University of Zurich[UZH]

The potential of daily adaptive proton therapy in the case of deforming anatomies

Master Thesis in Physics

Thallis Alves Santos

Supervised by
Prof. Dr. Jan Unkelbach, UZH
Dr. Francesca Albertini, PSI
Advisor
Dr. Katarzyna Czerska, PSI

March 9, 2025

Abstract

This thesis investigates the efficacy of adaptive proton therapy for head and neck cancer patients, emphasizing the critical aspects of target coverage and organ-at-risk (OAR) sparing. The research evaluates the applicability of existing structure quality assurance (QA) protocol in the context of deformable registration, considering whether current methods are suitable for handling anatomical changes during treatment. Additionally, the study compares the dose distributions of different adaptive treatment configurations, implemented through two treatment planning systems (TPS) and the two available Gantries (G2 and G3) at the Paul Scherrer Institute (PSI). By analyzing a series of five patient cases, this study provides in-depth insights into the advantages and limitations of adaptive proton therapy. The results reveal that adaptive proton therapy consistently outperforms conventional treatment approaches by offering significant improvements in daily simulated delivered dose, better tumor coverage, and enhanced sparing of critical structures. These findings were statistically validated through the Wilcoxon test, confirming the adaptive strategy's superiority over traditional methods in all tested configurations (adaptive vs. conventional treatments). Notably, while adaptation in G2 demonstrated superior dose distributions in most cases, adaptation in G3 also showed promising results, with satisfactory simulated treatment dose distributions and performing better than conventional approaches. This research underscores the potential of adaptive proton therapy to optimize treatment for head and neck cancer patients, particularly in the face of anatomical changes or uncertainties in patient positioning during the course of treatment.

Keywords: adaptive proton therapy, head and neck cancer, treatment planning systems, deformable registration, proton therapy

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Jan Unkelbach and Francesca Albertini, for their guidance throughout this journey. Thanks to all the colleagues at the Paul Scherrer Institute for their invaluable support and encouragement throughout this work. I am especially grateful to Katarzyna for her constructive and detailed feedback, as well as her insightful discussions, which significantly enriched this project. Her support came at a crucial time and was greatly appreciated.

A special thanks goes to my former colleagues Márcio, Bruna, and Tamara, who, despite not being with me in person now, taught me so much and never doubted my success in this new foreign country.

Although my family is far away, their love and encouragement have been a constant source of strength. I've lost count of how many times I spent the 2 hours commuting talking to my mom. I am deeply thankful for their unwavering support, even from a distance. I love you all so much.

I would also like to express special and infinite thanks to my girlfriend, Débora, whose emotional (and financial) support has been crucial throughout this journey. Her patience, encouragement, and belief in me have been a constant source of motivation. I would not be here without her.

As the song says, "*I gotta have roots before branches, to know who I am before I know who I wanna be, and faith to take chances, to live like I see: a place in this world for me.*". I've found my roots, taken chances, and built the path forward, all thanks to the people who crossed my path.

Finally, I acknowledge the collaborative efforts and resources provided by the CPT, which were essential for the completion of this work.

Contents

I Introduction 1

1 Motivation 1

II Materials and Methods 4

2 Gantry 2 x Gantry 3 5
2.1 Gantry 2 . . . 5
2.2 Gantry 3 . . . 6

3 Cases 7
3.1 Case 1 . . . 8
3.2 Case 2 . . . 9
3.3 Case 3 . . . 10
3.4 Case 4 . . . 12
3.5 Case 5 . . . 13

4 Treatment Plans in FIonA 14
4.1 Structure QA . . . 14
4.2 Adaptive simulation process . . . 15

5 Treatment plans in RayStation® 15
5.1 Performing adaptive treatment simulation in RayStation® . . . 16

6 Statistical Analysis 17
6.1 Deviation analysis . . . 17
6.2 Wilcoxon test . . . 17

III Results 17

7 Structure QA 18
7.1 Case with good propagation results . . . 18
7.2 Case with challenging propagation results . . . 19

8 Comparison between adaptive and non-adaptive approaches 21
8.1 Comparison between adaptive in G2 (full RO) vs non adaptive in G2 and non adaptive in G3 . . . 21
8.1.1 Case with favorable adaptive performance . . . 21
8.1.2 Case with limited adaptive performance . . . 23
8.1.3 Summary of the second and third configuration . . . 24
8.2 Comparison between adaptive G2 vs adaptive in G3 vs non adaptive in G3 . . . 26
8.2.1 Case with favorable adaptive performance . . . 26
8.2.2 Case with limited adaptive performance . . . 28
8.2.3 Summary of third and fourth configuration . . . 29

9 Comparison between FIonA and RayStation **30**
9.1 Case with favorable adaptive performance 31
9.2 Case with limited adaptive performance 33
9.3 Summary of the first configuration 36

IV Discussion **38**

V Conclusions & Outlook **40**

VI Appendix **41**

A OAR metrics **41**

B Structure QA - all cases **43**
B.1 Case 2 43
B.2 Case 4 44
B.3 Case 5 45

C Comparison between different configurations - all cases **47**
C.1 Comparison between adaptive in G2 (full RO) vs non adaptive in G2 and non adaptive in G3 47
C.1.1 Case 1 47
C.1.2 Case 3 48
C.1.3 Case 5 50
C.2 Comparison between adaptive G2 vs adaptive in G3 vs non adaptive in G3 . . . 52
C.2.1 Case 1 52
C.2.2 Case 3 53
C.2.3 Case 5 54
C.3 Comparison between FIonA and RayStation 56
C.3.1 Case 1 56
C.3.2 Case 3 59
C.3.3 Case 5 61

D Dose Volume Histograms **65**

List of Figures

1 Difference of the depth dose curve with and without a 2 cm material of different density for photons (left) and protons (right). 1
2 Workflow diagram of a generic daily adaptive proton therapy treatment [1]. . . . 2
3 Workflow scheme of DAPT at PSI. The gray phases are performed in ADAPT [1]. 3
4 In-house developed Gantry 2. 6
5 Commercial Gantry 3. 6
6 The split-target technique, here presented for Planning Target Volume (PTV) but the same logic can be applied to CTV [2]. 7
7 Fields configuration and planned dose distribution for Case 1 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. 8
8 Fields configuration and planned dose distribution for Case 2 (Image taken from FIonA). CTV_5412 in blue, and CTV_6996 in orange. 10
9 Fields configuration and planned dose distribution for Case 3 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. 11
10 Rigid (left) and deformable (right) registration of daily synCT and planning CT for fraction fraction 10 of Case 3 (between daily synCT and planning CT). The change in head and chin position is evident in the case of rigid image registration (RIR). 11
11 Fields configuration and planned dose distribution for Case 4 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. The arrows are pointing to Carotids. 12
12 Fields configuration and planned dose distribution for Case 5 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. The arrows are pointing to submandibulars. 13
13 Absolute volume differences for Case 1 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc. 19
14 Relative volume differences for structures bigger than 10 cc for Case 1 and all simulated daily fractions, displayed from the smallest structure (left) to the largest (right). The red line indicates the QA threshold of 10%. 19
15 Absolute volume differences for Case 3 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc. 20
16 Relative volume differences for structures bigger than 4 cc for Case 3 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%. . 20
17 Case 4: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels 22
18 Case 4: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 22
19 Case 2: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels 23
20 Case 2: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 24

21 Case 4: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels. 27
22 Case 4: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 27
23 Case 2: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels. 28
24 Case 2: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 29
25 FIonA daily target dose differences, D98% and V95%, for Case 4. 31
26 RayStation® daily target dose differences, D98% and V95%, for Case 4. 32
27 FIonA daily dose differences for some selected organs of Case 4, the red line represents the dose constraint defined in the clinical goal. 32
28 RayStation® daily dose differences for some selected organs of Case 4, the red line represents the dose constraint defined in the clinical goal.. 33
29 FIonA daily target dose differences, D98% and V95%, for Case 2. 34
30 RayStation® daily target dose differences, D98% and V95%, for Case 2. 34
31 FIonA daily dose differences for some selected organs of Case 2, the red line represents the dose constraint defined in the clinical goal. 35
32 RayStation® daily dose differences for some selected organs of Case 2, the red line represents the dose constraint defined in the clinical goal. 35
33 Absolute volume differences for Case 2 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc. 43
34 Relative volume differences for structures bigger than 10 cc for Case 2 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%. . 43
35 Absolute volume differences for Case 4 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc. 44
36 Relative volume differences for all structures for Case 4 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%. 45
37 Absolute volume differences for Case 5 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc. 46
38 Relative volume differences for structures bigger than 8 cc for Case 5 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%. . 46
39 Case 1: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels 47
40 Case 1: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 48
41 Case 3: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels 49

42 Case 3: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 49
43 Case 5: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels 51
44 Case 5: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 51
45 Case 1: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels. 52
46 Case 1: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 52
47 Case 3: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels. 53
48 Case 3: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 54
49 Case 5: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels. 55
50 Case 5: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal. 55
51 FIonA daily target dose differences, D98% and V95%, for Case 1. 56
52 RayStation® daily target dose differences, D98% and V95%, for Case 1. 57
53 FIonA daily dose difference for some selected organs of Case 1, the red line represents the dose constraint defined by MD. 58
54 RayStation® daily dose difference for some selected organs of Case 1, the red line represents the dose constraint prescription defined in the clinical goal. 58
55 FIonA daily target dose differences, D98% and V95%, for Case 3. 59
56 RayStation® daily target dose differences, D98% and V95%, for Case 3. 59
57 FIonA daily dose differences for some selected organs of Case 3, the red line represents the dose constraint defined in the clinical goal. 60
58 RayStation® daily dose differences for some selected organs of Case 3, the red line represents the dose constraint defined in the clinical goal. 61
59 FIonA daily target dose differences, D98% and V95%, for Case 5. 62
60 RayStation® daily target dose differences, D98% and V95%, for Case 5. 62
61 FIonA daily dose differences for some selected organs of Case 5, the red line represents the dose constraint defined in the clinical goal. 63
62 RayStation® daily dose differences for some selected organs of Case 5, the red line represents the dose constraint defined in the clinical goal. 63
63 Results for FIonA (left) and RayStation® (right) for Dmax in all level targets. . 64
64 DVH for lower-dose target (CTV_5412) for Case 1 for all configurations. 66
65 DVH for mid-dose target (CTV_5940) for Case 1 for all configurations. 67
66 DVH for high-dose target (CTV_6996) for Case 1 for all configurations. 68
67 DVH for right cochlea for Case 1 for all configurations. 69
68 DVH for lower-dose target (CTV_5412) and for high-dose target (CTV_6996) for Case 2 for all configurations. 70
69 DVH for superior pharyngeal constrictor for Case 2 for all configurations. 71

70 DVH for lower-dose target (CTV_5412) for Case 3 for all configurations. 72
71 DVH for mid-dose target (CTV_5940) for Case 3 for all configurations. 73
72 DVH for high-dose target (CTV_6996) for Case 3 for all configurations. 74
73 DVH for oral cavity for Case 3 for all configurations. 75
74 DVH for lower-dose target (CTV_5412) for Case 4 for all configurations. 76
75 DVH for mid-dose target (CTV_5940) for Case 4 for all configurations. 77
76 DVH for high-dose target (CTV_6996) for Case 4 for all configurations. 78
77 DVH for right carotid for Case 4 for all configurations. 79
78 DVH for lower-dose target (CTV_5412) for Case 5 for all configurations. 80
79 DVH for mid-dose target (CTV_5940) for Case 5 for all configurations. 81
80 DVH for high-dose target (CTV_6996) for Case 5 for all configurations. 82
81 DVH for left submandibular for Case 5 for all configurations. 83

List of Tables

1 Target Dose Metrics. The doses are presented as percentages of the target prescription dose. *Since the high-dose target is within the middle- and low-dose targets, the value is the same (76.96 Gy RBE), independent of the target level. . 5
2 Number of fields with specified targets as well as gantry and couch angles for Case 1 (Gantry 2 in red, Gantry 3 in blue). 8
3 Number of fields with specified targets as well as gantry and couch angles for Case 2 (Gantry 2 in red, Gantry 3 in blue). 9
4 Number of fields with specified targets as well as gantry and couch angles for Case 3 (Gantry 2 in red, Gantry 3 in blue). 10
5 Number of fields with specified targets as well as gantry and couch angles for Case 4 (Gantry 2 in red, Gantry 3 in blue). 12
6 Number of fields with specified targets as well as gantry and couch angles for Case 5 (Gantry 2 in red, Gantry 3 in blue). 13
7 Summary of available data for adaptive simulation process 15
8 Summary of treatment configurations simulated for each patient. Hybrid RO refers to plans optimized including range uncertainty. Full RO refers to plans optimized including both, setup and range uncertainty. In addition, in brackets used treatment machine is specified. 16
9 Comparison between conventional treatments, at G2 and G3. The percentages indicate the metrics that performed better for each respective Gantry or showing No Difference (N.D.) (not statistically significant). 25
10 RayStation, Configuration 2: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach. 25
11 RayStation, Configuration 3: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach. 26
12 RayStation, Configuration 4: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach 30
13 Comparison between adaptive treatments, at G2 and G3. The percentages indicate the metrics that performed better for each respective Gantry or showing No Difference (N.D.) (not statistically significant). 30

14 FIonA, Configuration 1: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach. 36
15 RayStation, Configuration 1: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach 37
16 Clinical goals for various anatomical structures. 41
17 Remarks in the dose metric table. 42

Part I
Introduction

1 Motivation

Proton therapy represents one of the most advanced forms of radiation treatment available today. Its precision stems from the unique physical properties of protons, which deliver the majority of their energy at a specific depth (known as the Bragg peak). This allows proton therapy to target tumors more accurately while sparing surrounding healthy tissues compared to conventional photon therapy. However, the effectiveness of proton therapy is highly sensitive to anatomical changes during treatment, more than for conventional photon therapy, as possible to see in figure 1. Factors such as tumor shrinkage, organ movement, or patient weight loss can alter the geometry of the target area and surrounding structures, compromising the precision of the delivered dose.

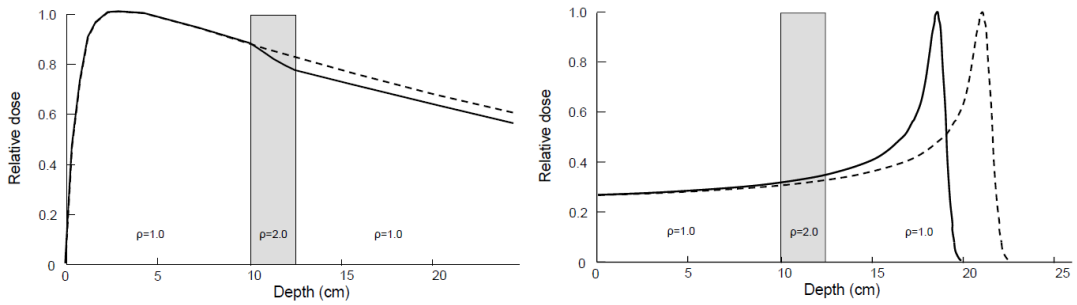

Figure 1: Difference of the depth dose curve with and without a 2 cm material of different density for photons (left) and protons (right).

To address these challenges, the concept of adaptive radiotherapy (ART) was introduced. ART involves modifying the treatment plan during the course of therapy to account for anatomical changes. This can be achieved in two primary ways: offline and online adaptation.

Offline ART adjusts treatment plans based on periodic imaging data collected during the treatment course. The periodicity can occur at established intervals or when significant anatomical differences are detected during treatment, a process that has been well-known in radiotherapy for a long time. However, studies have shown that not all patients benefit equally from this approach, highlighting the importance of patient selection for optimizing resources and clinical outcomes [3].

Online ART, on the other hand, enables real-time modifications based on daily imaging. This approach offers significant advantages by allowing the treatment plan to be updated immediately before each session, ensuring precise targeting despite day-to-day anatomical variations. However, the implementation of online adaptive treatment presents significant challenges. It requires acquiring an image suitable for dose calculation, redefining target/organ structures, replanning, and performing quality assurance (QA) checks within a matter of minutes, all while the patient is lying on the treatment couch, as illustrated in figure 2. During this process, the patient must remain still, which can be difficult due to discomfort or natural biological movements. Moreover, this workflow demands a lot of professionals in the control room, working together to ensure optimal and safe daily treatment delivery.

The concept of adaptive proton therapy (APT) extends these principles to proton therapy. APT is currently not commercially available, but it can be clinically developed utilizing daily imaging and advanced computational techniques to online daily adapt treatment plans, as illustrated in the diagram in figure 2. This approach, as said, can be particularly beneficial for proton therapy due to its higher sensitivity to density changes compared to photon therapy (as ilustrated in figure 1). Despite its potential benefits, implementing online APT is as complex as, if not more so than, adaptive photon therapy. The additional complexity arises from the precise dose deposition characteristics of protons, which make them highly susceptible to even minor variations in anatomy.

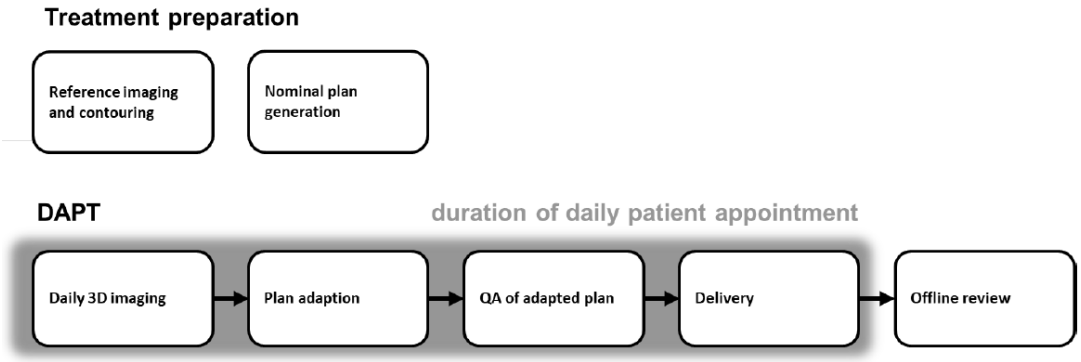

Figure 2: Workflow diagram of a generic daily adaptive proton therapy treatment [1].

At Paul Scherrer Institute (PSI) in Switzerland, a clinical workflow for daily adaptive proton therapy (DAPT) has been implemented using an in-house developed adaptive module, ADAPT (figure 3). It is worth mentioning that ADAPT treatment series consists of two plans: template and fallback. Template plan serves as a basis for daily plan reoptimization, while fallback plan is used as a backup solution in case of any issues encountered during the daily fractions. These two plans can be either identical or different in terms of beam arrangement, optimization constraints or used margins. Once the DAPT treatment series is assigned, the ADAPT tool is used to facilitate the preparation, reoptimization, and management of the online adaptation process, ensuring that a daily adapted treatment can be delivered while the patient remains on the treatment couch. This tool simplifies the adaptation workflow, allowing for online adjustments, enhancing treatment precision and effectiveness [1]. The workflow was initially implemented for the cranial region, within rigid anatomical areas [4]. All the process was time optimized to be done in order of minutes [5].

Head and neck cancers (HN) highlight the critical need for adaptive approaches in proton therapy. The head and neck region is anatomically complex, containing numerous critical structures such as the spinal cord, brainstem, optic nerves, salivary glands, and swallowing muscles. These structures are highly sensitive to radiation, and even small deviations in dose can lead to severe side effects, including xerostomia (dry mouth), dysphagia (difficulty swallowing), or neurological damage. Furthermore, HN patients frequently experience significant anatomical changes during treatment, such as tumor shrinkage and weight loss, which can alter dose distributions and compromise treatment outcomes [6].

However, adapting the current DAPT workflow for HN requires modifications, as the head and neck region is no longer a rigid anatomical area. The initial implementation of DAPT was optimized for tumors located within head, where rigid registration (RIR) is sufficient to handle position and anatomical variations. RIR is well-suited for anatomical sites that do not undergo significant structural changes, such as the cranial region, as it allows for translations and rotations but does not account for tissue deformations. However, in regions where organs

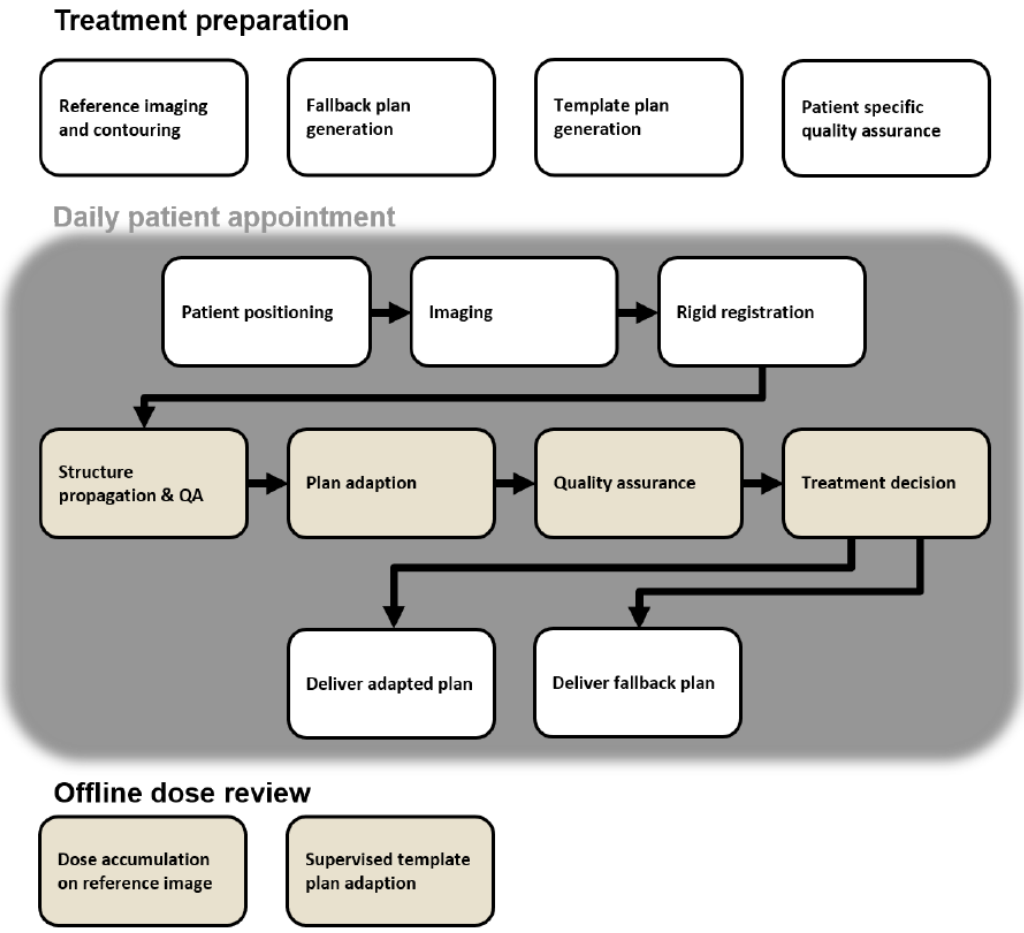

Figure 3: Workflow scheme of DAPT at PSI. The gray phases are performed in ADAPT [1].

shift, deform, or shrink, such as in abdominal or head and neck cases, RIR alone is insufficient to accurately represent anatomical changes during treatment.

To address this limitation, deformable image registration (DIR) would be more suitable for adaptive radiotherapy of HN. Unlike rigid registration, DIR accounts for complex anatomical deformations, such as tumor shrinkage or displacement of critical structures. This capability is essential for accurately propagating organ contours and reoptimizing daily plan, ensuring that treatment remains precise despite of the significant anatomical variations in the head and neck region.

The motivation for this work lies in addressing the current challenges of proton therapy for head and neck cancers (HN) and explore the potential of daily adaptive proton therapy (DAPT) workflow developed at PSI. Traditional proton therapy depends on static imaging obtained weeks before treatment, which limits its ability to adapt anatomical changes during the treatment course. DAPT aims to overcome these limitations by incorporating daily imaging, deformable registration, and online plan adaptation, thereby improving tumor targeting, minimizing radiation exposure to healthy tissues, and enhancing overall treatment dose distribution. This project investigates the application of DAPT for HN using clinical data from patients treated at PSI Gantry 3 with daily cone-beam computed tomography (CBCT) scans acquired for daily positioning. These images were further used to generate synthetic CTs (synCTs) for study purposes. These were subsequently aligned with planning images via deformable registration, enabling daily structures propagation and daily plan reoptimization, simulating DAPT

implementation for HN cases.

In conclusion, this work aims to demonstrate the benefits of DAPT in addressing the challenges posed by anatomical changes in HN. Through the integration of advanced imaging and computational tools, we seek to establish DAPT as a viable approach for improving treatment precision and efficacy in proton therapy, ultimately advancing the standard of care for patients with head and neck cancer.

Part II
Materials and Methods

Five head and neck cancer cases were previously treated at PSI, utilizing the planning system *Eclipse*^{TM} (Varian Medical Systems, Palo Alto, California, USA) at Gantry 3 (G3), a commercial proton therapy gantry equipped with integrated Cone Beam CT (CBCT) for imaging.

In this master thesis, adaptive treatments were simulated for all five cases, necessitating replanning using two treatment planning systems: an in-house Flexible Ion-planning Application (FIonA) and RayStation® 2023B (RaySearch Laboratories, Stockholm, Sweden). The treatment plans were created for two distinct gantry systems: Gantry 2 (G2), an in-house developed proton therapy system with both FIonA and RayStation® available for planning, and Gantry 3 (G3), a commercially available gantry from Varian Medical Systems (Palo Alto, California, USA), which supports planning exclusively with RayStation®. This dual-system setup allowed for a comprehensive exploration of alternative planning strategies and enabled a performance comparison of adaptive treatment techniques across different planning systems and gantry configurations.

The treatment plans in FIonA were generated for G2, following methodology of daily adaptive proton therapy (DAPT) workflow developed at PSI. Although the ADAPT system itself was not utilized, the adaptive treatment process was manually simulated in FIonA using deformable image registration algorithm from Velocity (Varian Medical Systems, Palo Alto, California, USA) to perform images registration and daily structures propagation.. Details are provided in section 4.

For the plans created in RayStation® 2023B, both G2 and G3 were employed, in order to explore different configurations and analyses. The daily adaptive workflow in RayStation® was performed using a script, mimicking in-house developed DAPT workflow behavior. Details can be found in section 5.

Using two different TPS's, FIonA and RayStation®, provided a way to study how daily adaptive plans can be delivery in different approaches. Taking advantage of the unique features of both systems, it was possible to better understand how planning methods can affect treatment dose distribution.

Every case was planned to meet QUANTEC [7] and DAHANCA [8] dose criteria for the organs at risk, using prescriptions provided by a medical doctor (MD). These prescriptions adhered to international recommendations while incorporating patient-specific considerations to address individual anatomical and clinical needs. For some cases, specific metrics were adjusted to ensure the plan achieved reasonable dose values, particularly for organs located within or near high-dose regions. All cases were planned using Multiple Field Optimization (MFO) with a Simultaneous Integrated Boost (SIB) technique over 33 treatment fractions. The prescribed dose levels for the CTVs were 54.12, 59.40, and 69.96 Gy RBE for the low, middle, and high-dose targets, respectively (except for Case 2, as explained in section 3.2). For each patient and each target, five metrics were analyzed across the three CTV dose levels, as presented in table 1, as well as dose-volume histograms (DVH) (appendix D).

The values for the OAR metrics are presented in appendix A, where case-specific adjustments to these metrics are also highlighted.

Table 1: Target Dose Metrics. The doses are presented as percentages of the target prescription dose. *Since the high-dose target is within the middle- and low-dose targets, the value is the same (76.96 Gy RBE), independent of the target level.

Metric	**Value**	**Description**
Dmin	90%	At Least 90% Dose At 100.0% Volume
D98%	98%	At Least 98% Dose At 98.0% Volume
Dmean	100%	At Least 100% Average Dose
V95%	95%	At Least 95.0% of Volume At 95% Volume
Dmax	110%*	At Most 110% Dose At 0.0%

2 Gantry 2 x Gantry 3

In this study, both G2 and G3 were utilized for planning head-and-neck (HN) cases. The standard-of-care treatment is typically performed on G3, primarily due to its capability for volumetric imaging and the use of a couch that is capable of applying position correction in 6 degrees of freedom, allowing for precise patient positioning. On the other hand, the adaptive workflow is currently being implemented on G2.

Having the access to two treatment planning systems, two treatment machines (G2 and G3) as well as an adaptive platform established at G2, a comprehensive overview of adaptive and conventional treatment configurations was possible, as described in section 5..

It is important to note some key differences between the two Gantries that may influence the simulated delivered dose distribution. These differences, primarily related to geometry and beam delivery characteristics, can result in varying treatment results between the two Gantries and should be considered during analysis.

2.1 Gantry 2

PSI was the first institute, and the only one for several years, worldwide to use Intensity Modulated Proton Therapy (IMPT). The development of G2 was a strategic choice, designed for an iso-centric compact layout with a diameter limited to 7.5 meters. This compact design enabled more precise treatments in a smaller space, contributing to the advancements in proton therapy technology at PSI [9].

The G2 system features a compact design with a range of motion that allows beam delivery from angles between $-30°$ and $180°$. While this differs from the full $360°$ rotation typically available in conventional photon gantries and some proton gantries, the couch rotation variying from $-180°$ to $180°$, complements the system's capabilities. By combining gantry and couch rotations, it is possible to achieve the desired beam entry direction for nearly all treatment indications, demonstrating the system's adaptability and efficiency despite its compact design.

In G2, the sweeper magnets are mounted before the last bending magnets (upstream scanning), generating a parallel beam. This keeps the spot size small across all energies (100-230 MeV, with a width of less than 3-4 mm). The nozzle can also be extended to shorten the air gap between the beam's exit window and the patient. The range shifter in this gantry is fully automated, allowing it to move in or out for each spot as needed. This feature improves the quality of treatment plans and provides greater flexibility in adapting to different clinical scenarios.

An in-room CT on-rails is used for daily positioning, planning and control CT scans. It even supports 4D imaging for capturing time-resolved scans. Additionally, there's an x-ray system mounted directly on the gantry, which allows for 2D Beam's Eye View (BEVs) acquisition in treatment position..

Figure 4: In-house developed Gantry 2.

2.2 Gantry 3

Gantry 3, a commercial gantry from Varian (Palo Alto, California, USA), has been in a clinical operation since July 2018. G3 offers treatment capabilities comparable to G2 but introduces some key differences. Unlike G2, G3 features full 360° gantry rotation and employs downstream scanning, with sweepers positioned after the last bending magnet. This configuration results in a divergent beam, a maximum field size of 30 x 40 cm^2 and a larger spot width compared to G2, although it still maintains precise proton steering using nine primary magnets around a vacuum pipe. With active scanning technology, G3 also can effectively target tumors of any three-dimensional shape [10].

Figure 5: Commercial Gantry 3.

The gantry is equipped with an integrated 360° co-rotating onboard CBCT imaging system that ensures accurate patient positioning directly in the treatment position without movement of the couch. Its capability to deliver large irradiation fields, rapid energy changes, and relatively high dose rates minimizes the treatment time per patient.

Beyond the gantry itself, the integrated couch can move in six degrees of freedom, enabling corrections in rotation as well as translations, unlike the G2 couch which only supports translational movements.

3 Cases

Each head and neck case was initially planned using a field arrangement matching the original treatment setup, with minor angle adjustments made as necessary.

According to the International Commission on Radiation Units and Measurements (ICRU) recommendations [11], margins are typically added to the Clinical Target Volume (CTV) to create the Planning Treatment Volume (PTV), these margins account for uncertainties in patient setup, organ motion, and other variabilities during treatment.

A split-target technique was used to reach certain sub-regions of the target. These sub-regions are typically named based on their anatomical location. For example, CTV_{UP} refers to the CTV within cranial region in the head, while CTV_{DOWN} often corresponds to lymph node regions. The CTV_{DOWN} region can be further subdivided into CTV_{DOWN_L} and CTV_{DOWN_R}, representing the left and right sides of the patient, respectively. Additionally, a hybrid region, such like CTV_{HYBRID_L}, can be defined, representing the union of parts of CTV_{UP} and CTV_{DOWN_L}.

Such segmentation allows for precise adaptation of the treatment fields to the specific anatomy and clinical goals. An exemplary illustration of such target sub-division is shown in figure 6.

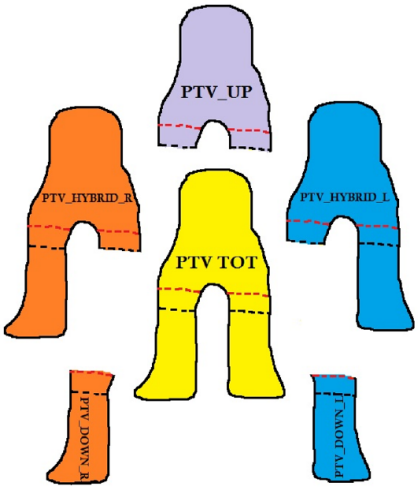

Figure 6: The split-target technique, here presented for Planning Target Volume (PTV) but the same logic can be applied to CTV [2].

Each patient had acquired daily cone-beam CT (CBCT) images for positioning purposes. However, while CBCT images are valuable for visualizing patient anatomy, they suffer from other limitations which make them unsuitable for proton dose calculations. To overcome these limitations, synthetic computed tomography (synCT) images is generated based on daily CBCTs,

providing a more accurate representation of the patient's current anatomy. These synCTs were generated for each patient by deforming respective planning CTs to match daily CBCTs, and preserving the Hounsfield Units (HU) of the CT. The system used for generating synCTs was Velocity (Varian Medical Systems, Palo Alto, California).

It is important to note that the synCTs used in this study were not generated as part of this work, they were rather created prior to the current analysis.

3.1 Case 1

The first case was a squamous cell nasopharyngeal carcinoma, for which a split-target technique was employed. Six fields were used to target different regions of the tumor and depending on the target location, various gantry and couch angles were applied to optimize the simulated delivered dose. The field configuration, along with the axial, sagittal, and coronal images, is shown in table 2 and figure 2.

Table 2: Number of fields with specified targets as well as gantry and couch angles for Case 1 (Gantry 2 in red, Gantry 3 in blue).

Field	**Gantry [°]**	**Couch [°]**	**Target**
1	180 / 180	0 / 0	CTV_{5412}
2	130 / 130	25 / 335	CTV_{UP}
3	50 / 50	0 / 0	CTV_{5412}
4	0 / 0	0 / 0	CTV_{DOWN}
5	50 / 310	180 / 0	CTV_{5412}
6	130 / 230	155 / 25	CTV_{UP}

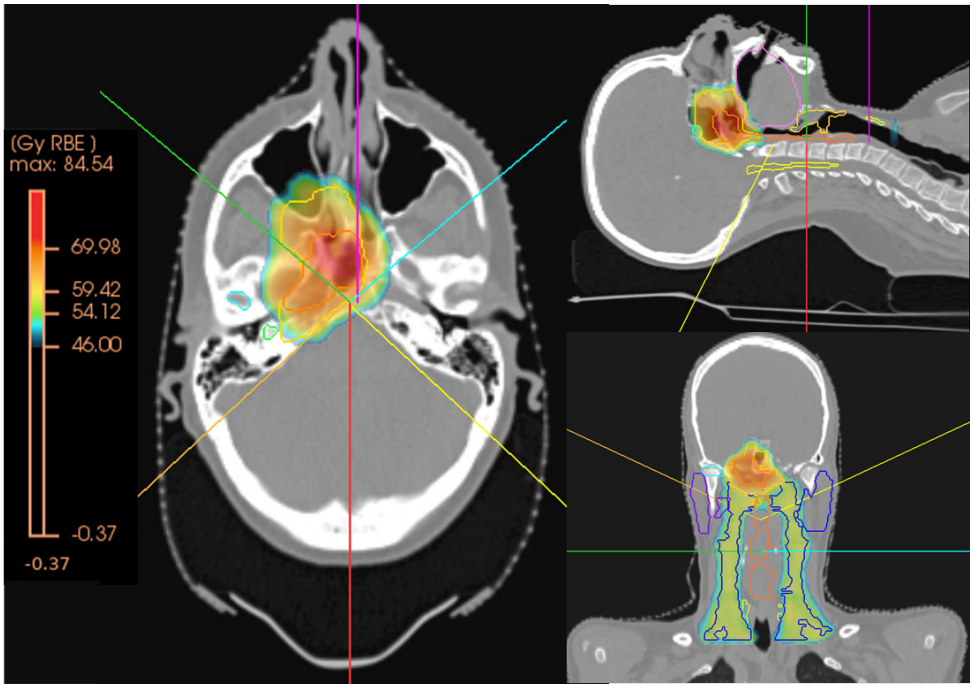

Figure 7: Fields configuration and planned dose distribution for Case 1 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange.

Prescription details and clinical goals are summarized in appendix A. For this particular case, fulfilling the dose constraint for the right cochlea presented a significant challenge. The organ had to receive a mean dose of less than 45.0 Gy RBE, yet it was located near the middle-level CTV, which required a prescribed dose of 59.4 Gy RBE. Consequently, the MD established a softer constraint for the right cochlea, setting a maximum dose of 54.0 Gy RBE to balance protection and target coverage.

A total of 27 daily CBCT images were acquired during the course of the treatment, and no replanning was required during this period.

3.2 Case 2

The second case was a re-irradiation of squamous cell carcinoma of the ethmoid in the nasopharyngeal region, and had only two target dose levels of 54.12 and 69.96 Gy RBE. Due to the smaller treatment area, only four fields (two anterior-oblique and two posterior-oblique) were used, all targeting the same region, as shown in table 3.

Table 3: Number of fields with specified targets as well as gantry and couch angles for Case 2 (Gantry 2 in red, Gantry 3 in blue).

Field	**Gantry [º]**	**Couch [º]**	**Target**
1	110 / 110	160 / 340	CTV_{5412}
2	40 / 40	120 / 300	CTV_{5412}
3	40 / 320	60 / 60	CTV_{5412}
4	110 / 250	20 / 20	CTV_{5412}

Prescription details and clinical goals are summarized in appendix A. For this case, achieving the dose constraint of a mean dose below 45.0 Gy RBE for the superior pharyngeal constrictor muscle represented a challenge as possible to see in the sagittal and coronal slices in figure 8. To try to achieve this constraint, the inferior part of the low dose target, corresponding to approximately three slices, had reduced coverage.

The dose constraints for the lenses also could not be fully achieved. However, considering the minimal clinical impact of overdosing these organs, efforts were made to reduce the dose as much as possible while prioritizing target coverage.

During the course of treatment, 33 CBCTs were acquired and no replanning was necessary.

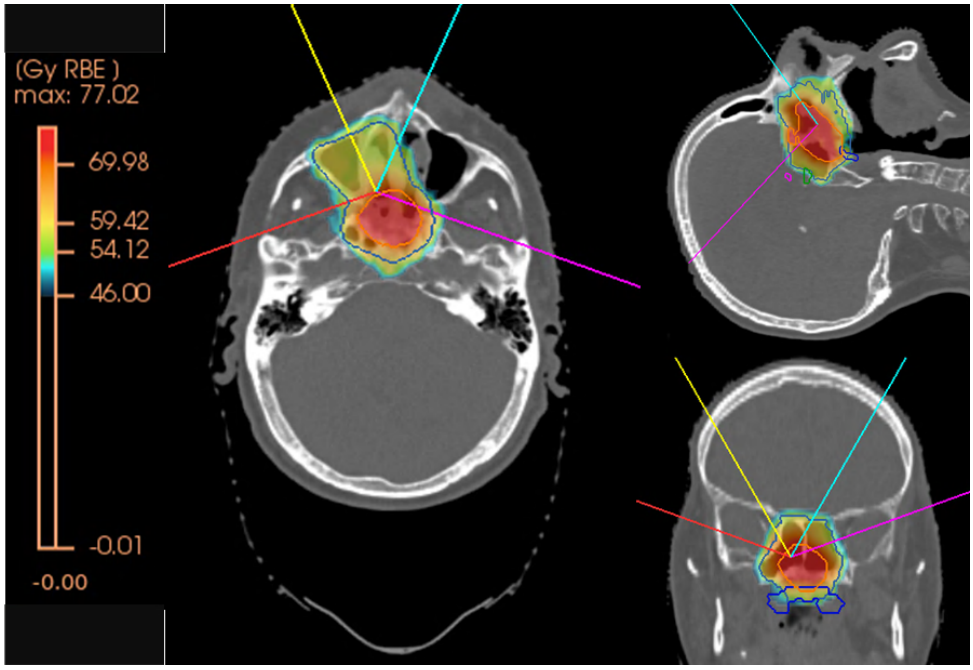

Figure 8: Fields configuration and planned dose distribution for Case 2 (Image taken from FIonA). CTV_5412 in blue, and CTV_6996 in orange.

3.3 Case 3

The third case was an adenoid cystic carcinoma of the palate, a rare tumor in the head and neck region. A split-target technique with six fields was used, each targeting different tumor region.

Table 4: Number of fields with specified targets as well as gantry and couch angles for Case 3 (Gantry 2 in red, Gantry 3 in blue).

Field	**Gantry [º]**	**Couch [º]**	**Target**
1	110 / 110	15 / 15	CTV_{UP}
2	55 / 55	15 / 15	CTV_{HYBRID_L}
3	30 / 30	0 / 0	CTV_{DOWN_L}
4	-30 / 330	0 / 0	CTV_{DOWN_R}
5	55 / 315	165 / 165	CTV_{HYBRID_R}
6	110 / 250	165 / 165	CTV_{UP}

Prescription details and clinical goals are summarized in appendix A. Due to the high complexity of this case, several constraints for OARs were challenging to be achieved during the planning process, particularly for the right eye and both parotids. Only the left parotid was able to receive a dose below the prescribed constraint, as the right parotid was located close to the high-dose target (69.96 Gy RBE). This is clearly shown in the axial and coronal slices of figure 9, where the right parotid is delineated with a green contour and the high-dose target is outlined with a white line.

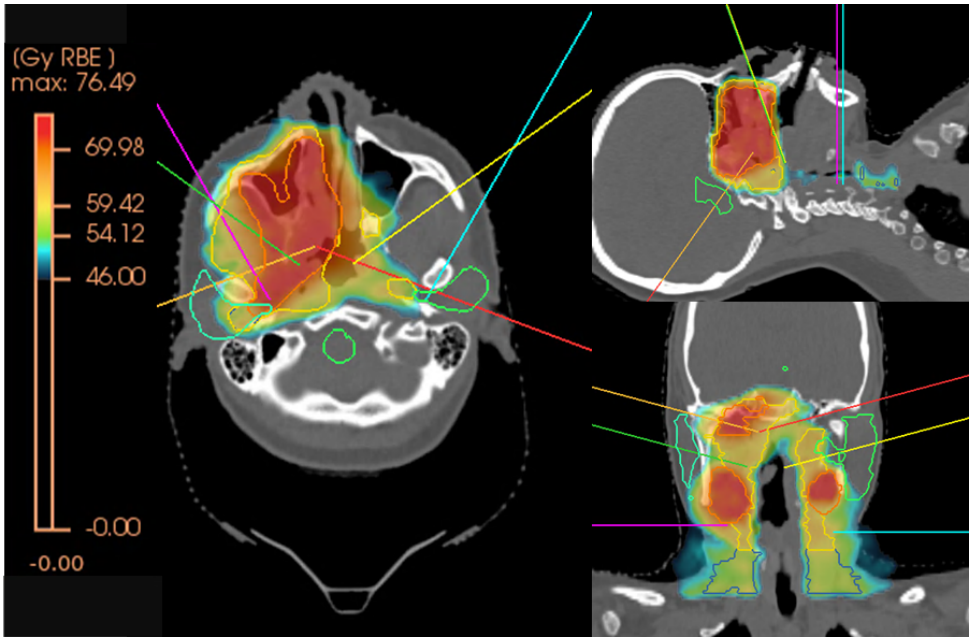

Figure 9: Fields configuration and planned dose distribution for Case 3 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange.

Figure 10: Rigid (left) and deformable (right) registration of daily synCT and planning CT for fraction fraction 10 of Case 3 (between daily synCT and planning CT). The change in head and chin position is evident in the case of rigid image registration (RIR).

The patient required two replannings during treatment. For daily adaptive treatment simulation, no replanning was performed; however, for the conventional treatment simulation, replanning was performed in the same way and considering the same timeline as during the treatment.

The first replanning CT was acquired on the day of fraction 5. Due to the time needed for planning, quality assurance and approvals, the replanned treatment was only implemented starting from fraction 9. Similarly, the second replanning CT was performed on the day of fraction 14, but the updated treatment plan was not delivered until fraction 21. This resulted in 13 fractions being treated using the second replanned plan.

It is important to highlight that for this case, difference in position became evident and challeging in the middle of the treatment, particularly between fractions 9 and 20, as illustrated in figure 10.

A total of 33 CBCTs were acquired, one for each fraction of the treatment. The synthetic CTs were generated using the initial planning CT as the reference, rather than the corresponding replanned CTs.

3.4 Case 4

The fourth case was a squamous cell hypopharyngeal carcinoma. Four fields (two anterior-oblique and two posterior-oblique) were used to treat this case, with the same target across all fields, as showed in table 5.

Table 5: Number of fields with specified targets as well as gantry and couch angles for Case 4 (Gantry 2 in red, Gantry 3 in blue).

Field	**Gantry [º]**	**Couch [º]**	**Target**
1	110 / 110	30 / 330	CTV_{5412}
2	70 / 70	0 / 0	CTV_{5412}
3	70 / 290	180 / 0	CTV_{5412}
4	110 / 250	150 / 30	CTV_{5412}

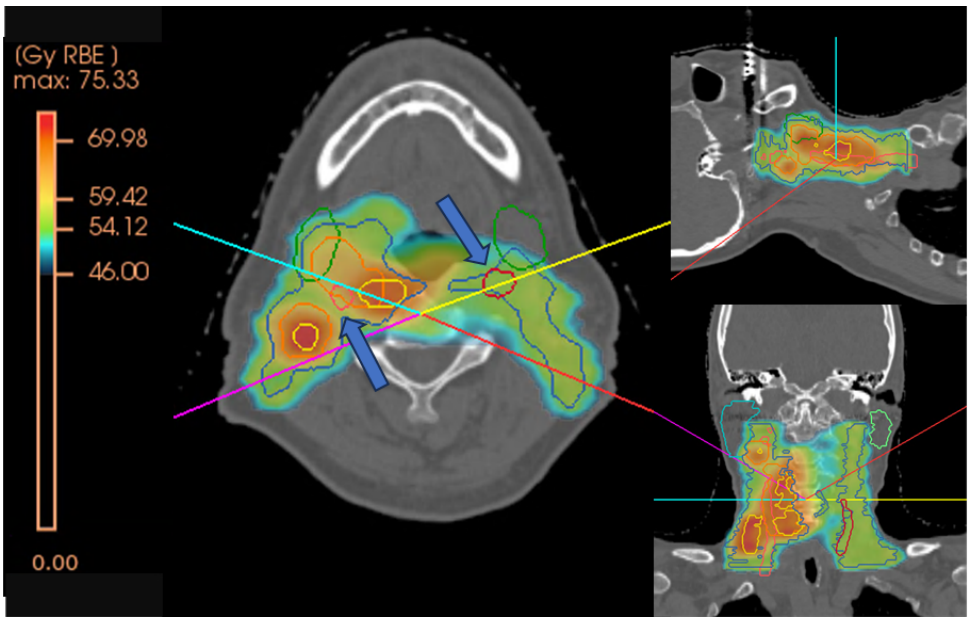

Figure 11: Fields configuration and planned dose distribution for Case 4 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. The arrows are pointing to Carotids.

Prescription details and clinical goals are summarized in appendix A. In this case, the carotid artery was partially inside the high-dose target, making it impossible to meet the constraint of a maximum dose of 45.0 Gy RBE. Similarly, the right submandibular gland and the right parotid gland were located near the high-dose target, resulting in mean doses exceeding the prescribed constraints of 30.0 Gy RBE and 26.0 Gy RBE, respectively, as shown in figure 11.

The patient needed to be replanned twice, similarly to Case 3. The first replanning CT was performed in the first treatment day and the replanned treatment was started at fraction 6. The second replanning CT was performed at fraction 14, and was not treated until fraction 28.

Daily CBCTs were acquired for all 33 fractions throughout the treatment.

3.5 Case 5

The fifth case was a sinonasal undifferentiated carcinoma, a rare and aggressive malignant tumor of the nasal cavity. Six fields with a split-target technique, as specified in table 6, were used to treat different target regions.

Table 6: Number of fields with specified targets as well as gantry and couch angles for Case 5 (Gantry 2 in red, Gantry 3 in blue).

Field	**Gantry [º]**	**Couch [º]**	**Target**
1	110 / 110	20 / 340	CTV_{POST_L}
2	50 / 50	0 / 0	CTV_{ANT_L}
3	20 / 20	0 / 0	CTV_{5412}
4	20 / 340	80 / 0	CTV_{5412}
5	50 / 310	180 / 0	CTV_{ANT_R}
6	110 / 250	160 / 20	CTV_{POST_R}

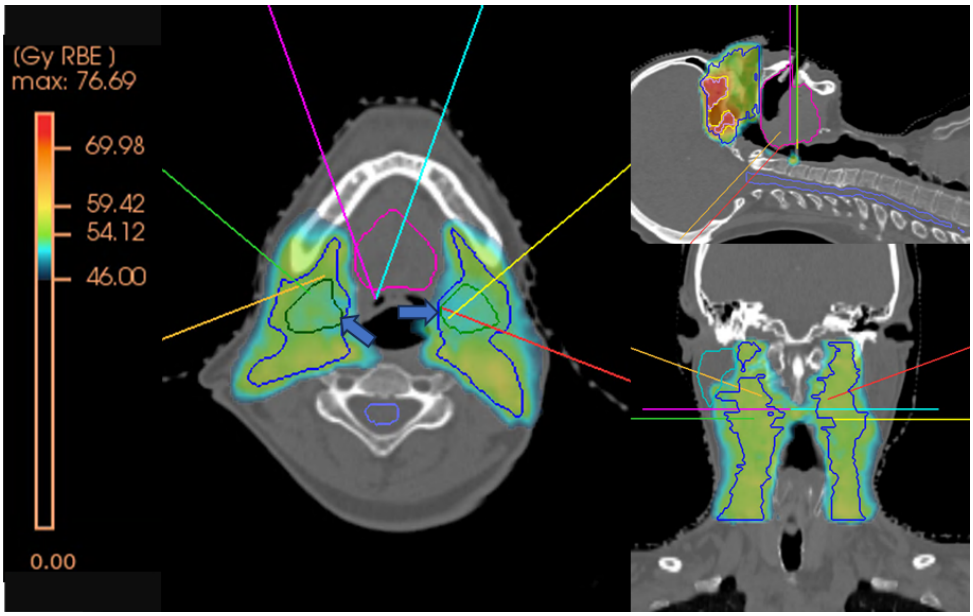

Figure 12: Fields configuration and planned dose distribution for Case 5 (Image taken from FIonA). CTV_5412 in blue, CTV_5940 in yellow and CTV_6996 in orange. The arrows are pointing to submandibulars.

Unlike in previous cases, target division was different than the one presented in figure 6. The CTV_{POST} of each side is the union between the CTV_{ANT} and the cranial part of the target that is entire in the nasal cavity (region of the high-dose target).

Prescription details and clinical goals are summarized in appendix A. For this case, achieving the mean dose of 54.0 Gy RBE for the submandibular glands was particularly challenging, as both structures are entirely within the low-dose target (54.12 Gy RBE). In addition, the right optical nerve, located within the high-dose target, achieved maximum dose of 72.0 Gy RBE, rather than 54.0 Gy RBE which is the standardly used constraint. A lower dose in the submandibular glands, located within the target, is visible in the dose distribution shown in figure 12.

No replanning was required for this patient and in total 31 daily CBCTs were acquired during the course of treatment.

4 Treatment Plans in FIonA

As HN patients are standardly planned and treated at G3, it was replicated the planning approach at G2 with the use of FIonA TPS and prepare the patients for DAPT workflow simulation as shown in figure 3. Depending on the treatment simulation approach, different planning considerations were applied and for each case, two treatment plans were created.

The fallback plan followed conventional planning approach with the plan optimization done on PTV (isotropic 4-5 mm margin around each CTV dose level), the template plan was optimized on CTV with reduced margin (isotropic 2 mm margin around each CTV dose level) and included range uncertainty of ±3%. This hybrid robust optimization approach results from FIonA TPS capability of including only range uncertainty in the optimization problem. No setup uncertainty was considered for these plans.

Robust optimization allows for the minimization of treatment margins by accounting for errors in the optimization. These advanced plans can ensure effective dose distribution while maintaining high-quality treatment outcomes. Studies have indicated that robust plans can lead to improved treatment outcomes by ensuring that critical structures remain within safe dose limits, even in the face of anatomical changes. This approach enhances treatment efficacy, minimizes the risk of adverse effects, and ultimately leads to better overall patient dose distribution [12].

Both plans were generated using the same fields and beam/couch arrangements. Each plan was optimized to meet the prescription provided by medical doctors (MD) and to replicate target and OARs doses achieved in clinical treatment plans.

4.1 Structure QA

In adaptive workflows, volume verification plays a crucial role in ensuring the accuracy of propagated structures, particularly when adapting treatment plans based on daily imaging. In the currently developed DAPT workflow, only rigid propagation of daily structures is performed, with a dedicated volume check consisting of two criteria: a relative difference of 10% and an absolute difference of 1 cc in relation with the volume from the reference set, planning CT. It has been verified that for anatomies with predominantly rigid behavior, these thresholds are sufficient [13].

In this work, it was investigate the feasibility and potential benefits of online adaptive therapy for HN cases, where anatomical changes can be more complex. Since the DAPT workflow was replicated in FIonA, now incorporating deformable registration, it is necessary to reassess

whether the existing volume QA criteria remain appropriate. The structure QA process ensures that the new structure's volume remains within the defined thresholds when compared to the propagated structure. If both criteria are met, the QA passes. If only one is satisfied, a warning is issued; however, the treatment may still proceed. For smaller structures, this often results in a noticeable relative percentage difference, whereas for larger structures, the absolute volume difference in cubic centimeters is the primary metric [13].

The volume analysis was performed in FIonA, using structures that were deformably propagated through Velocity (Varian Medical Systems, Palo Alto, California, USA). For each structure and each daily image, the calculated volume was compared to the reference structures, such as targets and organs at risk (OARs), delineated by the physicians on the planning CT.

4.2 Adaptive simulation process

As discussed, from this point forward, the available data for each HN case consists of the components outlined in table 7, i.e. daily synCT images, corresponding propagated daily structure sets, and two treatment plans, i.e., fallback and template, comprising a DAPT treatment series.

Table 7: Summary of available data for adaptive simulation process

Data Type	**Description**
Image	Daily synthetic CT images for each fraction.
Structure Sets	Corresponding daily structure sets for each treatment fraction.
Treatment Plans	Two treatment plans: fallback plan and template plan.

For each fraction, template plan was reoptimized on daily synCT to simulate daily plan adaptation and fallback plan was recalculated to simulate conventional treatment approach. In particular, daily plan reoptimization involved adjustment of spot positions, resulting in a different spot list and MU than the initial template plan, which can lead to variations in dose distribution [14]. On the other hand, the recalculation of a fallback plan without reoptimization, retained the original list of spots and calculated MU as in the initial fallback plan.

This treatment simulation approach closely mirrors the DAPT workflow developed at PSI, as previously mentioned. This procedure was conducted for every patient and for each treatment fraction. It is important to highlight that for patients who had fewer than 33 synCT images available, the last synCT substituted the remaining fractions.

5 Treatment plans in RayStation®

RayStation® (RaySearch Laboratories, Stockholm, Sweden) utilizes a different optimization process compared to FIonA, with distinct optimization algorithms and calculation methods.

In RayStation®, robust optimization (RO) can be selectively applied to specific regions of interest (ROIs), unlike FIonA, where RO is always applied to the entire set of structures. In addition RayStation® can robustly optimize plans accounting for both setup and range uncertainty, therefore, a distinct approach was adopted for the planning process using this software.

Thanks to the possibility of using two treatment planning systems and two treatment machines available at PSI, four different configurations of daily adapted treatments were simulated.

In total, for each case six treatment plans were created to cover all possible treatment scenarios as specified in table 8.

Table 8: Summary of treatment configurations simulated for each patient. Hybrid RO refers to plans optimized including range uncertainty. Full RO refers to plans optimized including both, setup and range uncertainty. In addition, in brackets used treatment machine is specified.

Configuration	TPS	Daily Adaptive	Conventional Fallback
1	FIonA	Hybrid RO (3%) in CTV+2mm (G2)	Conventional plan in PTV (CTV+4/5mm) (G2)
	RayStation®	Hybrid RO (3%) in CTV+2mm (G2)	Full RO 3mm,3% in CTV (G2)
2	RayStation®	Full RO 2mm,3% in CTV (G2)	Full RO 3mm,3% in CTV (G2)
3	RayStation®	Full RO 2mm,3% in CTV (G2)	Full RO 3mm,3% in CTV (G3)
4	RayStation®	Full RO 2mm,3% in CTV (G3)	Full RO 3mm,3% in CTV (G3)

The first configuration reflects the established concepts implemented in FIonA, focusing on a comparative analysis between adaptive planning and the conventional standard of care in RayStation®, which employs full robust optimization (RO) on the CTV. The adaptive plan was optimized on the CTV with a reduced margin and included only range robustness, following the approach used in FIonA.

In the second configuration, also performed for Gantry 2, the full RO optimization was applied to both treatment simulations but with reduced setup uncertainty for the daily adaptive scenario. This configuration explored the potential of reducing setup margins in adaptive treatments, investigated its impact on target coverage and compared to the conventional treatment.

The third configuration explored the differences between adaptive treatments with full RO and reduced setup uncertainty versus full RO with standard setup settings for two treatment machines, i.e., G2 and G3. Due to their technological differences, it was interesting to compare whether the adaptive treatment on G2 would outperform the conventional treatment on the commercial G3. The differences between machines are highlighted in section 2.

Finally, the fourth configuration extends the investigation of adaptive planning exclusively at G3, it highlighted differences between adaptive treatment simulation with full RO and reduced setup uncertainty versus full RO with standard setup settings.

The configurations can also be analyzed in a cross-comparative manner. For example, the second and third configurations can be examined collectively to highlight differences between conventional treatment at G2 and G3. Similarly, third and fourth configurations can be grouped to compare the dose distribution of the two adaptive approaches. The primary distinction between these configurations lies in the gantry technological capabilities and the spot width characteristics of the respective systems.

All the plans were planned using the same beams arrangement and field-specific targets target as in FIonA.

5.1 Performing adaptive treatment simulation in RayStation®

Scripting in RayStation® (RaySearch Laboratories, Stockholm, Sweden) facilitates the automation of repetitive tasks, such as loading and registering images, performing calculations, and verifying patient data. This functionality enables users to create custom workflows that streamline processes, minimize manual errors, and save time. Given that no commercial solutions are currently available for proton therapy in the context of adaptive treatment simulation, scripting becomes essential. To implement an online adaptive workflow, mimicking earlier developed DAPT, custom scripts were developed at PSI in collaboration with RayStation´s developers to ensure accurate image registration, streamline adaptive treatment workflows, and provide the flexibility needed for personalized adjustments during treatment.

The script first verifies if the patient has two 3D images: one planning CT and one daily CT and that there are two distinct plans optimized on the planning CT, a primary template and fallback, each with correct naming convention. The initial step is to generate new structures on the daily CT, in this case, daily synCT. Three types of segmentation are available: rigid, deformable, and deep learning. For this work, deep learning segmentation, a machine learning model built in Raystation®, was used for OARs propagation. Once structure propagation is completed, template plan reoptimization and fallback plan recalculation is performed on daily image to simulate adaptive and conventional treatment scenarios, respectively.

6 Statistical Analysis

6.1 Deviation analysis

For all the analyses, the absolute difference between the analyzed values was defined as:

$$\delta = Value_{calculated} - Value_{prescription} \tag{1}$$

The dose is measured in Gy RBE, with the prescribed dose explicitly defined in absolute units as follows: CTV_5412 = 54.12 Gy RBE (corresponding to 100%), in accordance with clinical standards. Regardless of whether the evaluated structure is a target or an organ at risk, all absolute differences (δ) are expressed as percentages. For constraints already defined as percentages (e.g., $V_{95\%} < 95\%$), δ was calculated by directly subtracting the percentage values.

All δ values were computed relative to the approved fallback plan, which serves as the baseline. This fallback plan represents the standard-of-care treatment, created during the initial planning phase and approved by the physician based on the planning CT. That is reflected in the box plots presented in part III, where the absolute relative differences between the achieved values (adaptive/conventional) and their respective fallback plan baselines is illustrated. These comparisons are performed for different simulated treatment scenarios, ensuring that all deviations are consistently assessed against the approved standard-of-care plan.

6.2 Wilcoxon test

Since the data distributions were generally not normal, a non-parametric statistical hypothesis test was applied. The Wilcoxon test was used to assess the statistical significance between two distributions, such as the dose simulated for daily adapted and conventional treatments. A p-value < 0.05 was considered statistically significant, indicating that the two distributions were likely different.

For each dose metric analyzed, two distributions of 33 fraction doses were obtained, corresponding to the adaptive and conventional treatment simulations. The Wilcoxon test was used to evaluate their relationship and determine whether there was a statistically significant difference between them.

As the Wilcoxon test is influenced by the median, the median values of the distributions were analyzed to determine which approach had a higher dose. Specifically, if a significant difference was found, it was possible to identify which treatment resulted in a higher or lower dose for the evaluated metrics. Namely, a target with better coverage was identified by higher coverage metrics (Dmin, D98, V95 and Dmean) and a lower maximum dose Dmax (see table 1), while a OAR metric with an average higher dose was considered worse for a given simulated treatment type.

Part III
Results

7 Structure QA

The volumes analyzed correspond to the structures delineated by physicians (Medical Doctor, MD) on the planning CT. These were compared to the structures propagated in Velocity (Varian Medical Systems, Palo Alto, California, USA) for every daily image available. Structures created specifically to support the treatment planning process, such as those used for optimization or geometrical purposes, were excluded from the analysis.

The accepted threshold for the structure QA was defined in absolute and relative values of 1 cc and 10%, successively. Larger structures, such as the oral cavity with a volume of around 90 cc, may face challenges in meeting absolut criteria, an absolute deviation of 1 cc represents only about 1% of the structure's volume, a clinically insignificant variation that could still lead to a failure in the absolute volume evaluation. On the other hand, smaller structures are more prone to failing percentage-based criteria due to their small volumes, for instance, a left lens with a volume of 0.2 cc in the planning CT would require its 10% volume difference to remain within a range of 0.18 cc to 0.22 cc, corresponding to just a few voxels. While such small differences are unlikely to have significant clinical implications, they may nonetheless cause the structure to fail the percentage-based evaluation.

For the volume analysis, the structures were first evaluated against the absolute criteria. Structures that did not meet these criteria were classified as "large volume" and the percentage-based criteria were subsequently assessed to determine if the daily variations fell within the currently accepted threshold of 10%.

The analysis was performed for all cases; however, two cases, one with good and another with less favorable propagation results, were selected for detailed discussion. All other cases are presented in appendix B.

7.1 Case with good propagation results

Case 1 demonstrated successful daily structure propagation, with all daily propagated structures remaining below the current threshold for all fractions.

The registration of daily images was performed without significant issues, as all images were closely aligned with the planning CT. As shown in figure 13, the absolute differences in volume were evaluated across all daily fractions compared to the reference volumes. For all analyzed structures, volume differences were assessed in both absolute and relative units. Structures with an absolute volume below 10 cc successfully met the 1 cc threshold criteria. For structures larger than 10 cc, the volume difference was subsequently evaluated in percentage units using the 10% threshold, figure 14. All remaining structures also satisfied this criteria. Therefore, for this particular case, the currently clinically implemented structure QA approach for rigid anatomies would be sufficient.

In summary, for the first case, all the structures met the proposed structure QA criteria, whether using relative or absolute criteria, across all the daily images.

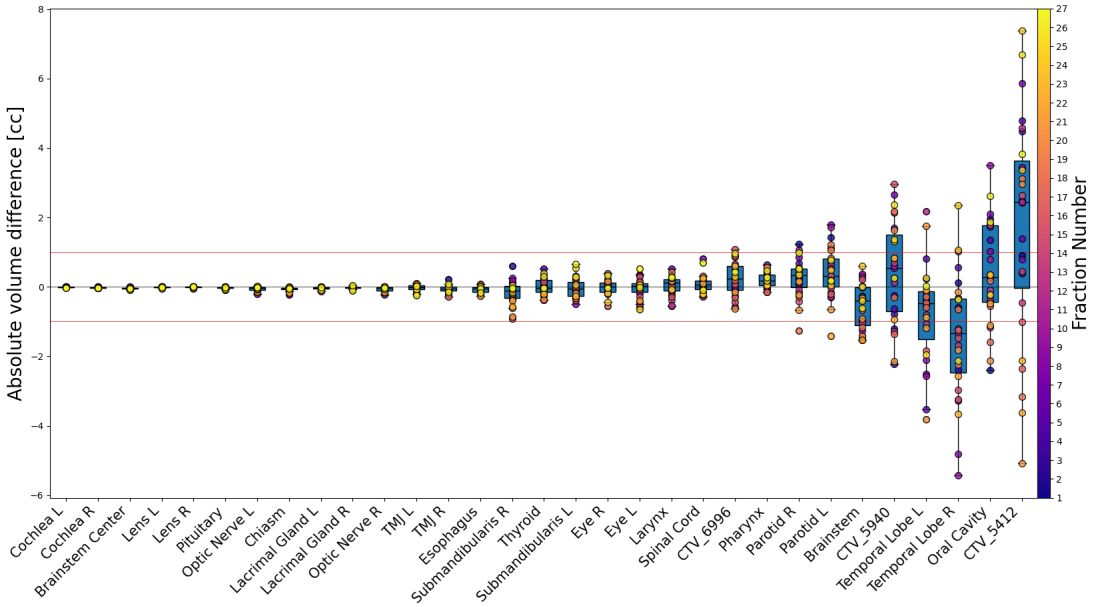

Figure 13: Absolute volume differences for Case 1 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc.

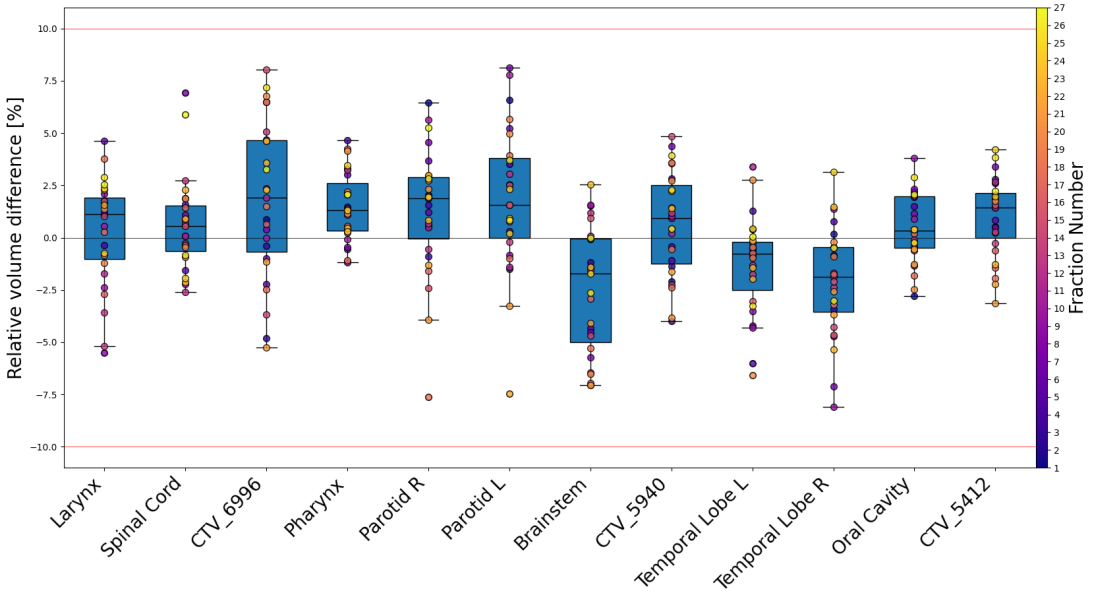

Figure 14: Relative volume differences for structures bigger than 10 cc for Case 1 and all simulated daily fractions, displayed from the smallest structure (left) to the largest (right). The red line indicates the QA threshold of 10%.

7.2 Case with challenging propagation results

As detailed in subsection 3.3, Case 3 required two replanning sessions during treatment. The first replanned CT (repCT) was acquired in fraction 5, with the treatment plan updated and applied starting from fraction 9. Similarly, the second repCT was performed on the day of fraction 14, and the corresponding updated plan was used from fraction 21 onwards.

The absolute difference criteria fail for structures larger than 4 cc, which are relatively small volumes compared to the other cases, as illustrated in figure 15. For instance, the esophagus, with a volume of 4.38 cc in the planning CT, fails the absolute difference criteria in three daily images.

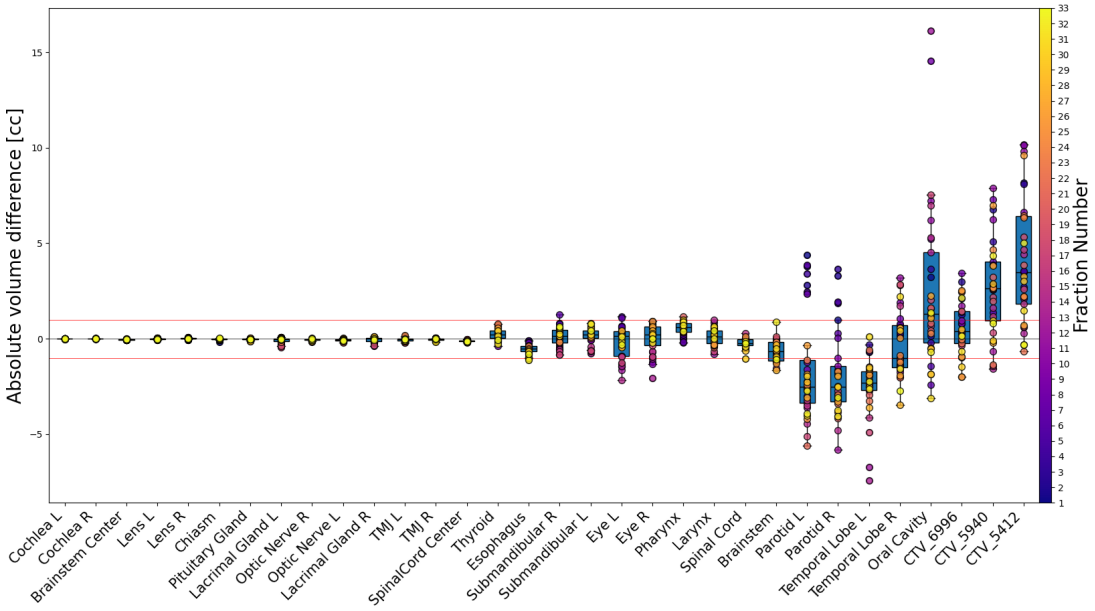

Figure 15: Absolute volume differences for Case 3 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc.

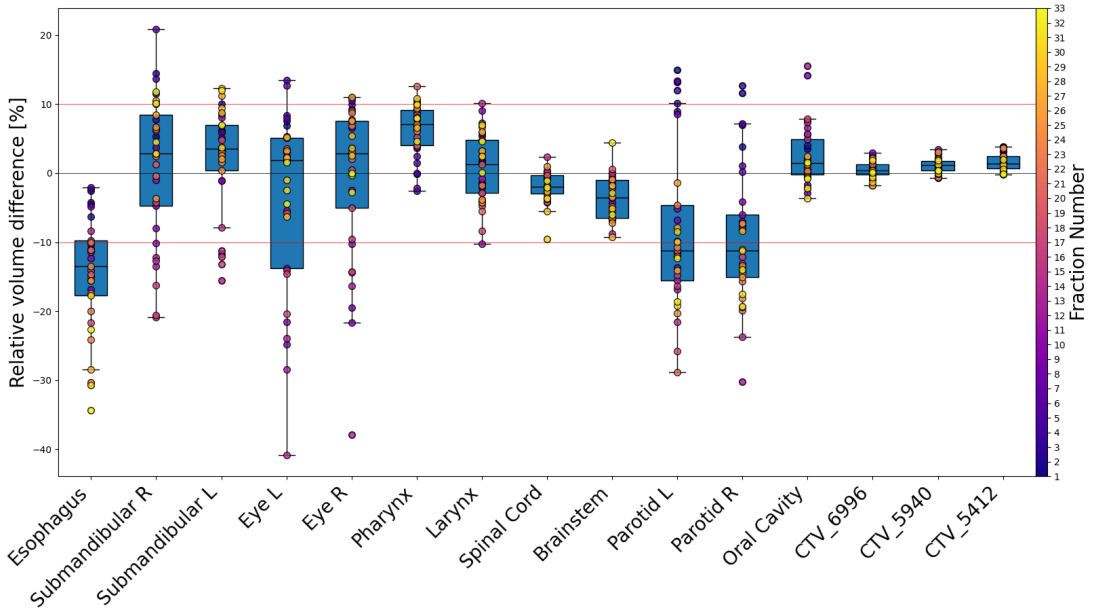

Figure 16: Relative volume differences for structures bigger than 4 cc for Case 3 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%.

Similarly, when evaluated using the relative difference criteria, figure 16, structures larger

than 4 cc also fail, indicating that these structures would not meet the requirements for the currently implemented structure QA in which one of the thresholds have to be passing.

In summary, for this case, it can be identified how many structures would fail the structure QA criteria, where the volume difference exceeds both ±1 cc and ±10%. Out of the 32 structures analyzed, 9 failed in at least one fraction, with 2 structures failing in more than 10 fractions, nearly one-third of the treatment. Based on these findings, it would be necessary to apply a more soft QA criteria of ±2.5 cc and ±35% for this case to ensure that all structures pass the QA evaluation across the daily fractions.

8 Comparison between adaptive and non-adaptive approaches

In this section, different datasets from the studied configurations (table 8) are presented to analyze the adaptive and conventional approaches in both gantries.

For each fraction, two treatment simulations were performed, daily adapted and conventional, both compared to respective fallback plan serving as a baseline (standard-of-care).

For each patient it was evaluated absolute differences between achieved and baseline values expressed in relative units. For OAR metric the dose constraints were also included in the plot, to identify where dose metrics either exceeded or significantly fell below the prescribed constraint values. As discussed in section 6.2, the Wilcoxon test results showed if the two distributions (adapted and conventional treatment simulation daily doses) are statistically significant.

Two specific cases are shown: one where the adaptive approach performed well and another where it showed less favorable results compared to conventional treatment. The results for all other cases can be found in appendix C, as well as in the summary of each subsection, where the Wilcoxon test result for all studied cases are presented.

8.1 Comparison between adaptive in G2 (full RO) vs non adaptive in G2 and non adaptive in G3

The second and third configuration were exclusively performed in RayStation® and explored its potential in adaptive treatments at G2 in relation to the conventional approach in both G2 and G3, as specified in table 8.

All results are presented as the absolute difference in relation to the baseline plan which in case of these configurations is fallback plan optimized with full RO (3mm,3%) settings at G3. That results from the fact that at PSI, G3 is standardly used for HN treatments and, therefore, referred to in the study as the standard of care.

The comparison of both configurations together allowed not only to evaluate the adaptive versus conventional treatment approach, but also the conventional approaches between both gantries, G2 and G3.

8.1.1 Case with favorable adaptive performance

For Case 4, the adaptive treatment demonstrated better target coverage and OAR sparing compared to both conventional treatments, particularly in fractions requiring replanning.

The conventional treatment in G2 demonstrated superior target coverage compared to the conventional plan in G3 (figure 17). However, the simulated daily dose to OARs was higher for half of the evaluated metrics on G2, indicating that, for this scenario, OAR sparing might

be more effective with the conventional treatment on G3 (when the adaptive approach is not considered), as shown in figure 18.

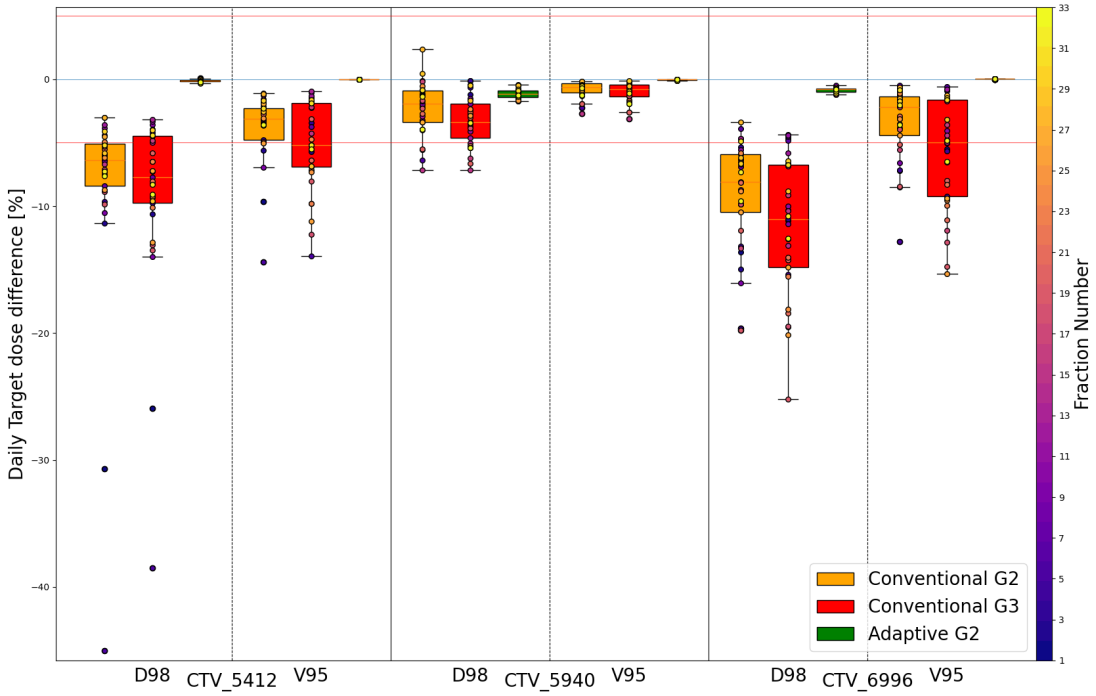

Figure 17: Case 4: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels

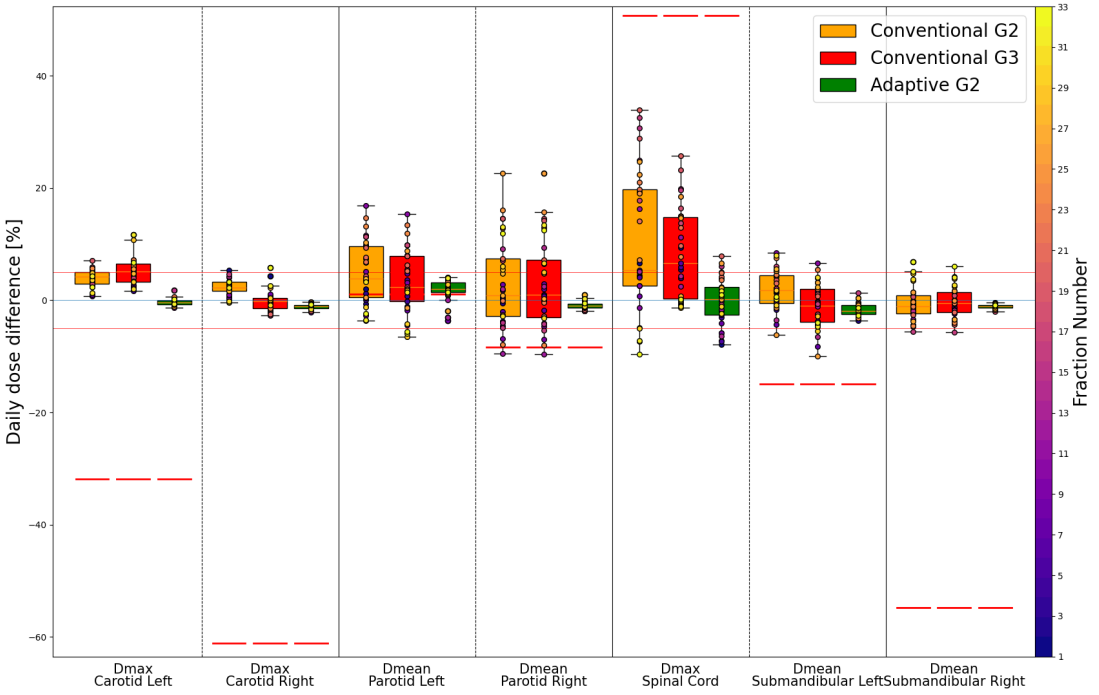

Figure 18: Case 4: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

In this case, both carotids, as well as the right parotid and right submandibular gland, were located near the high-dose target, as discussed in section 3.4. For both configurations, the adaptive daily dose was lower for these three structures. However, no significant difference between the conventional treatments in both gantries was observed for the right parotid and submandibular gland, as shown in figure 18.

The Wilcoxon test comparing the two conventional treatments revealed that 53.3% of the target metrics showed no significant difference, while the remaining metrics indicated better target coverage for the conventional treatment on G2. For the OAR metrics, 33.3% exhibited no significant difference, whereas 50.0% demonstrated a lower simulated delivered dose with the conventional treatment on G3.

When comparing the adaptive approach to both conventional treatments, the Wilcoxon test yielded similar results for both configurations. All metrics for both configurations demonstrated equal or superior dose distribution with the adaptive approach, both in terms of OAR sparing and target coverage. Notably, in this case there was no analyzed metric where either of the conventional approaches outperformed the adaptive treatment.

8.1.2 Case with limited adaptive performance

For Case 2, as shown in figures 19 and 20, the adaptive plan appears comparable to the conventional approaches in most metrics.

The target dose in the adaptive approach showed an improvement, particularly in the D98% metric, when compared to the conventional treatment simulation at G3. In contrast, the conventional treatment simulation at G2 demonstrated similar coverage. However, the average dose difference across all distributions did not exceed 1%, as shown in figure 19.

For the selected OARs shown in figure 20, no clear trend was observed in the simulated delivered dose with the adaptive treatment compared to the conventional treatments. In particular, the superior pharyngeal constrictor, a restrained organ for this case, showed no advantage in terms of D_{mean} when compared to conventional treatment in G2.

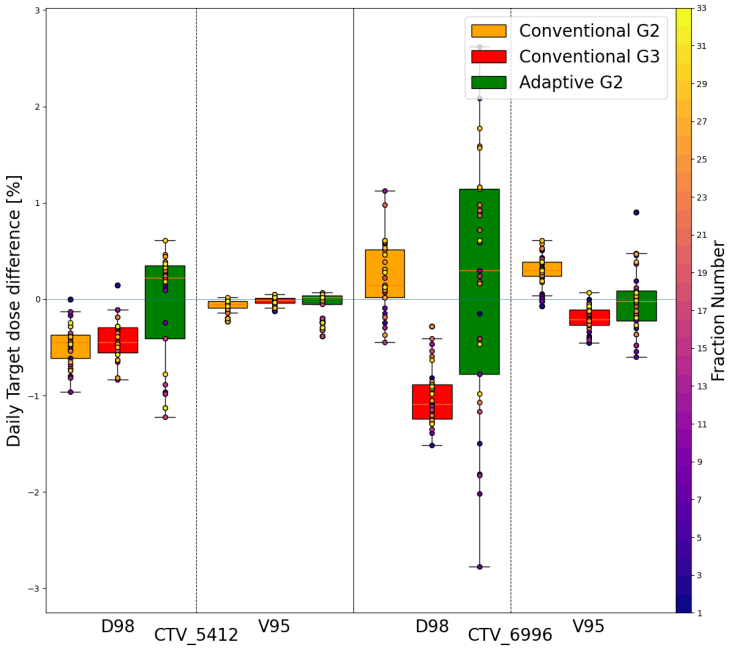

Figure 19: Case 2: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels

For Case 2, the comparison of the two conventional approaches showed that conventional treatment in G2, could be better than in G3. Where for target metrics, 70% had better coverage for G2, and for OAR metrics, 20% showed no significant difference between the conventional treatment and 55% showed a lower simulated dose in G2 conventional treatment. Table 9 has a summary of all conventional daily dose metrics.

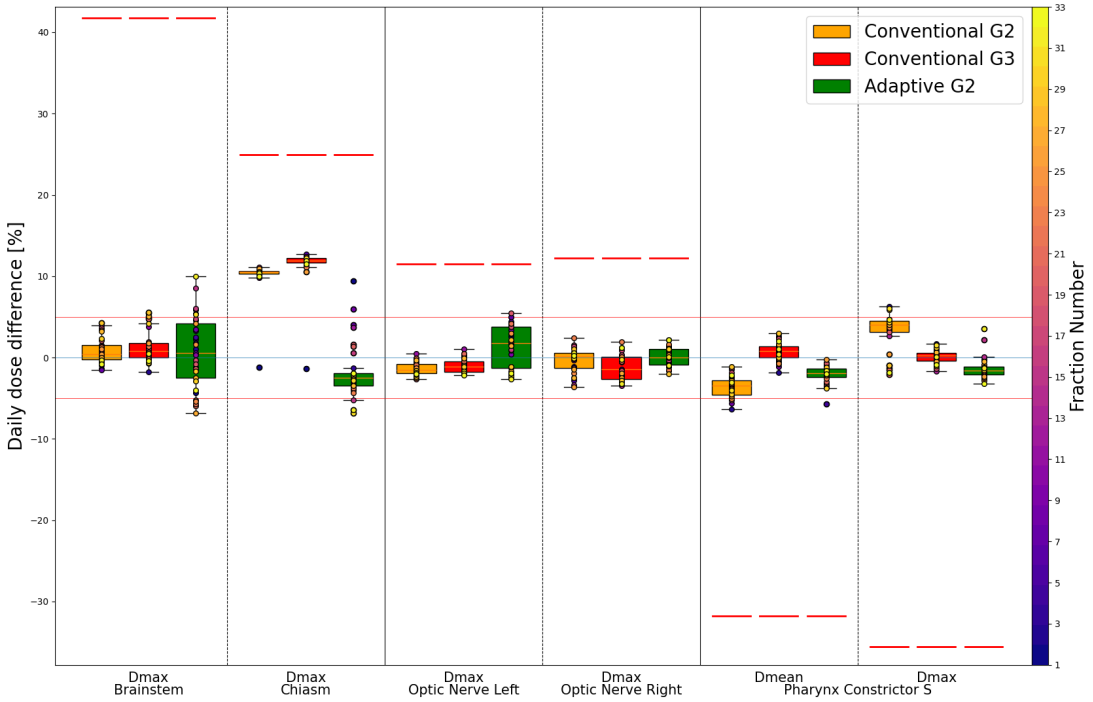

Figure 20: Case 2: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

The Wilcoxon test showed that the adaptive approach demonstrated no notable improvement for target coverage in both configurations, with roughly half of the target metrics showing no significant difference in the daily simulated doses between the adaptive approach and the conventional in both Gantries. For OAR metrics, the adaptive approach yielded better dose distribution compared to conventional treatments. On average, 25.0% of the OAR metrics exhibited no significant difference, whereas 42.5% indicated a reduction in average daily dose with the adaptive approach.

8.1.3 Summary of the second and third configuration

The second and third configuration allowed not only to compare adaptive and conventional treatment approaches, but also to compare conventional treatments simulated at G2 and G3. In all analyzed cases (see appendix C), the differences between these two machines were not pronounced, and their detailed comparison is presented in table 9.

Table 9: Comparison between conventional treatments, at G2 and G3. The percentages indicate the metrics that performed better for each respective Gantry or showing No Difference (N.D.) (not statistically significant).

Case	**Metric**	**Conv. G3**	**N.D.**	**Conv. G2**
1	Target	60.0%	20.0%	20.0%
	OAR	60.0%	11.4%	28.6%
2	Target	30.0%	0.0%	70.0%
	OAR	25.0%	20.0%	55.0%
3	Target	53.3%	40.0%	6.7%
	OAR	37.9%	24.1%	37.9%
4	Target	0.0%	53.3%	46.7%
	OAR	50.0%	33.3%	16.7%
5	Target	13.3%	20.0%	66.7%
	OAR	33.3%	15.4%	51.3%
Mean (SD)	Target	31.3% (25.6%)	26.7% (20.5%)	42.0% (28.0%)
	OAR	41.3% (13.8%)	20.9% (8.5%)	37.9% (15.9%)

When comparing the adaptive treatment with both conventional approaches is possible to see that for all the Cases most metrics would have a benefit with adaptive treatment, i.e. the daily dose distribution of the treatment could be improved by online adaptation. The comparison between the adaptive and both conventional approaches was largely similar, with the most significant difference between tables 10 and 11 being observed in target coverage for Case 2 and Case 5. These cases showed a larger difference between the conventional treatments, as indicated by table 9.

Table 10: RayStation, Configuration 2: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach.

Case	**Metric**	**Adaptive**	**N.D.**	**Total**
1	Target	86.7%	6.7%	93.3%
	OAR	68.6%	14.3%	82.9%
2	Target	20.0%	50.0%	70.0%
	OAR	40.0%	30.0%	70.0%
3	Target	100.0%	0.0%	100.0%
	OAR	41.4%	51.7%	93.1%
4	Target	86.7%	13.3%	100.0%
	OAR	94.4%	5.6%	100.0%
5	Target	40.0%	6.7%	46.7%
	OAR	64.1%	12.8%	76.9%
Mean (SD)	Target	66.7% (34.6%)	15.3% (19.9%)	82.0% (23.3%)
	OAR	61.7% (22.4%)	22.9% (18.4%)	84.6% (12.1%)

Table 11: RayStation, Configuration 3: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach.

Case	Metric	Adaptive	N.D.	Total
1	Target	46.7%	46.7%	93.3%
	OAR	54.5%	18.2%	72.7%
2	Target	50.0%	50.0%	100.0%
	OAR	45.0%	20.0%	65.0%
3	Target	100.0%	0.0%	100.0%
	OAR	62.1%	34.5%	96.6%
4	Target	100.0%	0.0%	100.0%
	OAR	77.8%	22.2%	100.0%
5	Target	80.0%	20.0%	100.0%
	OAR	64.1%	7.7%	71.8%
Mean (SD)	Target	75.3% (26.0%)	23.3% (24.3%)	98.7% (3.0%)
	OAR	60.1% (12.6%)	24.3% (9.5%)	81.5% (15.7%)

8.2 Comparison between adaptive G2 vs adaptive in G3 vs non adaptive in G3

The Configuration 4, as described in section 5, was set up to provide an idea of a possible adaptive method in G3. It aims to give a first analysis of the adaptive daily dose distribution and to evaluate the potential improvements from establishing the adaptive process in this gantry.

The configuration 4 can be paired with the results of configuration 3 since both of configurations have the same conventional plan. Now is possible to access both adaptive approaches, in G2 and in G3 against the standard-of-care of HN cases in G3 at PSI. All results are presented as the absolute difference in relation to the baseline plan which is fallback plan optimized with full RO (3mm,3%) settings at G3.

8.2.1 Case with favorable adaptive performance

For Case 4, the adaptive approach in G2 and in G3 demonstrated its ability to improve target coverage, accounting for patient daily changes and reducing the need for replanning.

In figure 21, the precise tumor coverage achieved through adaptation during the course of treatment is clearly visible. However, the comparison between the two adaptive approaches are not clearly evident.

For the OARs, most metrics showed a lower daily simulated dose for the adaptive approach in G2. Specifically, the right parotid and submandibular gland had a dose reduction of less than 3% of dose with the adaptive approach in G2, when compare to adaptive in G3, as shown in figure 22.

The Wilcoxon test results indicated a superior dose distribution for the adaptive treatment in G3. Only 6.7% of target metrics showed no significant difference, while the remaining metrics demonstrated better coverage with the adaptive approach. For OAR metrics, 52.6% showed no significant difference, and 47.4% indicated a lower dose with the adaptive treatment. In summary, all analyzed metrics showed better or equal results for the adaptive approach in G3 compared to conventional in G3.

When comparing the two adaptive treatments, adaptation in G2 resulted in a superior daily dose distribution. Specifically, 26.7% of target metrics showed no significant difference, while

53.3% demonstrated better coverage with adaptation in G2. For OAR metrics, this difference was even more pronounced with 27.8% showing no significant difference and 61.1% indicating a lower dose for adaptive treatment in G2.

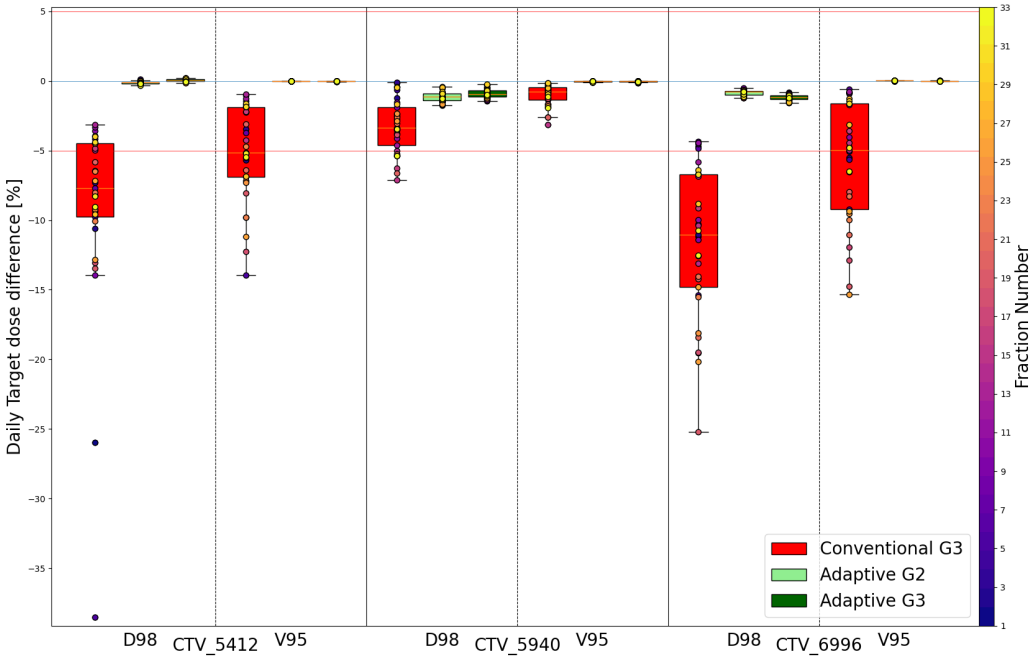

Figure 21: Case 4: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels.

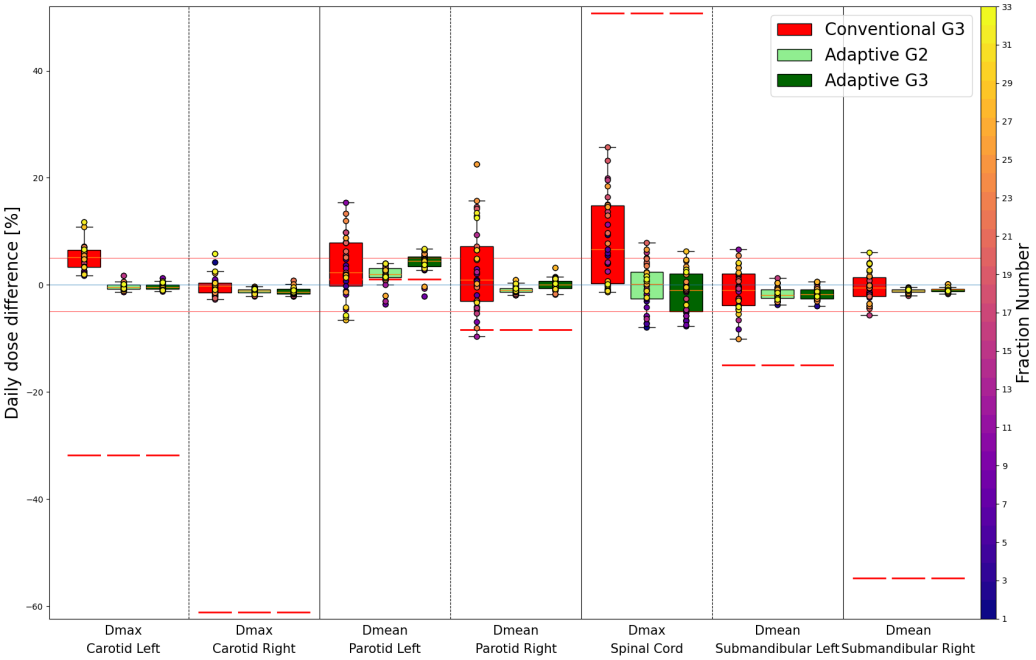

Figure 22: Case 4: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

8.2.2 Case with limited adaptive performance

For Case 2, the adaptive treatment in G3 showed for some metrics a better daily dose distribution in relation to conventional standard-of-care treatment.

As shown in figure 23, the target coverage in this adaptive approach showed an improvement in relation to the conventional treatment. However, the target metrics show slightly lower coverage compared to the adaptive approach in G2, although the differences are less than 1% of the prescription dose.

The figure 24 shows the distribution of daily doses for a set of selected OARs, the three plots are very similar, with exception for the chiasm, structure that in both adaptive scenarios received a significant lower dose compared to conventional. For this specific case, the superior pharyngeal constrictor metrics showed a lower dose for both adaptive approaches, with a reduction of less than 2% of the prescription dose for the adaptive approach in G2.

The Wilcoxon test showed that 40.0% of target metrics had no difference between the adaptive and conventional approach in G3, and 50.0% showed a better coverage with adaptation. For the OAR metrics 5% of metrics showed no significant difference and 50% showed lower simulated delivered dose to adaptive approach.

When comparing both adaptive treatments, adaptive treatment in G2 seems to be better for this Case, 70% of target metrics distributions had no significant difference and the 30% rest had better coverture for G2. For OAR metrics, 45% had no significant difference and 35% lower dose with adaptive in G2, as shown in the table 13.

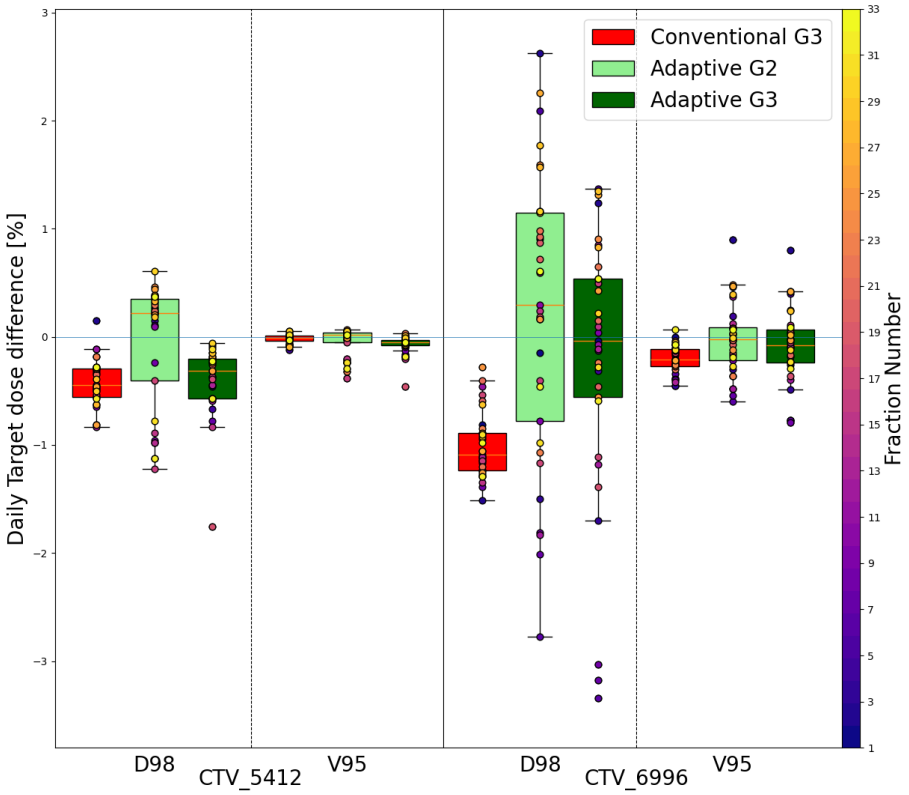

Figure 23: Case 2: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels.

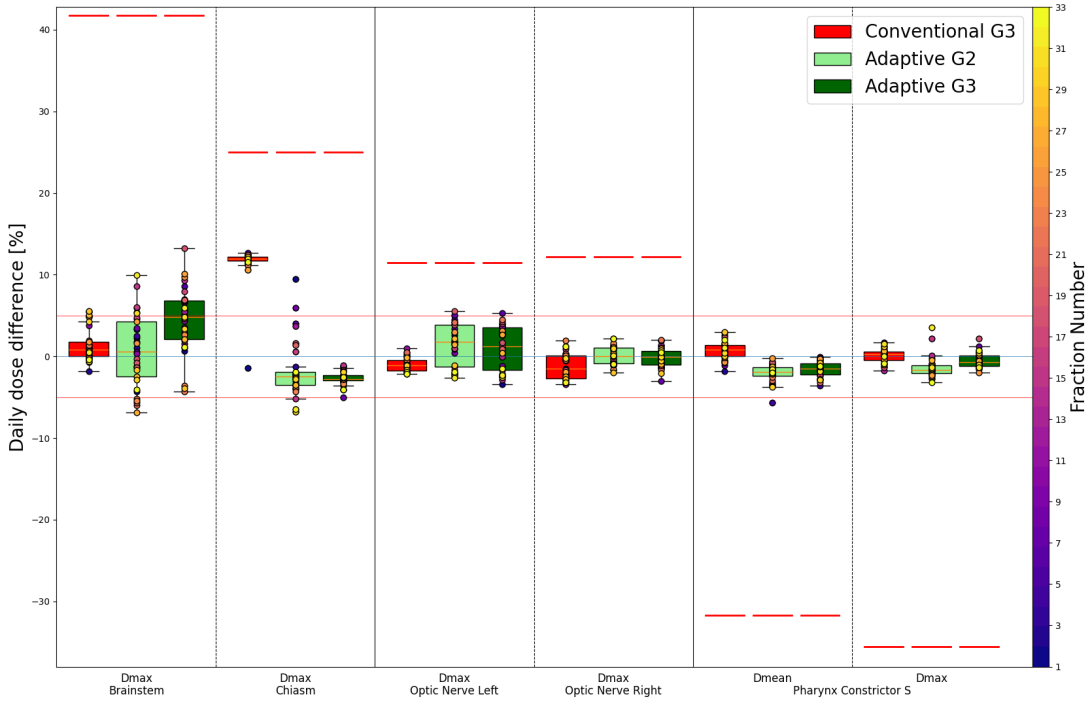

Figure 24: Case 2: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

8.2.3 Summary of third and fourth configuration

The results obtained in Configuration 3, which cross-compared the gantries with the adaptive approach in G2 and the conventional approach in G3, were already discussed and presented in section 8.1.3.

The configuration 4 compared the adaptive treatment with a reduced setup uncertainty (2%, 3mm) to the standard-of-care treatment in G3. The results demonstrated promising dose distribution favoring the adaptive treatment in G3.

As shown in table 12, the percentage of metrics where target coverage was improved or where the OAR received a lower dose with the adaptive approach is highlighted. The results were consistent with other configurations, with a notable improvement in adaptive treatment, especially for patients requiring replanning. Notably, in Case 3 and Case 4, the values were almost always close to 100%, indicating that the adaptive treatment could provided better or equal dose distribution specially in cases where replanning is needed.

Comparing the third and fourth configurations allowed a closer examination of the behavior between the two adaptive approaches. Namely, in G2, where the adaptive process is implemented, and in G3, where the current standard treatment for HN cases is applied. Comparing both treatments also give an ideia of the benefits in implementing an adaptive process in G3.

Table 13, show the percentage of metrics where the dose distribution showed a better target coverage or lower OAR dose for each Gantry. In contrast to the comparison with conventional treatments in both Gantries (table 9) the adaptive approach showed a difference, where, except for Case 1, the simulated dose distribution with adaptation in G2 appear to yield better results.

Adaptation in G3 may not necessarily be required, although it could still lead to better results compared to the conventional approach, and its implementation would represent a valuable enhancement to the treatments at PSI.

Table 12: RayStation, Configuration 4: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach

Case	**Metric**	**Adaptive**	**N.D.**	**Total**
1	Target	66.7%	0.0%	66.7%
	OAR	60.0%	22.9%	82.9%
2	Target	50.0%	40.0%	90.0%
	OAR	50.0%	5.0%	55.0%
3	Target	80.0%	20.0%	100.0%
	OAR	48.3%	41.4%	89.7%
4	Target	93.3%	6.7%	100.0%
	OAR	47.4%	52.6%	100.0%
5	Target	46.7%	40.0%	86.7%
	OAR	59.0%	15.4%	74.4%
Mean (SD)	Target	67.3% (19.8%)	21.3% (18.5%)	88.7% (13.7%)
	OAR	53.5% (5.6%)	26.9% (18.5%)	80.4% (17.0%)

Table 13: Comparison between adaptive treatments, at G2 and G3. The percentages indicate the metrics that performed better for each respective Gantry or showing No Difference (N.D.) (not statistically significant).

Case	**Metric**	**Adapt. G3**	**N.D.**	**Adapt. G2**
1	Target	40.0%	26.7%	33.3%
	OAR	42.9%	31.4%	25.7%
2	Target	0.0%	70.0%	30.0%
	OAR	20.0%	45.0%	35.0%
3	Target	6.7%	13.3%	80.0%
	OAR	27.6%	13.8%	58.6%
4	Target	20.0%	26.7%	53.3%
	OAR	11.1%	27.8%	61.1%
5	Target	20.0%	6.7%	73.3%
	OAR	12.8%	25.6%	61.5%
Mean (SD)	Target	17.3% (15.3%)	28.7% (24.7%)	54.0% (22.7%)
	OAR	22.9% (12.9%)	28.7% (8.9%)	48.4% (16.8%)

9 Comparison between FIonA and RaySta-tion

As detailed in table 8, this study examined four configurations, analyzing different adaptive and conventional approaches. The results from FIonA and RayStation can be analyzed together in the first configuration, as both involve adaptive planning with hybrid robustness optimization. In the first configuration, the adaptive plan was designed with an optimization target defined as the CTV, with a smaller margin (2 mm). In contrast, the conventional plan in FIonA utilized PTV optimization, whereas in RayStation®, the plan was optimized directly on the CTV with full robustness optimization (3 mm, 3%), as previously mentioned.

Here, the observed differences between the two TPS are primarily attributed to the dose calculation algorithm, analytical in FIonA versus Monte Carlo (MC) in RayStation®, as well as

differences in the optimization process, with RayStation's optimization algorithm being superior. These factors contribute to variations in dose distribution and plan robustness between the two systems.

Two representative cases are presented here: one where the adaptive approach performed well and another where it did not lead to significant improvements. The remaining cases are detailed in appendix C.

9.1 Case with favorable adaptive performance

Case 4 required two replannings during treatment, with the first replanned CT performed on the first day of treatment. As shown in figures 25 and 26, there was a clear underdosage in the low-dose target during the initial treatment days, particularly before the first replanned fraction, delivered on the sixth day of treatment. The adaptive approach improved target coverage for both TPS, addressing the low coverage observed during the initial fractions, which corresponded to the period when the need for replanning had already been identified but not yet implemented.

Although no significant differences are apparent between the target coverages in figures 25 and 26, the DVH in the figure 75 of appendix D shows improved dose conformity for the mid-dose target in the adaptive treatment in RayStation®.

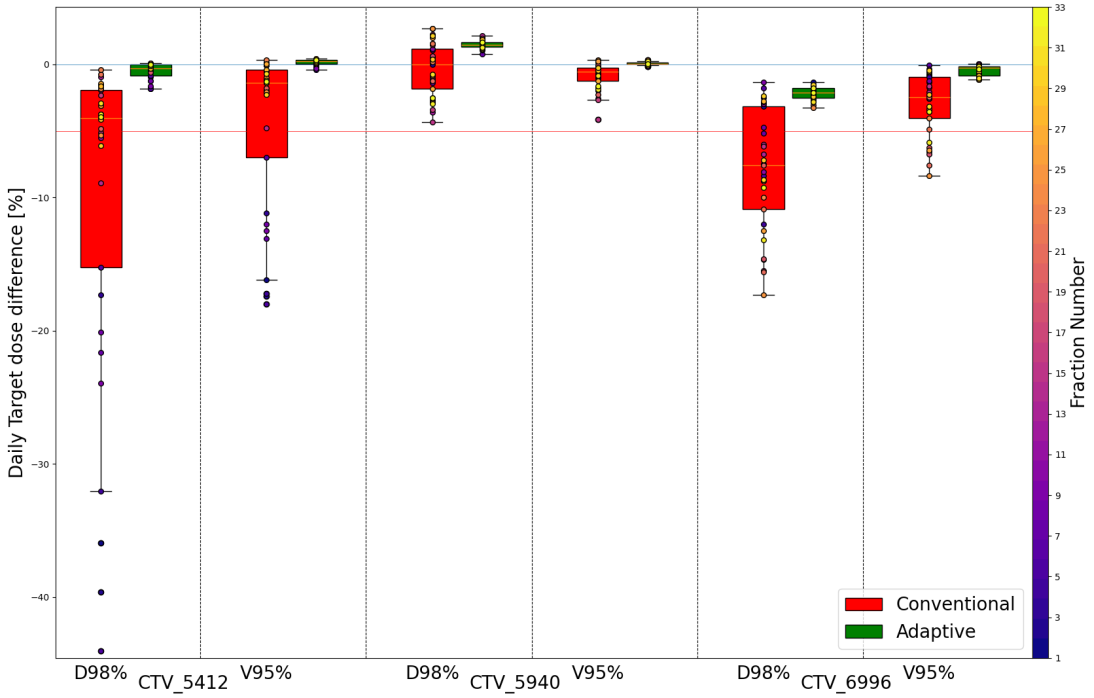

Figure 25: FIonA daily target dose differences, D98% and V95%, for Case 4.

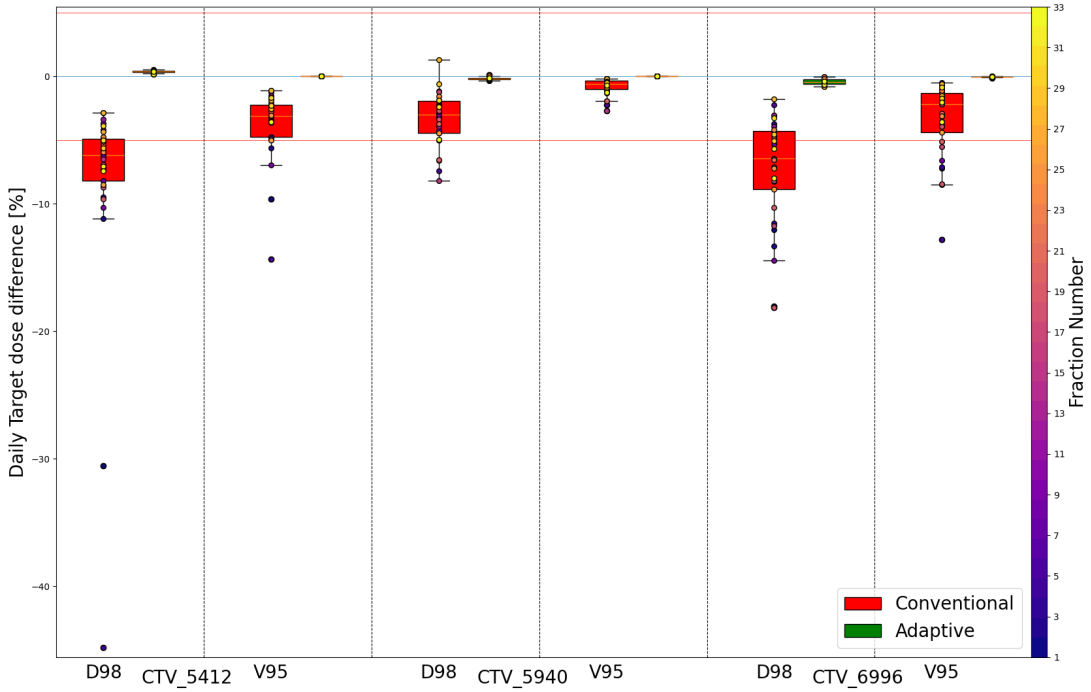

Figure 26: RayStation® daily target dose differences, D98% and V95%, for Case 4.

The simulated delivered dose to the OARs shows a noticeable improvement with the adaptive approach for both TPS, as illustrated in figures 27 and 28. Across all metrics, the dose was consistently lower with adaptation, even for organs adjacent to the high-dose target, such as the right carotid and the right submandibular.

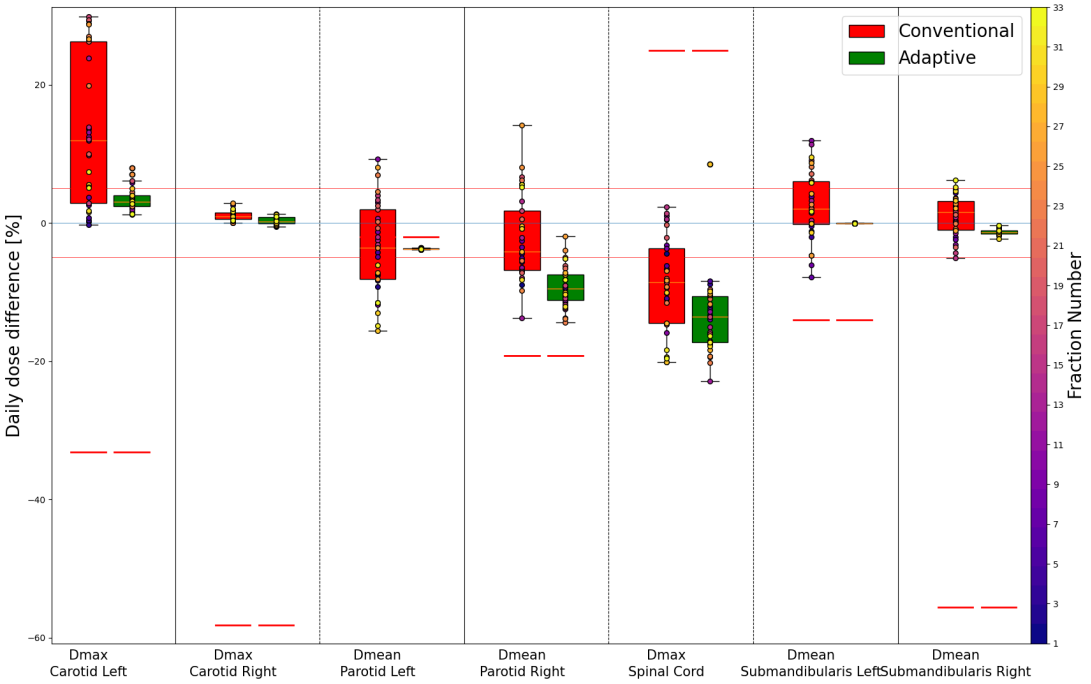

Figure 27: FIonA daily dose differences for some selected organs of Case 4, the red line represents the dose constraint defined in the clinical goal.

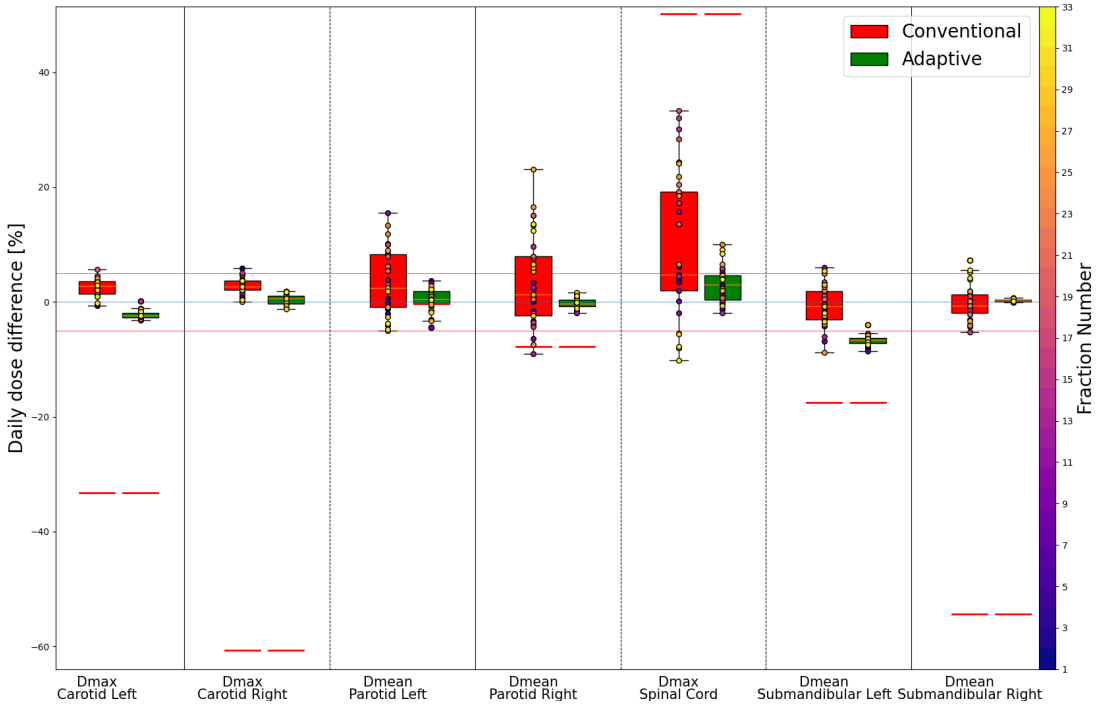

Figure 28: RayStation® daily dose differences for some selected organs of Case 4, the red line represents the dose constraint defined in the clinical goal..

The Wilcoxon test demonstrated again similar results for both TPS and significant improvements with adaptive treatment (also seen in tables 14 and 15 in section 9.3).

For the target metrics, both TPS demonstrated an average of 90.0% of metrics with improved or equal coverage. For the OAR metrics, all of them showed either a lower or equal simulated delivered dose with adaptive treatment for both TPS.

9.2 Case with limited adaptive performance

For the Case 2 the simulated doses for adaptive and conventional treatment scenarios, and two TPS, are presented in figures 29, 30, 31 and 32.

For this case, all target coverage metrics were better with adaptive treatment for both TPS. However, in the RayStation® results, a slightly difference was observed for V95% of the high-dose target, where the adaptive treatment showed a little lower dose, with a difference of approximately 0.5%, as showed in figure 30.

Regarding the OARs, figure 31 illustrates that, in FIonA, the adaptive approach resulted in a higher maximum dose to the superior pharyngeal constrictor for nearly every fraction. On the other hand, the mean dose was consistently lower with the adaptive treatment. This effect was not observed in the RayStation® results (figure 32), where both the maximum and mean doses for the pharyngeal constrictor were reduced with the adaptive approach. Interestingly, this dose reduction in RayStation® results corresponds to a small reduction in the coverage of the high-dose target, likely due to prioritizing sparing of this OAR.

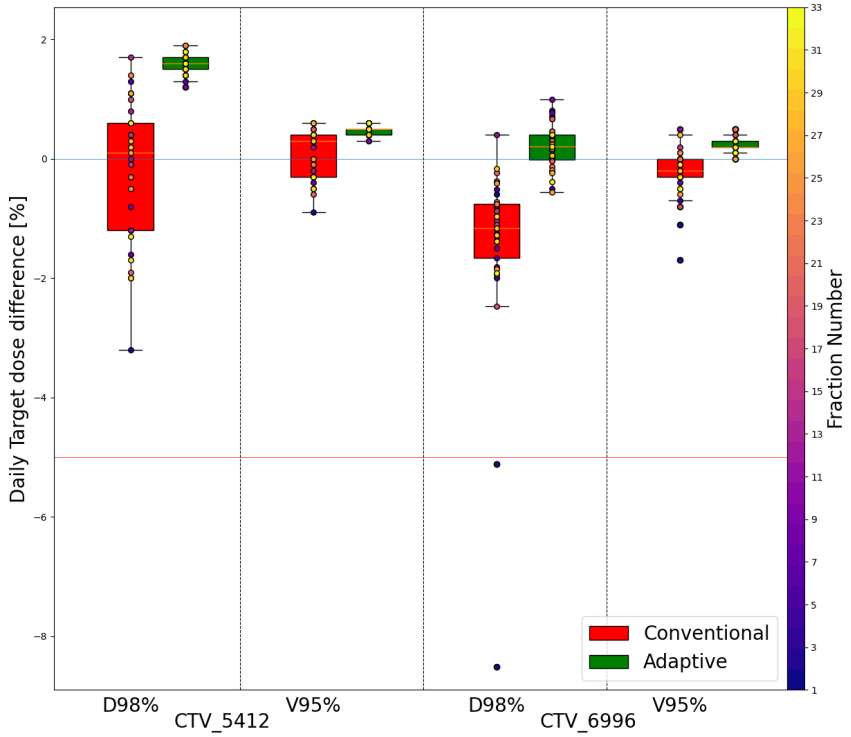

Figure 29: FIonA daily target dose differences, D98% and V95%, for Case 2.

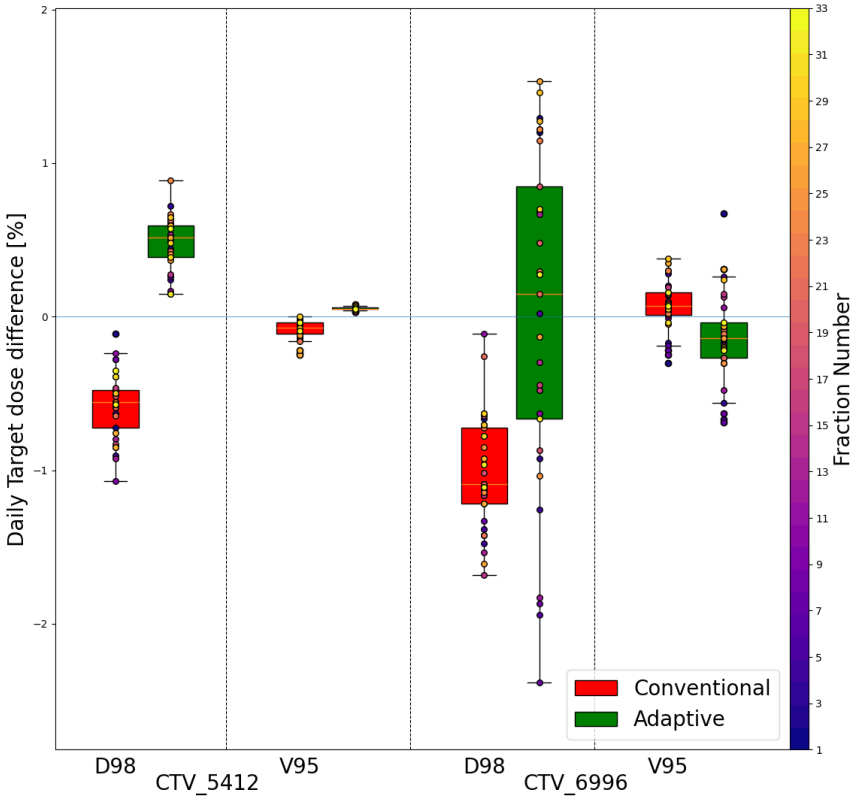

Figure 30: RayStation® daily target dose differences, D98% and V95%, for Case 2.

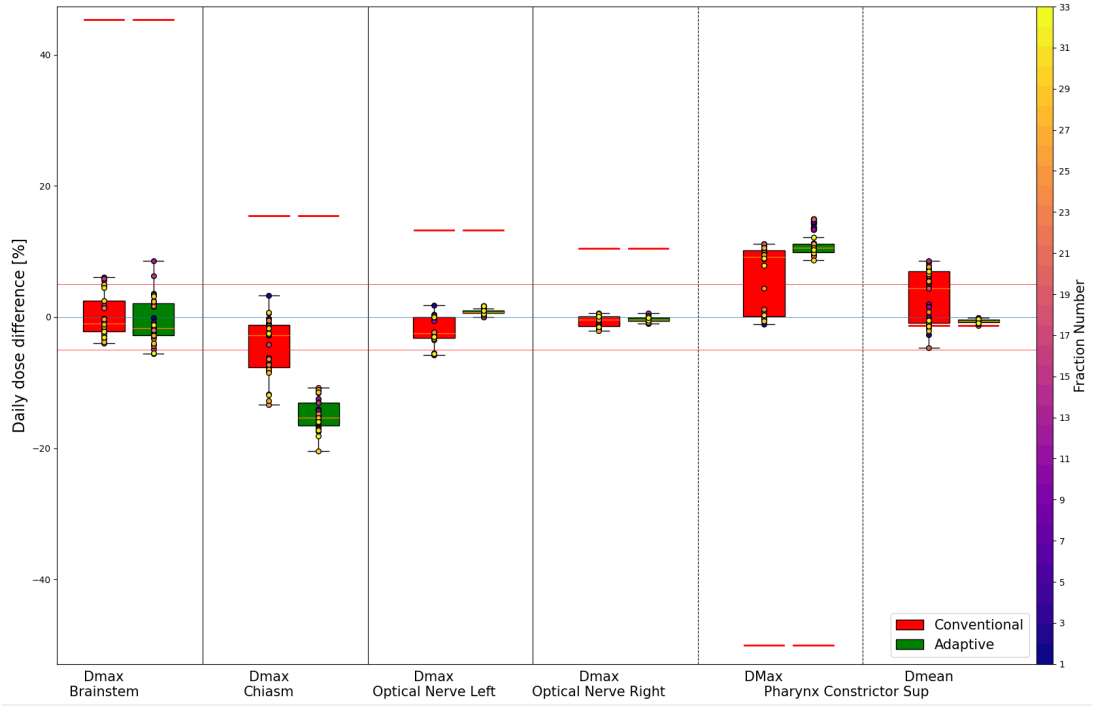

Figure 31: FIonA daily dose differences for some selected organs of Case 2, the red line represents the dose constraint defined in the clinical goal.

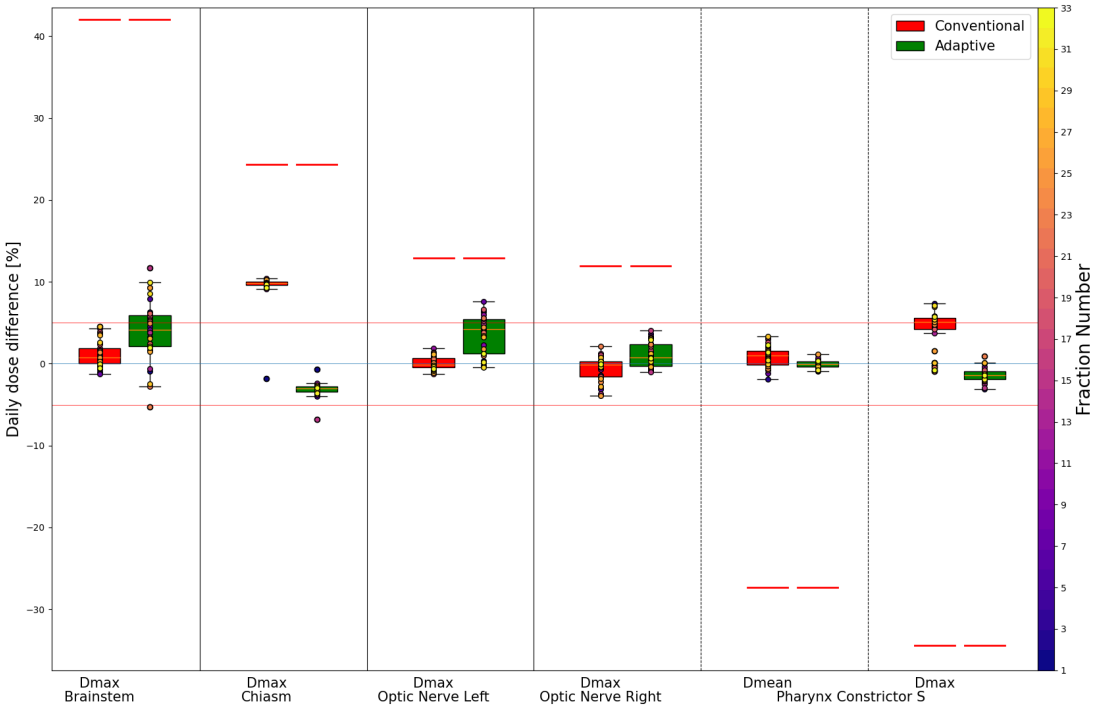

Figure 32: RayStation® daily dose differences for some selected organs of Case 2, the red line represents the dose constraint defined in the clinical goal.

Using the Wilcoxon test, it was observed that the results from the two TPS were similar in terms of the simulated daily dose distribution. The adaptive approach in both TPS demonstrated

an improvement in target coverage. Regarding the target dose, the adaptive treatment showed an improvement, with on average 75.0% of the metrics indicating better or comparable coverage with adaptive treatment. For the OAR dose, the simulated delivered dose was better for the adaptive approach in approximately 57.5% of the cases, showing no significant difference between the conventional and adaptive treatment in this case.

9.3 Summary of the first configuration

As discussed, the first configuration was performed for both TPS, in order to estimate differences between two planning systems, with one of them currently used clinically for daily adaptive proton treatments.

All Wilcoxon test results are summarized in the tables 14 and 15. Analyzing the two tables reveals that the results of both TPS were comparable for OARs metrics. However, for target metrics, a lower average was observed in FIonA compared to RayStation®, which is more evident in the target DVH plots in appendix D, where the FIonA results differ qualitatively from those obtained with RayStation®.

When analyzing the adaptive approach with hybrid RO in relation to the conventional approach, both TPS showed similar results for OAR doses. Except for Case 2, the simulated delivered dose to most OARs was lower in the adaptive approach compared to conventional treatment. Regarding target coverage, all cases demonstrated better coverage with the adaptive treatment, a result consistent with the other configurations studied..

Table 14: FIonA, Configuration 1: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach.

Case	**Metric**	**Adaptive**	**N.D.**	**Total**
1	Target	60.0%	6.7%	66.7%
	OAR	71.4%	11.4%	82.9%
2	Target	70.0%	0.0%	70.0%
	OAR	30.0%	30.0%	60.0%
3	Target	73.3%	26.7%	100.0%
	OAR	72.4%	13.8%	86.2%
4	Target	73.3%	13.3%	86.7%
	OAR	93.8%	6.3%	100.0%
5	Target	66.7%	6.7%	73.3%
	OAR	66.7%	10.3%	76.9%
Mean (SD)	Target	68.7% (5.6%)	10.7% (10.1%)	79.3% (13.8%)
	OAR	66.9% (23.1%)	14.3% (9.2%)	81.2% (14.6%)

Table 15: RayStation, Configuration 1: Percentage of dose metrics either better (statistically significant) in Adaptive or showing No Difference (N.D.) (not statistically significant) when compared to conventional approach

Case	**Metric**	**Adaptive**	**N.D.**	**Total**
1	Target	93.3%	6.7%	100.0%
	OAR	60.0%	17.1%	77.1%
2	Target	70.0%	10.0%	80.0%
	OAR	55.0%	0.0%	55.0%
3	Target	100.0%	0.0%	100.0%
	OAR	44.8%	41.4%	86.2%
4	Target	86.7%	6.7%	93.3%
	OAR	62.5%	37.5%	100.0%
5	Target	93.3%	0.0%	93.3%
	OAR	53.8%	17.9%	71.8%
Mean (SD)	Target	88.7% (11.5%)	4.7% (4.5%)	93.3% (8.2%)
	OAR	55.2% (6.8%)	22.8% (16.9%)	78.0% (16.7%)

Part IV
Discussion

This study investigated the feasibility and advantages of daily adaptive proton therapy for head and neck cancer patients, with a particular focus on target coverage and organ-at-risk (OAR) sparing. The findings demonstrate that adaptive proton therapy consistently outperforms conventional treatment approaches, as shown by statistically significant improvements in dose distribution. These results align with prior research indicating the potential of adaptive strategies to improve treatment accuracy, particularly in cases with anatomical changes.

The first explored aspect was the investigation if currently implemented DAPT workflow could support deforming anatomies, i.e., HN cases, and if the structure QA framework would be applicable also beyond rigid anatomical regions. The results confirm that the current QA protocol is insufficient for deformable anatomies, as three out of five cases did not meet the established threshold criteria when adaptive planning steps were followed. This finding underscores the necessity for a more robust QA framework tailored to deformable image registrations.

In this study, the propagated structures were visually checked; however, no visual inspection was performed by a medical doctor (MD), a step that is typically included in the DAPT process developed at PSI, crucial specially in the target automatic segmentation [15]. Additionally, verifying the center of mass (COM) for each daily propagated structure might be particularly relevant for deformable registrations, as it is already part of the current structure QA at PSI. Assessing COM shifts could provide additional insights into structure propagation accuracy and might be considered in future evaluations [1].

While the development of an alternative QA strategy was not the primary focus of this work, the findings suggest that refining the current framework could be beneficial. Future research should investigate a broader set of cases to identify potential trends and further validate the applicability of existing thresholds. Additionally, visual inspection by MDs, along with the integration of complementary QA steps such as COM analysis, or even the division in small and large structures, could contribute to a more robust evaluation of deformable registration accuracy in clinical practice. Given that studies have shown larger structure volume variations when using deformable registration [16], careful validation remains essential to ensure its reliability in adaptive workflows.

Moreover, this study considered various treatment configurations, accounting for the availability of different treatment planning systems (TPS) and delivery machines. Across all configurations, the adaptive approach demonstrated advantages over the conventional approach, regardless of the treatment planning system (TPS) or treatment machine used. This advantage was observed for both target coverage and OAR sparing.

For cases requiring replanning during treatment (Cases 3 and 4), the adaptive approach successfully retrieved planned dose distributions, demonstrating that offline replanning would not have been necessary if the adaptive workflow had been followed. As shown in tables 10, 11, 12, 14 and 15 the adaptive approach performed equally well or better in nearly 100% of the analyzed metrics compared to the conventional approach. Even in cases where no replanning was required, approximately 80% of the analyzed metrics showed improvements with daily adaptation.

These results indicate that daily adaptation enhances treatment dose distribution and provides a robust alternative to offline replanning, which is time-consuming and may lead to suboptimal dose delivery during replanning fractions. However, actual treatment delivery depends on additional factors such as intrafraction motion and machine operation, which were not within the scope of this study.

It was further investigated whether differences between treatment machines influenced the effectiveness of the adaptive approach for the selected study cases. The analysis showed that in 28.7% of cases, there was no significant difference between the two gantries when considering both targets and OARs. However, in 54.0% of target metrics and 48.4% of OAR metrics, the adaptive approach in G2 outperformed that in G3. This suggests that the specific capabilities of G2 may contribute to enhanced treatment precision in adaptive workflows (as shown in table 13).

However, the dose differences were minimal for target metrics, whereas they were more pronounced for OAR metrics. As shown in section 8.2 and appendix C, the variation did not exceed 1% of the prescription dose (approximately 0.54 Gy RBE) for target metrics, while for OARs, it could exceed 10% (approximately 5.41 Gy RBE) for some metrics. This is likely due to the different beam delivery characteristics of G2, as discussed in section 2.

Adaptation at G3 also leads to significant improvements over conventional methods, and further refinement of its adaptive protocols could enhance treatment dose distribution even further. These findings suggest that integrating adaptive strategies as a standard practice in G3 could provide clinically relevant benefits and improve overall treatment efficacy.

In contrast, when comparing conventional treatment between the two gantries, no clear trend was observed (as shown in table 9). This result aligns with expectations, as both gantries are already used clinically with well-established conventional treatment methods that have demonstrated good outcomes in clinical practice.

When comparing both TPS, RayStation demonstrated a better daily simulated dose distribution compared to FIonA. With a more advanced calculation algorithm, the commercial TPS can provide a fast daily adaptive simulation. Although it is not currently implemented in the clinical framework of DAPT, it could serve as a valuable tool for further studies and the development of new clinical procedures in the future, particularly for online adaptation in G3.

Part V
Conclusions & Outlook

This thesis evaluated the benefits of daily adaptive proton therapy for head and neck cases. Additionally, the study assessed the current structure QA implemented for rigid registration and investigated its applicability to deformable registration. Furthermore, the work compared different treatment configurations using two treatment planning systems and the two available gantries at PSI, G2 and G3.

The current structure QA, as expected, is not sufficient for deformable anatomies. In three out of the five cases, the current QA process, which uses thresholds of 1 cc and/or 10%, would fail if the steps of the aDAPT protocol were followed, and new threshold values were found for every failed case. This result proofs the need for a more robust QA framework capable of addressing deformable registrations. Although alternative QA solutions were not proposed in this study, developing such solutions should be a focus of future research.

The results demonstrate the significant advantages of adaptive proton therapy in daily dose distribution, tumor coverage, and critical structure sparing. Across all configurations, adaptive strategies consistently outperformed conventional methods, as confirmed by the Wilcoxon test, which showed statistically significant improvements in both target coverage and OAR sparing. These findings highlight the promising clinical potential of adaptive proton therapy, especially for cases involving anatomical changes, where the replanning in the conventional treatment was needed.

In particular, the adaptive treatment in G2 showed superior dose distribution, while G3 also had satisfactory results and, outperformed the conventional method. These results suggest that integrating adaptive strategies into routine practice in G3 can provide better-tailored treatments and improve therapeutic daily dose distribution.

Building on these findings, further investigations could explore a larger patient cohort, as this study was limited to only five cases. Additionally, all pre-processed synthetic CT (synCT) images were generated in Velocity (Varian Medical Systems, Palo Alto, California, USA) and since the algorithms and methods used to create synCTs vary between software, an important next step would be to evaluate different synCT generation algorithms and their effects, for example, in RayStation® by incorporating into the adaptive process.

Furthermore, while this research focused on head and neck cases, future studies should explore the application of adaptive strategies in other anatomical regions where deformable registration plays a crucial role, such as the abdominal region.

In conclusion, this thesis contributes to the growing body of evidence supporting adaptive proton therapy as a significant advancement in proton therapy for head and neck cancer. By combining technological capabilities with clinical applications, it paves the way for more personalized, precise and effective cancer treatments, ultimately improving patient outcomes and quality of life.

Part VI

Appendix

A OAR metrics

The clinical goals were defined by MDs for each patient individually and based on QUANTEC [7] and DAHANCA [8] guidelines. Most of the metrics are available in DAHANCA. Between the two sources, the stricter value was considered for table 16.

Table 16: Clinical goals for various anatomical structures.

Structure	**Metric**	**Acceptable Value**
Brainstem	DMax	54.00 Gy RBE
Carotid	DMax	40.00 Gy RBE
Chiasm[(1)]	DMax	54.00 Gy RBE
Cochlea[(2)]	DMean	45.00 Gy RBE
Esophagus	DMean	30.00 Gy RBE
Eye	DMax	45.00 Gy RBE
Hippocampus	D40.0%	7.20 Gy RBE
Lacrimal Gland	DMean	25.00 Gy RBE
Larynx	DMax	40.00 Gy RBE
Lens[(3)]	DMax	7.00 Gy RBE
Lips	DMean	20.00 Gy RBE
Mandible	DMax	72.00 Gy RBE
Optic Nerve[(4)]	DMax	54.00 Gy RBE
OralCavity[(5)]	DMean	30.00 Gy RBE
Parotid[(6)]	DMean	26.00 Gy RBE
Pharynx[(7)]	DMean	50.00 Gy RBE
Pharyngeal Constrictors L	DMean	55.00 Gy RBE
Pharyngeal Constrictors M	DMean	55.00 Gy RBE
Pharyngeal Constrictors S[(8)]	DMean	55.00 Gy RBE
Pituitary[(9)]	DMean	20.00 Gy RBE
SpinalCord	DMax	45.00 Gy RBE
Submandibular	DMean	35.00 Gy RBE
Temporomandibular Joint[(10)]	DMax	54.00 Gy RBE
Temporal Lobe	DMax	72.06 Gy RBE
Thyroid	DMean	40.00 Gy RBE

These metrics were not always achieved depending on the anatomy and the target location. For some structures the constraint was altered for some cases, in order to allow a better target coverage. The metrics constraint values for each case are represented in table 17.

Table 17: Remarks in the dose metric table.

No.	**Structure**	**Remark**
(1)	Chiasm	For Cases 3 and 5: Dmax $<$ 56 Gy RBE.
(2)	Cochlea	For Cases 1 and 2, a constraint for Dmax was accepted instead of Dmean. For Case 1: Dmax $<$ 20 Gy RBE (left cochlea) and Dmax $<$ 35 Gy RBE (right cochlea). For Case 2: Dmax $<$ 40 Gy RBE for both cochleae.
(3)	Lens	For Cases 3 and 5: Dmean $<$ 7 Gy RBE.
(4)	Optic Nerve	For Case 2: Dmax $<$ 58 Gy RBE. For Cases 3 and 5, only the right optic nerve was altered: Dmax $<$ 60 Gy RBE and $D_{2\%} <$ 72.06 Gy RBE, respectively.
(5)	Oral Cavity	For Case 3: Dmax $<$ 79.80 Gy RBE was set alongside the Dmean constraint.
(6)	Parotid	For Case 1: Dmax $<$ 30 Gy RBE was set alongside the Dmean constraint. For Cases 3 and 4, the left parotid constraint: Dmean $<$ 20 Gy RBE.
(7)	Pharynx	For Case 1: Dmax $<$ 72.06 Gy RBE.
(8)	Pharyngeal Constrictors S	For Case 2: Dmean $<$ 46 Gy RBE.
(9)	Pituitary	For Case 1: Dmean $<$ 45 Gy RBE. For Case 2: Dmax $<$ 35 Gy RBE.
(10)	Temporomandibular Joint	For Case 3: Dmax $<$ 72.06 Gy RBE. For Cases 4 and 5: Dmax $<$ 68 Gy RBE.

B Structure QA - all cases

B.1 Case 2

For the second case, daily image registration was performed without any issues, similar to the first case.

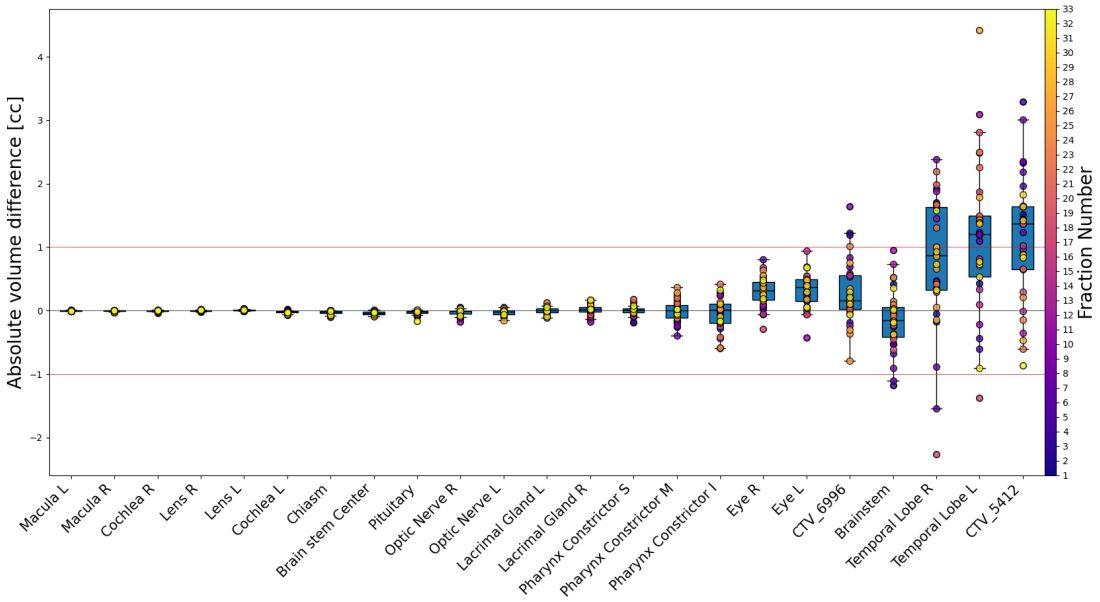

Figure 33: Absolute volume differences for Case 2 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc.

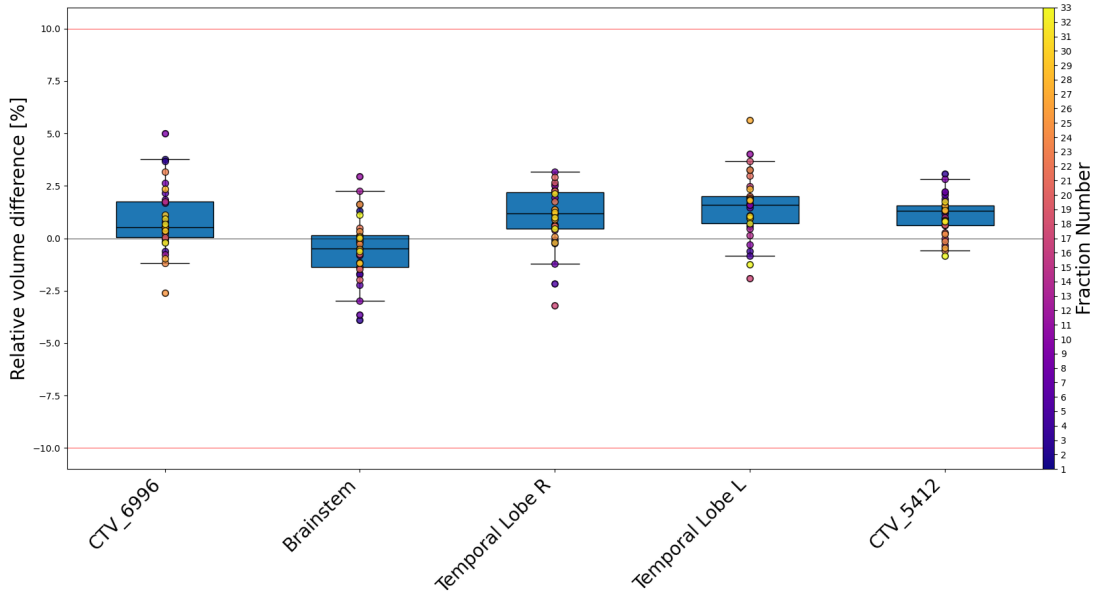

Figure 34: Relative volume differences for structures bigger than 10 cc for Case 2 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%.

As shown in figure 33, structures larger than 10 cc occasionally exceeded the 1 cc threshold for absolute volume differences for some of the daily fractions. For structures larger than 10 cc, the volume difference was subsequently evaluated in percentage units using the 10% threshold, ensuring that all larger structures consistently met the QA requirements across all daily images, as illustrated in figure 34.

Overall, all deformably propagated volumes successfully passed the QA criteria established for structure evaluation. It suggests that the thresholds defined for rigid registration were also appropriate and applicable for Case 2, given the simplicity and reproducibility of this case. For more complex head and neck (HN) scenarios, it would expect larger differences in propagated volumes due to greater anatomical variability.

B.2 Case 4

This fourth case, like Case 3, required two replannings during the treatment course. The first replanned CT was acquired on the first day of treatment, with the updated treatment plan implemented starting from fraction 6. Later, a second replanned CT was considered necessary, which was acquired on fraction 14, and the updated plan was used from fraction 28 until the end of the treatment.

In this case, even smaller structures failed to meet the absolute volume criteria. For instance, the hyoid, which had a volume of 3.3 cc in the planning CT, showed a significant volume discrepancy in the fifth fraction, where the absolute volume difference was 1.4 cc, which corresponds to an approximate relative difference of 30.3%.

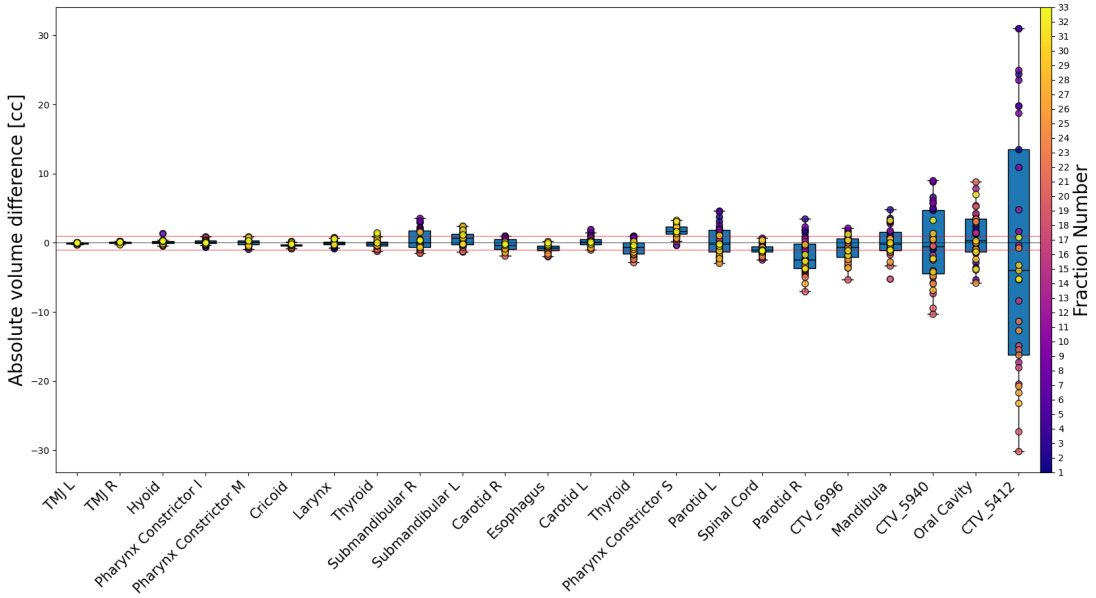

Figure 35: Absolute volume differences for Case 4 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc.

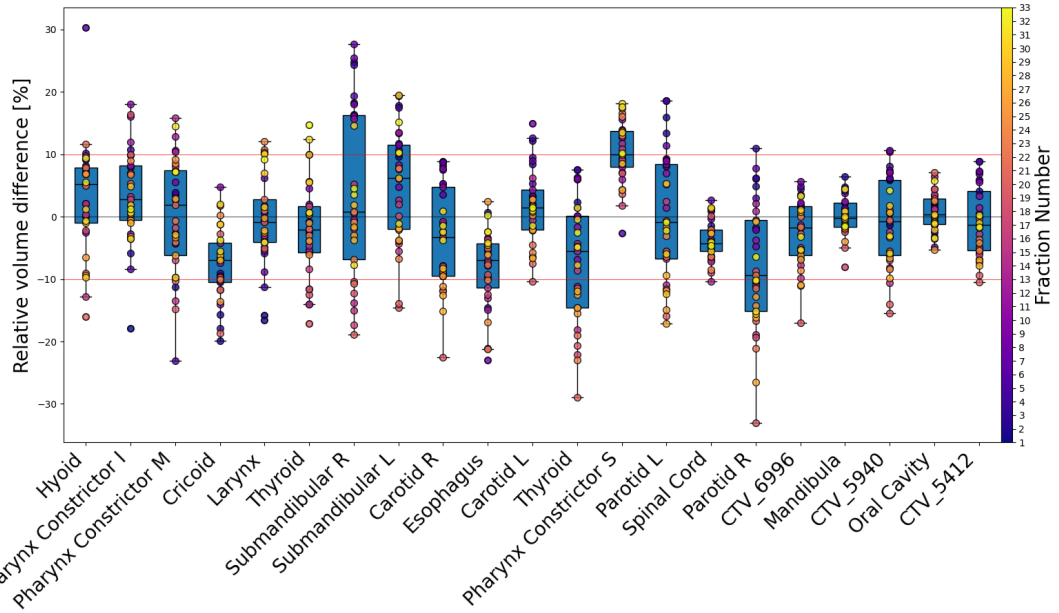

Figure 36: Relative volume differences for all structures for Case 4 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%.

In this case, for the 25 structures, the QA fails for 16 in at least one fraction, and for 6 structures it fails in more then 10 fractions. This case showed the biggest variation between the propagated structures. Based on these findings, it would be necessary to apply a more soft QA criteria of ±1 cc and ±34% for this case to ensure that all structures pass the QA evaluation across the daily fractions.

B.3 Case 5

For this final case, the absolute difference QA fail for structures larger than 8 cc, as shown in figure 37. Even when applying the relative criteria, as illustrated in figure 38, the structure volumes still fall outside the acceptable thresholds defined for structure QA.

Interestingly, the fractions that fail in both criteria appear to be random and are not associated with any specific period of the treatment, unlike the pattern observed in replanned cases.

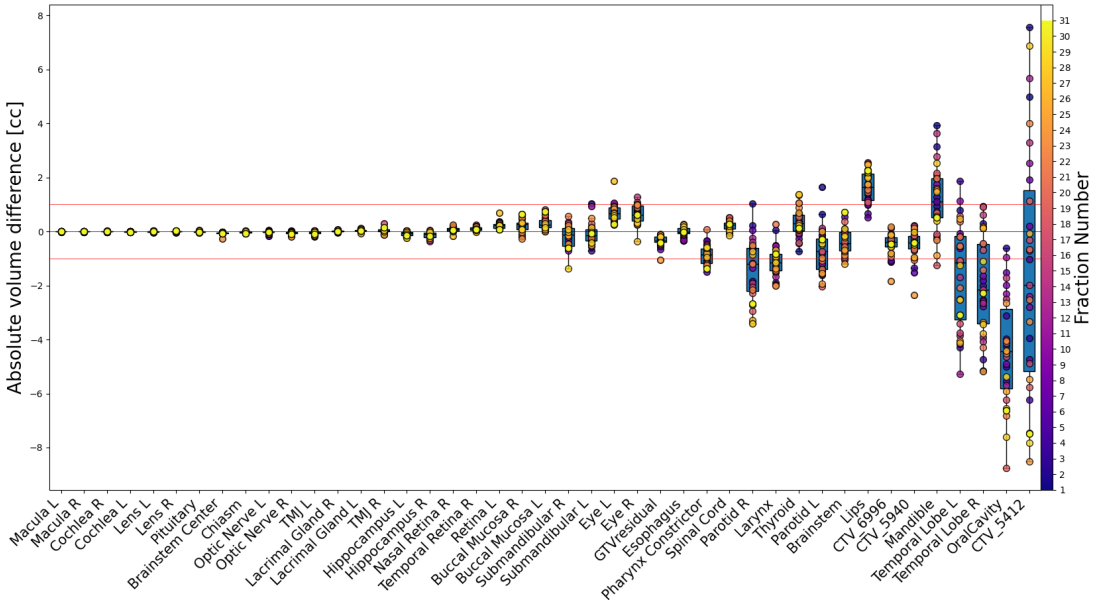

Figure 37: Absolute volume differences for Case 5 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 1 cc.

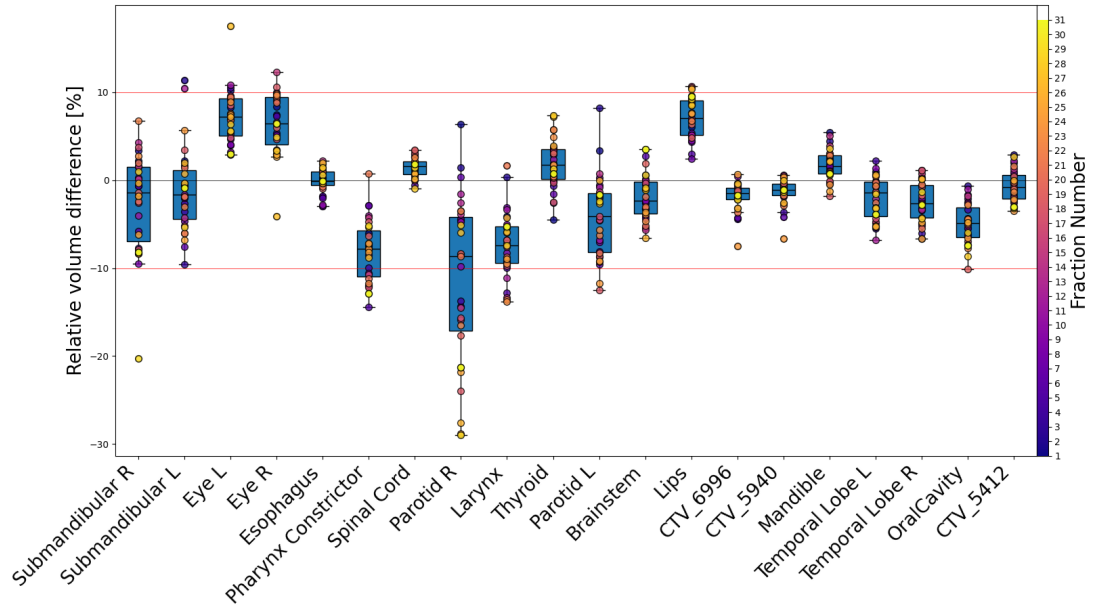

Figure 38: Relative volume differences for structures bigger than 8 cc for Case 5 and all simulated daily fractions, displayed from the smallest (left) to the largest (right) structure in the structure set. The red line indicates the QA threshold of 10%.

Even though this plan did not undergo replanning during the treatment, the propagated structures showed significant variation across the daily images. Out of the 44 structures analyzed, 11 failed the structure QA in at least one fraction, with 2 of them failing in more than 10 fractions. To ensure that all structures pass the QA evaluation across the daily fractions, a more soft QA criteria of ± 1 cc and $\pm 28\%$ would need to be applied for this case.

C Comparison between different configurations - all cases

C.1 Comparison between adaptive in G2 (full RO) vs non adaptive in G2 and non adaptive in G3

C.1.1 Case 1

The graphs below show boxplots representing the results obtained for all three treatment scenarios, adaptive at G2, conventional at G2 and conventional at G3. Figures 39 and 40 show daily doses for target and OAR.

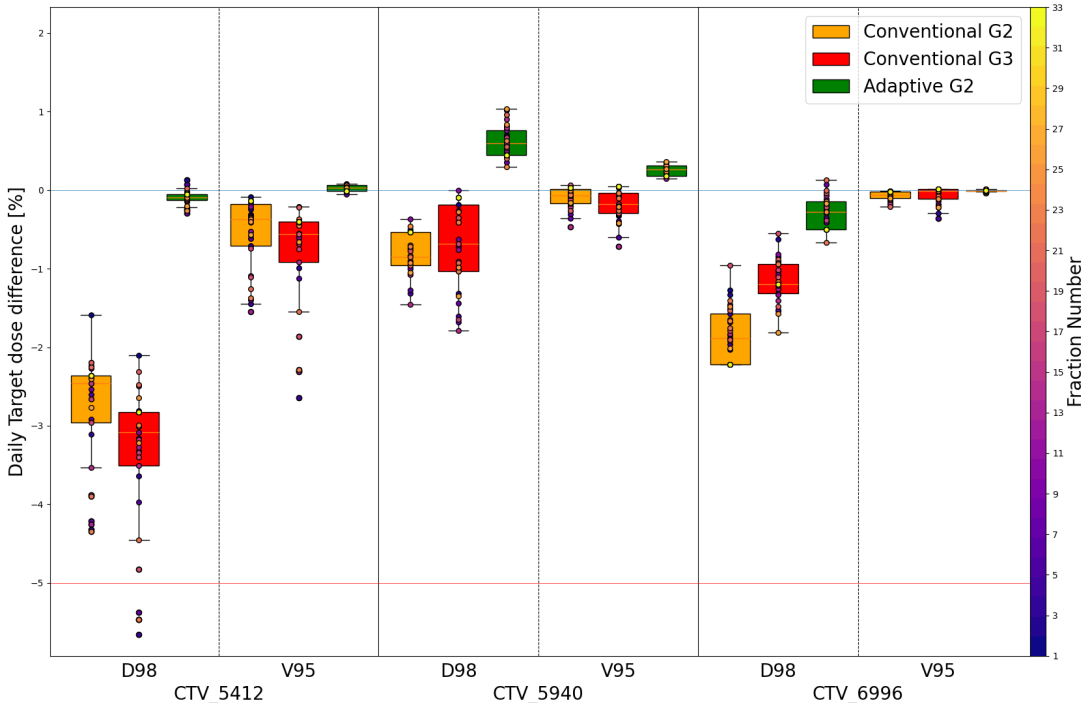

Figure 39: Case 1: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels

For this case, the target coverage with the adaptive treatment in G2 was improved compared to both conventional treatments. As shown in figure 39, the V95% metric for the high-dose target was no much higher with the adaptive approach, in fact, when applying the Wilcoxon test, no significant difference was observed when comparing with the conventional treatment in G3, while on conventional in G2, the coverage was lower than that achieved with the adaptive results.

Regarding the OARs, the adaptive treatment also demonstrated a reduction in the simulated delivered dose. This is particularly noteworthy for the right cochlea, which is located very close to the mid-dose target. The simulation of the adaptive treatment successfully deliver the planned dose, whereas the conventional approaches showed a more divergent distribution, with the simulated delivered dose occasionally exceeding the planned and approved dose by more than +10% (15.2 cGy RBE/fraction).

In comparing the two conventional approaches, the results highlighted differences between the two conventional treatment simulations, favoring the treatment in G3. For target metrics,

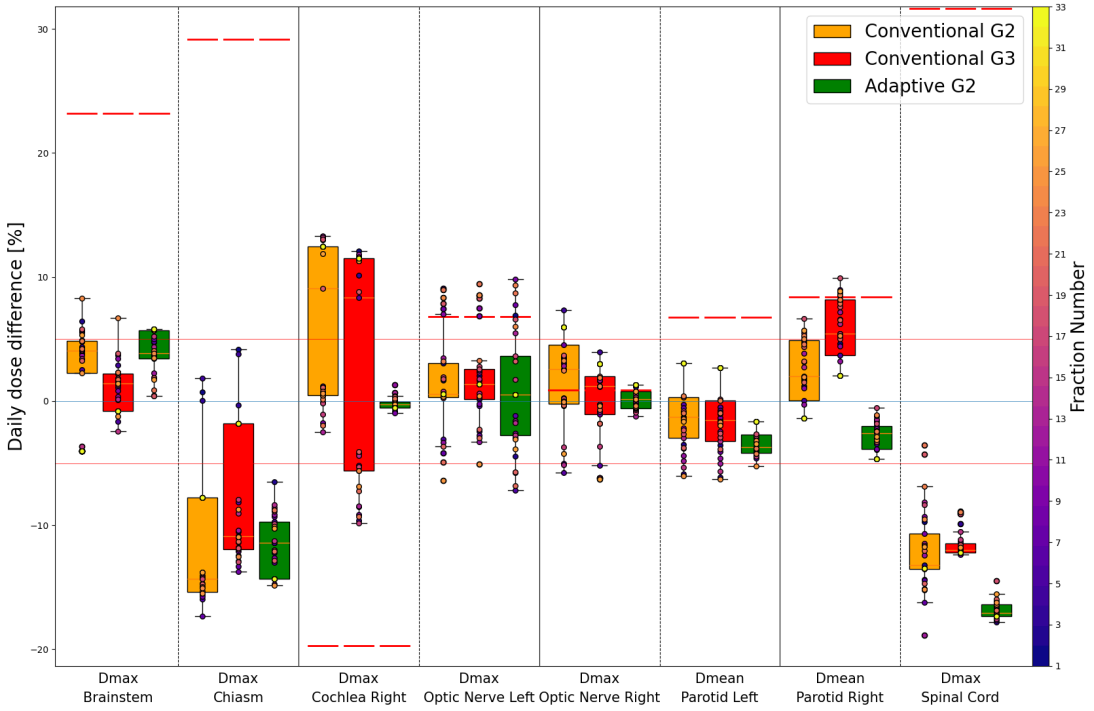

Figure 40: Case 1: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

20.0% of the metrics showed no significant difference, while 60.0% indicated better coverage with the conventional treatment in G3. The difference was similar for OAR, where 11.4% of the metrics showed no significant difference, and 60.0% demonstrated a lower simulated delivered dose with the treatment in G3. The comparison between the two conventional approach are showed in table 9.

The Wilcoxon test revealed a difference between the two configurations, since the conventional treatment in G3 provided better results compared to the conventional approach in G2.

In the second configuration, the adaptive approach demonstrated superior target coverage compared to the conventional treatment in G2. Only 6.7% of the target metrics showed no significant difference, while 86.7% indicated better coverage with the adaptive approach. Regarding OAR metrics, 14.3% of the metrics showed no significant difference, and 68.6% demonstrated a lower simulated delivered dose with the adaptive treatment.

For the third configuration, although the difference was not as pronounced as in the second configuration, the adaptive treatment still exhibited better dose distributions. Specifically, 46.7% of the target metrics showed no significant difference, while 46.7% indicated better coverage with the adaptive treatment. Regarding OAR metrics, 22.9% showed no significant difference, and 51.4% demonstrated a lower dose with the adaptive treatment.

C.1.2 Case 3

When comparing the two conventional treatment approaches, the conventional treatment simulation performed with G3 exhibited better target coverage, as shown in figure 41. Specifically, 40.0% of the target metrics showed no significant difference between Gantries 2 and 3, while 53.3% indicated better coverage with G3. However, no statistically significant difference in the simulated delivered OAR dose was observed. 24.1% of the metrics showed no significant differ-

ence, while the remaining metrics were evenly distributed between the conventional treatments on both Gantries.

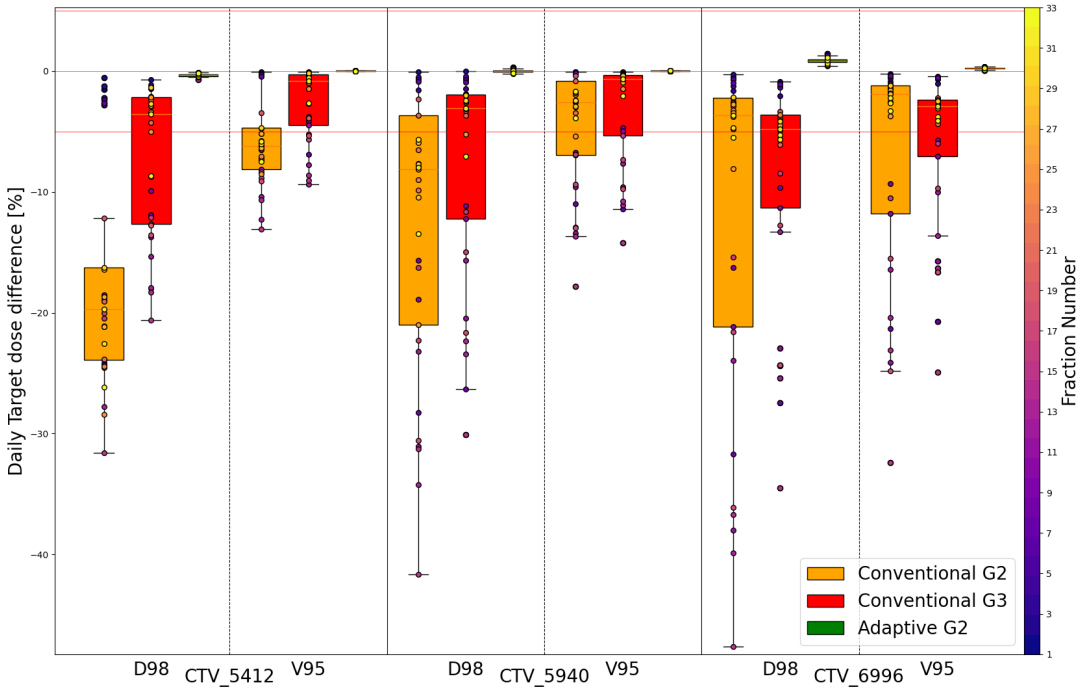

Figure 41: Case 3: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels

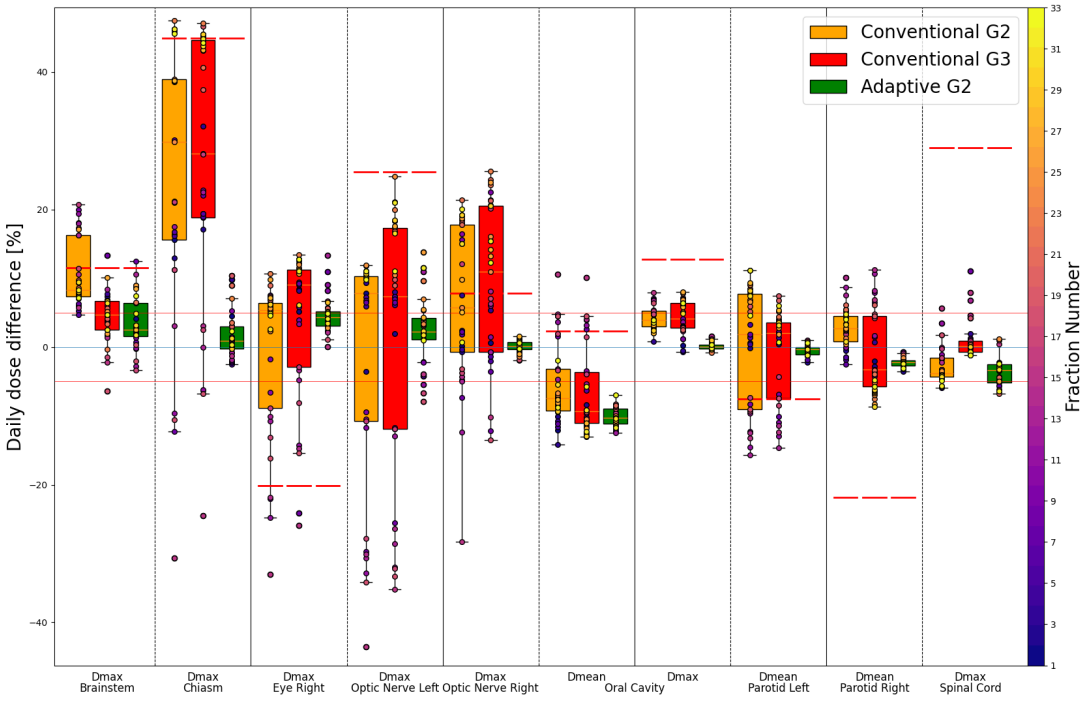

Figure 42: Case 3: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

The daily doses on OAR illustrated in figure 42 highlights a more precise and less dispersed dose distribution with the adaptive treatment. In contrast, conventional treatments generally had a more scattered and higher dose, occasionally exceeding the defined cllinical goals.

The Wilcoxon test showed that both configurations had similar results. All target metrics for both configurations demonstrated better target coverage for the adaptive approach. For the OAR metrics, an average of 94.8% of the metrics exhibited a lower or equal simulated delivered dose with the adaptive approach when compared to conventional treatment in both Gantries.

C.1.3 Case 5

For the fifth case, the two configurations showed some differences in the results.

The conventional treatment in G2 provided better target coverage and lower OAR doses compared to the conventional treatment in G3. This is showed in figure 43, where the D98% values for the mid- and high-dose targets are significantly higher with conventional treatment in G2, even when compared to the adaptive approach. Similarly, the OAR doses were sometimes lower with the conventional approach in G2, as illustrated in figure 44.

In fact, the Wilcoxon test revealed that only 23.1% of the target metrics exhibited no significant difference between the two conventional approaches, while 61.5% indicated better target coverage with the conventional treatment on G2. Regarding the OAR metrics, 15.4% showed no significant difference between the two conventional treatments, whereas 51.3% demonstrated a lower simulated delivered dose with G2.

Even with the better performance in conventional treatment in G2, in some cases, this approach did not outperform the adaptive approach, which showed a better target coverage and lower OAR simulated delivered dose.

The Wilcoxon test revealed some differences between the two configurations, especially in the target metrics, similar to Case 1.

For the conventional treatment in G2, the adaptive approach showed modest improvements. While 6.7% of target metrics exhibited no significant difference, 40.0% indicated better coverage with the adaptive approach. Regarding OAR metrics, the adaptive approach demonstrated superior performance, with 12.8% showing no significant difference and 64.1% indicating a lower simulated delivered dose compared to the conventional approach.

In the third configuration, the adaptive approach showed more notable improvements over the conventional treatment in G3. For target metrics, 22.2% showed no significant difference, while the remaining metrics favored the adaptive approach for better target coverage. The results for OAR metrics in this configuration were consistent with those of the second configuration, with 7.7% showing no significant difference and 64.1% indicating a lower simulated delivered dose with the adaptive treatment.

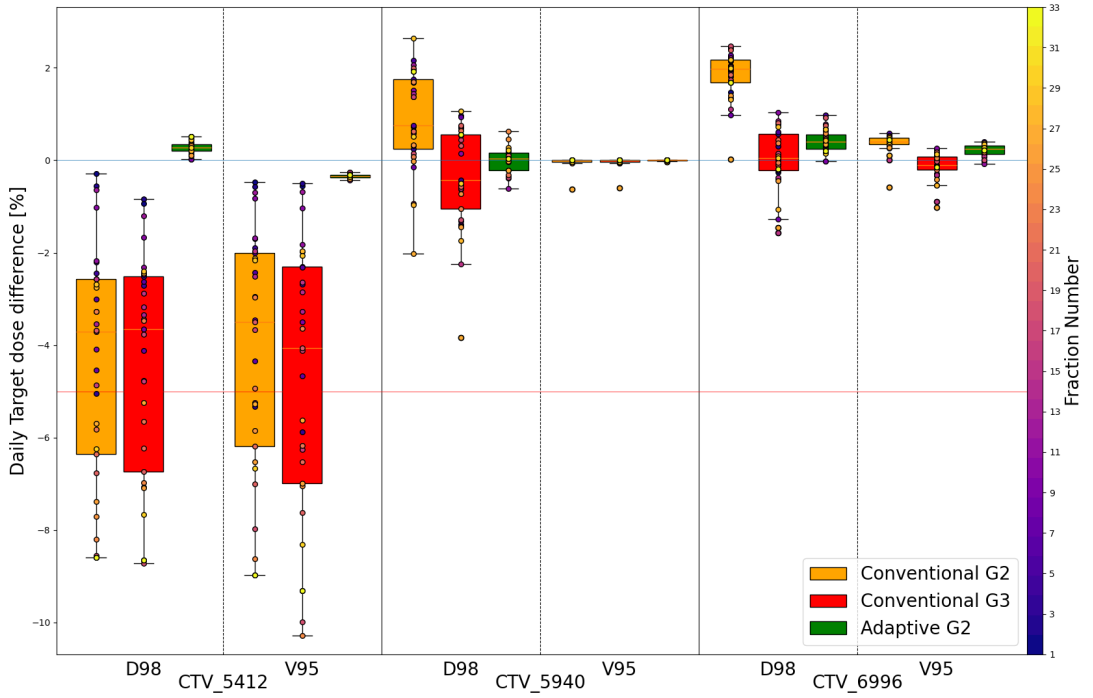

Figure 43: Case 5: Daily target dose differences (D98% and V95%) between adaptive and two conventional treatment approaches for all three target dose levels

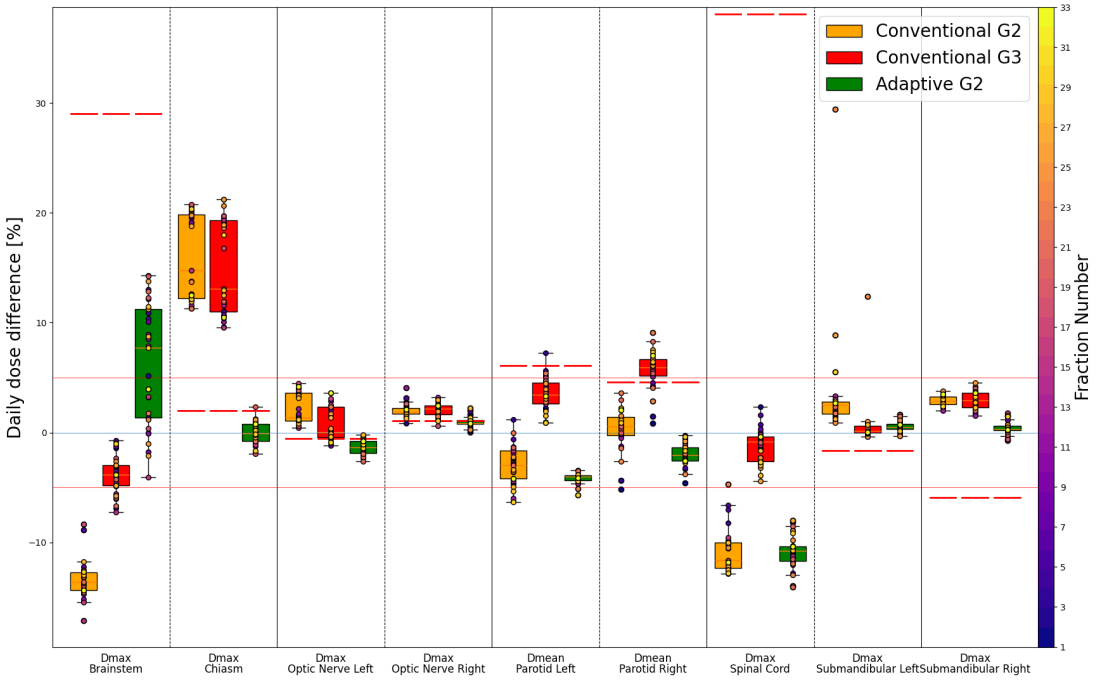

Figure 44: Case 5: Daily dose differences for some selected organs between adaptive and two conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

C.2 Comparison between adaptive G2 vs adaptive in G3 vs non adaptive in G3

C.2.1 Case 1

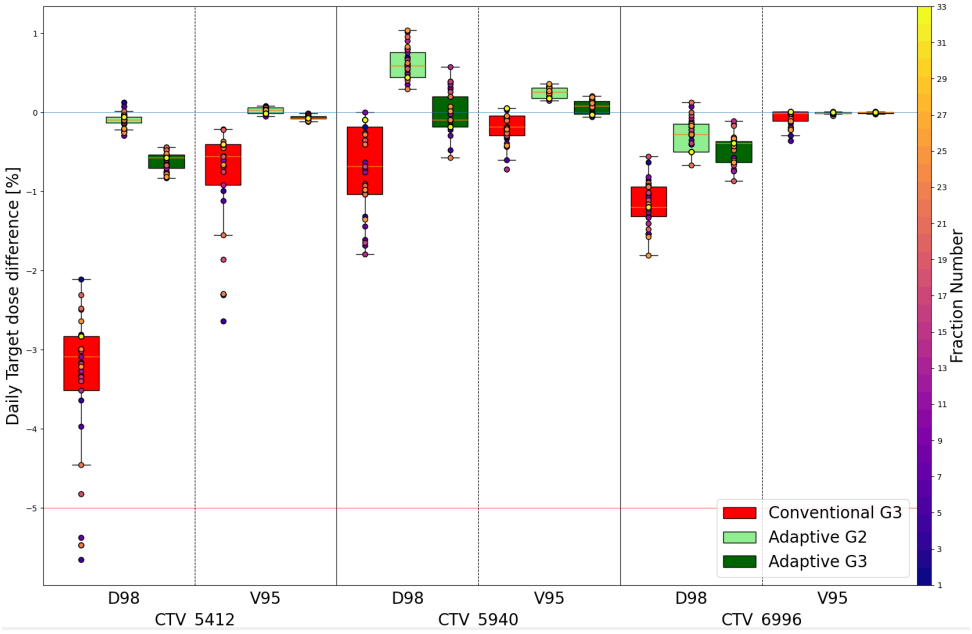

Figure 45: Case 1: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels.

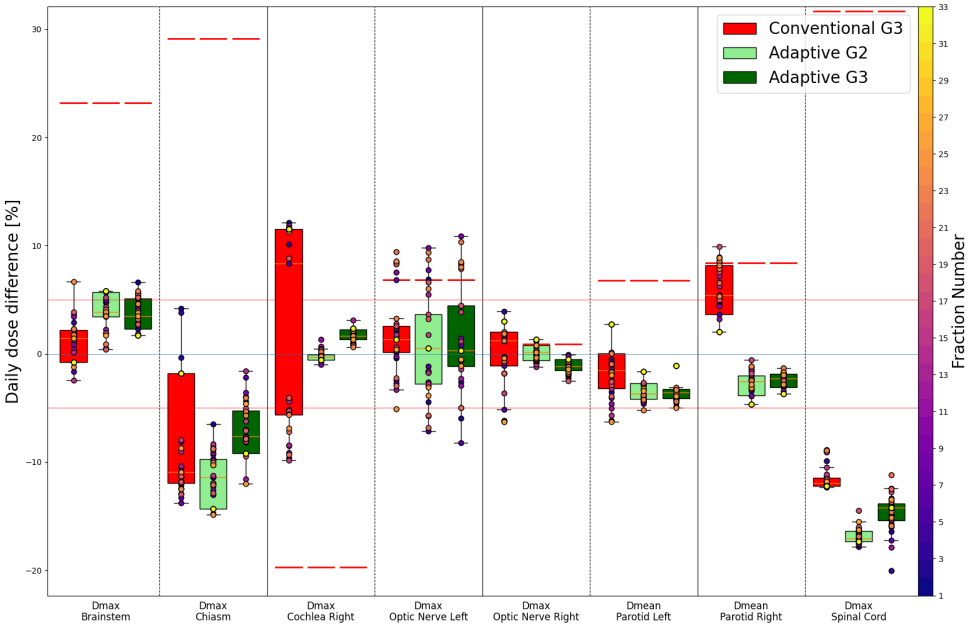

Figure 46: Case 1: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

As showed in the figures 45 and 46, the adaptive treatment in G3 also shows a better coverage and OAR sparing in relation to conventional standard-of-care treatment.

Regarding target coverage, 66.7% of the metrics had a better result with the adaptive treatment. For OAR, 22.9% of the metrics had no significant difference between the two approaches, adaptative in G3 and standard-of-care, and in 60% the dose could be lowered with adaptive treatment.

When comparing the two adaptive approaches, the results favored adaptation in G3. For the target, 26.7% of metrics showed no significant difference between the two approaches, while 40.0% indicated improved coverage with adaptive treatment in G3. Notably, six metrics (D98% and V95%) presented in figure 45 were the only ones where adaptation in G2 performed better. For the OARs, 31.4% of metrics showed no significant difference between the approaches, while 42.9% exhibited a lower dose with adaptive treatment in G3.

C.2.2 Case 3

For Case 3, both adaptive approaches improve simulated dose distribution compared to the current standard of care. As this replanned case experienced low coverage at the beginning of the treatment, adaptation in G3 effectively accounted for daily variations and had a better-optimized daily dose, similar to the performance of adaptive treatment in G2.

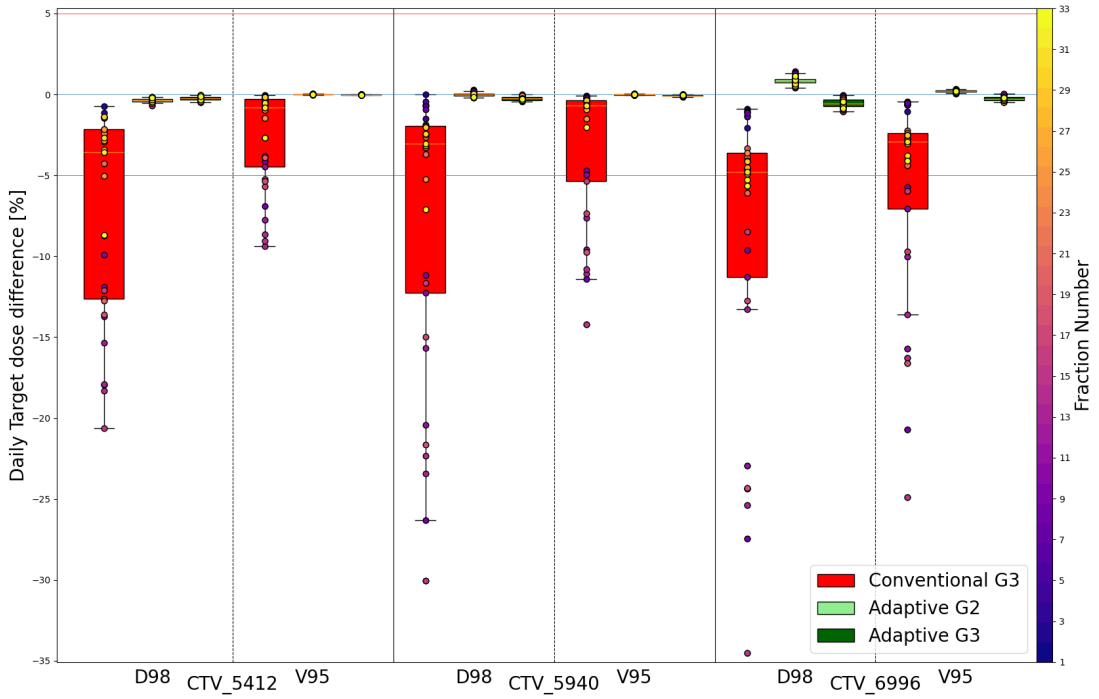

Figure 47: Case 3: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels.

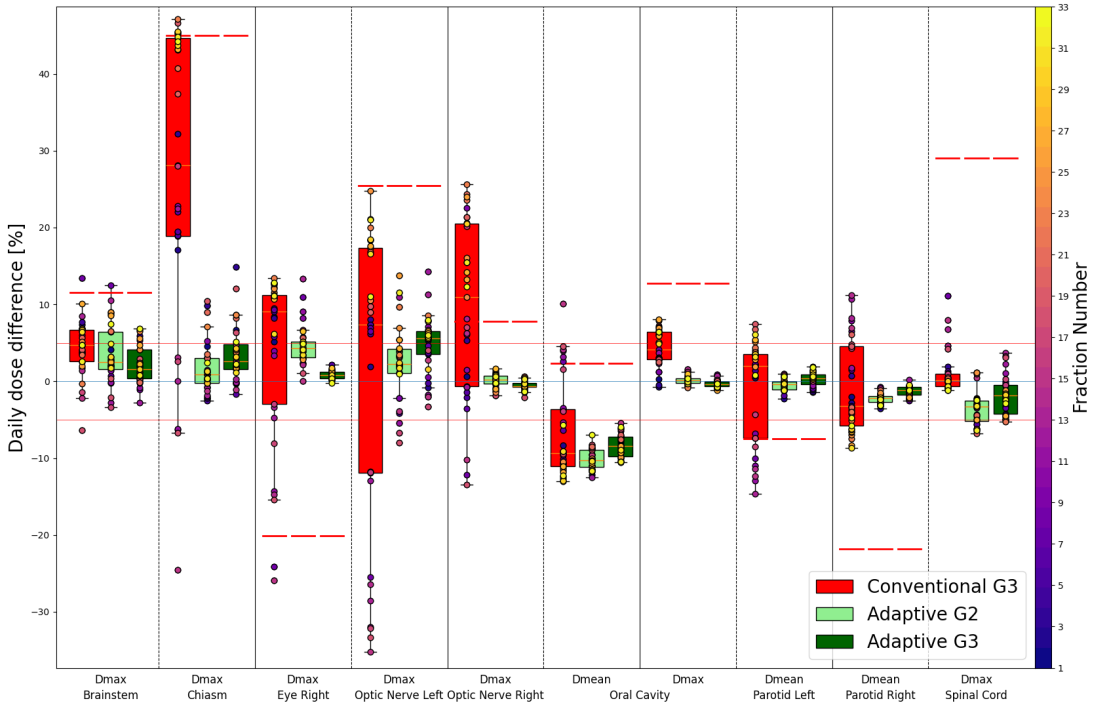

Figure 48: Case 3: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

The Wilcoxon test demonstrated better dose distribution for the adaptive treatment, with 20.0% of target metrics showing no significant difference and the remaining metrics indicating improved coverage for the adaptive approach. For OAR metrics, 41.4% showed no significant difference, while 48.3% exhibited a lower dose with adaptive treatment. In particular, the conventional approach resulted in higher doses to the chiasm and optic nerves, which were located close to the target in this case, as shown in figure 48.

When comparing the two adaptive approaches, as shown in figures 47 and 48, adaptation in G2 provided better target coverage and OAR sparing. Namely, only 13.3% of target metrics showed no significant difference, while 80.0% demonstrated improved coverage with adaptive G2. For OAR metrics, 13.8% showed no significant difference, and 58.6% indicated a lower dose with adaptive G2. In figure 72 in appendix D, a better coverage of the high-dose target can be observed in the adaptive treatment of configuration 3 (G2) compared to the one of configuration 4 (G3).

C.2.3 Case 5

For Case 5, the adaptation in G3 also showed to have superior dose distribution in relation to conventional standard-of-care, as showed in figures 49 and 50.

In figure 49, a lower coverage in the mid- and high-dose target regions can be seen for the adaptive approach in G3. The Wilcoxon test revealed that 40.0% of the target metrics showed no significant difference, while 46.7% exhibited a lower dose with adaptive treatment in G3. Even though the conventional in G3 had a lower coverage, as discussed in section 8.1.

In terms of OAR metrics, similar findings were observed compared to the other configurations. Specifically, 15.4% of the metrics showed no significant difference, and 59.0% indicated a lower dose with the adaptive treatment.

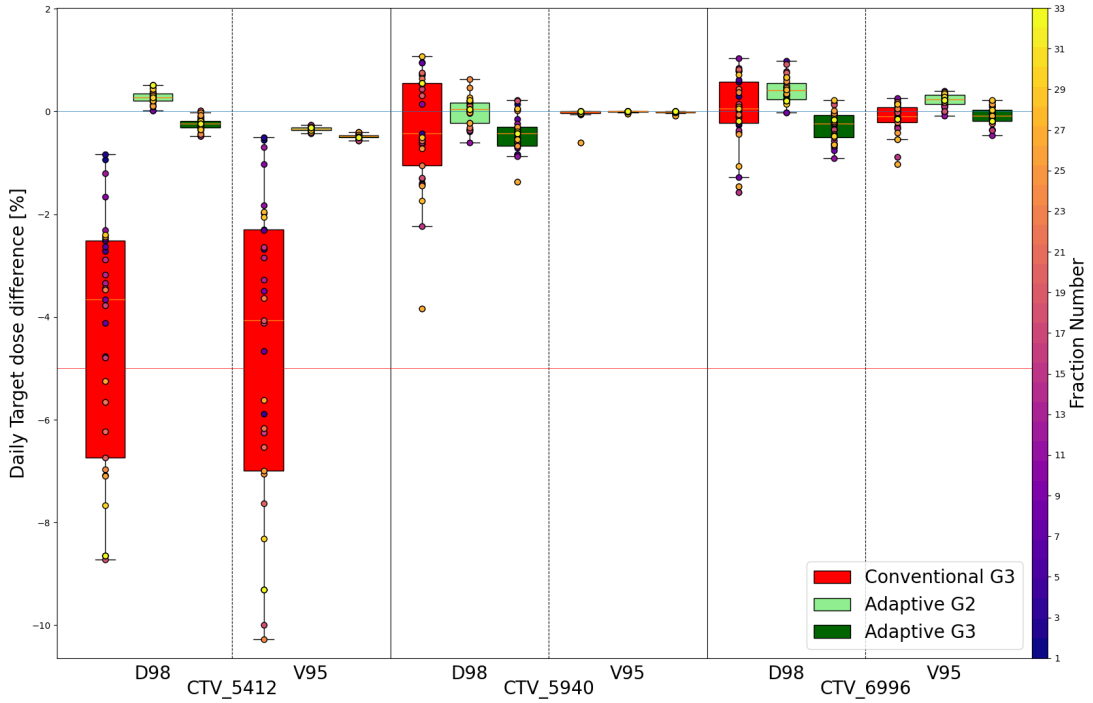

Figure 49: Case 5: Daily target dose differences (D98% and V95%) between two adaptive and conventional treatment approaches for all three target dose levels.

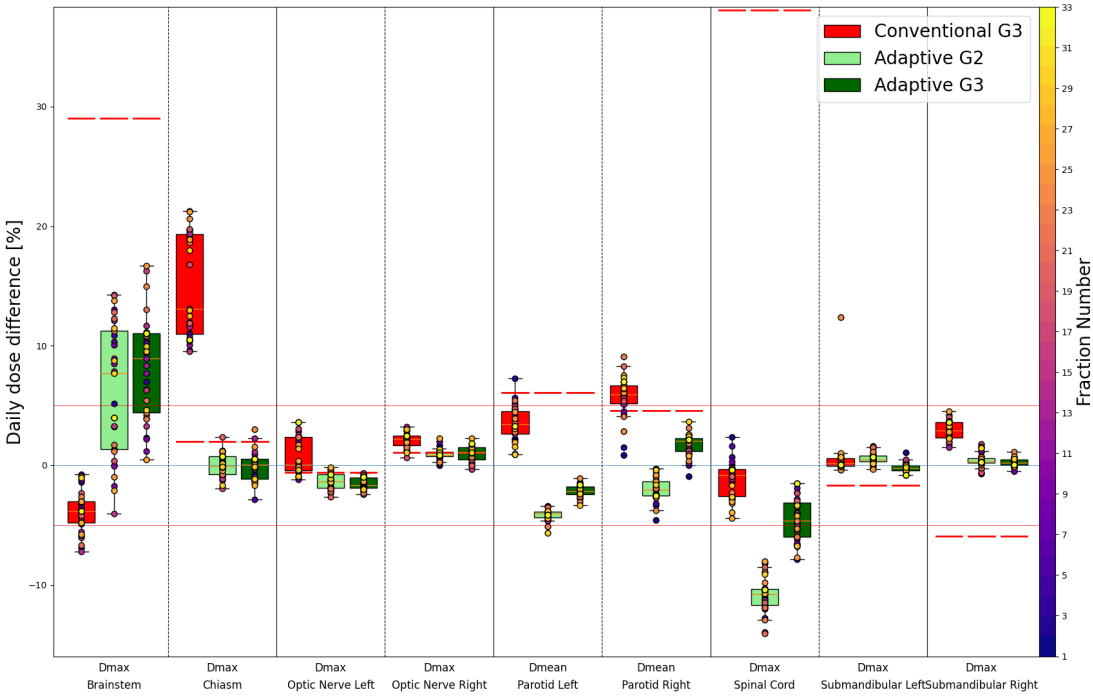

Figure 50: Case 5: Daily dose differences for some selected organs between two adaptive and conventional treatment approaches. The red line represents the dose constraint defined in the clinical goal.

Interestingly, when comparing the two adaptive approaches, adaptation in G2 once again resulted in a better dose distribution. Only 6.7% of the target metrics showed no significant

difference, while 73.3% demonstrated an improvement in target coverage with adaptive treatment in G2. For OAR metrics, 25.6% showed no significant difference, while 61.5% had a lower dose for adaptive treatment in G2.

C.3 Comparison between FIonA and RayStation

C.3.1 Case 1

For the first case, most of the metrics demonstrated relatively small dose variations when compared to the planned dose, for both conventional and adaptive treatment scenarios, for both TPS. The results of RayStation® had a few difference in relation to the findings in FIonA.

As presented in figures 51 and 52, RayStation® achieved a better target coverage for all levels where in FIonA, there was a slight reduction of approximately 1% (around 5.40 Gy RBE) in mean dose for the D98% and V95% metrics for the high-dose target in the adaptive treatment, figure 51, which is reflected in figure 66 of appendix D is possible to see the low shoulder in the daily DVH of FIonA.

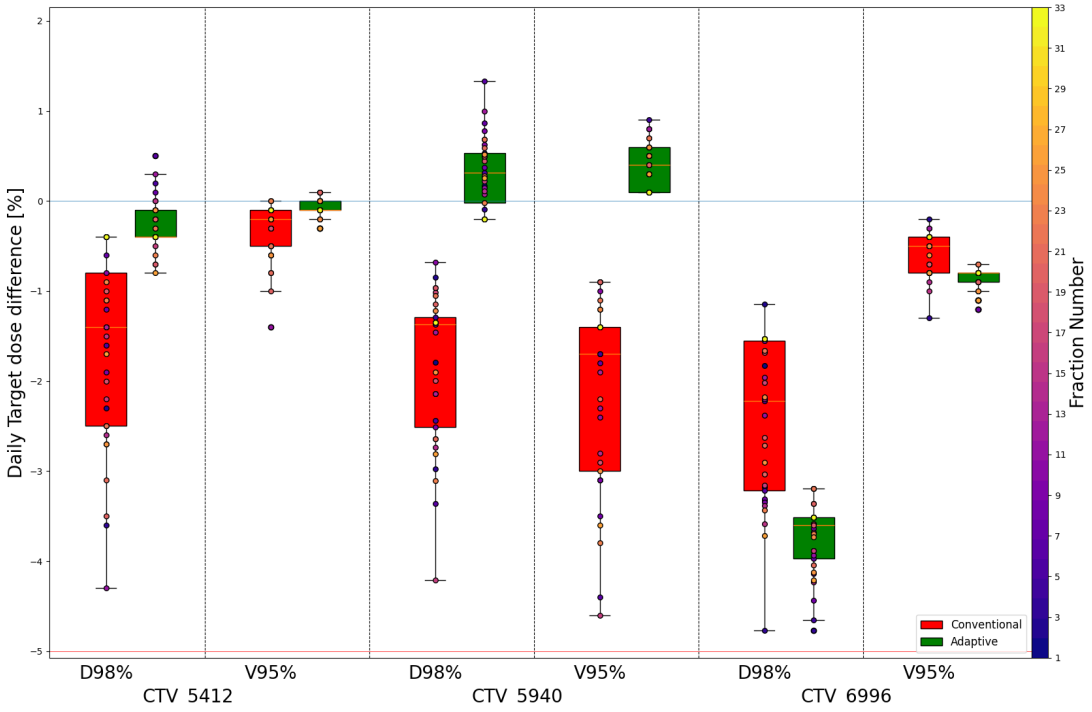

Figure 51: FIonA daily target dose differences, D98% and V95%, for Case 1.

For critical structures such as the brainstem, spinal cord and chiasm, the doses were consistently lower in the adaptive treatment in FIonA, as shown in the graph in figure 53. However, in RayStation® (figure 54), the doses to the brainstem and chiasm were higher with the adaptive treatment. An overdose of approximately 5% for the brainstem and 10% for the chiasm was observed. This could suggest that the larger setup robustness in the conventional plan was more effective in sparing organs far from the target. Nonetheless, the simulated delivered dose in the adaptive treatment remained below 10% of the dose constraint, indicating that it was still within acceptable limits.

Regarding the right cochlea, in the initial plan the clinical goal was not achieved, with Dmax higher that anticipated. In the adaptive treatment for both TPS, the dose showed limited variation across the daily treatment sessions, indicating a good simulated delivery pattern.

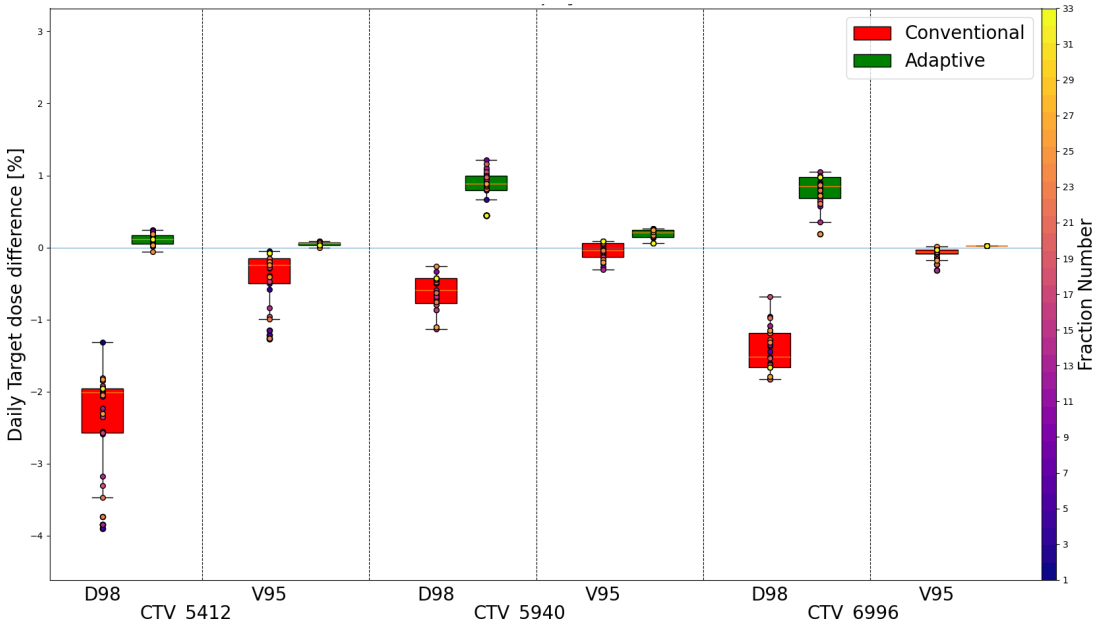

Figure 52: RayStation® daily target dose differences, D98% and V95%, for Case 1.

In RayStation®, the simulated delivered dose in the adaptive treatment was lower than in the conventional treatment. Conversely, in FIonA, the conventional treatment generally had a lower dose to the cochlea, but the dose remained inconsistent and still failed to meet the prescribed constraint. The lower dose in the right cochlea in the conventional treatment is likely due to the recalculation of the treatment plan without considering anatomical changes. Since conventional treatment was based on a fixed plan, density changes in the tissues could have led to this discrepancy, resulting in a loss of target coverage, particularly in the high-gradient region between the middle-dose target and the right cochlea, as seen in figure 51. This, in turn, caused a reduced dose to the right cochlea.

Applying the Wilcoxon test to compare the two approachs revealed differences in the results between the two TPS, but in general, the adaptive plan, irrelevant of the TPS, was always superior than conventional, either by providing a higher dose to the target or a lower dose to OARs.

Regarding the target, the number of metrics where the adaptive plan had a better or comparable simulated delivered dose with adaptive treatment in relation to conventional treatment differed between the two TPS. In RayStation®, 93.3% of the metrics showed better or comparable target coverage in adaptive treatment. In contrast, in FIonA, this number was 60.0%, primarily due to lower coverage in the high-dose target, where conventional treatment could have provided better result. For the OAR metrics the both TPS results were similar, an average of 80.0% showed a lower or comparable simulated delivered dose with adaptive treatment in relation to conventional treatment.

All the results can be seen in the tables 14 and 15 in section 9.3.

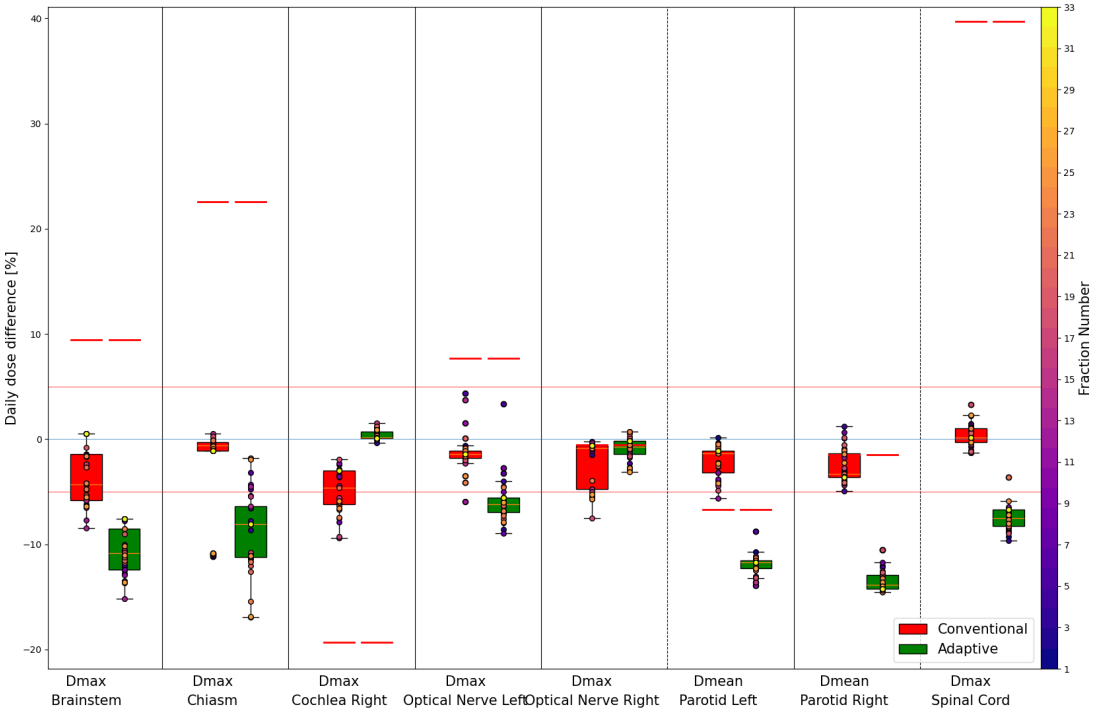

Figure 53: FIonA daily dose difference for some selected organs of Case 1, the red line represents the dose constraint defined by MD.

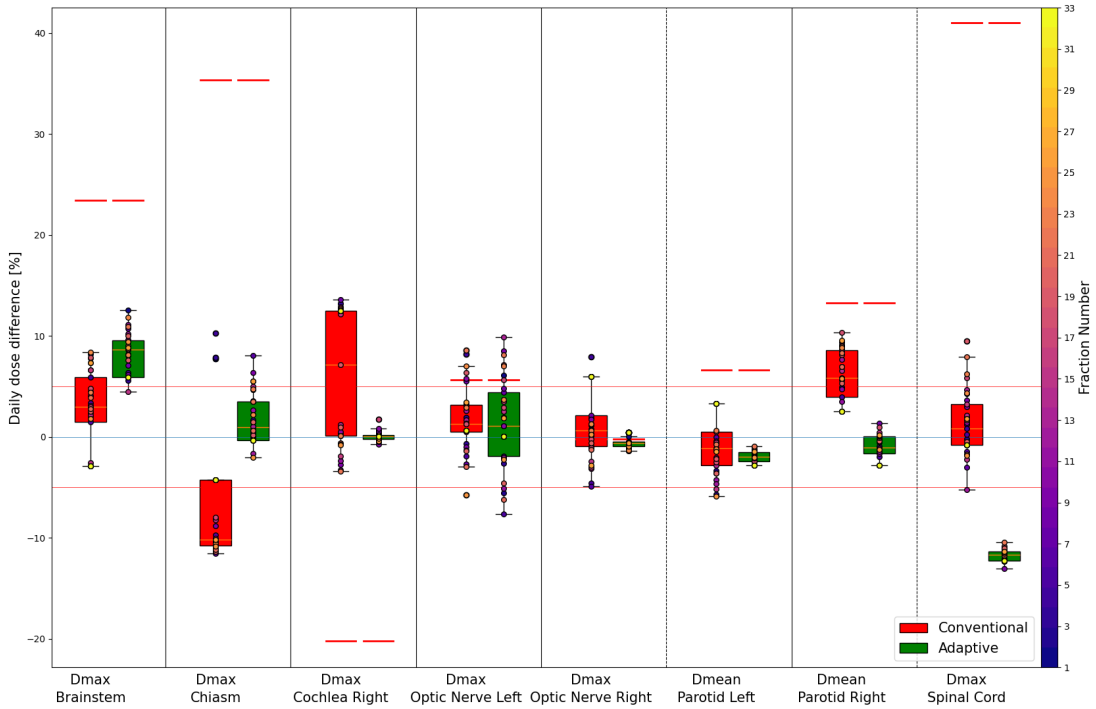

Figure 54: RayStation® daily dose difference for some selected organs of Case 1, the red line represents the dose constraint prescription defined in the clinical goal.

C.3.2 Case 3

Since the third case had been replanned twice during the course of treatment, the simulated conventional dose takes into account the two replanned dose distributions and the daily fractions where the treatment would be delivered. The simulated adapted dose was, on other hand, simulated with the same first plan made in the planning CT.

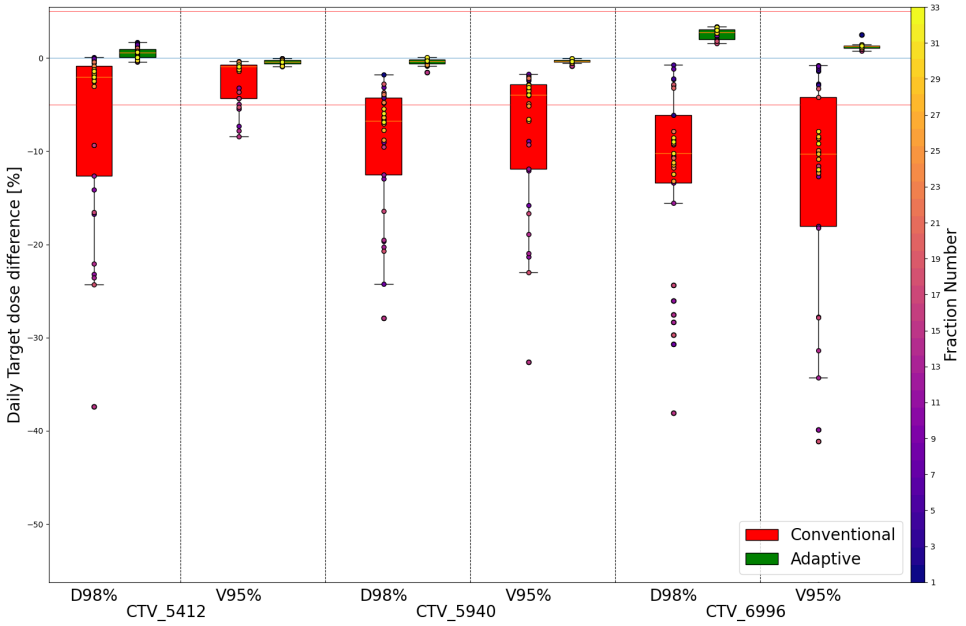

Figure 55: FIonA daily target dose differences, D98% and V95%, for Case 3.

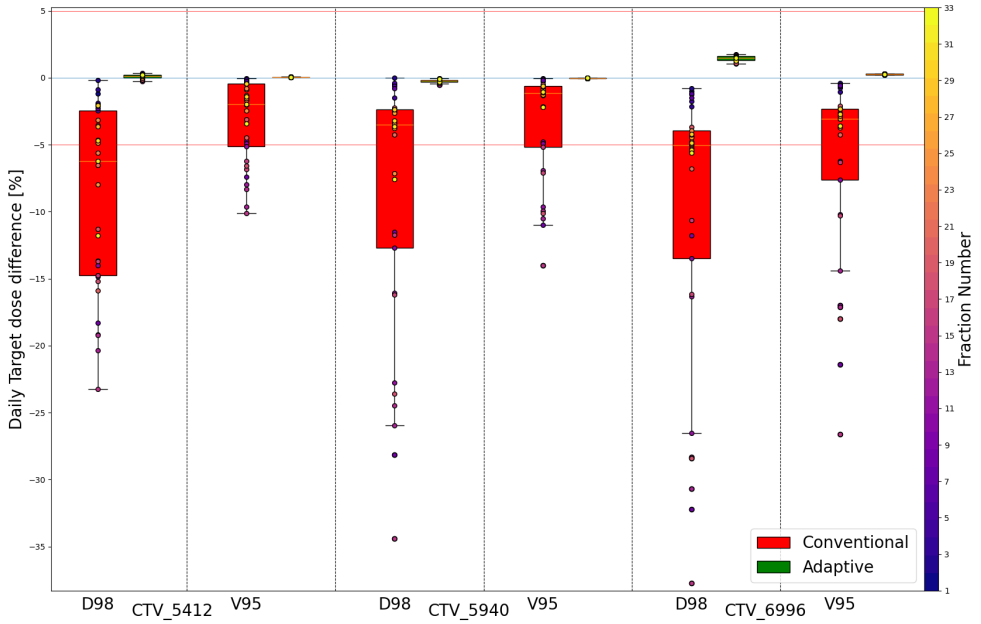

Figure 56: RayStation® daily target dose differences, D98% and V95%, for Case 3.

In the graph shown in figure 55 and 56, a significant improvement in target coverage can be observed when using the adaptive approach for the two TPS. Notably, the lower coverage is

evident during the initial treatment fractions, highlighting the need for replanning early in the treatment, as identified in fraction 5. In appendix D, in figures 70, 71 and 72, it is possible to notice lower coverage in some daily fractions and for all target levels, suggesting a suboptimal target coverage during the period between the acquisition of the repCT and the application of the first replanned fraction.

For the OARs, the adaptive approach generally resulted in lower doses compared to the conventional approach. In the figures 57 and 58, the conventional treatment had higher doses in 6 out of the 10 selected metrics.

It is worth noting that, in this case, the right eye and optic nerve were located near the high-dose target volume. In FIonA, neither the right eye nor the optic nerve showed significant differences between the two treatment approaches. In RayStation®, however, the dose to the right optic nerve showed an improvement with the adaptive treatment.

The Wilcoxon test showed similar results between the two TPS, highlighting the favorable dose distributions of the adaptive approach.

For the target metrics, all metrics demonstrated equal or better coverage with the adaptive treatment in both TPS. For the OAR metrics, both TPS showed that 86.2% of the metrics were either better or comparable with the adaptive approach in comparison to conventional treatment (tables 14 and 15 in section 9.3).

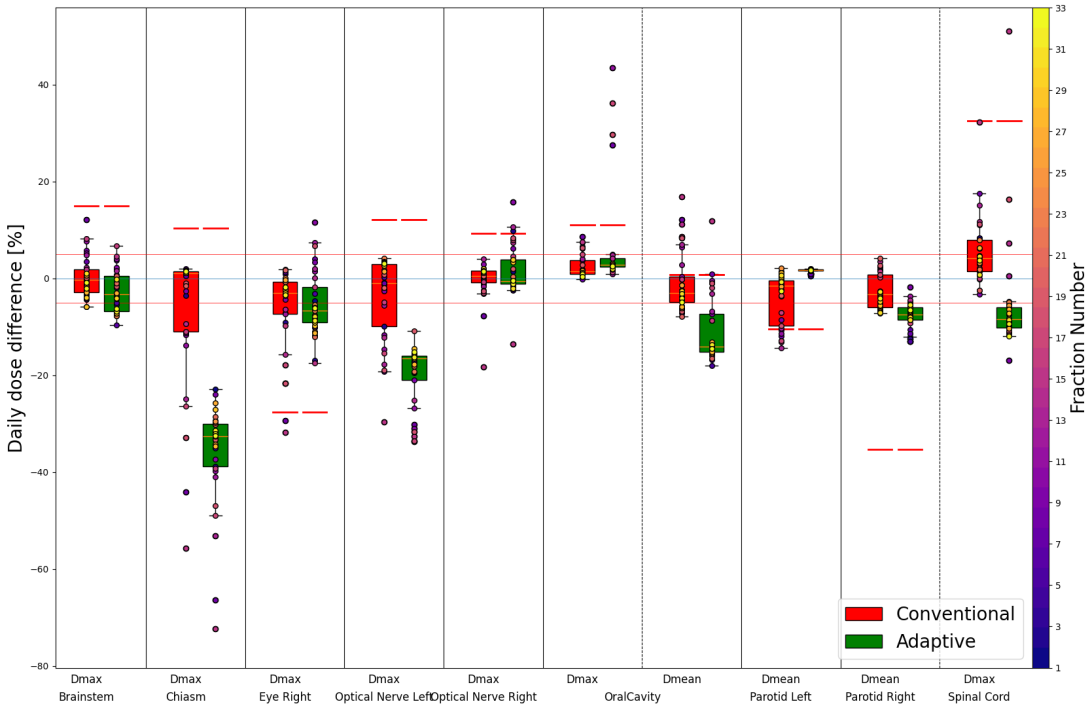

Figure 57: FIonA daily dose differences for some selected organs of Case 3, the red line represents the dose constraint defined in the clinical goal.

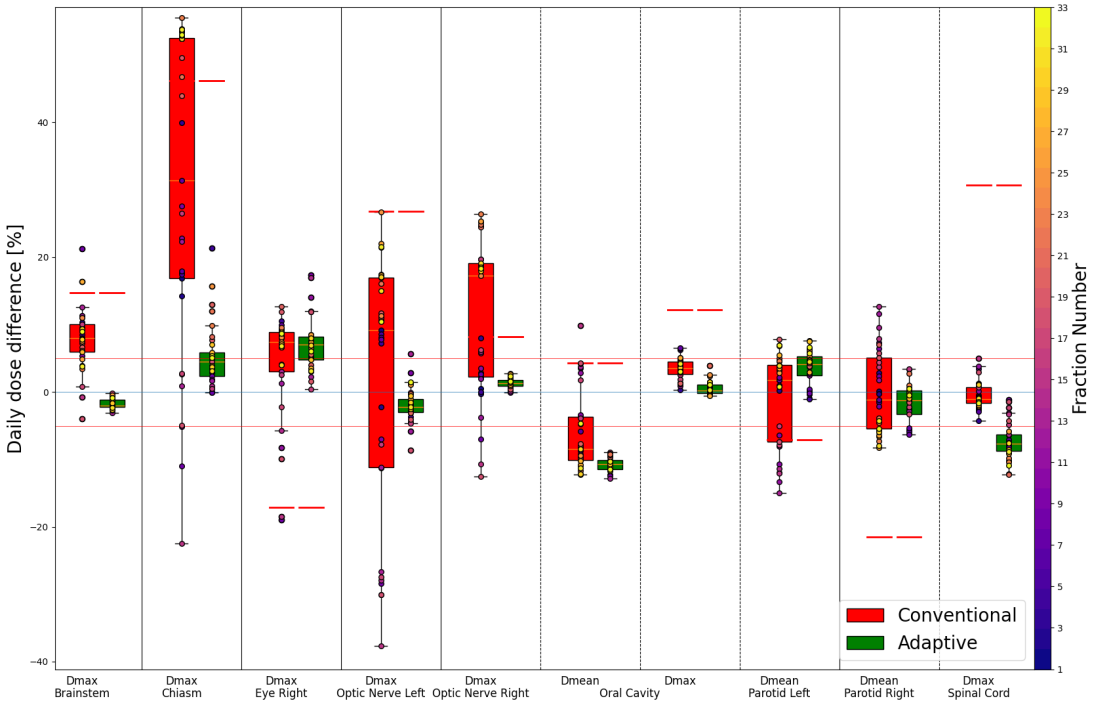

Figure 58: RayStation® daily dose differences for some selected organs of Case 3, the red line represents the dose constraint defined in the clinical goal.

C.3.3 Case 5

For the last case, a clear improvement in the daily target dose coverage is evident when using the adaptive approach, whether in FIonA or RayStation®. In the graph shown in figures 59 and 60, the values of D98% and V95% for the adaptive treatment demonstrate significantly better target coverage for all three target levels. This improvement is particularly notable in the high-dose target in FIonA, where the adaptive approach shows a bigger difference compared to the conventional treatment, as seen in figure 80 of appendix D. For RayStation®, the low-dose target is highly undercovered in the conventional approach.

For the OARs, among the selected metrics in figure 61, a higher dose was observed with the adaptive approach only for the left optic nerve, with differences of less than 2% in the FIonA results.

It is important to highlight that the submandibular glands were entirely within the low-dose target. With the adaptive approach, the simulated delivered dose remained consistent with the planned dose. In contrast, for the conventional treatment, the simulated delivered dose was higher. In RayStation®, as shown in figure 62, a higher simulated delivered dose for this organ can even be observed towards the end of the treatment.

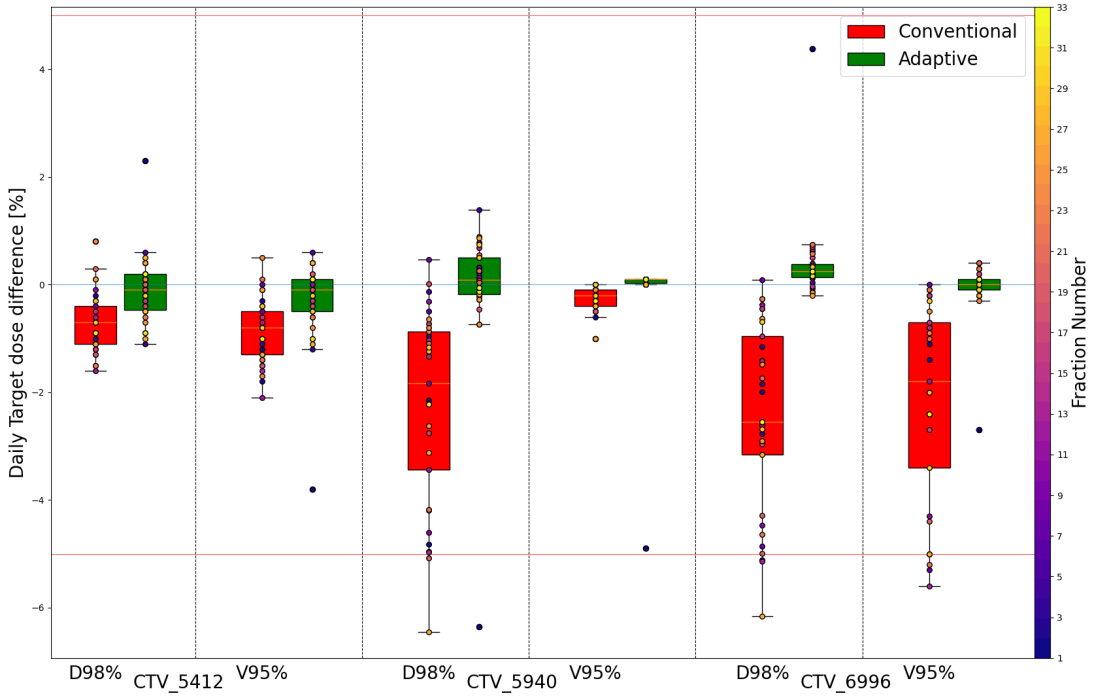

Figure 59: FIonA daily target dose differences, D98% and V95%, for Case 5.

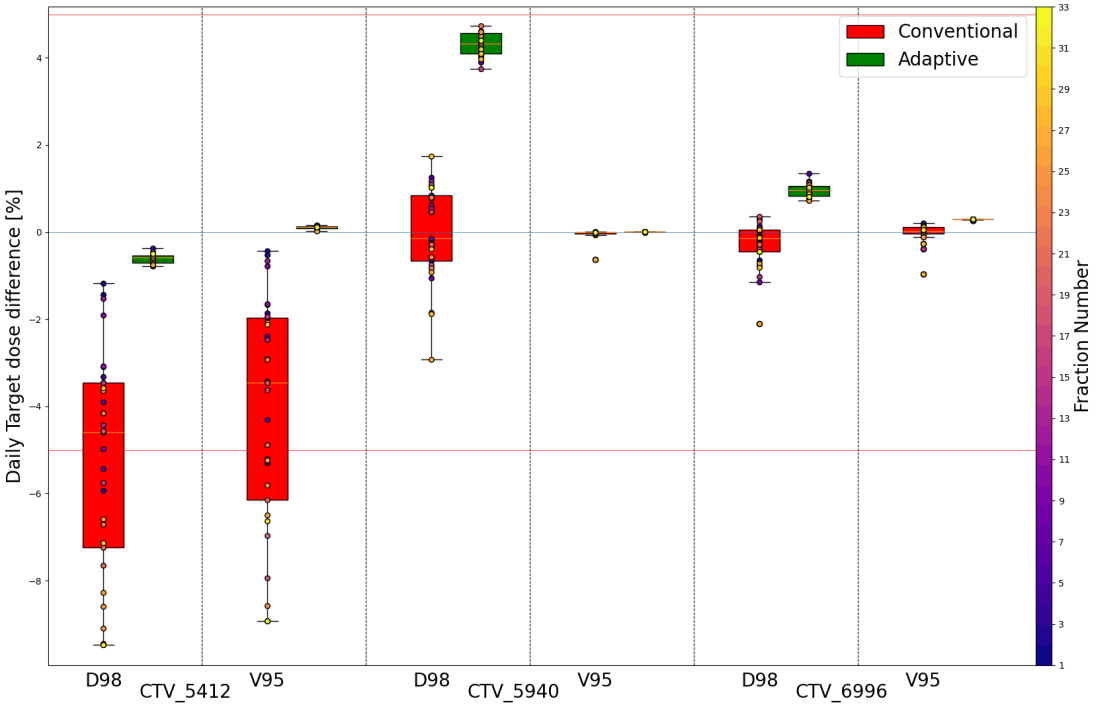

Figure 60: RayStation® daily target dose differences, D98% and V95%, for Case 5.

The Wilcoxon test demonstrated differences regarding the target coverage between the adaptive and conventional treatments for both TPS.

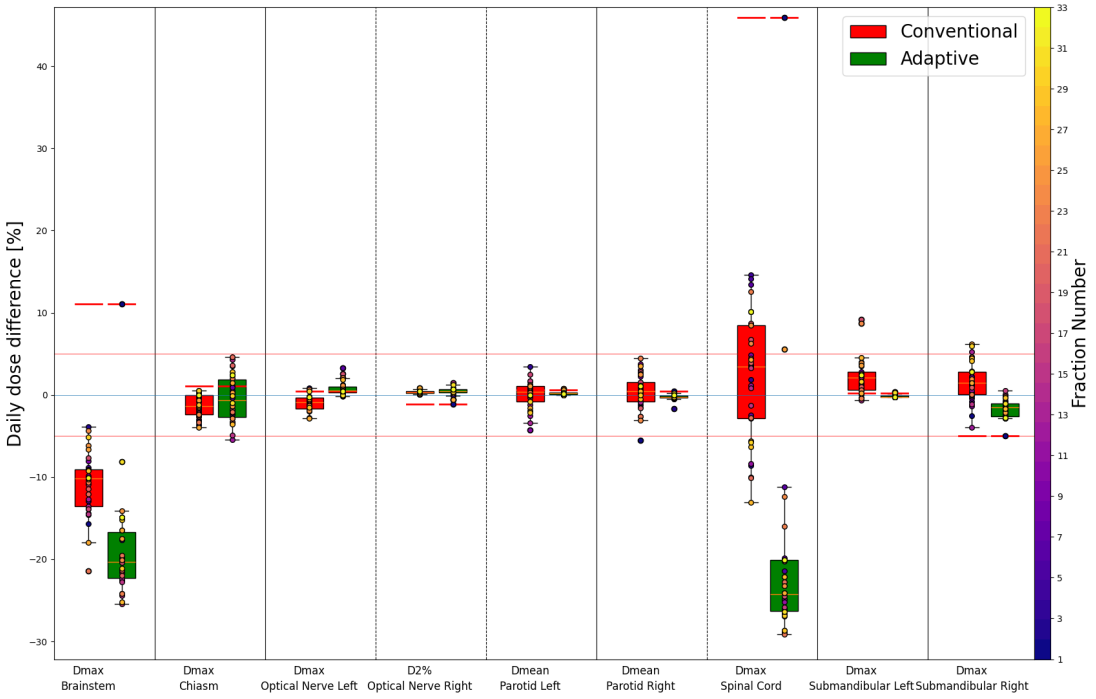

Figure 61: FIonA daily dose differences for some selected organs of Case 5, the red line represents the dose constraint defined in the clinical goal.

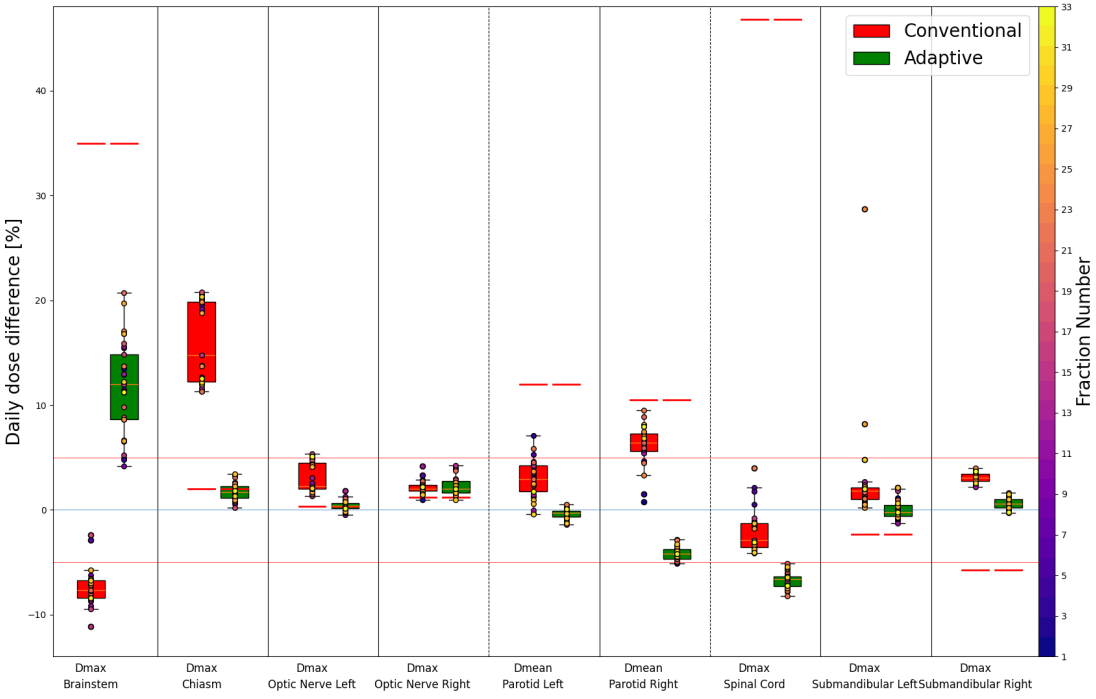

Figure 62: RayStation® daily dose differences for some selected organs of Case 5, the red line represents the dose constraint defined in the clinical goal.

93.3% of target metrics in RayStation® and 73.3% in FIonA showed better or equal coverage with the adaptive treatment. This difference is mainly to the daily maximum dose in FIona that was comparable higher than the conventional approach, as showed in figure 63. If Dmax would

be not considered, the two results would be equal, with 91.7% of coverage metrics showing a better or equal resilt for adaptive treatment. For the OAR metrics, an average of 74.4% showed a lower or equal dose with the adaptive treatment.

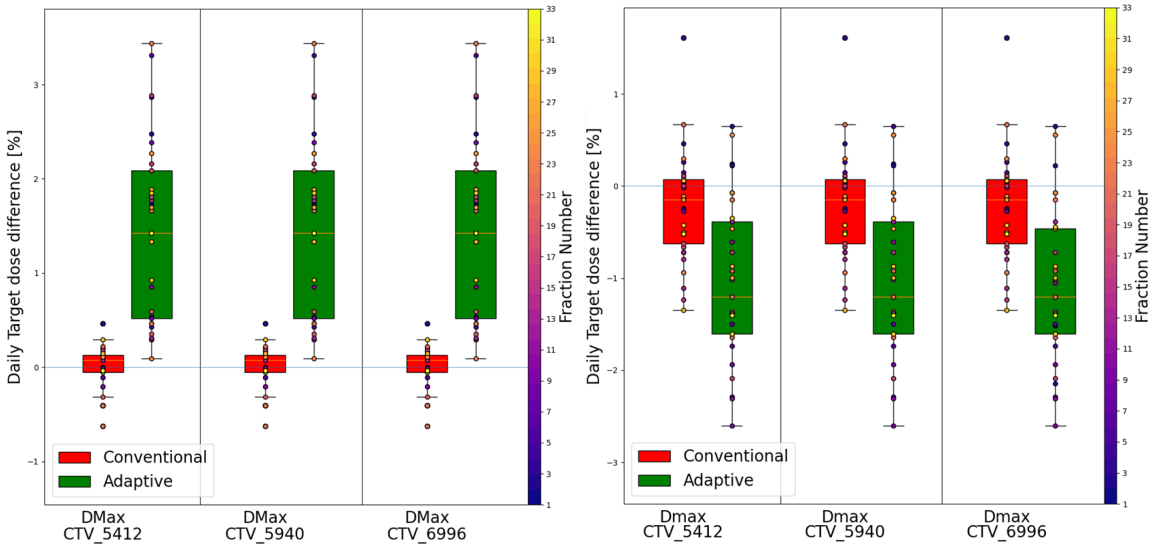

Figure 63: Results for FIonA (left) and RayStation® (right) for Dmax in all level targets.

D Dose Volume Histograms

Each daily fraction had a unique dose-volume histogram (DVH) for every structure involved. These individual DVHs were exported and compiled into a single comparative plot, allowing for a qualitative analysis of the dose distribution between the adaptive and conventional approaches for each configuration.

The DVH plots included in this appendix present, for each structure, both the planned dose in the planning CT and the simulated dose for each daily fraction plotted all together with color-band reflecting minimum, maximum and median DVH curve throughout the simulated treatment course. Although a large number of structures were analyzed, the focus was placed on the targets for each case to facilitate a better comparison of the target coverage of the different configurations. Additionally, the DVH of one OAR was also selected for each case.

The DVHs were plotted in terms of daily doses, as the variations in daily DVHs reflect changes in dose distributions caused by daily anatomical changes. The number of fractions is 33 for every Case, resulting in daily doses of 1.64, 1.8, and 2.12 Gy RBE for the low-, mid-, and high-dose targets (CTV_5412, CTV_5940 and CTV_6996), respectively.

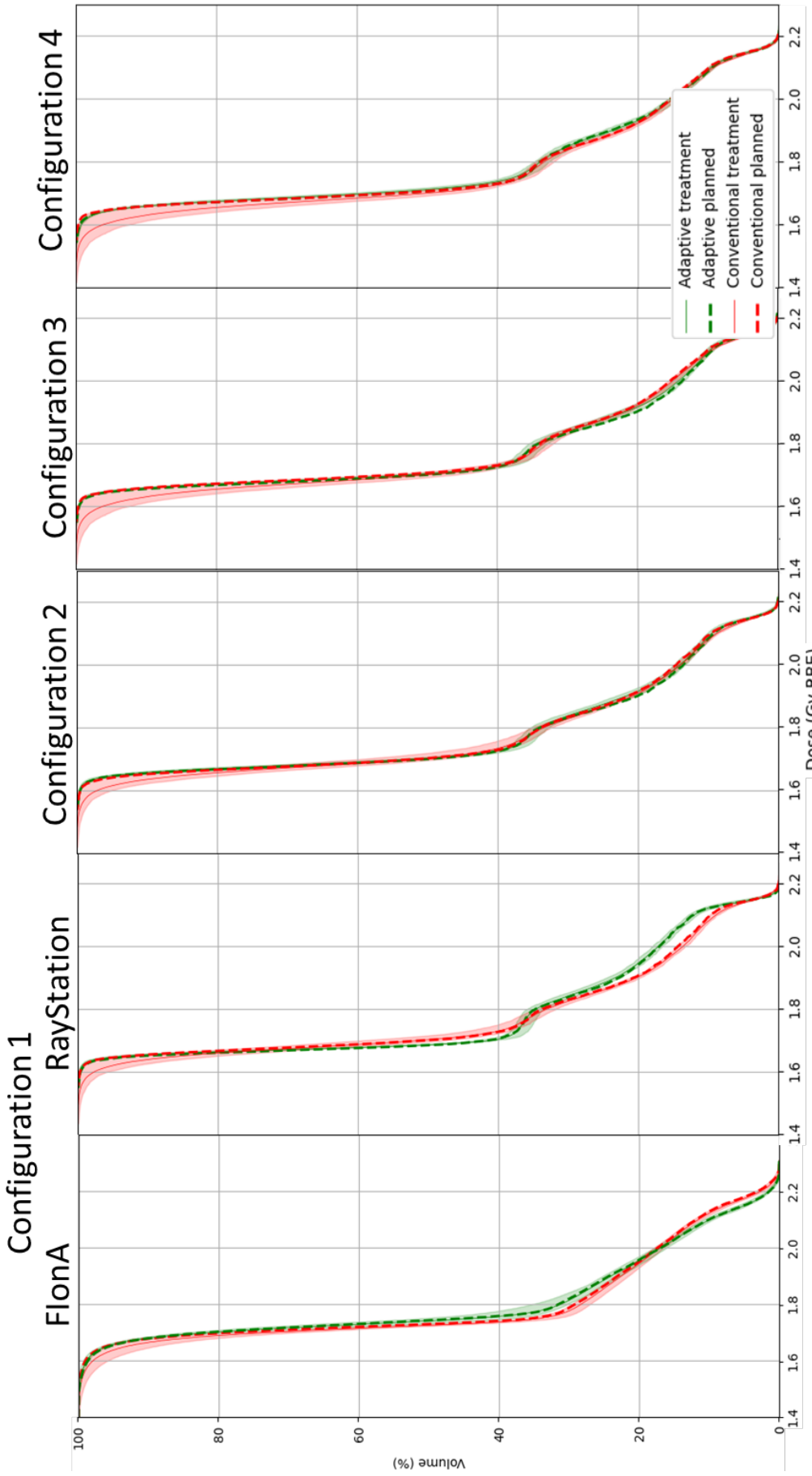

Figure 64: DVH for lower-dose target (CTV_5412) for Case 1 for all configurations.

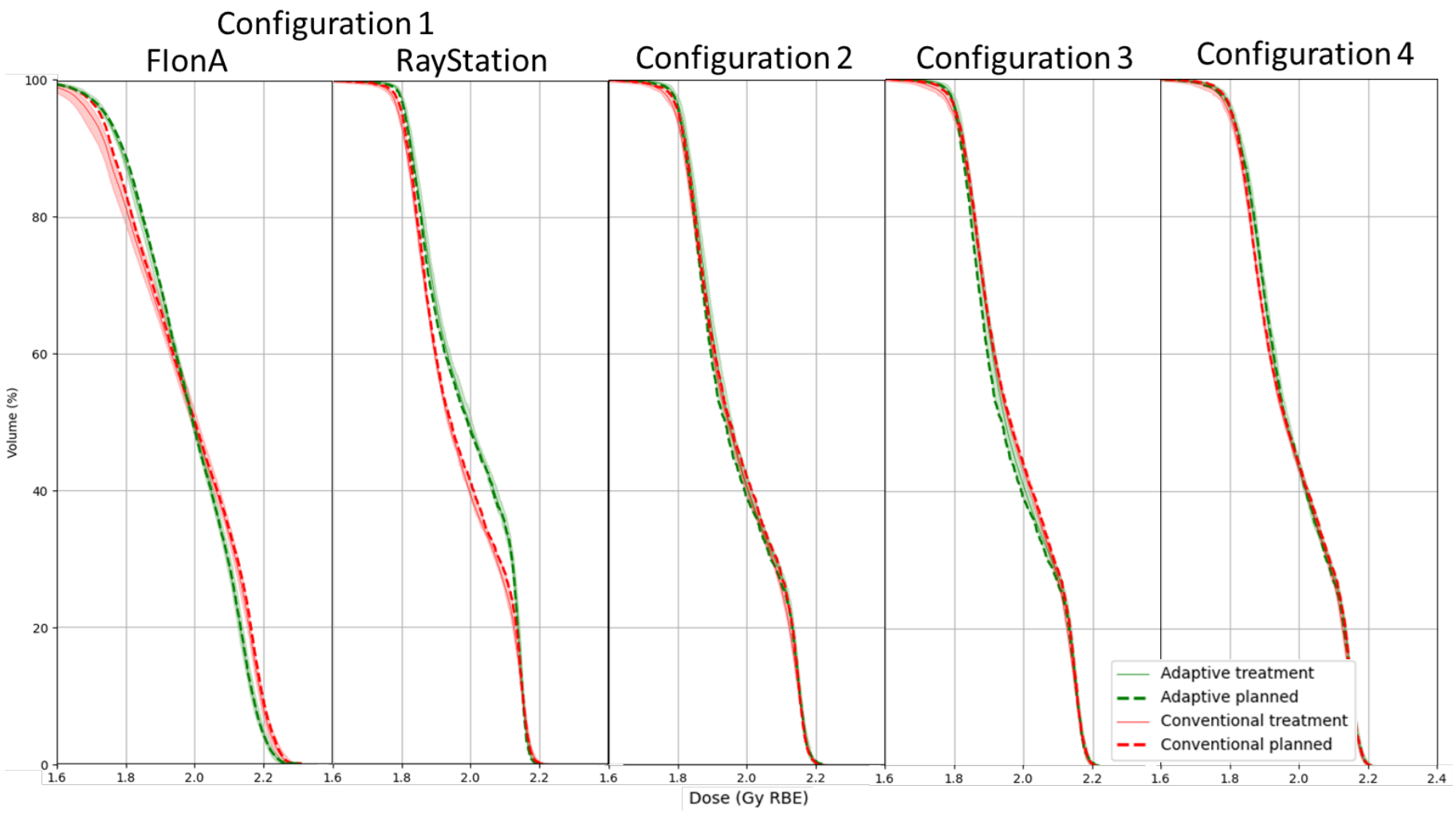

Figure 65: DVH for mid-dose target (CTV_5940) for Case 1 for all configurations.

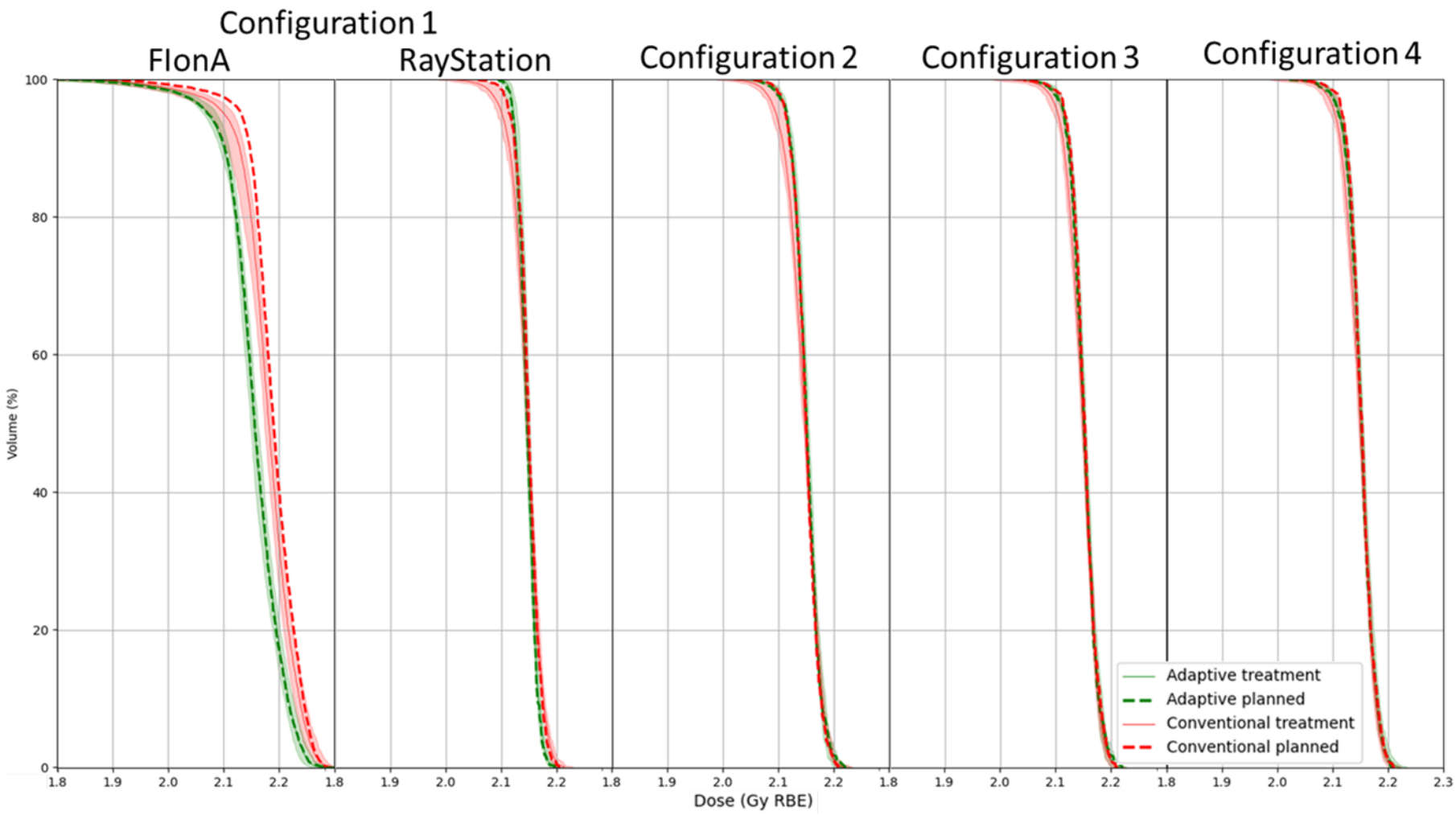

Figure 66: DVH for high-dose target (CTV_6996) for Case 1 for all configurations.

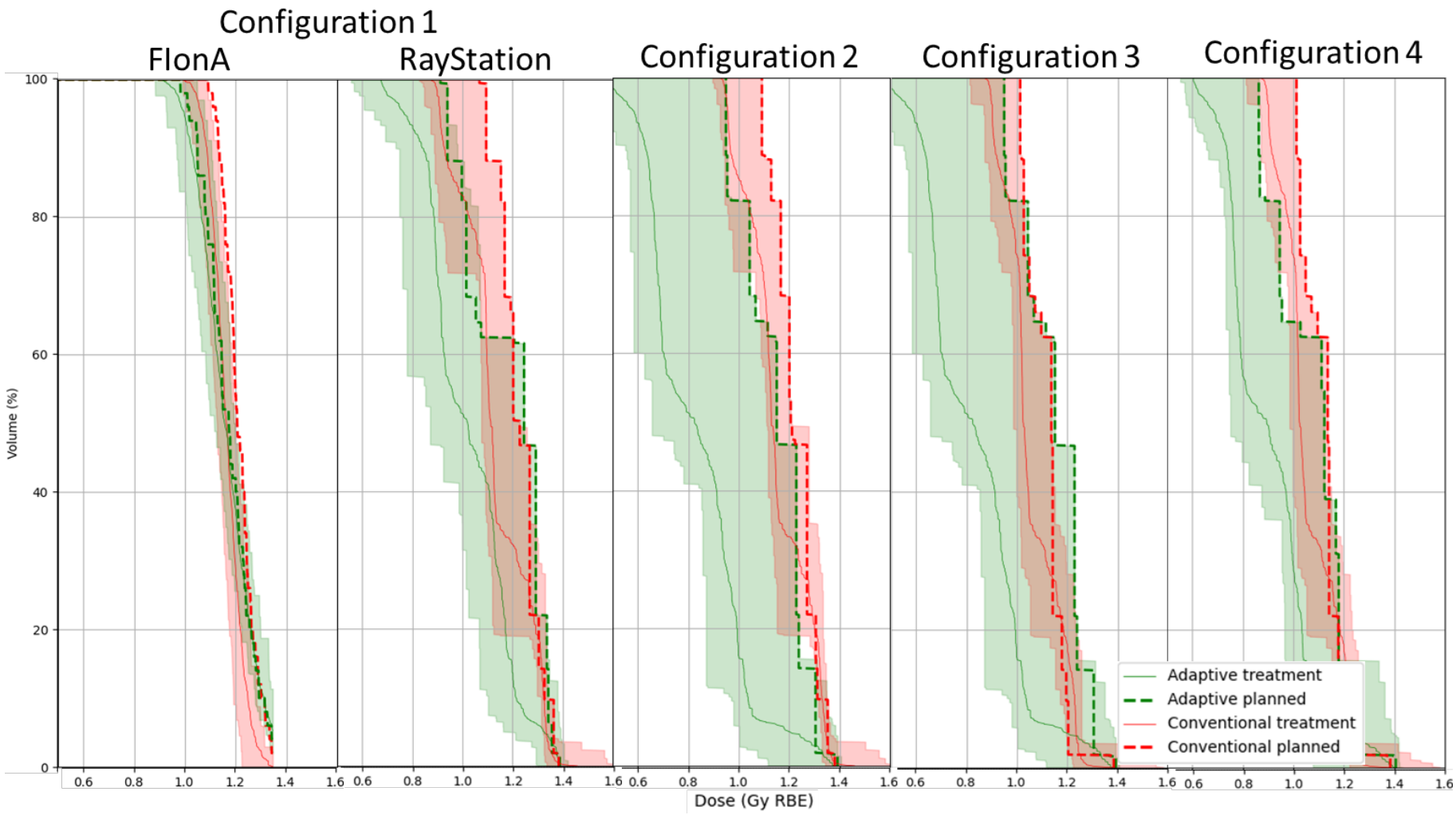

Figure 67: DVH for right cochlea for Case 1 for all configurations.

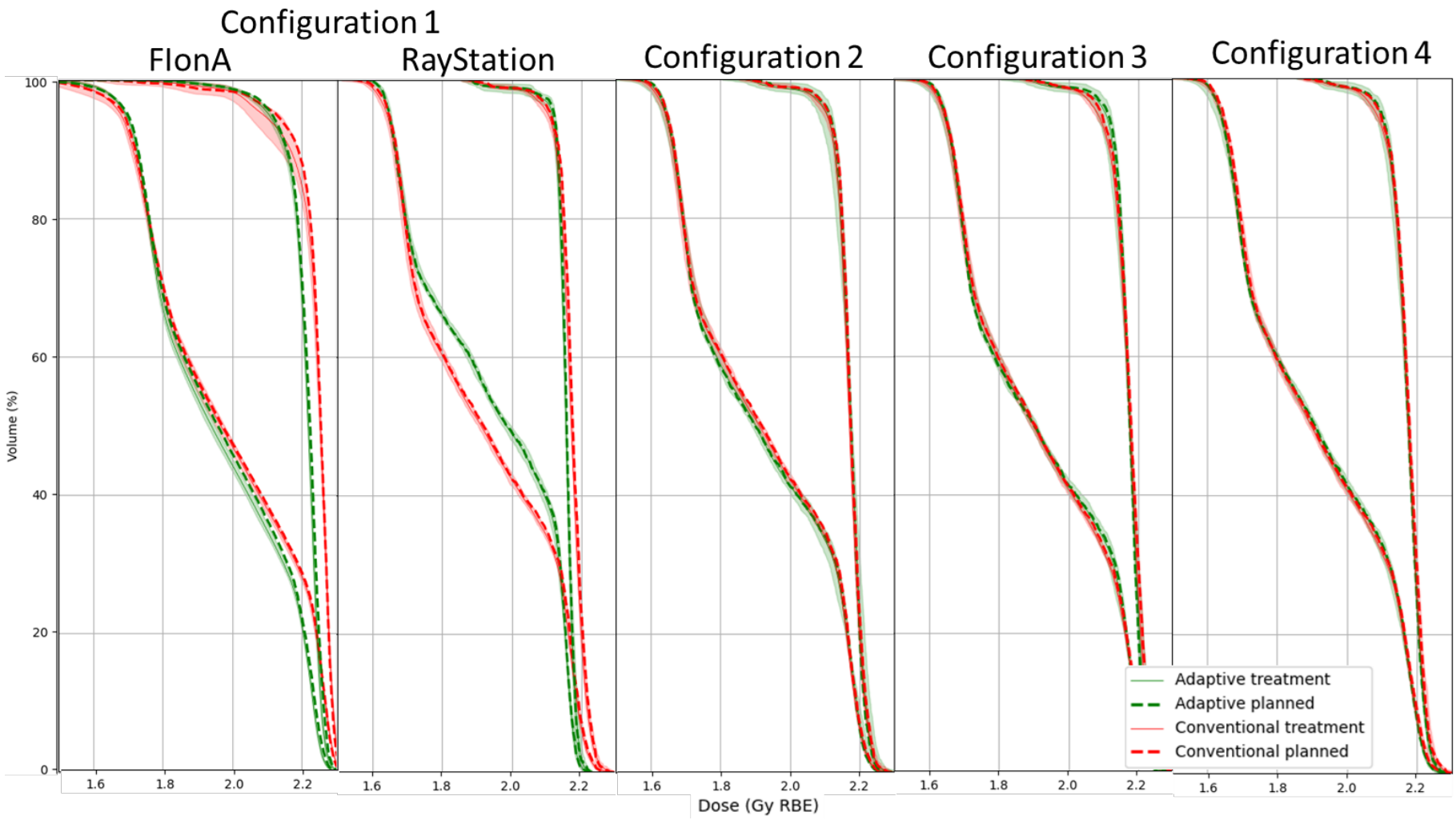

Figure 68: DVH for lower-dose target (CTV_5412) and for high-dose target (CTV_6996) for Case 2 for all configurations.

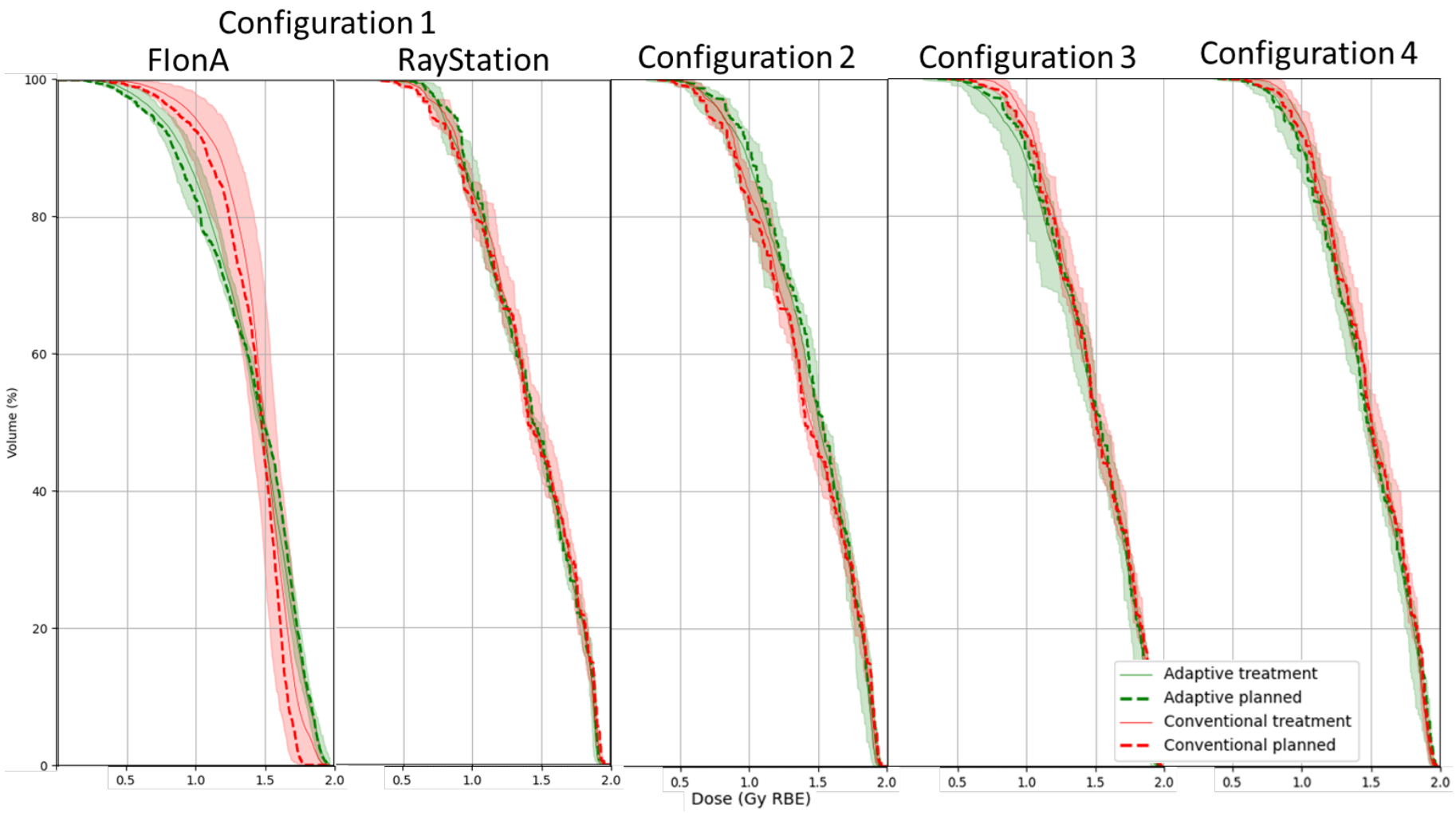

Figure 69: DVH for superior pharyngeal constrictor for Case 2 for all configurations.

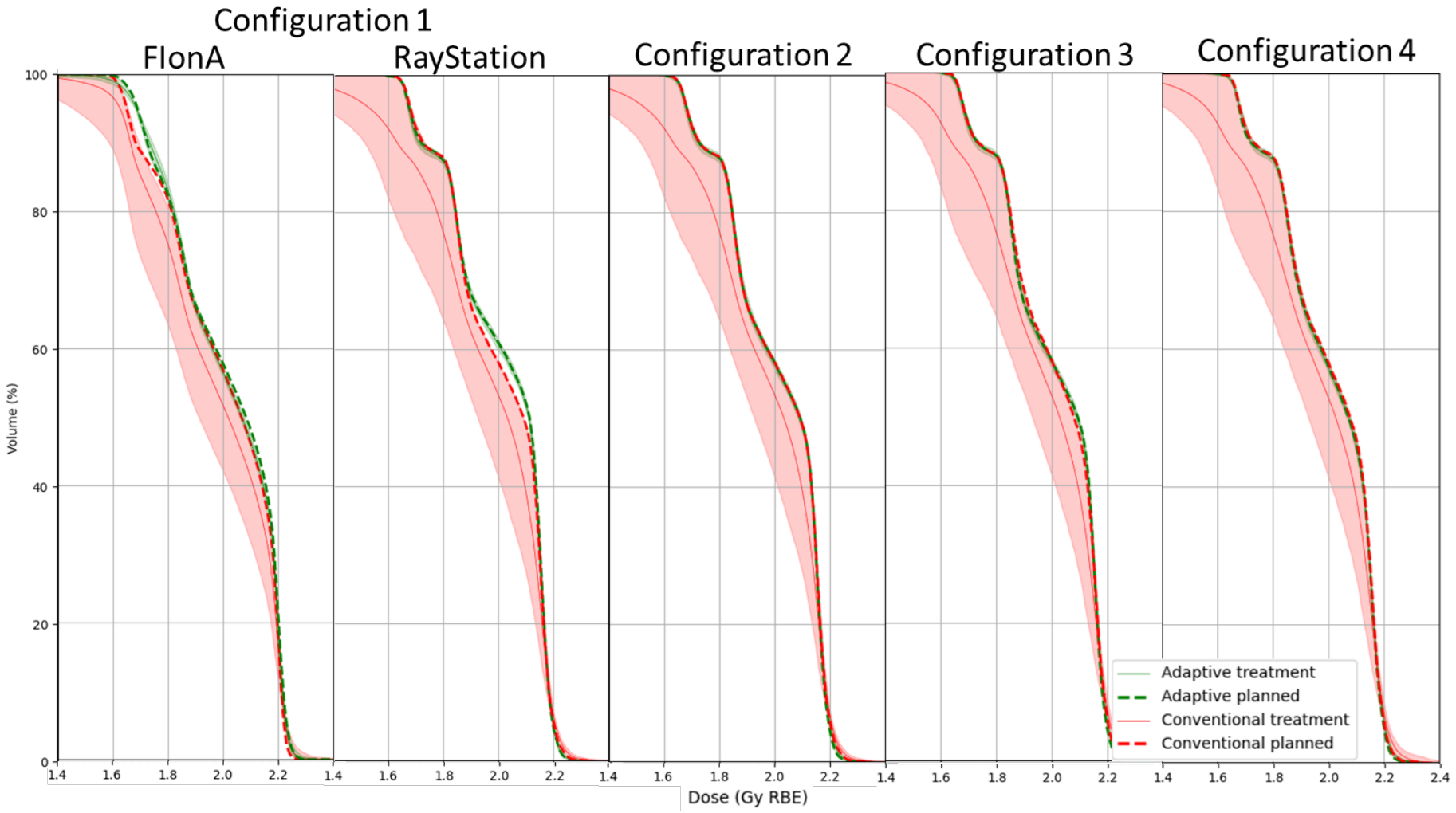

Figure 70: DVH for lower-dose target (CTV_5412) for Case 3 for all configurations.

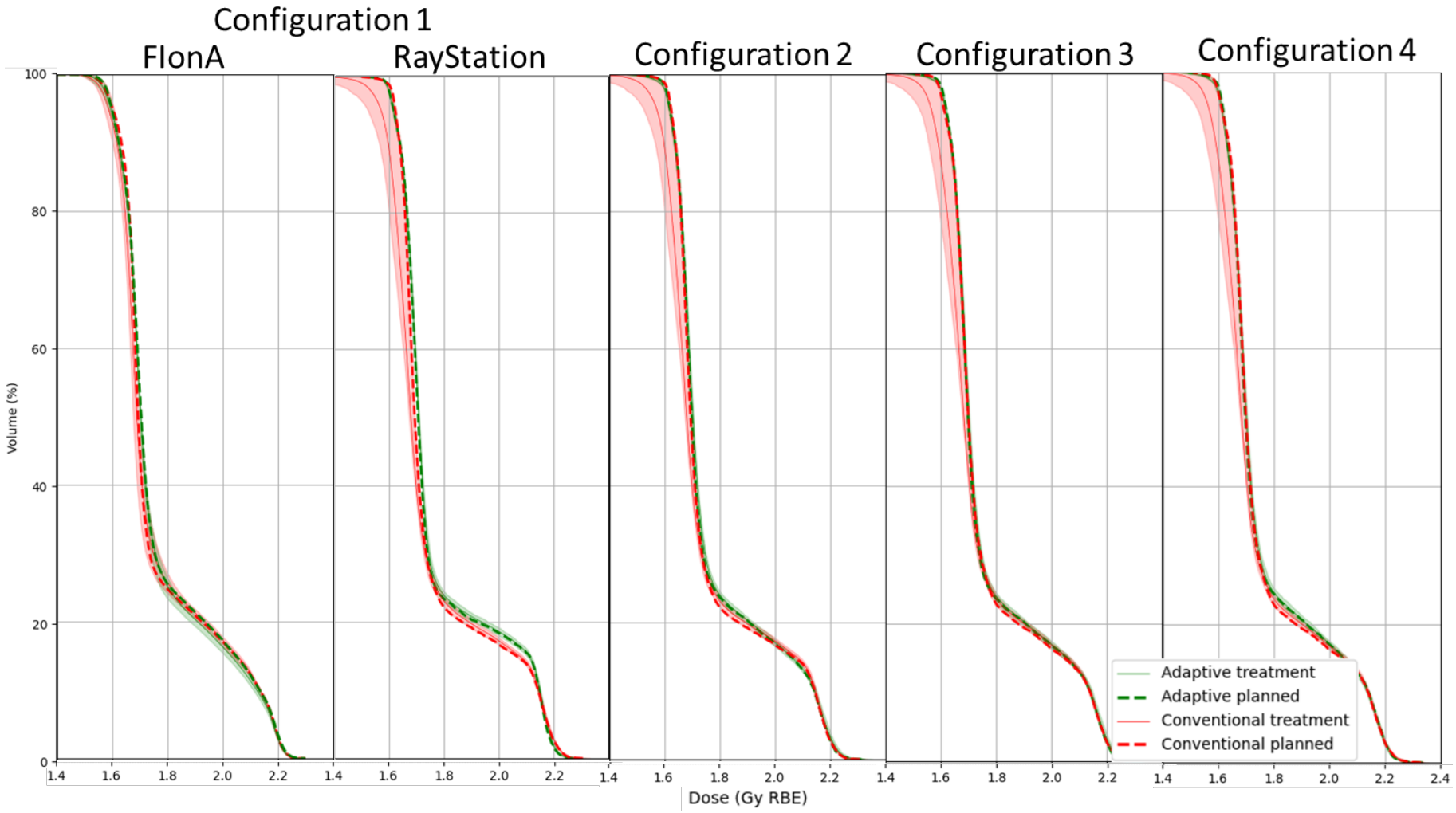

Figure 71: DVH for mid-dose target (CTV_5940) for Case 3 for all configurations.

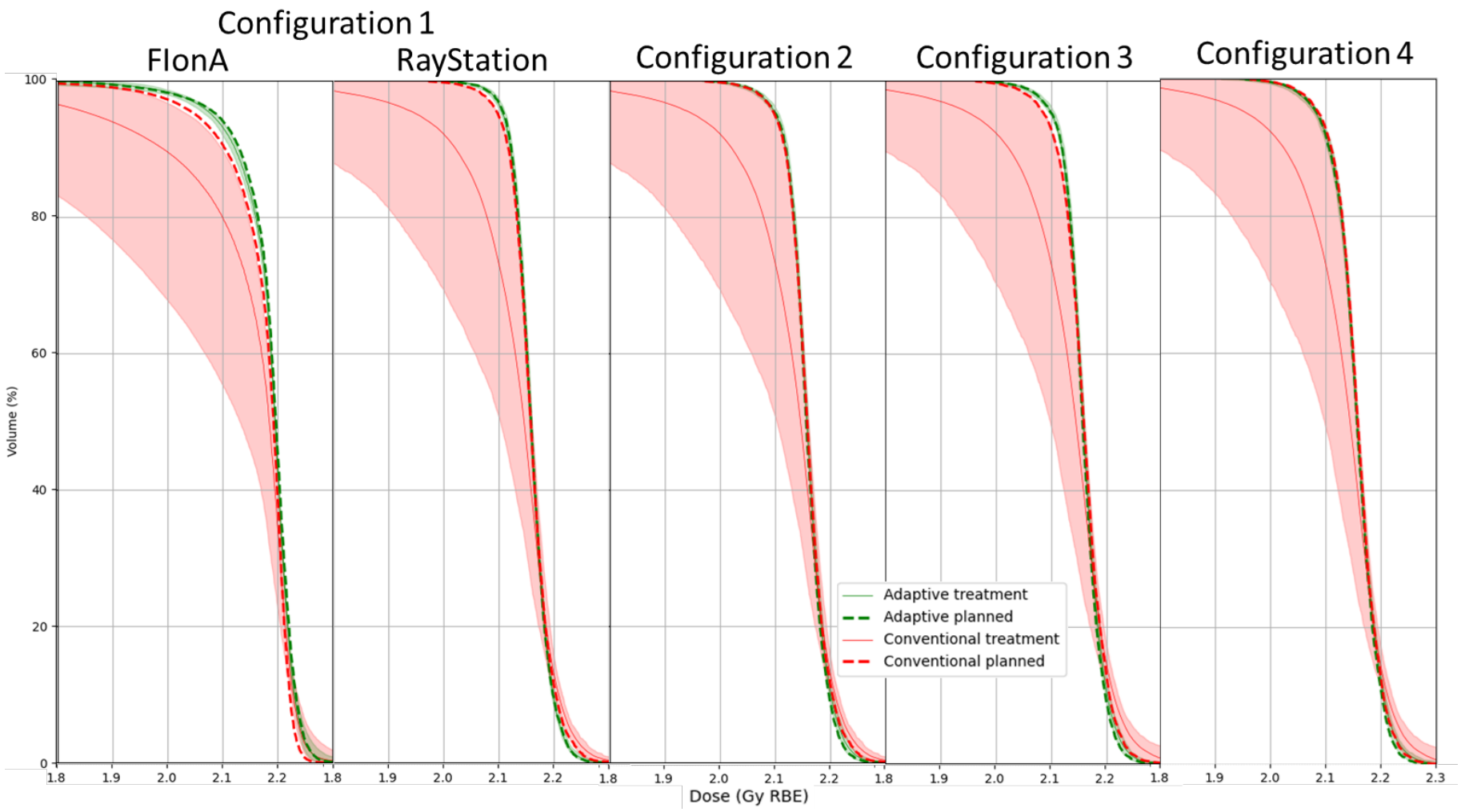

Figure 72: DVH for high-dose target (CTV_6996) for Case 3 for all configurations.

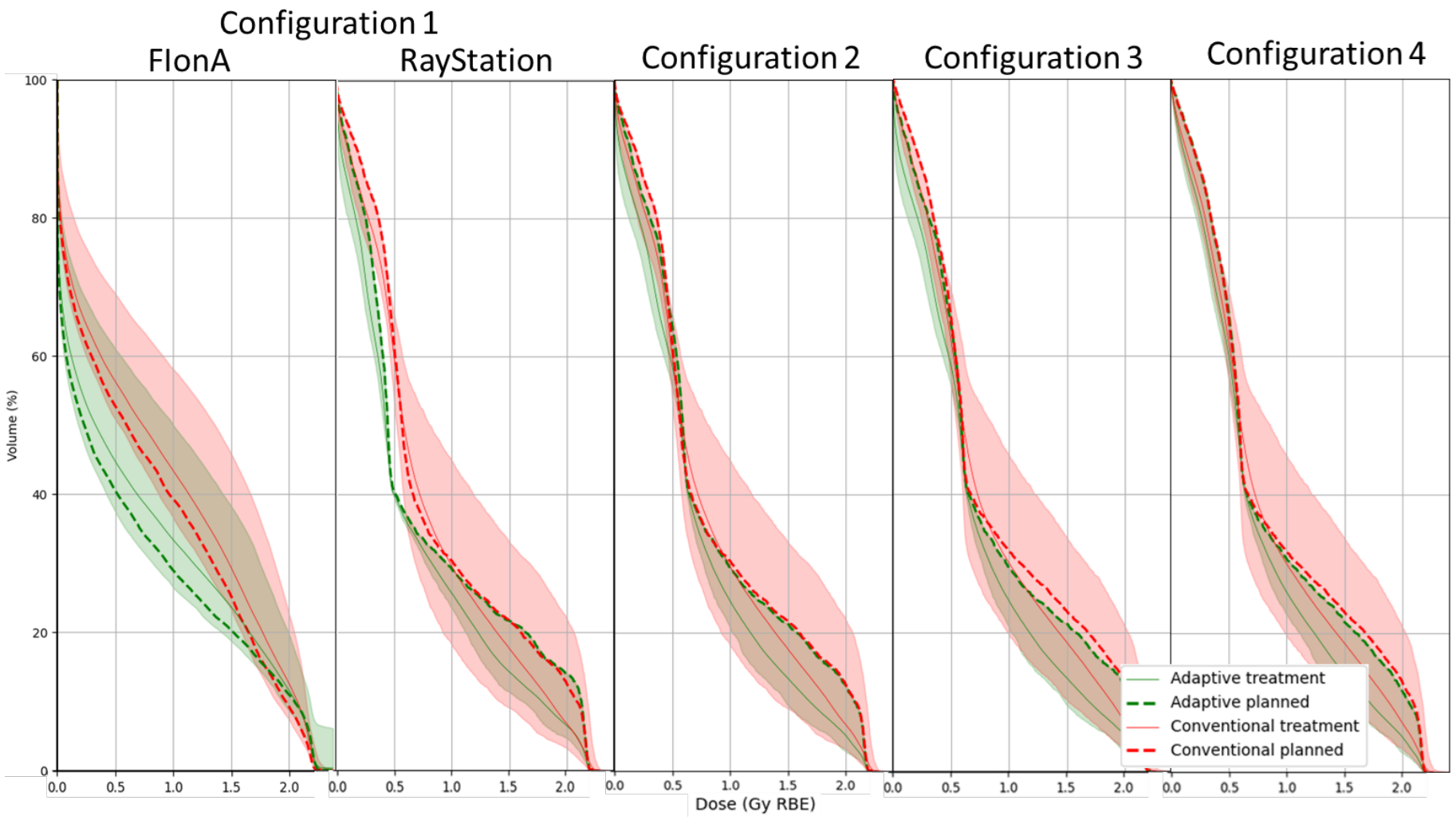

Figure 73: DVH for oral cavity for Case 3 for all configurations.

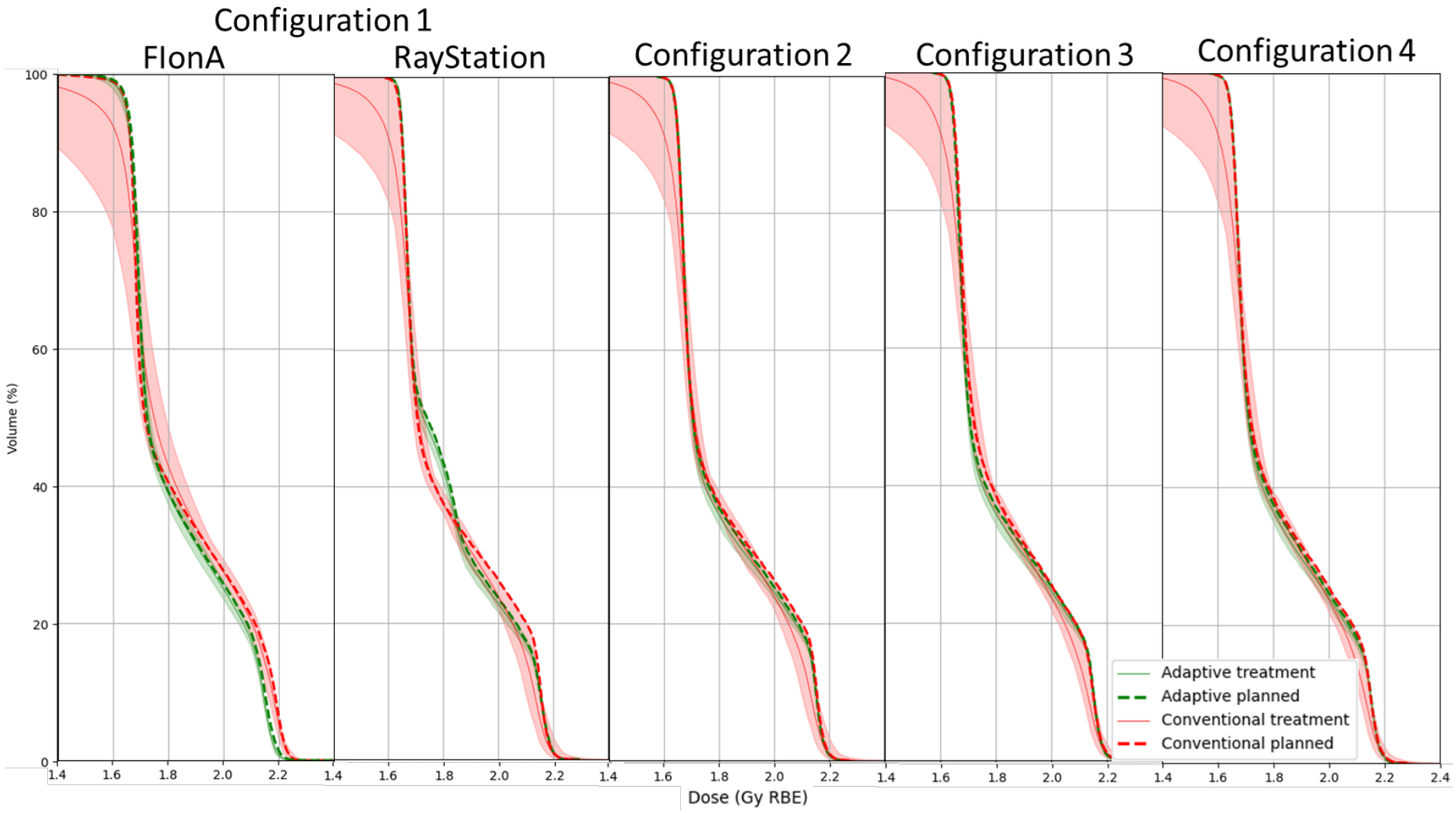

Figure 74: DVH for lower-dose target (CTV_5412) for Case 4 for all configurations.

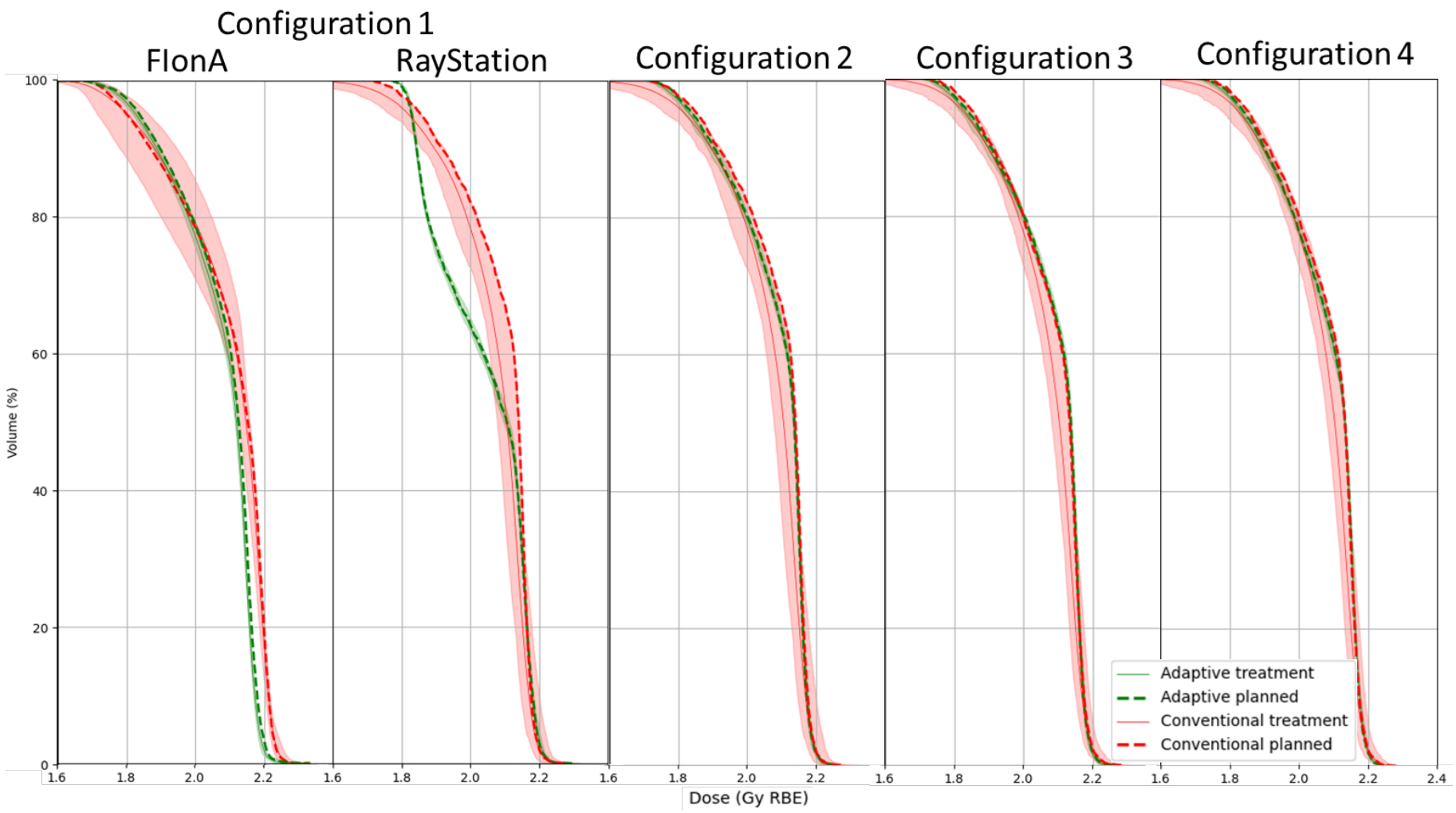

Figure 75: DVH for mid-dose target (CTV_5940) for Case 4 for all configurations.

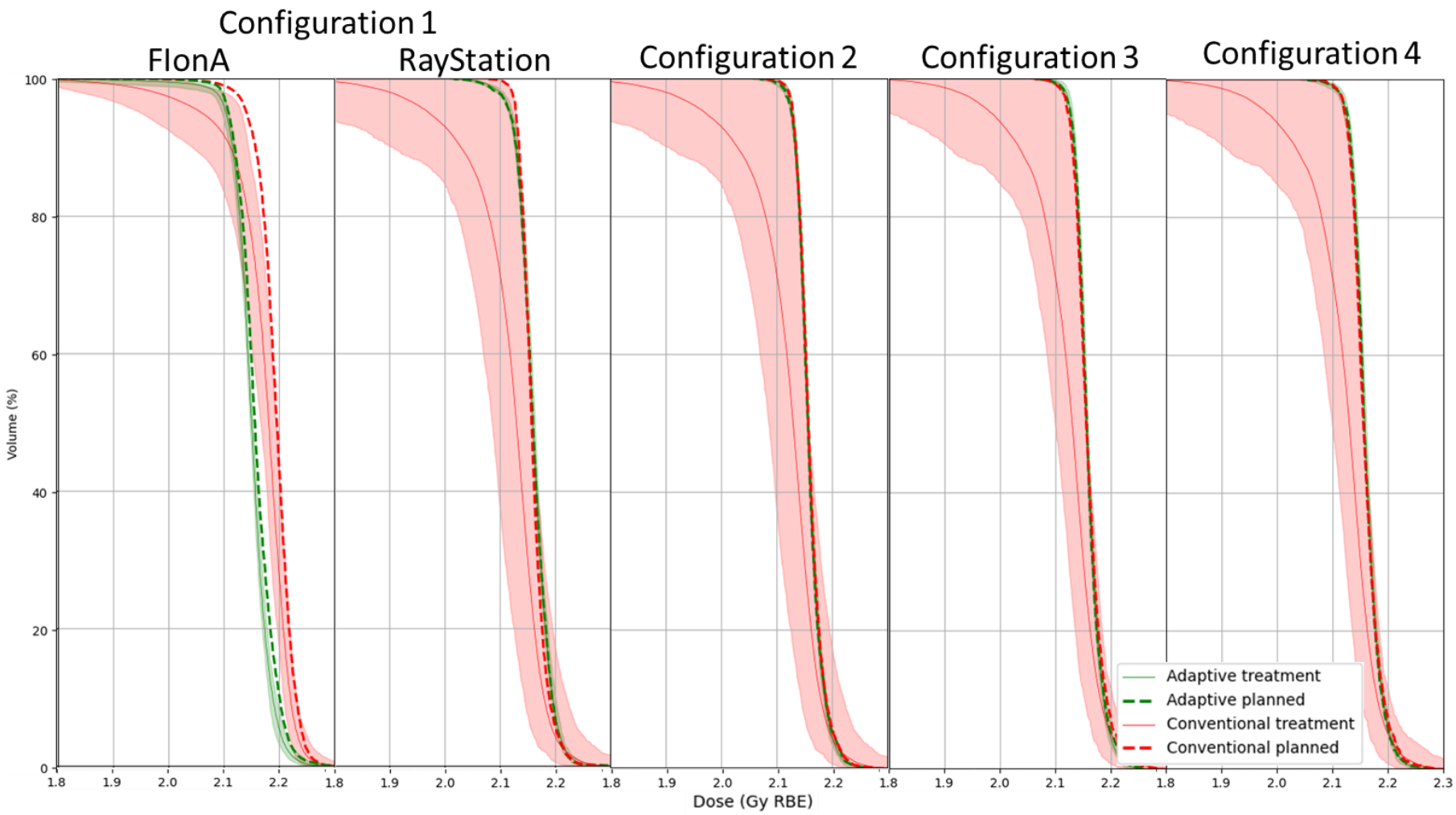

Figure 76: DVH for high-dose target (CTV_6996) for Case 4 for all configurations.

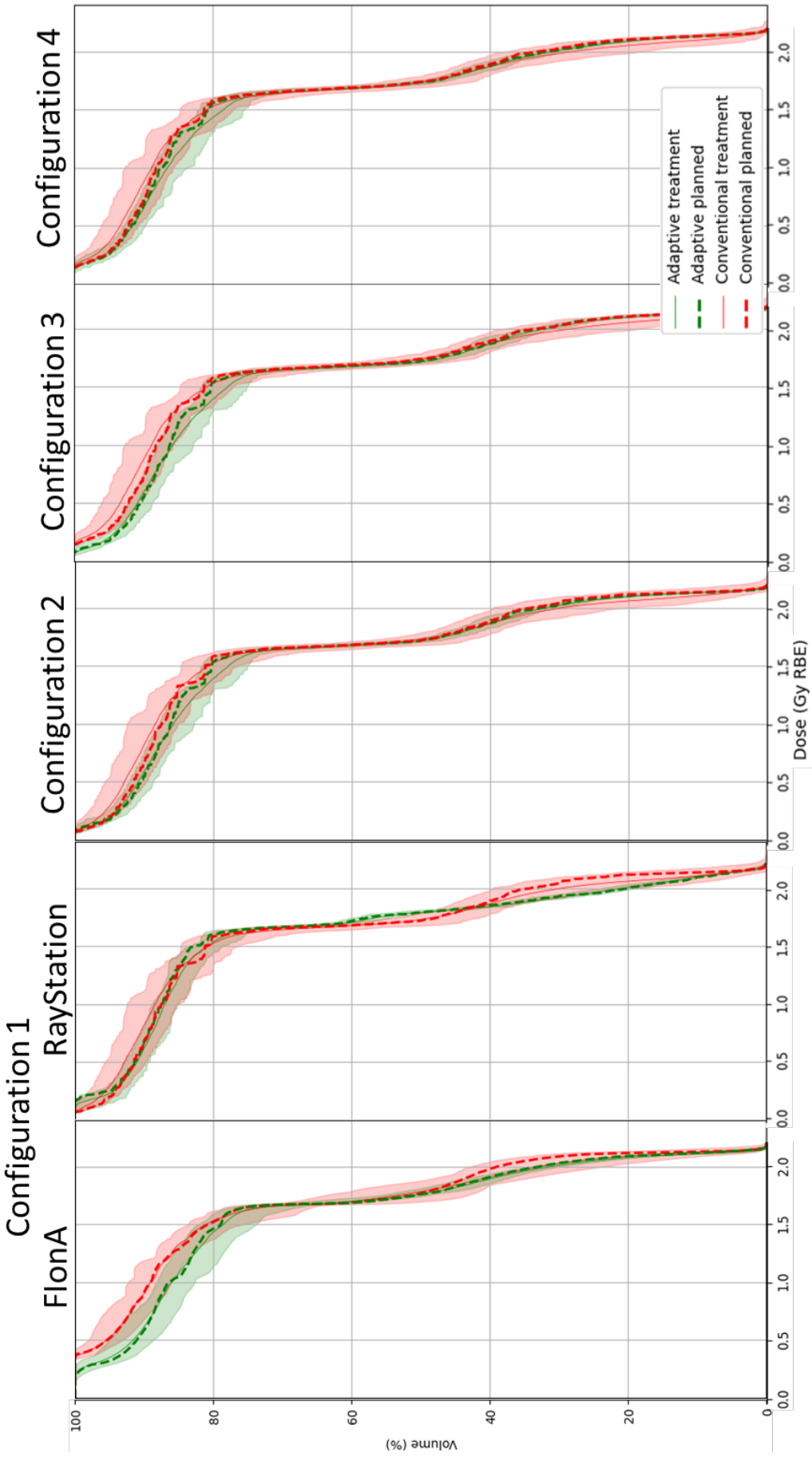

Figure 77: DVH for right carotid for Case 4 for all configurations.

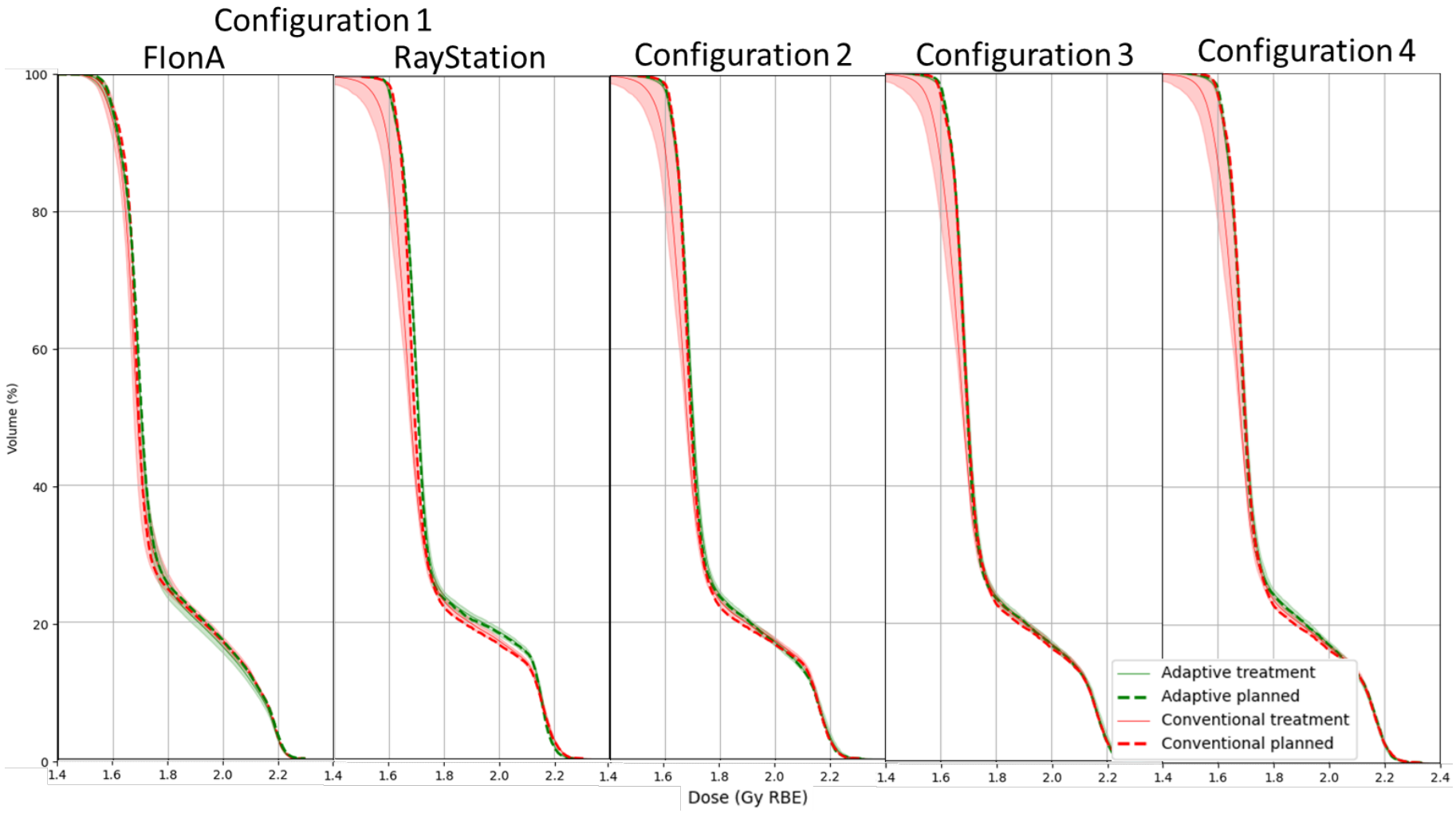

Figure 78: DVH for lower-dose target (CTV_5412) for Case 5 for all configurations.

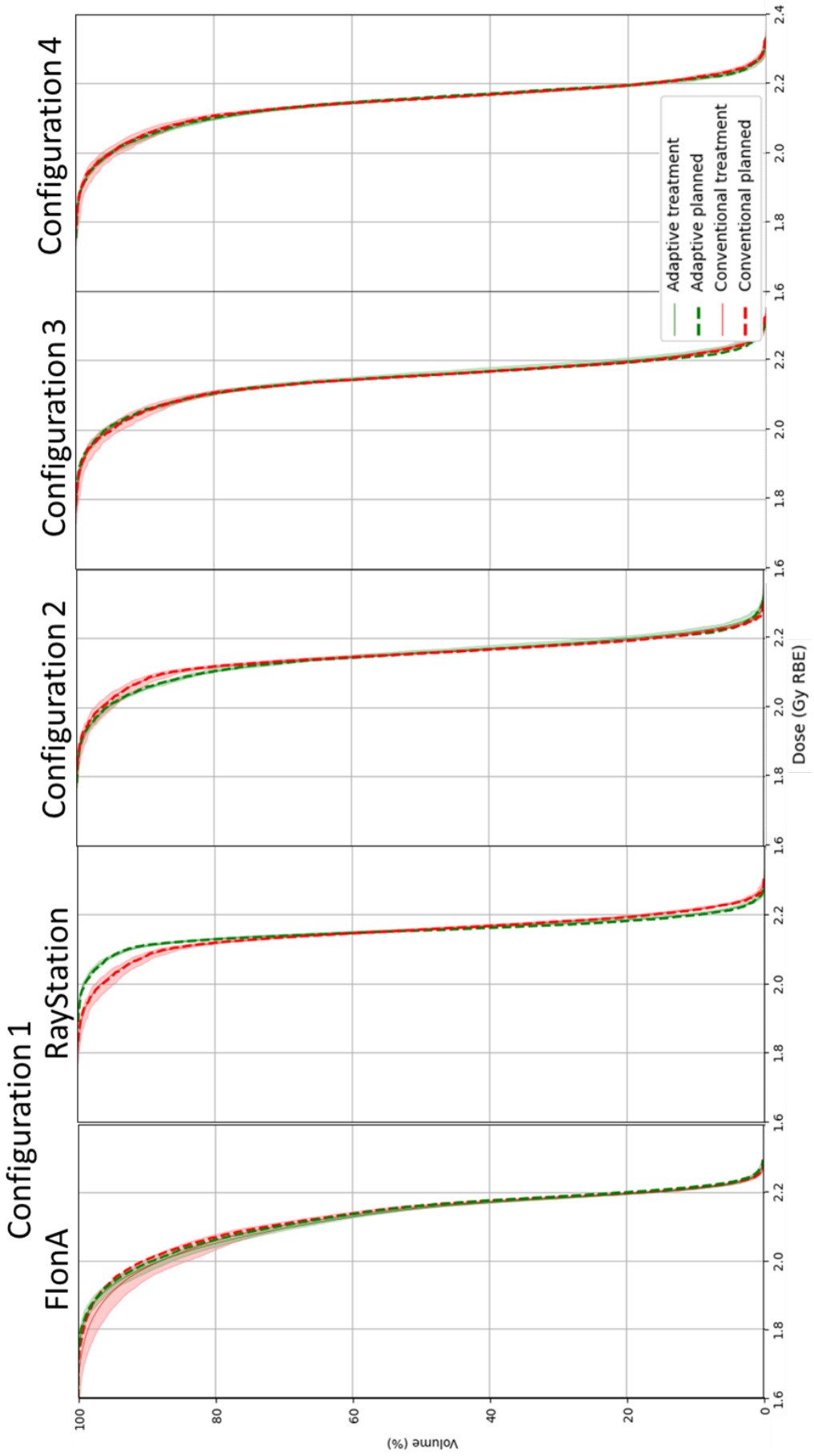

Figure 79: DVH for mid-dose target (CTV_5940) for Case 5 for all configurations.

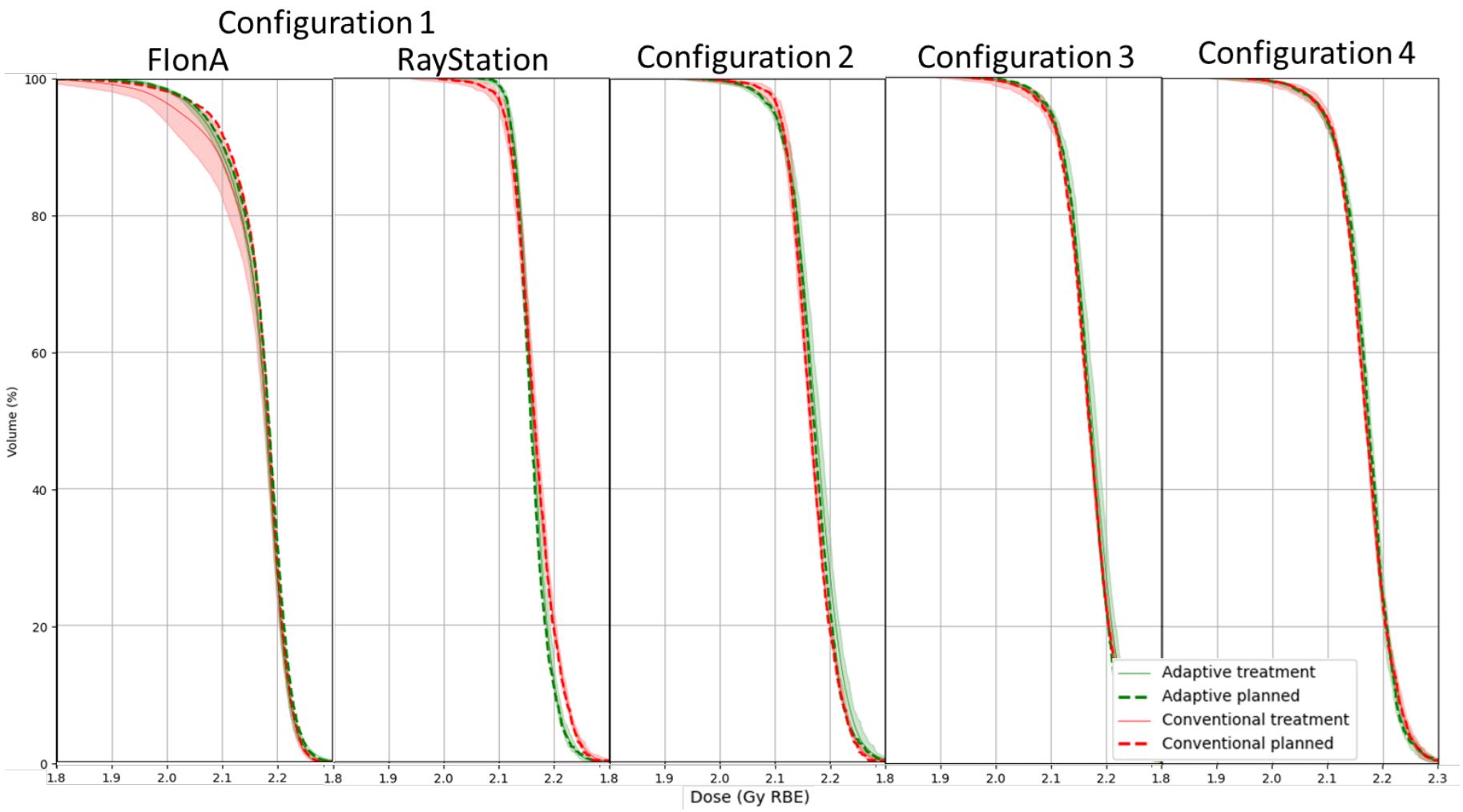

Figure 80: DVH for high-dose target (CTV_6996) for Case 5 for all configurations.

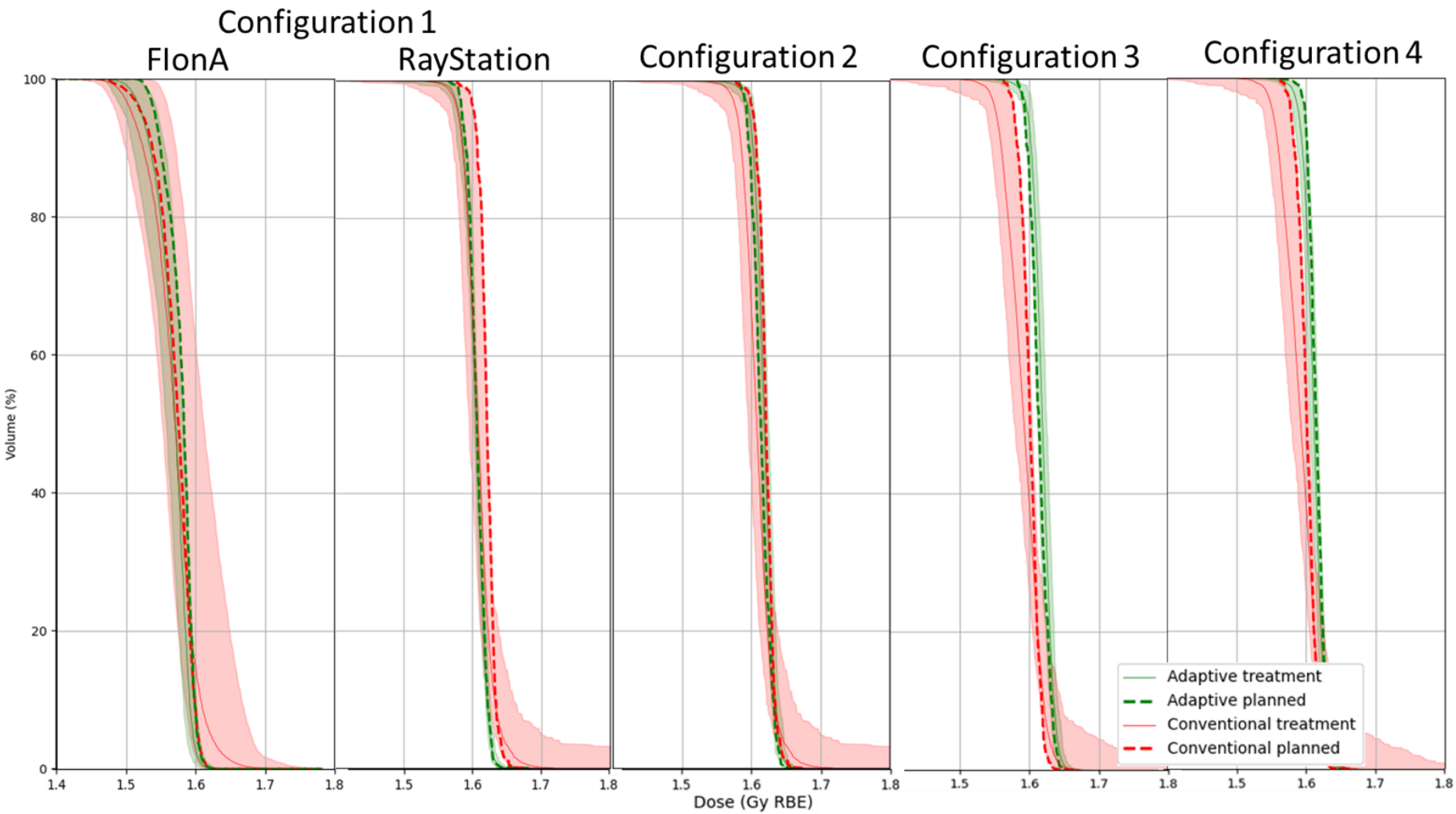

Figure 81: DVH for left submandibular for Case 5 for all configurations.

References

[1] Michael Matter. *Technical Implementation of Daily Adaptive Proton Therapy.* PhD thesis, ETH Zurich, 2020.

[2] Lisa Stefanie Fankhauser. A comprehensive analysis of uncertainties and their potential clinical consequences for head and neck cancer. Master's thesis, ETH Zurich, 2023.

[3] Michalis Aristophanous, Eric Aliotta, Philip Lichtenwalner, Shira Abraham, Mohammad Nehmeh, Amanda Caringi, Peng Zhang, Yu-Chi Hu, Pengpeng Zhang, Laura Cervino, Daphna Gelblum, Sean McBride, Nadeem Riaz, Linda Chen, Yao Yu, Kaveh Zakeri, and Nancy Lee. Clinical experience with an offline adaptive radiation therapy head and neck program: Dosimetric benefits and opportunities for patient selection. *International Journal of Radiation*, 119(5):1557–1568, 2024.

[4] Francesca Albertini, Katarzyna Czerska, Miriam Vazquez, Ilija Andaca, Barbara Bachtiary, Rico Besson, Alessandra Bolsi, Anne Sophie Bogaert, Evangelia Choulilitsa, Jan Hrbacek, Sisse Jakobsen, Dominic Leiser, Michael Matter, Alexandre Mayor, Gabriel Meier, Andre Nanz, Lena Nenoff, David Oxley, Dorota Siewert, Benno Andreas Rohrer Schnidrig, Andreas Johan Smolders, Hubert Szweda, Michelle Van Heerden, Carla Winterhalter, Antony John Lomax, and Damien Charles Weber. First clinical implementation of a highly efficient daily online adapted proton therapy (dapt) workflow. *Physics in Medicine and Biology*, 69(10):105003, 2024.

[5] Michael Matter, Lena Nenoff, Gabriel Meier, Damien Charles Weber, Antony John Lomax, and Francesca Albertini. Intensity modulated proton therapy plan generation in under ten seconds. *ACTA ONCOLOGICA*, 58(20):1435–1439, 2019.

[6] Jerry L Jr 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. *Int J Radiat Oncol Biol Phys.*, 59(4):960–970, 2004.

[7] Lawrence B. Marks, Ellen D. Yorke, Andrew Jackson, Randall K. Ten Haken, Louis S. Constine, Avraham Eisbruch, Søren M. Bentzen, Jiho Nam, and Joseph O. Deasy. Use of normal tissue complication probability models in the clinic. *International Journal of Radiation Oncology Biology Physics*, 76(3, Supplement):S10–S19, 2010.

[8] S. Mogens Bentzen, Jens Overgaard, Jørgen Johansen, Erik Andersen, Hanne Primdahl, and Cai Grau. Dahanca radiotherapy guidelines, 2020.

[9] Eros Pedroni, Ralph Beapark, Terence Böhringer, Adolf Coraz, Jürgen Duppich, Sven Forss, David George, Martin Grossmann, Gudrun Goitein, Christian Hilbes, Martin Jermann, Shixiong Lin, Antony Lomax, Marco Negrayus, Marco Schippers, and Goran Kotrie. The psi gantry 2: a second generation proton scanning gantry. *Zeitschrift für Medizinische Physik*, 14:25–34, 2004.

[10] A. Koschik, C. Bula, J. Duppich, M. Grossmann, J. M. Schippers, and J. Welte. Gantry 3: Further development of the psi proscan proton therapy facility. *Proceedings of the 6th International Particle Accelerator Conference (IPAC'15)*, pages 2700–2702, 2015.

[11] International Commission on Radiation Units and Measurements (ICRU). Icru report 50: Prescribing, recording, and reporting photon beam therapy, 1993. Available: `https://www.icru.org/`.

[12] Wei Liu, Xiaodong Zhang, Yupeng Li, and Radhe Mohan. Robust optimization of intensity modulated proton therapy. *Physics in Medicine and Biology*, 60(19):7437–7449, 2015.

[13] Marianna Canonaco. Towards daily adaptive proton therapy: tolerances evaluation for quality assurance checks and dose back calculation. Master's thesis, Politecnico di Milano, 2022.

[14] Mislav Bobić, Evangelia Choulilitsa, Hoyeon Lee, Katarzyna Czerska, Jeppe B. Christensen, Alexandre Mayor, Sairos Safai, Brian A. Winey, Damien C. Weber, Antony J. Lomax, Harald Paganetti, Konrad P. Nesteruk, and Francesca Albertini. Multi-institutional experimental validation of online adaptive proton therapy workflows. *Physics in Medicine Biology*, 69(16):165021, 2024.

[15] Andreas Johan Smolders, Evangelia Choulilitsa, Katarzyna Czerska, Nicola Bizzocchi, Reinhardt Krcek, Antony John Lomax, Damien Charles Weber, and Francesca Albertini. Dosimetric comparison of autocontouring techniques for online adaptive proton therapy. *Physics in Medicine and Biology*, 68(17):175006, 2023.

[16] Xin Li, Yuyu Zhang, Linghong Zhou, Yinghua Shi, Shuyu Wu, Yang Xiao, Xuejun Gu, and Xin Zhen. Comprehensive evaluation of ten deformable image registration algorithms for contour propagation between ct and cone-beam ct images in adaptive head neck radiotherapy. *PLOS ONE*, 12(4):e0175906, 2017.

[17] Francesca Albertini. *Planning and Optimizing Treatment Plans for Actively Scanned Proton Therapy: evaluating and estimating the effect of uncertainties.* PhD thesis, ETH Zurich, 2011.

[18] Michael Velec, Joanne L. Moseley, B. Math., Cynthia L. Eccles, Tim Craig, Michael B. Sharpe, Laura A. Dawnson, and Kristy K. Brock. Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *Int J Radiat Oncol Biol Phys.*, 80(1):265–272, 2011.

[19] Lena Nenoff. *Technical Implementation of Daily Adaptive Proton Therapy.* PhD thesis, ETH Zurich, 2020.

[20] Lena Nenoff, Michael Matter, Marjolaine Charmillot, Serge Krier, Klara Uher, Damien Charles Weber, Antony John Lomax, and Francesca Albertini. Experimental validation of daily adaptive proton therapy. *Phys. Med. Biol.*, 66(20), 2021.

[21] Uwe Schneider, Eros Pedroni, and Antony John Lomax. The calibration of ct hounsfield units for radiotherapy treatmnet planning. *Phys. Med. Biol.*, 41:111–124, 1996.

[22] Yuta Kai, Takeo Arimura, Ryuhei Toya, Masanori Tsuboi, and Masato Hiraoka. Comparison of rigid and deformable image registration for nasopharyngeal carcinoma radiotherapy planning with diagnostic position pet/ct. *Japanese Journal of Radiology*, 38(3):256–264, 2020.

[23] Pingfang Tsai, Yu-Lun Tseng, Brian Shen, Christopher Ackerman, Huifang Zhai, Francis Yu, Charles B. Simone, J. Isabelle Choi, Nancy Y. Lee, Rafi Kabarriti, Stanislav Lazarev, Casey L. Johnson, Jiayi Liu, Chin-Cheng Chen, and Haibo Lin. The applications and pitfalls

of cone-beam computed tomography-based synthetic computed tomography for adaptive evaluation in pencil-beam scanning proton therapy. *Cancers*, 15(20):5101, 2023.