Mitigating anatomical changes in head and neck radiotherapy

Ying Zhang

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Medical Physics and Biomedical Engineering

University College London

March 24, 2022

Supervisors:

Gary Royle – Proton and Advanced Radiotherapy Group, UCL

Esther Bär – Proton and Advanced Radiotherapy Group, UCL

I, Ying Zhang, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This thesis aims to address the issue of anatomical changes during the treatment of head and neck cancer patients. Intensity-modulated proton therapy (IMPT) shows advantages in delivering a conformal dose to the target while minimizing the dose to the adjacent normal tissue. However, in the delivery of IMPT, patients not being static can lead to dosimetric discrepancies. This thesis explores different techniques to reduce the influence of uncertainty from anatomical changes (anatomical uncertainty).

In Chapter 3, anatomical variations are incorporated to reveal the robustness of a plan, thus improving the selection of a robust plan. The benefit is demonstrated in all beam arrangements used in the study. In Chapter 4, a probability model is developed to simulate major anatomical deformations at each weekly time point based on population data. For overall anatomical uncertainty prediction during treatment, the PM reduces anatomical uncertainty from 3.72$\pm$0.46 mm (no model) to 0.81$\pm$0.56 mm on average. In Chapter 5, an individual model is developed to refine the systematic prediction of population data using individual progressive information. Compared with no model, whose average anatomical uncertainty and gamma index are 4.47$\pm$1.23 mm and 93.87$\pm$0.83% at week 6, respectively, the refined individual model reduces anatomical uncertainty to 1.89$\pm$1.23 mm and improves the gamma index to 96.16$\pm$1.84% at week 6. In Chapter 6, a proposed flexible strategy creates upfront predicted replans on the predicted individual geometries from the RIM. This application of the RIM reduces the parotid D_{mean} by 0.53 Gy on average, achieving the minimum benefit of non-delay treatment, while the dosimetric of other organs has no significant difference with the standard replanning technique ($p>0.05$). This

prediction-based replanning improves clinical workflow efficiency.

This research integrates computational methods into IMPT treatment for head and neck patients. New perspectives for mitigating anatomical uncertainty are provided and discussed in this thesis.

Acknowledgements

I would like first to express my sincere gratitude to my supervisors. Firstly, to Gary Royle provided me with the opportunity to work with a group of excellent people to finish my PhD project. His guidance and support accompanied me throughout the PhD. To Esther Bär for all the patience, time, support, and encouragement you gave to me, and for all the knowledge and skills you shared with me. To Stacey McGowan, although you were only being my supervisor for less than one year, you guided me into this field. Without my supervisors, this work would not have been possible.

Then I would like specially thank my colleagues, collaborators and friends: Jailan Alshaikhi, Wenyong Tan, Richard Amos, and Matthew Lowe for their valuable input to my work. The time spent with you has been my most precious experience in the past three years.

Thank you for the amazing people I shared the office with. Boyue Ding, Charles-Antoine Collins-Fekete, Megan Wilson, Reem Ahmad, Yaru Pang and Zhuoyan Shen always being there for chats, coffee breaks, drinks and parties. I feel really lucky to have all of you.

Thank you for my dearest friends. Ruijiao Dai, even though we are not in the same place, you are always there to chat, share, and support. Yuting Huo often shares her lovely daughter to light me up.

A big thank you to my family members who always supported me, assured me and comforted me when I was upset. Especially to my partner, Yanze Liu, for taking good care of me and always cheering me up. Thank you for being always there for me.

Contents

List of Figures

List of Tables

Acronyms

AAHUD Average absolute HU difference.

AM Average model.

CBCT Cone-beam CT.

C-DIR evaluation The DIR uncertainty evaluation based on contour.

$\mathbf{CT}_t^{\mathbf{PM},i}$ i-th predicted CT of the PM at week t.

$\mathbf{CT}^{\mathbf{sNRV}}$ CT simulated a sNRV.

CTV Clinical target volume.

$\mathbf{D_2}$ The minimum dose to the hottest 2% volume.

$\mathbf{D_{95}}$ The minimum dose to 95% of target volume.

$\mathbf{D_{max}}$ The maximum dose in the volume.

$\mathbf{D_{mean}}$ The mean dose in the volume.

dCT Deformed rCT (rCT is the float image in registration).

dCT-DIR evaluation The difference between accumulated dose of dCTs and corresponding pCT is used to evaluation DIR uncertainty, referred to as dCT-DIR evaluation.

DDVH Dose deviation volume histogram.

DIR Deformable image registration.

DT Distance transformation.

DTA Distance to agreement.

DVF Deformation vector field.

DVH Dose volume histogram.

GTV Gross target volume.

H&N Head and neck.

HU Housfield unit.

IGRT Image-guided radiotherapy.

IMPT Intensity modulated proton therapy.

MADD Mean value of voxel-wise absolute dose difference.

MC Monte Carlo simulation.

MSD Mean surface distance.

NPC Nasopharyngeal carcinoma.

NTCP Normal tissue complication probability.

OARs Organs at risk.

PC Principal component.

PCA Principal component analysis.

pCT Planning CT.

PD_r The predicted image set to create the predicted replan).

PDF Probability density function.

PM Probability model.

PTV Planning target volume.

QA Quality assurance.

RBE Relative biological effectiveness.

rCT Weekly repeat CT.

R-DIR evaluation The DIR uncertainty evaluation based on proton spot range.

RIM Refined individual model.

RPF Reactive predicted plan on flexible week.

RPS Reactive predicted plan on scheduled week.

RSP Relative stopping power.

SLD Spot location deviation.

sNRVs Small non-rigid variations.

SPS Scheduled predicted plan on scheduled week.

SVF Stationary velocity field.

$\mathbf{V}_{95}$ The percent volume that received at least 95% of the prescription dose.

$\boldsymbol{v}_{a,t\to p}$ $\boldsymbol{v}_{t\to p}$ in the atlas space.

$\boldsymbol{v}_{p\to a}$ The SVF between pCT and the atlas.

$\boldsymbol{v}_{p\to t}$ The SVF that registered rCT_t to pCT.

$\boldsymbol{v}_{t\to p}$ The SVF that registered pCT to rCT_t.

WD Wasserstein distance.

WEPL Water equivalent path length.

WSLD Weighted spot location deviation.

$\Delta\mathbf{WSLD}_t^{\mathbf{res}}$ The difference between the estimated anatomical uncertainty from a model and actual anatomical uncertainty at treatment week t in WSLD.

Publications

The work presented in this thesis has resulted in the following outputs.

Journal papers:

- **Ying Zhang**, Stacey McGowan Holloway, Megan Zoë Wilson, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. ”DIR-based models to predict weekly anatomical changes in head and neck cancer proton therapy”. Physics in Medicine & Biology 67, no. 9 (2022): 095001.

- **Ying Zhang**, Jailan Alshaikhi, Richard A. Amos, Matthew Lowe, Wenyong Tan, Esther Bär and Gary Royle. ”Improving workflow for adaptive proton therapy with predictive anatomical modelling: A proof of concept.” Radiotherapy and Oncology, 173 (2022): 93-101.

- **Ying Zhang**, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. ”A probability model for anatomical robust optimisation in head and neck cancer proton therapy.” Physics in Medicine & Biology (2022).

- **Ying Zhang**, Jailan Alshaikhi, Richard A. Amos, Wenyong Tan, Yaru Pang, Gary Royle and Esther Bär. ”Pre-treatment analysis of non-rigid variations can assist robust IMPT plan selection for head and neck patients.” Medical Physics, 49 (2022):7683-7693.

- Bär, Esther, Charles-Antoine Collins-Fekete, Vasilis Rompokos, **Ying Zhang**, Mark N. Gaze, Alison Warry, Andrew Poynter, and Gary Royle. ”Assessment of the impact of CT calibration procedures for proton therapy planning on paediatric treatments.” Medical physics (2021).

Oral presentations:

- **Ying Zhang**, Jailan Alshaikhi, Richard Amos, Wenyong Tan, Gary Royle, Esther Baer. Small Non-rigid Variations can Assist Robust IMPT Plan Selection for Head and Neck Patients. American Society for Radiation Oncology 2021.

- **Ying Zhang**, Jailan Alshaikhi, Richard A. Amos, Wenyong Tan, Gary Royle , Esther Bär. Systematic progression changes can assist robust IMPT plan selection for head and neck patients. European Society for Radiotherapy and Oncology 2022.

- Esther Bär, Charles-Antoine Collins-Fekete, Vasilis Rompokos, **Ying Zhang**, Mark N. Gaze, Alison Warry, Andrew Poynter, Gary Royle. CT calibration for precise proton therapy planning in children. European Society for Radiotherapy and Oncology 2021.

- Yaru Pang, Michael Kosmin, Zhuangling Li, Xiaonian Deng, Zihuang Li, Xianming Li, **Ying Zhang**, Boyue Ding, Gary Royle, Spyros Manolopoulos. MRI-based tumour control probability in Glioblastoma treated with radiotherapy. European Society for Radiotherapy and Oncology 2021.

Poster presentations:

- **Ying Zhang**, Megan Z.Wilson,Jeffrey Liu,Jailan Alshaikhi, Gary Royle, Stacey M.Holloway. Can Proton Water Equivalent Path Length calculations be used instead of full dose recalculation for determining when to adapt a plan? PPRIG Proton Physics Workshop 5, National Physical Laboratory(NPL),UK. Feb 2019.

- **Ying Zhang**, Jailan Alshaikhi, Richard Amos, Wenyong Tan, Gary Royle, Esther Baer. First Application of Predictive Model to Assist Adaptive Proton Therapy. American Society for Radiation Oncology 2021.

Chapter 1

Introduction

1.1 Head and neck proton therapy

Radiotherapy is often delivered as a definitive or adjuvant treatment for cancer patients. Over 75% of head and neck (H&N) patients undergo radiotherapy as part of their treatment [1]. Because cancer cells in the H&N region are often surrounded by sensitive organs and structures, such as parotid glands, oral cavity, brainstem and optic chiasm, the radiation damages on these structures are unavoidable, leading to complications such as dysphagia (swallowing difficulties) and dysgeusia (taste changes). These complications can degrade patients' quality of life for a long time. Intensity-modulated proton therapy (IMPT) as the state-of-art radiotherapy technique has advantages in delivering a conformal dose to the target while minimizing the dose to the adjacent normal tissue [2, 3, 4, 5], exploiting the steep falloff of the Bragg peak. The potential benefits of proton therapy over photon therapy on H&N cancer treatment have been revealed in the literature [2, 3, 5]. However, this precise delivery technique has inherent sensitivity to uncertainties, which are especially common in H&N cancer treatment.

1.2 Uncertainties in proton therapy and evaluation

Uncertainties degrade the quality of treatment. Systematically exploring uncertainties can help design congruous mitigation methods. Several studies have reported the magnitude of uncertainties in proton therapy. This section summarises their uncertainty evaluation method and conclusions. Uncertainties are divided into 1)

uncertainty before planning, 2) uncertainty in planning, 3) uncertainty during treatment and 4) additional uncertainty from using deformable image registration.

1.2.1 Uncertainties before planning

The uncertainty exists in contour delineation. Till today, no image modalities can accurately pinpoint the tumour cells. Therefore, the contours of the gross tumour volume (GTV) and the clinic treatment volume (CTV) are only the subjective judgment of oncologists through the visual surrogate of the present tumour. Delineation uncertainty reported by Rasch *et al.* (2010) [6] was 3.3 mm for high-risk CTV and 4.9 mm for low-risk CTV based on co-registered MRI-CT of nasopharyngeal carcinoma (NPC) patients. Similar 3.2 mm delineation uncertainty for CTVs was reported by Aznar *et al.* (2017) using PET/CT of Hodgkin lymphoma [7].

The uncertainty also exists in tumour prescription dose and normal tissue tolerance. Theoretically, patient-specific prescriptions should be given based on the characteristics of tumour cells, for example, the distribution of clonogenic and hypoxia tumour cells [8, 9, 10], tumour sensitivity to radiotherapy fraction [11], size and tumour genotype [12]. However, these microscopic markers have yet to be widely used in prescription. For organs at risk (OARs) constraints, especially serial organs, the headroom of dose constraints is unknown because no clinical trial will be allowed to explore the dose limitation with the risk that might be imposed on patients' quality of life. In summary, the planned dose objectives come with uncertainty in the first place.

1.2.2 Uncertainties in planning

The uncertainty originates from the CT-based plan due to the CT imaging technique. Noise, CT artefacts, beam hardening, and density heterogeneity lead to the inaccuracy in Hounsfield Units (HU), which will be converted to relative stopping power (RSP) for dose calculation, specifically referred to as range uncertainty. The magnitude of range uncertainty depends on dose calculation algorithms. Yang *et al*. (2012) [13] reported that range uncertainty can be as high as 5% for lung, 2.4% for bone and 1.6% for soft tissue in a Monte Carlo simulation (MC) based dose

calculation algorithm. The typical value of 3.5% is generally used in H&N cancer treatment. When the analytical dose calculation is used, an additional 2.2% uncertainty will be brought in from physics and CT conversion [14]. Yepes *et al*. (2018) [15] reported that an analytical dose calculation can result in a 10% overestimation of the target dose and up to a 10 Gy underestimation of the dose to some OARs. Even though a MC-based dose calculation is more accurate than an analytical dose calculation, it still has uncertainty in modelling the degradation of the Bragg peak in heterogeneous media.

1.2.3 Uncertainties during treatment

The irradiation is delivered in fractions during radiotherapy to maximise the destruction of tumour cells while minimising damage to healthy tissue. However, while patients are assumed to be static at the planning stage, this fractionation introduces the following uncertainties:

Firstly, the uncertainty comes from beam reproducibility and patient setup. The magnitude adopted is centre-specific. For example, the MD Anderson Proton Therapy Center in Houston, the Loma Linda University Medical Center and the Roberts Proton Therapy Center at the University of Pennsylvania use 3 mm, while the University of Florida Proton Therapy Institute uses 1.5 mm. Please note that these margins are not fully generic and can be adjusted based on the treatment site and location of critical structures.

Secondly, the uncertainty comes from anatomical variations (anatomical uncertainty), including small non-rigid variations (sNRVs) and progressive changes. As H&N cancer is the focus of this thesis, the anatomical changes in this specific site are exclusively reviewed. For sNRVs, nasal filling, jaw movement, neck folds, spine flexion and shoulder position changes are common during treatment [16, 17]. However, its dosimetric impacts on proton therapy plans have yet to be revealed. For progressive changes, Wenyong *et al*. (2013) [18] reported that the tumour volume shrinkage of 20 nasopharynx cancer patients ranged from 20% to 60% during the treatment. The average reduction was 36.5%. OARs also lose cells under irradiation leading to complications such as dysphagia (swallowing difficulty) and

dysgeusia (taste changes), often accompanied by weight loss and the shrinkage of patient's outline [19]. Bhide *et al*. (2010) [20] showed that the parotid volume of 20 H&N patients decreased with a reduction rate between 21.3% and 42%, and an average of 2.3 mm medial shift occurred by the fourth week of treatment. With these anatomical changes during the course of radiotherapy, dose degradation is unavoidable. Kraan *et al*. (2013) [21] evaluated 10 oropharyngeal cancer patients and showed that anatomical changes led to an average 2% and 2.2% reduction in the D98% of high-risk CTV and low-risk CTV, respectively, and the maximum increase in brainstem dose can reach 9.2 Gy. Wu *et al*. (2017) [22] showed that, in 10 oropharyngeal patients, CTV mean doses were reduced by up to 7%, while an increase was shown in the right parotid with a range from 5% to 8%. Heukelom *et al*. (2019) [23] measured the influence of dose discrepancy in terms of normal tissue complication probability (NTCP) and concluded that anatomical changes can lead to $>5\%$ increase in NTCP for dysphagia and other toxicities.

1.2.4 Uncertainty in deformable image registration

Image-guided radiotherapy (IGRT), which incorporates imaging techniques during each treatment session, is often used in H&N cancer radiotherapy to identify the target position and track anatomical changes. The generated images can be used to analyse the target volume changes and the delivered dose. However, the images during the treatment are acquired at different timeframes, on different machines, and even in different modalities. Although they are aligned on a reference coordinate, the pixel or voxel on the same coordinate may not necessarily represent the same anatomical structure. Deformable image registration (DIR) is used to find the spatial correspondence between two images. During registration, the algorithm aims to find an optimal transformation to maximise the similarity between the two images. DIR has been widely explored for clinical applications. Veiga *et al*. (2015) [24] demonstrated the feasibility of using Niftyreg [25] as the DIR tool to calculate the dose distribution on cone-beam CTs (CBCT) for H&N patients. In their study, the planning CT was deformed to cone-beam CTs (reference image) to correct the HU for dose calculation. The transformation of DIR can also be used to warp the

dose distribution of a weekly CT to the planning CT. In an ideal DIR, the dose distribution on a weekly CT and the warped dose distribution on the planning CT should be the same. Heukelom *et al*. (2019) [23] and McCulloch *et al*. (2018) [26] used the transformations between weekly CTs and the planning CT (reference image) to warp the weekly doses in the reference frame of the planning CT and accumulate the dose. The differences between the planned and the accumulated doses were used to explore the impact of uncertainties during the treatment, measured by NTCP. Also, Tsiamas *et al*. (2018) [27] and Kranen *et al*. (2013) [28] exploited the transformation of DIR to model geometric changes mathematically.

Despite its great use in the clinic, deformable image registration (DIR) also introduces uncertainty. Nenoff *et al*. (2020) [29] reported that the average DIR uncertainty, evaluated using six different DIR algorithms, was 7.9% in PTV V_{95} (The per cent volume that received at least 95% of the prescription dose), and the variation between DIR methods on the accumulated dose on PTV V95 was 8.7%. Therefore, a DIR algorithm needs to be carefully assessed before use, especially where massive changes happen [30, 31].

1.2.5 Uncertainty evaluation methods

In the aforementioned studies, anatomical changes (changes in tumour and/or normal organ structures) and dose discrepancies were used in the uncertainty evaluation. Water equivalent path length (WEPL) can also evaluate uncertainty. It has been proposed as a surrogate of dose distribution in the literature [32, 33] to avoid the time-consuming and computational-expensive dose calculation. Kim *et al*. (2017) [32] quantified the anatomical uncertainty by measuring the WEPL changes on the distal edge of tumour volume using only one beam direction. Holloway *et al*. (2017) [33] evaluated the uncertainty by measuring the WEPL changes in the CTV with different beam angles for robust optimisation. Figure 1.1 simulates the limitations of these methods. Figure 1.1 a) shows a beam path in which the HU changes but retains the same WEPL on the proximal and the distal edge of the CTV. However, the accumulated dose on the proton path is different, as shown in figure 1.1 b). Scenarios like this might happen in the nodal area, where the HUs are het-

erogeneous. When the neck is tilted or rotated, the HUs within the target area may exchange their position. Figure 1.1 c) shows another scenario with increased HU values beyond the CTV distal edge, resulting in the same WEPL in the CTV but a different accumulated dose. This potentially leads to an overdose in normal tissue, as shown in figure 1.1 d). If only the WEPLs in the CTV were used to evaluate the uncertainty, errors like the one described here would not be noticed. Additionally, they did not factor in that spot weights can also affect uncertainty evaluation [34]. In Chapter 2, these problems were addressed by using the spot location from the proton plan and adding the spot weights in uncertainty evaluation.

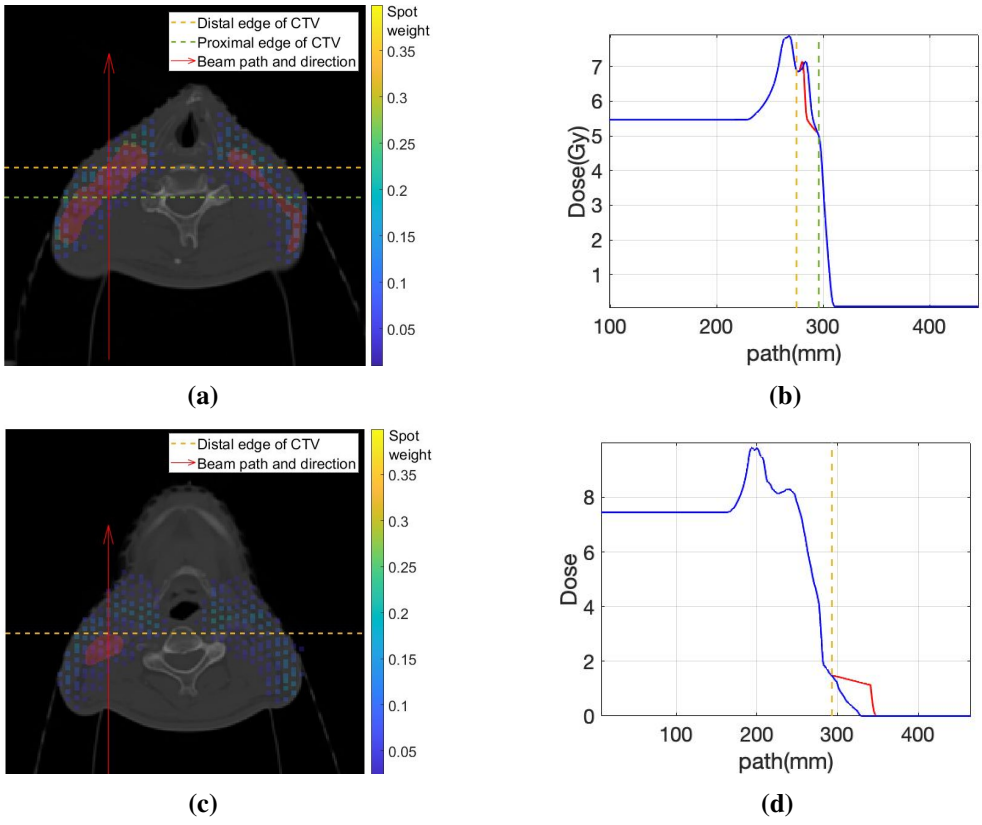


Figure 1.1: The limitations of using the WEPL of specific points or areas to evaluate the consequences of anatomical deformations. The images shown are slices from the planning CT, and the little spots are the proton spots delivered by the treatment system, with weights indicated by the colour code. The dose is calculated only considering the indicated red line of the spots. (a) and (b): The HU along the beam path first increased, then decreased, resulting in the same WEPL but different dose distributions. (c) and (d): The HU beyond the distal edges of nodal CTV are increased by 3.5%, consequently changing the dose distribution in the fall-off region of the distal dose.

1.3 Uncertainty mitigation methods

Referring to the aforementioned uncertainty sources, this section focuses on uncertainty mitigation methods. Before treatment, physicists make a mask using a plastic material that becomes soft and pliable when heated in warm water to help H&N patients reduce the movement above the shoulder during treatment. Markers are attached to the mask to help identify the treatment isocenter. For the residual uncertainties, the term "robustness" is used to measure the insensitivity of a plan to the defined uncertainty. The robustness of a treatment plan refers to two properties: first, the CTV should receive the prescribed dose despite the errors that may occur; and second, normal tissue constraints should be satisfied despite the potential errors in treatment planning or delivery [35].

1.3.1 Margin expansion

The conventional way to improve the robustness of a radiotherapy plan is to use a safety margin that expands the CTV to form the planning target volume (PTV). A popular margin recipe used in photon therapy was introduced by van Herk *et al.* (2000) [36]:

$$\text{Margin} = 2.5\sum + 1.64(\sigma - \sigma_p). \tag{1.1}$$

Where $\sum$ is the standard deviation of the systematic uncertainty, σ is the total standard deviation of the random errors combined with the beam penumbra width, and σ_p is the standard deviation describing the penumbra. This recipe allows a minimum of 95% of the prescription dose to be delivered to the CTV for 90% of patients. However, this solution is sub-optimal for proton therapy: 1) The conformal expansion leads to a sub-optimal trade-off between the target coverage and the OARs sparing. The margin is either too large, which gives extra burdens to OARs, or not sufficient to cover the tumour. 2) Uniform dose distribution in PTV is not achievable in reality. 3) The PTV is based on the so-called static dose cloud approximation, which assumes that the dose distribution would not be affected by geometric changes in patients as long as the CTV is within the PTV. However, this

assumption is violated in IMPT [35, 37]. Lowe *et al.* (2020) [38] illustrated that the PTV cannot cover the dose distortion, and as much as 8.4% dose degradation was found in the CTV of H&N patients. 4) In 3D IMPT, the spots are required to cover the target and weights are adjusted to get the optimal distribution. Different spot weights can lead to the same distribution, but different robustness [34]. The PTV concept did not exploit the advantage of this technique, and an unnecessary compromise was made by using the PTV [35].

1.3.2 Robust optimisation

Instead of taking uncertainties into account to create a safety margin, new approaches have been suggested to explicitly include the effect of geometrical uncertainties directly in the treatment plan optimisation [39, 40, 41] by modifying the cost functions. Known as probabilistic treatment planning or robust optimisation, these strategies are based on a stochastic approach that uses probability density functions, generally normal distribution [41] or uniform distribution [42], to describe the uncertainties. The optimisation that includes setup and range uncertainty [38] is referred to as conventional robust optimisation, in which setup uncertainty is generally modelled by a few millimetres of rigid shift, and the range uncertainty from HU is modelled by uniformly changing HU value, typically of the order of a few percentages. However, except for the setup and range uncertainty, patients also experience anatomical changes during treatment. These anatomical changes cannot be simplified as rigid translations, and thus the conventional robust optimisation cannot cover the dose degradation from anatomical changes. In very recent studies, multiple CT images were involved in the optimiser to account for anatomical uncertainty, referred to as anatomical robust optimisation. Mesías *et al.* (2019) [43] used the CT images of the first two weeks to account for anatomical random changes, e.g. shoulder movement or neck tilt. Yang *et al.* (2020) [44] used the image from the first adaptation to include progressive anatomical changes. They both concluded that anatomical robust optimisation improved plan robustness toward anatomical changes and reduced the number of plan adaptations for H&N patients. However, the required multiple scanning will give patients an extra dose and

increase the workload for clinicians. Using acquired CT images during treatment will also limit the creation of robust plans at the early planning stage. Aiming to address these limitations, an anatomical model was built in Chapter 4 to produce the potential deformations with corresponding probability before the treatment. This model provided a potential solution for anatomical robust optimisation.

1.3.3 Robust evaluation

After robust optimisation, robust evaluation is still needed. Set up and range uncertainty are included in the conventional robust evaluation, while the robust evaluation, including inter-fractional anatomical variations, often uses images acquired during the course of treatment [33, 45, 46], and as such, it can only inform the planning process for a portion of the treatment delivery. Holloway *et al*. (2020) [33] added the anatomical variations into robust evaluation based on CBCTs at each fraction. The impact of different beam angles on the robustness of a plan was presented in this paper. Based on the plan evaluation, McGowan *et al*. (2015) [47] proposed to create a robustness plan database to assist in finding a more robust planning approach. However, patients have their own unique geometries, on which the impact of uncertainty can vary. A pre-treatment robust evaluation, including setup uncertainty, range uncertainty and anatomical uncertainty based on individual geometry, would be more helpful in informing clinical decisions. In this thesis, the small non-rigid variations, which are unavoidable during treatment, are included in the patient-specific robust evaluation in Chapter 3.

1.3.4 Adaptive radiotherapy

Compared with setup and range uncertainty, anatomical uncertainty is more complex and can lead to more severe dose discrepancies. Thus, current research focuses on strategies to mitigate the influence of anatomical changes [43, 48, 49, 50, 51]. Adaptive radiotherapy is proposed to this end. Two strategies for delivering adaptive proton therapy are available: offline and online adaptation. Online adaptation is a state-of-art technique intended for same-day application. However, limited by the current computational speed, online adaptation either compromises the accuracy or

constrains the optimiser. As a result, the online adapted plan is always inferior to a new treatment plan. Matter *et al.* (2019) [49] used the analytical pencil beam scanning algorithm to generate a plan in ten seconds. However, the analytical calculation is less accurate than a MC-based algorithm. Studies using constrained MC [50, 51] reported that the calculation time of an adapted plan online ranged from 3 to 22 minutes [50, 51, 48]. Bobić *et al.* (2021) [51] constrained the optimizer by only adjusting the beamlet positions, energies and beamlet weights to produce adapted plans. They reported a median adjustment time of 12 minutes, excluding the time for DIR. Lalonde *et al.* (2021) [48] only adjusted the weight of the beamlets to produce adapted plans. Their median adjust time was also 12 minutes but included the time for DIR. While online adaptation is considered superior to offline adaptation because of no treatment delay, no clinical solution exists for pre-delivery quality assurance (QA), and the reported time required to calculate the online adapted plan is currently limiting its application in terms of total treatment time per patient [52].

Offline adaptation is often triggered by clinically meaningful criteria based on centre-specific protocols. Tumour coverage is generally one of the criteria. Mesías *et al.* (2019) [43] triggered a plan adaptation if 98% volume of CTV cannot be covered by 95% of the prescription dose. In the TORPEdO trial (A phase III trial of proton therapy versus intensity-modulated radiotherapy for multi-toxicity reduction in oropharyngeal cancer; CRUK/18/010) conducted in the UK, they will replan if 99% volume of CTV cannot be covered by 90% of the prescription dose and 95% of the CTV volume cannot be covered by 95% of the prescription dose [53]. Besides tumour control, radiation-induced toxicities are also a great concern in the clinic. Brouwer *et al.* (2016) [54] suggested that an extra 3 Gy mean dose on the parotids can be of clinical relevance to severe toxicity, such as xerostomia, which can have a long-term impact on patients' quality of life after treatment [55]. These dose-relevant criteria require routine dose verification based on the daily or weekly images, which is time-consuming in a busy clinical workflow. In literature, different replan surrogates also exist, including weight loss [56], body contour changes [57], and HU changes in parotid glands [58]. However, replanning based on dosimetric

changes is still the gold standard in the clinic.

Even though offline adaptation is generally used in proton therapy facilities, the process challenges clinical workflow efficiency. Whilst plans are adapted, patients must either continue treatment with an existing sub-optimal plan or face interruptions to treatment. The latter may be particularly undesirable for rapidly growing tumours such as squamous carcinomas of the H&N [59]. Furthermore, this reactive approach to plan adaptation can create an unpredictable workload for treatment planning staff, the medical physics team who perform patient-specific plan QA, and radiation oncologists who review and approve the plans.

Adaptive plans that can be prepared in advance would be beneficial to the clinical workflow: 1) A replan can be delivered as soon as needed due to the ability to perform patient-specific QA/verification before adaptation is required, for example, on a predicted CT, which triggered a replan. 2) For patients, there is no gap in treatment or delivery of a few sub-optimal fractions while a replan is calculated, approved, and verified through QA. 3) For workflow, it allows for ease of scheduling patient-specific QA along with machine QA. To this end, we need the facilitation of predictive anatomical models, e.g. using the predicted images of an anatomical model to create an adaptive plan in advance, as presented in Chapter 6.

1.4 Anatomical models and their application in proton therapy

Anatomical models have been proposed in the literature as potential solutions to uncertainty mitigation. Generally, there are two types of anatomical models: population model and individual model.

1.4.1 Population model

A population model can explore the pattern of anatomical changes based on population data. Panagiotis *et al.* (2018) [27] fed all the data acquired during the treatment into the principal component analysis (PCA) to model the anatomical changes of H&N patients. However, they ignored that anatomy can change progressively over

time, and no predictive ability was shown in this model. Yu *et al.* (2016) [60] used an intra-patient model to generate artificial deformations for DIR evaluation. They applied PCA on the deformation vector fields (DVF) of each daily time point to capture the dominant modes of deformation, called principal components (PC). Each generated DVF can be represented by a linear combination of PCs. They assumed that the probability density function (PDF) of the coefficients α of each PC follows a standard normal distribution. However, they did not validate if the PDF of α from their dataset was consistent with this assumption, and the probability of each predicted DVF was not revealed. To date, it is still challenging to quantify the probability of a certain type of anatomical deformation to arise during the treatment course.

1.4.2 Individual model

An individual model is built based on individual patient data. One particular application of the individual model is to create a patient-specific PTV. Thörnqvist *et al.* (2013) [61] applied PCA to the motion of prostate CTVs for each patient. 13 patients, each with $9-10$ CTs, were included in this study. The created union of simulated shapes covered 95% of CTV changes when 4 PCs were used in this patient-specific model. Xu *et al.* (2014) [62] also proposed two margins for prostate cancer treatment based on a PCA model. One was the optimized PTV margin that was iterative updated until the predefined coverage criteria were satisfied, while the other was a dosimetric margin between the CTV and the treated volume. Both methods showed an advantage in target coverage compared to conformal margin expansion. However, their expanded margin cannot be obtained before treatment because patient-specific images acquired during treatment were all included in their model.

Another application is to use images acquired during the first F fractions and predict the anatomical changes of the following fractions [28, 63]. Kranen *et al.* (2013) [28] explored using the systematic deformation of the first F fractions for adaptive intervention. They deformed the planning CT using: 1) the average deformation from the previous 10 fractions as the single intervention strategy; and 2) the

average deformation from the previous week as the weekly intervention strategy. However, the progressive changes between fractions were not considered in their study. Chetvertkov *et al*. (2016) [54] modelled the patient-specific anatomical variations using regularized PCA. They assumed a positive linear correlation between the coefficients of PC and fractions due to the fact that anatomical changes are progressive. However, they required at least half of the total fraction for a reasonable estimation. That is the main disadvantage of this individual model, as patients must have already begun to deform through treatment for the model to develop. Ideally, an anatomy predictive model that can provide accurate predictions before treatment would benefit clinical practice the most. However, images acquired during treatment are only a snapshot of the anatomy. Interfraction variations and the influence of acute toxicities on patient eating during treatment also determine the weekly anatomical changes. An alternative is to develop a model that captures the systematic anatomical changes based on population data. Each patient's model could then be refined as patient-specific data is acquired over the course of treatment. The above two models were only assessed based on the misalignment of anatomical landmarks, and no adaptive plan was created on the predicted anatomy for further evaluation. Therefore, the application of anatomical models for adaptive radiotherapy still needs to be explored.

1.5 Research aims and objectives

Although setup uncertainty and range uncertainty have been considered in robust evaluation and optimisation, anatomical changes in proton therapy can still lead to severe dosimetric discrepancies. To fully assess the anatomical uncertainty for H&N cancer, DIR is required. Because different DIR algorithms come with different magnitudes of uncertainty, to which proton therapy is inherently sensitive, a DIR needs to be carefully evaluated for a specific application. This thesis starts with the DIR evaluation.

1. Evaluate the selected DIR algorithm for mathematical modelling. The objectives of this project are:

- Assess the DIR uncertainty using contours-related metrics.
- Develop a new uncertainty evaluation method based on weighted spot location deviation(WSLD) and apply it for DIR uncertainty evaluation.
- Develop a workflow to assess the accumulated DIR uncertainty based on dose distribution. Justify the use of the DIR algorithm by comparing the accumulated DIR uncertainty with the accumulated anatomical uncertainty.

The DIR evaluation validates the feasibility of using the selected algorithm to capture anatomical deformations. In this thesis, the anatomical changes are divided into 1) sNRVs (such as neck tilt and spine flexion) that are random and not patient-specific and 2) progressive changes that are dependent on individual features. The DIR is first used to capture possible sNRVs because the dosimetric impact of sNRVs has yet to be explored. sNRVs are included in the robust evaluation to guide beam selection based on individual geometry.

2. To explore the role of sNRVs in robust evaluation. The objectives of this project are:
   - Build a distribution of possible sNRVs based on population data.
   - Assess the dosimetric discrepancies caused by sNRVs.
   - Incorporate sNRVs in the robust evaluation to guide the beam arrangement selection.
   - Validate quantitatively that the sNRVs-based evaluation is better than the conventional evaluation method.

When a large patient dataset is available, limiting the number of included sNRVs to the most common ones helps to improve computational efficiency. For that purpose, the DIR is used to build an anatomical model based on population data. The model produces potential anatomical deformations considering the time dependence of the progressive changes during treatment and estimates the probability of each type of deformation, referred to as the probability model (PM). The

dosimetric influences of sNRVs from the real deformations are used as the benchmark to validate the PM. The PM can be used for anatomical robust optimisation.

3. To assist anatomical robust optimisation, the objectives of this project are:

    - Develop a probability model (PM) based on PCA to model major deformations at each weekly time point in patients.
    - Quantify the probability of each type of anatomical deformation based on population data.
    - Validate the feasibility of using the PM to measure anatomical uncertainty.

Anatomical robust optimisation can reduce the need for adaptive planning during treatment. However, the robust optimisation alone may not be adequate to account for anatomical changes during treatment [34]. An individual model was developed for prospective offline replanning.

4. To predict the anatomical changes during the treatment for an individual, the objectives of this project are:

    - Develop an average model (AM) based on population data to predict the systematic progressive changes of each week before treatment.
    - Refine the prediction by adding patient-specific progressive information from the data acquired during the course of treatment as the refined individual model (RIM).
    - Evaluate the models using the average absolute HU differences(AAHUD), contours, WSLD and IMPT dose distributions.

The evaluation of the RIM model demonstrated its ability to predict individual anatomical changes. Thus, it was investigated for clinical use. The RIM model was explored to assist with replanning in advance.

5. To maximise the use of the individual model for offline replanning. The objectives of this project are:

- Find the best strategy to use the RIM for offline replanning; three different strategies are compared as follows:

  1) Scheduled predicted plan on the scheduled week: replans are optimised on predicted images of weeks 3 and 5 and applied to weeks 3/4 and 5/6, respectively.

  2) Reactive predicted plan on the flexible week: a plan is optimised on a predicted CT, which triggers a replan, and applied flexibly as soon as the verification CT indicates that a plan adaptation is necessary.

  3) Reactive predicted plan on the scheduled week: a plan is optimised on a predicted CT, which triggers a replan, and applied from scheduled week 3.

- Compare the reactive predicted replan with the standard replan based on the accumulated dose distribution.

1.6 Impact and novelty of the work

The research presented in this thesis focuses on mitigating anatomical uncertainty. The following aspects are novel:

- Use WSLD to evaluate uncertainty: Firstly, compared to the previously published WEPL-based uncertainty evaluation methods, the influence of spot weight on the robustness of a plan was considered. Additionally, the radiation target is generally divided into sub-targets in proton planning to exploit multi-field optimisation. This can be reflected in the spot location of each beam field. Using WSLD improves evaluation accuracy.

- Evaluate the dosimetric impact of sNRVs and validate the necessity of including sNRVs in robust optimisation/evaluation. In recent literature, the dosimetric impact of anatomical changes was reported. However, they did not separately report the dosimetric impact of sNRVs and progressive changes. The underlying causes of these two types of change are different. Hence their corresponding mitigating strategies should be different. Progressive changes

have trends to follow. Thus, using adaptive radiotherapy can potentially bring more benefits to patients. Whereas sNRVs are unpredictable and not patient-specific but can lead to unacceptable discrepancies. sNRVs are more suitable to be accounted into robust optimisation as setup uncertainty and range uncertainty. The method of including sNRVs in the robust evaluation was proposed to help select the best beam arrangement. This application can potentially spare OARs and reduce the replanning rate.

- Build two anatomical models, the PM and the RIM, considering the time dependence of the progressive changes in H&N patients. The PM and the RIM aim to assist anatomical robust optimisation and predictive replanning, respectively.
  - The population-based probability model for H&N patients is proposed for anatomical robust optimisation. The PM generates major deformations at each weekly time point during treatment. Each potential deformation can be quantified by probability. The PM has the best performance in terms of estimating anatomical uncertainty.
  - Aiming at the limitations of the previous individual models, which either required at least half fractions to provide a reasonable estimation or ignored the progressive changes between fractions, the RIM considering the time dependence of anatomical changes is proposed. In this model, the systematic progressive changes of each week are applied to a new patient before treatment, and the prediction is refined as treatment goes. This model greatly reduces the requirement of imaging frequency and can capture progressive changes in time.
- Develop the first application of the RIM to create replan in advance. In the literature, the proposed models were only assessed based on the misalignment of anatomical landmarks, and no potential application was evaluated based on dose distribution. This thesis provides potential predictive replanning strategies based on the refined individual model. The results show that

the predicted plan adaptation technique achieves similar coverage of CTVs and reduces the parotid dose compared to the standard replanning strategy. Prediction-based replanning enables adaptive therapy to be delivered without treatment gaps or sub-optimal fractions, as can occur during a standard re-planning strategy. The operational logistics of a busy clinical practice may also benefit from this improved workflow efficiency.

Chapter 2

Evaluation of Deformable Image Registration

In this chapter, diffeomorphic deformable image registration implemented in the NifTK was evaluated for building anatomical models.

The work in this chapter was incorporated in the following outputs:

Ying Zhang, Megan Z.Wilson,Jeffrey Liu, Jailan Alshaikhi, Gary Royle, Stacey M.Holloway. ”Can Proton Water Equivalent Path Length calculations be used instead of full dose recalculation for determining when to adapt a plan?” PPRIG Proton Physics Workshop 5, National Physical Laboratory(NPL),UK. Feb 2019.

Ying Zhang, Stacey McGowan Holloway, Megan Zoë Wilson, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. ”DIR-based models to predict weekly anatomical changes in head and neck cancer proton therapy”. Physics in Medicine & Biology 67, no. 9 (2022): 095001.

Ying Zhang, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. ”A probability model for anatomical robust optimisation in head and neck cancer proton therapy.” Physics in Medicine & Biology (2022).

Esther Bär, Charles-Antoine Collins-Fekete, Vasilis Rompokos, Ying Zhang, Mark N. Gaze, Alison Warry, Andrew Poynter, Gary Royle. CT calibration for precise proton therapy planning in children. European Society for Radiotherapy and Oncology 2021.

Esther Bär, Charles-Antoine Collins-Fekete, Vasilis Rompokos, Ying Zhang, Mark N. Gaze, Alison Warry, Andrew Poynter, and Gary Royle. "Assessment of the impact of CT calibration procedures for proton therapy planning on paediatric treatments." Medical physics (2021).

Contribution of Authors: Jailan Alshikhi created the proton plans. Wenyong Tan provided the clinical data and valuable input. Gary Royle provided valuable input concerning the clinical importance and impact of this work. Esther Bär and Stacey McGowan Holloway supervised the project and guided the development of ideas, methods, results, and conclusions of this work. All other work presented in this chapter was done by myself. Esther Bär used the methods proposed in the chapter to evaluate the range uncertainty in CT calibration.

2.1 Introduction

In image-guided radiotherapy (IGRT), although images acquired during treatment are aligned on a reference coordinate, the pixel or voxel on the same coordinate may not necessarily represent the same anatomical structure. Image registration is the process of aligning different images into the same coordinate system with matched imaging contents. It is composed of three main components: 1) a transformation metric, 2) a similarity metric (cost function) 3) optimiser. The optimiser drives the algorithm to find the best result of a similarity metric, such as the maximisation of mutual information between images, through an iterative procedure to obtain the optimal transformation. In deformable image registration (DIR), the transformation is a spatially variant vector field. Therefore, it allows a non-uniform mapping between images. DIR has been proposed to correct HU for CBCT, warp and accumulate dose to the reference planning CT, and model geometric changes during treatment. However, no transformation can achieve a 100% match. DIR uncertainty exists. Nenoff *et al.* (2019) [29] compared six different DIR algorithms and showed that the variation of DIR uncertainty on the accumulated dose of PTV V95 between the different algorithms was high (8.7%). The additional uncertainty caused by a DIR algorithm can be up to 26.3% in PTV V_{95}. Therefore, a DIR method needs

to be carefully evaluated before applying it to a specific application, especially a proton application, as the inherent physical characteristics of protons make the dose distribution of a proton plan more sensitive to DIR uncertainty than that of a photon plan.

Generally, DIR uncertainty was evaluated based on contour and dose distribution. For DIR evaluation based on contour, the contours of the floating image are propagated to the reference frame using the transformation between two images. The differences between the propagated and the reference contours measure DIR uncertainty [64]. For DIR evaluation based on dose distribution, the dose calculated on a weekly CT is deformed to the planning CT using the transformation [29]. The dose-volume histogram (DVH) differences between the dose distribution on the weekly CT and the warped dose distribution on the planning CT represent DIR uncertainty. However, the above two methods are based on the assumption that the contours on weekly CT are absolutely consistent with the contours on the planning CT, which is not true in reality, as mentioned in Chapter 1.

The DIR uncertainty has yet to be thoroughly evaluated for building anatomical models for H&N proton therapy. Aiming at this limitation, this chapter evaluates the uncertainty of the DIR algorithm provided by Niftyreg [1]. The evaluation is conducted based on: 1) contours; 2) proton spot location. To improve the accuracy of uncertainty evaluation based on proton range, a new uncertainty evaluation method based on the spot range and spot weight is developed and applied to DIR uncertainty evaluation; 3) dose distribution. A workflow to assess the accumulated DIR uncertainty based on the dose distribution is proposed and compared with the accumulated anatomical uncertainty to justify the use of this DIR algorithm. This workflow eliminates the delineation error.

[1] https://cmiclab.cs.ucl.ac.uk/mmodat/niftyreg

2.2 Methods and materials

2.2.1 Deformable image registration tool

NiftyReg, which is an open-source DIR tool developed by CMIC at UCL (http://cmic.cs.ucl.ac.uk/home/software), has been used in the CT-CBCT registration for proton therapy [24, 30]. The diffeomorphic image registration of NiftyReg [65] was selected for this study because its deformation field is invertible, differentiable, and the inverse is also differentiable. As a result, its transformation leads to one-to-one voxel mappings, which inherently preserve the underlying topology. Besides, the diffeomorphic setting enforces consistency under compositions of the deformations. If the deformations are diffeomorphic, then the result of the composition will also be diffeomorphic.

The diffeomorphic image registration is a B-spline-based method implemented in NiftyReg. Spline-based transformations assume that a set of corresponding control points can be identified in the source and target images. The control points are adjusted until the alignment between the warped image and the reference image is maximised. The transformation is computed through a cubic B-spline interpolation from the lattice of control points overlaid on the reference image. The diffeomorphism is parameterized by a stationary velocity field (SVF) using the exponential map [66]. DVF $\boldsymbol{\phi}$ can be expressed as equation 2.1

$$\boldsymbol{\phi} = \exp(\boldsymbol{v}). \tag{2.1}$$

$\boldsymbol{v}$ represents the SVF in the diffeomorphic image registration [65].

While in many registration approaches, the calculation of the inverse transformation is computationally expensive, and invertibility cannot be guaranteed, the inverse transformation $\boldsymbol{\phi}^{-1}$ can be easily calculated in diffeomorphic image registration using equation 2.2

$$\boldsymbol{\phi} = \exp(\boldsymbol{v}) \quad \Rightarrow \quad \boldsymbol{\phi}^{-1}(x) = \exp(-\boldsymbol{v}). \tag{2.2}$$

It should be noted that the rigid registration was performed based on a bony

match before the DIR. The rigid registration area was masked by the body contour that was large enough to cover the whole PTV.

2.2.2 Data

Five H&N patients with a planning CT (pCT) and a weekly repeat CT (rCT$_t$), where t (t=0,1,2,3,...) represents the week of CT scanning, were recruited. They experienced weight loss of 17.21%, 4.84%, 6.98%, 4.17% and 17.8% during the treatment, respectively. Contours in the pCT and rCT$_t$ were manually delineated by an oncologist. For each patient, rCT$_t$ was deformed to pCT to produce $\boldsymbol{v}_{p\to t}$, where p stands for pCT and t stands for the week when the rCT was acquired (rCT was rigidly registered to the pCT before the DIR). dCT is the deformed rCT using $\boldsymbol{v}_{p\to t}$. rCT, dCT and pCT share the same isocenter.

Proton plans were created for the five patients using three beam angles (60°, 300°, 180°) in Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA), complying with the original radiotherapy protocol. The dosimetric goals for all plans in this study are summarised in table 2.1.

Table 2.1: Dosimetric goals of the treatment plans created in this study.

Structure	Metric	Dosimetric Goals
High-risk CTV	D_{95} (The minimum dose to 95% of target volume)	$>$ 95% of prescription dose(72.6 Gy)
Low-risk CTV	D_{95}	$>$ 95% of prescription dose(63 Gy)
High-risk CTV	D_2 (The minimum dose to the hottest 2% volume)	$<$ 107% of prescription dose
Spinal cord	D_{max} (The max dose in the volume)	$<$45 Gy
Brainstem	D_{max}	$<$55 Gy
Chiasm	D_{max}	$<$55 Gy
Optical Nerve	D_{max}	$<$55 Gy
Parotid glands	D_{mean} (The mean dose in the volume)	$<$26 Gy
Oral cavity	D_{mean}	$<$40 Gy
Larynx	D_{mean}	$<$40 Gy
Cochlear	D_{mean}	$<$45 Gy

2.2.3 Deformable image registration evaluation methods

The influence of DIR uncertainty was investigated from 1) the weekly changing trend and 2) the accumulated influence during the treatment.

As the magnitude of progressive changes in H&N patients increases during treatment, leading to difficulty in finding the correspondence between two images, the hypothesis is that the overall DIR uncertainty will increase along with treatment.

The weekly changing trend of DIR uncertainty was investigated based on contour and proton spot range.

2.2.3.1 Deformable image registration evaluation based on contour

The evaluation based on contour was used to observe the weekly changing trends. The selected contours on rCT_t were propagated using the deformation vector field $\boldsymbol{v}_{p\to t}$. The propagated contours were compared with the corresponding contours on the pCT (gold standard) to evaluate the DIR uncertainty at each weekly time-point, referred to as C-DIR evaluation. In the C-DIR evaluation, the mean surface distance (MSD) was chosen to quantify the contour differences because it is a metric generally used in the DIR evaluation, and the quantification is given in millimetres, which can be compared with voxel size to justify the acceptance of DIR error. MSD indicates the average difference between the gold standard surface S and the propagated surface S'.

$$MSD = \frac{1}{n_S + n'_S} \left(\sum_{a=1}^{n_S} DT(a, S') + \sum_{b=1}^{n'_S} DT(b, S) \right), \tag{2.3}$$

where distance transformation (DT) calculates the minimum distance between a point on surface S and surface S' using

$$DT(a, S') = min(\| \mathbf{a} - \mathbf{b} \|) \;\; \mathbf{a} \in S, \;\; \mathbf{b} \forall S'. \tag{2.4}$$

The contours included in the C-DIR evaluation were low-risk CTV, high-risk CTV and parotid glands. These structures commonly change their shape and volume during treatment.

2.2.3.2 Deformable image registration evaluation based on proton spot range

In this section, an estimation of the spot location within a patient, derived from the treatment plan file and CT image information, was developed to evaluate uncertainty. DIR uncertainty evaluation is one of the applications, referred to as R-DIR evaluation.

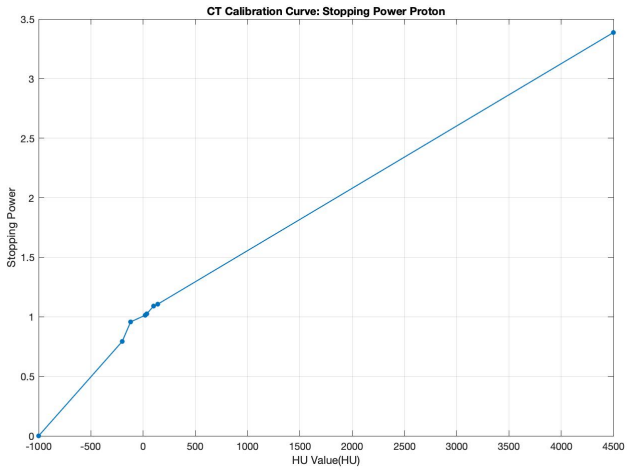


Figure 2.1: The calibration curve between HU and RSP.

In the pencil beam scanning technique, doses are delivered spot by spot. The spot positions (X,Y) and energy/layer (Z) can be extracted from the plan files of the Varian treatment planning system. (X,Y) are recorded relative to the isocenter (the centre of the target) in the gantry coordinate system. (X,Y) with a beam angle can specify a beam path. The beam energy (Z) determines the depth of a spot on the beam path by calculating the WEPL using equation 2.5.

$$WEPL = \sum_{i,j,k \in S} RSP_{i,j,k} \cdot d_{i,j,k}, \tag{2.5}$$

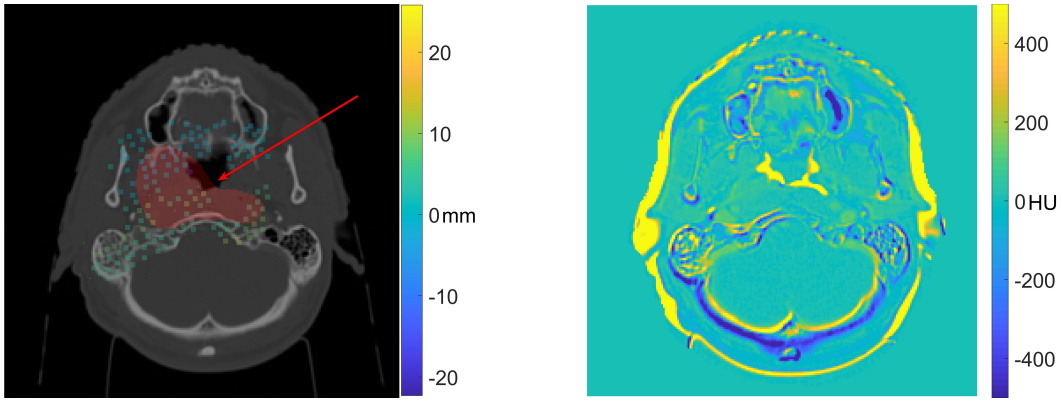
where S is a set of voxels which contain the beam path. $\mathrm{RSP}_{i,j,k}$ is the voxel-wise relative stopping power estimated from the CT number using a clinical calibration curve, shown in figure 2.1. $d_{i,j,k}$ is the path length of the beam inside voxels (i,j,k), estimated by a ray tracing algorithm [67]. The beam-lines were assumed parallel in this study.

The deviation of each spot (SLD) on the beam path is calculated using

$$\sigma(\boldsymbol{r}) = |\boldsymbol{r}_{\text{uncertainty}} - \boldsymbol{r}_{\text{reference}}|, \tag{2.6}$$

where $\boldsymbol{r}$ is a spot position in the CT. $\boldsymbol{r}_{\text{reference}}$ is a spot location in the reference frame. $\boldsymbol{r}_{\text{change}}$ is the corresponding spot location under uncertainty.

To visually analyse the uncertainty, the spot error map was developed to intuitively present the deviation (in the beam direction) of each spot relative to its original position in the pCT. Figure 4 shows an example of the spot error map of the beam angle 60°. The deviations are due to the anatomical changes during the treatment. The colour bar on the right side represents the error calculated using equation 2.6.



(a) **(b)**

Figure 2.2: A slice of the spot error map between the pCT and CT_6. (a) is the spot error map of the beam angle 60°. The red area is the radiation target of beam angle 60°. Positive values mean spots go deeper along the beam path, and negative values mean spots stop at shallower places. (b) is the image difference between the pCT and CT_6, as the reference for the spot error map.

Because spot position and weight both affect dose distribution [34], spot weights were added to the uncertainty evaluation to improve accuracy. The weighted SLD (WSLD) is presented in equation 2.7. Without having to calculate the dose distribution, the WSLD is more effective in describing the consequences of uncertainty.

$$\mathrm{WSLD} = \sum_{\boldsymbol{r}} \sigma(\boldsymbol{r}) \cdot w_r, \ \sum w_r = 1, \tag{2.7}$$

where w_r is the normalized spot weight. $\sigma(\boldsymbol{r})$ is the deviation of the corresponding spot on the beam path as defined in equation 2.6.

To evaluate the uncertainty of the DIR algorithm, the WSLD between dCT_t and their corresponding pCT was calculated, using the pCT as the reference in equation

2.7. dCT_t is the deformed rCT in the reference frame of the pCT and should have the exact spot locations as the pCT in an ideal DIR algorithm. The evaluation based on WSLD was used to observe the weekly changing trends.

2.2.3.3 Deformable image registration evaluation based on accumulated dose distribution

The accumulated influence of DIR uncertainty during treatment was investigated based on dose distribution.

This study proposed to calculate the dose distribution on dCT_t and add them up as the accumulated dose of dCTs (dCTs should have the same dose distribution as the pCT in an ideal DIR algorithm). The difference between the accumulated dose of dCTs and the corresponding planning dose measures the accumulated influence of DIR uncertainty during treatment, referred to as dCT-DIR evaluation. Because dCTs use the same contours as the pCT, the delineation error was removed in the dCT-DIR evaluation.

To justify the use of the DIR algorithm, the accumulated DIR uncertainty was intended to compare with the accumulated anatomical uncertainty based on dose distribution. For easy implementation, the difference between the accumulated dose distribution of weekly CTs (weekly propagated dose distribution in the reference frame of the pCT) and the corresponding planning dose distribution was calculated to represent the accumulated uncertainty from anatomical changes plus DIR (A+DIR). The uncertainty evaluation results of A+DIR versus DIR should lead us to the same conclusion as anatomical changes versus DIR. If the dose difference caused by the DIR uncertainty is much smaller than the A+DIR uncertainty, this error will be accepted for DIR-based modelling.

The workflow of accumulating weekly dose distributions and dCT's dose distributions is shown in figure 2.3.

The evaluation metrics used in the dCT-DIR evaluation were: 1) the difference between the accumulated dose of dCTs and the planning dose on the clinically-concerned dose metrics. 2) the gamma index between the planning dose and the accumulated dose of dCTs. 3) voxel-wise absolute dose difference between the

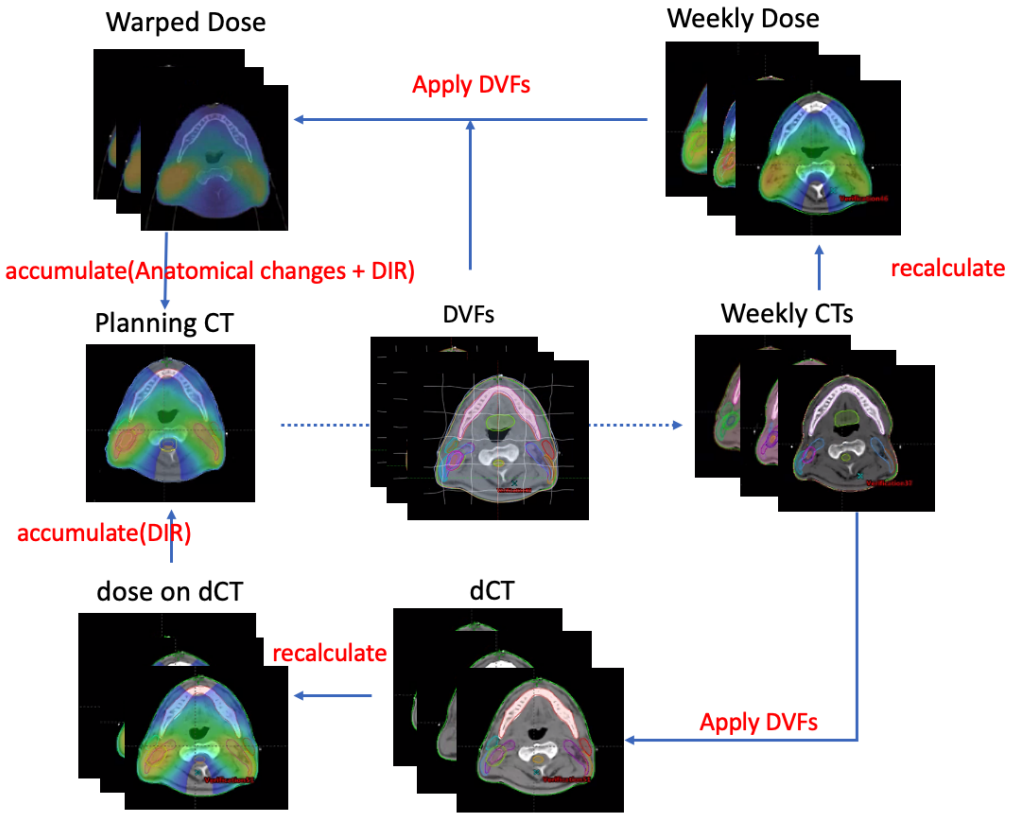


Figure 2.3: The workflow of accumulating weekly dose distributions and dCT's dose distributions.

planning dose and the accumulated dose of dCTs. Three evaluation metrics were used as follows:

1. Difference between the accumulated dose of dCTs and the planning dose was calculated for the clinically-concerned dose metrics, including D_{95} of CTVs, D_{mean} of parallel organs and D_{max} of serial organs.

2. Gamma-index (γ) was used to evaluate the dose distribution difference between the accumulated dose of dCTs ($\mathbf{D}_{cal}$) and the planning dose ($\mathbf{D}_{ref}$) using acceptance criteria [68]. It combines two important dose difference measurements: absolute dose difference and distance to agreement (DTA). DTA finds the distance between a dose point in the planning dose distribution and the nearest point with the same dose value in the accumulated dose of dCTs. Absolute dose difference is sensitive in high-dose gradient regions, while DTA

is sensitive in low-dose gradient regions. The gamma index complements the two measurements. The gamma index at one dose point can be calculated using

$$\gamma(\mathbf{r}_{\text{ref}}) = \min\{\Gamma(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}})\} \forall \{\mathbf{r}_{\text{cal}}\}, \tag{2.8}$$

where

$$\Gamma(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}}) = \sqrt{\frac{r^2(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}})}{{\Delta d_M}^2} + \frac{\delta^2(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}})}{{\Delta D_M}^2}}, \tag{2.9}$$

with

$$r(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}}) = |\mathbf{r}_{\text{cal}} - \mathbf{r}_{\text{ref}}|, \tag{2.10}$$

and

$$\delta(\mathbf{r}_{\text{cal}}, \mathbf{r}_{\text{ref}}) = \mathbf{D}_{\text{cal}}(\mathbf{r}_{\text{cal}}) - \mathbf{D}_{\text{ref}(\mathbf{r}_{\text{ref}})}. \tag{2.11}$$

In the above equations, $\mathbf{r}_{\text{ref}}$ is the reference point, $\mathbf{r}_{\text{cal}}$ is the calculated point on the accumulated dose of dCTs, Δd_M is the distance difference criterion and ΔD_M is the dose difference criterion. As the passing criteria, this study used $\gamma(\mathbf{r}_{\text{ref}}) < 1$ the calculation passes, otherwise fails. The criteria used were Δd_M=2 mm and ΔD_M=2%.

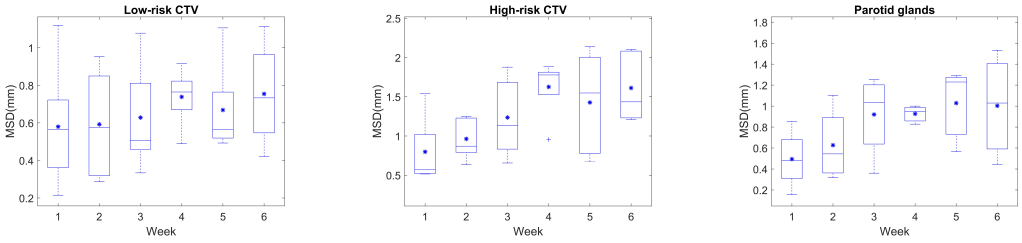
3. The voxel-wise absolute dose difference between the planning dose and the accumulated dose can be visually demonstrated in the dose-deviation-volume histogram (DDVH). The voxel-wise absolute dose difference caused by the DIR uncertainty and the A+DIR uncertainty was compared in DDVH. The organs included in this analysis were targets (low-risk CTV, high-risk CTV), serial OARs (spinal cord, brainstem, chiasm), and parallel OARs (parotid glands, oral cavity, larynx).

   To quantify the DDVH, the area below the DDVH curve was calculated, representing the mean value of the voxel-wise absolute dose difference (MADD).

2.3 Results

2.3.1 Deformable image registration evaluation based on contour

The weekly changes in MSD across the five patients are shown in figure 2.4. The DIR uncertainty shows an increasing trend during the treatment. The maximum MSD of all the structures involved is below 3 mm (slice thickness).



(a) **(b)** **(c)**

Figure 2.4: The weekly MSD between the deformed contours in dCTs and the corresponding contours in the pCT for high-risk CTV, low-risk CTV and parotid glands. In the box plot, the horizontal lines indicate the median value, and the asterisks indicate the mean value.

2.3.2 Deformable image registration evaluation based on proton spot range

The WSLD of DIR uncertainty across the test dataset is shown in figure 2.5. In individual cases, the minimum and the maximum WSLD of 0.44 mm and 2.17 mm were found ($<$ slice thickness of 3 mm), respectively. The average WSLD with 95% confidential interval (CI) across the five test patients increased from 0.86$\pm$0.14 mm (week 1) to 1.33$\pm$0.48 mm (week 6). The weekly average was 1.03$\pm$0.23 mm, close to the pixel size of 0.98 mm.

2.3.3 Deformable image registration evaluation based on accumulated dose distribution

The differences between the planning dose and the accumulated dose of dCT_t on clinically-concerned dose metrics are shown in figure 2.6. The figure shows that the influence of DIR uncertainty on the dose metric differences of these organs can be

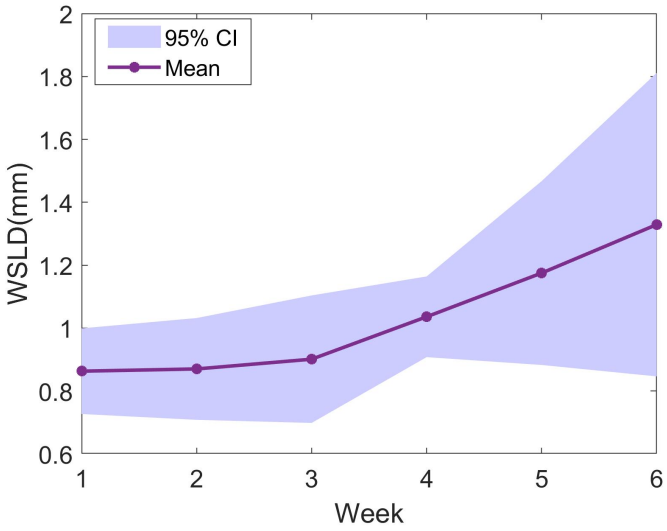


Figure 2.5: The weekly WSLD between dCTs and the corresponding pCT. The result is estimated in average WSLD with 95% CI over the 5 test cases.

controlled within 2 Gy.

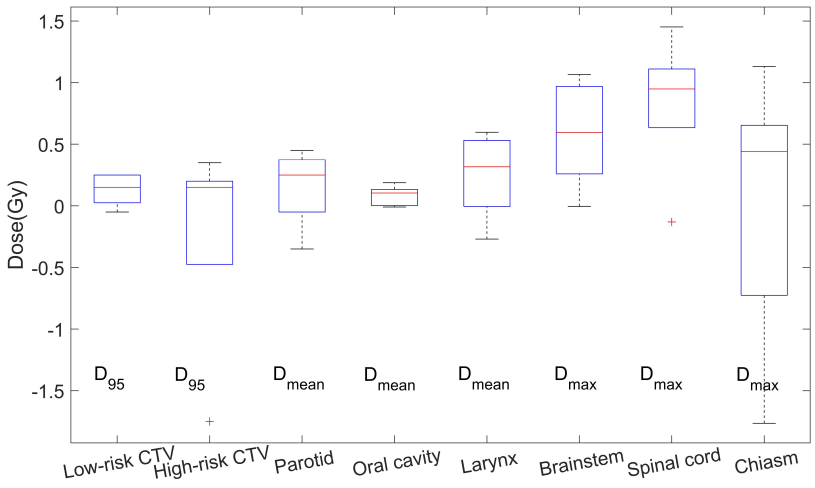


Figure 2.6: The differences between the planning dose and the accumulated dose of dCT_t on clinically-concerned dose metrics.

The gamma index was calculated between the planning dose and the accumulated dose of dCTs for each patient, as shown in figure 2.7. The minimum gamma index of 97.73% corresponds to patient 5, who had a weight loss of 17.8%. Deduced from here, the DIR-based application on patients with larger anatomical changes would introduce more DIR uncertainty than those with less anatomical changes.

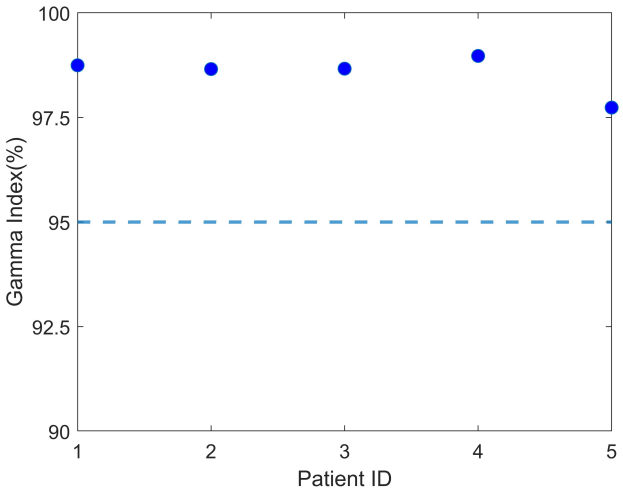


Figure 2.7: The gamma index between the accumulated dose from dCTs and the planning dose for each patient.

The DDVHs that demonstrate the influence of the A+DIR uncertainty and the DIR uncertainty are compared in figure 2.8 for organs of interest, including two target volumes: low-risk CTV and high-risk CTV; three serial OARs: chiasm, brainstem and spinal cord; and three parallel OARs: parotid glands, oral cavity and larynx. From visual assessment, the dosimetric discrepancy from the DIR uncertainty was much smaller than from the anatomical uncertainty. Significant differences between the influence of the DIR uncertainty and the A+DIR uncertainty were found on OARs.

The MADD (mean±95% confidence interval) quantified the average DDVH area across the five patients for each organ of interest. The MADD of the A+DIR uncertainty and the DIR uncertainty are compared in table 2.2. Anatomical uncertainty can increase the MADD of parallel OARs, serial OARs and CTVs by at least 3 times, around 2 times, and around 1.5 times, respectively.

Table 2.2: The quantification of A+DIR uncertainty and DIR uncertainty using the MADD for each organ.

Uncertainty	MADD (mean + 95 CI%) (Gy)							
	CTVs		Parallel OARs			Serial OARs		
	Low-risk CTV	High-risk CTV	Parotid glands	Oral cavity	Larynx	Brainstem	Spinal cord	Chiasm
A+DIR	1.36±1.19	1.35±1.19	2.52±4.3	1.53±3.43	2.96±3.59	1.98±1.11	1.67±1.74	3.3±7.5
DIR	0.56±0.22	0.53±0.55	0.6±0.3	0.33 ±0.19	0.74±0.34	0.74±0.73	0.47±0.29	1.02±0.93

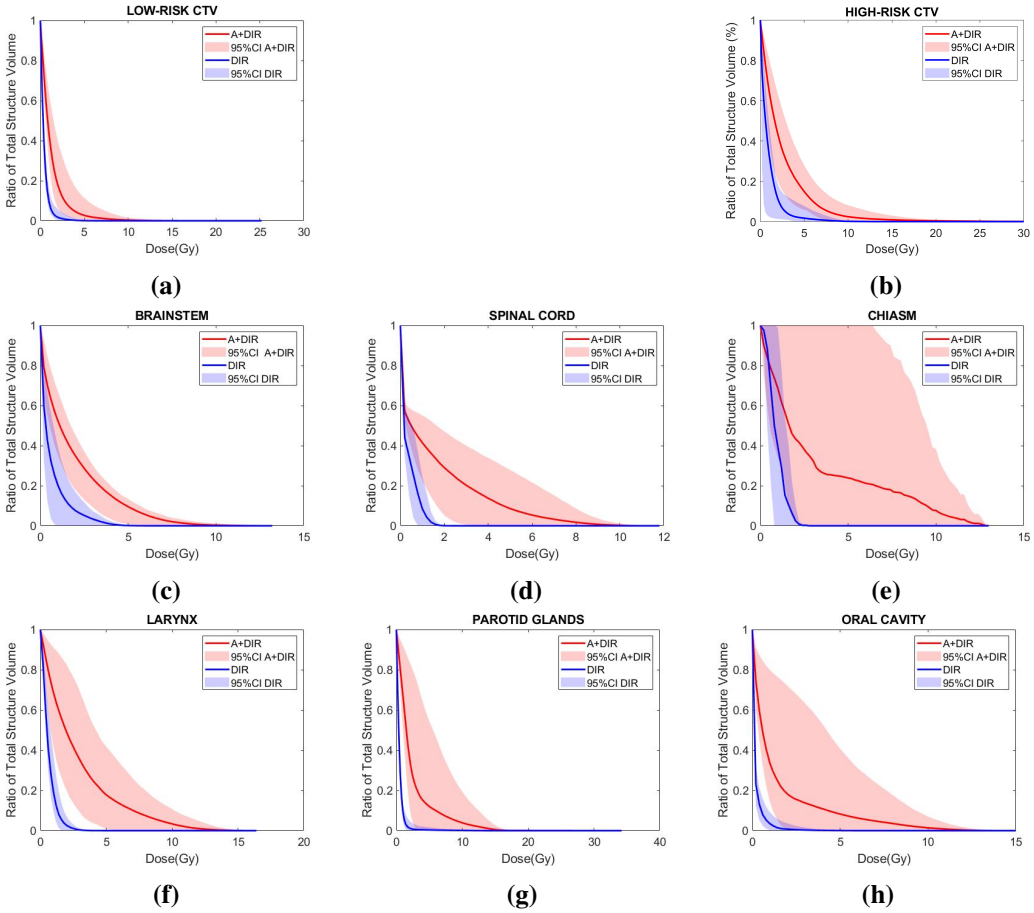


Figure 2.8: The comparison between the accumulated dose discrepancy from DIR uncertainty and A+DIR uncertainty on DDVH across the five patients. The shaded area is the variation between the patients.

2.4 Discussion

This chapter quantifies the DIR uncertainty for building anatomical models for H&N proton therapy. The DIR algorithm was evaluated from three aspects: 1) in the C-DIR evaluation, the maximum MSD (3-dimensional measurement) was less than 3 mm (slice thickness). 2) in the R-DIR evaluation, the average WSLD during the course of treatment was 1.03 mm, close to the pixel size (0.98 mm). 3) in the dCT-DIR evaluation, the influence of the accumulated DIR uncertainty on the clinically-concerned dose metrics can be controlled within 2 Gy. The gamma index between the planning dose and the accumulated dose of dCTs was above 97.73% (95% is the standard passing rate generally accepted [69, 70]). Furthermore, compared with the MADD caused by the A+DIR uncertainty, the MADD caused by the

DIR uncertainty was small. These results justified the rationality of accepting this DIR uncertainty for anatomical models.

The C-DIR and the R-DIR evaluation were used to explore the weekly changing trend of the DIR uncertainty. Their results revealed that the DIR uncertainty increases along with the treatment, validating that tissue shrinkage increases the difficulty in finding the anatomical correspondence between two images.

In the dCT-DIR evaluation, the dose deviations caused by the DIR uncertainty in the parallel OARs were the smallest (close to the vertical line of 0), in contrast to the serial OARs, where the largest deviations were present (see figure 2.8). Among the serial OARs, the dose on the chiasm is more sensitive to DIR uncertainty than the brainstem and spinal cord because the chiasm volume is small. The workflow proposed in the dCT-DIR evaluation eliminated the delineation uncertainty. It improved the evaluation accuracy regarding the accumulated influence of the DIR on each organ (especially on small organs, which are more sensitive to delineation error).

In the dCT-DIR evaluation based on MADD analysis, the dose deviations caused by the DIR uncertainty were compared with the A+DIR uncertainty. Overall, the dosimetric discrepancy from the DIR uncertainty was smaller than from the anatomical uncertainty on both mean value and confidence intervals. Because plans were created to ensure target coverage, the MADD differences between the DIR uncertainty and the A+DIR uncertainty on CTVs were smaller than those on OARs.

Spot error map is a useful tool developed for uncertainty analysis. It gave us an intuitive visual view of the possible spot location variations, which can guide the use of beam angles and the design of objectives in optimisation. For example, the error map can capture the 'dangerous spots' with high variations, which might damage critical normal tissues. Therefore, clinicians can avoid those spot positions or increase the weight of normal tissue protection in the optimisation procedure.

Previous studies used the WEPL changes of a beam to specific points or areas to evaluate the uncertainty. However, the contribution of spot weights on the uncertainty evaluation was ignored. To address this limitation, WSLD was proposed

as a surrogate of dose distribution, ensuring computing accuracy and efficiency. Although WSLD is an easier way to quantify the uncertainty, it is less effective at reflecting the changes of each organ at risk because proton spots in a plan are distributed around the radiation target.

2.5 Conclusion

This study aims to evaluate the diffeomorphic image registration of NiftyReg for building anatomical models for H&N proton therapy. The evaluation was based on contour, proton spot location and dose distribution. The weekly DIR uncertainty was mostly smaller than the slice thickness. The accumulated influence of DIR uncertainty measured by the gamma index was above 95% on average.

The WSLD was proposed to evaluate uncertainty and applied to DIR uncertainty evaluation. This new evaluation based on WSLD does not rely on contour selection and only gives a single value, making it easy to analyse. This WSLD tool is also used to measure anatomical uncertainty and validate the anatomical model in Chapters 4 and 5.

The workflow of evaluating the accumulated influence of DIR uncertainty reduced the influence of delineation error, thus improving the evaluation accuracy. The methods and steps presented in this study can directly be applied to other DIR algorithm evaluations.

Chapter 3

Inclusion of non-rigid variations of head and neck patients for IMPT plan robust evaluation

This chapter explored the way to include small non-rigid variations in the pre-treatment robust evaluation to assist IMPT beam selection.

The work in this chapter resulted in the following outputs:

Ying Zhang, Jailan Alshaikhi, Richard Amos, Wenyong Tan, Gary Royle, Esther Bär. Small Non-rigid Variations can Assist Robust IMPT Plan Selection for Head and Neck Patients. American Society for Radiation Oncology 2021.

Ying Zhang, Jailan Alshaikhi, Richard A. Amos, Wenyong Tan, Yaru Pang, Gary Royle and Esther Bär. "Pre-treatment analysis of non-rigid variations can assist robust IMPT plan selection for head and neck patients." Medical Physics, 49 (2022):7683-7693

Contribution of Authors: Jailan Alshikhi created the proton plans. Wenyong Tan provided the clinical data and valuable input. Richard Amos, Yaru Pang, Virginia Marin Anaya and Gary Royle provided valuable input concerning the clinical importance and impact of this work. Esther Bär supervised the project and provided valuable input. I developed the main ideas of the study, including the design of methods, code implementation, tool development for data processing and the final analysis.

3.1 Background

IMPT offers the potential to limit doses to normal tissues for H&N cancer patients [2, 3, 4, 5, 71, 72]. However, anatomical variations in the radiation area increase dosimetric uncertainty during treatment delivery [22, 21]. Wu *et al.* (2017) [22] showed that CTV mean doses were reduced by up to 7% in 10 oropharyngeal patients. Heukelom *et al.* (2019) [23] revealed that the dosimetric influence of anatomical changes led to >5% NTCP increase for dysphagia and the other toxicities. While these dosimetric influences are often correlated with progressive changes, studies have yet to reveal the dosimetric impact of small non-rigid variations (sNRVs), such as neck folds, neck tilts and spine flexions, on proton therapy plans. These sNRVs cannot be simplified as rigid translations, and unlike progressive changes that are patient-specific, sNRVs occur randomly.

Current research in H&N proton therapy delivery focuses on developing adaptive strategies to mitigate the influence of progressive anatomical changes. In clinical practice, offline adaptive planning strategies are applied when a certain threshold of dose to a critical structure is reached [54, 43]. This method is effective, but delays in implementing adaptive re-plans exist due to the time required for imaging, replanning, plan approval, and plan verification. Online adaptation is an aspirational technique intended for same-day application. However, due to the current computational speed, online adaptation either compromises the accuracy or constrains the optimiser [49, 51, 48]. The median adjustment time of 12 minutes was reported for online adaptation based on MC simulation [51, 48]. When plans are adapted either online or offline, the patient position may be different from the position in the image. sNRVs not captured during imaging will still be present.

In addition to adaptive planning strategies that mitigate the dosimetric impact of anatomical variability, evaluation of plan robustness is also used [47, 73]. Set up and range uncertainty are taken into consideration in conventional robust evaluation. Treatment plan evaluation, including inter-fractional anatomical variations, often uses images acquired during the course of treatment [46, 45, 33], and as such, they can only inform the planning process for a portion of the treatment delivery. A

more complete robust evaluation, including the possible sNRVs before treatment, is crucial to design a plan that is robust towards these anatomical changes. Because sNRVs are not patient-specific, they can be included in robust evaluations to provide additional information before treatment.

This chapter aims to: 1) build a distribution of possible sNRVs based on population data. 2) assess the dosimetric impact of random sNRVs on the dose delivered from IMPT plans; 3) incorporate sNRVs into the robust evaluation to account for anatomical uncertainty before treatment. Although only sNRVs are considered in the new robust evaluation, this study evaluates its effectiveness in indicating plan robustness to inter-fractional anatomical changes, including both sNRVs and progressive changes. 4) quantitatively validate the benefit of the new evaluation method. This new evaluation technique is compared to the conventional robust evaluation with the gold-standard evaluation (after-treatment evaluation that used weekly repeated CTs) as the reference for quantification.

3.2 Methods

3.2.1 Patient data

Twenty NPC patients with weekly repeat CT and fifteen oropharynx cancer patients with weekly CBCT who received photon therapy were recruited retrospectively. The deformations between week 0 (planning CT) and week 1 of treatment for all 35 patients were obtained, creating a distribution of possible sNRVs based on the method described in 3.2.2. Examples of sNRVs are shown in Appendix B.1. Four nasopharynx patients who had weekly repeat CTs were randomly selected as the test dataset, where the 35 sNRVs were applied to their planning CT (see section 3.2.2).

This study evaluated the robustness of IMPT plans towards the uncertainties (see section 3.2.3) applied to the test dataset based on the following scenarios: 1) plans with different numbers of fields from 3 fields to 5 fields; 2) plans with different beam angles. The different beam arrangements used are listed in the upper part of table 3.1 and illustrated in appendix B.2. The targets (both tumour and nodal

area) were split for different fields in these IMPT plans. All plans were robustly optimized using ± 3 mm setup and $\pm 3.5\%$ range uncertainty in Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). A relative biological effectiveness (RBE) of 1.1 for proton beams was used. The dosimetric goals for all plans in this study are summarised in the lower part of table 3.1. A plan was deemed robust (stop optimisation) if the goals set for the CTVs and serial organs were fulfilled for all 12 dose distributions (3 mm orthogonal shifts combined with the $\pm 3.5\%$ range error) as well as the nominal scenario (the error-free distribution).

Table 3.1: Plan beam arrangements and dosimetric goals used in this paper.

Plan beam arrangements	
Beam arrangements	Angle
$3B_{45}$	45 180 315
$3B_{60}$	60 180 300
$4B_{110}$	60 110 250 300
$4B_{120}$	60 120 240 300
5B	60 110 180 250 300
Dosimetric goals of the treatment plans	
Structure	Metric Goal Under Uncertainty
High-risk CTV	$D_{95} > 95\%$ of prescription dose (72.6 Gy)
Low-risk CTV	$D_{95} > 95\%$ of prescription dose (63 Gy)
High-risk CTV	$D_2 < 107\%$ of prescription dose
Spinal cord	D_{max} <45 Gy
Brainstem	D_{max} <55 Gy
Chiasm	D_{max} <55 Gy
Optical Nerve	D_{max} <55 Gy
Structure	Metric Goal in Nominal
Parotid glands	D_{mean} <26 Gy
Oral cavity	D_{mean} <40 Gy
Larynx	D_{mean} <40 Gy
Cochlea	D_{mean} <45 Gy

3.2.2 Extracting small non-rigid variations from CT images

Anatomical variations during the first week of treatment are predominately due to sNRVs, whereas progressive changes (weight loss, tumour shrinkage) are less significant [74, 75, 76]. Thus, the anatomical changes in the first week from a cohort of patients can be seen as the representatives of a distribution of possible sNRVs.

The sNRVs of a cohort of patients (see section 3.2.1) were captured using the diffeomorphic image registration. To apply the deformation between groups of subjects, the stationary velocity fields (SVFs) of diffeomorphic image registration

between week 0/week 1 need to be projected to an atlas space, in which all the SVFs have the same position and resolution. The atlas was obtained from a group-wise registration which spatially normalised a cohort of patients. The procedure of generating the atlas is illustrated in B.3. In the procedure of the projection, the planning CT (pCT) of each patient was the reference geometry, and the weekly repeat CT acquired during the first treatment week (rCT_t) was registered to the pCT to produce $\boldsymbol{v}_{p\to t}$, where p stands for pCT and t stands for the week when the weekly CT was acquired (in this case $t = 1$). Then, each patient's pCT was registered to the atlas to produce $\boldsymbol{v}_{p\to a}$, where a stands for atlas. $\boldsymbol{v}_{a\to p}$ transformed the inter-patient velocity fields $\boldsymbol{v}_{p\to t}$ into the atlas space using

$$\boldsymbol{v}_{a,p\to t} = \boldsymbol{v}_{a\to p} \circ \boldsymbol{v}_{p\to t} \circ \boldsymbol{v}_{a\to p}^{-1}, \quad p \forall P. \tag{3.1}$$

P includes all the patients' data used in this study.

Then $\boldsymbol{v}_{a,t\to p}$ was transformed into the space of an individual patient $\tilde{p}$ using

$$\boldsymbol{v}_{\tilde{p}\to t} \approx \boldsymbol{v}_{a\to\tilde{p}}^{-1} \circ \boldsymbol{v}_{a,p\to t} \circ \boldsymbol{v}_{a\to\tilde{p}}. \tag{3.2}$$

The deformation $\boldsymbol{v}_{\tilde{p}\to t}$ was used for warping the pCT to simulate a sNRV. Finally, in order to warp the pCT, the transformation must be directed from the predicted anatomy to the pCT. This can be simply achieved by reversing the SVFs using

$$\boldsymbol{v}_{t\to\tilde{p}} = -\boldsymbol{v}_{\tilde{p}\to t}. \tag{3.3}$$

The warped image $\mathrm{CT}^{\mathrm{sNRV}}$ for patient $\tilde{p}$ was acquired from:

$$\boldsymbol{\phi}_{t\to\tilde{p}} = \exp(\boldsymbol{v}_{t\to\tilde{p}}), \tag{3.4}$$

$$\mathrm{CT}_{\tilde{p}}^{\mathrm{sNRV}} = \boldsymbol{\phi}_{t\to\tilde{p}}(\mathrm{pCT}), \tag{3.5}$$

with t = 1 for all the equations above. This method produced 35 $\mathrm{CT}^{\mathrm{sNRV}}$s for each

patient to represent the possible sNRVs.

3.2.3 Robustness evaluation

The 35 sNRV scenarios of each test patient were included in the robustness evaluation using CT^{sNRV}s. For the four test patients, the dose distributions of IMPT plans were calculated under each robustness scenario. 1) The robust evaluation based on the sNRV scenarios was compared with 2) the conventional evaluation that only included rigid translations. A probability analysis was used in these two before-treatment evaluation techniques to rank the robustness of IMPT plans for each dose metric listed in the lower part of table 3.1 (see figure 3.1).

3.2.3.1 Robustness evaluation scenarios

For the proposed evaluation method using sNRV scenarios (1), this study simulated the isocenter shift for each cardinal direction (x_n,y_n,z_n) following the Gaussian distribution with mean $\mu = 0\,\text{mm}$ and standard deviation $\sigma = 1.5\,\text{mm}$ [45] on the 35 CT^{sNRV}s. This was done to calculate the perturbed dose distribution caused by the sNRVs and setup uncertainty since the CT^{sNRV}s have the same isocenter as the pCT. The so found 35 dose distributions for each IMPT plan were included in this sNRV+setup evaluation.

The conventional evaluation (2) was simplified to only include the setup uncertainty by applying the same isocenter shifts used in the sNRV+setup evaluation to the pCT. This way, 35 perturbed dose distributions per IMPT plan were included to evaluate the robustness of the plan.

3.2.3.2 Probability analysis for robust evaluation

The workflow for the sNRV+setup evaluation (1) and the conventional evaluation (2) is illustrated in figure 3.1. Each considered dose metric Dx (e.g. D_{95}) would have the corresponding perturbed dose metrics under the uncertainty scenarios. The nominal dose metric is subtracted from the perturbed dose metrics to form a

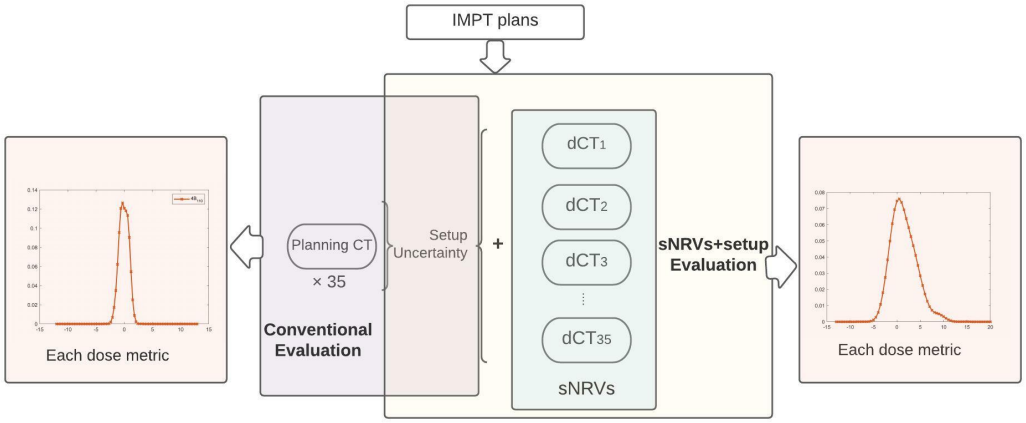


Figure 3.1: The workflow of the sNRV+setup evaluation and the conventional evaluation for each IMPT plan. Both evaluation methods produce 35 perturbed dose distributions. Each considered dose metric (Dx, e.g. D_{95}) would have the corresponding perturbed dose metrics under the uncertainty scenarios. The nominal dose metric is subtracted from the perturbed dose metrics to form a distribution of dose metric discrepancies (ΔDx) experienced across the uncertainty scenarios. The light yellow box indicates that IMPT plans were calculated on the uncertainty scenarios from sNRVs plus setup in the sNRV+setup evaluation. The light purple box indicates that IMPT plans were calculated on the setup uncertainty scenarios in the conventional evaluation. The dose distribution of ΔD_{95} in the conventional evaluation and the sNRV+setup evaluation was plotted in the left organ box and the right organ box, respectively, as an illustration.

distribution of dose metric discrepancies ΔDx experienced across the uncertainty scenarios.

The upper and lower boundaries of dose metrics in the evaluation can be demonstrated by the shaded areas in the nominal DVHs as an indicator of worst-case scenarios. It was also suggested in the literature [77] to include a probability approach in robust analysis. For this, the distance between the probability distribution of ΔDx under uncertainty and its ideal probability distribution (Dirac delta function, the dose metrics do not change even under uncertainty) was calculated using the Wasserstein distance (WD):

$$\mathrm{WD}(U,I) = \int_{-\infty}^{\infty} |U(x) - I(x)| dx, \tag{3.6}$$

where U and I are the probability distribution functions of ΔDx under uncer-

tainty and its ideal distribution, respectively. The WD measures the effort required to convert the distribution U into the I. The smaller the WD, the more robust a plan for this dose metric.

A two-sample t-test was used to determine if there is a significant difference between the distribution of ΔDx in the sNRV+setup evaluation and the conventional evaluation. $p < 0.05$ was taken as the significance level.

3.2.3.3 Performance analysis of robust evaluations

To investigate the effectiveness of sNRVs in indicating the plan robustness to inter-fractional anatomical changes before treatment, the dose discrepancy between the accumulated dose using weekly CTs and the nominal dose was taken as the gold standard. In the gold standard evaluation, the dose distributions of the IMPT plans with different beam arrangements were calculated on 6 weekly CTs of each test patient. Because the accumulated dose is generally used in treatment evaluation and related to prognostics, the weekly dose was accumulated in the reference frame of the pCT using the DIR algorithm of Niftyreg, referred to as Accu_{Nom}. Assuming that sNRVs are present in all treatment weeks, Accu_{Nom} should reflect that. In the weekly dose calculation, although the isocenter was determined using the information from the rigid registration, the setup error (both rigid and sNRVs) still existed. Thus, the difference between Accu_{Nom} and the nominal plan, referred to as ΔD_{st}, represents the influences from actual progression uncertainty and setup uncertainty (both rigid and sNRV). ΔD_{st} was used in the gold standard evaluation to assess the robustness of a plan.

Because different beam arrangements were used in this study, the robustness of beam arrangements can be ranked, referred to as robustness ranking. In the sNRV+setup evaluation and the conventional evaluation, the WD was used in robustness ranking for each dose metric. In the gold standard evaluation, ΔD_{st} was used in robustness ranking. To quantify the performance of the sNRV+setup evaluation and the conventional evaluation, the consistency C of the robustness ranking for a dose metric between an evaluation X and the gold standard G was measured by the weighted ranking discrepancy, defined as

$$C = \sqrt{\sum_i (RP_X(B_i) - RP_G(B_i))^2 \times w(B_i)}, \tag{3.7}$$

$RP_X(B_i)$ represents the robustness ranking position of a beam arrangement B_i in evaluation X. X is either the sNRV+setup evaluation or the conventional evaluation. The corresponding RP_X is RP_s in the sNRV+setup evaluation and RP_c in the conventional evaluation. $RP_G(B_i)$ represents the robustness ranking position of a beam arrangement B_i in the gold standard evaluation. $w(B_i)$ is the robustness weight of beam arrangement B_i in the gold standard evaluation to give a higher weight to the most distinguished beam arrangement. The weighting $w(B_i)$ is calculated based on ΔD_{st} using

$$w(B_i) = \frac{\sum_{j,j\neq i} |\Delta D_{st}(B_i) - \Delta D_{st}(B_j)|}{\sum_i \sum_{j,j\neq i} |\Delta D_{st}(B_i) - \Delta D_{st}(B_j)|}. \tag{3.8}$$

$|RP_X(B_i) - RP_G(B_i)|$ was used to identify whether the inclusion of sNRVs can benefit all beam arrangements. If $|RP_s(B_i) - RP_G(B_i)| \leq |RP_c(B_i) - RP_G(B_i)|$ for a dose metric, then this dose metric of beam arrangement B_i supports that the sNRV+R evaluation is better for robust evaluation, compared to the conventional evaluation. Regarding the 12 dose metrics listed in the lower part of table I for each beam arrangement, this study calculated the percentage of dose metrics that supports the inclusion of sNRV in robust evaluation based on the results of the four test patients.

The consistency C was used to assess the role of sNRVs in robust evaluation based on all beam arrangements. The consistency of the robustness ranking between the sNRV+setup evaluation and the gold standard evaluation was referred to as $C_{\text{sNRV+setup\&G}}$. The consistency of the robustness ranking between the conventional evaluation and the gold standard evaluation was referred to as $C_{\text{C\&G}}$. When $C_{\text{sNRV+setup\&G}}$ is smaller or equal to $C_{\text{C\&G}}$ ($R_{\text{sNRV}} \geq 0$) for a matrix, it supports the conclusion that sNRVs play a positive role in robust evaluation. C = 0 means that the robustness ranking in a before-treatment robust evaluation is the same as the gold standard robust evaluation.

3.3 Result

3.3.1 Dosimetric influences caused by small non-rigid variations

This section demonstrates the additional dosimetric discrepancies caused by the sNRVs on an exemplary patient.

An example of the dose distribution variations caused by a sNRV is shown in figure 3.2. The red arrows indicate the areas where the dose has fallen under 95% of the prescription dose.

Figure 3.3 compares the sNRV+setup evaluation with the conventional evaluation on an exemplary patient (patient 1) based on dose metrics. The upper and lower boundaries of dose metrics in the sNRV+setup evaluation (3.3a) and the conventional evaluation (3.3b) are indicated by the shaded areas in figure 3.3a) and b) separately. It can be observed that the additional sNRVs widen the bandwidth compared to the conventional robust evaluation. The comparisons of the dose metrics for this patient based on box plots are shown in figures 3.3c) to h). Dose metrics for the different plans with different beam arrangements are shown in the same figures as box plots. By comparing the boxplot of c)–e) (sNRV+setup evaluation) to f)–h) (conventional evaluation) in figure 3.3, the mean values of the CTVs' D_{95} in the sNRV+setup evaluation are lower than those in the conventional evaluation ranging from -1.57% to -0.95% (the range shows the differences between different beam arrangements). The mean values of parotid D_{mean}, oral cavity D_{mean} and larynx D_{mean} are higher than the values in the conventional evaluation, ranging from 1.02 Gy to 1.82 Gy, 0.52 Gy to 0.70 Gy, and 0.84 Gy to 3.18 Gy, respectively. The mean values of D_{max} of the spinal cord, optical nerves and the chiasm between the two evaluations only have slight differences, less than 0.6 Gy.

Figure 3.3 only partially demonstrates the Dx under uncertainty. Figure 3.4 plots the probability distribution of ΔDx in the conventional evaluation and the sNRV+setup evaluation on high-risk CTV D_{95} and parotid D_{mean}, respectively, for patient 1. Figure 3.4 shows the influence caused by the sNRVs on the probability distribution of ΔDx from different beam arrangements. The robustness of a beam arrangement is presented by the closeness of the probability curve of beam arrange-

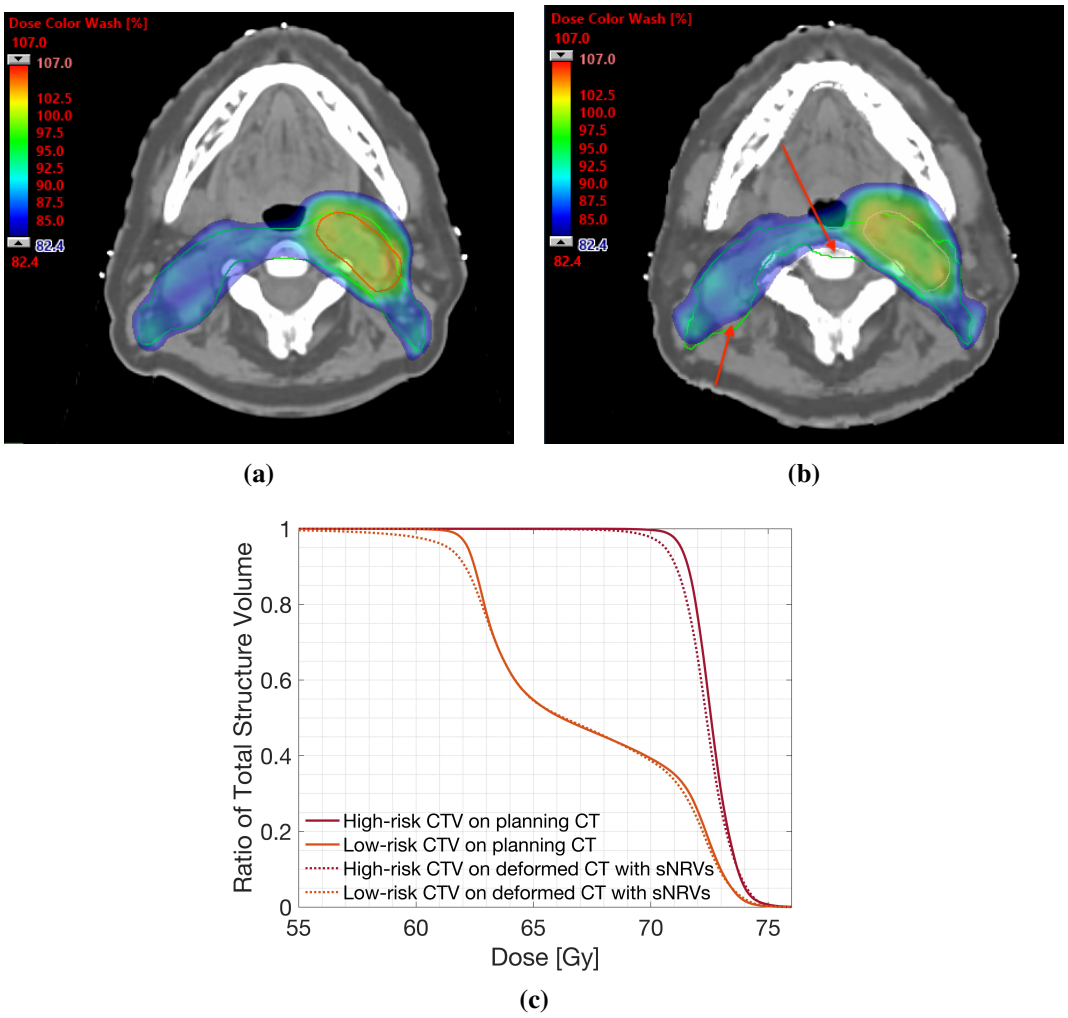


Figure 3.2: An example of the dose distribution variations caused by a random sNRV. (a) is the dose distribution on the planning CT, and (b) is the dose distribution of the same slice on a CT^{sNRV}. The green contours in the images are the Low-risk CTV. The colour bar was chosen to mask out doses lower than the 95% prescription dose of Low-risk CTV (82.4% is corresponding to the 95% prescription dose of Low-risk CTV). The red arrows indicate the areas of underdosage caused by the sNRV. (c) presents the difference in the DVH caused by the sNRV.

ments to the Dirac delta function. For the high-risk CTV, the $3B_{60}$ plan is the most robust (the ΔD_{95} curve of the $3B_{60}$ is the closest to the Dirac delta function, indicated as the dashed vertical line) in the sNRV+setup evaluation (figure 3.4b)), as opposed to the conventional evaluation, where this beam arrangement is found to be the less robust one (figure 3.4a)). For the parotid glands, in the conventional evaluation (figure 3.4c)), the $4B_{120}$ is the most robust beam arrangement, while in

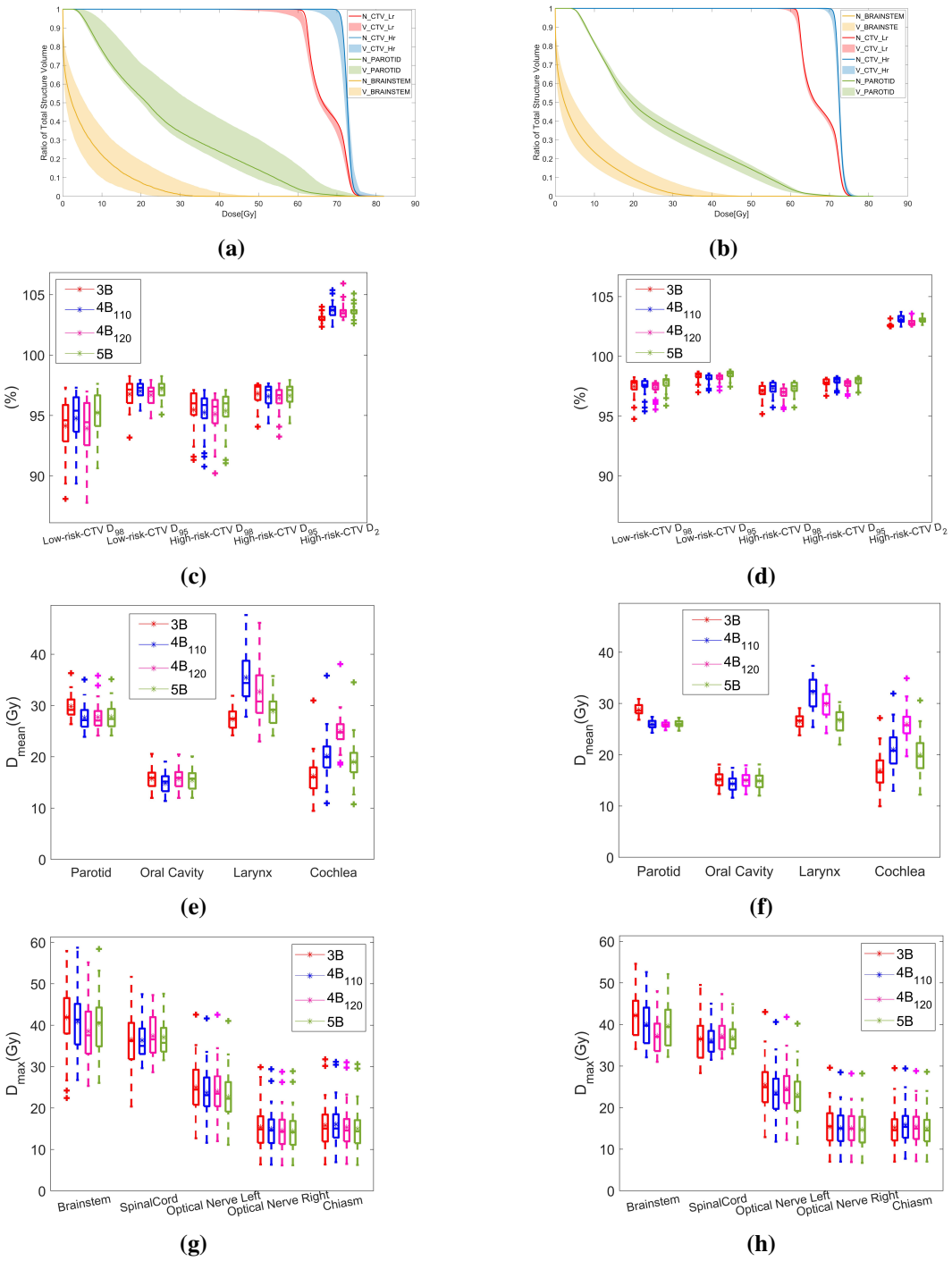


Figure 3.3: The comparison between the sNRV+setup evaluation and the conventional evaluation on patient 1. (a)-(b) show the shaded DVH from the $4B_{120}$ beam arrangement in the sNRV+setup evaluation and in the conventional evaluation, respectively. The solid line represents the DVH of the nominal plan (N in the legend), and the shaded area indicates the lower and upper boundary of the dose metrics in the respective evaluation caused by the variations (V in the legend). (c)-(h) visually summarise the statistics under the respective uncertainty using box plots. The horizontal lines in the box plot indicate the median dose metric among 36 scenarios (including the nominal scenario and 35 uncertainty scenarios defined in robust evaluation). The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The asterisks are used to indicate the mean value of the dose metrics. (c), (e) and (g) are the boxplots of Dx in the sNRV+setup evaluation. (d), (f) and (h) are the boxplots of Dx in the conventional evaluation.

the sNRV+setup evaluation (figure 3.4d)), the $3B_{60}$ is the most robust.

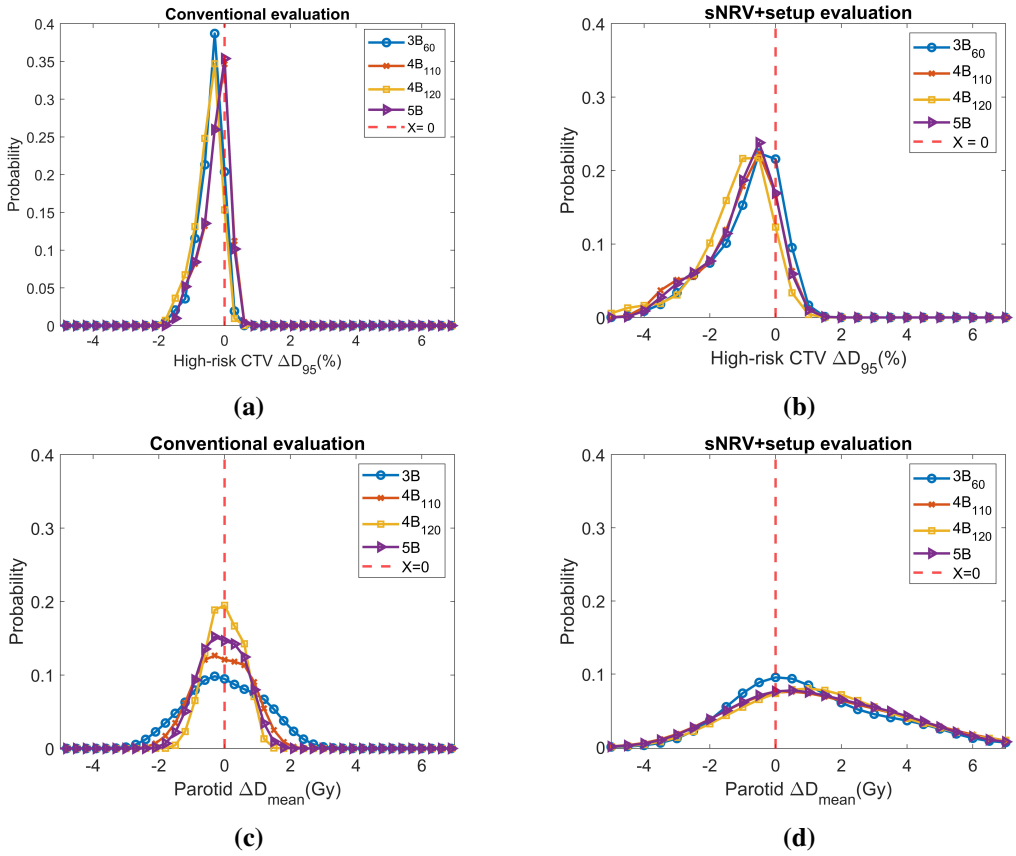


Figure 3.4: Probability distributions of ΔDx in the conventional evaluation and in the sNRV+setup evaluation on patient 1. (a) Probability distribution of ΔDx in the conventional evaluation on the high-risk CTV D_{95}. (b) Probability distribution of ΔDx in the sNRV+setup evaluation on the high-risk CTV D_{95}.(c) Probability distribution of ΔDx in the conventional evaluation on the parotid D_{mean}. (d) Probability distribution of ΔDx in the sNRV+setup evaluation on the parotid D_{mean}.

3.3.2 Robust evaluation analysis

The dosimetric evaluation in table 3.2 shows the detailed numbers of the sNRV+setup evaluation and the conventional evaluation for the exemplary patient 1. The results of the remaining three patients are listed in appendix B.4. Among four patients (490 scenarios), a maximum difference was observed in the sNRV+setup evaluation to the nominal dose of: 9.37% dose degradation on the D_{95} of CTVs, increase in parotid D_{mean} by 11.87 Gy, increase in larynx D_{mean} by 15.04 Gy, increase in brainstem D_{max} by 20.82 Gy, increase in spinal cord D_{max} by

20.96 Gy. For CTVs, 4 patients all had scenarios where the CTV D_{95} fell below 95%, 47 out of 490 scenarios in total. In contrast, in the conventional evaluation, a maximum difference was observed to the nominal dose of: 7.58% dose degradation on D_{95} of the CTVs, increase in parotid D_{mean} by 4.88 Gy, increase in larynx D_{mean} by 6.13 Gy, increase in brainstem D_{max} by 13.5 Gy, increase in spinal cord D_{max} by 12.9 Gy.

Please note that the worst-case CTV coverage (D_{95}) under setup uncertainty can drop below 95% in some cases. To generate 35 scenarios in conventional robust evaluation, the isocenter was shifted following a Gaussian distribution with mean μ= 0 mm and standard deviation σ= 1.5 mm. This results in multiple scenarios that can be used for statistical analysis, rather than the only 12 scenarios usually encountered during robust optimisation with 3mm orthogonal shifts and $\pm$3.5% range error. While the usual 3 mm option was still used to optimise the plan, the additional shifts created with the Gaussian distribution were used for the evaluations. Using this Gaussian distribution may result in scenarios where the shift exceeds 3 mm. However, only 4/490 scenarios were below 95%. Those scenarios only happened to patient 3 whose target volume was located close to the skin, making this particular patient more sensitive to setup uncertainty. This comparison between the sNRV+setup evaluation and the conventional evaluation in table 3.2 demonstrated the additional dose discrepancy that sNRVs can lead to.

3.3.3 Performance analysis of robust evaluations.

Based on table 3.2 and the table in appendix B.4, the calculated percentages of dose metrics satisfying $|RP_s(B_i) - RP_G(B_i)| \leq |RP_c(B_i) - RP_G(B_i)|$ are 91.7%, 79.2%, 75%, 75% and 75% for $3B_{45}$, $3B_{60}$, $4B_{110}$, $4B_{120}$, 5B respectively, showing that the inclusion of sNRVs is beneficial to the robust evaluation of all beam arrangements.

Table 3.2 intuitively compares different evaluations based on the RP. The RP difference was quantified as the consistency of robustness ranking between the two evaluation methods in table 3.3. Table 3.3 (from the last row) shows that the worst P_{sNRV} for one patient is 58.33% (still $>$ 50%) for patient 4, and P_{sNRV} is above 75% in 3/4 patients. Overall, P_{sNRV} is 77.1% for all patients. The P_{sNRV} in the last col-

Table 3.2: The sNRV+setup evaluation, the conventional evaluation and the gold standard evaluation for patient 1. p-values of the two-sample t-test are calculated between the distribution of ΔDx in the sNRV+setup evaluation and in the conventional evaluation. $p<0.05$ represents that ΔDx in the sNRV+setup evaluation and ΔDx in the conventional evaluation are taken from significantly different distributions. In the sNRV+setup evaluation and the conventional evaluation, the beam arrangements were ranked based on the WD for each ROI matrix. In the gold standard evaluation, the beam arrangements were ranked for each ROI matrix based on ΔD_{st}. RP is the robustness ranking position of a beam arrangement for a matrix.

	sNRV+setup Evaluation						Conventional Evaluation				p-value	Gold standard evaluation		
ROI/Matrix	Plan	Nominal	Min	Max	WD	RP_s	Minmum	Maximum	WD	RP_c		$Accu_{Nom}$	ΔD_{st}	RP_G
Low-risk CTV	$3B_{60}$	98.57	93.17	98.25	1.68	4	96.83	98.57	0.37	3	1e-10	97.62	-0.95	3
D_{95}(%)	$4B_{110}$	98.25	95.4	97.94	1.22	1	96.83	98.25	0.32	2	1e-9	97.62	-0.63	2
	$4B_{120}$	98.25	94.76	97.94	1.57	3	96.83	98.25	0.29	1	1e-11	97.3	-0.95	4
	5B	98.57	95.08	98.25	1.4	2	97.14	98.57	0.39	4	1e-10	97.94	-0.63	1
High-risk CTV	$3B_{60}$	97.66	94.08	97.66	0.83	1	96.14	97.66	0.46	3	0.01	97.66	0	1
D_{95}(%)	$4B_{110}$	97.66	94.35	97.66	1.08	4	96.42	97.8	0.28	1	1e-5	97.38	-0.28	3
	$4B_{120}$	97.66	93.25	97.66	1.05	3	96.14	97.52	0.52	4	1e-4	97.38	-0.28	3
	5B	97.66	94.35	97.93	1.04	2	96.42	97.8	0.32	2	1e-5	97.38	-0.28	3
High-risk CTV	$3B_{60}$	102.89	102.34	103.99	0.18	1	102.75	103.58	0.11	1	0.9	101.79	-1.1	4
D_2(%)	$4B_{110}$	103.44	102.34	105.37	0.45	3	103.03	104.27	0.25	4	0.64	103.17	-0.28	2
	$4B_{120}$	102.89	102.89	105.92	0.66	4	102.75	103.86	0.17	2	1e-6	102.89	0	1
	5B	103.44	102.62	105.1	0.28	2	103.17	104.13	0.18	3	0.93	102.89	-0.55	3
Parotid	$3B_{60}$	28.62	26.36	36.31	1.68	1	26.68	30.74	0.77	4	1e-3	31.48	2.86	1
D_{mean}(Gy)	$4B_{110}$	26.11	23.89	35.07	2.04	4	24.41	27.51	0.55	3	1e-3	31.82	5.72	3
	$4B_{120}$	25.94	24.14	35.81	1.97	3	24.85	26.77	0.34	1	1e-4	31.1	5.16	2
	5B	26.29	24.15	35.14	1.8	2	24.88	27.42	0.43	2	1e-4	32.24	5.95	4
Oral	$3B_{60}$	15.17	11.98	20.57	1.6	2	12.42	18.17	1.14	3	0.24	15.56	0.38	2
D_{mean}(Gy)	$4B_{110}$	14.33	11.38	19.06	1.56	1	11.69	17.53	1.13	2	0.26	14.7	0.37	1
	$4B_{120}$	15.05	11.99	20.46	1.63	4	12.36	18.01	1.08	1	0.13	15.71	0.66	4
	5B	14.87	11.99	20.07	1.62	3	12.1	18.18	1.16	4	0.23	15.27	0.39	3
Larynx	$3B_{60}$	26.6	24.18	31.89	1.71	1	24.09	29.39	1.05	1	0.17	28.96	2.36	1
D_{mean}(Gy)	$4B_{110}$	32.21	27.81	47.64	4.22	3	25.66	37.63	2.55	4	1e-3	42.42	10.2	3
	$4B_{120}$	30.11	22.99	46.12	4.27	4	24.64	34.01	1.97	3	0.02	44.4	14.29	4
	5B	26.78	24.11	35.75	2.76	2	22.31	30.6	1.72	2	1e-3	33	6.22	2
Cochlea	$3B_{60}$	15.53	9.45	30.97	2.38	2	9.09	26.29	2.65	2	0.83	13.79	-1.74	4
D_{mean}(Gy)	$4B_{110}$	20.21	10.95	35.83	2.53	4	12.53	31.56	3.01	4	0.63	18.51	-1.69	3
	$4B_{120}$	25.11	18.49	38.08	1.89	1	19.38	34.61	2.24	1	0.41	23.43	-1.68	2
	5B	19.02	10.75	34.52	2.5	3	11.72	30.11	2.88	3	0.65	17.46	-1.56	1
Brainstem	$3B_{60}$	43.02	22.42	57.93	5.83	3	34.72	55.22	4.2	4	0.57	40.84	-2.18	4
D_{max}(Gy)	$4B_{110}$	40.26	26.75	58.72	5.98	4	32.46	52.96	4.18	3	0.75	40.54	0.29	2
	$4B_{120}$	36.43	25.33	55.17	5.57	1	30.73	47.73	3.39	1	0.25	37.63	1.2	3
	5B	39.51	26.06	58.4	5.79	2	32.41	52.31	4.07	2	0.71	39.71	0.21	1
Spinal	$3B_{60}$	36.68	20.36	51.68	5.6	4	28.38	49.58	4.22	4	0.96	33.25	-3.43	4
D_{max}(Gy)	$4B_{110}$	35.11	29.63	47.5	3.54	2	30.81	44.31	2.57	2	0.34	34.92	-0.2	1
	$4B_{120}$	36.65	28.64	47.28	4.08	3	31.85	47.35	2.97	3	0.95	34.59	-2.07	3
	5B	35.47	31.55	47.61	3.11	1	31.87	43.87	2.31	1	0.13	36.56	1.09	2
Chiasm	$3B_{60}$	23.68	12.7	42.56	4.88	4	12.28	42.38	4.78	4	0.9	22.79	-0.89	3
D_{max}(Gy)	$4B_{110}$	22.29	11.61	41.64	4.52	2	11.29	40.09	4.57	1	0.67	21.41	-0.88	2
	$4B_{120}$	22.9	12	42.53	4.62	3	11.7	41.3	4.63	3	0.96	21.9	-0.99	4
	5B	21.62	11.11	41.03	4.51	1	10.72	39.62	4.61	2	0.88	20.77	-0.85	1
Optic nerve left	$3B_{60}$	13.97	6.37	29.86	4.05	4	6.27	28.87	3.72	4	0.56	16.44	2.47	4
D_{max}(Gy)	$4B_{110}$	13.6	6.32	29.38	3.75	3	6.2	27.7	3.49	3	0.52	15.3	1.7	1
	$4B_{120}$	13.44	6.15	28.74	3.65	2	6.14	27.44	3.47	1	0.79	15.64	2.19	3
	5B	13.23	6.18	28.9	3.65	1	5.83	27.33	3.48	2	0.61	15.16	1.93	2
OPTIC_NERVEV_R	$3B_{60}$	13.97	6.37	31.77	4.22	4	6.37	28.87	3.54	4	0.3	19.74	5.77	4
D_{max}(Gy)	$4B_{110}$	14.37	6.92	23.76	3.14	3	6.77	28.47	3.47	3	0.33	19.81	5.44	1
	$4B_{120}$	13.75	6.5	23.25	3.09	2	6.15	27.85	3.4	1	0.44	19.31	5.56	2
	5B	13.42	6.24	22.81	3.06	1	6.02	27.62	3.43	2	0.4	18.99	5.57	3

umn shows that $C_{\text{sNRV+setup\&G}}$ was always better or equal to $C_{\text{C\&G}}$ for the parotid glands, larynx, cochlea, spinal cord, chiasm. In the example of patient 1 from table 3.2, the WD of the sNRV+setup evaluation indicated that $3B_{60}$ was the most robust beam arrangement for the parotid D_{mean}, while the conventional evaluation indi-

cated that $4B_{120}$ was the most robust. The ΔD_{st} from the gold standard evaluation validated that $3B_{60}$ indeed was the most robust beam arrangement for the parotid D_{mean}.

3.4 Discussion

Dose distributions in proton therapy are more sensitive to geometric changes than in photon therapy. However, in previously published robust evaluation methods, the impact of anatomical changes was not considered before treatment. This study demonstrated the dose degradation on CTVs and OARs caused by the sNRVs and showed that including the sNRVs in the robust evaluation is beneficial.

3.4.1 The use of small non-rigid variations for robust beam selection

In the validation of sNRVs' role in robust evaluation, the dose discrepancy that represents the influence of inter-fractional anatomical changes and isocenter shifts was used as the gold standard. The consistency of robustness ranking showed that $C_{sNRV+setup\&G}$ was smaller than $C_{C\&G}$, especially on the low-risk CTV, parotid glands, larynx and spinal cord, which are closely related to outline changes and neck motions, and also on small structures that are sensitive to sNRVs such as cochlea and chiasm, supporting that sNRVs play a positive role in robust evaluation in terms of indicating robustness to inter-fractional anatomical changes.

No beam arrangement is insensitive to the sNRVs as anatomical changes were not considered in robust optimisation for any beam arrangement. The method proposed in this study can still assist in selecting robust beam arrangements for proton plans without anatomical robust optimisation. In table II and table B1 in Appendix B, the p-values between the distributions of ΔDx in the sNRV+setup evaluation and the conventional evaluation showed that the sNRVs mainly influenced the probability distribution of CTVs ΔD_{95} and parotid ΔD_{mean}. Because the highest priority of the robust optimisation for the four test patients in this study was to ensure the target coverage, similar performance of $D_{95\%}$ based on ΔD_{st} was found on different beam arrangements, with differences smaller than 2%. To best demonstrate the ad-

Table 3.3: Consistency of robustness ranking between two robust evaluation methods. When $C_{sNRV+setup\&G}$ is smaller or equal to $C_{C\&G}$ ($R_{sNRV} \geq 0$), the result supports the sNRV+setup evaluation and the numbers of $C_{sNRV+setup\&G}$ are highlighted in bold. The percentage of cases that support sNRVs in robust evaluation is referred to as P_{sNRV}. The P_{sNRV} of the final row summarizes the percentage of the ROI matrix that supports the sNRVs+setup evaluation ($R_{sNRV} \geq 0$) for each patient. The P_{sNRV} of the final column summarizes the percentage of patients in which the sNRVs+setup evaluation is better than the conventional evaluation ($R_{sNRV} \geq 0$) for each ROI metric. $P_{sNRV}(\%) > 75$ are marked by an asterisk (based on the last column, at least more than half of the patients agree, thus 3/4 is used).

ROI Matrix	Patient 1		Patient 2		Patient 3		Patient 4		$P_{sNRV}\ (\%) = \frac{\sum Patient(R_{sNRV} \geq 0)}{\sum Patient}$
	$C_{sNRV+setup\&G}$	$C_{C\&G}$	$C_{sNRV+setup\&G}$	$C_{C\&G}$	$C_{sNRV+setup\&G}$	$C_{C\&G}$	$C_{sNRV+setup\&G}$	$C_{C\&G}$	
Low-risk CTV $D_{95}(\%)$	**1**	2.12	1.15	0.58	**0**	0.87	0.87	0.71	50
High-risk CTV $D_{95}(\%)$	**0.58**	1.73	**0.58**	1.15	**1.41**	1.41	1.73	0.79	75*
High-risk CTV $D_{2}(\%)$	2.43	2.04	1.41	1.08	1.46	0.79	**1.50**	1.73	25
Parotid D_{mean}(Gy)	**1.10**	2.19	**0**	2.25	**0.78**	1.73	**0.72**	0.72	100*
Oral D_{mean}(Gy)	**0**	2.19	**1.06**	2.08	**1.55**	1.55	1.48	0	75*
Larynx D_{mean}(Gy)	**0**	0.71	**0.61**	0.61	**0**	0.74	**0**	0	100*
Cochlea D_{mean}(Gy)	**1.72**	1.72	**1**	2.30	**0**	0	**0.75**	1.73	100*
Brainstem D_{max}(Gy)	1.51	1.13	**0.73**	1.99	**0**	0	1.53	0.84	50
Spinal D_{max}(Gy)	**0.69**	0.69	**0.62**	1	**0**	0	**1.46**	1.73	100*
Chiasm D_{max}(Gy)	**0.76**	1	**1.92**	2.02	**0**	0	**1.52**	1.73	100*
Optic nerve left D_{max}(Gy)	**1.25**	1.40	**1.77**	2.05	**1.73**	1.73	1.68	0.86	75*
Optic nerve right D_{max}(Gy)	1.35	1.22	**1.10**	1.27	**1.73**	1.73	**1.54**	1.54	75*
$P_{sNRV}\ (\%) = \frac{\sum ROI(R_{sNRV} \geq 0)}{\sum ROI}$	75*		83.33*		91.67*		58.33		77.1*

vantage of the sNRV+setup evaluation over the conventional evaluation, the beam arrangement was selected based on the impact of the sNRVs on the parotid ΔD_{mean} as an illustration. The parotid ΔD_{mean} is also closely related to toxicity such as xerostomia which can have a long-term impact on patients' quality of life. For example, for patient 1, a similar parotid D_{mean} was achieved using $4B_{110}$ and $4B_{120}$. If $4B_{120}$ was selected based on WD, the accumulated parotid D_{mean} would reduce by 0.7 Gy, corresponding to 1 fraction of D_{mean} delivered to the parotid glands. Other organs can be used for beam selection as well, for example, the oral cavity D_{mean} in the sNRV+setup evaluation has the exact same rank as the gold standard. Also, the rank of the chiasm D_{max} in the sNRV+setup evaluation indicated the most robust beam arrangement as the gold standard evaluation.

There were two interesting scenarios worth noticing. In different beam arrangements for patient 1, even though the nominal parotid D_{mean} of $3B_{60}$ was the highest, the accumulated dose was lower than $4B_{110}$ and 5B because $3B_{60}$ was the most robust beam arrangement (the lowest WD) under the sNRV+setup uncertainty. The ΔD_{st} of $3B_{60}$ showed that $3B_{60}$ controlled the ΔD_{st} of the parotid D_{mean} within 3 Gy, which corresponds to 10% NTCP difference [54] and is used to trigger a replan to protect the parotid glands. This case indicated that beam selection based on robust evaluation can potentially reduce the replan rate, something that needs further investigation in the future. For patient 3, even though the nominal parotid D_{mean} of $4B_{120}$ was higher than in the 5B beam arrangements, the accumulated dose was the lowest because $4B_{120}$ was the more robust beam arrangement. A message that clearly emerges here is that the best nominal plan may not be the best plan during treatment.

The impact of different beam angles on the robustness of a plan can be analysed on patient-specific geometry using our method. The results can be used to create a robustness plan database to assist in finding a more robust planning approach, as presented by McGowan *et al.* (2015) [47] and Malyapa *et al.* (2016) [73].

3.4.2 The potential use of small non-rigid variations in clinic

The distribution of sNRVs has the potential to be used in other clinical applications.

Firstly, this study found that the sNRV that leads to the most dose discrepancy varies from patient to patient and from beam arrangement to beam arrangement. For example, the robustness ranking in parotid D_{mean} based on the WD of the sNRV+setup evaluation was $4B_{120} > 3B_{60} > 5B$ for patient 1, while $3B_{60} > 4B_{120} > 5B$ for patient 4. This is because individuality was considered in the method by using the patient-specific image. This approach can help clinicians avoid the set-ups with the sNRVs that can lead to unacceptable dose distributions for a specific patient using a specific beam arrangement.

Secondly, the acquired sNRVs can potentially assist in better estimating the truly delivered accumulated dose using weekly CTs. The sNRVs can be randomly allocated to each weekly CT, with 5 sNRVs per weekly CT. These deformed weekly CTs can be used to estimate the daily dose distribution under the influence of sNRVs. Repeating this procedure can reveal the range of potential accumulated doses for the whole treatment.

Thirdly, this study presented the possibility of including sNRVs from a patient population in robust analysis, which also indicated the potential to be used in robust optimisation. Mesías *et al.* (2019) [43] included the first two weekly CTs of patients into robust optimisation to account for the sNRVs, suggesting that sNRVs can reduce the need for adaptation. They indicated that the first two weekly CTs can be replaced by a series of CT images scanned before treatment. Li *et al.* (2015) [34] considered weekly CTs in the robust evaluation. Yang *et al.* (2020) [44] added the adaptive planning CTs into robust optimisation. However, their methods relied on the acquisition of CT images during treatment, which limited the creation of a robust plan at the early planning stage. In contrast to their patient-specific approach, an atlas-based technique is suggested here. While this approach is not patient-specific but based on the assumption that sNRVs are mainly random, there are some advantages. First, this method does not require the acquisition of a series of CT images of the same patient pre-treatment, therefore saving imaging dose and reducing workload. Secondly, assuming that sNRVs can be reasonably represented using this method, deformed images with the sNRVs can be prepared in advance

and fully exploit the benefits of robust optimisation with multiple CTs. This will be investigated in future studies.

It should be mentioned that the inclusion of a large patient cohort (many sNRV scenarios) would require recalculating the treatment plan many times. For efficiency, the included sNRVs should be limited to the most common/frequent ones. The most common sNRVs can be found, for example, by using the anatomical models, which apply principal component analysis (PCA) to anatomical deformations of a patient cohort to estimate the likelihood of a certain anatomical deformation to happen (please refer to Chapter 4). By only including the most likely principal components of the deformation in the robust evaluation, the number of recalculated plans can be reduced while still representing well the sNRVs. This trade-off will be explored in future work.

The concept presented here can be adapted to different scenarios. This study did not factor in immobilization equipment and patient characteristics such as age, size, disease staging and physical condition. All those factors are likely to influence the possibility and the amplitude of a specific anatomical change to arise during the treatment. While this is not yet considered in this study, the presented approach has the potential to do so. If sufficient patient data are available to build the atlas, the patient data can be stratified into groups based on immobilization devices and patient characteristics before performing a robust evaluation.

3.4.3 Limitations

For the purpose of showing the feasibility, plans with different beam arrangements were only created for the four test patients. Further validation of the method will be conducted on a large number of patients. Another limitation of this work is that the impact of the DIR uncertainty was ignored in the robust evaluations. This study assumed that the DIR uncertainty would equally affect the robust evaluations for different beam arrangements.

When validating the role of sNRVs in robust evaluation, this study did not take the range uncertainty from Hounsfield units (HU) into account because range uncertainty is an isolated source considered in the robust evaluation and is solely

based on the CT calibration. It is also justified by the small dosimetric impact from CT number range uncertainty compared to setup uncertainty and anatomical uncertainty[33, 47]. Therefore, it should only have small influences on the results of the comparison, which established that sNRV should be considered in the robust evaluation as a component of random set-up errors, not just rigid setups. However, to fully evaluate the plan, the range uncertainty should be used along with the rigid setups and sNRVs.

3.5 Conclusion

This study aims to demonstrate the additional dose discrepancy arising from the sNRVs and the influence of the sNRVs on robust plan evaluation. The deformations during the first week were used to build the distribution of possible sNRVs. A maximum difference was observed in the sNRV+setup evaluation to the nominal dose of: 9.37% dose degradation on D_{95} of CTVs and 11.87 Gy increase on parotid (D_{mean}). These dose metrics are unacceptable in the clinic and can trigger plan adaptation. Benefiting from this atlas-based method, patients' unique geometry was considered. This study found that the sNRV that can lead to the maximum difference varies from patient to patient and from beam arrangement to beam arrangement. One application of this analysis is to help clinicians avoid worse setups.

This study also used a probability analysis to select robust plans against sNRVs. Based on the quantitative validation, this study concluded that the sNRVs-based evaluation is better than the conventional evaluation. The percentages of dose metrics satisfying $|RP_s(B_i) - RP_G(B_i)| \leq |RP_c(B_i) - RP_G(B_i)|$ were at least 75% for all beam arrangements, showing that the inclusion of sNRVs is beneficial to the robust evaluation of all beam arrangements. In the measurement of the robustness ranking consistency with the gold standard evaluation, the sNRV+setup evaluation was better or equal to the conventional evaluation in 77.1% of cases, particularly better on parotid glands, larynx, chiasm, cochlea, and spinal cord.

The method provided in this study can potentially provide multiple images for anatomical robust optimisation.

Chapter 4

Probability model for head and neck patients

This chapter built an anatomical model to assist with anatomical robust optimisation for head and neck patients.

The work in this chapter resulted in the following outputs:

Ying Zhang, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. "A probability model for anatomical robust optimisation in head and neck cancer proton therapy." Physics in Medicine & Biology (2022).

Contribution of Authors: Jailan Alshikhi created the proton plans. Wenyong Tan provided the clinical data and valuable input. Gary Royle provided valuable input concerning the clinical importance and impact of this work. Esther Bär supervised the project. I developed the idea of PM and model evaluation, including developing codes and tools for data processing and the final analysis.

4.1 Background

The anatomical robust optimisation has been investigated to include small non-rigid variations (sNRVs) and (or) progressive changes to improve the plan robustness against anatomical uncertainty [43, 44]. However, it either requires multiple CT scans before planning [43] or needs the images acquired during treatment [44]. While multiple scanning gives the extra imaging dose to H&N patients and affects the efficiency of a busy proton therapy practice, the dependence on images acquired

during the treatment also compromises the benefits of reducing the replanning rate. The current online adaptation techniques also need to consider sNRVs in the planning optimisation because the reported time requirement (the median adjustment time of 12 minutes based on MC simulation) [50, 51] is a burden for patients lying on the couch and waiting for the optimisation to be finished. During this waiting time, the patient's position may be different from the position in the image. sNRVs not captured during imaging will still be present.

Including sNRVs in robust evaluation/robust optimisation using the method in Chapter 3 can be inefficient when a large patient cohort is available. Therefore, the included sNRVs should be limited to the most common/frequent ones. To this end, an anatomical model based on principal component analysis (PCA) becomes the focus of this study. PCA finds the best orthogonal basis, the principle components (PCs), whose variance of the projections of the data are ranked from the greatest to the smallest. Thus, it is possible to restore information using a limited number of PCs that describe the majority of anatomical deformations. The basics of PCA were detailed in Lever *et al.* (2017) [78]. This type of model simulates the possible geometric variations from a population of patient data. Thus, it removes the requirement of multiple scanning and the dependence on acquiring verification CTs during treatment for anatomical robust optimisation. Anatomical robust optimisation can be done at the planning stage using the predicted images. Several mathematical models have been proposed to account for anatomical changes [27, 60]. Yu *et al.* (2016) [60] assumed that the probability density function (PDF) of the coefficients α of each PC follows a standard normal distribution. However, they did not validate if the PDF of α from their dataset is consistent with this assumption, and the probability of each predicted deformation vector field (DVF) was not revealed. To date, it is still challenging to quantify the probability of a certain type of anatomical deformation to arise during the treatment course.

This study aims to: 1) develop a probability model (PM) at each weekly time point to address the limitations of the previous model[60]. 2) quantify the probability of each type of anatomical deformation based on population data. 3) validate

the feasibility of the PM to measure anatomical uncertainty.

4.2 Method and material

4.2.1 Patient data

Twenty NPC patients were recruited retrospectively. Each patient underwent a planning CT (pCT) and a weekly repeat CT (rCT_t), where t (t=0,1,2,3,...) represents the week of CT scanning. Because this dataset is relatively small, the leave-one-out strategy was used to build the model for more accurate performance. The leave-one-out strategy was applied for 5 test patients (one patient was held out as the test set, and the remaining patients were used as the training set to build the model, this process was repeated 5 times).

For all 5 test patients, an original (nominal) IMPT treatment plan was generated using Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). All plans generated throughout this study were robustly optimised with ± 3 mm setup and $\pm 3.5\%$ range uncertainty for CTVs and OARs. A RBE of 1.1 for proton beams was used. The dosimetric goals for all plans in this study are summarised in table 4.1. A plan was deemed acceptable if the goals set for the CTV and serial organs were fulfilled in the nominal scenario (the error-free distribution) as well as all 12 dose distributions (3 mm orthogonal shifts combined with the $\pm 3.5\%$ range error) in a robust evaluation. More clinical characteristics of the patients can be found in the papers of Tan *et al*. [18, 79].

4.2.2 Probability model

In the clinic, the magnitude of uncertainty is estimated from population data [28, 80, 36, 81]. To capture the major deformations in a population, this study developed the PM to statistically model the anatomical changes of the population based on PCA.

The weekly stationary velocity fields of diffeomorphic image registration (SVFs) $\boldsymbol{v}_{p\to t}$ between the pCT and the rCT_t of the training data were used as input to build the PM. The procedure was divided into the following steps and repeated for each treatment week t.

Table 4.1: Dosimetric goals of the treatment plans created in this study

Structure	Goal under uncertainty
High-risk-CTV	$D_{95} > 95\%$ of prescription dose (72.6 Gy, 33 fractions)
Low-risk-CTV	$D_{95} > 95\%$ of prescription dose (63 Gy, 33 fractions)
CTV	$D_2 < 107\%$ of prescription dose
Spinal cord	$D_{max} < 45$ Gy
Brainstem	$D_{max} < 55$ Gy
Chiasm	$D_{max} < 55$ Gy
Structure	Goal in Nominal
Parotid glands	$D_{mean} < 26$ Gy
Oral cavity	$D_{mean} < 40$ Gy
Larynx	$D_{mean} < 40$ Gy
Proton planning information: MFO planning; spot spacing size: 5mm; energy range:70 MeV – 250MeV; range shifter: 5cm; dose calculation algorithm: Piencel beam scanning (PBS); optimisation algorithm: Nonlinear Universal Proton Optimiser.	

1. The inter-fraction SVF $\boldsymbol{v}_{p\to t}$ of patients was projected into the atlas space as $\boldsymbol{v}_{a,p\to t}$ using 3.1.

2. The average SVF for treatment week t in the atlas space was calculated as the expectation value E of the deformation $\boldsymbol{v}_{a,p\to t}$ of the training dataset.

$$\mathbf{E}(\boldsymbol{v}_{a,p\to t}) = \frac{1}{N_p}\sum_{pi}\boldsymbol{v}^{pi}_{a,p\to t}, \tag{4.1}$$

where N_p is the number of patients used in this model, and pi is the patient index.

3. The random deformation of each patient at week t in the atlas space can be calculated as follows:

$$\boldsymbol{v}^{\text{rand},pi}_{a,p\to t} = \boldsymbol{v}^{pi}_{a,p\to t} - E(\boldsymbol{v}_{a,p\to t}). \tag{4.2}$$

4. The random deformations of all training patients at week t were composed to a random deformation matrix in the atlas space, referred to as $\boldsymbol{v}^{\text{rand}}_{a,p\to t}$, which was represented approximately using

$$\boldsymbol{v}_{a,p\to t}^{\text{rand}} = (\boldsymbol{v}_{a,p\to t}^{\text{rand},1}, \boldsymbol{v}_{a,p\to t}^{\text{rand},2}, \cdots, \boldsymbol{v}_{a,p\to t}^{\text{rand},N_p}) \approx \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,L} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N_p,1} & \alpha_{N_p,2} & \cdots & \alpha_{N_p,L} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \vdots \\ \boldsymbol{V}_L \end{bmatrix}. \tag{4.3}$$

$\boldsymbol{V}_l$ is the PC vector, also called eigenvector. $\alpha_{i,l}$ is the coefficient of the l-th eigenvector belonging to the i-th training set. L is the number of eigenvectors used to build the model. L was chosen to be able to represent 90% of population variations. Each column of the α matrix represents the coefficients of one eigenvector.

5. The the probability density function (PDF) of α of an eigenvector was annealed using kernel density estimation [82]. The estimation used

$$p_l(\alpha) = \frac{1}{N_p \cdot \sigma_l \sqrt{2\pi}} \sum_{i=1}^{N_p} \exp^{-(\alpha-\alpha_{i,l})^2/2\sigma_l^2}, \tag{4.4}$$

$$\sigma_l = \frac{1.06}{{N_p}^{0.2}} \sqrt{\frac{\sum_i^{N_p} (\alpha_{i,l} - \frac{1}{N_p}\sum_i^{N_p} \alpha_{i,l})^2}{N_p - 1}}, \tag{4.5}$$

where p_l is the probability distribution of the coefficients of the l-th eigenvector. The comparison between the real distribution of a column of α and the annealed distribution is shown in figure 4.1.

6. The sampled numbers from p_l formed the l-th column of the $\tilde{\alpha}$ matrix. Because the distribution was estimated from a limited training dataset, sampling extended the coefficients to capture all the possible random anatomical changes resulting in

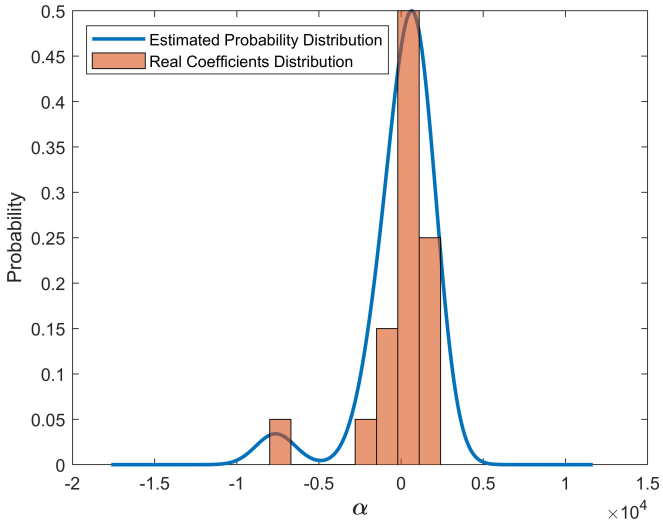


Figure 4.1: The comparison between the real distribution of a column of α and the annealed distribution.

$$\tilde{\boldsymbol{v}}^{\text{rand}}_{a,p\to t} = (\tilde{\boldsymbol{v}}^{\text{rand},1}_{a,p\to t}, \tilde{\boldsymbol{v}}^{\text{rand},2}_{a,p\to t}, \cdots, \tilde{\boldsymbol{v}}^{\text{rand},N_s}_{a,p\to t}) \approx \begin{bmatrix} \tilde{\alpha}_{1,1} & \tilde{\alpha}_{1,2} & \cdots & \tilde{\alpha}_{1,L} \\ \tilde{\alpha}_{2,1} & \tilde{\alpha}_{2,2} & \cdots & \tilde{\alpha}_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\alpha}_{N_s,1} & \tilde{\alpha}_{N_s,2} & \cdots & \tilde{\alpha}_{N_s,L} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \vdots \\ \boldsymbol{V}_L \end{bmatrix}. \tag{4.6}$$

Each row of the predictive matrix $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}} = (\tilde{\alpha}_{i,1}, \tilde{\alpha}_{i,2}, \cdots, \tilde{\alpha}_{i,L})$ in equation 4.6 was multiplied with $\boldsymbol{V}$ to form a predicted random deformation for week t, represented by $\tilde{\boldsymbol{v}}^{\text{rand},i}_{a,p\to t}$. i is the index of the predicted random deformations. N_s is the number of samples.

7. A deformation of the PM for week t is:

$$\boldsymbol{v}^{\text{PM},i}_{a,p\to t} = E(\boldsymbol{v}_{a,p\to t}) + \tilde{\boldsymbol{v}}^{\text{rand},i}_{a,p\to t}, \quad i\forall(1 \sim N_s). \tag{4.7}$$

8. Each deformation $\boldsymbol{v}^{\text{PM},i}_{a,p\to t}$ was transformed into the space of an individual patient using

$$\boldsymbol{v}_{p\to t}^{\mathrm{PM},i} = \boldsymbol{v}_{a\to p}^{-1} \circ \boldsymbol{v}_{a,p\to t}^{\mathrm{PM},i} \circ \boldsymbol{v}_{a\to p}. \tag{4.8}$$

9. To warp the pCT, the transformation must be directed from the predicted anatomy to the pCT. The deformation $\boldsymbol{v}_{t\to p}^{\mathrm{PM},i}$ needs to be reversed using:

$$\boldsymbol{v}_{t\to p}^{\mathrm{PM},i} = -\boldsymbol{v}_{p\to t}^{\mathrm{PM},i}. \tag{4.9}$$

10. The warped image $\mathrm{CT}_t^{\mathrm{PM},i}$ was acquired using:

$$\boldsymbol{\phi}_{t\to p}^{\mathrm{PM},i} = \exp(\boldsymbol{v}_{t\to p}^{\mathrm{PM},i}), \tag{4.10}$$

$$\mathrm{CT}_t^{\mathrm{PM},i} = \boldsymbol{\phi}_{t\to p}^{\mathrm{PM},i}(\mathrm{pCT}), \tag{4.11}$$

where $\mathrm{CT}_t^{\mathrm{PM},i}$ is the i-th predicted image of week t.

N_s predicted images can be obtained for week t. Considering that eigenvectors are orthogonal, the probability distribution of their coefficients is independent. Therefore, the probability of predicted images with specific $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ can be calculated by the joint probability

$$P(\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}) = p_1(\tilde{\alpha}_{i,1}) \cdot p_2(\tilde{\alpha}_{i,2}) \cdot \ldots \cdot p_L(\tilde{\alpha}_{i,L}), \qquad \sum_{i}^{N_s} P(\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}) = 1. \tag{4.12}$$

4.2.3 Model evaluation

4.2.3.1 Model evaluation based on weighted spot location deviation.

The total WSLD considered the probability estimated from the extended population was calculated using

$$\mathrm{WSLD}_t^{\mathrm{PM}} = \sum_{i=1}^{N_s} (\mathrm{WSLD}^{\mathrm{CT}_t^{\mathrm{PM},i}} \cdot P(\boldsymbol{\alpha}_{\boldsymbol{i}})). \tag{4.13}$$

$\mathrm{WSLD}^{\mathrm{CT}_t^{\mathrm{PM},i}}$ is the WSLD from the predicted image $\mathrm{CT}_t^{\mathrm{PM},i}$ at week t. $P(\boldsymbol{\alpha}^{\boldsymbol{i}})$ is the probability of the predicted image $\mathrm{CT}_t^{\mathrm{PM},i}$. N_s is the number of samples produced by the PM.

The WSLD was applied to evaluate the anatomical uncertainty from sNRVs, total anatomical uncertainty, and residual anatomical uncertainty from the PM.

- The anatomical uncertainty from sNRVs and total anatomical uncertainty simulated by the PM. The PM statistically summarised the probability of anatomical changes for each week. The progressive changes induced by the radiation in the first week were not significant [18]. Therefore, the WSLD estimated by the PM in the first week of treatment represented the influence of sNRVs such as tongue movement, shoulder positioning or small rotations. The WSLD from later fractions was the combined influence of sNRVs and progressive changes (total anatomical uncertainty).

- The residual anatomical uncertainty from the PM. The difference between the estimated anatomical uncertainty from the PM and the actual anatomical uncertainty of each week was used to evaluate the accuracy of the PM. It is referred to as the residual anatomical uncertainty ($\Delta\mathrm{WSLD}_t^{\mathrm{res}}$) from the PM, see equation 4.14,

$$\Delta\mathrm{WSLD}_t^{\mathrm{res}} = \mathrm{WSLD}_t^{\mathrm{real}} - \mathrm{WSLD}_t^{\mathrm{PM}}, \tag{4.14}$$

  where $\mathrm{WSLD}_t^{\mathrm{real}}$ is the actual anatomical uncertainty calculated by the WSLD between rCT_t and the pCT, which also corresponds to the residual anatomical uncertainty of no model. $\mathrm{WSLD}_t^{\mathrm{PM}}$ is the anatomical uncertainty at week t estimated by the PM. Ideally, the model should approach a ΔWSLD of 0 for each treatment week t.

4.2.3.2 Model evaluation based on dose distribution.

Anatomical deformations lead to dose variations. The dose on the deformed images was recalculated using the original IMPT plan. Then 1) the actual dose variations

from the training dataset were compared with 2) the dose variations simulated by the PM. The deformations of an exemplary patient in the first week were chosen to demonstrate the dosimetric influence of small non-rigid variations.

1) To obtain the actual dose variations, the actual deformations of 20 patients in the first week were applied to the planning CT of the exemplary patient to obtain 20 actual sNRVs. The IMPT plan of this exemplary patient was applied to the 20 actual sNRVs to calculate the dose variations. These variations can be illustrated as dose volume histogram (DVH) bands in the nominal DVHs of organs, referred to as actual DVH bands. The perturbed dose metrics were calculated for each considered dose metric Dx (e.g. D_{95}). The perturbed dose metrics subtracted the nominal dose metrics to obtain the dose metric discrepancy ΔDx.

2) To obtain the dose variations simulated by the PM, 20 CT_1^{PM} (CT of the PM in the first week) were selected following the joint probability distribution of the PM for the exemplary patient. The same IMPT plan was also applied to the 20 CT_1^{PM} to create DVH bands, referred to as simulated DVH bands from the PM. The dose metric discrepancy simulated by the PM was referred to as ΔDx'.

A two-sample t-test was used to determine if there is a significant difference between the distribution of ΔDx and ΔDx'. $p < 0.05$ was taken as the significance level.

4.3 Results

The exemplary patient's slice images from the planning CT and 2 predicted CTs of the PM in the first and the sixth week are shown in figure 4.2.

4.3.1 Model evaluation based on weighted spot location deviation.

The WSLD of the anatomical uncertainty is estimated from the PM. The result is shown in average WSLD with 95% confidence interval (CI) (see figure 4.3). The estimated anatomical uncertainty from sNRVs accounted for a range uncertainty of 2.18±0.19 mm. The estimated total anatomical uncertainty (from sNRVs and progressive changes) can reach 3.09±0.26 mm at week 6.

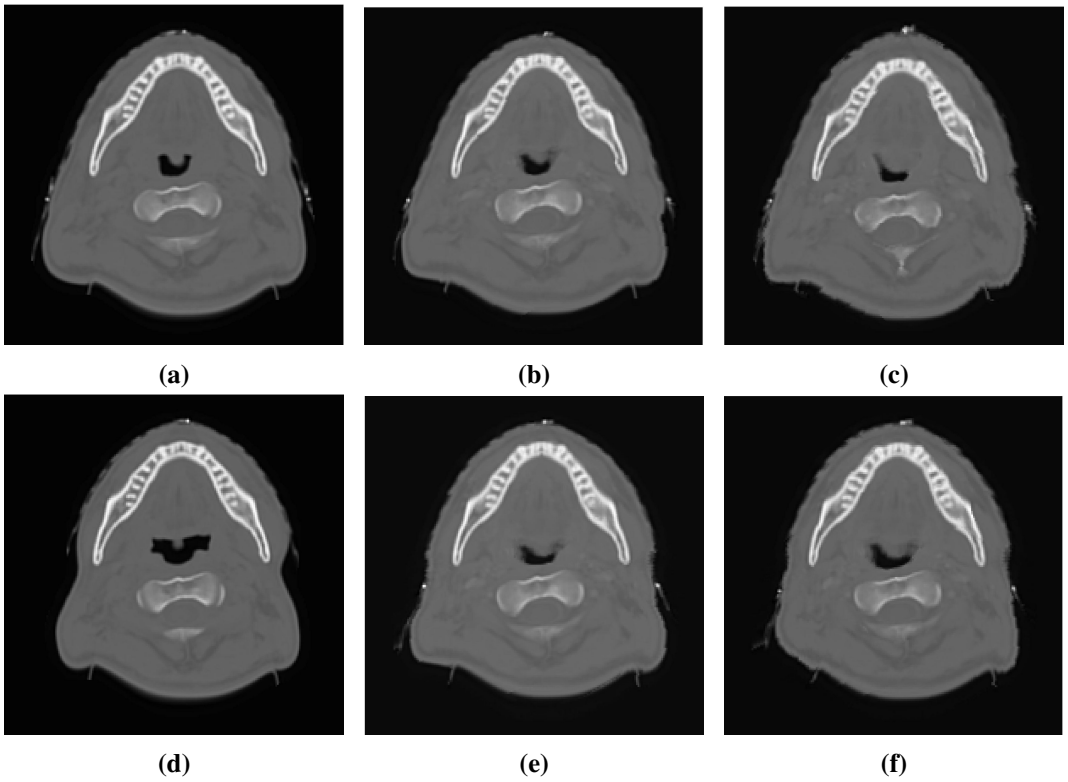


Figure 4.2: The exemplary patient's slice images from the rCT_1, rCT_6 and 2 predicted CT of the PM in the first week and the sixth week. (a) shows a slice image from the rCT_1 of the exemplary patient. (b)-(c) show slice images from 2 predicted CTs of the PM in the first week. (d) shows the same slice image from the rCT_6 of the exemplary patient. (e)-(f) show slice images from 2 predicted CTs of the PM in the sixth week.

The residual anatomical uncertainty from no model and the PM ($N_s = 100$) are compared in figure 4.4. In no model predicted images were replaced by the planning CT. When the anatomical uncertainty estimated from the PM was considered, the residual anatomical uncertainty was reduced from 4.47 ± 1.23 mm (no model) to 1.49 ± 1.08 mm (PM) at week 6, achieving a significant improvement as compared to no model.

The comparison of individual cases between the actual WSLD (using rCTs) and the estimated WSLD from the PM are listed in table 4.2.

4.3.2 Model evaluation based on dose distribution.

For the exemplary patient, the actual DVH bands in the first week and the simulated DVH bands from the PM in the first week are shown in figure 4.5. Supporting the

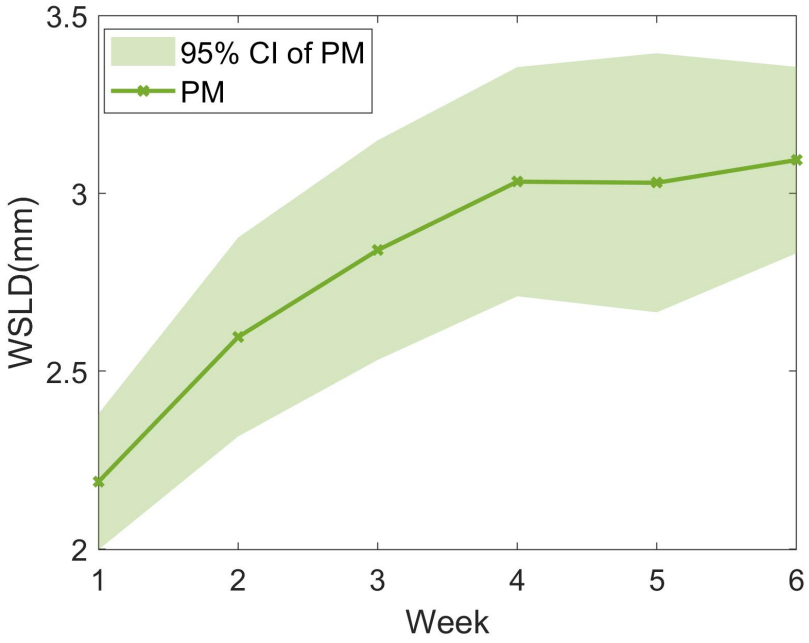


Figure 4.3: The anatomical uncertainty estimated from the PM in WSLD for each week. The result is estimated in average WSLD with 95% CI across the 5 test cases. The WSLD in the first week presents the uncertainty from sNRVs. The WSLD in the following weeks evaluates the combined effect of anatomical uncertainty from sNRVs and progressive changes (total anatomical uncertainty).

Table 4.2: The WSLD caused by the actual anatomical deformations (using rCTs) and the WSLD estimated by the PM in each test patient and week.

Week	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5	
	rCTs(mm)	PM(mm)	rCTs(mm)	PM(mm)	rCTs(mm)	PM(mm)	rCTs(mm)	PM(mm)	rCTs(mm)	PM(mm)
1	1.72	2.12	1.96	1.88	2.57	2.53	2.32	2.14	1.89	2.27
2	2.52	2.42	2.07	2.18	1.99	2.71	2.30	3.11	2.54	2.55
3	3.43	2.75	2.15	2.30	2.59	3.23	2.57	3.20	3.49	2.72
4	4.93	2.65	2.78	2.65	2.94	3.62	3.44	3.14	4.69	3.10
5	5.62	2.97	2.73	2.30	4.02	3.53	2.91	3.15	5.57	3.20
6	5.23	2.78	3.12	3.07	5.13	3.63	2.62	2.89	6.27	3.10

rationality of the PM, the simulated DVH bands of the PM demonstrated similar variations as the actual DVH bands.

The maximum, minimum, mean value (μ) and standard deviation(σ) of the ΔDx from the actual sNRVs and the ΔDx' from the simulated sNRVs of the PM are listed in table 4.3 with their p-values (between ΔDx and ΔDx').

The range of dose metric variations caused by the actual anatomical deformations in the first week was from -1.46 % to -0.05 % (low-risk CTV D_{95}), from -2.15 Gy to 6.83 Gy (parotid glands D_{mean}) and from -3.98 Gy to 12.59 Gy (spinal cord D_{max}) for low-risk CTV D_{95}, parotid glands D_{mean} and spinal cord D_{max}, re-

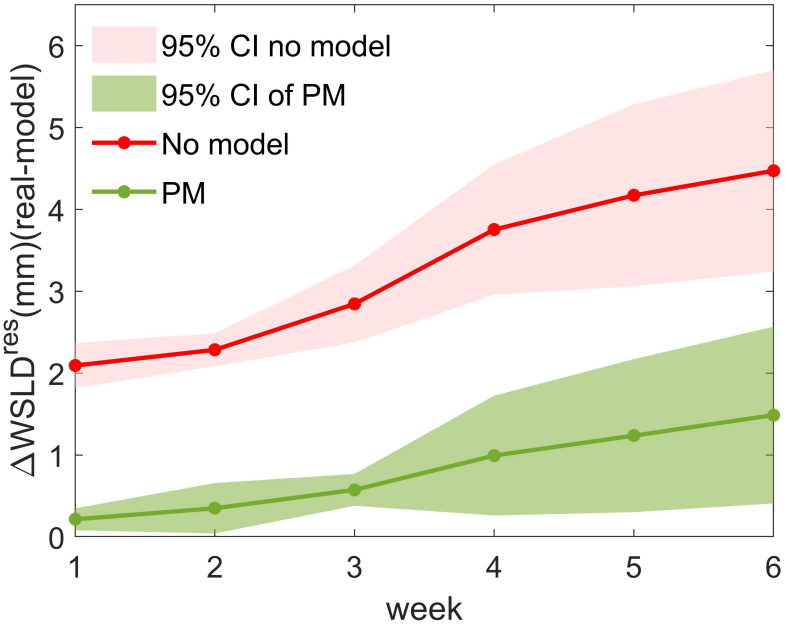


Figure 4.4: The residual anatomical uncertainty in WSLD. The residual anatomical uncertainty comes from no model, in which predicted images were replaced by planning CT, and the PM ($N_s = 100$) were compared. The result indicates the average difference with 95% CI between the estimated WSLD from models and the actual WSLD across the 5 test dataset.

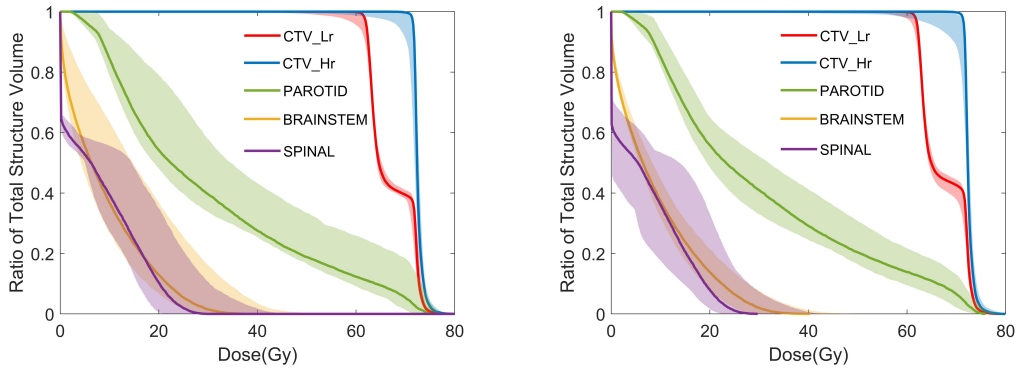


Figure 4.5: Evaluation of the PM based on dose distribution. a) shows the bandwidth from the actual DVH band in the first week. b) shows the bandwidth simulated from the PM in the first week. The initial N in the legend represents the nominal plan. The initial V in the legend represents the variations.

spectively. While the range of the dose metric variations simulated by the PM was from -0.97% to 0% (low-risk CTV D_{95}), from -2.35 Gy to 5.83 Gy (parotid glands D_{mean}) and from -3.04 Gy to 12.03 Gy (spinal cord D_{max}) for low-risk CTV D_{95},

Table 4.3: Dose metrics discrepancy (perturbed dose metrics - nominal dose metric) from the actual sNRVs (ΔDx) and the simulated sNRVs from the PM (ΔDx') are listed in maximum value, minimum value, mean value (μ) and standard deviation(σ). The p-values are calculated between ΔDx and ΔDx'.

	Low-risk CTV (ΔD_{95}(%))		High-risk CTV(ΔD_{95}(%))		Parotid Glands(ΔD_{mean}(Gy))		Brainstem(ΔD_2(Gy))		Spinal(ΔD_2(Gy))	
	Actual	Simulated	Actual	Simulated	Actual	Simulated	Actual	Simulated	Actual	Simulated
Maximum	-0.05	0	-0.19	0	6.83	5.81	8.91	4.98	12.59	12.03
Minimum	-1.46	-0.97	-3.12	-5.83	-2.15	-2.35	-5.08	-5.12	-3.98	-3.04
μ	-0.45	-0.43	-1.02	-0.88	1.95	1.4	-0.08	0.04	0.47	-0.15
σ	0.31	0.2	0.91	0.86	2.51	1.86	2.49	2.51	3.26	3.42
p-value	0.69		0.61		0.32		0.84		0.49	

parotid glands D_{mean} and spinal cord D_{max}, respectively. The p-values indicated that no significant difference was found between ΔDx and ΔDx' for all investigated dose metrics.

4.4 Discussion

In this chapter, an anatomical model was developed to evaluate anatomical uncertainty and quantify the probability of an anatomical deformation to arise during treatment. The model accuracy was evaluated based on WSLD and dose distribution.

The PM can simulate the small random variations in the first treatment week with an average error of 0.21±0.13 mm. For overall anatomical uncertainty prediction, the PM can reduce anatomical uncertainty from 4.47±1.23 mm (no model) to 1.49±1.08 mm at week 6 (see figure 5.5). The PM was considered suitable for estimating anatomical uncertainty.

For the dose metrics in table 4.3, the simulated dose metric range of the PM in the first week is basically within the range of the actual anatomical deformations, and the p-values between ΔDx and ΔDx' are > 0.05. supporting that the PM is feasible to simulate the anatomical variations.

For the purpose of validating the model, only 20 patients with weekly CT imaging, which is used less frequently in routine clinics than CBCT, were included to reduce the error from HU corrections when calculating the spot location. The procedure of using CBCT images to build the model is the same, except that the DIR error between CT and CBCT might be different and needs to be evaluated individually. The DIR error between CT and CBCT has been investigated in the literature

[83, 84].

The inclusion of more scenarios in the model training can improve the probability estimation. In probabilistic treatment planning or robust optimisation, uncertainty scenarios are often described using uniform distribution [42] or normal distribution [41] in the cost function. However, it is difficult to correspond one uncertainty scenario with a probability. This work exploited the independence between PCs to calculate the probability for each predicted CT. This can be used to design the cost functions of anatomical robust optimisation.

In very recent studies, plan robustness against anatomical changes was investigated by anatomical robust optimisation. Mesías *et al*. (2019) [43] and Yang *et al*. (2020) [44] both concluded that this method improved plan robustness toward anatomical changes and reduced the number of plan adaptations for H&N patients. However, Mesías *et al*. (2019) [43] required multiple scanning to produce extra CT images before treatment for robust optimisation. It will give extra imaging dose to the H&N patient and add a burden to a busy clinic. Yang *et al*. (2020) [44] used the image from the first plan adaptation to include the progressive anatomical changes in the second adaptive plan. However, it limits the creation of a robust plan at the early planning stage. To overcome these limitations, the PM was developed based on population data to capture systematic progressions and comprehensive random deformations of H&N patients, making it possible to include anatomical changes before treatment without extra burden [42]. For online adaptation, their results were based on the static images acquired several minutes (the median reported adaptation time: 12 minutes) before treatment application[48]. Considering the possibility of small patient movements during the waiting time, sNRVs can be included in robust optimisation for current online adaptation techniques. Such changes can, for example, be inferred from the here suggested PM in the first treatment week.

4.5 Conclusion

The PM used PCA to capture the major deformations of each treatment week. Exploiting the orthogonal of eigenvectors, a solution to calculate the probability of a certain type of anatomical change was given and applied to the anatomical uncertainty evaluation based on WSLD. In this WSLD-based evaluation, the PM can predict the anatomical uncertainty during the course of radiotherapy with an average error of 0.81 ± 0.56 mm. In the model evaluation based on the dose distribution, p-values for all investigated dose metrics indicated that no significant difference was found between the dose metrics discrepancy caused by the real deformations and the simulated deformations of the PM in the first week.

As the PM can accurately predict anatomical uncertainty during the course of radiotherapy, it has great potential to be used in clinical applications, such as anatomical robust optimisation.

Chapter 5

Individual anatomical model for head and neck patients

This chapter built anatomical models for prospective offline replanning for head and neck patients.

The work in this chapter resulted in the following outputs:

Ying Zhang, Stacey McGowan Holloway, Megan Zoë Wilson, Jailan Alshaikhi, Wenyong Tan, Gary Royle and Esther Bär. "DIR-based models to predict weekly anatomical changes in head and neck cancer proton therapy". Physics in Medicine & Biology 67, no. 9 (2022): 095001.

Ying Zhang, Jailan Alshaikhi, Richard A. Amos, Wenyong Tan, Gary Royle, Esther Bär. Systematic progression changes can assist robust IMPT plan selection for head and neck patients. European Society for Radiotherapy and Oncology 2022.

Contribution of Authors: Jailan Alshikhi created the proton plans. Stacey McGowan Holloway and Megan Zoë Wilson helped with the idea of building the AM. Wenyong Tan provided the clinical data and valuable input. Gary Royle provided valuable input concerning the clinical importance and impact of this work. Esther Bär supervised the project. I developed the idea of RIM and model evaluation, including developing codes and tools for data processing and the final analysis.

5.1 Background

Robust optimisation can help to mitigate uncertainty in treatment delivery. However, robust optimisation alone may not be adequate to account for changes during treatment [34]. Mitigating the dose discrepancy of progressive anatomical changes still needs the involvement of plan adaptation. In offline replanning, patients must either continue treatment with an existing sub-optimal plan or face interruptions to treatment. This reactive approach of adaptive therapy triggered by clinical criteria poses workflow challenges for busy clinical practice. While online adaptive proton therapy is considered superior to offline adaptation because the online adapted plans are intended for same-day application, a predictive model can be exploited to prepare an adaptive plan in advance and without the need for real-time QA, which is one of the most challenging aspects of online adaptation.

The individual model aims to provide accurate progressive information for individuality. The produced predicted images can be used for the prospective replanning. In previously proposed models, Kranen *et al*. (2013) [28] did not consider any progressive anatomical changes between fractions. Chetvertkov *et al*. (2016) [54] required at least half of the total fraction to provide a reasonable estimation for progressive changes.

To build an anatomical model that can accurately and promptly provide predicted images for prospective replanning, the objectives of this study are 1) To develop an average model (AM) based on population data to predict the weekly systematic progression changes before treatment. 2) To refine the prediction by adding the patient-specific progressive information from the data acquired during the course of treatment, as the refined individual model (RIM). 3) To evaluate the models using HU differences, contours, proton spot location deviations and IMPT dose distributions.

5.2 Method and material

5.2.1 Patient data

Twenty NPC patients were recruited prospectively. Each patient underwent a planning CT (pCT) and a weekly repeat CT (rCT_t), where t (t=0, 1, 2, 3, ...) represents the week of CT scanning. Contours in the pCT and weekly verification CTs were manually delineated by an oncologist. Five patients were held separately as a test set, and the model was built using the remaining 15 patients.

For all 5 test patients, an original (nominal) IMPT treatment plan with three beam fields (60°, 180°, 300°) was generated using Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). All plans generated throughout this study were robustly optimised with ±3 mm setup and ±3.5% range uncertainty for CTVs and critical OARs. A RBE of 1.1 for proton beams was used. The dosimetric goals for all plans in this study are summarised in table 5.1. A plan was deemed acceptable if the goals set for the CTV and serial organs were fulfilled in the nominal scenario (the error-free distribution) as well as all 12 dose distributions (3 mm orthogonal shifts combined with the ±3.5% range error) in a robust evaluation. More clinical characteristics of the patients can be found in the papers of Tan *et al.* [18, 79].

Table 5.1: Dosimetric goals of the treatment plans created in this study

Structure	Goal under uncertainty
High-risk-CTV	$D_{95} > 95\%$ of prescription dose (72.6 Gy, 33 fractions)
Low-risk-CTV	$D_{95} > 95\%$ of prescription dose (63 Gy, 33 fractions)
CTV	$D_2 < 107\%$ of prescription dose
Spinal cord	$D_{max} < 45$ Gy
Brainstem	$D_{max} < 55$ Gy
Chiasm	$D_{max} < 55$ Gy
Structure	Goal in Nominal
Parotid glands	$D_{mean} < 26$ Gy
Oral cavity	$D_{mean} < 40$ Gy
Larynx	$D_{mean} < 40$ Gy
Proton planning information: MFO planning; spot spacing size: 5mm; energy range:70 MeV – 250MeV; range shifter: 5cm; dose calculation algorithm: Piencel beam scanning (PBS); optimisation algorithm: Nonlinear Universal Proton Optimiser.	

5.2.2 Anatomical models

The models were built at each weekly time point, considering the time dependence of the progressive changes. The AM used the average deformation of each week for prediction. The RIM further refined the prediction of the AM by adding the deformation difference at the early treatment between the actual deformation acquired during the treatment and the average deformation. This deformation difference represents the progressive difference between individuals.

5.2.2.1 Average model

The first model implemented here was the AM. The weekly SVFs between the pCT and the rCTs of the training data in the atlas were used as input. The produced predicted CTs presented systematic progression changes during the course of treatment. The procedure was divided into three steps and repeated for each treatment week t.

1. The SVF of the AM for week t in the atlas space was calculated as the expectation value E of the deformation $\boldsymbol{v}_{a,p\to t}$ of the training dataset

$$\boldsymbol{v}^{\mathrm{AM}}_{a,p\to t} = \mathbf{E}(\boldsymbol{v}_{a,p\to t}) = \frac{1}{N_p}\sum_{pi}\boldsymbol{v}^{pi}_{a,p\to t}, \tag{5.1}$$

   where N_p is the number of patients used in this model, and pi is the patient index.

2. The deformation $\boldsymbol{v}^{\mathrm{AM}}_{a,p\to t}$ was transformed into the space of an individual patient using

$$\boldsymbol{v}^{\mathrm{AM}}_{p\to t} = \boldsymbol{v}^{-1}_{a\to p} \circ \boldsymbol{v}^{\mathrm{AM}}_{a,p\to t} \circ \boldsymbol{v}_{a\to p}. \tag{5.2}$$

3. The predicted patient-specific deformation $\boldsymbol{v}^{\mathrm{AM}}_{t\to p}$ was used for warping the pCT to generate the predicted anatomy. It can be simply obtained by reversing the SVF $\boldsymbol{v}^{\mathrm{AM}}_{p\to t}$ using

$$\boldsymbol{v}^{\mathrm{AM}}_{t\to p} = -\boldsymbol{v}^{\mathrm{AM}}_{p\to t}. \tag{5.3}$$

The predicted CT of the AM at treatment week t ($\mathrm{CT}_t^{\mathrm{AM}}$) can be acquired using

$$\boldsymbol{\phi}_{t\to p}^{\mathrm{AM}} = \exp(\boldsymbol{v}_{t\to p}^{\mathrm{AM}}), \tag{5.4}$$

$$\mathrm{CT}_t^{\mathrm{AM}} = \boldsymbol{\phi}_{t\to p}^{\mathrm{AM}}(\mathrm{pCT}). \tag{5.5}$$

The AM only considered systematic deformations. Random deformations (progressive variations between patients) can be included by adding individual random deformations using newly acquired weekly CTs of the individual patient during the treatment to gradually refine the prediction of the following weeks, leading to the RIM, as described in section 5.2.2.2.

5.2.2.2 Refined Individual Model

In this section, the RIM, which is based on the AM but includes the individual random deformations of a patient, was proposed to further improve the prediction. This study assumed that patients share the basic deformation trend during the treatment (AM), e.g., the progressive changes are rapid at the early treatment and then slow down, but with an individual baseline. This baseline as a constant can be corrected in the RIM using the deformation difference between the actual deformation of the patient acquired during the early treatment and the average deformation of the AM and applied to the prediction of the remaining treatment course. Hence, the RIM assumes that if the shrinkage of the parotid for one patient is visibly more severe compared to the average at fraction i, then the parotid shrinkage of the following fractions is more severe than the average with the same magnitude.

To build the RIM, the AM was applied to the patient's pCT first. The procedure to refine the prediction is as follows:

1. The accurate deformation between the pCT and rCT_t during the early treatment was captured, referred to as $\boldsymbol{v}_{t\to p}$. The update started from week 2 because the progressive changes in the first week were less significant[74, 75, 76].

2. The individual random deformation $\boldsymbol{v}^{\text{ind}}_{t+i\to p}$ for the remaining fractions can be obtained by

$$\boldsymbol{v}^{\text{ind}}_{t+i\to p} = \boldsymbol{v}_{t\to p} - \boldsymbol{v}^{\text{AM}}_{t\to p} \ , i = 1...(n-t), \tag{5.6}$$

where n is the total number of treatment weeks.

3. The deformation field $\boldsymbol{v}^{\text{RIM}}_{t+i\to p}$ for the following fractions as predicted by the RIM can be calculated as

$$\boldsymbol{v}^{\text{RIM}}_{t+i\to p} = \boldsymbol{v}^{\text{AM}}_{t+i\to p} + \boldsymbol{v}^{\text{ind}}_{t+i\to p} \ \ i = 1...(n-t). \tag{5.7}$$

When treatment starts, individual data can be obtained and used in the RIM to gradually update the predicted anatomy. In clinical practice, most H&N plan adaptations occur around the 3rd or 4th week of treatment. This study picked $t = 2, 3$ as examples. When t = 2, the model was referred to as RIM_2. When t = 3, the model was referred to as RIM_3.

5.2.3 Model evaluation

5.2.3.1 Model evaluation based on CT numbers

To assess the anatomical models, the difference image between the predicted image and the corresponding rCT was quantified using the AAHUD within the patient's outline.

5.2.3.2 Model evaluation based on contours

The contours in the predicted images are the propagated contours obtained by applying the deformations of the models to the contours in the pCT. The contour differences between the predicted contours and the manually delineated contours in rCTs (gold standard) were quantified using the 3-dimensional MSD for each week. The contours included in the evaluation were low-risk CTV, high-risk CTV and parotid glands. These structures commonly change their shape and volume during treatment.

5.2.3.3 Model evaluation based on weighted spot location deviation.

The WSLD was applied to evaluate 1) the influence of the systematic anatomical progression on the spot position and 2) the residual anatomical uncertainty.

- The systematic progression uncertainty simulated by the AM. The AM captured the systematic progressive changes of a patient cohort. Therefore, the WSLD estimated by the AM showed the consequence of the systematic progressive changes in the training patient cohort. pCT was used as a reference in equation 2.6.

- The residual anatomical uncertainty from models. The difference between the estimated anatomical uncertainty from the models and the actual anatomical uncertainty at each week was used to evaluate the accuracy of the models. $\Delta\text{WSLD}_t^{\text{res}}$ was calculated using equation 5.8,

$$\Delta\text{WSLD}_t^{\text{res}} = \text{WSLD}_t^{\text{real}} - \text{WSLD}_t^{\text{model}}, \tag{5.8}$$

  where $\text{WSLD}_t^{\text{real}}$ is the actual anatomical uncertainty calculated by the WSLD between rCT_t and the pCT, which is also corresponding to the residual anatomical uncertainty of no model. $\text{WSLD}^{\text{model}}$ is the anatomical uncertainty estimated by a model. The best model should approach a $\Delta\text{WSLD}_t^{\text{res}}$ of 0 for each treatment week t.

5.2.3.4 Model evaluation based on dose distribution.

The nominal plan was recalculated on the rCTs and the predicted weekly CTs. The gamma index was used to evaluate the dose difference between the dose distribution on a rCT and the corresponding predicted weekly CT [68]. A relatively stringent criterion of 2 mm/2% (a stringent criterion can lead to a smaller passing rate) and the acceptable passing rate of 95% were used in this study because they are the paired parameters generally used [85, 86]

5.3 Results

5.3.1 Anatomical model evaluation based on Hounsfield Units

In this section, the image difference on HU between rCT_6 and the corresponding predicted CT_6 is compared from the 5 test patients. For visual assessment, figure 5.1 shows a slice of the image difference of a test patient.

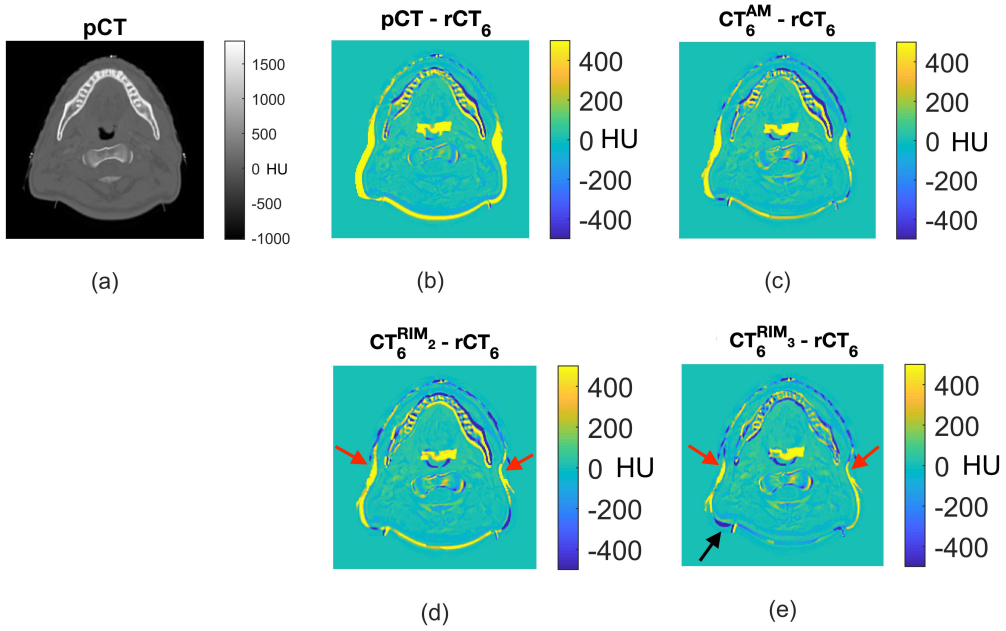


Figure 5.1: Comparison between different anatomical models using a representative example patient. (a) shows a slice from the pCT of a patient in the test dataset. (b) shows the difference image between the pCT and rCT_6, without the application of any anatomical model. (c) is the difference image between the predicted CT from the AM and rCT_6. (d) is the difference image between the predicted CT from the RIM_2 model and rCT_6. (e) is the difference image between the predicted CT from the RIM_3 model and rCT_6.

The shrinkage from the pCT to rCT_6 is indicated by the yellow area in figure 5.1 2b). This shrinkage leads to protons travelling further and causes a dose discrepancy as a result. From visual assessment, with the AM, the yellow area is reduced in 2c). The RIM predicted more accurately the anatomical changes of this patient in the area pointed by the red arrows. The refinement from the RIM_3 further reduced this difference but overestimated the posterior shrinkage, indicated by the black arrow.

The weekly AAHUD (no model, AM, RIM_2, and RIM_3) over all test patients with approximately 8 million voxels in total and a special case with approximately 2 million voxels are analyzed and shown in figure 5.2a) and b), respectively. Because the deformation of the 2nd or the 3rd week was used to refine the model, the AAHUD of the RIM_2 is shown from week 3 to week 6, and the AAHUD of the RIM_3 is shown from week 4 to week 6.

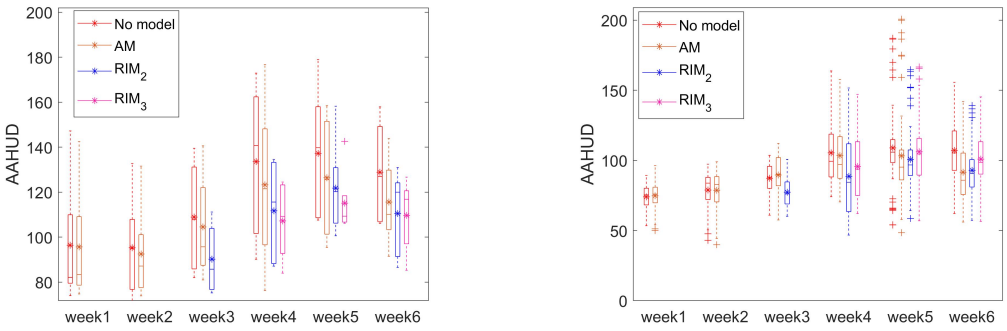


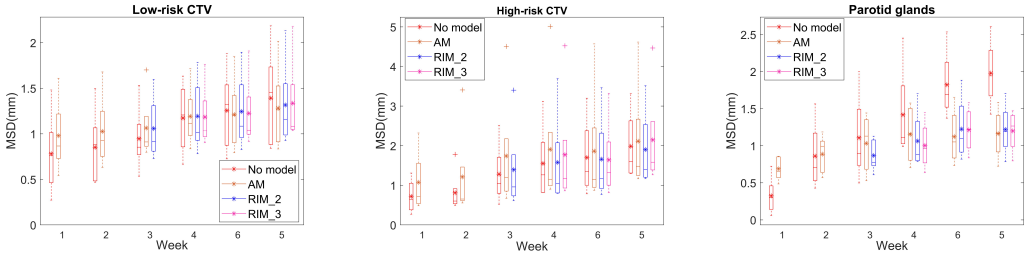
Figure 5.2: Boxplot of the AAHUD analysis: a) shows the average AAHUD from the 5 test cases; b) the AAHUD from a special case. The range shows the AAHUD of different image slices. The horizontal lines in the box plot indicate the median value, and the asterisks indicate the mean value.

In the special case (figure 5.2b)), no improvement was observed from the RIM_3 compared to the RIM_2, with only small HU differences between the two models. On average, compared to no model, the AM, the RIM_2, and the RIM_3 reduced the AAHUD by 13.6 HU, 18.4 HU, and 19.2 HU, respectively at week 6. The RIM_3 captured more characteristics of the individual anatomical changes and had a higher predictive ability than the RIM_2.

5.3.2 Model evaluation based on contours.

The weekly MSD between the predicted contours of the models and the corresponding contours in the rCT_t are shown for high-risk CTV, low-risk CTV and parotid glands in figure 5.3. When the predicted contours from the models were used, the average MSD of the parotid glands can be reduced from 1.98 mm (no model) to 1.16 mm (AM), 1.21 mm (RIM_2) and 1.19 mm (RIM_3) at week 6. This study also found a very slight improvement in low-risk CTV. The average MSD of low-risk CTV can

be reduced from 1.39 mm (no model) to 1.28 mm (AM) and 1.34 mm(RIM_3) at week 6. No improvement was found in high-risk CTV. Although the AM is slightly better than the RIM on the contour-based evaluation, the advantage is not significant (<0.07 mm).



(a) **(b)** **(c)**

Figure 5.3: The weekly MSD between the predicted contours of the models and the corresponding contours in the rCT_t for high-risk CTV, low-risk CTV and parotid glands. In the box plot, the horizontal lines indicate the median value, and the asterisks indicate the mean value.

5.3.3 Model evaluation based on weighted spot location deviation.

In this section, the range differences were estimated using the spot location of the treatment plans from the 5 test patients.

The WSLD originating from the systematic progression uncertainty estimated from the AM is shown in figure 5.4. The uncertainty from the systematic progression steadily increased from 0.64±0.05 mm at week 1 to 2.07±0.20 mm at week 6.

The average WSLD from the residual anatomical uncertainties from the models and the corresponding 95% CI were compared in figure 5.5. When the uncertainty estimated from the predicted images of the models was considered, the residual anatomical uncertainty was reduced from 4.47±1.23 mm (no model) to 1.89±0.96 mm (RIM_3), 2.24±1.13 mm (RIM_2), and 2.41±1.12 mm (AM) at week 6, achieving significant improvements as compared to no model.

The comparison of individual cases regarding the residual anatomical uncertainty between the four models (including no model) is listed in appendix C. The

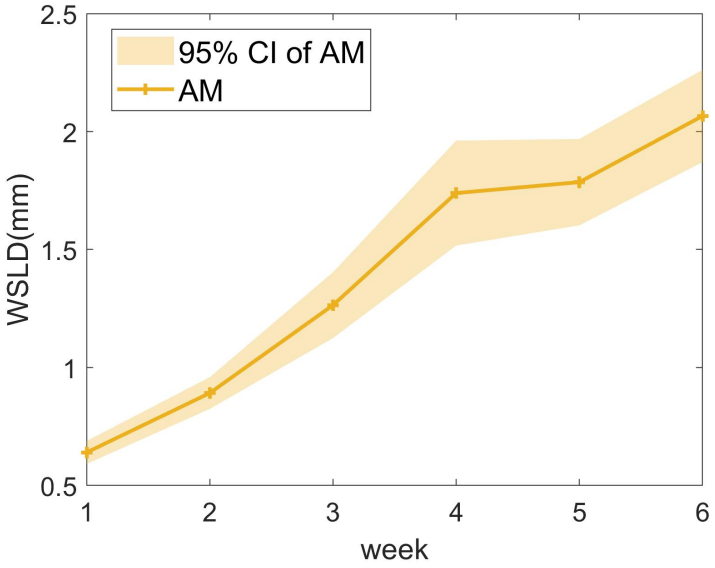


Figure 5.4: The systematic progression uncertainty estimated from the AM for each week. The result is estimated in average WSLD with 95% CI across the 5 test cases.

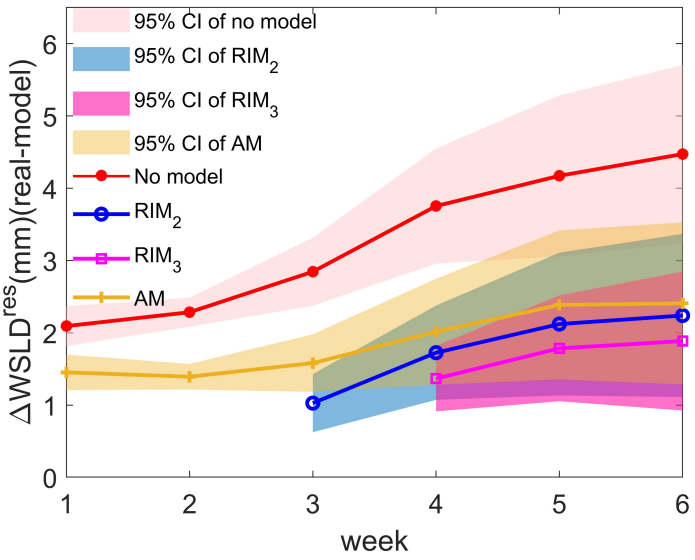


Figure 5.5: The residual anatomical uncertainty in WSLD. The residual anatomical uncertainty from the AM, the RIM_2 and the RIM_3 are compared. The graph shows the average difference with 95% CI between the estimated WSLD from the models and the actual WSLD across the 5 test dataset.

uncertainty of the DIR is also listed as a reference.

A summary of model uncertainty based on WSLD is listed in table 5.2.

Table 5.2: Summary of model evaluation based on WSLD over all test patients.

Week	Uncertainty (Mean±95%CI) (mm)				
	Model (Residual anatomical uncertainty)				DIR
	(No model)	AM	RIM_2	RIM_3	
1	**2.09±0.28**	1.45±0.24	-	-	0.86±0.14
2	**2.29±0.20**	1.39±0.18	-	-	0.87±0.16
3	**2.85±0.47**	1.58±0.40	1.03±0.38	-	0.90±0.20
4	**3.75±0.80**	2.01±0.73	1.73±0.65	1.37±0.45	1.04±0.13
5	**4.17±1.11**	2.39±1.03	2.12±0.99	1.79±0.73	1.18±0.30
6	**4.47±1.23**	2.41±1.12	2.24±1.13	1.89±0.96	1.33±0.48

5.3.4 Model evaluation based on dose distribution.

Figure 5.6 shows the DVH curves for the dose distribution of a test patient (figure 5.1) at week 6 from the rCT_6, the AM, the RIM_2, the RIM_3 and no model (using the pCT). The DVH of the RIM_3 was the closest to the DVH of the rCT_6. The worst performance in the OARs was observed without using a model.

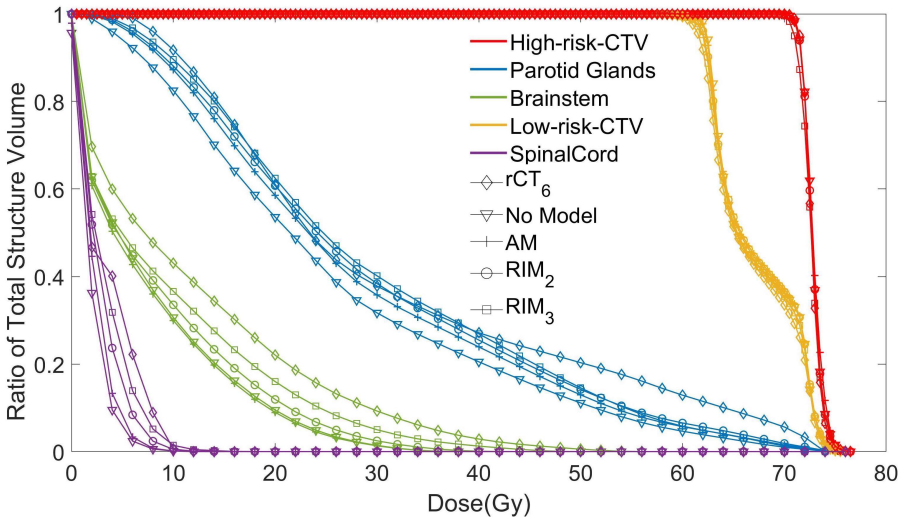


Figure 5.6: DVH curves for the dose distribution of a test patient (figure 5.1) at week 6 from the rCT_6, the AM, the RIM_2, the RIM_3 and no model (using the pCT).

The results of the gamma analysis between the dose distribution on rCTs and the corresponding predicted weekly CTs from no model (predicted images were replaced by the pCT), the AM, the RIM_2 and the RIM_3 were listed for each patient in table 5.3. The number of cases of which the gamma index $< 95\%$ was reduced from 9 (no model) to 6 (AM), to 4 (RIM_2) and 2 (RIM_3). The average gamma index among 5 test patients was improved from 93.87±2.48 % (no model) to 96.16±1.84

% (RIM_3) at week 6.

Table 5.3: The gamma index between the dose distribution on rCTs and on the corresponding predicted weekly CTs from no model (predicted images were replaced by planning CT), the AM, the RIM_2 and the RIM_3 for each test patient and each week. The gamma indexes below 95%, the generally accepted standard passing rate, are highlighted in bold. The gamma indexes of week 1 and week 2 are only listed for no model and the AM as the RIM models updated from week 3.

Id	week 1 (%)		week 3 (%)			week 5 (%)			
	No model	AM	No model	AM	RIM_2	No model	AM	RIM_2	RIM_3
1	98.1	98.1	**93.2**	**94.1**	95.3	**91.2**	**93.4**	**94.6**	96.1
2	99.3	98.8	99.1	98.7	99.2	98.0	98.2	98.6	98.2
3	96.4	97.2	96.1	96.4	96.2	**91.4**	**91.8**	**91.6**	**91.5**
4	97.7	97.8	95.8	95.2	96.5	96.6	96.8	96.7	96.9
5	98.5	98.1	96.6	97.3	98.0	**93.4**	95.6	96.0	96.4
mean	98.01	98.01	96.16	96.33	97.06	**94.09**	95.16	95.50	95.81
CI	0.85	0.47	1.71	1.44	1.23	2.45	2.07	2.12	2.06

Id	week 2 (%)		week 4 (%)				week 6 (%)			
	No model	AM	No model	AM	RIM_2	RIM_3	No model	AM	RIM_2	RIM_3
1	95.3	95.6	**93.4**	95.5	96.6	97.5	**91.2**	**93.9**	**94.6**	97.5
2	99.3	99.2	98.2	97.9	98.4	98.4	97.3	97.9	98.2	97.3
3	95.0	96.7	95.5	96.9	94.8	96.0	**91.5**	**92.4**	**91.4**	**92.2**
4	97.3	97.0	95.7	94.8	96.3	96.6	97.1	96.7	97.5	97.7
5	98.2	98.2	**94.1**	96.7	97.5	97.3	**92.2**	**94.9**	95.6	96.0
mean	97.03	97.34	95.37	96.39	96.72	97.16	**93.87**	95.18	95.48	96.16
CI	1.49	1.14	1.47	0.98	1.09	0.72	2.48	1.76	2.15	1.84

5.4 Discussion

In this chapter, two different anatomical models were developed and compared. The AM is a basic model only used to evaluate systematic anatomical uncertainty. The RIM is a further refinement based on the AM, with the suggested use in offline adaptive treatment planning. The model accuracy was evaluated based on AAHUD, contours, WSLD and dose distributions.

5.4.1 Model evaluation

From the AAHUD comparison, on average, the RIM predicted the anatomical changes with the highest accuracy compared to the AM or no model. This observed outcome was expected because the deformation differences include the progressive variations between patients. It is important to note that small non-rigid variations (sNRVs), such as jaw movement and shoulder position changes, will also be included in the deformation differences to update the model. If the magnitude of the

sNRVs was greater than the progressive variations, the RIM can be inferior to the AM, as shown in figure 5.2b). The particular patient shown in this case was very slim at the start of treatment and had <5% weight loss. The sNRVs can be more predominant, making the RIM less effective. Nevertheless, the observed differences between the AM and the RIM were small.

The model evaluation based on contours showed that the models are more effective in predicting the changes of parotid glands. Because patients were not stratified based on CTV features, and CTV location and size are diverse in this dataset, predicting the changes of the CTVs is challenging. This contour-based model evaluation assumes that the contours are perfect on all CTs. In this study, the inter-observer variation was eliminated as a single physician contoured the organs, and the intra-observer variability was minimized by using a copy-and-modification strategy[18]. In reality, intra-rater variability exists and can lead to an increase in the evaluation metrics.

As shown in figure 5.4, the effect of anatomical progressions in the first week of treatment was not significant, justifying the approach of refining the model from week 2 onward. The RIM_3 can reduce the anatomical uncertainty from 4.47 ± 1.23 mm (no model) to 1.89 ± 0.96 mm at week 6 (see figure 5.5).

In table 5.3, the average gamma indexes throughout the 5 test patients between the dose distribution on rCTs and the corresponding predicted weekly CTs from the AM and the RIM are all above 95% for each week, which is the standard passing rate generally accepted [69, 70]. Also, the average gamma index was improved from $95.18\pm1.76\%$ (AM) to 96.16 ± 1.84 (RIM_3) at week 6. Combining figure 5.2 and 5.5, this study demonstrated that the RIM can be gradually refined during the treatment and can potentially serve as a routine monitor to update the prediction and prepare for adaptive intervention if necessary.

In comparing individual gamma values in table 5.3 and individual WSLD in table C3 in appendix C, there is a high level of consistency of 80% in terms of selecting the best prediction, validating the feasibility of using WSLD as an evaluation tool.

5.4.2 Study limitations

The models were built based on a relatively small sample data set of 15 patients, and analyses were performed on 5 patients. The weight loss of these 5 patients ranged from 4% to 18% (The weight loss of the training dataset ranged from 2% to 12%), including patients with small and severe anatomical changes. Exploiting the current dataset, the sensitivity of the AM towards the training data is measured based on repeatedly sampling random training data for the same patient in the Appendix C.2. The 95% CI of different measurements is less than 0.12 mm. However, it still cannot totally remove the concerns of over-fitting when the model is directly applied to another dataset. Finding the optimal parameters to build models for patients with CBCT data is underway. It will enable a relatively large dataset to be included for modelling, thus avoiding the risk of over-fitting. Further validation of the model and the estimation of its sensitivity thereof will be conducted on a larger cohort of patients. When a large dataset is available, patient stratification can be used to improve the model's accuracy. The model can be built based on a cohort of patients with the same characteristics and applied to the same type of patients. The features that might be related to the anatomical changes have been revealed [87, 88, 89, 90, 91]. Assuming the model is built based on a large dataset with patient stratification, the accuracy of the model should be only limited by the DIR uncertainty and sNRVs uncertainty.

Another limitation of the presented study is that the patients used to build and evaluate the models have received photon therapy. This study assumed that patients undergoing proton therapy would have similar anatomical changes as those who received photon therapy. Further validation of the model will be conducted on a cohort of patients treated with IMPT.

5.5 Conclusion

This study has presented and analysed two anatomical models for H&N patients. The AM captures the systematic progressive changes during the treatment. The RIM is based on the AM but then updated using the patient's progression during

the treatment. Based on the evaluation of CT numbers, contours, range and dose distribution, this study concluded that the RIM gradually refined the prediction of anatomical changes. The exploration of the potential clinical application, using the RIM to prepare offline adaptive plans in advance, is demonstrated in Chapter 6.

In the contour-based evaluation, the average MSD of the AM was slightly better than that of the RIM. However, the advantage was not significant. Considering that parotid glands and low-risk CTV are closely related to outline changes, the predicted contours of these two organs were still chosen for the application of the RIM for prospective replanning in Chapter 6.

Chapter 6

Application of refined individual model in offline adaptive therapy

This chapter demonstrated the concept of using the refined individual model (RIM) to prepare an adaptive plan in advance. The work in this chapter resulted in the following outputs:

Ying Zhang, Jailan Alshaikhi, Richard Amos, Wenyong Tan, Gary Royle, Esther Bär. First Application of Predictive Model to Assist Adaptive Proton Therapy. American Society for Radiation Oncology 2021.

Ying Zhang, Jailan Alshaikhi, Richard A. Amos, Matthew Lowe, Wenyong Tan, Esther Bär and Gary Royle. "Improving workflow for adaptive proton therapy with predictive anatomical modelling: A proof of concept." Radiotherapy and Oncology, 173 (2022): 93-101.

Contribution of Authors: Jailan Alshikhi created the proton plans. Wenyong Tan provided the clinical data and valuable input. Richard A. Amos and Matthew Lowe provided valuable input concerning the clinical importance and impact of this work. Esther Bär and Gary Royle supervised and guided the project. All other work was performed by myself.

6.1 Background

Anatomical models for H&N patients have been proposed in previous studies [28, 63, 60]. Yu *et al.* (2016) [60] exploited the model to generate artificial defor-

mations for DIR evaluation. Kranen *et al.* (2013) [28] generated CTs to represent the systematic changes of previous fractions for plan adaptation. However, they only assessed their model based on the misalignment of the anatomical landmarks. The clinical application of anatomical models has not yet been fully explored and evaluated.

The current offline adaptation technique is a reactive approach leading to an unpredictable workload for treatment planning staff, the medical physics team who perform patient-specific plan QA, and radiation oncologists who review and approve the plans. Also, patients must either continue treatment with an existing sub-optimal plan or face interruptions to treatment whilst plans are adapted. Adaptive plans that can be prepared in advance would be beneficial to the clinical workflow: 1) The adapted replan can be delivered as soon as it is needed due to the ability to perform patient-specific QA/verification before the adaptation is required, for example, on a predicted CT, which triggered a replan. 2) For the patient, there is no gap in treatment or the delivery of a few sub-optimal fractions while the replan is calculated, approved, and verified through QA. 3) For the workflow, the option to prepare adaptive plans in advance allows for easier scheduling of patient-specific QA around machine QA, maintenance, and other demands for beam time.

This chapter explores the use of the refined individual model (RIM) to generate adaptive proton therapy plans in advance. Three different strategies are compared to find a strategy that can maximise the use of the RIM, benchmarked against a standard reactive clinical replanning technique.

6.2 Materials and methods

6.2.1 Refined individual model

In this chapter, the leave-one-out cross-validation strategy was applied to 20 NPC patients who had a planning CT and weekly CTs during the treatment to build the RIM. To predict a deformation for the remaining patient, the average deformation of the training population (n=19) was applied to the patient's planning CT. The RIM was then updated based on the patient's progression during treatment. This process

was repeated for 10 randomly selected patients. It still follows that each validation patient is independent from the training population used to create the average model.

6.2.2 Patient data

Ten validation patients were included in this study. Each patient had a planning CT (pCT), weekly verification CTs and predicted weekly CTs. Contours in the planning CT and weekly CTs were manually delineated by an oncologist. None of the IMPT plans presented in this study was applied during the clinical radiotherapy treatment of these patients. Instead, this is a retrospective study using the patients' imaging data. As tumour location and size are diverse in this dataset, predicting the change to the high-risk CTV (tumour) is challenging. The RIM is most effective in predicting the patient outline and parotid gland positions. Hence, for all OAR contours and the low-risk CTVs (nodal area) affected by neck changes, the predicted contours from contour propagation were used. For the high-risk CTV, the initial CTV of the planning CT was used in the predicted plan to ensure target coverage. In this study, plan adaptation was triggered with the aim of protecting the parotid glands, following the TORPEdO trial (A phase III trial of proton therapy versus intensity-modulated radiotherapy for multi-toxicity reduction in oropharyngeal cancer; CRUK/18/010) [53]. When the difference (between the original plan calculated on the planning CT and a weekly verification CT) of D_{mean} to both parotid glands was larger than 3 Gy [54], a replan was created. Detailed clinical information of this cohort of patients can be found in [18].

For all 10 validation patients, an original (nominal) IMPT treatment plan was generated using Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). All plans generated throughout this study were robustly optimised with ± 3 mm setup and $\pm 3.5\%$ range uncertainty for CTVs and critical OARs. A RBE of 1.1 for proton beams was used. The dosimetric goals and priorities for all plans in this study are summarised in table I [92, 93, 94, 5]. Further details can be found in Appendix D.1.

The nominal plan was recalculated on the weekly verification CTs and evaluated to identify the need for adaptation. Adaptation was required for 9 out of 10

patients (exception: patient ID 1 referred to in later tables), and adapted plans were generated using the CTs that triggered the replan.

A plan was deemed acceptable if the goals set for the CTV and serial organs were fulfilled in the nominal scenario (the error-free distribution) as well as all 12 dose distributions (3mm orthogonal shifts combined with the $\pm 3.5\%$ range error) in a robust evaluation. For dose distributions calculated on weekly CT images, the DIR algorithm of Niftyreg was used to accumulate the dose in the reference frame of the planning CT.

Table 6.1: Dosimetric goals of the treatment plans created in this study

Structure	Goal under uncertainty	Priority
High-risk-CTV	$D_{95} > 95\%$ of prescription dose (72.6 Gy, 33 fractions)	2
Low-risk-CTV	$D_{95} > 95\%$ of prescription dose (63 Gy, 33 fractions)	2
CTV	$D_2 < 107\%$ of prescription dose	2
Spinal cord	$D_{max} < 45$ Gy	1
Brainstem	$D_{max} < 55$ Gy	1
Chiasm	$D_{max} < 55$ Gy	1
Structure	Goal in Nominal	
Parotid glands	$D_{mean} < 26$ Gy	3
Oral cavity	$D_{mean} < 40$ Gy	3
Larynx	$D_{mean} < 40$ Gy	3

6.2.3 Adaptive planning using the refined individual model

Three different adaptation strategies were proposed to use the predicted images from the RIM to create adapted proton plans. These strategies enable adaptive plans to be created in advance of necessity, streamlining the clinical workflow and facilitating QA.

6.2.3.1 Scheduled predicted plan on scheduled week

The first adaptation strategy consists of generating two adapted plans after the RIM has been updated at week 2 of treatment. This study chose to use two plan adaptations following the paper of Wu *et al.* (2009) [95], which demonstrated very limited gains from increasing the replanning frequency from 2 to 6. The predicted CTs of week 3 (PD_3) and 5 (PD_5) were used to create two predicted plans. The plan created on PD_3 was intended for the treatment of week 3 and week 4, whereas the plan created on PD_5 was intended to be applied in week 5 and week 6. For verification, the

predicted replans can be recalculated on the corresponding weekly CTs before delivery, ensuring the applicability and suitability of the plan. This adaptation strategy is referred to as SPS adaptation.

6.2.3.2 Reactive predicted adaptation strategies

Reactive predicted adaptation strategies included 1) reactive predicted plan on a flexible week and 2) reactive predicted plan on a scheduled week.

- Reactive predicted plan on a flexible week. The second adaptation strategy comprised one plan adaptation. For that strategy, the nominal treatment plan was recalculated on each of the weekly predicted images, and the resulting dose distribution was assessed. The prospective replan was created in the predicted image PD_r where the recalculated dose distribution met the conditions required to trigger plan adaptation. The adapted plan, referred to as the reactive predicted plan, was then applied as soon as the verification CT collected during treatment triggered plan adaptation. Note that during this strategy, the adapted plan is not necessarily applied in the week that predicted adaptation but rather applied flexibly whenever an adaptation is triggered in a verification CT. For verification purposes, the predicted replan can be recalculated to the verification CT to confirm if the plan satisfies the dosimetric goals. This adaptation strategy is referred to as RPF adaptation.

  Because plan adaptation was triggered to protect the parotid glands, the overlap of the predicted contours (used in the predicted plan adaptation strategy) and the real contours (used in the standard replanning technique, mentioned in 6.2.3.3) for the parotid glands was measured by the Dice similarity coefficient (DSC) and compared with no model, where the contours in the planning CT replaced the predicted contours.

$$\mathrm{DSC_{PG}} = \frac{|PG_p \cap PG_r|}{|PG_r| + |PG_p|}, \tag{6.1}$$

  where PG_p and PG_r represent the binary masks of the parotid glands in a predicted CT and the corresponding real CT, respectively.

Although the predicted replan was applied flexibly in this strategy, the predicted replan week can be compared with the actual replan week to evaluate the predictive ability of the RIM.

- Reactive predicted plan on a scheduled week. Brown *et al.* (2016) [96] and Wu *et al.* (2008) [97] showed that plan adaptations occur around the 3rd or 4th week of treatment for H&N patients. To lighten the workload of weekly dose monitoring, one adaptation to the fixed week can be used as an alternative option. The TORPEdO trial[53], conducted by the Christie NHS Foundation Trust in Manchester, UK, was designed to repeat the planning CT scan during week 3 to evaluate the need for replanning. Therefore, this strategy applied the reactive predicted plan from week 3. For verification purposes, the plan can be recalculated to the verification CTs to confirm if the plan satisfies the dosimetric goals. This adaptation strategy is referred to as RPS adaptation.

6.2.3.3 Standard replan

For comparison, the standard adaptation plan was optimised on the verification CT which triggered a replan and applied to the treatment of the following week, representing a delay of 5 fractions before implementing the replan.

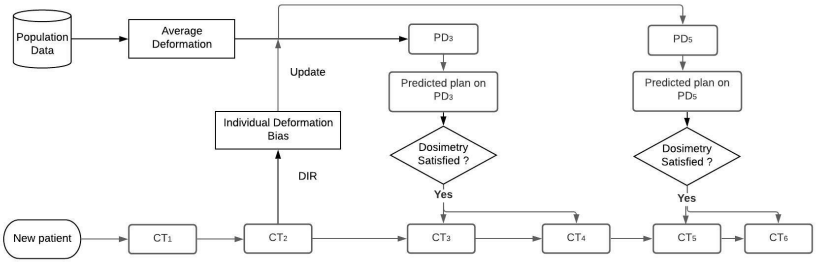
The workflows of the adaptation strategies mentioned above are shown in figure 6.1.

The dosimetric details of all plans generated for this study are summarised in table 6.2.

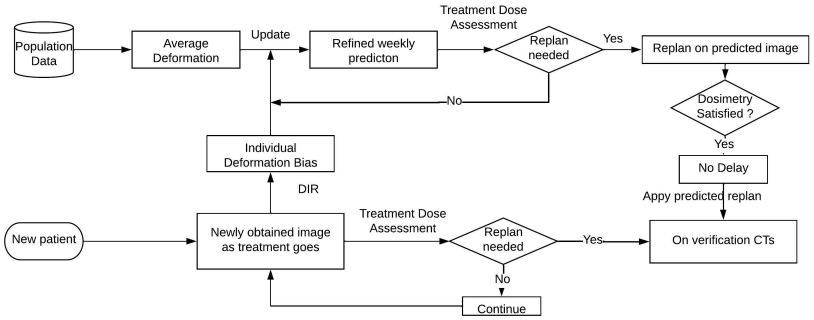
6.2.4 Plan evaluation using accumulated dose metrics

For the standard replanning technique and the three proposed strategies using predicted anatomy, the dose distributions on the weekly CTs were calculated and deformed to the planning CT to accumulate the dose, allowing an evaluation of the delivered dose to the patient. The accumulated dose using the standard replanning technique is taken as the gold standard.

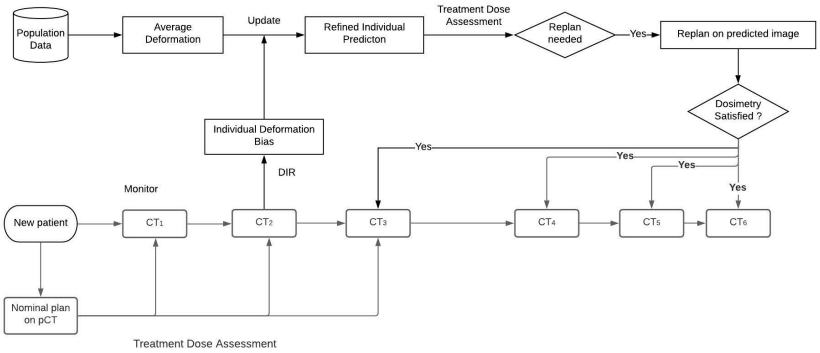
This study first compared the SPS and the RPF using five patient data sets (IDs 1–5). This comparison can reveal if one additional adaptation can improve the



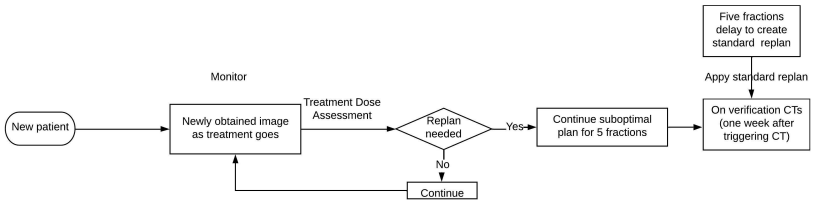
(a) The workflow of the SPS strategy. In the SPS strategy, predicted plans are created on PD_3 and PD_5 and applied to the treatment of week 3/4 and week 5/6 respectively.



(b) The workflow illustrating the RPF strategy. In the RPF strategy, the predicted replan is created on the predicted image which triggers a replan. The dose distributions on the weekly verification CT are monitored all the time during the treatment, and the predicted plan is applied whenever adaptation is required.



(c) The workflow of the RPS strategy. The predicted replan is created on the predicted image which triggers a replan. The reactive predicted plan is applied from week 3.



(d) The workflow of the standard adaptation. During the standard adaptation, the adapted plan is calculated on the CT that triggers a replan and applied in the following week, representing a delay of 5 fractions.

Figure 6.1: The workflows of the four adaptive IMPT strategies.

Table 6.2: Dose metrics of proton plans created for the 10 patient data sets. Does metrics are from the CTs where the plans have been optimised (the planning CT for the nominal plan, the adaptation-triggering repeat CT for the standard replanning and the adaptation-triggering predicted CT for the predicted adaptation). Please note that the patient with ID 1 has no row for the standard replan and flexible replan because no replan was triggered from the verification CT of that patient.

ID	Plan	High-risk CTV $D_{95\%}$(%)	Low-risk CTV $D_{95\%}$(%)	D2	Spinal cord D_{max}(Gy)	Brainstem D_{max}(Gy)	Chiasm D_{max}(Gy)	Parotid D_{mean}(Gy)	Oral Cavity D_{mean}(Gy)	Larynx D_{mean}(Gy)
plans of five patient for best strategy selection.										
1	Nominal plan	98.75	98.86	103.05	30.55	45.69	36.5	27.36	6.8	30.93
	SPS	98.85	99.02	103.24	28	46.15	33.9	28.88	6.7	28.35
		98.8	98.98	103.34	29.43	46.49	34.46	29.18	7.48	28.22
2	Nominal plan	98.28	98.66	102.92	34.49	40.08	22.06	25.53	14.24	26.24
	Standard replan	98.53	98.65	102.33	34.48	37.84	16.4	24.71	12.14	22.84
	RPF/RPS	98.3	98.53	102.97	36.89	37.1	16.36	25.15	14.6	21.55
	SPS	98.3	98.53	102.97	36.89	37.1	16.36	25.15	14.6	21.55
		98.36	98.34	102.58	36.23	37.1	17.38	27.25	14.68	22.6
3	Nominal plan	98.75	98.59	102.24	33.13	39.16	23.88	22.89	10.88	11.77
	Standard replan	98.7	98.54	102.24	33	39.5	23.6	23	10.76	12.28
	RPF/RPS	98.54	98.65	102.15	33	38.5	23.29	22	8.94	9.47
	SPS	98.55	98.67	102.29	28.2	33.3	22.51	20	9.17	10.42
		98.59	98.71	102.44	27.75	34.48	22.62	21.5	9.04	10.59
4	Nominal plan	98.42	98.82	102.96	32.9	34.4	32.02	24.53	7.43	22.18
	Standard replan	98.68	98.79	102.05	15.51	38.59	19.18	22	6.07	14.02
	RPF/RPS	98.57	98.70	102.83	17.46	37.8	18.7	23	6.8	17.6
	SPS	98.57	98.79	102.71	17.6	36.88	19.45	23.9	6.68	16.97
		98.61	98.65	103.14	16.71	37.55	20.54	23	6.7	17.81
5	Nominal plan	98.82	98.67	102.02	34.56	40.75	26.21	23.63	10.65	16.08
	Standard replan	98.87	98.65	101.82	35	40.56	26.4	23.85	10.6	16.28
	RPF/RPS	98.65	98.71	102.27	36.4	42.5	29.3	22.5	12.46	16.33
	SPS	98.73	98.63	102.18	34.7	40.44	26.26	22	12	16.13
		98.71	98.76	102.89	35.43	42.34	27.03	22.5	12.26	16.17
plans of five patient to verify the result of best strategy.										
6	Nominal plan	98.5	98.65	102.7	36	38.93	34.37	26.13	15.38	30
	Standard replan	98.4	98.57	101.3	35	42.38	20	30.1	13.29	22.71
	RPF/RPS	98.5	98.65	102.81	33	42.4	37	27.21	15	29.9
7	Nominal plan	98.6	98.27	102.58	36.53	40	26.86	27.04	13.19	19.64
	Standard replan	98.75	98.68	102.46	28.15	36.5	19.07	25.58	11.1	16.84
	RPF/RPS	98.68	98.55	102.76	34	42.82	26.65	27.4	11.37	17.54
8	Nominal plan	98.37	98.63	102.48	35.47	36.01	23.43	27.24	10.6	19.77
	Standard replan	98.6	98.5	101.97	32.95	37.68	24.24	26.96	9.92	15.43
	RPF/RPS	98.43	99.00	102.33	24.88	38.84	24.63	26.84	10.77	17.06
9	Nominal plan	98.75	98.76	102.31	11.84	42.58	42.63	22.66	8.83	9.2
	Standard replan	98.57	98.37	102.88	11.75	38.46	34.33	22.6	9.86	12.3
	RPF/RPS	98.74	98.79	102.47	8.74	40.17	40.82	22.67	8.3	7.93
10	Nominal	98.62	98.28	102.01	29.2	38.76	36.3	27.64	7.1	21.44
	Standard replan	98.45	98.42	102.05	31.75	42.19	33.57	28.74	8.52	17.33
	RPF/RPS	98.52	98.61	102.04	31.73	39.63	40.27	28	7.73	20.34

accumulated dose distributions and reveal the accuracy of the RIM. Ten patient data sets (IDs 1–10) were then used in the reactive strategies (the RPF and the RPS) and compared to the standard replanning technique for further investigation.

Equation 6.2 was used to calculate the dose metrics difference between the accumulated dose and the planned dose.

$$\Delta \mathrm{Dx} = \mathrm{Dx}_{\mathrm{accu}} - (\mathrm{Dx}_n * F_n + \mathrm{Dx}_r * F_r). \tag{6.2}$$

Dx represents a dose metric, ΔDx is the dose metric difference between the accumulated dose and the planned dose, Dx_{accu} is the dose metric of the accumulated dose, Dx_n is the dose metric of the nominal plan, and F_n is the number of fractions to which the nominal plan is applied. Dx_r is the dose metric of the adapted plan, and F_r is the number of fractions to which the replan is applied. The planned dose is represented by the sum $Dx_n * F_n + Dx_r * F_r$.

The dose metrics of the plans (nominal plan, the standard replan, the SPS, the RPF and the RPS) that are used for plan comparison are the same as in table 6.1. A two-sample t-test was used to determine if there is a significant difference between the distribution of ΔDx in two adaptation strategies, with $p < 0.05$ taken as the significance level.

6.3 Results

6.3.1 Comparison between scheduled predicted plan on scheduled week and reactive predicted plan on flexible week

To compare the SPS and the RPF, both techniques were applied to the CT imaging data of the five patients(IDs, 1–5). Table 6.3 shows the nominal plan and the accumulated doses of the nominal plan with no adaptation, the SPS and the RPF for each patient except patient 1. For patient 1, unlike patients 2-5, the dose recalculation on the weekly verification CTs and predicted CTs did not trigger a replan. Therefore, no plans were generated for the RPF.

This study found that the CTV coverage (D_{95}) remains similar between the SPS, the RPF and the nominal plan, with differences smaller than 1%. For the serial organs, three strategies (no adaptation, the SPS and the RPF) all showed similar accumulated dose metrics, with differences within tolerance (<3 Gy). The largest differences were observed in the accumulated parotid D_{mean} between no adaptation and each adaptive strategy (the SPS and the RPF). Compared to no adaptation, the SPS and the RPF decreased the accumulated D_{mean} to the parotid glands by up to 3.82 Gy and 3.95 Gy, respectively.

In figure 6.2, the SPS and the RPF are compared for the five patients in terms

Table 6.3: Nominal and accumulated dose metrics of the five patients used to compare the SPS and the RPF. Please note that for patient 1, no replan was triggered according to the replan criteria. The dose metrics for the parotid glands are highlighted in bold. The asterisk in plan column indicates the numbers in the same row are the accumulated dose.

ID	Plan	High-risk CTV D_{95} (%)	Low-risk CTV D_{95} (%)	D_2 (%)	Spinal cord D_{max} Gy	Brainstem D_{max} Gy	Chiasm D_{max} Gy	Parotid D_{mean} Gy	Oral Cavity D_{mean} Gy	Larynx D_{mean} Gy
1	Nominal plan	98.75	98.86	103.05	30.55	45.69	36.5	27.36	6.8	30.93
	No Adaptation *	97.93	98.89	103.17	28.56	42.39	27.5	29.61	6.9	31.14
	SPS*	97.82	98.89	102.34	28.31	42.62	28.79	**29.8**	7.04	29.49
2	Nominal plan	98.28	98.66	102.92	34.49	40.08	22.06	25.53	14.24	26.24
	No Adaptation*	96.6	95.71	102.89	33.39	40.63	22.16	31.87	17.45	32.03
	RPF*	96.6	96.35	102.62	35.81	41.27	19.45	**28.92**	17.39	28.43
	SPS*	96.4	96.67	102.62	35.49	40.71	19.58	**28.72**	17.4	28.12
3	Nominal plan	98.75	98.59	102.24	33.13	39.16	23.88	22.89	10.88	11.77
	No Adaptation *	97.97	98.89	102.07	30.59	35.69	24.58	24.53	10.76	12.36
	RPF*	98.05	98.89	102.07	31.34	38.28	23.8	**22.77**	10.48	11.27
	SPS*	98.01	98.89	102.07	30.02	36.69	22.93	**22.98**	10.5	11.77
4	Nominal plan	98.42	98.82	102.96	32.9	34.4	32.02	24.53	7.43	22.18
	No Adaptation*	97.74	97.94	102.62	30.47	29.68	29.29	28.36	8.62	25.09
	RPF*	97.82	96.98	102.07	24.36	33.89	22.04	**25.03**	8.31	21.39
	SPS*	97.86	97.62	102.34	24.24	34.04	22.73	**25.52**	8.19	21.34
5	Nominal plan	98.82	98.67	102.02	34.56	40.75	26.21	23.63	10.65	16.08
	No Adaptation*	98.09	98.57	102.34	39.36	42.39	28.91	28.61	12.19	19.99
	RPF*	98.01	98.57	102.07	35.86	42.18	28.89	**24.66**	12.24	19.5
	SPS*	97.97	98.57	102.07	35.04	42.57	26.87	**24.79**	12.24	19.9

of dose metric differences ΔDx. The mean values across the five patients of the dose metric differences for CTVs, parallel and serial OARs were relatively small (<0.5 Gy/%), with no clear trend as to which strategy produced plans with minimal dose differences.

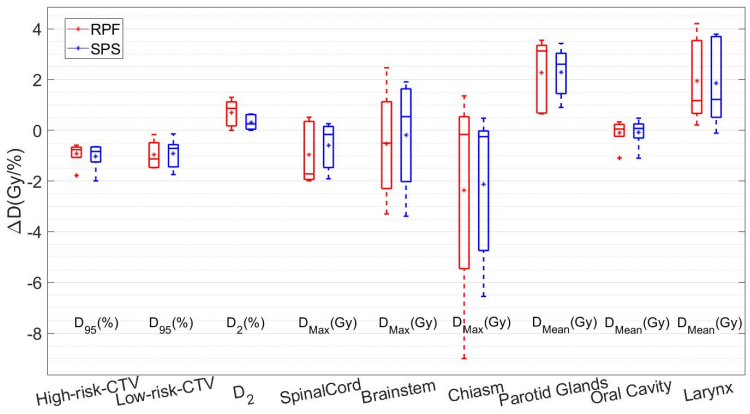


Figure 6.2: Comparison of the dose metric differences for 5 patients observed in the SPS and the RPF. The horizontal lines in the box plot indicate the median dose metric difference among the five patients, and the asterisks indicate the mean difference. D_{max} and D_{mean} are given in units of Gy.

The statistical analysis using the two-sample t-test also demonstrated that the

difference between the SPS and the RPF was not significant ($p>0.05$ for all investigated dose metrics), as shown in table 6.4. Given the findings and the workload in clinical proton therapy, it appeared that a practical solution is to continue with the reactive predicted adaptation strategy, requiring only 1 replan.

Table 6.4: p-value of the two-sample t-test between the distribution of ΔDx (see equation 1) in the RPF and the SPS with mean value(μ) and standard deviation(σ). Statistically significant differences between plans (taken as $p<0.05$) are highlighted in bold.

Statistics	High-risk CTV D_{95} (%)	Low-risk CTV D_{95} (%)	D_2 (%)	Spinal cord D_{max} Gy	Brainstem D_{max} Gy	Chiasm D_{max} Gy	Parotid D_{mean} Gy	Oral Cavity D_{mean} Gy	Larynx D_{mean} Gy
p value with mean value(μ) and standard deviation(σ) of five patients									
RPF($\mu \pm \sigma$)	-0.904± 0.446	-0.771± 1.161	-0.138± 0.228	-0.969± 1.257	-0.543± 2.247	-2.382± 4.265	1.747± 1.237	1.098± 1.132	1.876± 1.78
SPS($\mu \pm \sigma$)	-0.994± 0.524	-0.589±0.904	-0.382±0.33	-0.752± 0.932	-0.342± 2.202	-2.171± 3.037	1.744± 0.705	1.099±1.131	1.819± 1.806
p value(SPS and RPF)	0.777	0.789	0.210	0.764	0.89	0.930	0.996	0.999	0.961

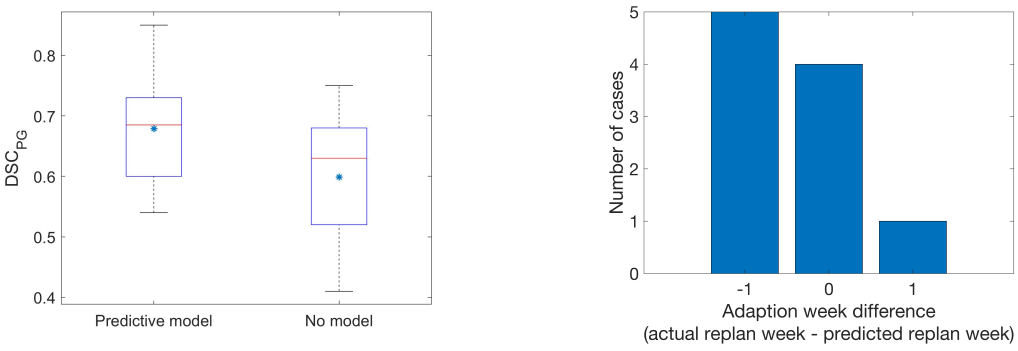
6.3.2 Comparison between reactive predicted adaptation strategies and the standard replan technique

The predictive power of the method is shown in figure 6.3, including DSC and replan week comparisons. The DSC of the parotid glands (DSC_{PG}) between predicted contours (used in the predicted plan adaptation) and the real contours (used in the standard replan) are compared with no model, in which planning contours and real contours in the standard replan are used to calculate the DSC, in figure 6.3a. Using the reactive predicted adaptation strategies, the DSC_{PG} was increased by 0.08.

The deviation of the actual replan week from the predicted replan week is shown in figure 6.3b. Of note, this study did not apply the predicted plan to the predicted week but applied it flexibly to the actual week that requires replanning. In 4/10 cases, the predicted images accurately predicted the replan week (one patient that did not need a replan is included in this scenario). In the remaining 6/10 cases, the predicted week and actual replan week differed by only 1 week.

In figure 6.4, the DVH differences between the dose of the nominal plan on a predicted CT and on the corresponding triggering repeat CT are illustrated for an exemplary patient. The DVH differences are small.

To further investigate the performance of the reactive predicted adaptation strategies, the RPF and the RPS strategy were compared with the standard replan



(a) **(b)**

Figure 6.3: The predictive ability of the RIM. (a) The comparison between the RIM and no model on DSC_{PG} for the ten validation patients. The horizontal lines in the box plot indicate the median value among the 10 patients, and the asterisks indicate the mean value. (b) shows the distribution of cases on adaptation week difference (between real replan week in the standard replan and the predicted replan week in the predicted plan adaptation).

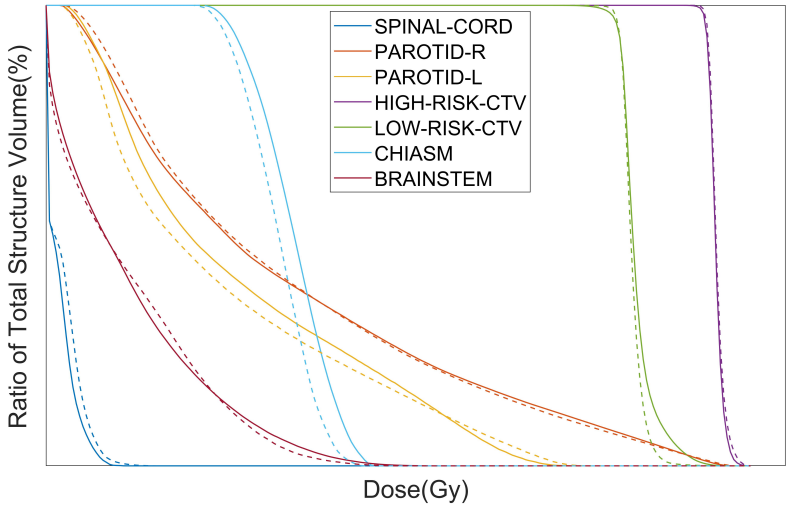


Figure 6.4: The DVH comparison between the dose of the nominal plan calculated on a predicted CT and on the corresponding triggering repeat CT for an exemplary patient. The solid lines belong to the dose on the predicted CT. The dashed lines belong to the dose on the triggering repeat CT.

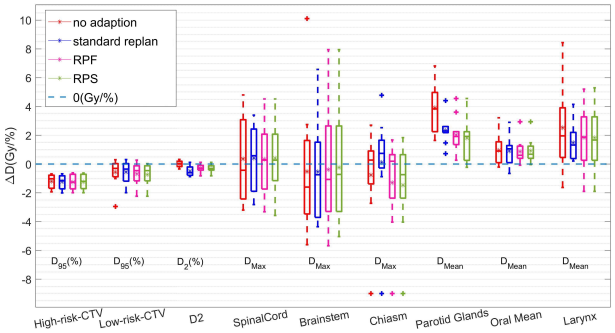
strategy based on the results of 10 patients (IDs, 1–10). Table 6.5 includes the nominal plan and the accumulated doses of the nominal plan with no adaptation, the standard replan, the RPF and the RPS. For patients who had a standard replan on week 3, the accumulated doses of the RPF and the RPS are the same. The results of the RPS are only shown in patients who did not have standard replan on week 3.

Table 6.5: Nominal and accumulated dose metrics with no adaptation, the RPF, the RPS and the standard replanning for the ten patients. Please note that for patient 1, no replan was triggered according to the replan criteria. The dose metrics for the parotid glands are highlighted in bold. The asterisk in the plan column indicates the numbers in the same row are the accumulated dose.

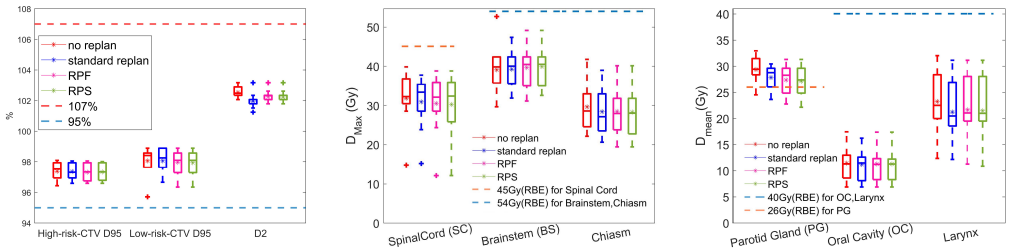
ID	Plan	High-risk CTV D_{95} (%)	Low-risk CTV D_{95} (%)	D_2 (%)	Spinal cord D_{max} Gy	Brainstem D_{max} Gy	Chiasm D_{max} Gy	Parotid D_{mean} Gy	Oral Cavity D_{mean} Gy	Larynx D_{mean} Gy
1	Nominal plan	98.75	98.86	103.05	30.55	45.69	36.5	27.36	6.8	30.93
	No Adaptation *	97.93	98.89	103.17	28.56	42.39	27.5	29.61	6.9	31.14
2	Nominal plan	98.28	98.66	102.92	34.49	40.08	22.06	25.53	14.24	26.24
	No Adaptation*	96.6	95.71	102.89	33.39	40.63	22.16	31.87	17.45	32.03
	Standard Replan*	97.11	96.67	101.79	35.45	40.54	20.64	**29.59**	16.24	26.89
	RPF*	96.6	96.35	102.62	35.81	41.27	19.45	**28.92**	17.39	28.43
3	Nominal plan	98.75	98.59	102.24	33.13	39.16	23.88	22.89	10.88	11.77
	No Adaptation *	97.97	98.89	102.07	30.59	35.69	24.58	24.53	10.76	12.36
	Standard Replan*	98.05	98.89	101.52	30.28	35.55	23.53	**23.65**	11.19	12.16
	RPF*	98.05	98.89	102.07	31.34	38.28	23.8	**22.77**	10.48	11.27
	RPS*	98.01	98.57	102.07	31.90	39.34	22.76	**22.15**	10.47	10.87
4	Nominal plan	98.42	98.82	102.96	32.9	34.4	32.02	24.53	7.43	22.18
	No Adaptation*	97.74	97.94	102.62	30.47	29.68	29.29	28.36	8.62	25.09
	Standard Replan*	97.86	97.62	101.79	23.85	32.26	27.3	**26.05**	8.11	20.84
	RPF*	97.82	96.98	102.07	24.36	33.89	22.04	**25.03**	8.31	21.39
5	Nominal plan	98.82	98.67	102.02	34.56	40.75	26.21	23.63	10.65	16.08
	No Adaptation*	98.09	98.57	102.34	39.36	42.39	28.91	28.61	12.19	19.99
	Standard Replan*	97.97	98.89	101.79	37.18	42.65	26.98	**26.08**	11.88	18.57
	RPF*	98.01	98.57	102.07	35.86	42.18	28.89	**24.66**	12.24	19.5
6	Nominal plan	98.5	98.65	102.7	36	38.93	34.37	29.65	15.38	30
	No Adaptation*	97.38	98.57	102.89	39.86	41.68	32.99	32.95	17.14	28.37
	Standard Replan*	97.03	98.25	101.79	37.74	41.93	32.98	**30.44**	16.15	27.41
	RPF *	96.79	97.94	102.34	38.81	43.56	31.84	**31.3**	16.29	28.04
7	Nominal plan	98.6	98.27	102.58	36.53	40	26.86	27.04	13.19	19.64
	No Adaptation*	97.03	98.57	102.34	36.77	39.05	28.31	30	12.98	23.13
	Standard Replan*	96.64	98.25	102.34	34.33	38.22	26.29	**28.87**	12.61	21.04
	RPF*	96.91	98.25	102.07	36.41	39.67	28.45	**28.07**	12.33	21.05
	RPS*	96.91	98.25	102.07	36.62	39.74	28.58	**27.49**	12.11	20.91
8	Nominal plan	98.37	98.63	102.48	35.47	36.01	23.43	27.24	10.6	19.77
	No Adaptation*	96.44	97.62	102.34	32.27	30.41	23.88	29.25	11.86	20.77
	Standard Replan*	96.6	97.94	102.07	33.2	31.9	22.67	**28.66**	11.78	19.51
	RPF*	96.6	97.94	102.34	29.13	31.14	23.99	**28.48**	11.98	20.91
	RPS*	96.8	97.94	102.07	25.85	32.59	23.42	**27.27**	12.17	19.73
9	Nominal plan	98.75	98.76	102.31	11.84	42.58	42.63	22.66	8.83	9.2
	No Adaptation*	96.95	98.25	102.62	14.78	52.68	41.76	27.14	9.45	17.64
	Standard Replan*	96.95	98.25	102.07	15.2	47.39	38.97	**25.2**	10.14	14.67
	RPF*	96.75	98.57	102.07	12.14	49.15	40.14	**24.86**	9.2	13.68
10	Nominal	98.62	98.28	102.01	29.2	38.76	36.3	27.64	7.1	21.44
	No Adaptation*	97.7	97.62	102.07	32.28	36.51	37.21	31.47	7.21	21.9
	Standard Replan*	97.54	96.98	101.24	33.62	39.52	37.66	**30.29**	7.06	20.12
	RPF*	97.78	97.3	101.79	32.92	35.07	38.78	**30.00**	7.86	21.07

In figure 6.5a), the RPF and the RPS were compared with the standard replanning technique and no adaptation for all 10 patients in terms of dose metric differences ΔDx. The mean and median differences of D_{95} between the accumulated dose and the planned dose for the CTVs were small, with a mean difference observed to the high-risk CTV of -1.20%, -1.23%, -1.25% and -1.24% for no adaptation, the standard replan, the RPF and the RPS, respectively. For parotid glands, on average, all adapted strategies can control the accumulated mean dose within a 3 Gy difference of the planned dose, with 2.34 Gy, 2.03 Gy and 1.81 Gy for the

standard replan, the RPF and the RPS respectively, compared with 3.91 Gy for no adaptation. Furthermore, the parotid D_{mean} using a reactive predicted adaptation strategy was generally lower than a standard replan, with 0.31 Gy and 0.53 Gy on average for the RPF and the RPS, respectively. To be noted here, the parotid D_{mean} of the predicted replan on the PD_r (either the RPF or the RPS) was observed to be lower than that of the standard replan on the verification CT in some cases (for example, patient 3 in table 6.2) because they were optimised on a different geometry. In figure 6.5a) that bias is removed using the dose metrics difference.



(a) Dose metric differences of the three adaptations.



(b) Accumulated D_{95} and D_2 for the high-risk and low-risk CTV.

(c) Accumulated D_{max} for the serial organs

(d) Accumulated D_{mean} for the parallel organs

Figure 6.5: Comparison of the dose metric differences and the accumulated dose metrics for 10 patients observed in no adaptation, the standard replan, the RPF and the RPS. The horizontal lines in the box plot indicate the median dose metric difference among 10 patients, and the asterisks indicate the mean difference. (a) shows the dose metric differences. D_{max} and D_{mean} are given in units of Gy. (b)–(c) shows the accumulated dose metrics. The dashed lines in (b)–(c) represent the defined clinical goals summarised in table 6.1.

Serial organs were observed to be more sensitive to geometric changes than parallel organs. In the standard replanning technique, the differences in the spinal cord D_{max} between the accumulated dose and the planned dose range from -2.80 Gy

to 3.41 Gy, consistent with the result reported by Noble *et al.* (2019) [98]. In the brainstem, those differences range from -4.58 Gy to 7.16 Gy. In the chiasm, the differences were observed between -9 Gy to 6.83 Gy.

Figure 6.5b)–6.5d) shows that all the accumulated dose metrics of CTVs and OARs satisfied the clinical goals set for the plans, with one exception that the parotid D_{mean} slightly exceeded the clinical goal. There are two reasons: firstly, the parotid D_{mean} of the nominal plan is already close to the clinical goal, with 25.82 Gy on average, and secondly, this cohort of patients all received chemotherapy, which causes severe shrinkage of the parotid glands during the treatment.

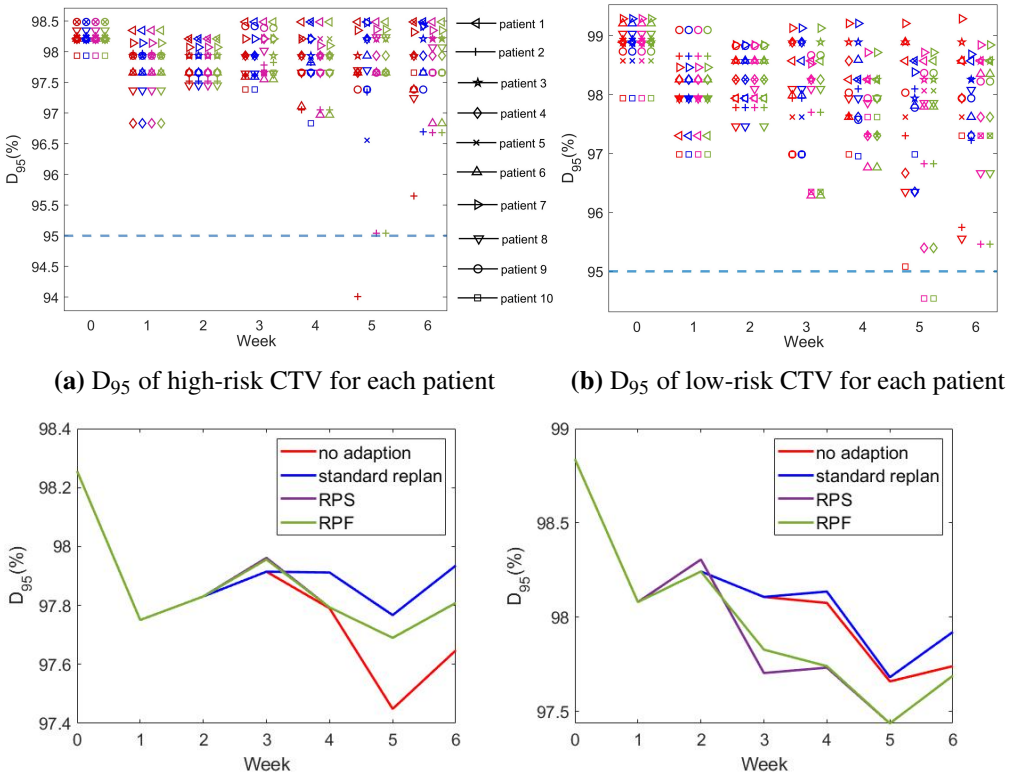
Statistical analysis using the two-sample t-test demonstrated that the accumulated dose difference between no adaptation and the standard replanning technique was significant ($p<0.05$) only in the parotid gland D_{mean} and $D_{2\%}$. The differences between the standard replan, the RPF, and the RPS were not significant (see table 6.6).

Table 6.6: p-value of the two-sample t-test between the distribution of ΔDx (see equation 1) in two adaptation strategies with mean value(μ) and standard deviation(σ). The capitals in parenthesis indicate no adaptation(N) and the standard replan(R). Statistically significant differences between plans (taken as $p<0.05$) are highlighted in bold.

Statistics	High-risk CTV D_{95} (%)	Low-risk CTV D_{95} (%)	D_2 (%)	Spinal cord D_{max} Gy	Brainstem D_{max} Gy	Chiasm D_{max} Gy	Parotid D_{mean} Gy	Oral Cavity D_{mean} Gy	Larynx D_{mean} Gy
No Adaptation($\mu \pm \sigma$)	-1.203± 0.491	-0.556± 0.96	0.007± 0.23	0.365± 3.029	-0.525±4.623	-0.767±3.267	3.913±1.773	0.945± 1.071	2.516± 3.009
Standard replan($\mu \pm \sigma$)	-1.236± 0.495	-0.545± 0.753	-0.466± 0.325	0.795± 2.181	-0.571± 3.715	0.848± 4.073	2.337± 0.989	0.989± 1.125	1.855± 1.248
RPF($\mu \pm \sigma$)	-1.252± 0.567	-0.685± 0.823	-0.295± 0.271	0.328± 2.307	-0.383± 4.036	-1.301± 3.201	2.03± 1.265	0.87± 0.867	1.846± 2.206
RPF($\mu \pm \sigma$)	-1.235± 0.550	-0.732± 0.786	-0.319± 0.266	0.418± 2.327	-0.237±3.969	-1.471±3.146	1.807±1.507	0.915± 0.86	1.824± 2.234
p value(N and R)	0.883	0.978	**0.001**	0.72	0.981	0.341	**0.024**	0.929	0.529
p value(R and RPF)	0.947	0.697	0.216	0.647	0.915	0.206	0.553	0.794	0.991
p value(RPF and RPS)	0.947	0.897	0.84	0.932	0.936	0.906	0.745	0.908	0.982

The weekly dose metrics changes of CTVs during the treatment are shown in figure 6.6 for each patient. In figure 6.6, CTV coverage is basically above 95% for all the adaptation methods, except for the variation of patient 10 at week 5. However, because the D_{95} of the low-risk CT of patient 10 at week 6 is above 95%, It can be deduced that this dose degradation is caused by small non-rigid variations such as neck tilt, which are difficult to subtract when the RIM is updated using individual information.

There are scenarios in which using the initial tumour contour of the planning



(a) D_{95} of high-risk CTV for each patient

(b) D_{95} of low-risk CTV for each patient

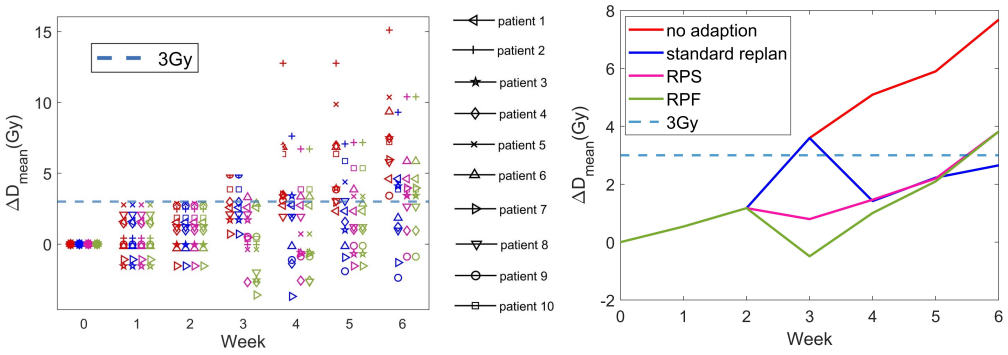
(c) Average weekly D_{95} of high-risk CTV across ten patients.

(d) Average weekly D_{95} of low-risk CTV across ten patients.

Figure 6.6: The weekly changes of D_{95} for high-risk and low-risk CTV for the 10 patients observed in no adaptation, the standard replan, the RPF and the RPS. Patients are represented using different markers listed in the middle. The dashed lines present the defined clinical goals summarised in table 6.1.

CT in the predicted plan shows better coverage than using the actual tumour contour in the standard replan, for example, patient 9 at week 5 in figure 6.6a (the position of the blue circle is below the pink circle and the green circle), it is because larger contours can mitigate the variations during the treatment. In the TORPEdo trial, therapeutic target volumes are not adapted according to changes in GTV. While in some other cases, the tumour displacement was extended outside the original contour [18], such as in patient 2 at week 5 in figure 6.6a (the position of the blue '+' is above the pink '+' and the green '+'). In such cases, the predicted replan using the initial contour is inferior to the standard replan using the actual tumour contour.

The weekly dose metrics changes of parotid glands during the treatment are shown in figure 6.7a). This study found that the average parotid D_{mean} in the RPS is



(a) D_{mean} difference of parotid glands for each patient

(b) Average weekly D_{mean} of parotid glands across ten patients

Figure 6.7: The weekly difference D_{mean} between weekly D_{mean} and D_{mean} of week 0 (planning CT) during the treatment with different adaptive strategies for the ten patients.

lower than the RPF. This benefit comes from applying the reactive predicted plans in advance for patients 3, 7 and 8. However, the reduced average parotid D_{mean} comes with a slight degradation in low-risk CTV D_{95}, see figure 6.6d). Because the reactive predicted plans were applied in advance to the triggering week, at which the shrinkage of low-risk CTV was less severe than in PD_r, the coverage of low-risk CTV was slightly degraded. However, this degradation was less than 1% for patients 3, 7 and 8. While initial high-risk CTV of planning CT was both used in the PD_r of the RPF and the RPS, no significant influence was made on the radiation coverage of high-risk CTV.

Figure 6.7b) shows that the average parotid ΔD_{mean} across ten patients is above 3 Gy from week 3. Therefore, week 3 can be justified as an appropriate time point for adaptation. In this cohort of patients, 6/10 patients had plan adaptations around the 3rd treatment week, and 2/10 patients had plan adaptations around the 4th treatment week, which is consistent with the conclusion from Brown *et al.* (2016) [96] and Wu *et al.* (2008) [97]. The complete table of the actual replan week and predicted replan week in the RPF can be found in Appendix D.

6.4 Discussion

Adaptive proton therapy can address the dose discrepancy caused by anatomical changes observed in H&N patients during the course of treatment. However, the offline adaptive strategy used in radiotherapy clinics can cause treatment delays and challenge workflow efficiency. This chapter exploits the RIM to prepare adaptive plans in advance. The results showed that the RPS can maximise the use of the RIM for prospective replanning, with similar CTVs coverage and reduced parotid dose compared to the standard replan.

Previous studies[51, 95] reported that the benefits of adapting the treatment should increase with an increasing replanning rate. However, the ΔDx of the SPS (using two adaptive plans) did not show significant benefits compared to the RPF (using one adaptive plan). One reason could be that the geometric changes of this patient cohort become stable in the last few fractions of treatment, resulting in the second predicted plan not significantly improving the results. A second reason could be that the prediction of the anatomical changes for some patients lags one week behind the actual anatomical changes. It should be noted that the benefit of the SPS is limited by the model accuracy, while the reactive adaptation technique (either the RPF or the RPS) is not due to its flexible application. Accurately predicting the week in which plan adaptation is an important next step in implementing a prospective offline adaptive therapy.

The comparison between the RPF and the RPS inspires a further potential application of anatomical models as follows: One could create predicted adaptive plans using each predicted weekly CT; then clinicians verify each predicted plan on CT_3 and CT_5 and choose the best plan for application. While this poses an extra workload to planning and QA, automated treatment planning [99] could be used.

In the RPF and the RPS strategies, although high-risk CTV and low-risk CTV coverage were slightly lower than the standard replan, the differences were statistically not significant ($< 1\%$). The CTV coverage degradation observed here is because the accurate, newly defined CTV contours were used in the standard replan, while the initial CTV contours from the nominal plan were used in the predicted re-

plans. This limitation can possibly be improved by using the contours in the latest weekly repeat CT or predicted CTV contours with improved accuracy after stratifying patients based on tumour features. The improved CTV contours can help further reduce the dose of OARs since the initial volume of CTVs is generally bigger than the CTVs adapted during treatment [18]. These works will be explored in the future.

In this work, the average benefit of a no-delay treatment on parotid D_{mean} is 0.31 Gy for the RPF and 0.53 Gy for the RPS. If the standard replan is applied to the same day of the RPF, the benefit of no-delay treatment from standard replan is on average 0.47 Gy (one fractional benefit is 0.09 Gy) to parotid D_{mean}. This benefit is 0.16 Gy higher than that from the RPF, indicating that the advantages of the RPF can enable adaptive therapy to be delivered without 3 sub-optimal fractions, but 0.06 Gy lower than from the RPS because in the RPS the adapted plans were applied in advance for some patients. There is one interesting scenario worth noticing. For patient 6, the predicted anatomy, which was used to create a prospective replan, suggested the need for adaptation due to a change in the mean parotid dose of $>$3 Gy. However, while the parotid ΔD_{mean} of the nominal plan calculated on the predicted anatomy was 3.05 Gy, the parotid ΔD_{mean} calculated on the triggering verification CT of week 3 was 6.35 Gy. A dramatic shrinkage of the parotid volume between week 2 and week 3 made the RIM less effective in this case. Despite this, when the predicted plan was applied on week 3, the parotid ΔD_{mean} on the verification CT of week 3 was reduced to 3.05 Gy. In this circumstance, clinicians might look to apply the predicted replan on week 3 and follow the standard replan procedure to create a new replan. This limitation can potentially be improved if patients are stratified by exploiting tumour-related features [87, 88] and outlining change-related features [89, 90, 91] based on a larger dataset. Another scenario worth discussing is where the dose recalculation on the predicted images indicated a parotid ΔD_{mean} that did not reach the triggering threshold but was close to it, for example, 2.75 Gy(RBE). In this case, this study suggests having an alternative plan available in case re-planning is triggered.

This study only considered 1-step adaptation for the proof of concept. In figure 6.7 a), the parotid ΔD_{mean} of some patients in week 4 is still more than 3 Gy(RBE) because of severe shrinkage. For example, patient 2 adapted the plan on week 3 with a parotid D_{mean} of 25 Gy(RBE). However, when the adapted plan was applied to week 4, the parotid D_{mean} increased to 32 Gy(RBE). In this case, a second adaptation is needed. If the model is updated every time when a new CT is acquired, then the prediction for a second adaptation is clearly possible.

In clinical practice, plans are robustly optimised to account for setup and range uncertainties, and while anatomical changes may not be included explicitly in the optimisation process, it is possible that some robustness to anatomical changes is provided by improved robustness to these setup and range uncertainties. Where the magnitude of setup errors in the robust optimisation is reduced, plans would likely be more sensitive to anatomical changes. As such, the application of predictive anatomical modelling to the design of a robust plan may allow for a reduced setup robustness margin, thereby improving dose conformity to some extent. Though there may be a trade-off between this margin and the number of plan adaptations required during a treatment course. Furthermore, the use of a predictive model as an additional error scenario in the robust optimisation may be of interest though further work is required to understand the detriment to the nominal plan of such robustness. The predictive nature of the approach can enable improved workflow management. It also should be noted that robust beam angle selection remains critical for avoiding anatomical variations such as nasal filling. Such variations cannot be modelled by deformations.

The data in this study was initially used to report the geometric and dosimetry changes of H&N patients who have received photon therapy. The mean parotid dose of the photon plans was 32.12 ± 11.2 Gy (mean ± standard deviation) in this cohort of patients, whereas the mean parotid dose of the proton plans in this study was 25.82 ± 2.33 Gy. Deduced from here, patients who underwent proton therapy might experience less toxicity. Therefore, weight loss due to acute toxicity, such as dysphagia and dysgeusia, could be lower, resulting in smaller magnitudes of

anatomical changes. This study may overestimate the number of patients who can benefit from using the predicted images of the RIM for offline adaptive therapy.

For this proof of concept, images from a cohort of 20 patients were used. This cohort was particularly suited for this study, having weekly CTs with manually delineated contours. In contrast, the use of weekly CBCT, which is more commonly available, is subject to additional DIR uncertainty. Secondly, the error from HU correction required by CBCT is removed. Weekly dose distributions can be directly calculated, and a replan can be directly created. In addition, the weekly dose metrics can be directly calculated using the manually delineated contours in weekly CTs.

Further model validation will be conducted on a larger cohort of IMPT patients with CBCT. The SVFs of the DIR between CT-CBCT can be directly used to build the model. The procedure is the same, except that the influence of the DIR uncertainty needs to be re-evaluated because the estimation of deformation fields is less accurate between CT-CBCT than CT-CT if the similarity metric of DIR is based on intensity. Another option is correct the HU of CBCT [100, 101] before DIR to improve the accuracy.

In the literature, anatomical models have only been used for DIR evaluation [60] or assessed based on the misalignment of the anatomical landmarks [63, 28]. This is the first demonstration of the potential of anatomical models in adaptive radiotherapy. Compared to online adaptation [50, 51, 48], this method can reduce the treatment time by preparing complete adaptive plans in advance without the need for real-time QA, which is one of the most challenging aspects of online adaptation. Furthermore, by exploiting novel auto-planning techniques [99], this method may allow for auto-replanning for adaptive IMPT.

6.5 Conclusion

This chapter explores the three strategies (the SPS, the RPF and the RPS) of using the RIM to prepare adaptive plans in advance. The RPS reduced the parotid D_{mean} by 0.53 Gy, achieving the benefit of non-delay treatment, while the dosimetric of other organs had no significant difference with the standard replanning technique

($p>0.05$ for all investigated dose metrics). Prediction-based replanning could potentially enable clinically acceptable adaptive therapy without treatment gaps or sub-optimal fractions to be delivered due to delays in the standard adaptive strategy, leading to an improved overall treatment course for patients. Furthermore, the ability to manage the adaptive therapy workflow prospectively with the predictive approach can increase the efficiency of a busy clinical proton therapy practice.

Chapter 7

Final remarks

Proton therapy has presented great potential in limiting the dose to normal tissues adjacent to the target region for H&N cancer patients. However, the sharp distal fall-off of the Bragg peak makes the dose distribution sensitive to uncertainty. With this precise delivery technique becoming widely available, there is a clear need to improve mitigation techniques to reduce the influence of uncertainty.

The anatomical changes commonly occur in the treatment of H&N cancer. Previous studies measured the anatomical uncertainty without discreetly categorising it into different sources. However, the underlying causes of sNRV and progressive changes are different. The sNRVs arise from non-rigid positioning errors, movements due to discomfort and spontaneous swallowing. They are more likely to occur randomly during the treatment. While progressive changes are due to radiation damage on cells and are patient-specific. The damage to tumour cells causes tumour shrinkage. The damage on normal cells not only leads to volume shrinkage but also causes acute toxicity, influencing chewing, swallowing and outline changing. Therefore, these two types of anatomical changes should be treated differently in terms of using mitigation techniques.

One approach to mitigate anatomical uncertainty is robust optimisation, which is accompanied by robust evaluation. In the conventional robust optimisation/evaluation, only rigid setup error and range uncertainty are considered. The inclusion of anatomical changes in anatomical robust optimisation/evaluation [43, 45] needs multiple scanning or relies on the acquisition of CT images during treatment.

It either gives the extra dose to patients and increases the workload for clinicians or limits the creation of a robust plan at the planning stage.

Offline replanning is generally used to account for anatomical uncertainty. However, patients must either continue treatment with an existing sub-optimal plan or face interruptions to treatment whilst plans are adapted. This reactive adaptive strategy also challenges workflow efficiency.

The work described in this thesis has contributed to the field by addressing the aforementioned issues using imaging computing techniques, DIR. The diffeomorphic image registration algorithm of Niftyreg was evaluated for building anatomical models for H&N proton therapy. Because anatomical models were built at each weekly time-point, the weekly changing trend and accumulated influence of the DIR were assessed. The results justified the use of this DIR algorithm. In this study, the uncertainty evaluation based on WSLD was proposed. WSLD gives a single value, making it easy to analyse. Besides, it combines spot location and spot weights, both of which affect dose distribution, thus accurately and effectively evaluating the uncertainty without having to calculate dose distribution.

Several potential applications for mitigating anatomical uncertainty were proposed in this thesis.

1. Include sNRVs into the robust evaluation to choose robust plans against anatomical uncertainty.

2. Develop the PM to generate the predicted images with high possibilities for anatomical robust optimisation.

3. Develop the RIM for prospective replanning. Different strategies were explored to maximise the use of RIM.

sNRVs are unavoidable during the treatment. Thus they are better to be considered in the robust evaluation/optimisation. The role of sNRVs in the robust evaluation was investigated in Chapter 3. This study used a probability approach to evaluate the robustness of a plan and provided a way to quantify its performance using robustness ranking consistency. The approach presented not only can help

clinicians avoid the sNRVs that can lead to unacceptable dose distributions but also help to select the robust beam arrangement. It was the first study to reveal the dosimetric influences of sNRVs and incorporate them into the robust evaluation to guide the selection of a robust IMPT plan. The study suggests that the inclusion of sNRVs can be beneficial to robust evaluation for all beam arrangements.

When a large patient cohort is available, it would be more efficient to limit the included sNRVs to the most common/frequent ones by using an anatomical model. In Chapter 4, the PM captured the major deformations at each weekly time point based on population data and has been validated as the most effective model in describing anatomical uncertainty with an error of 0.81 ± 0.56 mm. The dose variations simulated by the PM in the first week (simulated sNRVs) have no significant difference from those caused by the actual sNRVs. This study gives a solution to calculate the probability of a certain type of anatomical change, which can be added to the cost function of the optimiser. Not limited to just including sNRVs in anatomical robust optimisation, anatomical robust optimisation can also account for progressive changes during treatment. However, both Mesías *et al*. (2019) [43] and Lalonde *et al*. (2021) [48] demonstrated that the robustness of anatomical robust optimisation comes with the price of increasing the integral dose. The anatomical robust optimisation using the PM needs to be explored to find the best balance between those. The potential images for anatomical robust optimisation can be from 1) the PM of each week, 2) the AM of each week plus the PM of the first week. The anatomical robust optimisation using the PM does not require multiple CT scanning for the same patient pre-treatment, therefore saving imaging dose and reducing workload.

The robust optimisation alone may not be adequate to account for progressive changes during the treatment. Adaptive planning is still suggested [34]. To prospectively prepare replans for offline adaptive radiotherapy, Chapter 5 built the RIM based on the AM and updated it using the patient's progressive information acquired during the treatment. The RIM effectively reduced anatomical uncertainty and can gradually refine the prediction. In this study, the predictive ability of the RIM was demonstrated based on CT numbers, contours, range and dose distribu-

tion. The validation steps and the steps for clinical translation can directly be applied to more complex future models.

Due to the limited patient data size used in this thesis, the RIM cannot accurately predict the replan week. To minimise the influence of the current accuracy of the RIM, a reactive strategy was chosen to create a predicted plan in advance. The results showed that the reactive strategy using the RIM can achieve the minimum benefit of no-delay treatment in terms of protecting parotid glands. The benefits of using the RIM to prepare adaptive plans are 1) improved treatment quality as there is no gap in treatment or the delivery of a few sub-optimal fractions. 2) streamlined workflow as the option to prepare adaptive plans in advance allows for easier scheduling of patient-specific QA along with machine QA, maintenance, and other demands for beam time. 3) compared with online adaptation, which is intended for same-day application. This application can reduce the treatment time by preparing complete adaptive plans in advance and improve the plan quality as the plan of online adaptation is never as good as offline adaptation due to limited computational capacity and constrained optimisers.

Along with the application, the anatomical uncertainty was evaluated from three aspects: 1) systematic progressive deformation of each week, 2) sNRVs uncertainty of the first week and 3) total anatomical uncertainty (from sNRVs and progressive changes) of each week.

One limitation of this study is that anatomical models were built based on a patient cohort treated with photon radiotherapy but collected retrospectively for proton study. However, the presented methods are easily transferable to a patient data set having received proton therapy. Another aspect of anatomical models that is of interest for future research is exploiting big data techniques. Either stratifying patients based on features in the planning stage or using deep learning methods to directly obtain the predicted geometry worth investigation[102, 103].

To summarise, the work presented in this thesis provides a better understanding of anatomical uncertainty and mitigation techniques. First, the influences of anatomical uncertainty are presented. Second, the anatomical models and the model

evaluation methods are given, which can guide the development of more complex future models. Third, the applications and potential applications of models are provided. The results might inspire future clinical applications.

Appendix A

Parameters used in DIR

Velocity field option was used. Other optimal parameters used for CT-CT image registration were investigated and shown in table A.1.

Table A.1: NiftyReg parameters for CT-CT registration.

Parameters	Value
Similarity Measure	NMI
Control Point Spacing	5mm
Bending Energy	10%

Other optimal parameters used for CT-atlas image registration were investigated and shown in table A.2.

Table A.2: NiftyReg parameters for CT-atlas CT registration.

Parameters	Value
Similarity Measure	NMI
Control Point Spacing	7mm
Bending Energy	1%

Other optimal parameters used for CT-CBCT image registration were investigated and shown in table A.3[104].

Table A.3: NiftyReg parameters for CT-CBCT registration.

Parameters	Value
Similarity Measure	NMI
Control Point Spacing	5mm
Bending Energy	3%

Appendix B

Pre-treatment analysis of non-rigid variations can assist robust IMPT plan selection for head and neck patients

B.1 Examples of small non-rigid variations.

B.2 Beam arrangements illustration.

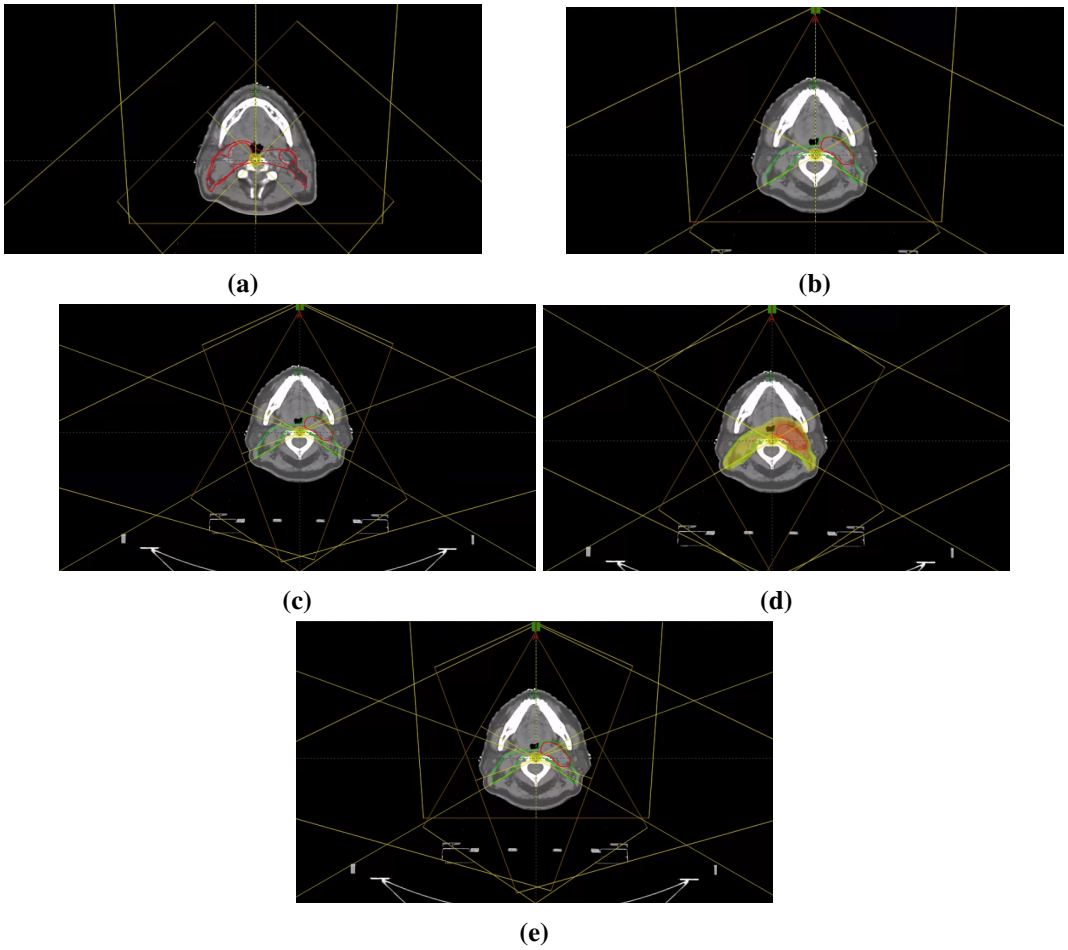


Figure B.1: Illustration of different beam arrangements. (a) - (e) are $3B_{45}$, $3B_{60}$, $4B_{110}$, $4B_{120}$, 5B, respectively.

B.3 Generation of average reference anatomy for statistical models

A groupwise registration algorithm was adapted from the NiftyReg package to generate an 'average' atlas. Group-wise registration can be used to spatially normalise a cohort of patients in a common space. Using N different patient CT images, the iterative algorithm consists of the following 6 steps:

1. Perform rigid registration between N-1 other CT I_i and an arbitrary reference image I^*. The warped image $I_i^{'} = I_i(\mathbf{T}_{\text{rigid}}, i)$. The template image is updated

as the average of all N images:

$$I^* = \frac{1}{N}(I^* + \sum_{i=1}^{N-1} I_i^{'}) \qquad I_i^{'} = I_i(\mathbf{T}_{\text{rigid}}) \tag{B.1}$$

2. Perform affine registration. All N images are registered to I^*, producing the affine transformations $\phi_{\text{aff},i}$ for each iteration.

3. Update I^* with a series of affine iterations. To enforce the mean of all the transformations to be the identity, the mean of the non-rigid components of the affine transformations $\phi_{\text{aff},i}$ is removed using $\phi_{\text{aff},i}^{'} = \phi_{\text{aff},i} - \frac{1}{N}\sum_i \phi_{\text{aff},i}$. I^* is updated using the average of the $\phi_{\text{aff},i}^{'}$:

$$I^* = \frac{1}{N}\sum_{i=1}^{N} \phi_{\text{aff},i}^{'}(I_i) \tag{B.2}$$

4. Steps ii) and iii) are repeated until there is no visual improvement of quality in I^*.

5. Perform diffeomorphic deformable image registration. All N images I_i are deformed to the current template image I^* using the stationary velocity field v_i.

6. Update I^* with a series of deformable image registration iterations. The spacing of the control points for the B-Spline velocity grid is gradually stepped down from coarse to fine (30mm to 8 mm) during the iteration. For each iteration, remove the mean of the velocity field from each transformation $v^{'} = v - \frac{1}{N}\sum_i v_i$ as before. The average image is computed as:

$$I^* = \exp\Big(\frac{1}{N}\sum_{i=1}^{N} v^{'}(I_i)\Big) \tag{B.3}$$

B.4 Robust evaluation table for other three patients.

Table B1: The sNRV+setup evaluation, the conventional evaluation and the gold standard evaluation for other three patients. p-values of the two-sample t-test are calculated between the distribution of ΔDx in the sNRV+setup evaluation and in the conventional evaluation. $p<0.05$ represents ΔDx in the sNRV+setup evaluation and ΔDx in the conventional evaluation are taken from significantly different distributions. In the sNRV+setup evaluation and the conventional evaluation, the beam arrangements were ranked based on the WD for each ROI matrix. In the gold standard evaluation, the beam arrangements were ranked for each ROI matrix based on ΔD_{st}. RP is the robustness ranking position of a beam arrangement for a matrix.

		sNRV+setup Evaluation					Conventional Evaluation				p-value	Gold standard evaluation		
ROI/Matrix	Plan	Nominal	Min	Max	WD	RP	Minmum	Maximum	WD	RP		$\mathrm{Accu}_{\mathrm{Nom}}$	ΔD_{st}	RP
patient 2														
Low-risk CTV	$3B_{45}$	98.25	92.22	97.94	1.6	3	96.67	98.25	0.36	4	1e-8	97.94	-0.32	3
D_{95}(%)	$3B_{60}$	98.25	93.17	98.25	1.92	4	96.03	98.25	0.3	2	1e-9	97.94	-0.32	3
	$4B_{120}$	98.25	93.81	98.25	0.97	1	95.71	98.25	0.36	3	1e-5	97.94	-0.32	3
	5B	98.57	94.44	98.57	1.57	2	96.98	98.57	0.26	1	1e-8	98.57	0	1
High-risk CTV	$3B_{45}$	98.48	95.73	98.48	0.87	2	96.56	98.48	0.4	2	1e-4	98.21	-0.28	2
D_{95}(%)	$3B_{60}$	98.48	93.8	98.21	1.03	4	96.56	98.48	0.42	3	1e-6	97.93	-0.55	4
	$4B_{120}$	98.48	94.9	98.21	0.89	3	96.56	98.48	0.47	4	1e-5	98.21	-0.28	2
	5B	98.48	95.45	98.48	0.77	1	96.56	98.48	0.34	1	1e-4	98.21	-0.28	2
High-risk CTV	$3B_{45}$	102.34	102.07	103.17	0.23	2	101.93	102.89	0.11	3	0.11	101.79	-0.55	2
D_2(Gy)	$3B_{60}$	102.62	102.34	103.17	0.12	1	102.48	103.03	0.03	1	0.93	102.07	-0.55	3
	$4B_{120}$	102.34	102.07	103.99	0.5	4	102.07	103.44	0.23	4	0.02	103.17	0.83	4
	5B	102.07	101.52	103.17	0.31	3	101.93	102.2	0.04	2	1e-5	102.34	0.28	1
Parotid	$3B_{45}$	21.69	19.13	29.05	1.22	1	18.72	26.57	1.05	4	0.01	24.67	2.98	1
D_{mean}(Gy)	$3B_{60}$	24.96	22.79	31.33	1.26	2	22.48	28.43	0.76	2	1e-3	28.2	3.24	2
	$4B_{120}$	23.07	20.8	31.53	1.86	3	20.97	24.98	0.77	3	1e-6	27.51	4.44	3
	5B	23.47	20.52	31.15	1.9	4	21.78	26.7	0.57	1	1e-4	28.6	5.12	4
Oral	$3B_{45}$	27.36	22.78	31.92	1.64	4	24.79	31.63	0.88	1	0.24	27.76	0.41	4
D_{mean}(Gy)	$3B_{60}$	12.15	9.46	16.01	1.37	2	9.59	15.67	0.93	2	0.22	12.13	-0.01	1
	$4B_{120}$	11.33	8.74	15.55	1.36	1	7.46	14.38	0.98	4	0.04	11.44	0.11	3
	5B	12.13	9.41	15.66	1.38	3	9.5	15.63	0.95	3	0.11	12.2	0.07	2
Larynx	$3B_{45}$	16.7	13.14	23.63	1.71	1	15.22	19.48	0.83	1	0.2	19.81	3.11	2
D_{mean}(Gy)	$3B_{60}$	17.13	13.61	23.81	1.76	2	15.18	20.12	0.9	2	0.11	19.6	2.47	1
	$4B_{120}$	16.62	11.21	32.2	3.44	4	9.32	23.07	2.13	4	0.01	25.44	8.82	4
	5B	17.76	14.44	26.44	2.42	3	15.53	21.5	0.98	3	0.01	21.17	3.41	3
Cochlea	$3B_{45}$	19.01	14.96	25.91	2.29	3	15.59	24.35	1.75	1	0.24	26.63	7.62	4
D_{mean}(Gy)	$3B_{60}$	13.33	8.24	25.87	2.39	4	7.72	20.88	2.27	4	0.45	13.75	0.42	3
	$4B_{120}$	21.67	18.09	29.09	1.87	1	17.98	32.11	1.99	2	0.73	21.31	-0.36	2
	5B	20.7	16.57	29.46	2.26	2	15.36	27.77	2.09	3	0.65	20.52	-0.18	1
Brainstem	$3B_{45}$	32.23	27.74	65.33	9.45	4	24.03	43.13	3.26	1	1e-6	35.4	3.16	4
D_{max}(Gy)	$3B_{60}$	39.17	30.07	59.99	5.37	2	26.57	50.67	3.95	3	0.11	38.91	-0.26	1
	$4B_{120}$	41.35	28.71	60.47	5.38	3	29.15	57.95	4.49	4	0.91	38.86	-2.49	3
	5B	43.21	31.31	63.29	4.11	1	30.91	53.71	3.79	2	0.9	41.77	-1.44	2
Spinal	$3B_{45}$	26.47	14.73	56.03	7.7	4	12.67	36.47	3.78	3	0.07	38	11.53	4
D_{max}(Gy)	$3B_{60}$	32.3	14.39	47.24	5.45	3	16	44	3.8	4	0.93	37.63	5.32	3
	$4B_{120}$	34.27	24.3	49.33	5.01	2	26.77	50.57	3.6	2	0.97	38.09	3.82	1
	5B	34.62	26.32	48.42	4.57	1	26.32	44.92	3.01	1	0.58	38.92	4.3	2
Chiasm	$3B_{45}$	29.45	23.11	53.5	9.33	1	15.55	44.05	5.56	4	1e-7	28.39	-1.07	3
D_{max}(Gy)	$3B_{60}$	27.1	23.23	51.1	10.31	2	14.8	41	4.96	2	1e-8	26.31	-0.8	2
	$4B_{120}$	26.06	22.68	50.93	11.29	4	14.86	41.06	5.22	3	1e-8	25.46	-0.6	1
	5B	27.92	25.06	51.89	10.53	3	16.02	41.32	4.76	1	1e-8	29.22	1.3	4
Optic nerve left	$3B_{45}$	20.4	9.17	48.61	4.78	2	8.7	34.9	4.94	4	0.38	21.33	0.93	3
D_{max}(Gy)	$3B_{60}$	23.6	11.56	49.63	4.78	1	11.1	38	4.75	2	0.62	24.1	0.51	2
	$4B_{120}$	25.02	12.81	50.15	5.11	4	12.42	39.02	4.87	3	0.35	25.28	0.26	1
	5B	24.76	13.03	50.04	5.04	3	12.76	38.56	4.58	1	0.7	26.13	1.38	4
Optic nerve right	$3B_{45}$	23.48	9.68	48.61	6.89	4	9.98	40.68	5.9	4	0.3	27.27	3.8	4
D_{max}(Gy)	$3B_{60}$	30.18	14.63	52.59	6.6	1	14.78	45.98	5.54	2	0.65	32.46	2.29	2
	$4B_{120}$	29.82	14.47	53.26	6.77	3	14.62	45.42	5.61	3	0.25	32.08	2.26	1
	5B	31.78	16.37	53.09	6.67	2	16.58	46.38	5.23	1	0.67	34.58	2.81	3

Table B1: Continued

ROI/Matrix	Plan	sNRV+setup Evaluation Nominal	Min	Max	WD	RP	Conventional Evaluation Minmum	Maximum	WD	RP	p-value	Gold standard evaluation $Accu_{Nom}$	ΔD_{st}	RP
patient 3														
Low-risk CTV	$3B_{60}$	98.57	95.4	98.57	1.26	3	95.56	98.57	0.66	2	1e-5	98.25	-0.32	3
D_{95}(%)	$4B_{120}$	98.89	95.4	98.89	1.15	1	95.4	98.89	0.52	1	1e-4	98.89	0	1
	5B	98.57	95.08	98.57	1.17	2	94.76	98.57	0.69	3	0.01	98.57	0	2
High-risk CTV	$3B_{60}$	98.48	91.87	98.48	1.21	1	92.7	98.48	0.75	2	0.01	98.21	-0.28	2
D_{95}(%)	$4B_{120}$	98.48	89.12	98.48	1.31	2	90.91	98.48	0.82	3	0.02	98.21	-0.28	2
	5B	98.48	91.87	98.48	1.33	3	93.39	98.48	0.6	1	1e-4	98.21	-0.28	2
High-risk CTV	$3B_{60}$	103.72	102.89	106.75	0.6	3	102.75	104.68	0.24	2	1e-3	103.72	0	1
D_2(Gy)	$4B_{120}$	103.44	102.62	105.37	0.52	2	102.75	104.41	0.35	3	0.13	102.89	-0.55	3
	5B	103.17	102.07	105.65	0.38	1	101.93	103.58	0.22	1	0.01	102.89	-0.28	2
Parotid	$3B_{60}$	30.15	27.49	40.84	2.12	2	27.16	34.09	1.09	1	1e-3	32.88	2.72	3
D_{mean}(Gy)	$4B_{120}$	29.75	27.1	41.62	2.12	1	26.69	34.1	1.3	3	1e-3	30.63	0.88	1
	5B	28.84	26.1	40.29	2.25	3	25.55	33.1	1.26	2	1e-3	30.7	1.87	2
Oral	$3B_{60}$	8.36	4.7	12.33	1.23	3	4.5	12.44	1.33	3	0.88	7.79	-0.58	2
D_{mean}(Gy)	$4B_{120}$	8.49	5.19	11.66	0.99	1	4.96	11.56	1.1	1	0.65	7.79	-0.7	3
	5B	7.31	4.03	10.91	1.12	2	3.83	10.87	1.17	2	0.93	6.76	-0.56	1
Larynx	$3B_{60}$	34.1	19.34	41.66	3.1	1	22.94	40.23	2.98	2	0.77	35.14	1.04	1
D_{mean}(Gy)	$4B_{120}$	34.57	16.73	49.62	4.69	3	21.91	40.93	3.33	3	0.21	37.32	2.75	3
	5B	32.1	16.69	43.46	3.36	2	21.77	37.78	2.66	1	0.3	33.46	1.36	2
Cochlea	$3B_{60}$	14.15	9.94	26.62	3.26	3	10.18	27.14	2.88	3	0.38	17.63	3.48	3
D_{mean}(Gy)	$4B_{120}$	17.9	14.59	29.85	2.39	2	14.67	30.42	2.52	2	0.64	20.8	2.9	2
	5B	18.93	15.57	30.68	2.33	1	15.67	31.21	2.5	1	0.65	21.7	2.77	1
Brainstem	$3B_{60}$	40.18	36.11	53.48	4.02	3	33.98	53.68	3.19	3	0.11	43.34	3.16	3
D_{max}(Gy)	$4B_{120}$	47.42	43.35	58.28	2.59	1	43.62	59.42	2.52	1	0.82	49.91	2.49	1
	5B	45.89	39.4	57.94	3.36	2	40.49	58.59	3.1	2	0.9	48.65	2.76	2
Spinal	$3B_{60}$	29.75	23.39	41.59	3.68	3	26.35	37.75	2.45	3	0.68	32.05	2.3	3
D_{max}(Gy)	$4B_{120}$	44.02	39.09	50.1	1.53	1	41.62	50.12	1.33	1	0.65	46.01	1.99	1
	5B	33.37	26.57	40.97	2.21	2	29.97	40.67	2.02	2	0.66	35.47	2.1	2
Chiasm	$3B_{60}$	33.94	22.24	50.15	6.05	3	20.64	45.54	5	3	0.15	36.41	2.47	3
D_{max}(Gy)	$4B_{120}$	37.59	23.16	51.88	5.89	2	23.69	48.09	4.56	2	0.15	38.41	0.82	2
	5B	36.68	22.29	50.36	5.74	1	22.68	46.48	4.45	1	0.14	36.74	0.06	1
Optic nerve left	$3B_{60}$	31.63	16.55	56.31	6.71	3	17.33	49.23	5.78	3	0.35	31.37	-0.26	1
D_{max}(Gy)	$4B_{120}$	31.84	18.14	53.26	5.48	1	18.64	46.44	4.94	1	0.55	30.94	-0.9	3
	5B	31.55	17.66	53.52	5.6	2	18.05	46.05	5.03	2	0.51	30.72	-0.83	2
Optic nerve right	$3B_{60}$	35.31	20.44	56.31	7.12	3	19.71	53.51	5.93	3	0.38	34.85	-0.46	1
D_{max}(Gy)	$4B_{120}$	31.84	18.14	53.26	5.82	1	18.74	43.84	4.47	1	0.22	28.55	-3.29	3
	5B	31.55	17.66	53.52	6	2	18.25	44.05	4.58	2	0.2	28.48	-3.07	2
patient 4														
Low-risk CTV	$3B_{60}$	98.57	96.03	98.25	1.04	1	97.62	98.57	0.33	3	1e-8	97.94	-0.63	2
D_{95}(%)	$4B_{120}$	98.57	96.35	98.25	1.04	2	97.78	98.57	0.2	1	1e-13	98.25	-0.32	1
	5B	98.89	96.03	98.57	1.18	3	98.1	98.89	0.22	2	1e-15	98.25	-0.63	3
High-risk CTV	$3B_{60}$	97.93	94.08	97.93	0.83	1	96.69	97.93	0.45	2	1e-3	97.38	-0.55	3
D_{95}(%)	$4B_{120}$	97.93	95.18	97.38	1.1	2	96.83	97.8	0.47	3	1e-9	97.66	-0.28	2
	5B	97.93	95.45	97.38	1.14	3	97.11	97.93	0.3	1	1e-13	97.93	0	1
High-risk CTV	$3B_{60}$	103.44	102.89	103.99	0.13	1	103.31	104.13	0.16	1	1e-3	102.62	-0.83	3
D_2(Gy)	$4B_{120}$	102.89	102.89	103.72	0.37	3	102.75	103.72	0.18	2	0.01	103.17	0.28	2
	5B	102.89	102.62	103.72	0.25	2	102.89	103.72	0.26	3	0.85	102.89	0	1
Parotid	$3B_{60}$	28.1	24.17	38.89	2.82	1	24.49	32.12	1.41	1	1e-3	32.27	4.17	1
D_{mean}(Gy)	$4B_{120}$	26.39	23.87	38.06	3.06	3	22.35	30.73	1.57	3	1e-4	31.27	4.88	2
	5B	26.27	23.86	37.73	3.01	2	22.38	30.42	1.48	2	1e-4	31.19	4.93	3
Oral	$3B_{60}$	17.65	14.44	23.25	1.28	1	15.69	21.34	1.01	3	0.17	18.63	0.98	3
D_{mean}(Gy)	$4B_{120}$	16.77	13.7	20.58	1.57	3	14.92	20.47	1	2	0.11	17.69	0.93	2
	5B	16.67	13.63	20.42	1.51	2	14.78	20.36	0.97	1	0.11	17.56	0.89	1
Larynx	$3B_{60}$	30.92	26.59	36.07	1.89	1	28.95	34.59	1.04	1	0.9	31.57	0.64	1
D_{mean}(Gy)	$4B_{120}$	32.33	24.09	41.39	3.01	3	28.8	39.73	2.1	3	0.12	36.49	4.16	3
	5B	28.8	22.69	34.62	2.2	2	26.51	33.41	1.24	2	0.34	31.06	2.27	2
Cochlea	$3B_{60}$	18.75	15.33	28.42	2.09	1	14.04	25.52	2.33	3	0.66	18.91	0.16	1
D_{mean}(Gy)	$4B_{120}$	20.49	17.18	27.62	2.31	3	17.04	26.75	2.11	2	0.98	20.78	0.29	2
	5B	21.13	17.98	28.24	2.23	2	17.9	27.23	2.05	1	0.89	21.46	0.33	3
Brainstem	$3B_{60}$	39.62	36.43	52.81	5.98	3	29.22	46.82	3.55	2	1e-6	38.38	-1.24	1
D_{max}(Gy)	$4B_{120}$	37.12	28.64	53.18	5.13	2	26.82	46.12	3.6	3	0.09	35.04	-2.08	3
	5B	37.59	30.04	52.92	4.85	1	27.89	45.09	3.28	1	0.04	35.68	-1.91	2
Spinal	$3B_{60}$	32.66	18.37	43.91	4.96	1	23.86	42.56	3.44	1	0.22	40.04	7.38	3
D_{max}(Gy)	$4B_{120}$	27.37	18.84	41.85	5.13	2	19.47	37.97	3.54	3	0.1	33.59	6.23	1
	5B	28.22	19.74	43.03	5.26	3	20.42	38.82	3.45	2	0.09	35	6.78	2
Chiasm	$3B_{60}$	37.02	39.55	59.74	16.08	3	24.92	47.62	4.1	3	1e-21	38.31	1.29	1
D_{max}(Gy)	$4B_{120}$	34.27	34.81	56.98	14	2	25.17	44.87	3.98	1	1e-17	37.08	2.81	3
	5B	35.45	34.6	56.7	13.66	1	24.85	45.85	4.03	2	1e-16	37.88	2.43	2
Optic nerve left	$3B_{60}$	38.86	23.5	56.91	6.32	3	26.36	49.76	5.48	2	0.86	38.49	-0.37	1
D_{max}(Gy)	$4B_{120}$	40.87	27.07	54.72	5.54	1	27.67	51.77	5.33	1	0.82	39.42	-1.45	2
	5B	39.68	26.87	54.36	5.7	2	25.88	50.98	5.52	3	0.91	38.21	-1.47	3
Optic nerve right	$3B_{60}$	38.86	23.5	59.41	6.68	3	25.76	51.76	5.04	3	0.34	37.62	-1.25	2
D_{max}(Gy)	$4B_{120}$	40.87	29.48	54.72	5.29	1	27.87	51.47	4.57	1	0.61	38.55	-2.32	3
	5B	39.68	27.49	54.36	5.56	2	26.58	50.58	4.72	2	0.63	38.67	-1.01	1

Appendix C

Anatomical model

C.1 Individual Cases

The WSLD for each model in each test patient and week is listed in table B1. The bold text marks the closest WSLD number estimated from models to the real WSLD number from rCT.

C.2 Sensitivity of model

Because the AM is the basic model, the sensitivity measurement was conducted on the AM using WSLD. To measure the sensitivity of the AM to the training data, we randomly selected one test patient. Then, we randomly selected 5x15 patients from the remaining 19 patients as the training set, resulting in 5 groups of sensitivity training data for the test patient. For each sensitivity training data set, the WSLD in the test patient was calculated ($WSLD_{AM}^{sensitivity}$). The 95% CI of $WSLD_{AM}^{sensitivity}$ is used as a measure for the sensitivity to the training data. The result shows only small differences between the 5 groups. Because there is a 15/19 chance that the sensitivity training data include the data used in the original training dataset, this measure only represents the sensitivity of the model based on this cohort of patients. Another group of patient data is required to fully measure the sensitivity of the model to the small set of training data.

Table B1: WSLD for each model in each test patient and week. The numbers of the models that are the closest to the corresponding number of rCT (no model) are highlighted in bold as the best prediction.

ID	rCT(No model)(mm)	AM(mm)	RIM_2(mm)	RIM_3(mm)
id = 1				
week1	1.72	0.57	-	-
week2	2.52	0.84	-	-
week3	3.43	1.18	**1.78**	-
week4	4.93	1.65	2.11	**2.92**
week5	5.62	1.70	2.18	**2.95**
week6	5.23	2.00	2.53	**3.25**
id = 2				
week1	1.96	0.60	-	-
week2	2.07	0.80	-	-
week3	2.15	1.06	**1.77**	-
week4	2.78	1.34	**2.00**	1.91
week5	2.73	1.47	**1.98**	1.91
week6	3.12	1.71	2.11	**2.15**
id = 3				
week1	2.57	0.66	-	-
week2	1.99	0.86	-	-
week3	2.59	1.32	**1.56**	-
week4	2.94	**1.91**	1.70	2.22
week5	4.02	1.79	1.56	**2.17**
week6	5.13	2.12	1.75	**2.24**
id = 4				
week1	2.32	0.73	-	-
week2	2.30	0.96	-	-
week3	2.57	1.23	**1.85**	-
week4	3.44	1.73	1.90	**2.00**
week5	2.91	1.87	2.01	**2.04**
week6	2.62	2.12	2.13	**2.22**
id = 5				
week1	1.89	0.65	-	-
week2	2.54	1.00	-	-
week3	3.49	1.53	**2.12**	-
week4	4.69	2.07	2.43	**2.89**
week5	5.57	2.10	2.49	**2.88**
week6	6.27	2.38	2.68	**3.07**

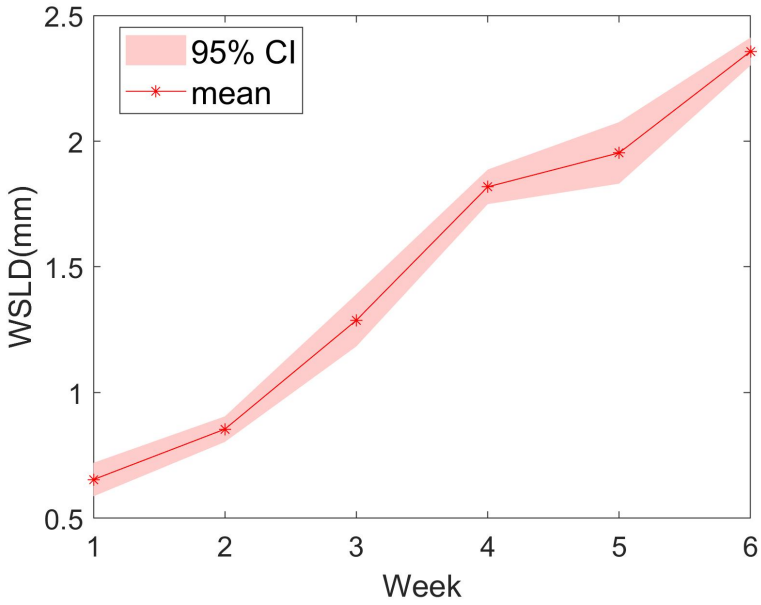


Figure C.1: The measurement of model sensitivity for AM. The WSLD of AM was calculated 5 times based on 5 groups of training data, referred to as $\text{WSLD}_{\text{AM}}^{\text{sensitivity}}$. The figure shows the mean value and 95% CI of the $\text{WSLD}_{\text{AM}}^{\text{sensitivity}}$.

Appendix D

Model application

D.1 Planning information

IMPT uses a multi-field optimization approach in which the target (both tumour, nodes and radiation area) is split for each beam field. All beam fields are optimised simultaneously so that their combined dose distribution covers the target volume homogeneously. In this study, IMPT treatment plans were generated using a five-field beam arrangement. The target split is shown in figure D.1. Beam directions and the corresponding target region are listed in table D1 with the beam database information.

Table D1: Beam directions and the corresponding target region. The colour is corresponding to the region illustrated in figure D.1. Beam database information is listed in the note.

Beam angle	Target region
60	grey, red
110	grey, red
180	orange, red
250	orange, grey
300	orange, grey
Note: spot spacing size: 5mm; energy range: 70 MeV – 250MeV.	

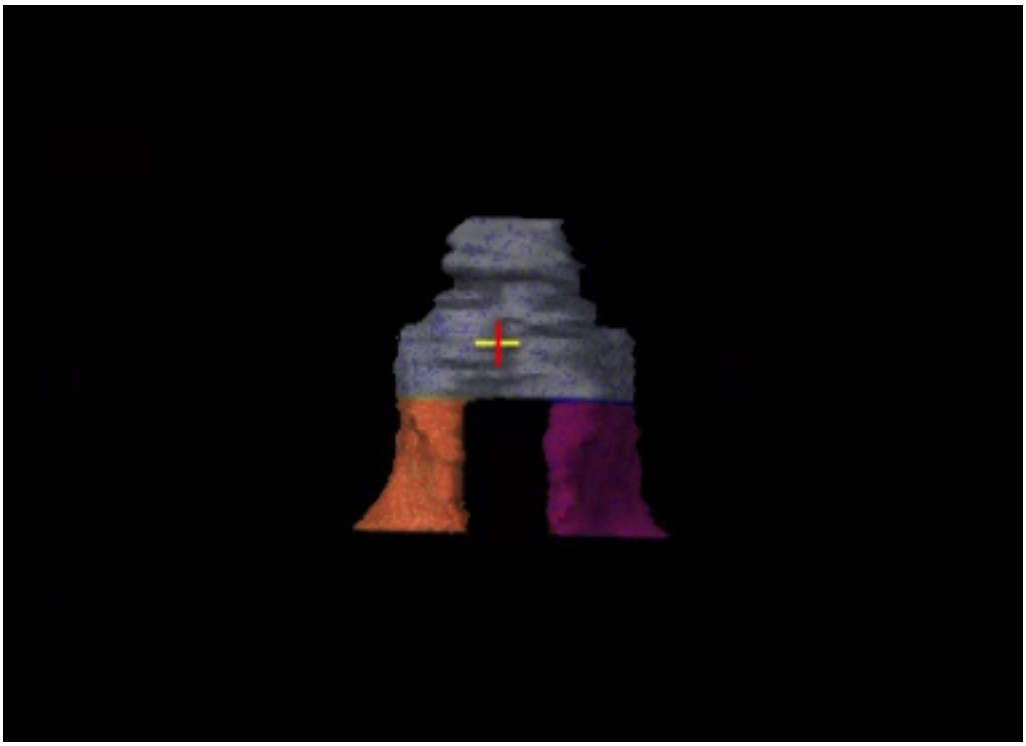


Figure D.1: Target split illustration.

D.2 Results of the model's predictive power evaluation on adaptation week.

The real week that requires a replan in the standard replan and the predicted week of replan using the predicted images of the RIM are listed in table D2.

Table D2: The real week that requires a replan in the standard replan and the predicted week of replan using the predicted images of the RIM.

ID	standard replan week	predicted replan week
1	None	None
2	3	3
3	4	4
4	3	4
5	3	4
6	3	4
7	4	3
8	5	6
9	3	3
10	3	4

Bibliography

[1] Ratko TA, Douglas GW, and de Souza JA. Radiotherapy treatments for head and neck cancer update. *Journal of the National Cancer Institute*, 2014.

[2] Marloes Steneker, Antony Lomax, and Uwe Schneider. Intensity modulated photon and proton therapy for the treatment of head and neck tumors. *Radiotherapy and Oncology*, 80(2), 2006.

[3] Tara A Van De Water, Antony J Lomax, Hendrik P Bijl, Marije E De Jong, Cornelis Schilstra, Eugen B Hug, and Johannes A Langendijk. Potential benefits of scanned intensity-modulated proton therapy versus advanced photon therapy with regard to sparing of the salivary glands in oropharyngeal cancer. *International Journal of Radiation Oncology* Biology* Physics*, 79(4), 2011.

[4] S. E. McGowan, N. G. Burnet, and A. J. Lomax. Treatment planning optimisation in proton therapy, 2013.

[5] Hideki Minatogawa, Koichi Yasuda, Yasuhiro Dekura, Seishin Takao, Taeko Matsuura andTakaaki Yoshimura, Ryusuke Suzuki, and et al. Potential benefits of adaptive intensity-modulated proton therapy in nasopharyngeal carcinomas. *Journal of applied clinical medical physics*, 22(1), 2021.

[6] Coen R N Rasch, Roel J H M Steenbakkers andIsabelle Fitton, Joop C Duppen, Peter J C M Nowak, Frank A Pameijer, Avraham Eisbruch, Johannes H A M Kaanders, Frank Paulsen, and Marcel van Herk. Decreased 3D observer

variation with matched CT-MRI, for target delineation in Nasopharynx cancer. *Radiation oncology*, 5(1), 2010.

[7] Marianne C Aznar, Theodore Girinsky, Anne Kiil Berthelsen, Berthe Aleman, Max Beijert, Martin Hutchings, Yolande Lievens, Paul Meijnders, Peter Meidahl Petersen, Deborah Schut, Maja V Maraldo, Richard van der Maazen, and Lena Specht. Interobserver delineation uncertainty in involved-node radiation therapy (INRT) for early-stage Hodgkin lymphoma: on behalf of the Radiotherapy Committee of the EORTC lymphoma group. *Acta Oncologica*, 56(4), 2017.

[8] Stefan J van Hoof, Joana B Verde, and Frank Verhaegen. Dose painting by dynamic irradiation delivery on an image-guided small animal radiotherapy platform. *The British journal of radiology*, 92(1095), 2019.

[9] Michele Avanzo, Joseph Stancanello, Giovanni Franchin, Giovanna Sartor, Rajesh Jena Annalisa Drigo, Andrea Dassie, Marco Gigante, and Elvira Capra. Correlation of a hypoxia based tumor control model with observed local control rates in nasopharyngeal carcinoma treated with chemoradiotherapy. *Medical physics*, 37(4), 2010.

[10] Karen Lim, Philip Chan, Robert Dinniwell, and Anthony Fyles. Cervical cancer regression measured using weekly magnetic resonance imaging during fractionated radiotherapy: radiobiologic modeling and correlation with tumor hypoxia. *International Journal of Radiation Oncology Biology Physics*, 70(1), 2008.

[11] Kevin Leder, Ken Pitter, Quincey LaPlant, Dolores Hambardzumyan, Brian D Ross, Timothy A Chan, Eric C Holland, and Franziska Michor. Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, 156(3), 2014.

[12] David G Kirsch, Max Diehn, Aparna H Kesarwala, Amit Maity, Meredith A Morgan, Julie K Schwarz, Robert Bristow, Sandra Demaria, Iris Eke,

Robert J Griffin, Daphne Haas-Kogan, Geoff S Higgins, Alec C Kimmelman, Randall J Kimple, Isabelle M Lombaert, Li Ma, Brian Marples, Frank Pajonk, Catherine C Park, Dörthe Schaue, Phuoc T Tran, Henning Willers, Brad G Wouters, and Eric J Bernhard. The future of radiobiology. *Journal of the National Cancer Institute*, 110(4), 2017.

[13] Ming Yang, X. Ronald Zhu, Peter C. Park, Uwe Titt, Radhe Mohan, Gary Virshup, James E. Clayton, and Lei Dong. Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. *Physics in Medicine and Biology*, 2012.

[14] H Paganetti. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in medicine and biology*, 57(11):R99, 2012.

[15] Pablo Yepes, Antony Adair, David Grosshans, Dragan Mirkovic, Falk Poenisch, Uwe Titt, Qianxia Wang, and Radhe Mohan2. Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy. *Physics in Medicine & Biology*, 63(4), 2018.

[16] Eric K Hansen, M Kara Bucci, Jeanne M Quivey, Vivian Weinberg, and Ping Xia. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*, 64(2), 2006.

[17] Tom Bruijnen, Bjorn Stemkens, Chris H J Terhaard, Jan J W Lagendijk, Cornelis P J Raaijmakers, and Rob H N Tijssen. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiotherapy and Oncology*, 130, 2019.

[18] Wenyong Tan, Yanping Li, and Guang Han. Target volume and position variations during intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma. *OncoTargets and therapy*, 6, 2013.

[19] Dan Fang Yan, Sen Xiang Yan, Qi Dong Wang, Xin Biao Liao, Zhong Jie Lu, and Yi Xiang Wang. Predictors for replanning in loco-regionally advanced nasopharyngeal carcinoma patients undergoing intensity-modulated radiation therapy: A prospective observational study. *BMC Cancer*, 2013.

[20] Shreerang A. Bhide, Mark Davies, Kevin Burke, Helen A. McNair, Vibeke Hansen, Y. Barbachano, I. A. El-Hariry, Kate Newbold, Kevin J. Harrington, and Christopher M. Nutting. Weekly Volume and Dosimetric Changes During Chemoradiotherapy With Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study. *International Journal of Radiation Oncology Biology Physics*, 2010.

[21] Aafke C.Kraan, Stevenvan de Water, David N.Teguh, AbrahimAl-Mamgani, Tom Madden, Hanne M.Kooy, Ben J.M.Heijmen, and Mischa S.Hoogeman. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. *International Journal of Radiation Oncology* Biology* Physics*, 87(5), 2013.

[22] Richard Y Wu, Amy Y Liu, Terence T Sio, Pierre Blanchard, Cody Wages, Mayankkumar V Amin, Gary B Gunn, Uwe Titt, Rong Ye, and Kazumichi Suzuki. Intensity-Modulated Proton Therapy Adaptive Planning for Patients with Oropharyngeal Cancer. *International Journal of Particle Therapy*, 4(2):26–34, 2017.

[23] Jolien Heukelom, Michael E Kantor, Abdallah S R Mohamed, Hesham Elhalawani, Esengul Kocak-Uzel, Timothy Lin, Jinzhong Yang, Michalis Aristophanous, Coen R Rasch, Clifton David, and FullerJan-Jakob Sonke. Differences between planned and delivered dose for head and neck cancer, and their consequences for normal tissue complication probability and treatment adaptation. *Radiotherapy and Oncology*, 142, 2019.

[24] Catarina Veiga, Ana Mónica Lourenço, Syed Mouinuddin, Marcel van Herk, Marc Modat, Sébastien Ourselin, Gary Royle, and Jamie R McClelland. To-

ward adaptive radiotherapy for head and neck patients: feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. *Medical physics*, 41(3), 2014.

[25] NiftyReg. Niftyreg groupwise registration.

[26] Molly M.McCulloch, Daniel G.Muenz, Matthew J.Schipper, Michael Velec, Laura A.Dawson, and Kristy K.Brock. A simulation study to assess the potential impact of developing normal tissue complication probability models with accumulated dose. *Advances in radiation oncology*, 3(4), 2018.

[27] Panagiotis Tsiamas, Hassan Bagher-Ebadian, Farzan Siddiqui, Chang Liu Christian A Hvid, Joshua P Kim, Stephen L Brown, Benjamin Movsas, and Indrin J Chetty. Principal component analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images. *Medical physics*, 45(12), 2018.

[28] Simon Van Kranen, Angelo Mencarelli, Coen Rasch Suzanne Van Beek, Marcel Van Herk, and Jan Jakob Sonke. Adaptive radiotherapy with an average anatomy model: Evaluation and quantification of residual deformations in head and neck cancer patients. *Radiotherapy and Oncology*, 109(3), 2013.

[29] Lena Nenoff, Cássia O.Ribeiro, Michael Matter Luana Hafner, Mirjana Josipovic, and Johannes A. Deformable image registration uncertainty for inter-fractional dose accumulation of lung cancer proton therapy. *Radiotherapy and Oncology*, 147, 2020.

[30] Catarina Veiga, Guillaume Janssens, Ching-Ling Teng, Thomas Baudier, Lucian Hotoiu, Jamie R McClelland, Gary Royle, Liyong Lin, Lingshu Yin, James Metz, Timothy D Solberg, Zelig Tochner, Charles B Simone, James McDonough, and Boon-Keng Kevin Teo. First clinical investigation of cone beam computed tomography and deformable registration for adaptive proton therapy for lung cancer. *International Journal of Radiation Oncology Biology Physics*, 95(1), 2016.

[31] Hualiang Zhong and Indrin J. Chetty. Caution Must Be Exercised When Performing Deformable Dose Accumulation for Tumors Undergoing Mass Changes During Fractionated Radiation Therapy. *International Journal of Radiation Oncology Biology Physics*, 97(1):182–183, 2017.

[32] Jihun Kim, Yang Kyun Park, Gregory Sharp, Paul Busse, and Brian Winey. Water equivalent path length calculations using scatter-corrected head and neck CBCT images to evaluate patients for adaptive proton therapy. *Physics in Medicine and Biology*, 62(1):59–72, 2017.

[33] S M Holloway, M D Holloway, and S J Thomas. A method for acquiring random range uncertainty probability distributions in proton therapy. *Physics in Medicine & Biology*, 63(1), 2017.

[34] Heng Li, Xiaodong Zhang, Peter Park, Wei Liu, Joe Chang, Zhong xing Liao, Steve Frank, Yupeng Li, Falk Poenisch, Radhe Mohan, Michael Gillin, and Ronald Zhu. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiotherapy and Oncology*, 114(3), 2015.

[35] Jan Unkelbach, Markus Alber, Mark Bangert, Rasmus Bokrantz, Timothy C.Y. Chan, Joseph O. Deasy, Albin Fredriksson, Bram L. Gorissen, Marcel Van Herk, Wei Liu, Houra Mahmoudzadeh, Omid Nohadani, Jeffrey V. Siebers, Marnix Witte, and Huijun Xu. Robust radiotherapy planning. *Physics in Medicine and Biology*, 63(22), 2018.

[36] 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. *International Journal of Radiation Oncology Biology Physics*, 2000.

[37] F. Albertini, E. B. Hug, and A. J. Lomax. Is it necessary to plan with safety margins for actively scanned proton therapy? *Physics in Medicine and Biology*, 2011.

[38] M Lowe, A Gosling, O Nicholas, T Underwood, E Miles, Y.-C. Chang, R A Amos, and N G Bu. Comparing proton to photon radiotherapy plans: UK consensus guidance for reporting under uncertainty for clinical trials. *Clinical Oncology*, 32(7), 2020.

[39] J Unkelbach and U Oelfke. Incorporating organ movements in inverse planning: assessing dose uncertainties by Bayesian inference. *Physics in Medicine and Biology*, 50(1), 2005.

[40] B Sobotta, M So¨hn, and M Alber. Robust optimization based upon statistical theory. *Medical physics*, 37(8), 2010.

[41] Rom´an Bohoslavsky, Marnix G Witte, Tomas M Janssen, and Marcel van Herk. Probabilistic objective functions for margin-less IMRT planning. *Physics in medicine and biology*, 58(11), 2013.

[42] Albin Fredriksson, Anders Forsgren, and Björn Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical Physics*, 38(3):1672–1684, 2011.

[43] Macarena Cubillos-Mesías, Esther G.C. Troost, Fabian Lohaus, Linda Agolli, Maximilian Rehm, Christian Richter, and Kristin Stützer. Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation. *Radiotherapy and Oncology*, 131:127–134, 2019.

[44] Zhiyong Yang, Xiaodong Zhang, Xianliang Wang, X. Ronald Zhu, Brandon Gunn, Steven J. Frank, Yu Chang, Qin Li, Kunyu Yang, Gang Wu, Li Liao, Yupeng Li, Mei Chen, and Heng Li. Multiple-CT optimization: An adaptive optimization method to account for anatomical changes in intensity-modulated proton therapy for head and neck cancers. *Radiotherapy and Oncology*, 142:124–132, 2020.

[45] Macarena Cubillos-Mesías, Esther G.C. Troost, Fabian Lohaus, Linda Agolli, Maximilian Rehm, Christian Richter, and Kristin Stützer. Quantifica-

tion of plan robustness against different uncertainty sources for classical and anatomical robust optimized treatment plans in head and neck cancer proton therapy. *The British journal of radiology*, 93(1107):20190573, 2020.

[46] Christina Hague, MRCP, FRCR Marianne Aznar, and et al. Inter-fraction robustness of intensity-modulated proton therapy in the post-operative treatment of oropharyngeal and oral cavity squamous cell carcinomas. *The British Journal of Radiology*, 93(1107), 2020.

[47] S. E. McGowan, F. Albertini, S. J. Thomas, and A. J. Lomax. Defining robustness protocols: A method to include and evaluate robustness in clinical plans. *Physics in Medicine and Biology*, 60(7):2671–2684, 2015.

[48] Arthur Lalonde, Mislav Bobić, Brian Winey, Joost Verburg, Gregory C.Sharp, and Harald Paganetti. Anatomic changes in head and neck intensity-modulated proton therapy: comparison between robust optimization and online adaptation. *Radiotherapy and Oncology*, 2021.

[49] Michael Matter, Lena Nenoff, Gabriel Meier, Damien C. Weber, Antony J. Lomax, and Francesca Albertini. Intensity modulated proton therapy plan generation in under ten seconds. *Acta Oncologica*, 58(10), 2019.

[50] P Botas, J Kim, B Winey, and H Paganetti. Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations. *Physics in Medicine and Biology*, 64(1), 2018.

[51] Mislav Bobić, Arthur Lalonde, Gregory C Sharp, Clemens Grassberger, Joost M Verburg, Brian A Winey, Antony John Lomax, and Harald Paganetti. Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy. *Physics in Medicine & Biology*, 2021.

[52] Kazumichi Suzuki, Michael T. Gillin, Narayan Sahoo, X. Ronald Zhu, Andrew K. Lee, and Denise Lippy. Quantitative analysis of beam delivery

parameters and treatment process time for proton beam therapy. *Medical Physics*, 38(7):4329–4337, 2011.

[53] J.Price, E.Hall, C.West, and D.Thomson. TORPEdO-A Phase III Trial of Intensity-modulated Proton Beam Therapy Versus Intensity-modulated Radiotherapy for Multi-toxicity Reduction in Oropharyngeal Cancer. *Clinical oncology (Royal College of Radiologists (Great Britain))*, 32(2), 2020.

[54] Charlotte L.Brouwer, Roel J H M Steenbakkers, Arjen van der Schaaf, Chantal T.C.Sopacua, Lisanne V.van Dijk, Roel G.J.Kierkels, Hendrik P.Bijl, Johannes G.M.Burgerhof, Johannes A.Langendijk, and Nanna M.Sijtsema. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiotherapy and Oncology*, 120(1), 2016.

[55] Trille Kjaer, Susanne Oksbjerg Dalton, Elo Andersen, Randi Karlsen, Anni Linnet Nielsen, Merete Kjaer Hansen, Kirsten Frederiksen, and Christoffer Johansen. A controlled study of use of patient-reported outcomes to improve assessment of late effects after treatment for head-and-neck cancer. *Radiotherapy and Oncology*, 119(2), 2016.

[56] Simon van Kranen, Olga Hamming-Vrieze, Annelisa Wolf, Eugène Damen, Marcel van Herk, and Jan Jakob Sonke. Head and neck margin reduction with adaptive radiation therapy: robustness of treatment plans against anatomy changes. *International Journal of Radiation Oncology* Biology* Physics*, 96(3), 2016.

[57] Sarah Weppler, Harvey Quon, Robyn Banerjee, Colleen Schinkel, and Wendy Smith. Framework for the quantitative assessment of adaptive radiation therapy protocols. *Journal of Applied Clinical Medical Physics*, 19(6), 2018.

[58] Claudio Fiorino, Giovanna Rizzo, Elisa Scalco, and Sara Broggi. Density variation of parotid glands during IMRT for head–neck cancer: Correla-

tion with treatment and anatomical parameters. *Radiotherapy and Oncology*, 104(2), 2012.

[59] BFCO. *Timely delivery of radical radiotherapy: guidelines for the management of unscheduled treatment interruptions*. The Royal College of Radiologists, fourth edition edition, 2019.

[60] Z. Henry Yu, Rajat Kudchadker, Lei Dong, Yongbin Zhang, Laurence E. Court, Firas Mourtada, Adam Yock, Susan L. Tucker, and Jinzhong Yang. Learning anatomy changes from patient populations to create artificial CT images for voxel-level validation of deformable image registration. *Journal of Applied Clinical Medical Physics*, 2016.

[61] Sara Thörnqvist, Liv B Hysing, Andras G Zolnay, Matthias Söhn, Mischa S.Hoogeman, Ludvig P Muren, and Ben J M Heijmen. Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model. *Acta Oncologica*, 52(7), 2013.

[62] Huijun Xu, Douglas J.Vile, Manju Sharma, J.James Gordon, and Jeffrey V.Siebers. Coverage-based treatment planning to accommodate deformable organ variations in prostate cancer treatment. *Medical physics*, 41(10), 2014.

[63] M A Chetvertkov, F Siddiqui, I Chetty J. Kim, A Kumarasiri, C Liu, and J J Gordon. Use of regularized principal component analysis to model anatomical changes during head and neck radiation therapy for treatment adaptation and response assessment. *Medical Physics*, 43(10), 2016.

[64] Kristy K. Brock, Sasa Mutic, Todd R. McNutt, Hua Li, and Marc L. Kessler. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132: Report. *Medical Physics*, 44(7), 2017.

[65] Jan Ehrhardt, René Werner, Alexander Schmidt-Richberg, and Heinz Handels. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE transactions on medical imaging*, 30(2), 2010.

[66] Brian C Hall. *Lie algebras and the exponential mapping*. Lie Groups, Lie Algebras, and Representations. Springer, New York, NY, 2003.

[67] Cheuk Fai Lui. Density heterogeneity tool for optimising beam angle selection in proton therapy, 2018.

[68] Daniel A Low, William B Harms, Sasa Mutic, and James A Purdy. A technique for the quantitative evaluation of dose distributions. *Medical physics*, 25(5), 1998.

[69] Lukasz Szczurek, Robert Juszkat, Jolanta Szczurek, Ilona Turek, and Piotr Sosnowski. Pre-treatment 2D and 3D dosimetric verification of volumetric arc therapy. A correlation study between gamma index passing rate and clinical dose volume histogram. *PLoS ONE*, 14(8), 2019.

[70] X. Jin, H. Yan, C. Han, Y. Zhou, J. Yi, and Congying Xie. Correlation between gamma index passing rate and clinical dosimetric difference for pretreatment 2D and 3D volumetric modulated arc therapy dosimetric verification. *British Journal of Radiology*, 88(1047), 2015.

[71] Nataniel H Lester-Coll and Danielle N Margalit. Modeling the Potential Benefits of Proton Therapy for Patients With Oropharyngeal Head and Neck Cancer. *International journal of radiation oncology, biology, physics*, 104(3), 2019.

[72] Tara A Water, Hendrik P Bijl, Cornelis Schilstra, Madelon Pijls-Johannesma, and Johannes A Langendijk. The Potential Benefit of Radiotherapy with Protons in Head and Neck Cancer with Respect to Normal Tissue Sparing: A Systematic Review of Literature. *The oncologist*, 16(3), 2011.

[73] Robert Malyapa, Matthew Lowe, Alessandra Bolsi, Antony J Lomax, Damien C Weber, and Francesca Albertini. Evaluation of robustness to setup and range uncertainties for head and neck patients treated with pencil beam scanning proton therapy. *International Journal of Radiation Oncology* Biology* Physics*, 95(1), 2016.

[74] Jerry L. Barker, Adam S. Garden, K. Kian Ang, Jennifer C. O'Daniel, He Wang, Laurence E. Court, William H. Morrison, David I. Rosenthal, K. S.Clifford Chao, Susan L. Tucker, Radhe Mohan, and Lei Dong. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *International Journal of Radiation Oncology Biology Physics*, 2004.

[75] Choonik Lee, Katja M Langen, Weiguo Lu, Jason Haimerl, Eric Schnarr, Kenneth J Ruchala, Gustavo H Olivera, Sanford L Meeks, Patrick A Kupelian, Thomas D Shellenberger, and Rafael R Mañon. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. *Radiotherapy and Oncology*, 89(1), 2008.

[76] Kristin Stützer, Annika Jakobi, Anna Bandurska-Luque, Steffen Barczyk, Carolin Arnsmeyer, Steffen Löck, and Christian Richter. Potential proton and photon dose degradation in advanced head and neck cancer patients by intratherapy changes. *Journal of Applied Clinical Medical Physics*, 2017.

[77] Victor Hernandez, Christian Rønn Hansen, Lamberto Widesott, Anna Bäck, Richard Canters, Marco Fusella, Julia Götstedt, Diego Jurado-Bruggeman, Nobutaka Mukumoto, Laura Patricia Kaplan, Irena Koniarová, Tomasz Piotrowski, Lorenzo Placidi, Ana Vaniqui, and Nuria Jornet. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans, 2020.

[78] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: Principal component analysis. *Nat Methods*, 14, 2017.

[79] Wenyong Tan, Jianzeng Ye, Ruilian Xu, Xianming Li, Wan He, Xiaohong Wang, Yanping Li, and Desheng Hu. The tumor shape changes of nasopharyngeal cancer during chemoradiotherapy: The estimated margin to cover the

geometrical variation. *Quantitative Imaging in Medicine and Surgery*, 6(2), 2016.

[80] J Unkelbach and U Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Physics in Medicine & Biology*, 49(17), 2004.

[81] Alessandra Bolsi, Antony J Lomax, Eros Pedroni, Gudrun Goitein, and Eugen Hug. Experiences at the Paul Scherrer Institute with a remote patient positioning procedure for high-throughput proton radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 71(5), 2008.

[82] Martin J Murphy, Francisco J Salguero, Jeffrey V Siebers, David Staub, and Constantin Vaman. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Medical physics*, 39(2), 2012.

[83] Xin Zhen, Hao Yan, Linghong Zhou, Xun Jia, and Steve B. Jiang. Deformable image registration of CT and truncated cone-beam CT for adaptive radiation therapy. In *Physics in Medicine and Biology*, volume 58, 2013.

[84] Catarina Veiga, Jamie McClelland, Syed Moinuddin, Ana Lourenço, Kate Ricketts, James Annkah, Marc Modat, Sébastien Ourselin, Derek D'Souza, and Gary Royle. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. *Medical Physics*, 41(3), 2014.

[85] Jong Min Park, Jung In Kim, So Yeon Park, Do Hoon Oh, and Sang Tae Kim. Reliability of the gamma index analysis as a verification method of volumetric modulated arc therapy plans. *Radiation Oncology*, 13(1), 2018.

[86] Liting Yu, Timothy L.S. Tang, Naasiha Cassim, Alexander Livingstone, Darren Cassidy, Tanya Kairn, and Scott B. Crowe. Analysis of dose comparison techniques for patient-specific quality assurance in radiation therapy. *Journal of Applied Clinical Medical Physics*, 20(11), 2019.

[87] Marta Bogowicz, Stephanie Tanadini-Lang, Matthias Guckenberger, and Oliver Riesterer. Combined CT radiomics of primary tumor and metastatic lymph nodes improves prediction of loco-regional control in head and neck cancer. *Scientific reports*, 9(1), 2019.

[88] Pernille Lassen, Jesper G Eriksen, Stephen Hamilton-Dutoit, Trine Tramm, Jan Alsner, and Jens Overgaard. Effect of HPV-associated p16INK4A expression on response to radiotherapy and survival in squamous cell carcinoma of the head and neck. *J Clin Oncol*, 27(12), 2009.

[89] Lisanne V.van Dijk, Charlotte L.Brouwer, Arjenvan der Schaaf, Johannes G.M.Burgerhof, Roelof J.Beukinga, Johannes A.Langendijk, Nanna M.Sijtsema, and Roel J.H.M.Steenbakkersa. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiotherapy and Oncology*, 122(2), 2017.

[90] Hubert S Gabryś, imageFlorian Buettner, imageFlorian Sterzing, imageHenrik Hauswald, and imageMark Bangert. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. *Frontiers in oncology*, 8, 2018.

[91] Davide Brivio, Yida David Hu, Danielle N. Margalit, and Piotr Zygmanski. Selection of head and neck cancer patients for adaptive replanning of radiation treatment using kV-CBCT. *Biomedical Physics and Engineering Express*, 4(5), 2018.

[92] Anne W. Lee, Wai Tong Ng, Jian Ji Pan, Chi Leung Chiang, Sharon S. Poh, Horace C. Choi, Yong Chan Ahn, Hussain AlHussain, June Corry, Cai Grau, Vincent Grégoire, Kevin J. Harrington, Chao Su Hu, Dora L. Kwong, Johannes A. Langendijk, Quynh Thu Le, Nancy Y. Lee, Jin Ching Lin, Tai Xiang Lu, William M. Mendenhall, Brian O'Sullivan, Enis Ozyar, Lester J. Peters, David I. Rosenthal, Giuseppe Sanguineti, Yoke Lim Soong, Yungan Tao, Sue S. Yom, and Joseph T. Wee. International Guideline on Dose Priori-

tization and Acceptance Criteria in Radiation Therapy Planning for Nasopharyngeal Carcinoma. *International Journal of Radiation Oncology Biology Physics*, 105(3):567–580, 2019.

[93] Gary D. Lewis, Emma B. Holliday, Esengul Kocak-Uzel, Mike Hernandez, Adam S. Garden, David I. Rosenthal, and Steven J. Frank. Intensity-modulated proton therapy for nasopharyngeal carcinoma: Decreased radiation dose to normal structures and encouraging clinical outcomes. *Head and Neck*, 38, 2016.

[94] Jaden D Evans, Riley H Harper, and Molly Petersen. The Importance of Verification CT-QA Scans in Patients Treated with IMPT for Head and Neck Cancers. *International journal of particle therapy*, 7(1), 2020.

[95] Qiuwen Wu, Yuwei Chi, Peter Y. Chen, Daniel J. Krauss, Di Yan, and Alvaro Martinez. Adaptive Replanning Strategies Accounting for Shrinkage in Head and Neck IMRT. *International Journal of Radiation Oncology Biology Physics*, 75(3):924–932, 2009.

[96] Elizabeth Brown, Rebecca Owen, Fiona Harden, Kerrie Mengersen, Kimberley Oestreich, Whitney Houghton, Michael Poulsen, Selina Harris, Charles Lin, and Sandro Porceddu. Head and neck adaptive radiotherapy: Predicting the time to replan. *Asia-Pacific Journal of Clinical Oncology*, 12(4), 2016.

[97] Q. Wu, S.R. Nandalur, K.R. Nandalur, D.J. Krauss, P.Y. Chen, D. Yan, and A.A. Martinez. What Time is Optimal for Replanning Head and Neck IMRT (HN-IMRT)? *International Journal of Radiation Oncology*Biology*Physics*, 72(1), 2008.

[98] David J Noble, Ping Lin Yeap, Shannon Y K Seah, Karl Harrison, Leila E A Shelley, Marina Romanchikova, Amy M Bates, Yaolin Zheng, Gillian C Barnett, Richard J Benson, Sarah J Jefferies, Simon J Thomas, Raj Jena, and Neil G Burnet. Anatomical change during radiotherapy for head and neck

cancer, and its effect on delivered dose to the spinal cord. *Radiotherapy and Oncology*, 130, 2019.

[99] Jasper Kouwenberg, Joan Penninkhof, Steven Habraken, Jaap Zindler, Mischa Hoogeman, and Ben Heijmena. Model based patient pre-selection for intensity-modulated proton therapy (IMPT) using automated treatment planning and machine learning. *Radiotherapy and Oncology*, 6, 2021.

[100] Yingzi Liu, Yang Lei, Tonghe Wang, Yabo Fu, Xiangyang Tang, Walter J. Curran, Tian Liu, Pretesh Patel, and Xiaofeng Yang. CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy. *Medical Physics*, 47(6), 2020.

[101] Richard Qiu, Yang Lei, Aparna Kesarwala, Kristin Higgins, Jeffrey D. Bradley, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Chest CBCT-based synthetic CT using cycle-consistent adversarial network with histogram matching. 2021.

[102] Zhi Cheng, Minoru Nakatsugawa, Chen Hu, Scott P. Robertson, Xuan Hui, Joseph A. Moore, Michael R. Bowers, Ana P. Kiess, Brandi R. Page, Laura Burns, Mariah Muse, Amanda Choflet, Kousuke Sakaue, Shinya Sugiyama, Kazuki Utsunomiya, John W. Wong, Todd R. McNutt, and Harry Quon. Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy. *Advances in Radiation Oncology*, 3(3), 2018.

[103] Joel Jaskari, Jaakko Sahlsten, Jorma Järnstedt, Helena Mehtonen, Kalle Karhu, Osku Sundqvist, Ari Hietanen, Vesa Varjonen, Vesa Mattila, and Kimmo Kaski. Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes. *Scientific Reports*, 10(1), 2020.

[104] C Veiga, J McClelland, S Moinuddin, K Ricketts, M Modat, S Ourselin, D D'Souza, and G Royle. Towards adaptive radiotherapy for head and neck

patients: validation of an in-house deformable registration algorithm. *Journal of Physics: Conference Series. IOP Publishing*, 489(1), 2014.