modalities, such as proton or brachytherapy. Existing proton therapy is already commonly using robust optimization for their treatment planning due to the strong non-linearity of proton dose deposition. Applying probabilistic target definition and optimization should also integrate well with the commonly used method of pencil beam scanning. Brachytherapy would be a very different implementation, requiring additional uncertainties such as source placement accuracy. However, being able to account for source placement inaccuracy or expected source drift, could lead to more robust dose plans.

Another avenue of potential exploration is biological optimization. With this approach, one directly tries to optimize for the expected tumor control probability (TCP) and normal tissue complication probability (NTCP) values in the optimization, rather than achieving a predefined target dose distribution. While this is a tempting approach, the most commonly used TCP models today are still linear-quadratic, which may not accurately represent the actual tumor response to radiation, based on the results of microbeam irradiation or partial volume irradiation studies, as well as the rarely observed but infamous, abscopal effect.

Regarding our specific implementation of the PTP approach, some upgrades could be made as well, such as improvements of the the mathematical formulation and engine, potentially improving

the speed of optimization and convergence ability of the calculations. As a highly active field of research, different optimization algorithms such as ROME and IPOpt, could be implemented, which could potentially improve the speed and accuracy of calculating the optimization result [149], [150]. This would also allow more complex and better scenario selection methods to be used. Robust optimization approaches for dose planning have been developed on specialized systems that can utilize hundreds of scenarios in various combinations of positioning, target rotation, density, or deformation realizations. While the optimization approach used in this work has been developed with these considerations in mind and could be expanded to account for these cases with relative ease, this would cause the already considerable computational burden to increase exponentially.

8.3 Conclusion

This thesis studies probabilistic target definition and incorporation in treatment planning over various disease sites. It presents a novel quantification of imaging uncertainties, a method of accounting for uncertainties using a probabilistic framework and a dose optimization approach that can directly incorporate these uncertainties into treatment planning. It also includes several applications of this workflow in different disease types: patients with glioblastoma, prostate cancer, head and neck cancers and whole brain irradiation with hippocampus sparing. In all these cases, probabilistic target definition and planning produced feasible dose plans that achieved comparable or better dose benchmarks than dose plans optimized using the classical approach. More importantly, the proposed probabilistic planning is considerably more versatile than the classical dose planning, as it is not limited by the use of margins. While the benefit of these dosimetric advantages would need to be validated in a clinical trial, existing knowledge of response to radiation therapy indicates a favorable, or at least comparable response. Overall, this

work provides extensive evidence supporting the use of a probabilistic approach in radiation therapy.

9 Appendix – Summary of Datasets

1) Sir Charles Gairdner FET Glioblastoma Study
    a) Presented in Chapters 4 and 5
    b) Study population: 24 patients with histologically confirmed glioblastoma grade IV
    c) A longitudinal prospective imaging cohort study in analysis of FET PET/CT imaging to predict prognosis and identify treatment-resistant tumors.
    d) Data: FET PET/CT imaging at baseline, 8 patients with additional retest FET PET/CT imaging, MRI, Classical RT Dose plans
2) The Netherlands Cancer Institute Prostate mp-MRI Repeatability Study
    a) Presented in Chapter 6
    b) Study population: 12 patients with histologically confirmed prostate cancer
    c) The study is a retrospective imaging cohort study acquired to evaluate novel treatment planning approaches in patients with prostate cancers.
    d) Data: MR-derived CT images of 12 patients, MP-MRI derived tumor likelihood maps, contours of prostate, bladder, rectum, and femoral heads
3) Centre Léon Bérard H&N Dataset
    a) Presented in Chapter 7.1
    b) Study population: 4 patients with different cancers of head and neck
    c) This dataset is a sample of patients selected for standard-of-care curative RT for H&N
    d) Data: Contrast enhanced CT with extensive segmentations of disease volumes and pertinent organs at risk
4) Universitaetsklinikum Tuebingen Hippocampus Sparing Dataset

a) Presented in Chapter 7.2
b) Study population: 2 patients with glioblastoma receiving whole brain irradiation with hippocampus sparing
c) This dataset is retrospectively evaluated cohort aimed to evaluate the viability of probabilistic planning in patients
d) Data: CT scans with contours of the hippocampi, metastasis, and pertinent organs at risk

5) Avastin Dataset
a) Presented in Chapter 7.3
b) Study population: 10 patients with head and neck squamous cell carcinoma
c) This phase I study aimed to test the safety of administering Avastin (bevacizumab) to patients with head and neck squamous cell carcinoma.
d) Data: FDG, FLT, Cu-ATSM and DCE CT imaging at baseline, post Avastin and mid chemo-RT.

10 References

[1] T. Landberg *et al.*, “ICRU Report 50: Prescribing, Recording and Reporting Photon Beam Therapy,” *J. Int. Comm. Radiat. Units Meas.*, vol. os26, no. 1, p. NP-NP, Sep. 1993.

[2] T. Landberg *et al.*, “ICRU Report 62: Prescribing, Recodring and Reporting Photon Beam Therapy (Supplement to ICRU Report 50),” *J. Int. Comm. Radiat. Units Meas.*, vol. os32, no. 1, p. NP-NP, Nov. 1999.

[3] V. Gregoire *et al.*, “ICRU Report 83: Prescribing, Recording and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT),” *J. ICRU*, vol. 10, no. 1, p. NP.2-NP, Apr. 2010.

[4] R. J. H. M. Steenbakkers *et al.*, “Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: A ‘Big Brother’ evaluation,” *Radiother. Oncol.*, vol. 77, no. 2, pp. 182–190, 2005.

[5] F. B. M. Van Herk, H. V. F. Vandaele, S. S. K. De Jaeger, and N. J. D. C. Carbaat, “Inter - observer variation of hippocampus delineation in hippocampal avoidance prophylactic cranial irradiation,” *Clin. Transl. Oncol.*, vol. 21, no. 2, pp. 178–186, 2019.

[6] S. Apisarnthanarax *et al.*, “Determining optimal clinical target volume margins in head-and-neck cancer based on microscopic extracapsular extension of metastatic neck nodes,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 3, pp. 678–683, 2006.

[7] J. Unkelbach *et al.*, “Robust radiotherapy planning,” *Phys. Med. Biol.*, vol. 63, no. 22, p. 22TR02, Nov. 2018.

[8] E. Grönlund, S. Johansson, A. Montelius, and A. Ahnesjö, “Dose painting by numbers based on retrospectively determined recurrence probabilities,” *Radiother. Oncol.*, vol. 122, no. 2, pp. 236–241, 2017.

[9] R. Jeraj, T. Bradshaw, U. Simoncic, U. Simon i , and U. Simoncic, “Molecular Imaging to Plan Radiotherapy and Evaluate Its Efficacy,” *J Nucl Med*, vol. 56, no. 11, pp. 1752–1765, 2015.

[10] C. Fiorino, M. Reni, A. Bolognesi, G. M. Cattaneo, and R. Calandrino, “Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning,” *Radiother. Oncol.*, vol. 47, no. 3, pp. 285–292, 1998.

[11] D. F. Dubois, B. R. Prestidge, L. A. Hotchkiss, J. J. Prete, and W. S. Bice, “Intraobserver and interobserver variability of MR imaging- and CT-derived prostate volumes after transperineal interstitial permanent prostate brachytherapy.,” *Radiology*, vol. 207, no. 3, pp. 785–789, Jun. 1998.

[12] A. C. Riegel *et al.*, “Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, no. 3, pp. 726–732, 2006.

[13] G. Galvais, “PhD - GANAR,” *PhD Propos.*, vol. 1, 2015.

[14] M. Lê, J. Unkelbach, N. Ayache, and H. Delingette, “Sampling image segmentations for uncertainty quantification,” *Med. Image Anal.*, vol. 34, pp. 42–51, 2016.

[15] M. Hatt *et al.*, “Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211,” *Med. Phys.*, vol. 44, no. 6, pp. e1–e42, 2017.

[16] J. Unkelbach *et al.*, “Robust radiotherapy planning,” *Phys. Med. Biol.*, vol. 63, no. 22, p. 22TR02, Nov. 2018.

[17] L. C. Fru *et al.*, “Potential role of the glycolytic oscillator in acute hypoxia in tumors.,” *Phys. Med. Biol.*, vol. 60, no. 24, pp. 9215–9225, 2015.

[18] C. Lin *et al.*, “Repeatability of quantitative 18 F-NaF PET: A multicenter study,” *J. Nucl. Med.*, vol. 57, no. 12, pp. 1872–1879, 2016.

[19] M. Niyazi *et al.*, “ESTRO-ACROP guideline ‘target delineation of glioblastomas,’” *Radiother. Oncol.*, vol. 118, no. 1, pp. 35–42, 2016.

[20] K. R. Swanson, J. Alvord E.C., and J. D. Murray, “A quantitative model for differential motility of gliomas in grey and white matter,” *Cell Prolif.*, vol. 33, no. 5, pp. 317–329, 2000.

[21] K. R. Swanson, C. Bridge, J. D. Murray, and E. C. Alvord, “Virtual and real brain tumors: Using mathematical modeling to quantify glioma growth and invasion,” *J. Neurol. Sci.*, vol. 216, no. 1, pp. 1–10, 2003.

[22] P. W. Yuen, K. Y. Lam, A. C. L. Chan, W. I. Wei, and L. K. Lam, “Clinicopathological analysis of local spread of carcinoma of the tongue,” *Am. J. Surg.*, vol. 175, no. 3, pp. 242–244, 1998.

[23] F. Eckert, K. Zwirner, S. Boeke, D. Thorwarth, D. Zips, and S. M. Huber, “Rationale for

combining radiotherapy and immune checkpoint inhibition for patients with hypoxic tumors," *Front. Immunol.*, vol. 10, no. MAR, pp. 1–14, 2019.

[24] L. Alic *et al.*, "Facilitating tumor functional assessment by spatially relating 3D tumor histology and In Vivo MRI: Image registration approach," *PLoS One*, vol. 6, no. 8, pp. 1–10, 2011.

[25] T. F. Mutanga, H. C. J. De Boer, V. Rajan, M. L. P. Dirkx, L. Incrocci, and B. J. M. Heijmen, "Day-to-day reproducibility of prostate intrafraction motion assessed by multiple kV and MV imaging of implanted markers during treatment," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 83, no. 1, pp. 400–407, 2012.

[26] J. Wu *et al.*, "Positioning errors and prostate motion during conformal prostate radiotherapy using on-line isocentre set-up verification and implanted prostate markers," *Radiother. Oncol.*, vol. 61, no. 2, pp. 127–133, 2001.

[27] L. Ma *et al.*, "Impact of millimeter-level margins on peripheral normal brain sparing for gamma knife radiosurgery," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 89, no. 1, pp. 206–213, 2014.

[28] J. A. Lee, C. Y. Kim, Y. J. Park, W. S. Yoon, N. K. Lee, and D. S. Yang, "Interfractional variability in intensity-modulated radiotherapy of prostate cancer with or without thermoplastic pelvic immobilization," *Strahlentherapie und Onkol.*, vol. 190, no. 1, pp. 94–99, 2014.

[29] M. Krengli *et al.*, "Reproducibility of patient setup by surface image registration system in conformal radiotherapy of prostate cancer," *Radiat. Oncol.*, vol. 4, pp. 1–10, 2009.

[30] K. Bell, M. Heitfeld, N. Licht, C. Rübe, and Y. Dzierma, "Influence of daily imaging on plan quality and normal tissue toxicity for prostate cancer radiotherapy," *Radiat. Oncol.*, vol. 12, no. 1, pp. 1–11, 2017.

[31] J. Dekker, T. Rozema, F. Böing-Messing, M. Garcia, D. Washington, and W. de Kruijf, "Whole-brain radiation therapy without a thermoplastic mask," *Phys. Imaging Radiat. Oncol.*, vol. 11, no. July, pp. 27–29, 2019.

[32] R. J. H. M. Steenbakkers *et al.*, "Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 2, pp. 435–448, 2006.

[33] Z. Chen, Z. Yang, J. Wang, and W. Hu, "Dosimetric impact of different bladder and rectum

filling during prostate cancer radiotherapy," *Radiat. Oncol.*, vol. 11, no. 1, pp. 1–8, 2016.

[34] E. J. Hall and A. J. Giaccia, *Radiobiology for the Radiologist*, 7th ed. Lippincott Williams & Wilkins, 2012.

[35] S. C. Formenti and S. Demaria, "Systemic effects of local radiotherapy," *Lancet Oncol.*, vol. 10, no. 7, pp. 718–726, 2009.

[36] M. Z. Dewan *et al.*, "Fractionated but not single-dose radiotherapy induces an immune-mediated abscopal effect when combined with anti-CTLA-4 antibody," *Clin. Cancer Res.*, vol. 15, no. 17, pp. 5379–5388, 2009.

[37] M. Moreau *et al.*, "Priming the Abscopal Effect Using Multifunctional Smart Radiotherapy Biomaterials Loaded with Immunoadjuvants," *Front. Oncol.*, vol. 8, no. March, pp. 1–8, 2018.

[38] D. G. Tempel, N. P. Brodin, and W. A. Tomé, "On the Inclusion of Short-distance Bystander Effects into a Logistic Tumor Control Probability Model," *Cureus*, vol. 10, no. 1, pp. 1–24, 2018.

[39] M. Baumann, M. Krause, and R. Hill, "Exploring the role of cancer stem cells in radioresistance," *Nat. Rev. Cancer*, vol. 8, no. 7, pp. 545–554, 2008.

[40] A. Brahme and A. K. Argren, "Optimal dose distribution for eradication of heterogeneous tumors," *Acta Oncol. (Madr).*, vol. 26, no. 5, pp. 377–385, 1987.

[41] D. Yan, S. Chen, D. J. Krauss, P. Y. Chen, P. Chinnaiyan, and G. D. Wilson, "Tumor Voxel Dose-Response Matrix and Dose Prescription Function Derived Using 18 F-FDG PET/CT Images for Adaptive Dose Painting by Number," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 104, no. 1, pp. 207–218, 2019.

[42] P. Vaupel, "Tumor microenvironmental physiology and its implications for radiation oncology," *Semin. Radiat. Oncol.*, vol. 14, no. 3, pp. 198–206, 2004.

[43] F. Kang *et al.*, "Inter-heterogeneity and intra-heterogeneity of αvβ3 in non-small cell lung cancer and small cell lung cancer patients as revealed by 68Ga-RGD2 PET imaging," *Eur. J. Nucl. Med. Mol. Imaging*, no. 127, 2017.

[44] A. A. Alizadeh *et al.*, "Toward understanding and exploiting tumor heterogeneity," *Nat. Med.*, vol. 21, no. 8, pp. 846–853, 2015.

[45] F. M. Frame *et al.*, "Tumor heterogeneity and therapy resistance - implications for future

treatments of prostate cancer," *J. Cancer Metastasis Treat.*, vol. 3, no. 12, p. 302, 2017.

[46] J. Y. Huang *et al.*, "Investigation of various energy deposition kernel refinements for the convolution / superposition method," *Med. Phys.*, vol. 121721, no. December, pp. 1–10, 2013.

[47] B. J. Blyth and P. J. Sykes, "Radiation-induced bystander effects: What are they, and how relevant are they to human radiation exposures?," *Radiat. Res.*, vol. 176, no. 2, pp. 139–157, 2011.

[48] J. J. Gordon, N. Sayah, E. Weiss, and J. V. Siebers, "Coverage optimized planning: Probabilistic treatment planning based on dose coverage histogram criteria," *Med. Phys.*, vol. 37, no. 2, pp. 550–563, 2010.

[49] M. A. Grotzer, E. Schültke, E. Bräuer-Krisch, and J. A. Laissue, "Microbeam radiation therapy: Clinical perspectives," *Phys. Medica*, vol. 31, no. 6, pp. 564–567, 2015.

[50] U. A. van der Heide, A. C. Houweling, G. Groenendaal, R. G. H. Beets-Tan, and P. Lambin, "Functional MRI for radiotherapy dose painting," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1216–1223, 2012.

[51] S. M. Bentzen and V. Gregoire, "Molecular Imaging–Based Dose Painting: A Novel Paradigm for Radiation Therapy Prescription," *Semin. Radiat. Oncol.*, vol. 21, no. 2, pp. 101–110, Apr. 2011.

[52] M. Witte, G. Shakirin, A. Houweling, H. Peulen, and M. Van Herk, "Dealing with geometric uncertainties in dose painting by numbers: Introducing the ΔvH," *Radiother. Oncol.*, vol. 100, no. 3, pp. 402–406, 2011.

[53] S. Supiot, A. Lisbona, F. Paris, D. Azria, and P. Fenoglietto, "« Dose-painting »: mythe ou réalité ?," *Cancer/Radiotherapie*, vol. 14, no. 6–7, pp. 554–562, 2010.

[54] A. Trofimov, J. Unkelbach, T. F. DeLaney, and T. Bortfeld, "Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands," *Pract. Radiat. Oncol.*, vol. 2, no. 3, pp. 164–171, 2012.

[55] K. Bush, I. M. Gagne, S. Zavgorodni, W. Ansbacher, and W. Beckham, "Dosimetric validation of Acuros® XB with Monte Carlo methods for photon dose calculations," *Med. Phys.*, vol. 38, no. 4, pp. 2208–2221, 2011.

[56] J. Sorriaux *et al.*, "Experimental assessment of proton dose calculation accuracy in

inhomogeneous media," *Phys. Medica*, vol. 38, pp. 10–15, 2017.

[57] E. E. Klein *et al.*, "Task group 142 report: Quality assurance of medical acceleratorsa," *Med. Phys.*, vol. 36, no. 9, pp. 4197–4212, 2009.

[58] C. C. Ling, P. Zhang, Y. Archambault, J. Bocanek, G. Tang, and T. LoSasso, "Commissioning and Quality Assurance of RapidArc Radiotherapy Delivery System," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 72, no. 2, pp. 575–581, 2008.

[59] A. Brahme, "Optimization of stationary and moving beam radiation therapy techniques," *Radiother. Oncol.*, vol. 12, no. 2, pp. 129–140, Jun. 1988.

[60] A. Brahme, J. E. Roos, and I. Lax, "Solution of an integral equation encountered in rotation therapy," *Phys. Med. Biol.*, vol. 27, no. 10, pp. 1221–1229, 1982.

[61] A. M. Cormack, "A problem in rotation therapy with X Rays," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 13, no. 4, pp. 623–630, 1987.

[62] P. Källman, A. Ågren, and A. Brahme, "Tumour and normal tissue responses to fractionated non-uniform dose delivery," *Int. J. Radiat. Biol.*, vol. 62, no. 2, pp. 249–262, 1992.

[63] T. R. Mackie *et al.*, "Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy," *Med. Phys.*, vol. 20, no. 6, pp. 1709–1719, Nov. 1993.

[64] T. Chan, T. Bortfeld, and J. N. Tsitsiklis, "SU-FF-T-53: A Robust Approach to IMRT Optimization," *Med. Phys.*, vol. 33, no. 6, p. 2061, 2006.

[65] J. Unkelbach, T. Bortfeld, B. C. Martin, and M. Soukup, "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning," *Med. Phys.*, vol. 36, no. 1, pp. 149–163, 2008.

[66] S. Prajapati, T. R. Mackie, and R. Jeraj, "Open-Source Medical Devices (OSMD) Design of a Small Animal Radiotherapy System," *J. Phys. Conf. Ser.*, vol. 489, no. 1, p. 012017, Mar. 2014.

[67] D. Wolff *et al.*, "Volumetric modulated arc therapy (VMAT) vs. serial tomotherapy, step-and-shoot IMRT and 3D-conformal RT for treatment of prostate cancer," *Radiother. Oncol.*, vol. 93, no. 2, pp. 226–233, 2009.

[68] T. Piotrowski, M. Skórska, and A. Jodda, "Tomotherapy – a different way of dose delivery in radiotherapy," vol. 16, no. 1, pp. 16–25, 2012.

[69] T. Landberg, P. Almond, J. M. V. Burgers, M. Busch, C. A. Joslin, and J. P. Paunier, "ICRU

Report 29: Dose Specification for Reporting External Beam Therapy with Photons and Electrons," *J. Int. Comm. Radiat. Units Meas.*, vol. os15, no. 2, p. NP-NP, Apr. 1978.

[70] M. Van Herk, "Errors and Margins in Radiotherapy," *Semin. Radiat. Oncol.*, vol. 14, no. 1, pp. 52–64, 2004.

[71] A. Bel, M. van Herk, and J. V. Lebesque, "Target margins for random geometrical treatment uncertainties in conformal radiotherapy," *Med. Phys.*, vol. 23, no. 9, pp. 1537–1545, Sep. 1996.

[72] P. Aaltonen *et al.*, "Specification of Dose Delivery in Radiation Therapy. Recommendations by the Nordic Association of Clinical Physics (NACP)," *Acta Oncol. (Madr).*, vol. 36, no. sup10, pp. 1–32, Jan. 1997.

[73] J. H. Killoran, H. M. Kooy, D. J. Gladstone, F. J. Welte, and C. J. Beard, "A numerical simulation of organ motion and daily setup uncertainties: Implications for radiation therapy," *Int. J. Radiat. Oncol.*, vol. 37, no. 1, pp. 213–221, Jan. 1997.

[74] J. C. Stroom, H. C. J. De Boer, H. Huizenga, and A. G. Visser, "Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 43, no. 4, pp. 905–919, 1999.

[75] M. Van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, "The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 4, pp. 1121–1135, 2000.

[76] M. Chu, Y. Zinchenko, S. G. Henderson, and M. B. Sharpe, "Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty," *Phys. Med. Biol.*, vol. 50, no. 23, pp. 5463–5477, 2005.

[77] T. C. Y. Chan, T. Bortfeld, and J. N. Tsitsiklis, "A robust approach to IMRT optimization," *Phys. Med. Biol.*, vol. 51, no. 10, pp. 2567–2583, 2006.

[78] J. Unkelbach, T. C. Y. Chan, and T. Bortfeld, "Accounting for range uncertainties in the optimization of intensity modulated proton therapy," *Phys. Med. Biol.*, vol. 52, no. 10, 2007.

[79] Z. Perkó, S. R. van der Voort, S. van de Water, C. M. H. Hartman, M. Hoogeman, and D. Lathouwers, "Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion," *Phys. Med. Biol.*, vol. 61, no. 12, pp. 4646–4664, 2016.

[80] S. Van Der Voort, S. Van De Water, Z. Perkó, B. Heijmen, D. Lathouwers, and M.

Hoogeman, “Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 95, no. 1, pp. 163–170, 2016.

[81] A. Fredriksson, A. Forsgren, and B. Hårdemark, “Minimax optimization for handling range and setup uncertainties in proton therapy,” *Med. Phys.*, vol. 38, no. 3, pp. 1672–1684, 2011.

[82] D. Fontanarosa *et al.*, “An in silico comparison between margin-based and probabilistic target-planning approaches in head and neck cancer patients,” *Radiother. Oncol.*, vol. 109, no. 3, pp. 430–436, 2013.

[83] R. Bohoslavsky, M. G. Witte, T. M. Janssen, and M. Van Herk, “Probabilistic objective functions for margin-less IMRT planning,” *Phys. Med. Biol.*, vol. 58, no. 11, pp. 3563–3580, 2013.

[84] M. G. Witte, J. J. Sonke, J. Siebers, J. O. Deasy, and M. Van Herk, “Beyond the margin recipe: The probability of correct target dosage and tumor control in the presence of a dose limiting structure,” *Phys. Med. Biol.*, vol. 62, no. 19, pp. 7874–7888, 2017.

[85] T. Win *et al.*, “Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer,” *Clin. Cancer Res.*, vol. 19, no. 13, pp. 3591–3599, 2013.

[86] T. J. Bradshaw, S. R. Bowen, N. Jallow, L. J. Forrest, and R. Jeraj, “Heterogeneity in Intratumor Correlations of 18F-FDG, 18F-FLT, and 61Cu-ATSM PET in Canine Sinonasal Tumors,” *J. Nucl. Med.*, vol. 54, no. 11, pp. 1931–1937, 2013.

[87] M. Hatt, F. Tixier, C. Cheze Le Rest, O. Pradier, and D. Visvikis, “Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 40, no. 11, pp. 1662–1671, 2013.

[88] F. J. Lagerwaard *et al.*, “Multiple ‘slow’ CT scans for incorporating lung tumor mobility in radiotheraphy planning,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 51, no. 4, pp. 932–937, 2001.

[89] N. Shusharina, D. Craft, Y. L. Chen, H. Shih, and T. Bortfeld, “The clinical target distribution: A probabilistic alternative to the clinical target volume,” *Phys. Med. Biol.*, vol. 63, no. 15, p. aacfb4, 2018.

[90] J. Unkelbach *et al.*, “The role of computational methods for automating and improving clinical target volume definition,” *Radiother. Oncol.*, vol. 153, pp. 15–25, 2020.

[91] T. Bortfeld, N. Shusharina, and D. L. Craft, “Probabilistic definition of the clinical target volume-implications for tumor control probability modeling and optimization,” *Phys. Med. Biol.*, vol. 66, no. 1, pp. 0–18, Nov. 2021.

[92] E. Sterpin, S. T. Rivas, F. Van den Heuvel, B. George, J. A. Lee, and K. Souris, “Development of robustness evaluation strategies for enabling statistically consistent reporting,” *Phys. Med. Biol.*, vol. 66, no. 4, p. 045002, Feb. 2021.

[93] S. Campbell *et al.*, “Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: Implications for the management of head-and-neck cancers,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 82, no. 2, pp. 574–581, 2012.

[94] S. C. M. M. Ten Eikelder, P. Ferjančič, A. Ajdari, T. Bortfeld, D. den Hertog, and R. Jeraj, “Optimal treatment plan adaptation using mid-treatment imaging biomarkers,” *Phys. Med. Biol.*, vol. 65, no. 24, Oct. 2020.

[95] E. Sterpin, S. T. Rivas, F. Van den Heuvel, B. George, J. A. Lee, and K. Souris, “Development of robustness evaluation strategies for enabling statistically consistent reporting,” *Phys. Med. Biol.*, vol. 66, no. 4, p. 045002, Feb. 2021.

[96] S. Teoh, B. George, F. Fiorini, K. Vallis, and F. Van den Heuvel, “Robustness assessment using probabilistic scenarios of intensity modulated proton therapy and volumetric arc therapy in non-small-cell lung cancer: an in-silico radiotherapy planning study,” *Lancet*, vol. 389, p. S94, 2017.

[97] W. U. Shipley *et al.*, “Proton Radiation as Boost Therapy for Localized Prostatic Carcinoma,” *JAMA J. Am. Med. Assoc.*, vol. 241, no. 18, pp. 1912–1915, 1979.

[98] B. Vanderstraeten, W. De Gersem, W. Duthoy, W. De Neve, and H. Thierens, “Implementation of biologically conformal radiation therapy (BCRT) in an algorithmic segmentation-based inverse planning approach,” *Phys. Med. Biol.*, vol. 51, no. 16, 2006.

[99] M. Hatt *et al.*, “Reproducibility of 18F-FDG and 3’-Deoxy-3’-18F-Fluorothymidine PET Tumor Volume Measurements,” *J. Nucl. Med.*, vol. 51, no. 9, pp. 1368–1376, 2010.

[100] B. F. Kurland *et al.*, “Test–retest reproducibility of 18F-FDG PET/CT uptake in cancer patients within a qualified and calibrated local network,” *J. Nucl. Med.*, vol. 60, no. 5, pp. 608–614, 2019.

[101] M. A. Lodge *et al.*, “Repeatability of 18F-FLT PET in a multicenter study of patients with high-grade glioma,” *J. Nucl. Med.*, vol. 58, no. 3, pp. 393–398, 2017.

[102] P. Y. Wen and S. Kesari, “Malignant Gliomas in Adults,” *N. Engl. J. Med.*, vol. 359, no. 5, pp. 492–507, Jul. 2008.

[103] D. R. Johnson and B. P. O’Neill, “Glioblastoma survival in the United States before and during the temozolomide era,” *J. Neurooncol.*, vol. 107, no. 2, pp. 359–364, 2012.

[104] M. Lundemann *et al.*, “Feasibility of multi-parametric PET and MRI for prediction of tumour recurrence in patients with glioblastoma,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 46, no. 3, pp. 603–613, Mar. 2019.

[105] H. Duffau and E. Mandonnet, “The ‘onco-functional balance’ in surgery for diffuse low-grade glioma: Integrating the extent of resection with quality of life,” *Acta Neurochir. (Wien).*, vol. 155, no. 6, pp. 951–957, 2013.

[106] S. Moller *et al.*, “Prognostic value of 18F-FET PET imaging in re-irradiation of high-grade glioma: Results of a phase I clinical trial,” *Radiother. Oncol.*, vol. 121, no. 1, pp. 132–137, Oct. 2016.

[107] C. Debus *et al.*, “Impact of 18F-FET PET on Target Volume Definition and Tumor Progression of Recurrent High Grade Glioma Treated with Carbon-Ion Radiotherapy,” *Sci. Rep.*, vol. 8, no. 1, p. 7201, Dec. 2018.

[108] K. J. Langen *et al.*, “Imaging of amino acid transport in brain tumours: Positron emission tomography with O-(2-[18F]fluoroethyl)-L-tyrosine (FET),” *Methods*, vol. 130, no. 2, pp. 124–134, 2017.

[109] K. J. Langen *et al.*, “O-(2-[18F]fluoroethyl)-l-tyrosine: uptake mechanisms and clinical applications,” *Nucl. Med. Biol.*, vol. 33, no. 3, pp. 287–294, 2006.

[110] H. Vees, S. Senthamizhchelvan, R. Miralbell, D. C. Weber, O. Ratib, and H. Zaidi, “Assessment of various strategies for 18F-FET PET-guided delineation of target volumes in high-grade glioma patients,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 36, no. 2, pp. 182–193, 2009.

[111] N. Verburg *et al.*, “Direct comparison of [11C] choline and [18F] FET PET to detect glioma infiltration: a diagnostic accuracy study in eight patients,” *EJNMMI Res.*, vol. 9, no. 1, p. 57, Dec. 2019.

[112] P. Heiss, S. Mayer, M. Herz, H. J. Wester, M. Schwaiger, and R. Senekowitsch-Schmidtke, “Investigation of transport mechanism and uptake kinetics of O-(2- [18F]fluoroethyl)-L-tyrosine in vitro and in vivo,” *J. Nucl. Med.*, vol. 40, no. 8, pp. 1367–1373, Aug. 1999.

[113] A. L. Grosu *et al.*, “An interindividual comparison of O-(2-[18F]fluoroethyl)-L- tyrosine (FET)- and L-[methyl-11C]methionine (MET)-PET in patients with brain gliomas and metastases,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 81, no. 4, pp. 1049–1058, 2011.

[114] V. Dunet, C. Rossier, A. Buck, R. Stupp, and J. O. Prior, “Performance of 18F-Fluoro-Ethyl-Tyrosine (18F-FET) PET for the Differential Diagnosis of Primary Brain Tumor: A Systematic Review and Metaanalysis,” *J. Nucl. Med.*, vol. 53, no. 2, pp. 207–214, 2012.

[115] B. Suchorska *et al.*, “Biological tumor volume in 18FET-PET before radiochemotherapy correlates with survival in GBM,” *Neurology*, vol. 84, no. 7, pp. 710–719, Feb. 2015.

[116] C. Debus *et al.*, “Feasibility and robustness of dynamic 18F-FET PET based tracer kinetic models applied to patients with recurrent high-grade glioma prior to carbon ion irradiation,” *Sci. Rep.*, vol. 8, no. 1, p. 14760, 2018.

[117] A. Benouaich-Amiel *et al.*, “Evaluation of O-(2-[18F]-Fluoroethyl)-L-Tyrosine in the Diagnosis of Glioblastoma,” *Arch. Neurol.*, vol. 67, no. 3, pp. 370–372, 2010.

[118] C. Lin *et al.*, “Repeatability of Quantitative 18F-NaF PET: A Multicenter Study,” *J. Nucl. Med.*, vol. 57, no. 12, pp. 1872–1879, 2016.

[119] J. M. Bland and D. G. Altman, “Agreement between methods of measurement with multiple observations per individual,” *J. Biopharm. Stat.*, vol. 17, no. 4, pp. 571–582, 2007.

[120] M. A. Lodge, “Repeatability of SUV in Oncologic 18 F-FDG PET,” *J. Nucl. Med.*, vol. 58, no. 4, pp. 523–532, Apr. 2017.

[121] L. M. Velasquez *et al.*, “Repeatability of 18F-FDG PET in a Multicenter Phase I Study of Patients with Advanced Gastrointestinal Malignancies,” *J. Nucl. Med.*, vol. 50, no. 10, pp. 1646–1654, 2009.

[122] A. J. De Langen *et al.*, “Repeatability of 18F-FDG Uptake Measurements in Tumors: A Metaanalysis,” *J. Nucl. Med.*, vol. 53, no. 5, pp. 701–708, 2012.

[123] J. K. Hoang, S. K. Das, K. R. Choudhury, D. S. Yoo, and D. M. Brizel, “Using FDG-PET to Measure Early Treatment Response in Head and Neck Squamous Cell Carcinoma: Quantifying Intrinsic Variability in Order to Understand Treatment-Induced Change,” *Am. J. Neuroradiol.*, vol. 34, no. 7, pp. 1428–1433, Jul. 2013.

[124] J. Gempt *et al.*, “Volumetric analysis of F-18-FET-PET imaging for brain metastases,” *World Neurosurg.*, vol. 84, no. 6, pp. 1790–1797, 2015.

[125] P. E. Kinahan and J. W. Fletcher, “Positron Emission Tomography-Computed Tomography Standardized Uptake Values in Clinical Practice and Assessing Response to Therapy,” *Semin. Ultrasound, CT MRI*, vol. 31, no. 6, pp. 496–505, Dec. 2010.

[126] M. Unterrainer *et al.*, “Towards standardization of18F-FET PET imaging: Do we need a consistent method of background activity assessment?,” *EJNMMI Res.*, vol. 7, pp. 1–8, 2017.

[127] M. Reuter, E. R. Gerstner, O. Rapalino, T. T. Batchelor, B. Rosen, and B. Fischl, “Impact of MRI head placement on glioma response assessment,” *J. Neurooncol.*, vol. 118, no. 1, pp. 123–129, 2014.

[128] E. Konukoglu, O. Clatz, P. Y. Bondiau, H. Delingette, and N. Ayache, “Extrapolating glioma invasion margin in brain magnetic resonance images: Suggesting new irradiation margins,” *Med. Image Anal.*, vol. 14, no. 2, pp. 111–125, 2010.

[129] A. D. Yock *et al.*, “Robustness Analysis for External Beam Radiation Therapy Treatment Plans : Describing Uncertainty Scenarios and Reporting Their Dosimetric Consequences,” 2019.

[130] I. M. Lips *et al.*, “Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): Study protocol for a randomized controlled trial,” *Trials*, vol. 12, no. December, 2011.

[131] C. Draulans *et al.*, “Primary endpoint analysis of the multicentre phase II hypo-FLAME trial for intermediate and high risk prostate cancer,” *Radiother. Oncol.*, vol. 147, pp. 92–98, 2020.

[132] L. G. W. Kerkmeijer *et al.*, “Focal Boost to the Intraprostatic Tumor in External Beam Radiotherapy for Patients With Localized Prostate Cancer: Results From the FLAME Randomized Phase III Trial,” *J. Clin. Oncol.*, vol. 39, no. 7, p. JCO.20.02873, 2021.

[133] M. A. Van Schie *et al.*, “Repeatability of dose painting by numbers treatment planning in prostate cancer radiotherapy based on multiparametric magnetic resonance imaging,” *Phys. Med. Biol.*, vol. 62, no. 14, pp. 5575–5588, 2017.

[134] C. V. Dinh *et al.*, “Magnetic resonance imaging for prostate cancer radiotherapy,” *Phys. Medica*, vol. 32, no. 3, pp. 446–451, 2016.

[135] C. V. Dinh *et al.*, “Multicenter validation of prostate tumor localization using multiparametric MRI and prior knowledge,” *Med. Phys.*, vol. 44, no. 3, pp. 949–961, 2017.

[136] A. Koritnik, “Verjetnostna segmentacija tumorjev v področju glave in vratu,” University of Ljubljana, 2019.

[137] M. S. Lee *et al.*, “Is Prostate Imaging Reporting and Data System Version 2 Sufficiently Discovering Clinically Significant Prostate Cancer? Per-Lesion Radiology-Pathology Correlation Study,” *Am. J. Roentgenol.*, vol. 211, no. 1, pp. 114–120, Jul. 2018.

[138] D. C. Johnson *et al.*, “Detection of Individual Prostate Cancer Foci via Multiparametric Magnetic Resonance Imaging,” *Eur. Urol.*, vol. 75, no. 5, pp. 712–720, May 2019.

[139] P. J. Houdt *et al.*, “Histopathological Features of MRI-Invisible Regions of Prostate Cancer Lesions,” *J. Magn. Reson. Imaging*, vol. 51, no. 4, pp. 1235–1246, Apr. 2020.

[140] H. Ariyaratne, H. Chesham, J. Pettingell, and R. Alonzi, “Image-guided radiotherapy for prostate cancer with cone beam CT: dosimetric effects of imaging frequency and PTV margin,” *Radiother. Oncol.*, vol. 121, no. 1, pp. 103–108, 2016.

[141] O. J. Gurney-Champion *et al.*, “Quantitative imaging for radiotherapy purposes,” *Radiother. Oncol.*, vol. 146, pp. 66–75, 2020.

[142] M. G. Witte, J. Sonke, and J. Siebers, “Beyond the margin recipe : the probability of correct target dosage and tumor control in the presence of a dose limiting structure Beyond the margin recipe : the probability of correct target dosage and tumor control in the presence of a dose limiting str,” *Phys. Med. Biol.*, p. aa87fe, 2017.

[143] A. Grosu *et al.*, “Whole-brain irradiation with hippocampal sparing and dose escalation on metastases: neurocognitive testing and biological imaging (HIPPORAD) – a phase II prospective randomized multicenter trial (NOA-14, ARO 2015–3, DKTK-ROG),” *BMC Cancer*, vol. 20, no. 1, p. 532, Dec. 2020.

[144] H. J. W. L. Aerts, P. Lambin, and D. De Ruysscher, “FDG for dose painting: A rational choice,” *Radiother. Oncol.*, vol. 97, no. 2, pp. 163–164, 2010.

[145] C. C. Ling *et al.*, “Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 3, pp. 551–560, 2000.

[146] D. Thorwarth, S. M. Eschmann, F. Paulsen, and M. Alber, “Hypoxia Dose Painting by Numbers: A Planning Study,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 1, pp. 291–300, 2007.

[147] S. Differding *et al.*, “Radiation dose escalation based on FDG-PET driven dose painting by numbers in oropharyngeal squamous cell carcinoma: a dosimetric comparison between TomoTherapy-HA and RapidArc,” *Radiat. Oncol.*, vol. 12, no. 1, p. 59, 2017.

[148] I. Madani *et al.*, “Maximum tolerated dose in a phase i trial on adaptive dose painting by numbers for head and neck cancer,” *Radiother. Oncol.*, vol. 101, no. 3, pp. 351–355, 2011.

[149] J. Goh and M. Sim, “Robust optimization made easy with ROME,” *Oper. Res.*, vol. 59, no. 4, pp. 973–985, 2011.

[150] L. T. Biegler and V. M. Zavala, “Large-scale nonlinear programming using IPOPT: An integrating framework for enterprise-wide dynamic optimization,” *Comput. Chem. Eng.*, vol. 33, no. 3, pp. 575–582, 2009.