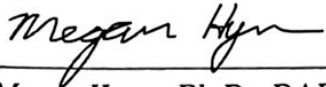


THESIS APPROVED BY

12/7/23
Date



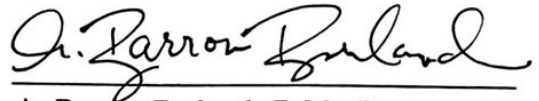
Megan Hyun, Ph.D., DABR



Michael Nichols, Ph.D.



Brian Hill, M.S., DABR



A. Barron Breland, D.M., Dean

EVALUATION OF TIMEPOINT ASSESSMENT SCHEMES
FOR OFFLINE ADAPTIVE RADIOTHERAPY OF THE
HEAD AND NECK USING VELOCITY

by

ANDREW FANNING

A THESIS

Submitted to the faculty of the Graduate School of Creighton
University in Partial Fulfillment of the Requirements for the degree of
Master of Science in Medical Physics in the Department of Physics

Omaha, NE
December 4, 2023

Abstract

The use of adaptive radiotherapy (ART) may improve care for those receiving head and neck cancer treatment by reducing dose to organs at risk (OARs) and allowing margin reduction for targets. Current literature investigates predetermined and triggered adaptation, but how to trigger and how often to assess remain open questions. So, a study of adaptive assessment schemes may help fill this gap in literature. One way to achieve this is through the creation of a reliable evaluation framework.

This study investigated the performance of midpoint and biweekly assessment of H&N treatment plans through a novel evaluation framework. Midpoint and biweekly assessment of H&N treatment plans was retroactively performed for three patients who received radiation treatment at the University of Nebraska Medical Center. These assessment schemes use a validated workflow based in Varian's Velocity software. Assessments triggered adaptive replanning if the dose to any of seven organs-at-risk was projected to exceed physician-determined dose constraints.

Once completed, the schemes were evaluated on their trustworthiness, dosimetric benefit, and efficiency. Trustworthiness was quantified through the ability of schemes to catch necessary replans and project final patient dose. Both schemes performed similarly when catching when replans were necessary, showing equal sensitivity and specificity metrics of 0.75 and 0.94-0.95, respectively. Midpoint and biweekly schemes predicted final delivered dose given replans with sensitivities of 0.95 and 0.86, respec-

tively. Midpoint assessment had a specificity of 0.5, while the biweekly assessment had insufficient data to calculate specificity.

Dosimetric benefit was quantified through comparison of the number of OARs whose dose constraints were violated without ART, with the total number seen when an assessment scheme was used. From four total OAR constraints being violated without adaptation, midpoint and biweekly assessment decreased this number to two and zero, respectively.

Lastly, the efficiency of both schemes was quantified as time-cost-effectiveness, where the cost was physicist time in hours. Regardless of the way benefit was quantified, the midpoint assessment performed more efficiently (i.e., with higher time-cost-effectiveness). This work shows that a clear picture of different modes of ART is given by this framework.

Acknowledgments

Firstly, I want to express my gratitude to my research advisor, Dr. Megan Hyun. I feel so lucky to have been able to learn from you and benefit from your expertise, kindness, and the high standards you set. I've always felt supported and believed in despite the many times I found a new and unique way to mess up the project. I'm also thankful for your patience throughout our many conversations distinguishing between adaptation, analysis, and assessment. Despite using way too many a-words, we found a way to adapt and analyze our assessments. Thank you so much.

I am also grateful to Dr. Michael Nichols, my research advisor on the Creighton side. I have a newfound appreciation for academic rigor, consistency, and relentless curiosity due to the example you set every day. Your advice in both research and academic capacities have helped me grow throughout the two years I've been at Creighton. Thank you, Dr. Nichols.

I would also like to thank Brian Hill, the third member of my thesis committee. I've learned so much about the clinical side of medical physics through the summer rotations you've helped construct and run. Through your help, my classmates and I have been able to contextualize the things we've seen in our rotations and grow in our knowledge. Thank you.

I'm thankful to my classmates for the friendships we've developed over my time at Creighton. My sanity is intact largely due to the levity many of you have brought

and I am so thankful for that. I can't wait to see the wonderful and successful medical physicists you all become.

Mom, dad, Cece, Zach, and Luke, thank you for the patience you've shown as I've grown throughout this program. I've felt nothing but unconditional love from all of you and that's given me the extra strength I've needed to make it through this degree. Thank you for believing in me.

Lastly, I would like to thank my girlfriend, Lillian, for her Herculean efforts of support and love. I would be nowhere near where I am today without you loving, supporting, and pushing me. You live out your values with reckless abandon every day and inspire me to do the same. It is such a blessing to know and love you. Thank you for everything.

Contents

Abstract	iii
Acknowledgments	v
1 Introduction	1
2 Background	5
2.1 Head & Neck Cancer	5
2.1.1 Diagnosis and Staging	7
2.1.2 Treatment Standard	8
2.2 Radiation Therapy of the Head and Neck	9
2.2.1 Radiation Therapy Workflow	11
2.2.2 Volumetric Modulated Arc Therapy (VMAT)	13
2.3 Deformable Image Registration	14
2.3.1 Objective Function	16
2.3.2 Transformation Model	20
2.3.3 Optimization Algorithm	22
2.3.4 DIR Quality Assurance	24
2.3.5 Summary	26
2.4 Adaptive Radiotherapy (ART)	27

3	Materials & Methods	32
3.1	Study Structure	33
3.1.1	Data Gathering	33
3.1.2	Patient Selection	34
3.1.3	Research Questions	34
3.2	Materials	35
3.2.1	Aria: Varian’s Record & Verify System	35
3.2.2	Eclipse: Varian’s Treatment Planning System	36
3.2.3	Velocity	37
3.3	Methods	38
3.3.1	Assessment Schemes	38
3.3.2	Research Question 1	49
3.3.3	Research Question 2	57
3.3.4	Research Question 3	58
4	Results & Discussion	61
4.1	Research Question 1	63
4.1.1	How well did schemes detect triggers?	63
4.1.2	How well do assessment schemes predict final delivered dose?	66
4.2	Research Question 2	68
4.3	Research Question 3	70
4.4	Discussion	71
4.5	Study Limitations	76
4.6	Future Work	77
5	Conclusions	79
A	Adaptive Workflow	82

B	aCT Creation Workflow	84
C	Replanning Instructions	91
D	Dose Accumulation Instructions	97
E	Copyright Permissions	102
F	Data Appendix	105

List of Tables

3.1	OARs used for Dose Analysis step of offline ART workflow. Dose constraints were taken from the original treatment plans of the patients in the study. Max means the maximum dose to the structure. Mean is the average dose delivered to the structure. *Constraint changed from patient to patient.	40
3.2	Comparison of Patient 2's dose constraints and $(D_{test})_{Mid,Wk4}$ for seven OARs.	42
3.3	Comparison of $(D_{test})_{Mid,Wk4}$ and re-optimized dose with dose constraints for Patient 2. The replan improves dose sparing to the OARs which may have violated their dose constraints (shown in red) while ensuring that target coverage is maintained.	43
3.4	The midpoint assessment's estimation for the total dose delivered to Patient 2 when a replan is performed at week four. When compared to the week four assessment, total triggers decreased and the OAR still triggering (shown in red) decreased in dose.	44
3.5	Week two's dose estimation for Patient 2 created through the biweekly use of the offline adaptive workflow.	47
3.6	Week four's dose estimation created through the biweekly use of the offline adaptive workflow.	48

3.7	Week six’s dose estimation created through the biweekly use of the offline adaptive workflow.	48
3.8	Truth and Test values for the midpoint assessment’s ”Detector” evaluation. Here, we note that any red colored numbers represent OAR triggers.	53
3.9	Truth and Test values for the evaluation of the midpoint assessment’s ability to predict final dose with adaptation. Here, we note that any red numbers represent OAR triggers.	56
3.10	The estimated time-cost associated with each step of the midpoint and biweekly assessment processes used for offline adaptive radiotherapy.	60
4.1	Comparison of final delivered dose with and without adaptive replanning when the midpoint assessment scheme was used. The last column shows the dose difference between the two.	69
4.2	Comparison of final delivered dose with and without adaptive replanning when the biweekly assessment scheme was used. The last column shows the dose difference between the two.	70
4.3	Dosimetric benefits seen when using the midpoint and biweekly assessment schemes on three patients with a total of 21 OARs.	70
4.4	Time-cost-effectiveness of midpoint and biweekly assessment schemes using three separate dosimetric benefit metrics.	70
4.5	Summary of results relevant to Research Question 1. Both sensitivity and specificity for the schemes’ ability to detect replan necessity and predict final dose are listed.	73

List of Figures

2.1	Anatomy of the head and neck region. Cancers of the head and neck most often form in the squamous cells of these organs. (2012) Terese Winslow LLC, U.S. Govt. has certain rights.	6
2.2	T-classification portion of the staging system for oral cancers as determined by the American Joint Committee on Cancer[17].	8
2.3	A Varian Edge linear accelerator at the University of Nebraska Medical Center (UNMC). Radiation is delivered via the gantry head to the patient positioned on the treatment table. The kV imager components are used to create cone-beam CT (CBCT) images which are used in this study. The MV imager can be used to capture a beam’s-eye-view of the radiation field.	10
2.4	Volume classification as defined by ICRU Report 50.	10
2.5	Standard of Care for external beam radiation therapy of the head and neck. Imaging and treatment delivery often take place over the course of 35 fractions for a H&N patient.	12

2.6	Fixed image (A) and moving image (B) were registered together using DIR algorithms. (C) shows the deformation vector field which was applied to the moving image in order for it to accurately deform to the shape of the fixed image. (D) displays the deformed moving image, which matches the fixed image exactly. Image taken from open access article [23].	15
2.7	(A) shows a simplified example of deformation using constrained regularization. Neighboring voxels move congruently with no overlap. (B) shows a simplified example of deformation with unconstrained regularization. Voxels are able to move freely from one another, which can lead to situations in which voxels overlap and create folds in the voxel grid. From Varian [26].	23
2.8	A flowchart for the deformable image registration process. Steps including the calculation of a similarity index, running of an optimization algorithm, and application of a transformation model to the moving image, are performed iteratively until a satisfactory similarity index value is met. Image taken from open access article [23].	24
3.1	Planned dose distribution for a patient receiving treatment for H&N cancer. The left side of the image shows which plan is open (here, it is called H & N PTV), as well as the structure set (contours) for the plan. a) shows an axial view of the patient’s dose distribution with a color wash. b) shows the plans resulting DVH, which helps clinicians visualize what levels of dose are delivered to which structures. c) shows a coronal (frontal) view of the patient’s dose distribution. d) shows a sagittal view of the dose distribution. Both a) and d) show the PTV contour in red.	37

3.2	Six steps of the offline ART workflow developed by Hyun et al. Each of these six steps can be carried out using Aria, Eclipse, and Velocity. See Appendix A for further details.	39
3.3	Sagittal view of a patient’s simulation CT, Week 4 CBCT, and the resulting aCT representing Week 4 anatomy. The aCT is created using Velocity’s DIR algorithm based on a cubic b-spline model.	41
3.4	Flowchart for the midpoint assessment scheme. This scheme, at most, requires one replan and one assessment to be performed. Final dose estimates only use information gathered using week four’s aCT image.	45
3.5	Flowchart of the creation of dose estimates using the biweekly assessment scheme. Plan numbers begin at 1 with the original plan. The *’s represent different replans based on when triggers occur. So, Plan 2 is created after a trigger on week two’s assessment. Plan 2*, however, didn’t trigger while week four’s assessment did trigger.	46
3.6	An example confusion matrix. The True positive (TP) represents a situation in which the diagnostic test reads positive and the true result is positive. True negatives (TN) are the same, but for negative results. False positives (FP) arise when the test reads as positive when the true result is negative. False negatives (FN) arise when the test reads as negative but the true result is positive. Sensitivity and specificity calculations are shown.	51
3.7	Confusion matrix displaying how well the midpoint assessment scheme caught necessary replans for Patient 2. All true OAR triggers were caught and one ‘false alarm’ (or false positive) was reported.	53

3.8	Confusion matrix which quantifies the ability of the midpoint assessment to predict final dose for Patient 2. Specificity was omitted due to a lack of TN and FP. Note that the rightmost column represents situations in which replanning ensured that OARs stayed under the plan goals.	57
4.1	Confusion matrix comparing midpoint assessment’s trigger results against ground truth dose values. The numbers indicate the total number of OARs that resulted in that particular outcome.	64
4.2	Confusion matrix comparing biweekly assessment’s trigger results against ground truth dose values. The numbers indicate the total number of OARs that resulted in that particular outcome.	65
4.3	Confusion matrix comparing midpoint assessment’s final dose projections against ground truth final dose projections. The numbers indicate the total number of OARs that resulted in that particular outcome.	67
4.4	Confusion matrix comparing biweekly assessment’s final dose projections against ground truth final dose projections.	68
4.5	Comparison of total end-of-treatment dose constraint violations for the midpoint and biweekly assessment schemes as seen in our three patient cohort.	69
E.1	Permission for Figure 2.1.	103
E.2	Permission for Figure 2.2.	104

Chapter 1

Introduction

Head and neck (H&N) cancer accounts for about one in 25 diagnoses of cancer in the United States [1]. In fact, the world sees about 1.1 million new cases reported per year [2]. Out of these cases, only about 60% of people survive after five years [3]. These levels of survival clearly call for active research into the improvement of treatment outcomes.

Highly specialized treatment plans are created for patients who are prescribed radiation therapy for H&N cancer, and clinicians go to great lengths in order to ensure that radiation is delivered exactly as intended. However, patients undergoing this treatment, typically along with chemotherapy, often lose weight over the course of their treatment. One study estimates that 50% of patients lose over 5% of their body weight during and after this combination of treatment [4]. Along with this weight loss comes shifts in the anatomical layout of a patient's body. Additionally, tumors may shift and shrink in response to treatment. Since treatment plans are created with an exact anatomical shape in mind, these changes may translate to unintended dosimetric consequences for target volumes and healthy tissue. In conventional radiation therapy, these effects are typically accounted for through the use of margins, which are usually not patient-specific, and may lead to higher doses to healthy organs for some patients.

From its inception, adaptive radiotherapy (ART) has attempted to mitigate the issues that arise from changing anatomy throughout a patient’s treatment course. The main idea behind ART is that treatment plans can be adapted throughout a treatment course in response to, or anticipation of, anatomical changes. Many studies have shown the benefits of ART with replanning at set intervals, boasting decreases in mean parotid gland dose and spinal cord max dose ranging from 0.6-6 Gray (Gy) and 0.1-4 Gy, respectively [5]. However, the choice to replan at set intervals may translate to a waste of time if these replans weren’t needed. This highlights a gap in the current body of literature — few works investigate the usefulness of quantitatively assessing whether a patient needs their treatment plan adapted. Moreover, no current work investigates how often these assessments should take place.

In order to determine the usefulness of adaptive assessment and the most efficient frequency (scheme) of adaptive assessment, large-scale multi-institutional studies must be performed. Studies of this size and nature require a simple and easily transferable workflow through which ART can be performed. They also require some framework with which evaluation of these assessment schemes can be performed reproducibly at each institution. Hyun et al. have developed an offline ART workflow (shown in Appendix A) using Varian’s Velocity software for all dose summation and image deformation, which are both critical steps in the ART process. It follows, then, that any institution with access to Velocity can implement this workflow.

A natural next step from this point is to develop an evaluation framework which quantifies the performance of Hyun’s offline ART workflow when used with different frequencies of assessment. This framework should be able to answer a few questions regarding these assessment schemes. First, it should be able to quantify how trustworthy an assessment scheme is in its estimation of delivered dose. When given a certain condition, a useful assessment scheme should reliably determine if the current course of treatment violates this condition. When replans are performed, a useful

assessment scheme should also lead to accurate predictions of a patient's final delivered dose. Second, the evaluation framework should quantify the benefit gained when using a certain assessment scheme. Third, it should be able to quantify the time cost-effectiveness of a given assessment scheme. The main goal of this project is to develop and test one such evaluation framework.

An additional benefit of this evaluation framework is that it creates a clear data set which clinicians can refer to when they are interested in using offline ART. This is particularly useful for low-resource clinics who may not be able to spare many extra work hours. When given a data set which conveys assessment schemes' trustworthiness, benefit, and efficiency, clinicians are equipped to make the best possible choice for their given clinic and scenario. This ties back into the need for large-scale multi-institutional trials. With a clear description of the performance and protocol of a given assessment scheme, those who are leading large-scale clinical trials have the information needed to decide which assessment scheme to use in a given study arm.

This project created one such evaluation framework and used it to quantify the performance of two different assessment schemes. This was done for three separate patients as a proof of concept. Both midpoint and biweekly assessment schemes performed comparably in their ability to detect necessary replans as well as predict final delivered dose to healthy tissue after replanning. Additionally, midpoint and biweekly assessment decreased the total number of organs-at-risk receiving unacceptable dose levels by 50% and 100%, respectively. Lastly, the midpoint assessment scheme was shown to be more time-cost-effective. It is the hope that this work will translate to an improvement of the assessment schemes used and facilitate their use on a larger scale.

The next section will describe the background necessary for understanding the following sections of this study. Descriptions of H&N cancer, radiation therapy, deformable image registration, and the history of adaptive radiotherapy will be given.

The following sections will include the materials & methods, results, and discussion of the study outcomes.

Chapter 2

Background

Adaptive radiotherapy (ART) rests on the shoulders of all the treatment modalities and strategies that precede it. So, in order to understand both ART in general and the content of this work, this chapter provides a baseline description of conventional radiotherapy and its areas of applicability for ART. This includes an understanding of head and neck cancer and relevant anatomy, radiation therapy, deformable image registration, and ART.

2.1 Head & Neck Cancer

Cancers of the head and neck (H&N) make up nearly 4% of all cancer in the United States and most often start in squamous cells [1] [6]. These thin and flat epithelial cells are found in tissue that forms the surface of the skin and lines hollow organs, of which the H&N region has many. Cancers affecting these cells are called squamous cell carcinomas and most often show up in areas such as the oral cavity, pharynx, larynx, nasal cavity, and salivary glands. Figure 2.1 depicts these specific areas of the body [1]. Cancers of the brain, eyes, esophagus, thyroid glands, and skin surrounding the head and neck are not classified as H&N cancer and are given specific classifications of their own.

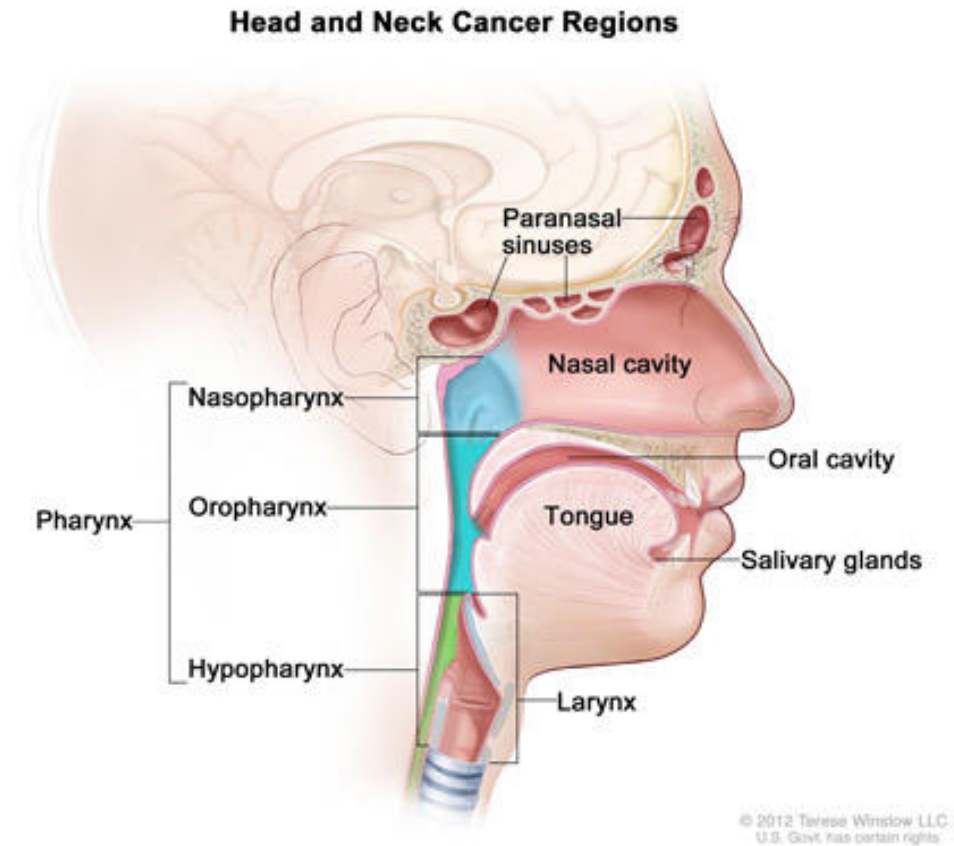


Figure 2.1: Anatomy of the head and neck region. Cancers of the head and neck most often form in the squamous cells of these organs. (2012) Terese Winslow LLC, U.S. Govt. has certain rights.

The most common cause of H&N cancer is the use of alcohol and tobacco [7]. This includes so-called secondhand smoke, as well as chewing tobacco. Other sources of H&N cancer include cancer-causing types of human papillomavirus (HPV), like HPV type 16 [8] [9] [10], occupational exposure to asbestos or formaldehyde [11] [12], and radiation exposure from previous medical or dental images [13]. Symptoms of H&N cancer vary depending on the affected area, but most often manifest as pain, bleeding, swelling, and malfunction in the affected area [7]. Since approximately 1 in every 25 cancer patients is a H&N cancer patient, treatment modalities and quality of care are continuously being researched. But, before any treatment is planned or delivered, accurate diagnosis and staging must take place.

2.1.1 Diagnosis and Staging

The diagnosis and staging steps in a patient's care journey help determine which types of treatment plan will work best for their particular case. Diagnosis can be performed through various tests, but most often begins with a physical exam [14]. The physician may palpate the H&N region and search for abnormalities visually. Additional tests like endoscopies and biopsies may be involved in the physical exam as well. The latter of these involves excising a small amount of suspected cancerous tissue so that it can be further investigated by a pathologist.

Along with the physical exam, diagnostic imaging may take place. These tests, which may include computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET) imaging, provide more tissue detail and determine if, and how far, cancer may have spread[14]. Lastly, tests for HPV are routinely performed since the prognosis for H&N cancer patients drastically improves when the cancer is HPV-induced. In fact, patients with HPV-induced H&N cancer have a disease-free, five-year survival rate of 85-90% compared to a rate of 25-40% otherwise[15].

After physical and imaging exams are completed, a diagnosis with appropriate staging is determined. Staging for H&N cancers can be complicated due to the need for different staging criteria for each kind of H&N cancer. Due to this level of complexity, this chapter will focus on staging of cancers in the oral cavity. This type of disease was chosen because each of the patients investigated in this study was diagnosed with a cancer of the oral cavity.

The standard staging framework is known as the TNM staging system [16]. The TNM system is broken up into 3 different categories, with the first being the T-classification. This describes the size and extent of the patient's primary tumor. The N-classification describes the number of nearby lymph nodes which have cancer. The M-classification describes whether the cancer has metastasized, or spread, to other parts of the body. Staging criteria for specific cancers and affected areas may vary, but

TX	Primary tumor cannot be assessed
Tis	Carcinoma <i>in situ</i>
T1	Tumor ≤ 2 cm and DOI ≤ 5 mm
T2	Tumor ≤ 2 cm, DOI > 5 mm and ≤ 10 mm <i>or</i> tumor > 2 cm and ≤ 4 cm and DOI ≤ 10 mm
T3	Tumor > 4 cm <i>or</i> any tumor with DOI > 10 mm
T4	
T4a	Tumor invades adjacent structures only (e.g., through cortical bone of mandible or maxilla, or involves the maxillary sinus or skin of the face)
T4b	Tumor invades masticator space, pterygoid plates, or skull base and/or encases the internal carotid artery

Figure 2.2: T-classification portion of the staging system for oral cancers as determined by the American Joint Committee on Cancer[17].

the staging system for oral cancers is shown in Figure 2.2. Various combinations of T, N, and M-classifications can be grouped into categories described by the American Joint Committee on Cancer[17]. These AJCC stages, 1-4, often correlate with how dangerous a patient’s disease is, and higher numbers often mean more aggressive treatment options. Additionally, the staging system for HPV-induced H&N cancers is different due to different prognoses and biological characteristics.

2.1.2 Treatment Standard

One of the most common treatments for H&N cancer, whether HPV-induced or otherwise, is chemoradiotherapy [7]. This treatment method uses a chemotherapy drug called cisplatin, delivered intravenously every few weeks, in conjunction with daily radiotherapy. Radiation treatment doses of 63-70 Gray (Gy) are most often prescribed [18]. These doses are delivered at a rate of around 2 Gy/day over the course of 6-7 weeks, in 30-35 treatments or "fractions". It is also important to note that patients receiving treatment for H&N cancer commonly lose weight throughout their treatment course. One study estimates that 50% of patients lose over 5% of their body weight during and after this combination of treatment [4]. This weight

loss may contribute to tumor volumes and healthy organs shifting in position relative to other anatomical structures.

As this study focuses on the topic of radiation therapy and the potential benefits of adaptive radiotherapy, the next section discusses the radiation therapy workflow in more detail. This is done to understand how changes to this process, such as treatment adaptation, help mitigate problems which commonly arise during radiation therapy of the head and neck.

2.2 Radiation Therapy of the Head and Neck

Radiation therapy is one of the most common forms of treatment for cancer [19]. It is estimated that 75% of patients with some form of H&N squamous cell carcinoma will benefit from radiation therapy [20]. External beam radiotherapy, which we will focus on for this project, uses a machine called a linear accelerator (LINAC), which is shown in Figure 2.3. These machines create ionizing radiation in the form of photons and electrons at megavoltage (MV) accelerating potentials. These beams of ionizing radiation are directed towards cancerous volumes in patients' bodies, which can result in both curative and palliative effects for patients.

These radiation beams are aimed so that they deliver the prescribed dose to a specified volume, the planning target volume (PTV), while trying to avoid high dose levels in healthy tissue. The PTV, and other relevant volumes, are defined by recommendations given by the International Commission on Radiation Units (ICRU) in their 50th report [21].

Figure 2.4 gives a visual depiction of this volume classification. The gross tumor volume (GTV), is defined as the visible or demonstrable extent and location of a tumor. This is typically determined through investigation of the images which were taken during the diagnosis process, like PET and MR images. The clinical target

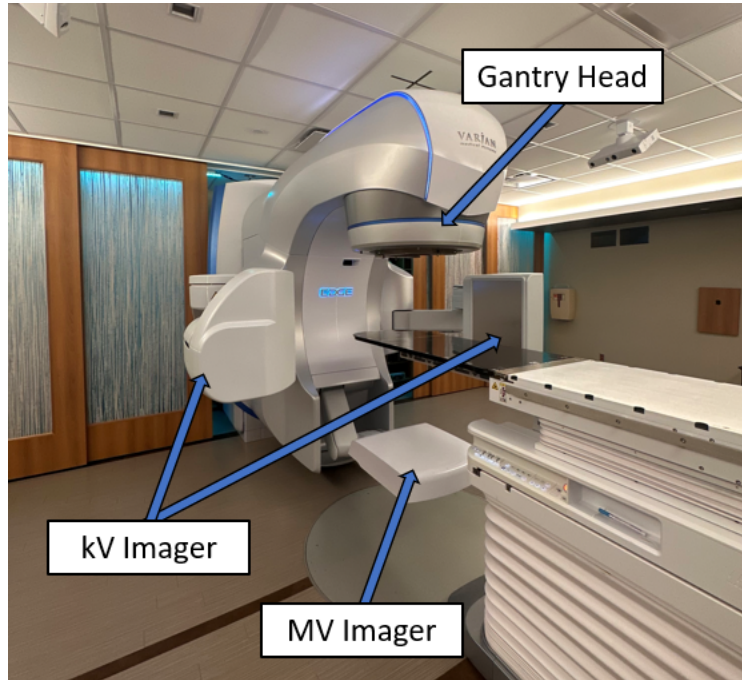


Figure 2.3: A Varian Edge linear accelerator at the University of Nebraska Medical Center (UNMC). Radiation is delivered via the gantry head to the patient positioned on the treatment table. The kV imager components are used to create cone-beam CT (CBCT) images which are used in this study. The MV imager can be used to capture a beam's-eye-view of the radiation field.

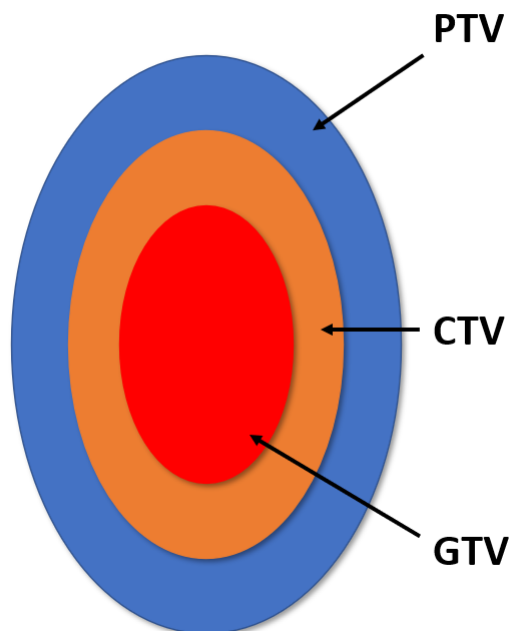


Figure 2.4: Volume classification as defined by ICRU Report 50.

volume (CTV) is a volume surrounding the GTV which is meant to account for the presence of microscopic disease around the visible tumor volume. The CTV is created as an expansion about the GTV. Then, the planning target volume (PTV) is yet another expansion about the CTV. This second expansion accounts for both movement of the tumor within the patient as well as mechanical limitations of the LINAC. Movement of the patient during treatment delivery (intrafractional) or the patient's body changing between fractions (interfractional) are both considered when creating a PTV.

Healthy tissues which may receive radiation dose are called organs at risk (OARs). Common OARs for H&N cancer treatment are parotid glands, the esophagus, pharynx, larynx, spinal cord, and other nearby organs due to their proximity to the treated area. Avoiding dose to healthy tissue is an incredibly important aspect of the treatment process, and satisfactory OAR dose sparing can often be achieved when the radiation therapy workflow is followed carefully.

2.2.1 Radiation Therapy Workflow

After diagnosis and staging of a patient's disease, if the prescribed treatment course includes radiation, the radiation therapy workflow as shown in Figure 2.5 begins with a simulation appointment. During this appointment, computed tomography (CT) images are taken of the patient while they are in their treatment position. This treatment position is commonly enforced through the use of various immobilization devices. Thermoplastic masks, which are fit directly to the patients face, neck and shoulder areas, along with bite blocks, were used for the patients in this study. Once this is completed, the treatment team uses this information to create a treatment plan which determines exactly how the LINAC will deliver radiation dose to the patient's PTV. This involves the use of a treatment planning system (TPS) like Varian's Eclipse, which allows the team to outline, or contour, important anatomical struc-

tures in the patient’s body. Then, the team can use the TPS to determine specifics of the treatment like the positioning of the LINAC, beam energies, and how to utilize beam shaping with multi-leaf collimators (MLCs) [22]. When done correctly, the treatment plan lays the groundwork for how the LINAC can deliver the prescribed radiation dose to the patient’s PTV.

When this treatment plan has undergone internal review, the patient comes in for treatment appointments. The treatment team aligns the patient in the LINAC vault and tries to closely mimic the patient’s setup from the simulation appointment. This is usually done through imaging at the treatment machine like a cone-beam CT (CBCT), typically using the kV imagers as seen in Figure 2.3. A CBCT image is taken and used for the purposes of image-guidance and alignment at the treatment machine. CBCT imaging makes use of a cone-shaped field and 2-dimensional detector which can reconstruct the patient volume within its field of view (FOV) after a single rotation.

After patient alignment is completed, dose is delivered. A patient may undergo different numbers of appointments, or fractions, depending on the type and severity of cancer and expected OAR dose, but H&N cancer patients typically receive their prescribed dose over 35 fractions. It is also important to note that alignment im-

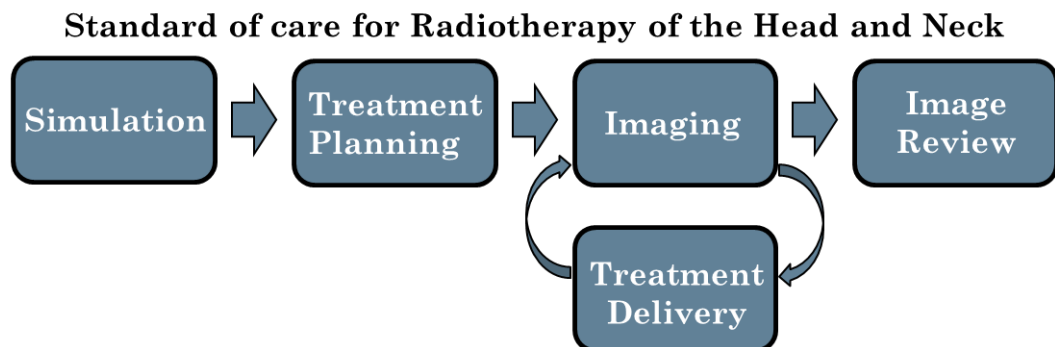


Figure 2.5: Standard of Care for external beam radiation therapy of the head and neck. Imaging and treatment delivery often take place over the course of 35 fractions for a H&N patient.

ages are routinely reviewed by physicians and medical physicists on the treatment team. This is done to ensure that the patient's body is positioned correctly and looks sufficiently similar to the patient's body on the day of simulation.

This workflow is the standard in the field of radiation therapy and can be used for multiple modalities of radiation therapy. Treatment types like 3D-conformal radiation therapy and simpler techniques which use static open-fields follow this workflow, but may fall short in terms of dose conformity and OAR dose in more complicated cases. These cases, often made difficult by the presence of many OARs, may necessitate the use of higher levels of treatment customization. This customization can be accomplished through more sophisticated radiation therapy techniques like intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT).

2.2.2 Volumetric Modulated Arc Therapy (VMAT)

IMRT and Volumetric Modulated Arc Therapy (VMAT) are advanced forms of external beam radiotherapy that have become standard of care for many treatment sites. IMRT boasts the capability to vary the intensity of radiation beams as they are being delivered. This is done through the use of dynamic MLCs which shape the field throughout treatment delivery. VMAT, then, takes the benefits of intensity modulated radiotherapy (IMRT) and elevates them. First, like IMRT, it has the ability to modulate the intensity of the treatment beam in time. It can also rotate the gantry along an arc while the treatment is ongoing. On top of this, it is able to vary the dose rate of the machine in addition to its MLC modulation of the treatment field [22]. When VMAT treatment is ongoing, the gantry is rotating continuously, and both MLCs and dose rate vary as the gantry traverses its predetermined arc. This allows for high levels of both efficiency and dose conformity. All patients in this study

were treated with VMAT due to the presence of many OARs and the need for high levels of dose conformity.

VMAT also commonly benefits from the use of image-guidance like CBCT imaging at the treatment machine. But, in order for these images to be useful in any way, they need to be spatially linked, or correlated, to the images used during treatment planning. Linking these images allows for information seen in a CBCT to inform any shifts that need to be made to ensure the patient is positioned properly. This link is created through a process called image registration.

2.3 Deformable Image Registration

As established in the previous section, image registration is a crucial part of the radiation therapy treatment process. Registration allows for the creation of shared coordinate spaces between images taken at different points in time or by different imaging techniques, meaning that daily alignment images like CBCTs can give clinicians information on necessary shifts at the treatment machine. This section will describe the basics of image registration with a focus on deformable image registration, as it is relevant to the field of adaptive radiotherapy.

To begin defining image registration, one can consider two sets of images [23]. The first is unmoving, or fixed, and is defined as $F(\vec{x})$, with \vec{x} being the 3-dimensional coordinate vector of a given voxel. The second is a moving set defined as $M(\vec{x}')$, with \vec{x}' being the 3-dimensional coordinate vector of voxels in the moving image. Image registration attempts to find the best transform $T(\vec{x}')$, operating on the coordinate vector of the moving image, that minimizes the difference between $F(\vec{x})$ and $M(T(\vec{x}'))$. The transformation can be thought of as the addition of the local position vector \vec{x}' of the moving image and a displacement vector $\vec{u}(\vec{x}')$. So, in the ideal case, an image registration operation could be defined as

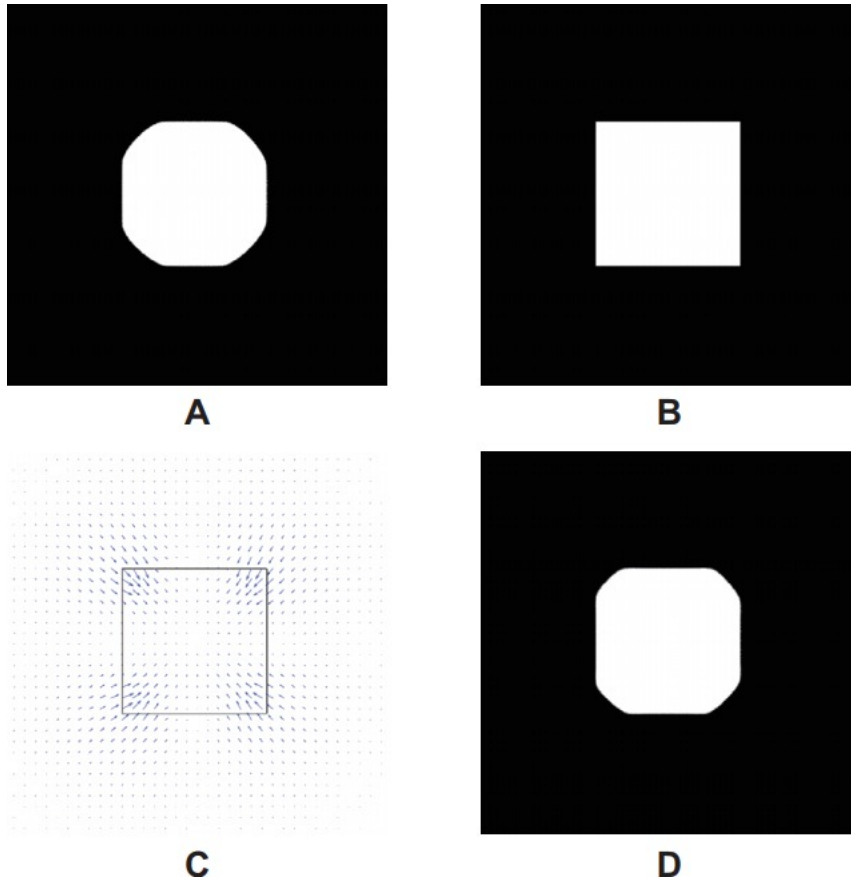


Figure 2.6: Fixed image (A) and moving image (B) were registered together using DIR algorithms. (C) shows the deformation vector field which was applied to the moving image in order for it to accurately deform to the shape of the fixed image. (D) displays the deformed moving image, which matches the fixed image exactly. Image taken from open access article [23].

$$F(\vec{x}) = M(T(\vec{x}')) = M(\vec{x} + \vec{u}(\vec{x}')), \quad (2.1)$$

where the two images share a coordinate space after registration. It is important to note that the displacement vector $\vec{u}(\vec{x}')$ acts on a single voxel, but the collection of all displacement vectors constitutes a deformation vector field (DVF) which describes the correspondance between fixed and moving images. Figure 2.6 shows an example of a deformation vector field which was created to deformably register the image in (B) to the image in (A).

For rigid registration, the transformation takes place through translation and rotation, which affords the transformation 6 degrees of freedom (DOF). This ensures that all voxel-to-voxel relationships are maintained. This is no longer the case with non-rigid (deformable) image registration (DIR). The voxel-to-voxel relationships can change, and pixels are often morphed such that the voxel volume is not constant. In DIR, the number of DOFs increases dramatically along with the necessary computational resources. This kind of registration is more useful in situations where weight loss, tumor shrinkage, and OAR (organs at risk) movement are a factor, or when the patient position is different in one image compared to another. When explaining DIR algorithms, it is useful to consider their three main components: the objective function, transformation model, and optimization algorithm. This section will describe each of these three components of a DIR algorithm.

2.3.1 Objective Function

An objective function can be thought of as a metric which quantifies the results of a given transformation or provides information on the two images' level of similarity. This objective function is either maximized or minimized through iterative processes, depending on the type of objective function used. In other words, DIR algorithms solve an optimization problem for an objective function. Most commonly, one will see objective functions based in intensity, image features, or a combination of the two [23].

Intensity-based objective functions operate under the assumption that intensity values of anatomical areas between image sets are similar. For this reason, intensity-based objective functions are typically used for single image sets obtained through a single modality. One of the most commonly used objective functions or registration metrics is mutual information (MI). This metric is used by Velocity, which is one of the main tools used in this project. This image registration technique can be thought of

as a way to maximize the amount of shared information between two images. Claude Shannon, the founder of information theory, defines information quantities through a metric heavily inspired by thermodynamic physics [24]. But, in order to understand Shannon’s information entropy, the idea of information must be established and its connection with probability solidified.

The information provided by an ‘event’ or, in this case, image voxel, is directly related to the degree to which its content is surprising. In other words, events with a high probability of occurring provide low levels of information, while highly unlikely events provide high levels of information. Specifically, the information content, or self-information (SI), can be defined as

$$SI = \log \left(\frac{1}{p(E)} \right), \tag{2.2}$$

where $p(E)$ is the probability of a given event occurring. One can see that the SI of an event with a 100% probability of occurrence is zero, meaning that no information is provided. It also immediately follows that the SI can be expressed in the more familiar form

$$SI = -\log(p(E)). \tag{2.3}$$

Entropy, then, quantifies the average amount of information gained when identifying the outcome of an event, or, intensity of a given image voxel. The entropy H of the fixed image’s collective voxel intensity, for example, can be expressed through the equation

$$H(Int.) = \mathbb{E}(SI(Int.)), \tag{2.4}$$

where \mathbb{E} is the expectation value operator. In full, then, the entropy of voxel intensity can be defined as

$$H(Int) = - \sum_{i \in \mathbb{X}} p(Int_i) \log(p(Int_i)), \quad (2.5)$$

where \mathbb{X} is the range of possible voxel intensities and Int_i is the i^{th} possible value of intensity. For reference, typical intensities in CT images are mapped to 12-bit gray scale, meaning values range from -1000 to 3095.

As a simple example, consider the information gained from observing both fair and unfair coin flips, and a fair dice roll. The fair coin flip has two possible outcomes, each with equal probability $p = \frac{1}{2}$. This means that observing an outcome of heads would provide

$$SI = -\log(p(E)) = -\log\left(\frac{1}{2}\right) = 1bit, \quad (2.6)$$

if a logarithmic base of 2 is used. This means that one bit of information was provided. It follows, then, that the entropy of an unknown event result is

$$H(X) = - \sum_{i=1}^2 p(x_i) \log_2(p(x_i)) = - \sum_{i=1}^2 \frac{1}{2} \log_2\left(\frac{1}{2}\right) = - \sum_{i=1}^2 \frac{1}{2} (-1) = 1. \quad (2.7)$$

with X representing the discrete variable given by the flip's outcome.

When considering an unfair coin flip, say, with a probability of 20% head and 80% tails, the entropy (or average information gain from a single flip) would be

$$H(X) = - \sum_{i=1}^2 p(x_i) \log(p(x_i)) = -(0.2 \times \log(0.2) + 0.8 \times \log(0.8)) = 0.72bits. \quad (2.8)$$

One can conclude that unfair coins, or uneven probability distributions, return lower entropy than uniform probability distributions.

If one were to perform the same calculations with a fair die (equal probability $p = \frac{1}{6}$), one would find answers of $SI = H(X) = 2.58$ for a single die roll. It follows that uniform distributions with more possible outcomes increase the entropy of a given event outcome.

A final concept important to understanding the definition of mutual information is the idea of joint entropy. Joint entropy $H(X, Y)$ can be described as a measure of uncertainty associated with the set of random variables X and Y . This value is given by the equation

$$H(X, Y) = - \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} P(x, y) \log(P(x, y)), \quad (2.9)$$

where $P(x, y)$ is the joint probability of the specific outcomes x and y , out of probability spaces \mathbb{X} and \mathbb{Y} , both occurring.

Given these established quantities, one can define mutual information (MI) conceptually as mutual dependence between two variables. In other words, MI describes the amount of information gained about one variable when the other variable is known. MI is given by the equation

$$MI(X; Y) = H(X) + H(Y) - H(X, Y), \quad (2.10)$$

where $H(X)$ and $H(Y)$ are the familiar entropies of the random variable sets X and Y and $H(X, Y)$ is the joint entropy of the two. This equation can equivalently be represented as

$$MI(X; Y) = \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} P(x, y) \log\left(\frac{P(x, y)}{P(x)P(y)}\right), \quad (2.11)$$

where $P(x, y)$ is the joint probability of the two instances of the variable sets X and Y , and $P(x)$ and $P(y)$ are the probabilities of x and y outcomes of the variables X

and Y . This equation, using variables from earlier in this section (like those used in Eq. 2.1), is expressed as

$$MI(F, M') = H(F) + H(M') - H(F, M'), \quad (2.12)$$

and

$$MI(F, M') = \sum_{x \in \Omega} \sum_{x' \in \Omega} P_{FM'}(x, x') \log\left(\frac{P_{FM'}(x, x')}{P_F(x)P_{M'}(x')}\right), \quad (2.13)$$

where F and M' represent the fixed and transformed moving image and Ω is the area of overlap between the two images.

Conceptually, in the case of two identical images being registered, their MI would be at the highest possible value. Equation 2.10 shows that this is the case when the joint entropy is minimized. Then, looking at Equation 2.9, this value reaches zero as the joint probability approaches 1. This means that joint entropy is minimized, and MI is maximized, when knowledge of one variable guarantees knowledge of the other. In other words, MI is maximized when the images are identical. For this reason, mutual information serves as a helpful objective function for DIR algorithms. For an objective function such as mutual information to be used, though, there needs to be some form of transformation of the moving image taking place.

2.3.2 Transformation Model

The next necessary component of a DIR algorithm is a transformation model. Transformation models can be thought of as the method through which the similarity index (i.e., mutual information) is maximized in a DIR algorithm. Transformation models fall under two main categories: parametric and non-parametric. One can think of a parametric transformation as one that can be modeled using a simple equation such as

$$f(x, y) = (x', y') = (x, y) + t(x, y), \quad (2.14)$$

where x and y are coordinates in the moving image, x' and y' are coordinates in the fixed image, and t is a parameter to be determined. This particular example is that of a single translation. So, the transformation can be described through specifying the single translation parameter ' t '. In other words, a parametric transformation can be expressed as a linear combination of basis functions. The coefficients of these basis functions determine the shape of the resulting deformation vector field (DVF), which depicts the transformation from moving image to fixed image. It is also important to note that the number of fitting parameters for a parametric transformation model depends on the specific model being used. Non-parametric models, on the other hand, cannot be expressed as a linear combination of basis functions. This means that each image element (i.e., voxel) needs its transformation to be specifically calculated, which drastically increases the computational resources required for a transformation model.

This section will focus mainly on parametric transformation models, since these are more relevant to the work at hand. One such parametric model, which is used in Velocity's image registration algorithm [25], is the cubic b-spline model. This transformation model describes the moving image's deformation as a linear combination of b-splines, or basis splines [23]. This means that a rectangular grid of the spline functions' control points of dimension (C_x, C_y, C_z) is laid over the moving image of dimension (N_x, N_y, N_z) , and voxels lying on control points have their deformation vectors directly calculated. This deformation vector field is expressed as $\vec{\phi}$. Voxels that do not lie directly on control points are deformed according to an interpolation of the deformation vector field $\vec{\phi}$ of the control points. Then, the deformation vector of a position vector \vec{x}' , $\vec{u}(\vec{x}')$, after k iterations, can be expressed as

$$U_k(\vec{x}') = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u)B_m(v)B_n(w)\vec{\phi}(i+l, j+m, k+n)_k, \quad (2.15)$$

where u , v , and w are the cardinal distances from the position vector \vec{x}' to the nearest control point (i, j, k) , and B_l represents the l^{th} basis of the spline function. More accurate DVFs, then, could be created by increasing the number of control points at the cost of additional computation time. Most parametric transformation models begin with a rigid registration as well. This is done since using rigid registration prior to deformation requires less computational resources as opposed to using only DIR.

An additional consideration when discussing transformation models is the type of regularization function used. A regularization function essentially imposes constraints on the types of deformations created by a deformation vector field in order to ensure the result is physically meaningful. The two main types of regularization function that are used in DIR are constrained and unconstrained. Figure 2.7 shows a simplified version of these two regularization functions. In constrained regularization functions, points in space are prevented from moving independently of their neighbors. This forces smooth DVFs to form and can simulate tissue deformation more accurately. When using unconstrained regularization functions, individual points move completely independently of one another, which may lead to points deforming over one another. This becomes an issue because DIR's main goal, in this context, is the simulation of real tissue deformation. For this reason, Velocity utilizes constrained regularization.

2.3.3 Optimization Algorithm

The last part of a DIR algorithm is the optimization algorithm. This portion of the algorithm is often proprietary, so the specifics of Velocity's optimization algorithm are not available to the public. Despite this, it is known that optimization algorithms

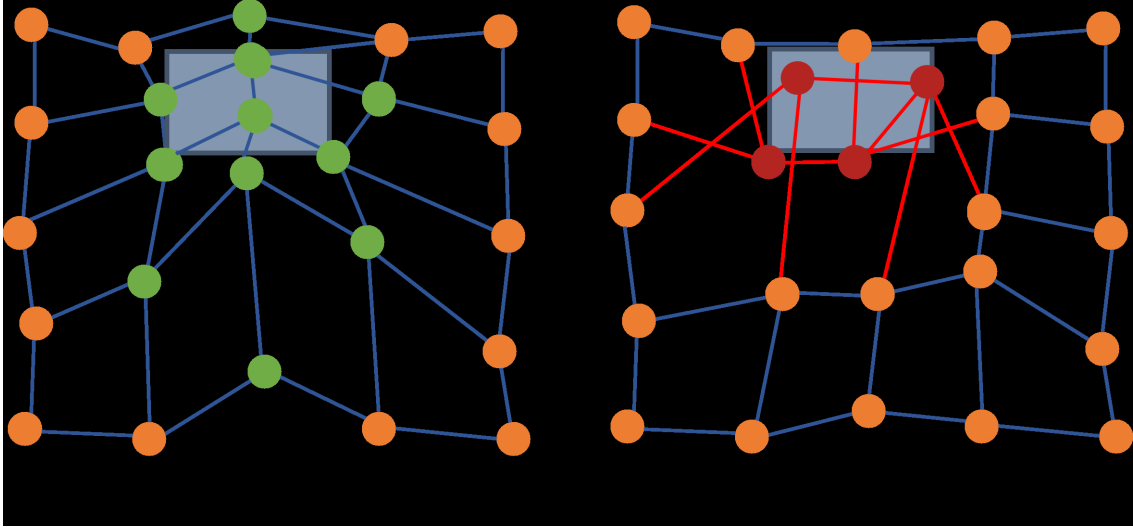


Figure 2.7: (A) shows a simplified example of deformation using constrained regularization. Neighboring voxels move congruently with no overlap. (B) shows a simplified example of deformation with unconstrained regularization. Voxels are able to move freely from one another, which can lead to situations in which voxels overlap and create folds in the voxel grid. From Varian [26].

attempt to find global maxima in the DIR algorithm’s similarity index, commonly mutual information. This is most commonly done through an iterative process which is seen in Figure 2.8. The process begins with one fixed and one moving image. The similarity index between these two images is then calculated. Based on this value, along with information from each individual image, the optimization algorithm tries to find a global maximum in the similarity index through variation of the transformation model’s parameters. After this is completed, the transformation model creates a deformation vector field, which is applied to the moving image. The similarity index between the fixed image and the newly deformed moving image is then calculated. Based on this new value and new image information, the optimization algorithm is run again. This cycle continues iteratively until some form of termination condition is met. This condition is most often a threshold value in the similarity index between the two images.

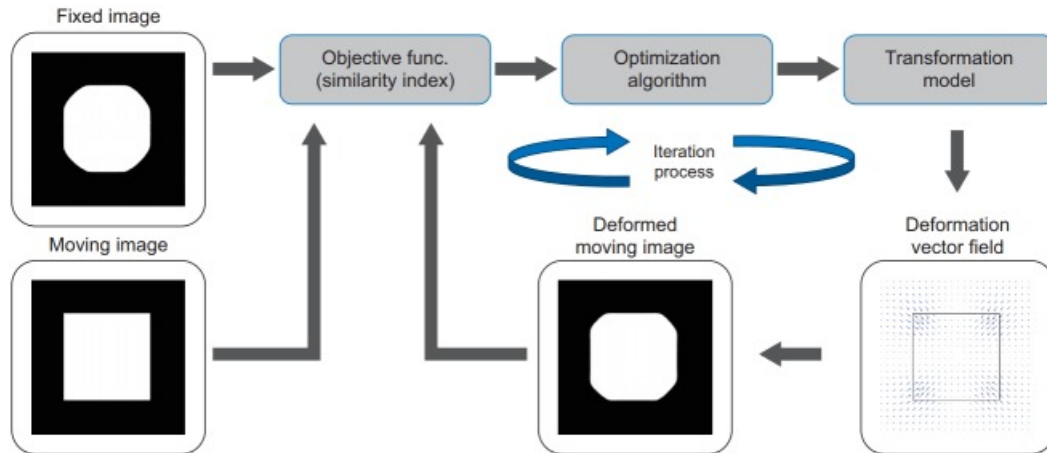


Figure 2.8: A flowchart for the deformable image registration process. Steps including the calculation of a similarity index, running of an optimization algorithm, and application of a transformation model to the moving image, are performed iteratively until a satisfactory similarity index value is met. Image taken from open access article [23].

2.3.4 DIR Quality Assurance

Once the optimization algorithm maximizes the desired similarity index, it is important that quality assurance (QA) procedures are performed. These are an incredibly important part of the DIR process as they confirm that the final product is reliable and high-quality. The current standards for QA procedures are listed in the American Association of Physicists in Medicine’s (AAPM) Task Group 132 (TG-132) report [27]. This report outlines how to mitigate and catch errors which may arise from data acquisition or the registration process itself. Data acquisition errors, like incorrect scan parameters or low initial image quality, can decrease the final accuracy of registration between images from the very start. Image data can be very highly varied, so one needs to evaluate both the appearance of registered images and the accuracy of the registration itself.

Quick tests, like visual inspection, are a good first step in the QA process. Common checks for visual inspection are unrealistic looking organs, mismatch between deformed and primary image, and other similar artifacts. As is the case with most

QA procedures, though, the specifics of implementation are largely left up to individual clinics. Tests like checking for negative Jacobian determinants are a helpful indicator of nonphysical transformations showing up in a registration process, according to TG-132 [27]. In short, a Jacobian determinant gives information on the local transformation experienced by a voxel in a moving image along with information on the relative change in its volume. Assuming the transformation is differentiable, its associated Jacobian determinant can take on any real value. Any determinant with a value greater than one indicates local volume increase, while determinants ranging from (0,1) indicate a local volume decrease. Jacobian determinants with negative values, however, indicate a coordinate inversion, which translates to image voxels occupying the same space or folding over each other. For the purposes of medical image registration, negative Jacobian determinants are generally undesirable. There may be small amounts of permissible negative values in areas of the body containing fluids and varying fill-levels, but high amounts of negative Jacobian determinants may indicate a poor transformation.

Another useful QA procedure is quantifying high-density voxel elasticity. Voxels within a body contour with Hounsfield units of 800-3000, which is the common range for bony anatomy, can be tested for their elasticity. Intuitively, these values should be quite low. It is known that elasticity is inversely proportional to rigidity, which means that an elasticity of zero indicates a completely rigid transform with no volume change. Velocity, the software used for this study, calculates elasticity and scales such that a value of 1 indicates 9% elasticity by voxel volume. They assert that any values above this number should be investigated manually. This means that inspection of problem areas in the registration should be performed in order to determine whether these areas will reasonably affect dose estimations. For instance, if the poor registration areas are near the shoulders, they are sufficiently far from the OARs which are being investigated.

In addition to QA being performed whenever the registration process is performed, QA is also done when software is installed, upgraded, as well as annually. TG-132 guidelines are followed to evaluate both rigid and DIR algorithms for accuracy and constancy using geometric and anthropomorphic digital phantoms provided by the task group. This process was performed at the University of Nebraska Medical Center (UNMC) and is done so annually.

2.3.5 Summary

The process of DIR is an integral part of adaptive radiation therapy treatments. DIR can be broadly broken up into three components: its objective function, transformation model, and optimization algorithm. Varian's Velocity software, which is an integral part of this document's work, uses a cubic b-spline transformation model whose objective function is mutual information.

DIR can be used for multiple tasks ranging from the fusion of PET and CT images, which aids in diagnosis of disease, to the tracking of anatomic and dosimetric changes of a patient over their treatment course. This latter capability proves incredibly informative when analyzing patients receiving radiation therapy for H&N cancer. If the anatomic changes a patient's body is undergoing are monitored by the treatment team, situations in which anatomic changes may impact treatment outcomes can be caught. For example, there may be situations in which a patient's weight loss causes their parotid glands to move into a high dose portion of the treatment field. When these situations arise, it may be useful to recreate the patient's treatment plan so that dose is delivered using more accurate information on the patient's anatomy. When formalized, treatment strategies like this fall under the umbrella of adaptive radiotherapy.

2.4 Adaptive Radiotherapy (ART)

As has been established in Section 2.2.1, radiation therapy is a fundamental part of modern cancer care. As medical physicists, much of our work is performed such that cancer treatment by radiation is performed as accurately and efficiently as possible. It is towards this end that medical physicists, alongside radiation oncologists and dosimetrists, create treatment plans which are specific to each individual patient. Depending on the specific mode of treatment, H&N plans are commonly administered over 6-7 weeks [28].

What is often underestimated is that patient anatomy is dynamic over the 6-7 weeks of treatment. Sometimes it is so dynamic that treatment margins can't fully account for anatomical changes. For instance, Raghaven et al. report that patients with head & neck squamous cell carcinoma (HNSCC) and having received IMRT treatment experienced significant changes in their GTV's size, citing volume decreases ranging from 29.5%-72% [29]. If a tumor shrinks significantly, the result may be unnecessary irradiation of healthy tissue where the tumor used to be. If a tumor shifts spatially, the result may be undercoverage of the tumor and more unnecessary irradiation of healthy tissue.

Weight loss, tumor shrinkage, and systematic shifts in the location and shape of organs-at-risk (OARs) can all impact the effectiveness of a specific treatment plan. Depending on the magnitude of the discrepancy between initial anatomy and current patient anatomy, this can defeat the purpose of patient-specific treatment plans. It would then be advantageous to change treatment plans as the patient's anatomy changes.

Adaptive radiotherapy (ART) is a technique that allows for the adaptation of a treatment plan based on observed or expected changes in patient anatomy. This technique promises a more precise and patient-specific treatment by changing a treatment plan either at the treatment machine or between fractions when a threshold in a spec-

ified metric is exceeded. Although adapted plans deliver radiation based on “day-of” or updated anatomy, the process is often costly in staff time and computation power. A time-cost-effectiveness analysis may help determine how the most benefit can be achieved with the least increase in physicist time.

Head and neck ART began with work by Yan et al. in 1997 [30]. They used electronic portal imaging device (EPID) imaging on 3 axes to account for set-up error and proposed that replanning should take place when “sufficient” deviation from planned dose was observed. The development of kilovoltage (kV) CBCT, megavoltage (MV) CBCT and CT-on-rails allowed for the tracking of both isocentric error (like Yan) and multipoint/multi-ROI (region of interest) displacement alongside volumetric alteration in soft tissue [31]. These play a major role in image-guided radiotherapy (IGRT) which provides insight into the magnitude of any anatomical changes. Replanning based on changes seen during daily imaging fits into Yan’s definition of ART quite well. According to Yan et al., ART, “customizes each patients’ treatment plan to patient-specific variation by evaluation and characterizing the systematic and random variations through image feedback and including them in adaptive replanning.” This level of ART was meant to keep PTV and OAR dose levels equal to the levels in the original treatment plan and treated any additional OAR sparing as an added benefit.

Modern practice views OAR sparing as the first purpose of ART. Large-scale clinical data for ART is sparse, but Castelli et al. give a review of some of the studies that have taken place for H&N cancer [5]. These studies tend to focus on ART strategies involving replans which happen at predetermined time or prescription dose points with no assessment to determine the necessity of a replan. Additionally, most studies decided to replan once or twice, with one exception being a study with weekly replans. Each study consistently reports the benefits of ART for maintaining dose

to tumor volumes. These studies clearly show that replanning, even if only done a single time, consistently improves tumor volume coverage and OAR sparing.

Many studies also investigate optimal timing and frequency of adaptive replanning but lack large-scale clinical data. Castelli et al. provide an overview of these studies that consistently show the benefits of ART for organs at risk [5]. For instance, Dewan et al., in their prospective study, report consistent decrease to both spinal cord and parotid gland dose for patients with locally advanced H&N cancer when replanning at fraction 20 of 35 [32]. They do note, however, that despite these many isolated studies, there is no international guideline stating how or when to apply ART. These studies range from 10-33 patients and replan between 1-6 times. Half of the listed studies in [5] used deformable image registration (DIR), the method of choice for this work, to perform their dosimetric analysis while the other half used alternative methods. It is important to note that the low patient numbers in these studies could be bolstered through multi-institutional trials.

One large-scale, multi-institutional, prospective clinical trial, the ARTFORCE trial, tests conventional radiotherapy (RT) versus what they call "redistributed" radiotherapy, as well as two different chemotherapy drugs [33]. The redistributed RT focuses on creating a GTV based on fluorodeoxyglucose (FDG) uptake as seen on a positron emission tomography (PET) image. This was done in order to decrease dose to healthy tissue. Towards that same end, they also re-image the patient using their simulation CT scanner during their second week of treatment. Using the new CT image, a new treatment plan is created, which will be delivered for the rest of the treatment course. This is another example of ART implementation with a predetermined replan schedule as opposed to predetermined assessment of a replan's necessity. Results for this study are not available as the trial is still ongoing.

There was also a survey showing that, out of 32 institutions in the Tata Memorial Hospital Radiotherapy Practicum, 92% listed head and neck as a site at which they

implemented ART. This large number was balanced out by the fact that hindrances like lack of equipment (48%), training (36%), and tools/management support (26%) prevent many institutions from implementing ART to the level which they would like [34]. This demonstrates that ART is, by nature, resource-heavy and time-consuming.

Overall, the existing literature suggests benefits of ART for both target volume coverage and OAR dose sparing. It's clear that the act of replanning, through the use of updated anatomical information, creates plans which are more highly tailored to the patient's current body. Specifics of whether a replan is warranted, however, does not consistently show up in the current body of work. In other words, existing studies do not often discuss how to determine whether a replan needs to happen in the first place. Without this information, existing ART methods are susceptible to replanning in situations which may not warrant a replan. For larger clinics with ample resources, this may not be an issue. For smaller clinics, however, this use of limited resources may disproportionately affect the workload of the treatment team and translate to no clinical benefit.

Additionally, research into how often assessment of a replan's necessity should take place is sparse. This differs from questions answered in earlier work, like that mentioned in Castelli's review, because those studies investigate how frequently replans should take place. The yet unanswered question, however, is how frequently assessment of a replan's necessity should take place if unnecessary replans are to be avoided. Again, this distinction proves important for low resource clinics. More frequent assessment translates to increased workload, and likely better results in terms of PTV coverage and OAR dose sparing. But, the shift from predetermined replan intervals to predetermined assessment intervals may help these clinics avoid unnecessary replan workload.

In order for the shift from predetermined replan intervals to predetermined assessment intervals to take place, there needs to be some way to define assessment criteria

and analyze the performance of different assessment intervals (schemes). This is necessary because clinics need to be confident when determining which assessment scheme works best for their needs. There also needs to be a straightforward ART workflow, usable across institutions, which allows for higher cohort numbers and statistically powerful conclusions on the varying schemes' benefits. The goal of this current study was to utilize one such workflow, developed by Hyun et al., and develop a set of performance metrics to quantify the success of different assessment schemes. This work demonstrates the feasibility of the developed performance metrics and the use of two different assessment schemes for inter-fractional ART for H&N cancers. This work also presents a time-cost-effectiveness analysis of the schemes' performances that low-resource clinics could use in their decision-making process.

Chapter 3 describes the specific structure and research questions of the current study, all relevant software used, as well as the methods used to answer these research questions. Following this, Chapter 4 lays out the resulting data, including discussion.

Chapter 3

Materials & Methods

In Chapter 2, the importance and potential benefits of adaptive radiotherapy (ART) were established. The current body of literature clearly points towards plan adaptation improving target volume coverage and organ-at-risk (OAR) sparing, especially in the case of H&N cancer patients [5]. Despite the number of studies involving ART, though, there are no large-cohort, multi-institutional, clinical studies. There is also the open question of which assessment scheme performs most efficiently. This may be the case due to the difficulty of interpreting results from multiple institutions which may be using completely separate clinical workflows. This highlights the importance of simple workflows which can be easily implemented in different clinics.

When investigating different assessment frequencies (schemes), it is also important to develop performance metrics that provide valuable information to clinicians who may be interested in implementing ART. A reliable assessment method should be able to accurately determine when an adaptive replan is warranted, create an accurate estimation of the actual dose delivered to patients, and improve OAR dose sparing while maintaining target volume coverage. Any metrics created to evaluate an assessment scheme should be able to convey this information.

The present study uses a simple workflow, as described in Appendix A, to retroactively implement adaptive assessments for three H&N cancer patients at two different frequencies: once at the midpoint of treatment and once every two weeks. These assessment frequencies, dubbed the midpoint assessment scheme and the bi-weekly assessment scheme, are then evaluated with three novel performance metrics. This study displays the ease with which Hyun et al.’s workflow can be implemented, and demonstrates the feasibility of the assessment schemes and performance metrics for larger-scale studies. An additional goal of the study was to compile the performance results of these assessment schemes and compare them against their estimated time/work requirements in a time-cost-effectiveness analysis. This dataset, then, can aid clinicians in low-resource clinics in determining which ART assessment scheme best fits their needs.

3.1 Study Structure

This section outlines how data were gathered, describes the patients selected for this study, and lays out the study’s research questions. Additional specifics of the study’s methodology are described in greater detail in subsequent sections.

3.1.1 Data Gathering

All data was gathered from the database used by UNMC’s radiation oncology department. This database is contained in Aria, which is the department’s Record & Verify system. This study was approved by UNMC’s Institutional Review Board (IRB) under Protocol #842-18-EP.

3.1.2 Patient Selection

Three patients were selected for this study. Each patient received VMAT treatment at the University of Nebraska Medical Center some time in 2020-2021 for a cancer of the oral cavity. The staging of the three patients, according to the TNM staging system, were T2N1M0, T2N1M0, and T2N0M0. Patients were selected because of noticeable anatomical changes seen on their daily CBCT images. These changes included tumor shrinkage, weight loss, and positional variation of OARs like the spinal cord. Ages of the patients included in this study were 57, 57, and 74 years old.

3.1.3 Research Questions

This study aims to quantify the performance of two assessment schemes, midpoint and biweekly, through the development and use of three novel performance metrics. The specifics of each performance metric, as well as assessment scheme, are described in subsequent sections. Answering the following questions describes how much the schemes' estimations can be trusted, the clinical benefit of using these assessment schemes, and the relationship between the work required for each scheme and its clinical benefit.

Research Question 1: Do the midpoint and biweekly assessment schemes accurately predict replan necessity and final delivered dose when compared to a retrospective best estimate?

Research Question 2: Does the use of the midpoint assessment scheme or biweekly assessment scheme for ART prevent the overdosing of OARs in all cases? If not, does it still improve OAR dose sparing?

Research Question 3: Which of these two assessment schemes is more time-cost-effective?

3.2 Materials

Varian (a Siemens Healthineers company) is one of the world's largest medical physics and radiation oncology technology providers[26]. This study utilized Varian products including Aria (Version 15.1), Eclipse (Version 15.6), and Velocity (Version 4.1).

3.2.1 Aria: Varian's Record & Verify System

Aria is Varian's record and verify system. These types of systems operate as complete radiotherapy information management systems which can interface with imaging systems, treatment planning computers, and treatment delivery systems [35]. This means they store all information relevant to patients and their treatment. This can include basic demographic information as well as specifics of the treatment like dose prescription, collimator settings, total treatment delivery time, and gantry angle.

They also allow for communication between these separate workspaces, meaning that treatment parameters can be sent from treatment planning computers to treatment devices and stored after the completion of treatment. Aria also integrates scheduling and billing into its system so that the total treatment workflow can be as streamlined as possible.

Aria allowed us to gather patient demographic information, treatment plans, CT images, CBCT images, dose prescriptions, dose delivered to the patient (expressed in monitor units, or, MUs) and dose constraints for each patient's treatment course. Aria's import/export workspace also allowed us to transfer relevant data from various workspaces and software like Eclipse and Velocity.

3.2.2 Eclipse: Varian’s Treatment Planning System

Eclipse is Varian’s treatment planning system (TPS). This system within Aria allows users to create treatment plans based on a patient’s specific needs after taking in information like dose prescriptions, anatomical information, and disease type. One of many useful workspaces in this system is the contouring workspace. Here one can outline relevant volumes, or regions-of-interest (ROIs), such as target volumes OARs. It is common that target volumes, specifically the gross tumor volume (GTV), are contoured by a physician while OARs are contoured either by dosimetrists or medical physicists who are creating the treatment plan.

Once contours are drawn and verified, they can be used in Eclipse’s dose optimization algorithm, in the case of IMRT and VMAT plans. This algorithm allows one to specify constraints with which the plan must comply. These constraints typically include minimum dose levels that must be delivered to tumor volumes and maximum dose levels for OARs. The optimization algorithm takes these into account via a cost function and iterates over varying parameters so that the cost function is minimized. Once this plan optimization is completed and the dose is calculated, properties like the plan’s Dose Volume Histogram (DVH) can be found in Eclipse. Figure 3.1 shows an example of the planned dose distribution for a patient, along with the resulting DVH, after their treatment plan was optimized.

Relevant patients were treated at UNMC prior to this study being conducted, so all contouring of OARs and target volumes was completed by professional clinicians. These patients’ original treatment plans were also created, approved, and delivered prior to the start of this study. In addition to this, Eclipse was used for the creation and optimization of adaptive replans, as well as dose calculation when adaptive replans were created.

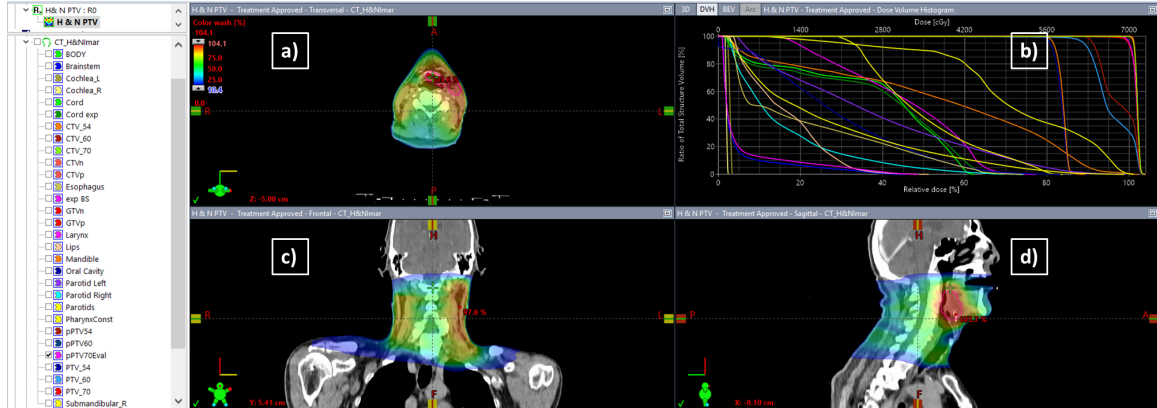


Figure 3.1: Planned dose distribution for a patient receiving treatment for H&N cancer. The left side of the image shows which plan is open (here, it is called H & N PTV), as well as the structure set (contours) for the plan. a) shows an axial view of the patient’s dose distribution with a color wash. b) shows the plans resulting DVH, which helps clinicians visualize what levels of dose are delivered to which structures. c) shows a coronal (frontal) view of the patient’s dose distribution. d) shows a sagittal view of the dose distribution. Both a) and d) show the PTV contour in red.

3.2.3 Velocity

Velocity is a tool that can be used for offline, or interfractional, adaptive radiotherapy. This piece of software is able to register multi-modality images, meaning that images from modalities like PET, CT, MRI, and CBCT can be spatially correlated [36]. This capability allows for the creation of synthetic images, called adaptive CTs (aCTs), which combine information from simulation CT images and daily CBCT images. These aCTs are crucial to the offline ART process as they provide information on changing patient anatomy through the use of CBCT images while maintaining high-fidelity information on tissue density from the simulation CT image.

Through Velocity’s creation of aCTs, the dose delivered by both the original and all re-optimized treatment plans can be calculated using images which are more representative of the patient’s evolving anatomy at any given time. This allows for more accurate estimations of the dose truly delivered to patients over the course of their treatment. Additionally, Velocity is able to accumulate dose over multiple images and treatment plans. This deformed-dose summation feature allows for the consideration

of more than just a single snapshot of a patient’s anatomy. The creation of aCTs and dose summation over multiple images were utilized for this study, both of which rely on deformable image registration.

These three components have been used to complete all work relevant to this study. Any institution with Aria, Eclipse, and Velocity is able to perform all the workflows described hereafter, which was the intention of their creation. Section 3.3 describes the work carried out in this study and familiarizes the reader with all necessary nomenclature used.

3.3 Methods

Tools like Aria, Eclipse, and Velocity have proven to be important pieces of a fruitful offline adaptive radiotherapy program in the previous section. Their uses have been outlined and their fit into this study has been defined. What follows is an in-depth description of how these tools were used to find answers to the three research questions detailed in Section 3.1.

3.3.1 Assessment Schemes

One goal of this study is to quantify the performance of two different assessment schemes for offline adaptive radiotherapy: the midpoint scheme and the biweekly scheme. It follows, then, that the protocol for these schemes must be laid out clearly. These schemes are meant to be applied clinically and should be viewed from the perspective of an interested clinician.

Each assessment scheme uses the general workflow developed by Hyun et al., as shown in Appendix A. This workflow, shown in Figure 3.2, provides direct instructions for a clinician wishing to determine whether an adaptive replan is warranted. The six steps of this workflow involve exporting data, generating aCTs through DIR, evalu-



Figure 3.2: Six steps of the offline ART workflow developed by Hyun et al. Each of these six steps can be carried out using Aria, Eclipse, and Velocity. See Appendix A for further details.

ating the quality of deformations, calculating dose, accumulating dose over multiple aCTs (if required), and analyzing dose. This workflow is explained more thoroughly in Appendix A. However, a few of these steps require some additional notes. Step two of this workflow, the creation of adaptive CT (aCT) images, was only carried out through the use of full field-of-view (full-FOV) CBCT images and simulation CT images. Since it is UNMC’s policy to take one full-FOV CBCT image per week of treatment, this allowed for one aCT to be created per week, in most cases. Two of the three patients investigated in this study had a single week in which no full-FOV CBCT was taken. Additionally, the evaluate deformations step involves QA checks of each aCT’s negative Jacobian determinants and high-density voxel elasticity. It was confirmed, through inspection of Velocity’s post-aCT-generation report, that no aCTs created for the purposes of this project had any negative Jacobian determinants. This confirms the absence of situations in which image voxels fold into the same space and represent unphysical transformations. Velocity’s post-aCT-generation report stated that scaled high-density voxel values ranged from 0.39 to 2.39. Each aCT with a high-density voxel elasticity above 1.0 was flagged and manually inspected. After this inspection, it was determined that areas with poor registration would not affect our OAR dose estimations.

Another important note to make is how the dose analysis step is carried out in this workflow. For the purposes of this project, dose analysis will determine whether an adaptive replan should take place. Both assessment schemes will only look at dose delivered to seven particular organs-at-risk (OARs). These OARs, along with their

clinical dose constraints, are laid out in Table 3.1. The choice of specific values and constraint types (i.e., mean vs max) come from the primary physician working with a particular patient. The rationale behind these choices most often comes from clinical trials. When the dose estimation for an OAR, as created by the adaptive workflow, is above the dose constraint, a replan is deemed necessary. Otherwise, no replan is triggered. It is important to note that, in this study, a violation to any of these seven OARs’ dose constraints is given equal weight.

After the offline adaptive workflow is carried out at any given point during the course of the patient’s treatment, a replan could be triggered. Appendix C provides specific instructions for how replans were carried out in this study. In short, the process uses Eclipse’s VMAT Optimization workspace and focuses on ensuring that the re-optimized treatment plan delivers sufficiently low dose to OARs which may have triggered the replan. Points of importance in this process include great care in ensuring that the correct dose constraints are input, taking note of which OARs are receiving too much dose according to the workflow’s estimation, and using the same plan normalization as the original treatment plan. For instance, if the dose to the PTV was normalized such that 95% of the tumor volume received 70 Gray (Gy) (V70 = 95%) for the original treatment plan, it should be normalized the same way after any re-optimizations.

OAR	Dose Constraint (cGy)
Spinal Cord	Max < 4500
Larynx	Mean < 4000 or 4500*
Left OR Right Parotid Gland	Mean < 2000 or 2600*
Pharynx	Mean < 5000-5500
Mandible	Max < 7200
Esophagus	Mean < 3000

Table 3.1: OARs used for Dose Analysis step of offline ART workflow. Dose constraints were taken from the original treatment plans of the patients in the study. Max means the maximum dose to the structure. Mean is the average dose delivered to the structure. *Constraint changed from patient to patient.

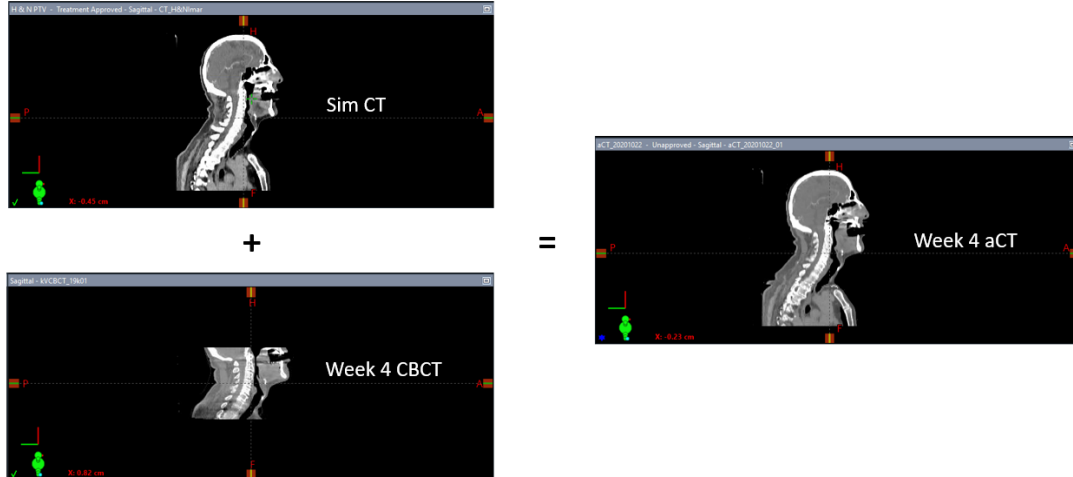


Figure 3.3: Saggital view of a patient’s simulation CT, Week 4 CBCT, and the resulting aCT representing Week 4 anatomy. The aCT is created using Velocity’s DIR algorithm based on a cubic b-spline model.

Midpoint Assessment Scheme

As implied by the name, a clinician using the midpoint assessment scheme carries out the general adaptive workflow once at the midpoint of a patient’s treatment course. As most H&N cases are treated over the course of seven weeks, this assessment happens most often during the fourth week of treatment.

The process begins in the fourth week, where a patient’s full-FOV CBCT from week four, treatment plan, and structure set (contours) are exported to Velocity from Eclipse. Then, an adaptive CT is created and verified. The aCTs are verified through visual inspection, by reviewing high-density voxel elasticity, and by checking the levels of negative Jacobian determinants seen in the transformation process. Figure 3.3 shows an example of an adaptive CT next to the patient’s simulation CT and week 4 CBCT.

Once the aCT is created and verified, the image, treatment plan, and structure set are sent back to Eclipse. Here, the dose delivered by the treatment plan is calculated. The dose delivered to the week four aCT, representative of the patient’s anatomy at that time, and scaled up to the full treatment length, constitutes the dose estimation

OAR & Constraint Type	Dose Constraint (cGy)	$(D_{test})_{Mid,Wk4}$ (cGy)
Spinal Cord Max	4500	4383.5
Larynx Mean	4000	3262.4
Left Parotid Mean	2600	3106.9
Right Parotid Mean	2000	961
Pharynx Mean	5000-5500	5103.9
Mandible Max	7200	7221
Esophagus Mean	3000	1321.2

Table 3.2: Comparison of Patient 2’s dose constraints and $(D_{test})_{Mid,Wk4}$ for seven OARs.

created by midpoint assessment. In other words, the midpoint estimation is the dose which would be delivered to the patient if the week four aCT represented the patient’s anatomy over the entire treatment course. This estimation will be given by

$$(D_{test})_{Mid,Wk4} = D_{Wk4,P1}, \quad (3.1)$$

where $D_{Wk4,P1}$ is the total dose delivered to week four anatomy by the original (first) treatment plan. The subscript test is used to denote dose estimations created through use of an assessment scheme as a clinician would use it.

Calculating the dose over the week four aCT allows for dose estimations to be created for each individual OAR. So, for the purposes of this study, this estimation turns into seven estimations for organs at risk. All further examples of this process will be taken from Patient 2’s data. Patient 2’s dose constraints and $(D_{test})_{Mid,Wk4}$ estimations are shown in Table 3.2. There are two instances in which the $(D_{test})_{Mid,Wk4}$ estimation is higher than the dose constraint (shown in red). So, the estimation for the left parotid gland’s mean dose and the mandible’s max dose are violating the dose constraint and triggering a replan.

In this example, following the assessment protocol outlined in Appendix A, one can see that the Patient 2 needs a replan, as evidenced by the estimated violations in the left parotid gland’s mean dose and mandible’s max dose. This means that

special attention is given to the constraints of the left parotid and mandible during the optimization process. Once the optimization process is completed, dose delivered by the re-optimized plan is calculated on the week 4 aCT and the plan is normalized in the same way that the original plan was. So, for instance, Patient 2’s original treatment plan was normalized such that 95% of the PTV received the prescription dose of 70 Gy. This means that the re-optimized plan should share this normalization. Table 3.3 shows the OAR doses calculated with the re-optimized plan on week 4’s aCT.

Once this plan is re-optimized, it would be used for the remainder of the patient’s treatment course, since the midpoint assessment scheme does not involve any further time-points for assessment and potential replanning. Then, an estimate of the final dose delivered to the patient is calculated based on week four’s aCT, the original treatment plan, and the re-optimized plan, if it was used. This dose summation is performed in Velocity using their ACTIVE Dosimetry Navigator. In-depth instructions for this process are provided in Appendix D. In the case of a replan being used, the estimate for the final dose delivered to the patient would be

$$(D_{test})_{mid,end} = 4 \times D_{Wk4,P1} + 3 \times D_{Wk4,P2}, \quad (3.2)$$

OAR	Dose Constraint (cGy)	$(D_{test})_{Mid,Wk4}$ (cGy)	Replan Dose (cGy)
Spinal Cord Max	4500	4383.5	4413
Larynx Mean	4000	3262.4	2596.6
Left Parotid Mean	2600	3106.9	2296.4
Right Parotid Mean	2000	961	949.7
Pharynx Mean	5000-5500	5103.9	5046.4
Mandible Max	7200	7221	7187.5
Esophagus Mean	3000	1321.2	1342.3

Table 3.3: Comparison of $(D_{test})_{Mid,Wk4}$ and re-optimized dose with dose constraints for Patient 2. The replan improves dose sparing to the OARs which may have violated their dose constraints (shown in red) while ensuring that target coverage is maintained.

OAR	Dose Constraint (cGy)	$(D_{test})_{mid,end}$ (cGy)
Spinal Cord	Max < 4500	4386
Larynx	Mean < 4000	3064
Left Parotid	Mean < 2600	2872
Right Parotid	Mean < 2000	968
Pharynx	Mean < 5000-5500	5025
Mandible	Max < 7200	7155
Esophagus	Mean < 3000	1271

Table 3.4: The midpoint assessment’s estimation for the total dose delivered to Patient 2 when a replan is performed at week four. When compared to the week four assessment, total triggers decreased and the OAR still triggering (shown in red) decreased in dose.

where $D_{Wk4,P1}$ is the weekly dose delivered by the original plan to week four’s aCT and $D_{Wk4,P2}$ is the weekly dose delivered by the re-optimized plan to week four’s aCT. We note here that the multiplication by four and three indicate the number of 5-day treatment weeks to which this dose is scaled. So, the $D_{Wk4,P1}$ is used to represent 20 fractions of treatment while $D_{Wk4,P2}$ represents 15 fractions. If no replan was used, the estimate for final dose would be

$$(D_{test})_{mid,end} = D_{Wk4,P1}, \quad (3.3)$$

where $D_{Wk4,P1}$ is scaled to the entire treatment length. Table 3.4 shows the final dose estimation for each OAR. So, in the case of Patient 2, the total amount of estimated triggers decreased from two to one, and the remaining trigger exceeded its dose constraint by a lower amount. Table 3.4 highlights the anticipated difference that the midpoint assessment scheme will make for this particular patient. Additionally, a flowchart depicting the overall process is shown in Figure 3.4.

Biweekly Assessment Scheme

Figure 3.5 shows an overview of the flow chart for the creation of dose estimates through the biweekly assessment scheme. The biweekly assessment scheme increases

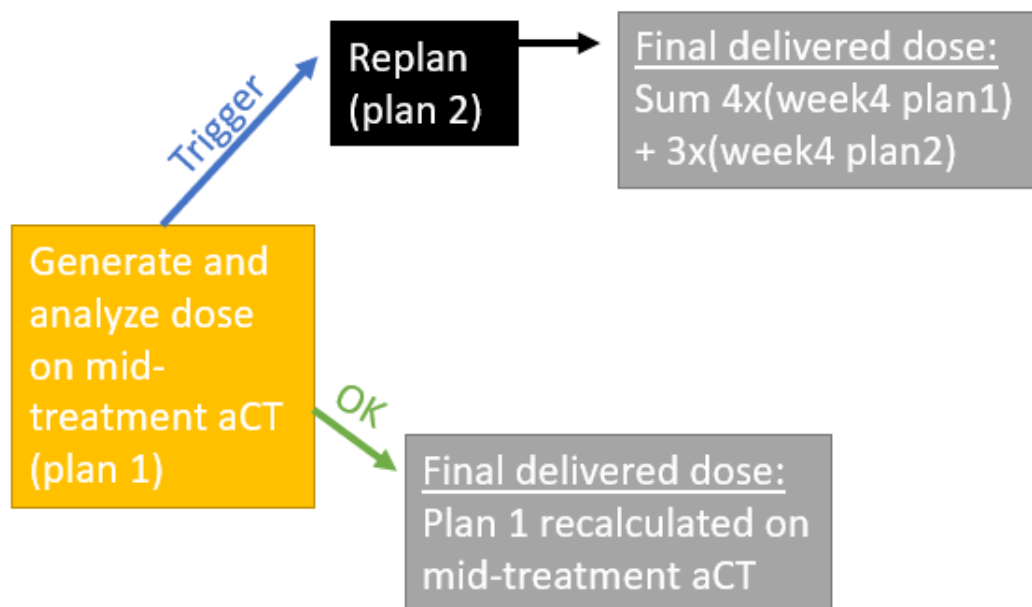


Figure 3.4: Flowchart for the midpoint assessment scheme. This scheme, at most, requires one replan and one assessment to be performed. Final dose estimates only use information gathered using week four’s aCT image.

assessment frequency to once every two weeks. This increases the minimum required work by three times when compared to the midpoint assessment scheme since three assessments take place. It may increase the workload even more if multiple replans are required. The expectation is that this process will lead to a decrease in OAR dose constraint violations at the end of treatment compared to the midpoint scheme or no assessment/adaptation.

The first assessment takes place during the second week of a patient’s treatment course. The week 2 full-FOV CBCT image, the original treatment plan, and structure set are exported to Velocity. Here, the week two aCT is created and verified. The aCT, along with treatment plan and structure set, is sent back to Eclipse and dose is calculated using the week two aCT. If any OAR dose estimates are above their constraints, a replan is triggered. The dose estimate created by this week two assessment is given by

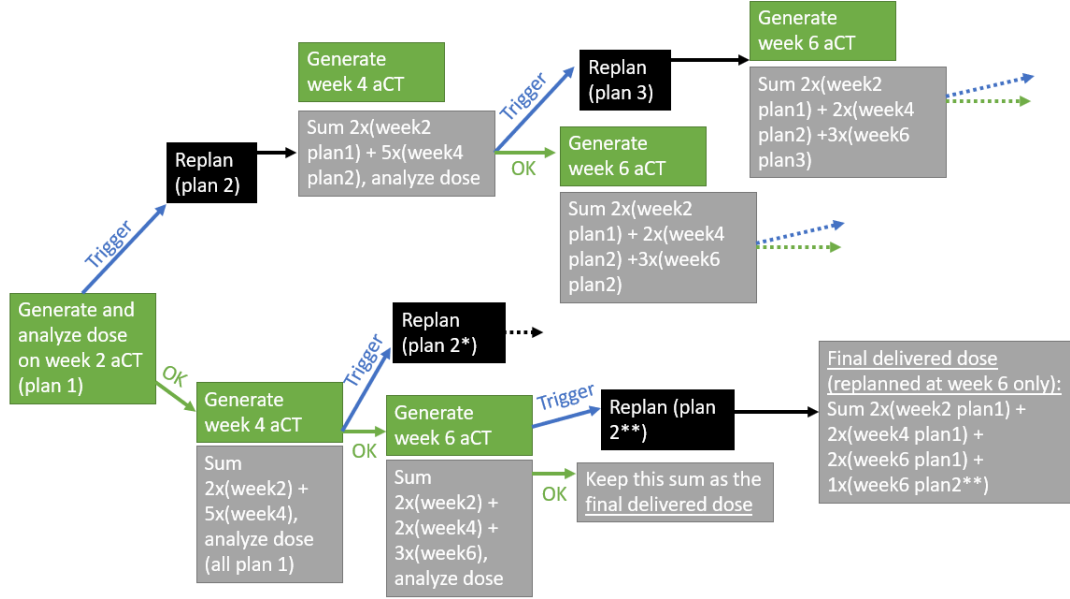


Figure 3.5: Flowchart of the creation of dose estimates using the biweekly assessment scheme. Plan numbers begin at 1 with the original plan. The *'s represent different replans based on when triggers occur. So, Plan 2 is created after a trigger on week two's assessment. Plan 2*, however, didn't trigger while week four's assessment did trigger.

$$(D_{test})_{Bi,Wk2} = D_{Wk2,P1}, \quad (3.4)$$

where $D_{Wk2,P1}$ is the total dose delivered to the week 2 aCT by the original treatment plan, scaled up to the entire treatment course. Table 3.5 shows the week 2 estimation of Patient 2 compared with their specific dose goals. We can see that the mean dose to the left parotid gland is estimated to violate the dose constraint, meaning that a replan is required. In this case, the original treatment plan is re-optimized, with special attention given to the dose to the left parotid gland, and the result is normalized identically to the original treatment plan. Once this plan is approved, it is used for the following two weeks (weeks 3 and 4).

At week four, the second assessment takes place. This assessment takes information from both week two and week four into account. An aCT is created for week four's anatomy in the same way as week two, but an additional dose summation step

is required. Specifically, a clinician would use two weeks of the dose estimation from week two, followed by five weeks of a dose estimation using week four anatomy and the re-optimized plan. This can be expressed as

$$(D_{test})_{Bi,Wk4} = 2 \times D_{Wk2,P1} + 5 \times D_{Wk4,P2}, \quad (3.5)$$

where $D_{Wk2,P1}$ is the weekly dose delivered by the original plan to week two's aCT (multiplying by 2 scales it to 10 fractions) and $D_{Wk4,P2}$ is the weekly dose delivered by the re-optimized plan to week four's aCT (multiplying by 5 scales it to 25 fractions). This is done as a best estimate for how long both of these plans were used. Then, this estimation determines whether an additional replan is required. Table 3.6 shows the resulting week four estimation for Patient 2. Given the results of this new estimation, an additional replan is required.

After the second re-optimization process, this new plan (Plan 3) is used over weeks five and six. At week six, the third assessment takes place. In a similar fashion to weeks two and four, the aCT is created and dose from Plan 3 is calculated. The resulting estimation, following the trend from week four, uses information from both previous assessment points. The equation for this assessment is

$$(D_{test})_{Bi,Wk6} = 2 \times D_{Wk2,P1} + 2 \times D_{Wk4,P2} + 3 \times D_{Wk6,P3}, \quad (3.6)$$

OAR	Dose Constraint (cGy)	$(D_{test})_{Bi,Wk2}$ (cGy)
Spinal Cord	Max < 4500	4386.7
Larynx	Mean < 4000	3271.3
Left Parotid	Mean < 2600	2638.1
Right Parotid	Mean < 2000	1091.9
Pharynx	Mean < 5000-5500	5086.9
Mandible	Max < 7200	7076.2
Esophagus	Mean < 3000	1651.6

Table 3.5: Week two's dose estimation for Patient 2 created through the biweekly use of the offline adaptive workflow.

OAR	Dose Constraint (cGy)	$(D_{test})_{Bi,Wk4}$ (cGy)
Spinal Cord	Max < 4500	4423
Larynx	Mean < 4000	2788
Left Parotid	Mean < 2600	2872
Right Parotid	Mean < 2000	951
Pharynx	Mean < 5000-5500	5013
Mandible	Max < 7200	7203
Esophagus	Mean < 3000	1256

Table 3.6: Week four’s dose estimation created through the biweekly use of the offline adaptive workflow.

where the multipliers indicate that $D_{Wk2,P1}$ and $D_{Wk4,P2}$ account for 10 fractions each and $D_{Wk6,P3}$ accounts for 15 fractions. The resulting estimation is shown in Table 3.7. As this shows another estimated dose constraint violation, one more re-optimization is required.

A final re-optimization takes place using the week 6 aCT and, after its approval, this plan is used for the remainder of the patient’s treatment course. Since no additional aCTs are generated or used, the final dose estimation for this patient is equivalent to the week six estimation. So,

$$(D_{test})_{Bi,end} = (D_{test})_{Bi,Wk6} = 2 \times D_{Wk2,P1} + 2 \times D_{Wk4,P2} + 3 \times D_{Wk6,P3}. \quad (3.7)$$

OAR & Constraint Type	Dose Constraint (cGy)	$(D_{test})_{Bi,Wk4}$ (cGy)
Spinal Cord Max	4500	4370
Larynx Mean	4000	2815
Left Parotid Mean	2600	2690
Right Parotid Mean	2000	1038
Pharynx Mean	5000-5500	4994
Mandible Max	7200	7188
Esophagus Mean	3000	1325

Table 3.7: Week six’s dose estimation created through the biweekly use of the offline adaptive workflow.

It is important to note the case of Patient 2 is one possible branch of the flow chart seen in Table 3.5. Using biweekly assessment for different patients may result in different weeks' estimates predicting OAR triggers, which changes the form of subsequent estimates. The most important idea when applying these schemes is to use all the updated anatomical information, as well as updated treatment plans, that are available, which will depend on the assessment scheme used.

In the case of Patient 2, we see a final estimation with the same trigger remaining. It even looks like the estimation is worse than where it started. But, this estimation fails to take into account the final replan (Plan 4). This is a potential source of deviation from the best possible estimate of delivered dose, which may lead to clinical estimations not faithfully representing whether dose constraints are truly violated. From a research perspective, then, these estimations should be compared to the best estimate so their accuracy can be quantified. More specifically, both schemes' ability to accurately determine whether OAR triggers are present, as well as the schemes' ability to estimate the final delivered dose should be investigated. Their level of accuracy may inform the decision-making process of interested clinicians when determining which assessment scheme may be of most use to them.

3.3.2 Research Question 1

The two previous sections have laid out the processes through which clinicians can employ offline adaptive radiotherapy at varying frequencies. It is also very important that these assessment schemes are able to be analyzed for their accuracy from a research perspective. Specifically, these assessment schemes should be evaluated by their ability to catch triggers when performing assessments and their ability to predict the final dose delivered to patients after assessment and replanning. These two modes of evaluation help to answer the first research question, which was:

Research Question 1: Do the midpoint and biweekly assessment schemes accurately predict replan necessity and final delivered dose when compared to a retrospective best estimate?

This will be done through the creation of two forms of "ground truth" estimates which are best estimates created using all available anatomical information (not just the anatomy at the weeks of assessment, which would be available to a clinician using that scheme). The idea behind these was to create the best possible estimate that corresponds to a scheme's replan detection and final dose estimates. Once these are created, the results from the two assessment schemes, from the perspective of a clinician, are compared to their corresponding ground truths. The first of these ground truths to be created determines how well a given assessment scheme detects necessary replans, or triggers.

How well do schemes detect triggers? (Detector Evaluation)

The first of these evaluations, the scheme's ability to catch triggers when assessing, can provide a clearer idea of whether a given scheme can be trusted to signal when replans should be carried out. In other words, this evaluation determined how well the assessment scheme detected triggers at a given point in time, given replans up until that point, but assuming no further replans. This ability will be quantified through the use of the diagnostic test format. Data was displayed in the form of a confusion matrix, which compares the binary outcome of a diagnostic 'test' with some 'true' determination. An example confusion matrix is shown in Figure 3.6. In all cases, True Positives (TP) and True Negatives (TN) are desirable results, while False Positives (FP) and False Negatives (FN) are to be avoided. The presence of TP and TN data points represent instances of the diagnostic test working properly.

		Truth	
		Pos	Neg
Test	Pos	TP	FP
	Neg	FN	TN
		Sens.	Spec.
		$TP/(TP+FN)$	$TN/(TN+FP)$

Figure 3.6: An example confusion matrix. The True positive (TP) represents a situation in which the diagnostic test reads positive and the true result is positive. True negatives (TN) are the same, but for negative results. False positives (FP) arise when the test reads as positive when the true result is negative. False negatives (FN) arise when the test reads as negative but the true result is positive. Sensitivity and specificity calculations are shown.

The two additional metrics shown in Figure 3.6, sensitivity and specificity, are helpful ways to summarize coherent results from the data. Sensitivity, which is given by

$$Sens. = \frac{TP}{TP + FN}, \tag{3.8}$$

provides information on how well positive cases are correctly identified. Specificity, given by

$$Spec. = \frac{TN}{TN + FP}, \tag{3.9}$$

describes the ability of the diagnostic test to report true negative cases as negative. These values will prove useful in the evaluation of the midpoint and biweekly assessment schemes.

For this first type of assessment evaluation, the estimations created by each assessment point (e.g., week four using the midpoint assessment scheme detailed in Section 3.3.1) will be used as the 'test'. However, a ground truth needs to be created so that the estimations' accuracy can be quantified. The ground truth should represent our best possible estimation of whether a patient's treatment course, if it were to continue with the most recent treatment plan created, violates dose constraints for the seven relevant organs-at-risk (OARs). This best estimate is obtained through the creation of additional aCTs, one for each week with an available full-FOV CBCT image, and the calculation of dose on these aCTs from the appropriate treatment plan. A simple example of ground truth is that which is used for evaluation of the midpoint assessment. Since no prior replans have taken place before the assessment at the midpoint of treatment, the ground truth is simply the dose delivered by the original treatment plan, accumulated over each available aCT. So, given that Patient 2 had seven week's worth of aCTs created, the ground truth for this evaluation is

$$(D_{truth})_{mid} = D_{P1,(Wk1-Wk7)}, \quad (3.10)$$

where $D_{P1,(Wk1-Wk7)}$ represents the dose delivered by the original treatment plan (P1) to the aCTs representing weeks one through seven. For this ground truth, we can define a positive as a situation in which it shows an OAR violation occurring. It follows, then, that a negative ground truth result is a situation in which an OAR does not trigger. Table 3.8 shows the truth and test estimated values for each OAR considered. Here we see a situation in which the midpoint assessment correctly catches the left parotid gland's overdosage. This data point corresponds to a true positive (TP) in our confusion matrix. We also see a false trigger on the mandible's dose. This corresponds to a false positive (FP). Largely, though, we see situations where the assessment scheme correctly predicts that no trigger will take place. These all

OAR	Truth ($D_{\text{truth}})_{\text{mid}}$ (cGy)	Test ($D_{\text{test}})_{\text{mid,Wk4}}$ (cGy)
Spinal Cord Max	4354	4383.5
Larynx Mean	3325	3262.4
Left Parotid Mean	2864	3106.9
Right Parotid Mean	1103	961
Pharynx Mean	5037	5103.9
Mandible Max	7095	7221
Esophagus Mean	1351	1321.2

Table 3.8: Truth and Test values for the midpoint assessment’s ”Detector” evaluation. Here, we note that any red colored numbers represent OAR triggers.

		Truth	
		Pos	Neg
Test	Pos	1	1
	Neg	0	5
		Sens.	Spec.
		1.0	0.83

Figure 3.7: Confusion matrix displaying how well the midpoint assessment scheme caught necessary replans for Patient 2. All true OAR triggers were caught and one ’false alarm’ (or false positive) was reported.

correspond to true negative (TN) results. The resulting confusion matrix, using the data from Patient 2 in Table 3.8, is shown in Figure 3.7.

Creating ground truths for the biweekly assessment scheme requires a bit more thought. Again, the goal of this first evaluation was to determine how well this assessment detects triggers at this point in time, given replans up to this point, but assuming no further replans. So, it was useful to define the process of creating ground truth estimates through Patient 2. Each week’s assessment will need a separate ground truth, and fortunately, the week two assessment’s ground truth is the same

as it was for midpoint assessment – Equation 3.10. This is the case because at week two, there are no possible replans to take into account.

Given that our Patient 2 treatment plan was re-optimized at week two, seen in Table 3.5, the ground truth for the week four assessment changes. Instead of only using the original treatment plan, the use of a second replan after week two needs to be accounted for. Specifically, the equation for the ground truth of week four’s assessment of Patient 2 is

$$(D_{truth})_{bi,Wk4} = D_{P1,(Wk1-2)} + D_{P2,(Wk3-7)}, \quad (3.11)$$

where $D_{P1,(Wk1-2)}$ represents the original treatment plan applied to the week one and week two aCTs. $D_{P2,(Wk3-7)}$ represents the re-optimized plan (Plan 2) applied to weeks three through seven’s aCTs. This ground truth tells us, with a best estimate, which OAR constraints would be violated if this treatment course continued with no further alteration.

Once again, we see that Patient 2’s plan was re-optimized at week four. This means that an additional alteration to the ground truth for week six’s assessment needs to be done. This week six ground truth is given by

$$(D_{truth})_{bi,Wk6} = D_{P1,(Wk1-2)} + D_{P2,(Wk3-4)} + D_{P3,(Wk5-7)}, \quad (3.12)$$

where $D_{P3,(Wk5-7)}$ is the third plan iteration (second re-optimization) applied to weeks five through seven’s aCTs. However, for example, if the week four assessment did not trigger a replan, the ground truth would remain the same as in Equation 3.11. So, the diagnostic test method allows us to evaluate how well the midpoint and biweekly assessment schemes detect when OAR triggers occur.

The second aspect of an assessment scheme’s ‘trustworthiness’, an integral part of Research Question 1, describes how well the schemes predict the final dose delivered to patients with all replans taken into account.

How well do assessment schemes predict final delivered dose?

The final dose estimate, for both biweekly and midpoint assessment schemes, takes into account all replans performed and is only created after the final assessment takes place. So, for the midpoint assessment, the test values correspond to those created by either Equation 3.2 or 3.3, depending on whether a replan was triggered. The biweekly test values are created after assessment and potential replanning at week six. So, for instance, the final dose estimation for Patient 2 is given by Equation 3.7. This is a particular situation in which a replan was triggered each of the three times the treatment plan was assessed.

These final dose estimates take into account anatomy which was used during assessment, which leads to a lower time requirement, but they may suffer in accuracy compared to the best possible estimate. This is why, from the research perspective, it is important to quantify the accuracy of these final estimates through comparing them against ground truth best estimates. These ground truths can be created through the use of all available aCTs and replans. So, for the midpoint assessment as applied to Patient 2, the ground truth is given by

$$(D_{truth})_{mid,end} = D_{P1,(Wk1-4)} + D_{P2,(Wk5-7)}, \quad (3.13)$$

meaning that the original treatment plan was applied over the first four weeks of treatment and, after triggering, the re-optimized plan was applied over weeks five through seven. Table 3.9 shows the final truth and test values for the midpoint assessment, which show the final OAR doses delivered to Patient 2. We can see the majority of OARs agree between truth and test. Specifically, we see that the final dose

OAR	Truth ($D_{\text{truth}})_{\text{mid,end}}$ (cGy)	Test ($D_{\text{test}})_{\text{mid,end}}$ (cGy)
Spinal Cord Max	4356	4376
Larynx Mean	3191	3200
Left Parotid Mean	2574	2885
Right Parotid Mean	1111	977
Pharynx Mean	5013	5023
Mandible Max	7110	7172
Esophagus Mean	1358	1269

Table 3.9: Truth and Test values for the evaluation of the midpoint assessment’s ability to predict final dose with adaptation. Here, we note that any red numbers represent OAR triggers.

estimate predicts that six OARs will remain below the appropriate dose constraints. These six OARs are instances of true positives in which an organ is estimated to stay under its plan goal with replans considered, and it stays under the goal according to the best possible estimate, or ground truth. The other OAR, the left parotid gland, is predicted to violate its dose constraint by the end of treatment, but the best estimate says it doesn’t. This corresponds to a false negative result. So, for Patient 2, these results can be displayed via confusion matrix in Figure 3.8.

The creation of ground truth for biweekly assessment works much in the same way. Particular care should be taken to ensure that, if a replan is triggered by the week six assessment, that the replan is applied to week seven’s aCT. This should be accounted for in the ground truth calculation of final dose. Since Patient 2’s treatment plan was re-optimized at week six, the ground truth for the final dose delivered is then

$$(D_{\text{truth}})_{\text{bi,end}} = D_{P1,(Wk1-2)} + D_{P2,(Wk3-4)} + D_{P3,(Wk4-6)} + D_{P4,(Wk7)}, \quad (3.14)$$

where four plans are used in total. Again, it should be noted that this ground truth is created due to the unique path this particular patient’s treatment course was on. Re-planning after each assessment is one of many possible ways the midpoint assessment scheme can go. While Patient 2 happened to follow the top path of the flowchart

		Truth	
		Pos	Neg
Test	Pos	6	0
	Neg	1	0
		Sens.	Spec.
		0.86	N/A

Figure 3.8: Confusion matrix which quantifies the ability of the midpoint assessment to predict final dose for Patient 2. Specificity was omitted due to a lack of TN and FP. Note that the rightmost column represents situations in which replanning ensured that OARs stayed under the plan goals.

in Figure 3.5, which resulted in this particular ground truth being used, going down any other path would have resulted in a new ground truth being calculated. Through the use of confusion matrices, the midpoint and biweekly assessment schemes can be evaluated through their sensitivity and specificity defined by staying under or going over OAR dose constraints. This adds to the body of information interested clinicians may refer to when determining which assessment scheme, if any, fits the needs of their clinic.

3.3.3 Research Question 2

The second research question this study posed relates to each assessment schemes' ability to decrease overall dose constraint violations. This question evaluated the effectiveness of these schemes; in other words:

Research Question 2: Does the use of the midpoint assessment scheme or bi-weekly assessment scheme for ART prevent the overdosing of OARs in all cases? If

not, does it still improve OAR dose sparing?

This was quantified through counting the number of true constraint violations seen with and without adaptive replanning. So, for Patient 2 and midpoint assessment, we can see that the total constraint violations seen without adaptation is 1. This is evident when looking at the truth column of Table 3.8. This can be compared against the truth column of Table 3.9, which shows the total constraint violations when adaptation is used. So, use of the midpoint assessment on Patient 2 decreased the dose constraint violations from one to zero for this patient.

Additionally, the magnitude of the decrease in dose may be useful to quantify since this may, in some situations, be more important than avoiding an OAR trigger. For instance, a situation in which the larynx mean dose decreases from 70 Gy to 50 Gy may be more useful than a situation in which the larynx mean dose decreases from 40.01 Gy to 39.99 Gy. The latter situation does prevent a trigger, if the dose constraint is set to 40 Gy, but may not translate to any meaningful difference in clinical outcomes. The former situation, on the other hand, may decrease the probability of toxicity considerably, even though it is still above the constraint value. Quantifying these values adds to the body of information clinicians can use when determining whether either assessment scheme may be of use to them.

The same thing can be done to quantify the benefit of the biweekly assessment scheme. Answering this research question gives clinicians a better idea of the true dosimetric benefit they can expect when employing either the midpoint or biweekly assessment scheme.

3.3.4 Research Question 3

The third research question asks about the time-cost-effectiveness of the midpoint and biweekly assessment schemes from the perspective of time-cost and dosimet-

ric benefit. According to the Center for Disease Control (CDC), cost-effectiveness analysis "is a way to examine both the costs and health outcomes of one or more interventions. It compares an intervention to another intervention (or the status quo) by estimating how much it costs to gain a unit of a health outcome, like a life year gained or a death prevented." [37]. We will attempt to perform this type of analysis (using physicist time as the cost) with the hope that it will answer the following question:

Research Question 3: Which of these two assessment schemes is more time-cost-effective?

In order to answer this question, time-cost and dosimetric benefit were quantified. The first of these was done through an investigation of each individual step of the midpoint and biweekly assessment schemes and its associated time-cost. It should be noted that, while there is a general flowchart for the process of midpoint and biweekly assessment, each application of these schemes can translate to no, few, or many replans. The amount of replans drastically changes the time-cost and may affect the time-cost of other steps as well. So, for the three patients investigated, we made the most conservative time-cost estimate. This means that we assume each case will have the maximum number of replans for that given scheme. In other words, midpoint assessment always triggers a replan and biweekly assessment triggers a replan three times. Table 3.10 outlines the time-cost estimates for each step of the midpoint and biweekly assessment schemes. As a note, additional sources of time-cost like physician re-contouring, re-simulating the patient, VMAT QA, and plan checks are not considered in this estimation of time-cost. These were left out because this estimation is meant to quantify the time-cost in physicist work hours only.

The quantification of dosimetric benefit was done a few different ways. First, the difference in final OARs whose constraints were violated were used to quantify benefit. This data was collected through the process outlined in the previous section. Second,

	Step	Time Cost (min)	
		Midpoint	Biweekly
Assessment & Replan	Export CBCT, simCT, etc. to Velocity	10	30
	Create aCT(s)	30	90
	aCT QA	10	30
	Calc plan on aCT	10	30
	Re-optimize plan	30	90
Final Dose Estimation	Export images & plans for final dose estimation	10	30
	Dose summation	30	90
Total Time Cost		130 min 2.17 hr	390 min 6.5 hr

Table 3.10: The estimated time-cost associated with each step of the midpoint and biweekly assessment processes used for offline adaptive radiotherapy.

the dosimetric change for OARs which were shown to violate dose constraints with no adaptation was used as an alternative benefit calculation. This was done by taking the difference between dose delivered to OARs without adaptation and dose delivered to OARs with adaptation. Then, an average dose difference for all triggered OARs was calculated. Third, the dosimetric change for all OARs was calculated as yet another conceptualization of dosimetric benefit. Again, an average of the dose change encountered when using either the midpoint or biweekly assessment scheme was calculated. Once both time-cost and dosimetric benefit are quantified, the relationship between them can be presented in a benefit-per-time (time-cost-effectiveness) format. This was done simply by dividing each benefit quantity by the time-cost estimate.

Chapter 4

Results & Discussion

The previous chapter described the structure of this study, the equipment used, the relevant research questions, and the process through which we attempt to answer these questions. Two assessment schemes for offline adaptive radiotherapy have been described. We also described how to quantify the performance of these assessment schemes. This includes evaluating the schemes' ability to accurately detect when replans are needed, as well as their ability to estimate the final dose delivered to patients. This information allows us to answer the research questions posed by this study. As a reminder, the three research questions which are answered in this section are given below.

Research Question 1: Do the midpoint and biweekly assessment schemes accurately predict replan necessity and final delivered dose when compared to a retrospective best estimate?

Research Question 2: Does the use of the midpoint assessment scheme or biweekly assessment scheme for ART prevent the overdosing of OARs in all cases? If not, does it still improve OAR dose sparing?

Research Question 3: Which of these two assessment schemes is more time-cost-effective?

It has also been established how to quantify whether these assessment schemes are translating to patient benefit. Study limitations, mainly that this is not a true prospective study, prevent the quantification of true clinical benefit in the form of clinical outcomes such as toxicity avoidance. But, knowledge of the dose delivered to patients allows for benefit to be described through a decrease in the number of organs-at-risk (OARs) whose dose constraints would otherwise be violated. Whether dose constraint violations are prevented, and if they aren't, whether the magnitude of the violation is decreased through use of a particular assessment scheme, provides information which can be used to answer Research Question 2.

Given that different assessment schemes require varying amounts of work, it is also important to quantify the relationship between the required time-cost for a particular scheme and its associated benefit. This is important because clinics are often required to be more efficient with their time and resources. Low-resource clinics may especially benefit from clear information connecting time investment to expected outcome. So, the time required to complete the various tasks for these schemes has been quantified and will be compared to the outcome and performance of the midpoint and biweekly assessment schemes. This information will aid in answering Research Question 3.

As a reminder, midpoint and biweekly assessment were performed on three patients who received VMAT treatment for head and neck (H&N) cancer at the University of Nebraska Medical Center between 2020 and 2021. These patients were selected, after their treatment course was completed, based on noticeable anatomical changes as seen on their daily CBCT images. These assessments were designed such that adaptive replanning was triggered whenever dose to seven particular OARs was estimated to be above the clinically-determined dose constraint. These OARs and constraints are shown in Table 3.1.

4.1 Research Question 1

The assessment schemes used in this study created dose estimations at each assessment point (e.g., weeks 2, 4, and 6 for biweekly) and created final dose projections for each patient after taking into account all replans. These are created in an attempt to detect when to replan and to understand what dose was delivered to patients after replanning. To evaluate these schemes, ground truth was established for both of the schemes as per the instructions in Section 3.3.2. These assessments' test values were compared against the ground truths in the diagnostic test format.

4.1.1 How well did schemes detect triggers?

All results in this section are quantified using the diagnostic test format. The description below interprets the contextual meaning of true positive, false positive, true negative, and false negative for the first analysis in Research Question 1.

True Positive: A True Positive is a situation in which an organ was projected to violate its plan goal and was confirmed to violate a plan goal.

False Positive: A False Positive is a situation in which an organ was projected to violate its plan goal and did NOT actually violate a plan goal.

True Negative: A True Negative is a situation in which an organ was NOT projected to violate its plan goal and did NOT actually violate a plan goal.

False Negative: A False Negative is a situation in which an organ was NOT projected to violate its plan goal and was confirmed to violate a plan goal.

The midpoint assessment takes place once, as close to the midpoint of treatment as possible, and creates a dose estimate which describes projected dose to a patient if the treatment course were to carry on from that point unchanged. This estimate uses anatomical information from the fourth week of treatment and calculates dose delivered to that anatomy. This is scaled up to represent the entire course of treat-

		$(D_{\text{truth}})_{\text{mid}}$	
		Pos	Neg
$(D_{\text{test}})_{\text{mid}}$	Pos	3	1
	Neg	1	16
		Sens.	Spec.
		0.75	0.94

Figure 4.1: Confusion matrix comparing midpoint assessment’s trigger results against ground truth dose values. The numbers indicate the total number of OARs that resulted in that particular outcome.

ment. If the projected dose to any of the seven OARs listed in Table 3.1 violates their associated dose constraints, then a replan is triggered. The accuracy of the midpoint assessment scheme’s triggers was quantified by comparing the dose estimates created by the scheme against an estimation which uses all available anatomical information on the patient over their treatment course. The results are shown in Figure 4.1.

The sensitivity of 0.75 means that, out of four total times a trigger should have happened, three actually did. In the same vein, the specificity of 0.94 means that, out of 17 OARs which should not have triggered a replan, 16 correctly did not trigger. To find the overall accuracy, we can divide the sum of true positives and true negatives by the total amount of data points. Overall, the midpoint assessment’s trigger detection accuracy, according to this test format, is $\frac{19}{21} = 0.90$. So, with the current sample size of 3 patients, we conclude that midpoint assessment correctly determined replan necessity approximately 90% of the time.

		$(D_{\text{truth}})_{\text{bi}}$	
		Pos	Neg
$(D_{\text{truth}})_{\text{bi}}$	Pos	6	3
	Neg	2	52
		Sens.	Spec.
		0.75	0.95

Figure 4.2: Confusion matrix comparing biweekly assessment’s trigger results against ground truth dose values. The numbers indicate the total number of OARs that resulted in that particular outcome.

Biweekly assessment follows a similar protocol when compared with midpoint assessment, but assesses more frequently and may result in more replans. Assessments take place on weeks two, four, and six, meaning that three separate detection attempts take place. As a result, the number of detection attempts triples for this particular analysis. Each assessment point only takes into account replans that have taken place prior and assumes no further replans. Again, the idea is to create an estimate of what dose would be delivered if the treatment course remained unchanged after that point in time. Results for the ability of the biweekly assessment scheme to detect triggers are shown in Figure 4.2. These numbers are virtually identical to those seen from the midpoint assessment. So, we can conclude that the sensitivity of 0.75 and specificity of 0.95 result in an overall trigger detection accuracy of $\frac{58}{63} = 0.92$. This means that biweekly assessment correctly determines replan necessity 92% of the time.

4.1.2 How well do assessment schemes predict final delivered dose?

Similar to the previous section, all results for the second analysis of Research Question 1 are displayed using the diagnostic test format, as summarized in the description below.

True Positive: A True Positive is a situation in which an organ is estimated to stay under its plan goal with replans considered AND it stays under the goal according to the 'truth'.

False Positive: A False Positive is a situation in which an organ is estimated to stay under its plan goal with replans considered AND it violates its plan goal according to the 'truth'.

True Negative: A True Negative is a situation in which an organ is estimated to violate its plan goal with replans considered AND it violates its plan goal according to the 'truth'.

False Negative: A False Negative is a situation in which an organ is estimated to violate its plan goal with replans considered AND it stays under the goal according to the 'truth'.

The midpoint assessment scheme, after its determination of triggers, creates a final dose projection based on whether a replan was created. If a replan was performed, the final dose estimate takes both the original plan and replan into account. In principle, if no replan took place, the final dose estimate simply scales the dose delivered by the original treatment plan to week four anatomy to the entire treatment course. However, the latter case was not realized in this study because each of the three patients had a replan triggered using midpoint assessment. These estimates are compared to ground truth estimates which use all available anatomical information and all replans to create the best final dose estimation. Cumulative results for the midpoint

		$(D_{\text{truth}})_{\text{mid,end}}$	
		Pos	Neg
$(D_{\text{test}})_{\text{mid,end}}$	Pos	18	1
	Neg	1	1
		Sens.	Spec.
		0.95	0.5

Figure 4.3: Confusion matrix comparing midpoint assessment’s final dose projections against ground truth final dose projections. The numbers indicate the total number of OARs that resulted in that particular outcome.

scheme’s final dose estimates are shown in Figure 4.3. The sensitivity of 0.95 means that, out of 19 total OARs which did not end with a dose constraint violation, 18 were correctly predicted as such. The specificity of 0.5 means that, out of two OARs which did end with a dose constraint violation, one of them was correctly predicted as such. This translates to an overall accuracy of $\frac{19}{21} = 0.90$. So, the midpoint assessment scheme correctly determined whether final dose delivered to OARs would violate dose constraints 90% of the time.

The biweekly assessment scheme, after all replans are taken into account, also creates an estimation of the final dose delivered to the patient. Results for the biweekly assessment’s final dose projection, as compared to the ground truth best estimate, are shown in Figure 4.4, whose main result is a sensitivity of 0.86. This value means that, out of 21 times in which the final dose to an OAR didn’t violate its dose constraint, 18 were correctly predicted as such. The lack of a specificity simply means that there were no OARs whose dose constraints were violated at the end of treatment. This also means that the accuracy of the biweekly assessment scheme’s final dose estimations is the same as the sensitivity of 0.86.

		$(D_{\text{truth}})_{\text{bi,end}}$	
		Pos	Neg
$(D_{\text{test}})_{\text{bi,end}}$	Pos	18	0
	Neg	3	0
		Sens.	Spec.
		0.86	N/A

Figure 4.4: Confusion matrix comparing biweekly assessment’s final dose projections against ground truth final dose projections.

4.2 Research Question 2

After either assessment scheme is carried out, it is important to quantify whether the schemes actually achieved the goal of decreasing OAR dose constraint violations. This is done through comparing the total amount of violations seen without adaptive replanning to the total amount of violations seen with adaptive replanning. The results for both the midpoint and biweekly assessment schemes are shown in Figure 4.5. Addition of a midpoint assessment decreased the total amount of OAR dose constraint violations from four to two. This is a 50% decrease in total violations. In addition, for midpoint assessment, the final dose values for the four OARs which experienced dose constraint violations both with and without adaptation in Table 4.1. This table compares the best possible estimation of dose delivered to triggered OARs without ART to the best possible estimation of dose to the same OARs with ART. Then, the difference between the two cases for each truly triggered OAR is displayed. For Patient 1, the act of replanning actually made the dose constraint violation for their larynx worse. This should be weighed against the trade-off of avoiding a trigger to the left parotid gland. In the cases of Patients 2 and 3, one

	Triggers w/o Adaptation	Triggers w/ Adaptation
Midpoint	4	2
Biweekly	4	0

Figure 4.5: Comparison of total end-of-treatment dose constraint violations for the midpoint and biweekly assessment schemes as seen in our three patient cohort.

OAR	Dose w/o Adaptation (cGy)	Dose w/ Adaptation (cGy)	Dose Difference (cGy)
Patient 1:			
Larynx	4064	4150	86
Left Parotid Gland	2000	1869	-131
Patient 2:			
Left Parotid Gland	2864	2566	-298
Patient 3:			
Right Parotid Gland	2323	2182	-141

Table 4.1: Comparison of final delivered dose with and without adaptive replanning when the midpoint assessment scheme was used. The last column shows the dose difference between the two.

dose constraint violation was seen and then avoided when adaptive replanning was used. The biweekly assessment, however, fully eliminates the four violations seen. Each of the four triggers patients experienced without adaptation were avoided when biweekly assessment was used. These results are shown in Table 4.2. This table shows consistent decrease in dose, below dose constraints, for every OAR that was shown to receive unacceptable levels of dose without assessment. So, complete eradication of dose constraint violations and consistent decrease in dose to these OARs is seen. The next section contextualizes the benefits seen when using these assessment schemes with their associated time-cost.

OAR	Dose w/o Adaptation (cGy)	Dose w/ Adaptation (cGy)	Dose Difference (cGy)
Patient 1:			
Larynx	4064	3972	-92
Left Parotid Gland	2000	1980	-20
Patient 2:			
Left Parotid Gland	2864	2572	-292
Patient 3:			
Right Parotid Gland	2323	1894	-429

Table 4.2: Comparison of final delivered dose with and without adaptive replanning when the biweekly assessment scheme was used. The last column shows the dose difference between the two.

Benefit	Midpoint	Biweekly
Final OAR Triggers Decrease	2	4
Average Triggered OAR Dose Decrease	1.19 Gy	2.08 Gy
Average Dose Decrease for all OARs	0.82 Gy	1.48 Gy

Table 4.3: Dosimetric benefits seen when using the midpoint and biweekly assessment schemes on three patients with a total of 21 OARs.

4.3 Research Question 3

Given the information in Table 3.10, we can quantify the relationship between time-cost and dosimetric benefit quite clearly. These results were gathered from the use of midpoint and biweekly assessment for three separate patients and are shown in Table 4.3. These results can then be used to create time-cost-effectiveness estimates for the midpoint and biweekly assessment schemes. These estimates are shown in Table 4.4.

Benefit	Midpoint Time-cost-effectiveness	Biweekly Time-cost-effectiveness
Trigger Decrease	0.92 $\frac{\text{Violations Avoided}}{\text{hr}}$	0.62 $\frac{\text{Violations Avoided}}{\text{hr}}$
Average Triggered OAR Dose Decrease	0.55 $\frac{\text{Gy Reduction}}{\text{hr}}$	0.32 $\frac{\text{Gy Reduction}}{\text{hr}}$
Average Dose Decrease for all OARs	0.38 $\frac{\text{Gy Reduction}}{\text{hr}}$	0.23 $\frac{\text{Gy Reduction}}{\text{hr}}$

Table 4.4: Time-cost-effectiveness of midpoint and biweekly assessment schemes using three separate dosimetric benefit metrics.

4.4 Discussion

Research Question 1 asks about the midpoint and biweekly assessment schemes' accuracy when predicting replan necessity and predicting final dose delivered to patients. Testing both schemes on their ability to detect necessary replans led to virtually identical results between the two. This is slightly counter-intuitive since biweekly assessment takes more anatomical information into account. Though not expected, this result implies that the use of either assessment scheme results in accuracy of over 90%. So, the baseline answer for Research Question 1 is yes, both schemes determine replan necessity and predict final delivered dose with approximately 90% accuracy. It should be noted that this performance may be the product of a small cohort size. Any statistically informed conclusions can only be made with a larger cohort size.

It could be argued that sensitivity may be more important than accuracy and specificity in the case of detecting replan necessity. The rationale behind this assertion is built on the potential consequences of a poor value of sensitivity or specificity. With poor sensitivity, OARs would end up being overdosed and going unnoticed. This likely translates to worse clinical outcomes and decreased quality of care. With poor specificity, OARs with appropriate dose levels trigger replans unnecessarily. This may deter clinicians from using these assessment schemes since unnecessary replans are a waste of time, but regardless, the downside of decreased quality of care outweighs time loss under most circumstances. This stance on sensitivity's importance gives a more nuanced answer to part of Research Question 1 as it describes the schemes' ability to catch OAR triggers which already exist. This is opposed to the answer accuracy and specificity may give, which is heavily influenced by the scheme correctly determining that OARs don't need a re-optimized plan. Using only sensitivity for this result, identical performance between assessment schemes is still seen. However, the ability to detect necessary replans is only 75%. With this taken into account, though, both assessment schemes still give clinicians a chance to improve patient care which would

have violated dose constraints without adaptive radiotherapy. After all, catching 75% of necessary replans is better than the 0% which is guaranteed when not using adaptive radiotherapy.

Testing both schemes on their ability to predict final delivered dose gives clinicians a better idea of how much they can trust the results of their adaptive replanning. This analysis showed more of a difference between the two assessments schemes in that midpoint assessment's sensitivity was higher than biweekly assessment. This means that midpoint assessment more accurately confirmed the acceptable dose levels of safely irradiated OARs when compared with biweekly assessment.

However, it is interesting to note that midpoint assessment's specificity of 0.5 shows that only 50% of OARs which truly received an unacceptable level of dose, even with replans, were correctly flagged as such. This is compared with the lack of specificity in the case of the biweekly assessment. What this indicates is that there were no data points in the right column of the confusion matrix in Figure 4.4. It follows, then, that there were no cases in which an OAR was reported to be overdosed, nor were there cases in which any OAR was truly overdosed. So, while the specificity of biweekly's final dose projection capability is undefined, it hints at the biweekly assessment scheme's superior performance in the avoidance of true OAR overdoses.

Additionally, just as the sensitivity was more important for detecting replan necessity, it could be argued that specificity, or lack thereof, is the more important descriptor for the estimation of final delivered dose. A low specificity, in this case, corresponds to an inability to accurately pinpoint overdosed OARs. This means that potential measures which may mitigate symptoms that come along with overdosed OARs would not be used in a timely manner. In other words, a low specificity negates the chance for using proactive treatment for toxicity mitigation. Low sensitivity, on the other hand, translates to the overuse of proactive treatment for safely irradiated

		Detect Replan Necessity	Predict Final Dose
Midpoint	Sensitivity	0.75	0.95
	Specificity	0.94	0.5
Biweekly	Sensitivity	0.75	0.86
	Specificity	0.95	N/A

Table 4.5: Summary of results relevant to Research Question 1. Both sensitivity and specificity for the schemes’ ability to detect replan necessity and predict final dose are listed.

OARs which may not need it. It may also cause unnecessary alarm for patients if these incorrect reports of OAR overdose are shared with them. While those outcomes are undesirable, the true presence of overdosed OARs with no knowledge of them may be worse.

Overall, the results show that both schemes detect replan necessity with virtually identical accuracy. Additionally, the results show that midpoint assessment may suffer in its ability to accurately determine whether OARs are overdosed after the completion of treatment. So, in terms of overall accuracy, both the midpoint and biweekly assessment schemes correctly determine the necessity of replans above 90% of the time. The midpoint and biweekly assessment schemes also accurately determined whether OARs’ final doses violated their dose constraints 90% and 86% of the time, respectively. A summary of the results relevant to Research Question 1 is shown in Table 4.5.

Research Question 2 asks about the dosimetric benefit when using either the midpoint or biweekly assessment scheme. When looking at reduction of OAR dose constraint violations, it is clear that biweekly assessment outperforms midpoint assessment. Additionally, in terms of actual dose reduction, two patients had a dose reduction to triggered OARs of 2.98 and 1.41 Gy. However, in Table 4.1, adaptive replanning actually increased the dose delivered to the triggered larynx for one patient. This seems to be the case because the midpoint assessment missed the larynx trigger and instead only triggered on the patient’s left parotid gland. Since the as-

assessment only indicated the left parotid gland trigger, re-optimization focused on the left parotid gland's dose constraint and treated the larynx as if it didn't need any extra attention.

This proves to be an issue in a few ways. First, the larynx's dose constraint is still violated, and the violation even worsens after replanning. Second, the magnitude of the violation to the larynx is larger than that of the left parotid gland. In fact, the left parotid gland only exactly reached the trigger dose of 2000 cGy with no adaptation, according to our best estimate. Intuitively, a good assessment scheme should more often catch violations which are larger compared to small violations. This may call into question the usefulness of an exact dosimetric threshold in the assessment criteria. It may be useful, then, to consider threshold criteria which lean more towards biological effect as opposed to a physical dose cutoff.

Overall, it was demonstrated that the midpoint and biweekly assessment schemes both decreased total instances in which OARs received unacceptable levels of dose by 50 and 100%, respectively. They both decreased the dose delivered to triggered OARs by an average of 1.19 and 2.08 Gy, respectively. So, both assessment schemes proved they provide dosimetric benefit to patients, and biweekly showed consistently superior benefits.

The third research question asks about the time-cost-effectiveness of both assessment schemes. Three different ways to quantify dosimetric benefit were used, and in each case, the midpoint assessment scheme was shown to be more cost-effective. Percent differences between midpoint and biweekly assessments' time-cost-effectiveness varied from 32.6 to 41.8%, depending on the dosimetric benefit metric used.

While this study doesn't have a large enough cohort size to come to statistically meaningful conclusions, this set of time-cost-effectiveness comparisons may give clinicians a better idea of how much benefit they can reasonably expect given a certain amount of available resources. Additionally, different metrics for dosimetric bene-

fit may be a major decision point for clinicians, depending on circumstance. For instance, a physician may be much more interested in reducing dose to all healthy tissue as opposed to just triggered OARs.

More importantly, though, these time-cost-effectiveness analyses should be used in conjunction with the results used to answer our other two research questions. From the beginning of this study, the overall goal was to create a framework for evaluation of different assessment schemes used for offline adaptive radiotherapy. The three research questions that were posed give a clear view of the overall performance of a given assessment scheme. In layman's terms, the first research question addressed how trustworthy each assessment scheme is. Depending on the clinician, these schemes may or may not achieve an acceptable level of overall accuracy. What is more important, though, is that our evaluation framework is able to present the accuracy of these schemes clearly.

The second research question addressed the benefit one can expect when using a given assessment scheme. More specifically, the major result for research question two is the overall decrease in OAR dose constraint violations by 50 and 100% for midpoint and biweekly assessment. This provides a simple measure of how much benefit can be expected when using a given assessment scheme, regardless of the amount of time and work involved. Clinics with ample resources may hold this evaluation in particularly high regard since this violation avoidance is in line with the objective of adaptive radiotherapy as a whole.

The third research question provided information on each scheme's time-cost-effectiveness. In other words, this question informs clinicians about how much dosimetric benefit can be expected per given time investment. This information may be particularly valuable to clinics with low resources, particularly in terms of physicist time. Additionally, the use of three different dosimetric benefit metrics allows

for multiple perspectives to be taken into account when evaluating an assessment scheme.

As an additional note, the time-cost-effectiveness results should be fairly robust when considering various computational environments and software versions. This is the case since the comparison of time-cost between assessment schemes was done in the same computational environment. So, the comparison is relative and should translate to other environments and softwares.

Overall, the evaluation framework created is able to answer multiple questions relevant to the performance of an adaptive assessment scheme. With this set of analyses defined and used for a small initial cohort, it stands to reason that larger-scale trials can be designed to find more statistically powerful results.

4.5 Study Limitations

There are a few points of this study which may be viewed as limitations. First, with a cohort size of three patients, it may be difficult to glean statistically meaningful results from our evaluations. However, the main goal of the study was to test whether these methods of assessment *evaluation* could provide useful data to clinicians who may be deciding whether to implement this offline ART workflow. This goal has been achieved through the framework built by the three research questions posed. Additionally, the results of this work point towards the ability to design larger-scale studies, even multi-institutionally, and evaluate the various assessment schemes through our methods as part of this design.

An additional limitation of our study was the choice of trigger threshold which signaled the necessity of a replan. Choosing an exact dosimetric threshold was an easy choice for a first step, but may not correlate exactly with clinical outcomes such as toxicity. For instance, an OAR which is 1 cGy over its trigger dose necessitates all

the work involved in the replanning workflow. This work may bring the dose to this OAR below its trigger dose but not translate to any meaningful clinical benefit. It follows that a trigger emphasizing clinical/biological benefit may prove more useful for this adaptive workflow.

Another limitation in our study is the inherent uncertainty associated with our ground truth estimations. This uncertainty comes both from deformation uncertainty and the fact that ground truth only uses one image per week. In the ideal case, there would be one aCT for each day of treatment, though this is impractical. Also, the best case scenario for mitigation of deformation uncertainty comes from rigorous QA protocols.

4.6 Future Work

A natural next step for this project would involve the switching of trigger values to biological quantities as opposed to dosimetric ones. For instance, the use of normal tissue complication probability (NTCP), or a relative change in NTCP, may prove more useful for toxicity mitigation. This, however, can best be demonstrated with a true prospective study. Further down the line, large-scale clinical trials would prove useful for the determination of this workflow and either assessment scheme's benefit for H&N cancer patients.

When discussing the use of new triggers, it may be helpful to consider the use of triggers for target volumes along with OARs. For instance, if a PTV dose constraint is that 100% of the volume receives 95% of the prescription dose, replans would be triggered if that constraint is violated. Even though replans created through this study's protocol are normalized identically to the original plan, many H&N plans have multiple PTVs. So, PTVs not used for normalization may end up being use-

ful for triggering replans. Then, extending to biological triggers, the tumor control probability (TCP) could also be used.

An additional step for future work is the development of a machine learning model which predicts if and when adaptive replanning should take place based on images from early in a patient's treatment course. This aligns with the spirit of bringing about the most clinical benefit for the least amount of work that we are interested in. In this same vein, the use of automatic planning and auto-contouring may speed up the process as well.

Chapter 5

Conclusions

The prevalence and severity of head and neck (H&N) cancer worldwide along with anatomical changes commonly seen throughout treatment highlight the importance of active research in the field of adaptive radiotherapy (ART). The implementation of ART has led to clear improvement in patient care by ensuring that a patient's anatomical changes are accounted for throughout their treatment. However, one of the many barriers to its widespread implementation is the time required to perform all required steps in the ART process. While the benefits are enticing, many clinics might not have the resources available for regular use of ART. Clinics may also not have sufficient information on different modes of implementation, which can prevent clinics from considering an offline ART program.

The resource problem is partially mitigated through the development of a simple adaptive workflow which can be used at varying frequencies according to a clinic's available resources. This workflow, developed by Hyun et al. as described in [Appendix A](#), uses Varian's Velocity software to perform assessments which determine whether a patient's treatment plan requires re-optimization. The ease of implementation and the versatility of using the workflow according to a clinic's means may lead to the possibility of large-scale trials for this mode of offline adaptive radiotherapy.

In order to combat the information issue, this study has worked to create an evaluation framework which can quickly and concisely give interested clinicians performance characteristics of different methods of ART. Specifically, this framework provides information on the trustworthiness, dosimetric benefit, and time-cost-effectiveness of certain modes of ART. Our study applied this evaluation framework to the midpoint and biweekly assessment schemes used for offline ART and demonstrated the usefulness of the data gathered through it.

Through investigating three patients, it was shown that both midpoint and biweekly assessment performed virtually identically when detecting whether replans needed to be performed. In both schemes' cases, 75% of instances where organs-at-risk (OARs) were receiving too much dose were flagged. Both schemes performed well when estimating the final dose delivered to a patient as well. The midpoint and biweekly schemes correctly determined whether OARs received too much dose after adaptive replanning 90% and 86% of the time respectively. These results show two trustworthy assessment schemes in action.

It was also shown that both schemes provided an overall dosimetric benefit. Midpoint and biweekly assessment led to a decrease in OAR dose constraint violations for each patient. In fact, the midpoint and biweekly schemes decreased overall dose constraint violations from four to two and zero, respectively. Additionally, when looking at OARs which would have been triggered with no adaptive replanning, midpoint assessment decreased their dose by an average of 1.19 Gy. Biweekly did the same by an average of 2.08 Gy. Both assessment schemes clearly improve the overall dose sparing for OARs which would have received too much dose otherwise.

Lastly, this framework provides information on the time-cost-effectiveness of both assessment schemes. The associated benefit of either scheme was quantified a few different ways, which allows for clinicians with varying priorities to accurately weigh their options. Time-cost-effectiveness was quantified through a decrease in overall

OAR dose constraint violations, the average dose decrease experienced by a triggered OAR, and the average dose decrease for all OARs considered. In each case, it was shown that the midpoint assessment scheme was more time-cost-effective.

As has been demonstrated, the creation and use of this framework provides a multifaceted view of the performance characteristics of two different adaptive assessment schemes. It also allows for clinicians to make informed decisions on their clinics' use of a certain assessment scheme. With that being said, certain alterations to the assessment schemes and future work also show much promise. First, the necessity of a replan could be based on a more biological metric. For instance, the use of normal tissue complication probability (NTCP) and its change over time may serve as a replan trigger which is expected to correlate more directly with clinical outcome. Additionally, dosimetric or biological emphasis on target coverage in addition to healthy tissue sparing may give a more complete view of a plan's performance with changing anatomy. This work has laid an important foundation that, combined with these alterations, may help increase the accessibility of H&N adaptive radiotherapy.

Appendix A

Adaptive Workflow

Megan Hyun, along with other UNMC physicists, have developed the following protocol for offline adaptive radiotherapy of the head and neck. The process is carried out through six steps which create an estimation of the true dose delivered to a patient during their treatment course. This appendix is referenced in multiple chapters of this study.

1. **Export Data:** Export the patient's original CT image, treatment plan, structure set (contours), and relevant CBCT images from the day of treatment from the Eclipse Treatment Planning System to Velocity.
2. **Generate Images:** Create adaptive CT images (aCTs) through the use of deformable image registration between the patient's simulation CT image and their relevant CBCT image. This process maps the reliable tissue density information from the simulation image on to the updated anatomy seen in the CBCT image. This allows for reliable dose calculation over the new anatomy.
3. **Evaluate Deformations:** After the dose accumulation is performed, Velocity reports a set of Jacobian determinants which can be analyzed for each deformation. This allows for the confirmation of the deformations' quality.

4. **Calculate Dose:** Export the newly created aCTs and their corresponding treatment plans to Eclipse and calculate dose.
5. **Accumulate Dose:** Export the resulting dose distributions and aCTs back to Velocity and accumulate dose from all relevant distributions. This means that each individual dose distribution is deformed back to the simulation CT image and a composite dose distribution is formed.
6. **Analyze Dose:** Compare dose points for various OARs and targets from the true generated composite dose to those from the planned dose distribution.

Appendix B

aCT Creation Workflow

This appendix describes the protocol used to create adaptive CT (aCT) images for offline adaptive radiotherapy. The majority of this process takes place in Velocity, Varian's image registration/comparison software. Any clinic with access to Velocity can create aCTs using this protocol. This appendix is referenced in Chapter 3.

Adaptive Dose Assessment Workflow

Edited for HN pilot redux 11/11/22 MAH

This workflow assumes some basic familiarity with Aria and Eclipse

1. **Export** the patient data to Velocity

Go to the Import/Export workspace in Aria, and select the Velocity export filter.

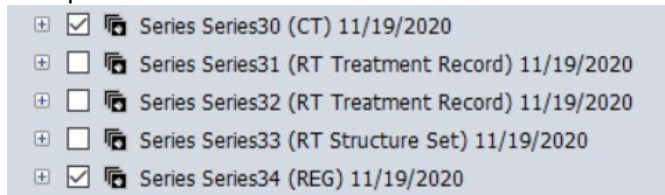


Deselect everything in the list.

Under the current course, select the patient's current treatment plan (if they have multiple, select the one treated on the days you want to analyze)



Under the DICOM folder, locate the study containing the treatment records and images, and select the CT and REG objects from the desired dates. They will probably not be listed in chronological order. You may select everything in the study if you like, it will just take some time to export.



Choose "Clear Structure Selection" and hit the blue arrow to continue

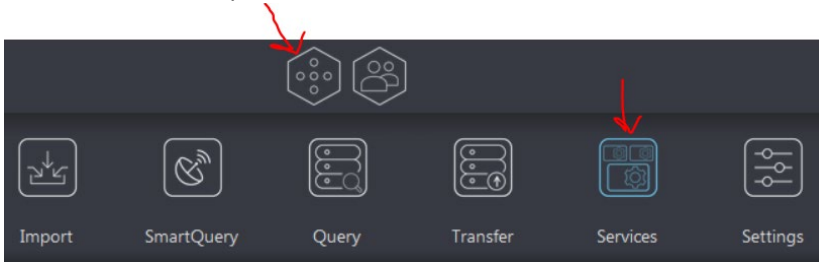


If the export was successful, it will tell you the number of images transferred in **green font**

Modality	Succeeded	Failed
CT	88	0
Structure Set	1	0
Plan	1	0
RT Dose	1	0
Total	91	0

- Log into **Velocity** and ensure the data has imported. Open the patient


This may take some time, so be patient. You can view the import progress in the Services part of the menu at the top

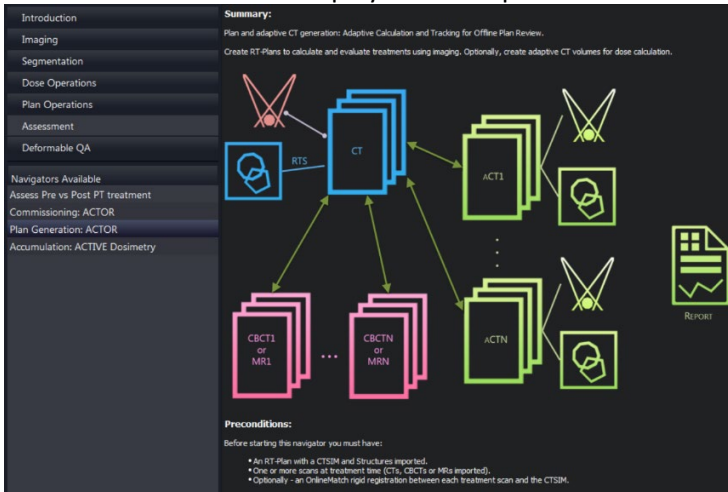


Import Service - Last 25 entries for the past 7 days		Auto Registration Service	Auto Segment Service	Send Service	Query Retrieve Move Se	
	Receive date	Calling AETitle	Import date	Patient Name	Patient ID	Message
✓	2020/06/16 14:06:11	VMSFSD		[REDACTED]	[REDACTED]	Import in progress
✓	2020/06/15 15:22:14	ALI_QUERY_S...	2020/06/15 15:22:21	[REDACTED]	[REDACTED]	Imported at 2020/06/15 15:22:21

Once the patient's data is imported, there should be NO lock symbol next to their name in the patient list on the home screen. Double click the patient entry to open them.

- Run the **Navigator**

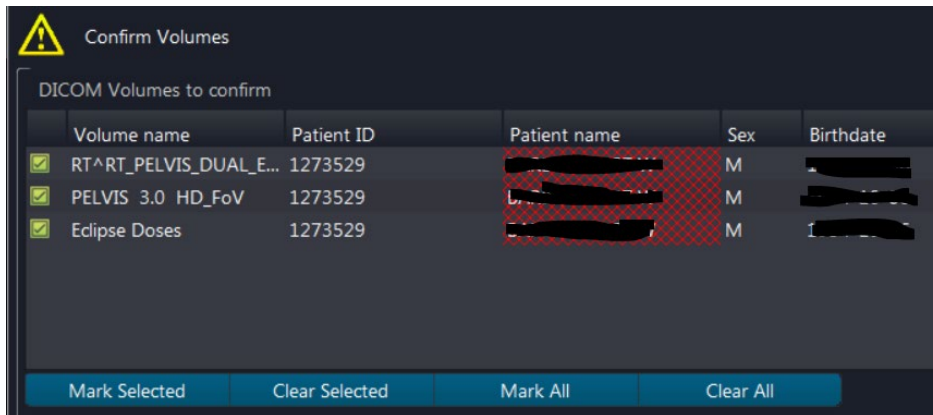
Click the Navigator symbol at the top  and choose "Assessment" then "Plan Generation: ACTOR". The screen will display all the required data for starting this Navigator. Hit Start



If it prompts you to confirm the imported volumes (not common), cancel the Navigator and click

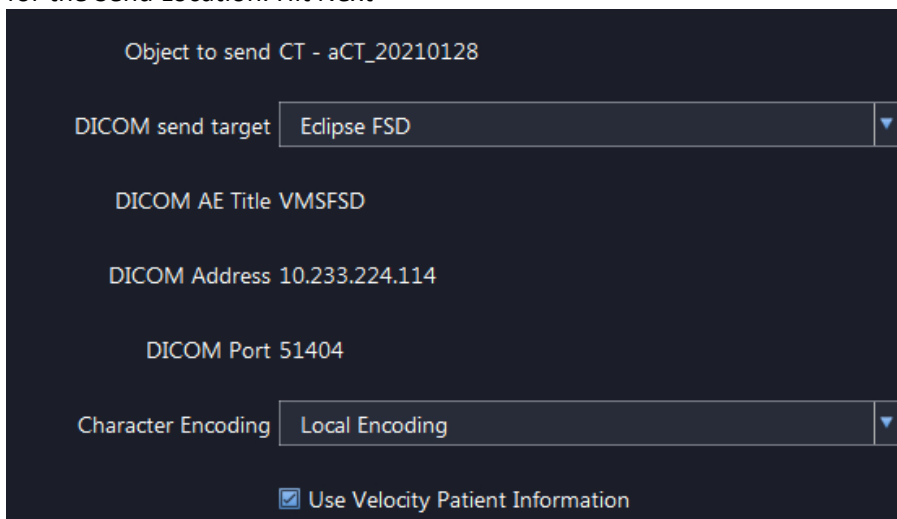


the Volumes icon and double click one of the red items. In the pop up window, click Mark All then OK. Re-enter the Navigator.



The Navigator will display each step and prompt you for information along the way.

First, select the send target. Select “Send from Grid” in the drop-down. Use the settings below for the Send Location. Hit Next

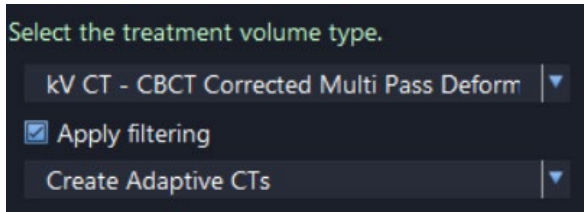


Second, select the plan you exported and hit Next


Third, select structures. Expand the structure set with the plus sign and highlight all structures in the list. Select the whole bar, not the green check box (the box is for display only). Hit next

Fourth, select up to 6 structures for QA of the DIR. These should be well-visualized structures like bony anatomy, spinal cord, bladder, NOT target structures. The BODY must be selected. For H&N, I recommend selecting at least the BODY, Cord, Esophagus, and individual Parotid Glands. You don’t need to select the maximum number. Hit next

Fifth, select the treatment volume type. The defaults should be selected (shown below) and hit next



Sixth, select the CBCT(s) for analysis. The names won't make sense (CBCTs usually start with "RT^RT", so go by the date of the image (which you double check in RT Summary in Aria, if you don't remember which CBCT dates you exported). Hit next

Seventh, select the registration from the day of treatment. Make sure the drop-down at the top has "Match name (choose the most recent registration)" selected. Turn on the ROI by clicking on the open box icon , and verify it covers the entire CBCT volume. Hit next

Eighth, run the workflow (hit Next). This process can take some time, especially if you are analyzing more than 1 CBCT (about 5 minutes per CBCT). There may be moments where the screen goes blank or the window says "not responding", but it is probably still going. If the system crashes, it will auto-save its progress.

Once it's done you'll see a Finished button at the bottom with green check marks showing all the completed steps

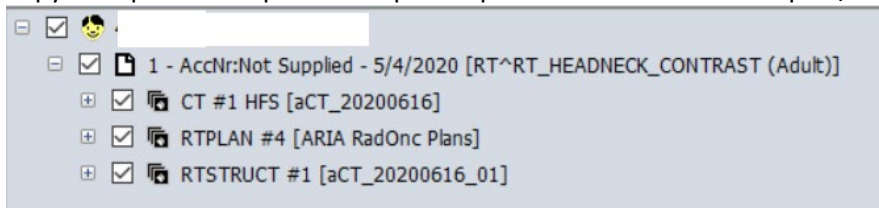


Data about the DIR will be displayed. You can use Velocity's tools to evaluate the DIR as well.

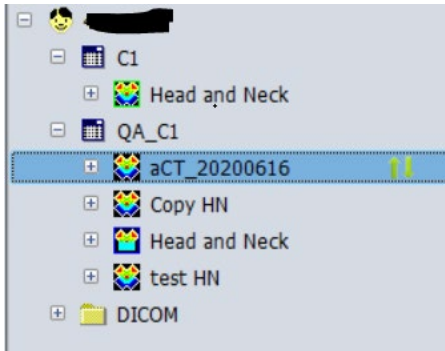
4. Import the aCT(s) into Eclipse

Before importing, **please create a new Research course specifically for this re-analysis**. These patients will likely already have aCTs in their QA course or an existing research course. Please name the new course something like "Research_pilot2022". Make sure the course status is Active and not Completed.

The Navigator should have automatically exported the adapted CTs (aCTs), structure set, and a copy or copies of the plan to Eclipse. Import these items in the Import/Export workspace



Move the aCT object into the course you created. The green arrows indicate an object that can be dragged between courses

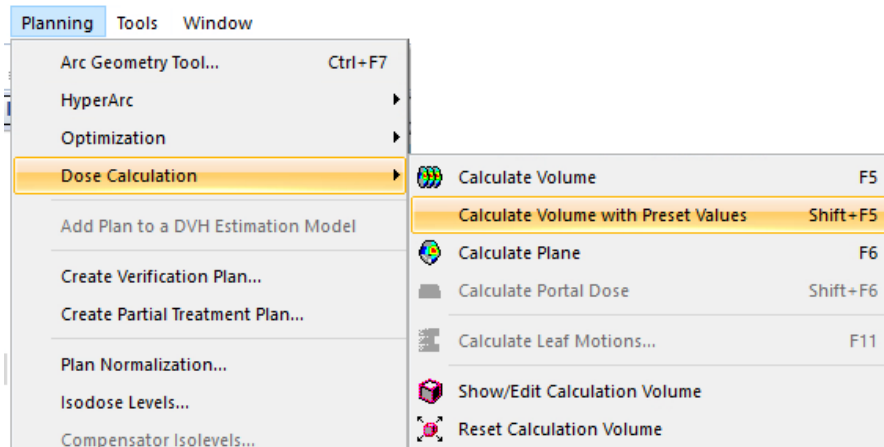


If, for some reason, you have trouble importing the plan and/or structures (as sometimes happens), try to import the aCTs alone. These will import into the DICOM folder.

5. Calculate plans

If everything in Step 5 was successful, you should only need to calculate each of the plans imported from Velocity. Leave the original fractionation.

If you were unable to import plans/structures, copy the original plan onto each aCT and place these new plans into a QA or Research course. Calculate them one by one using preset values (the MUs from the original plan). You may be prompted to confirm the CT scanner – make sure to confirm you’re selecting the correct scanner from the original sim.



Select Imaging Device ×

The image series is not connected to an imaging device defined in the system.


Select the imaging device:

Original imaging device

Manufacturer:
 Model name:
 Device serial number:

Selected imaging device

Manufacturer:
 Model name:
 Device serial number:

 The selected scanner does not match the original one. You should create a corresponding CT scanner in the RT Administration with the scanner specific calibration curves.

Calculate Dose with Preset Values ×

Plan

Preset Values

Field ID	Preset		Reference Point		MU
	Ref. Point	MU	ID	Dose [cGy]	
01	<input type="checkbox"/>	<input checked="" type="checkbox"/>			289.300
02	<input type="checkbox"/>	<input checked="" type="checkbox"/>			253.500

Absolute Dosimetry

Number of Fractions
 Dose per Fraction cGy
 Treatment Percentage %
 Plan Normalization Value %

Enter ClearCheck to obtain the dose metrics to enter into the spreadsheet. Copy the original ClearCheck template and make a new one labeled Research. You will likely need to update the expanded cord structure name for some reason. Enter the values in the spreadsheet Megan sent out. Megan will add the original doses for comparison.

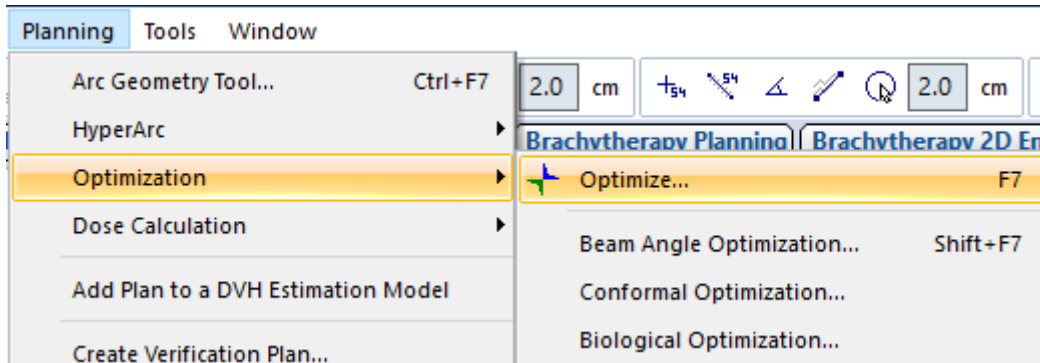
Appendix C

Replanning Instructions

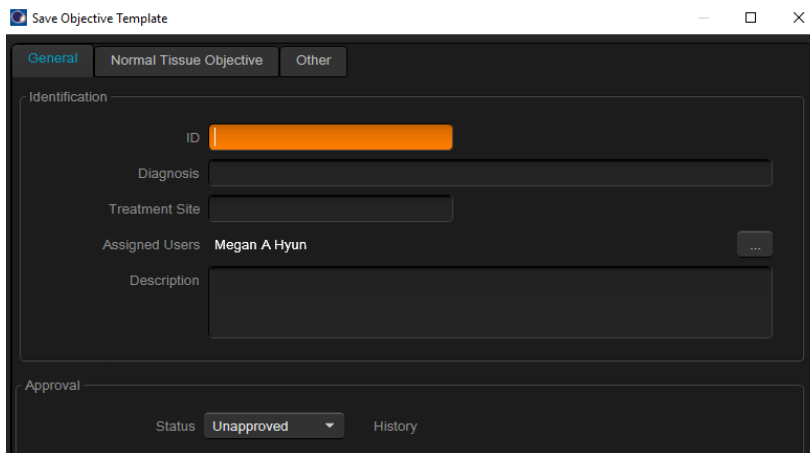
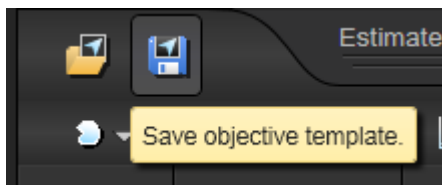
This appendix contains instructions for how to go about the replanning process following the completion of the workflow outlined in [Appendix A](#). This appendix is referenced in [Chapter 3](#).

Timepoint Study Re-Planning Instructions

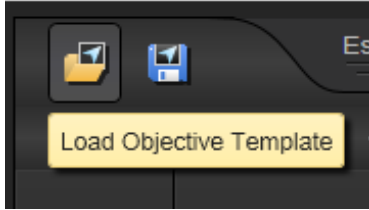
1. Copy original plan into QA/Research course
2. Open Optimizer for that plan



3. Make note of the original plan objectives (take screenshots of optimizer AND clearcheck) so you can ensure they're imported correctly and to guide replanning
4. Make note of which plan goals (in clearcheck) are violated/triggered for the assessment plan so you know what to focus on in the re-optimization
5. Save patient's objectives as a template, add a description



6. Pull up the aCT used for analysis (e.g., week 1 aCT for weekly assessment with an immediate trigger), and open the optimizer



7. Only check “unapproved” and find the template you saved

The screenshot shows a window titled "Select Objective Template". At the top, there are filter buttons for "Unapproved", "Reviewed", "Approved", and "Retired", with "Unapproved" selected. A search bar is also present. Below the filters is a table of objective templates.

ID	Approval	Users	Diagnosis	Treatment Site	Description
SDSFF	Unapproved	Patel, Purvi			
shen	Unapproved	Rosales, Erni R			
solberg	Unapproved	Granatowicz, Andrew D			
SRS_dz20	Unapproved	dzheng			
steinkraus	Unapproved	agranatowicz			
steinkraus2	Unapproved	agranatowicz			
SW_9Gyx3	Unapproved	Shuo Wang			
T1_HN_objectives	Unapproved	Megan A Hyun			Objectives from H&N plan timepoint study patient T1
test	Unapproved	Physicist			

Below the table is an "Objectives" section with a table of parameters:

Template Structure ID	Plan Structure ID	Color	Type	Vol [%]	Dose [cGy]	gEUD a	Priority
Cord	Cord	Green					
Cord exp	Cord exp1	Green					
			Upper	0.0	4500		75
			Upper	0.0	4000		75
GTVp	GTVp	Orange					
			Lower	100.0	7000		100
Larynx	Larynx	Pink					
			Upper	16.4	4812		40
			Upper	0.0	6200		100

To the right of the objectives table is a DVH plot showing Volume [%] on the y-axis (0.0 to 100.0) and Dose [cGy] on the x-axis (0 to 6000). The plot shows several curves with colored markers (blue, yellow, green, pink, red) at various dose points. Vertical dashed lines are drawn at 28.6, 57.1, and 85.7 cGy. At the bottom of the window are "Select" and "Cancel" buttons.

8. Verify the loaded objectives against the original plan (ideally through a screen capture of the originals)

	ID/Type	cm ³	Vol [%]	Dose[cGy]	Actual Dose [cGy]	Priority	gEUD a			
	GTVp	31.5								
	Lower	31.5	100.0	7000	7000	100				x
	pPTV56	120.0								
	Upper	0.0	0.0	6075	6252	200				x
	Upper	2.4	2.0	5800	5968	120				x
	Lower	115.2	96.0	5600	5483	220				x
	Lower	120.0	100.0	5320	3601	0				x
	pPTV60	344.1								
	Upper	0.0	0.0	7010	7286	150				x
	Upper	3.4	1.0	6500	6919	100				x
	Lower	344.1	100.0	5700	4850	75				x
	Lower	330.3	96.0	6000	5980	100				x
	PTV_70H&N	125.2								
	Upper	0.0	0.0	7210	7427	130				x
	Lower	120.2	96.0	7000	6975	200				x
	Cord	21.9								
	Upper	0.0	0.0	4000	4181	75				x
	Cord exp	76.4								

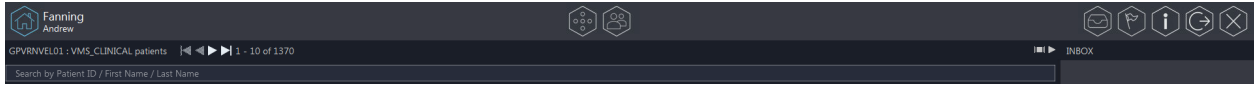
	ID/Type	cm ³	Vol [%]	Dose[cGy]	Actual Dose [cGy]	Priority	gEUD a				
	Cord exp	76.4									
	Upper	0.0	0.0	4500	4787	75		x			
	Larynx	19.4									
	Upper	0.0	0.0	6200	6107	100		x			
	Upper	3.2	16.4	4812	4677	40		x			
	Mean			3850	3918	40		x			
	Upper gEUD			4900	5385	30	40.0	x			
	Mandible	68.0									
	Upper	0.0	0.0	7100	7137	35		x			
	Oral Cavity	127.6									
	Upper	0.0	0.0	6300	6303	45		x			
	Mean			1048	1099	25		x			
	Parotid_L	22.6									
	Mean			1850	1875	40		x			
	Parotid_R	18.6									
	Mean			1850	1933	45		x			


Appendix D

Dose Accumulation Instructions

This appendix provides instructions for the "dose accumulation" step of the workflow described in Appendix [A](#). From the clinical perspective, this is only used when multiple weeks' aCTs are employed. It is most often used from the research perspective to accumulate dose over all available anatomical information. This appendix is referenced in Chapter [3](#).

1. Log into Velocity and type in the MRN of the patient you're working with. When patient pops up, double click on them.



2. At the top of the screen, you will find a button that looks like this.  Click this button to open up the Navigator menu.
3. In the Navigator menu, click Assessment → Accumulation: ACTIVE Dosimetry → Start

Introduction

Imaging

Segmentation

Dose Operations

Plan Operations

Assessment

Deformable QA

Navigators Available

Assess Pre vs Post PT treatment

Commissioning: ACTOR

Plan Generation: ACTOR

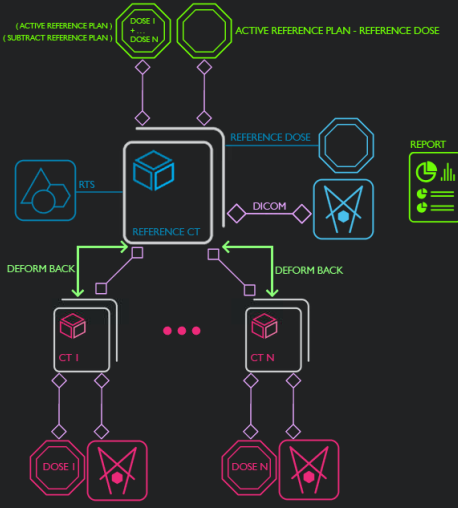
Single Plan Generation: ACTOR

Accumulation: ACTIVE Dosimetry

Summary:

Treatment Accumulation: Adaptive Calculation and Tracking for In-Vivo Estimation of dosimetry.

Accumulate estimated doses using image registrations. Optionally, compare to a reference dose.



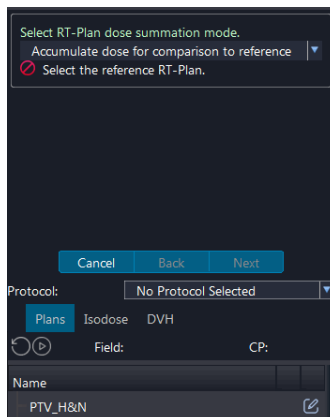
Preconditions:

Before starting this navigator you must have:

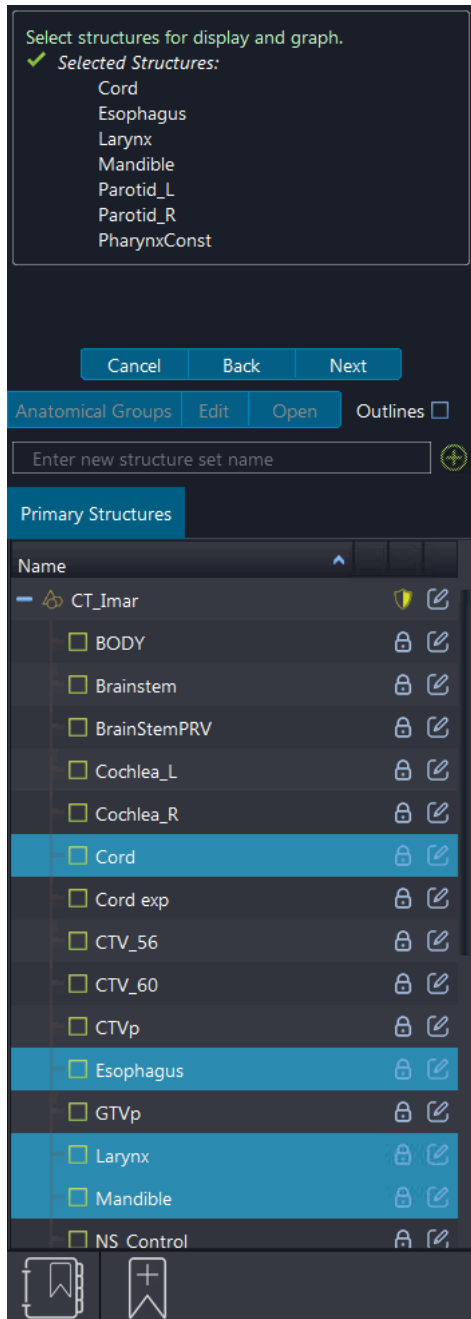
- An original RT_Plan with CTSIM and Structures imported.
- One or more Adaptive CT Dose Volumes for one or more Adaptive CTs.
- A Deform back registration between the original CT and the synthetic CT.

Cancel
Start

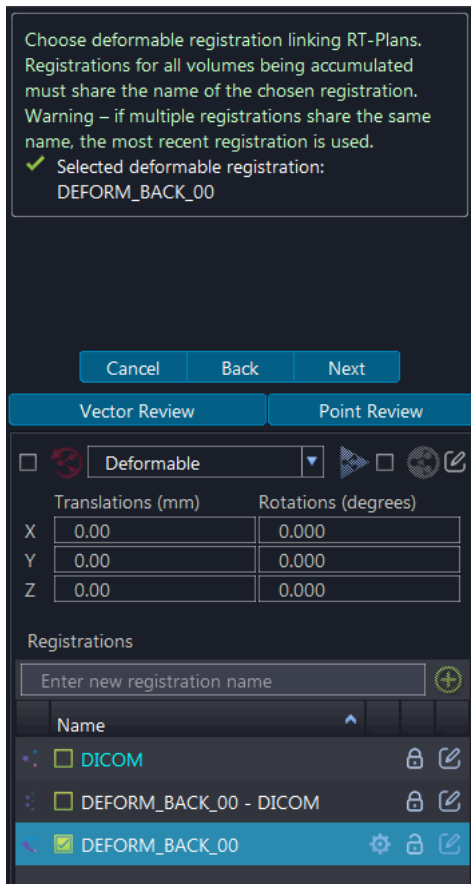
4. The Navigator will ask you to “Select RT-Plan dose summation mode.” You should select **Accumulate dose for comparison to reference**. Then, you will select the reference RT-Plan. This is the patient’s original treatment plan. This often has a name similar to “PTV_H&N”.



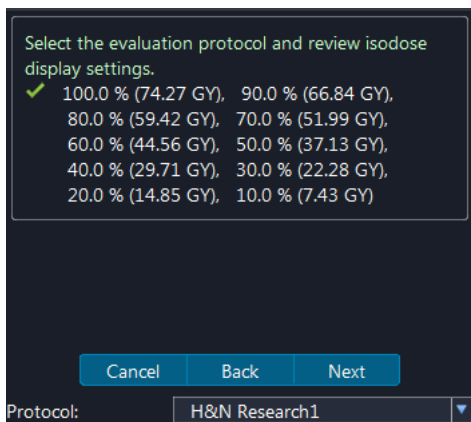
- Now, you must select structures for display and graph. These should be the OARs you wish to investigate. Open the list of contours from the appropriate imaging set and select all those you wish to use. For our purposes, this was spinal cord, mandible, left & right parotids, pharynx, esophagus, and larynx.



- Next, you are prompted to “Select RT-Plans to accumulate.” Here, choose all relevant plans from the provided list.
- Now, “Choose deformable registration linking RT-Plans.” Here, always choose **DEFORM_BACK_00**.



8. Next, you must “Select RT-Plan scaling method.” The two available choices are “Equal scaling: average method” and “External scaling performed: Direct sum method.” If all plans you wish to sum are kept at the length of the full treatment course, (e.g., 35 fx) choose the average method. If you’ve changed the number of fractions in a plan externally, choose the direct sum method.
9. Now, you must “Select the evaluation protocol and review isodose display settings.” If you’ve created a protocol for evaluation, select it in the protocol dropdown menu. You are also able to create a protocol from this dropdown menu, if desired.



10. All that is left is to “Create Accumulated Dose file” by clicking Next. Once accumulation is completed, a report will be displayed showing the accumulated dose to the selected structures.

It should be noted that the dose values from this process are shown under the **ACTIVE_REF_PLAN_NAME** section. So, in this case, the relevant doses are displayed under the ACTIVE_PTV_H&N section.

Selected Protocol: H&N Research1. Prescription: 70 GY									
Targets					Organs at Risk				
					Cord: Max Dose <= 45 GY Esophagus: Mean Dose <= 30 GY Larynx: Mean Dose <= 40 GY Mandible: Max Dose <= 72 GY Parotid_L: Mean Dose <= 20 GY Parotid_R: Mean Dose <= 20 GY PharynxConst: Mean Dose <= 55 GY				
Eclipse Doses Histogram Statistics									
	Name	# Bins	Type	Volume (cm ³)	Min/Mean/Max			Constraints	
■	Cord	1024	ORGAN	22.2 cm ³	1.96 GY	29.27 GY	41.81 GY	✓	Max Dose <= 45 GY (41.81 GY)
■	Esophagus	4096	ORGAN	30.7 cm ³	0.00 GY	6.97 GY	56.07 GY	✓	Mean Dose <= 30 GY (6.97 GY)
■	Larynx	1024	ORGAN	20.7 cm ³	25.02 GY	39.77 GY	60.62 GY	✓	Mean Dose <= 40 GY (39.77 GY)
■	Mandible	1024	ORGAN	71.2 cm ³	1.96 GY	32.10 GY	71.13 GY	✓	Max Dose <= 72 GY (71.13 GY)
■	Parotid_L	1024	ORGAN	23.2 cm ³	2.77 GY	18.79 GY	65.11 GY	✓	Mean Dose <= 20 GY (18.79 GY)
■	Parotid_R	1024	ORGAN	19.3 cm ³	3.41 GY	19.10 GY	64.97 GY	✓	Mean Dose <= 20 GY (19.10 GY)
■	PharynxConst	1024	ORGAN	11.2 cm ³	24.47 GY	49.66 GY	73.62 GY	✓	Mean Dose <= 55 GY (49.66 GY)
ACTIVE_PTV_H&N Histogram Statistics									
	Name	# Bins	Type	Volume (cm ³)	Min/Mean/Max			Constraints	
■	Cord	1024	ORGAN	22.2 cm ³	1.88 GY	31.10 GY	41.84 GY	✓	Max Dose <= 45 GY (41.84 GY)
■	Esophagus	4096	ORGAN	30.7 cm ³	0.00 GY	6.58 GY	53.82 GY	✓	Mean Dose <= 30 GY (6.58 GY)
■	Larynx	1024	ORGAN	20.7 cm ³	26.16 GY	39.72 GY	60.90 GY	✓	Mean Dose <= 40 GY (39.72 GY)
■	Mandible	1024	ORGAN	71.2 cm ³	2.15 GY	32.71 GY	70.97 GY	✓	Max Dose <= 72 GY (70.97 GY)
■	Parotid_L	1024	ORGAN	23.2 cm ³	3.31 GY	19.80 GY	65.29 GY	✓	Mean Dose <= 20 GY (19.80 GY)
■	Parotid_R	1024	ORGAN	19.3 cm ³	3.69 GY	20.31 GY	63.43 GY	⊗	Mean Dose <= 20 GY (20.31 GY)
■	PharynxConst	1024	ORGAN	11.2 cm ³	26.56 GY	50.31 GY	73.44 GY	✓	Mean Dose <= 55 GY (50.31 GY)

Appendix E

Copyright Permissions



THESIS COPYRIGHT PERMISSION FORM

Title(s) of the Image(s): Terese Winslow LLC owns the copyright to the following image(s):

Title(s) of illustration(s): Head and Neck Cancer Regions 2012

Description of the Work: Terese Winslow LLC hereby grants permission to reproduce the above image(s) for use in the work specified:

Thesis title: Evaluation of Timepoint Assessment Schemes for Offline Adaptive Radiotherapy of the Head and Neck Using Velocity

University: Creighton University

Digital Object Identifier (DOI) and Journal ISSN, if available: N/A

License Granted: Terese Winslow LLC hereby grants limited, non-exclusive worldwide print and electronic rights only for use in the Work specified. Terese Winslow LLC grants such rights "AS IS" without representation or warranty of any kind and shall have no liability in connection with such license.

Restrictions: Reproduction for use in any other work, derivative works, or by any third party by manual or electronic methods is prohibited. Ownership of original artwork, copyright, and all rights not specifically transferred herein remain the exclusive property of Terese Winslow LLC. Additional license(s) are required for ancillary usage(s).

Credit must be placed adjacent to the image(s) in the following format:

~~For the National Cancer Institute © (copyright year)~~ ²⁰¹² Terese Winslow LLC, U.S. Govt. has certain rights

Permission granted to:

Name: Andrew Fanning

Mailing address: 1238 N. Saint Joseph Ave., Hastings, NE 68901

Email address: ajf46786@creighton.edu

Phone number: 401-465-4980

Signature Andrew Fanning Date 10/30/2023
Andrew Fanning

Signature Terese Winslow Digitally signed by TERESE WINSLOW
Date: 2023.10.30 16:18:13 -05'00'
Terese Winslow, CMI, Member

Terese Winslow LLC, Medical Illustration
714 South Fairfax Street, Alexandria, Virginia 22314
(703) 836-9121
terese@teresewinslow.com
www.teresewinslow.com

Figure E.1: Permission for Figure 2.1.


Date ▼	Article Title	Publication	Type Of Use	Price	Status	Expiration Date	Order Number
30-Oct-2023	Changes in the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging of Head and Neck Cancer: Rationale and Implications	Current Oncology Reports	Thesis/Dissertation	0.00 \$	Completed 		5658980516753

Figure E.2: Permission for Figure 2.2.

Appendix F

Data Appendix

This appendix provides the data found for each patient. This involves the "detector" quantity used to determine if a replan should take place and the final dose projection with all replans considered. The first table displays the results for the midpoint assessment scheme. The "detector" quantity was created using midpoint anatomy and the final dose projection was created using information from all plans that were used for a particular patient. The second table displays the results for the biweekly assessment scheme. The "detector" quantity was created 3 times per patient at weeks 2, 4, and 6. Again, the final dose projection was created using information from all plans that were used for a particular patient. In all cases, red colored numbers indicate values above the given dose constraint.

Data for Midpoint Assessment Scheme

Detector Analysis	PATIENT 1 Midpoint Detector			Final Dose Analysis	PATIENT 1 Midpoint Final Dose		
Dose Constraints	Truth	Test	Detector Result	Dose Constraints	Truth	Test	Final Dose Result
Cord Max (cGy) < 4500	4323	4442.8	TN	Cord Max (cGy) < 4500	4202	4269	TP
Larynx mean (cGy) < 4000	4064	3904.3	FN	Larynx mean (cGy) < 4000	4150	3929	FN
Left parotid mean (cGy) < 2000	2000	2143.5	TP	Left parotid mean (cGy) < 2000	1869	1903	TP
Right parotid mean (cGy) < 2600	1712	1903.5	TN	Right parotid mean (cGy) < 2600	1976	1993	TP
Pharynx mean (cGy) < 5000-5500	5093	4991.3	TN	Pharynx mean (cGy) < 5000-5500	5111	4992	TP
Mandible max (cGy) < 7200	7123	6527.6	TN	Mandible max (cGy) < 7200	7104	7068	TP
Esophagus mean (cGy) < 3000	718	670.7	TN	Esophagus mean (cGy) < 3000	700	662	TP
Dose Constraints	Truth	Test	Detector Result	Dose Constraints	Truth	Test	Final Dose Result
Cord Max (cGy) < 4500	4354	4383.5	TN	Cord Max (cGy) < 4500	4356	4376	TP
Larynx mean (cGy) < 4000	3325	3262.4	TN	Larynx mean (cGy) < 4000	3191	3200	TP
Left parotid mean (cGy) < 2600	2864	3106.9	TP	Left parotid mean (cGy) < 2600	2574	2885	FN
Right parotid mean (cGy) < 2000	1103	961	TN	Right parotid mean (cGy) < 2000	1111	977	TP
Pharynx mean (cGy) < 5000-5500	5037	5103.9	TN	Pharynx mean (cGy) < 5000-5500	5013	5023	TP
Mandible max (cGy) < 7200	7095	7221	FP	Mandible max (cGy) < 7200	7110	7172	TP
Esophagus mean (cGy) < 3000	1351	1321.2	TN	Esophagus mean (cGy) < 3000	1358	1269	TP
Dose Constraints	Truth	Test	Detector Result	Dose Constraints	Truth	Test	Final Dose Result
Cord Max (cGy) < 4500	4095	4154.4	TN	Cord Max (cGy) < 4500	2844	2841	TP
Larynx mean (cGy) < 4500	3757	3730.4	TN	Larynx mean (cGy) < 4500	3985	4045	TP
Left parotid mean (cGy) < 2600	2240	2407.4	TN	Left parotid mean (cGy) < 2600	2272	2474	TP
Right parotid mean (cGy) < 2000	2323	2161	TP	Right parotid mean (cGy) < 2000	2182	2093	TN
Pharynx mean (cGy) < 5000-5500	4998	4997.5	TN	Pharynx mean (cGy) < 5000-5500	4911	4949	TP
Mandible max (cGy) < 7200	7083	7188.5	TN	Mandible max (cGy) < 7200	7084	7164	TP
Esophagus mean (cGy) < 3000	2161	2656.2	TN	Esophagus mean (cGy) < 3000	2002	2047	TP

Data for Biweekly Assessment Scheme

Detector Analysis

WEEK 2				PATIENT 1 Biweekly Detector				PATIENT 2 Biweekly Detector				PATIENT 3 Biweekly Detector			
Dose Constraints	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result
Cord Max (cGy) < 4500	4323	4381.5	TN	4354	4386.7	TN	4095	4130	TN	4388	4423	TN	1992	1981	TN
Larynx mean (cGy) < 4000	4064	4004.6	TP	3325	3271.3	TN	3757	3734.9	TN	2768	2788	TN	4246	4252	TN
Left parotid mean (cGy) < 2000	2000	1907.9	FN	2654	2638.1	TP	2240	2358.1	TN	2719	2872	TP	2302	2527	TN
Right parotid mean (cGy) < 2600	1712	1823.1	TN	1103	1091.9	TN	2323	2447.9	TP	1035	951	TN	1894	1878	TN
Pharynx mean (cGy) < 5000-5500	5093	5142.7	TN	5037	5086.9	TN	4998	5040.4	TN	5011	5013	TN	4875	4907	TN
Mandible max (cGy) < 7200	7123	6550.7	TN	7095	7076.2	TN	7083	7237.2	FP	7206	7203	TP	7105	7176	TN
Esophagus mean (cGy) < 3000	718	720.3	TN	1351	1651.6	TN	2161	2465.9	TN	1320	1256	TN	2047	2041	TN
WEEK 4				PATIENT 1 Biweekly Detector				PATIENT 2 Biweekly Detector				PATIENT 3 Biweekly Detector			
Dose Constraints	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result
Cord Max (cGy) < 4500	4199	4190	TN	4388	4423	TN	1992	1981	TN	4388	4423	TN	1992	1981	TN
Larynx mean (cGy) < 4000	3994	3909	TN	2768	2788	TN	4246	4252	TN	2768	2788	TN	4246	4252	TN
Left parotid mean (cGy) < 2000	2028	1988	FN	2719	2872	TP	2302	2527	TN	2028	2084	TP	2302	2403	TN
Right parotid mean (cGy) < 2600	2056	2068	TN	1035	951	TN	1894	1878	TN	2056	2030	TN	1894	1870	TN
Pharynx mean (cGy) < 5000-5500	5039	4992	TN	5011	5013	TN	4875	4907	TN	5039	5085	TN	4875	4883	TN
Mandible max (cGy) < 7200	7078	7071	TN	7206	7203	TP	7105	7176	TN	7078	7118	TN	7105	7136	TN
Esophagus mean (cGy) < 3000	655	629	TN	1320	1256	TN	2047	2041	TN	655	660	TN	2047	2068	TN
WEEK 6				PATIENT 1 Biweekly Detector				PATIENT 2 Biweekly Detector				PATIENT 3 Biweekly Detector			
Dose Constraints	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result
Cord Max (cGy) < 4500	4199	4189	TN	4375	4370	TN	1992	1985	TN	4375	4370	TN	1992	1985	TN
Larynx mean (cGy) < 4000	3994	4054	FP	2789	2815	FP	4246	4260	TN	3994	4054	FP	4246	4260	TN
Left parotid mean (cGy) < 2000	2028	2084	TP	2578	2690	TP	2302	2403	TN	2028	2084	TP	2302	2403	TN
Right parotid mean (cGy) < 2600	2056	2030	TN	1065	1038	TN	1894	1870	TN	2056	2030	TN	1894	1870	TN
Pharynx mean (cGy) < 5000-5500	5039	5085	TN	4995	4994	TN	4875	4883	TN	5039	5085	TN	4875	4883	TN
Mandible max (cGy) < 7200	7078	7118	TN	7153	7188	TN	7105	7136	TN	7078	7118	TN	7105	7136	TN
Esophagus mean (cGy) < 3000	655	660	TN	1360	1325	TN	2047	2068	TN	655	660	TN	2047	2068	TN
Final Dose Analysis				PATIENT 1 Biweekly Final Dose				PATIENT 2 Biweekly Final Dose				PATIENT 3 Biweekly Final Dose			
Dose Constraints	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result	Truth	Test	Detector Result
Cord Max (cGy) < 4500	4184	4189	TN	4387	4370	TN	1992	1985	TN	4184	4189	TN	1992	1985	TN
Larynx mean (cGy) < 4000	3972	4054	FP	2780	2815	FP	4246	4260	TN	3972	4054	FP	4246	4260	TN
Left parotid mean (cGy) < 2000	1980	2084	TP	2572	2690	TP	2302	2403	TN	1980	2084	TP	2302	2403	TN
Right parotid mean (cGy) < 2600	2031	2030	TN	1054	1038	TN	1894	1870	TN	2031	2030	TN	1894	1870	TN
Pharynx mean (cGy) < 5000-5500	5031	5085	TN	5009	4994	TN	4875	4883	TN	5031	5085	TN	4875	4883	TN
Mandible max (cGy) < 7200	7097	7118	TN	7156	7188	TN	7105	7136	TN	7097	7118	TN	7105	7136	TN
Esophagus mean (cGy) < 3000	658	660	TN	1345	1325	TN	2047	2068	TN	658	660	TN	2047	2068	TN

References

- [1] National Cancer Institute. Head and neck cancers. <https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet>, May 2021.
- [2] M. D. Mody, J. W. Rocco, S. S. Yom, R. I. Haddad, and N. F. Saba. Head and neck cancer. *The Lancet*, 398:2289–2299, 2021.
- [3] S. B. Chinn and J. N. Myers. Oral cavity carcinoma: current management, controversies, and future directions. *Journal of Clinical Oncology*, 33:3269–3276, 2015.
- [4] Z. A. Zaid, M. K. Neoh, Z. A. M. Daud, N. B. M. Yusop, Z. Ibrahim, Z. A. Rahman, N. Jamhuri, and A. Z. A. A’zim. Weight loss in post-chemoradiotherapy head and neck cancer patients. *Nutrients*, 14:548, 2022.
- [5] J. Castelli, A. Simon, C. Lafond, N. Périchon, B. Rigaud, É. Chajon, B. D. Bari, M. Özşahin, J. Bourhis, and R. d. Crevoisier. Adaptive radiotherapy for head and neck cancer. *Acta Oncologica*, 57:1284–1292, 2018.
- [6] R. L. Siegel, K. D. Miller, H. Fuchs, and A. Jemal. Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, 71:7–33, 2021.
- [7] D. Hashim, E. M. Genden, M. R. Posner, M. Hashibe, and P. Boffetta. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. *Annals of Oncology*, 30:744–756, 2019.

- [8] A. K. Chaturvedi, E. A. Engels, R. M. Pfeiffer, B. Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M. T. Goodman, M. Sibug-Saber, W. Cozen, L. Liu, C. F. Lynch, N. Wentzensen, R. C. Jordan, S. F. Altekruse, W. F. Anderson, P. S. Rosenberg, and M. L. Gillison. Human papillomavirus and rising oropharyngeal cancer incidence in the united states. *Journal of Clinical Oncology*, 29:4294–4301, 2011.
- [9] D. J. Adelstein, J. A. Ridge, M. L. Gillison, A. K. Chaturvedi, G. D’Souza, P. E. Gravitt, W. H. Westra, A. Psyrrri, W. M. Kast, L. A. Koutsky, A. R. Giuliano, S. Krosnick, A. Trotti, D. E. Schuller, A. A. Forastiere, and C. D. Ullmann. Head and neck squamous cell cancer and the human papillomavirus: summary of a national cancer institute state of the science meeting, november 9-10, 2008, washington, d.c. *Head Amp; Neck*, 31:1393–1422, 2009.
- [10] M. L. Gillison, G. D’Souza, W. H. Westra, E. A. Sugar, W. Xiao, S. Begum, and R. P. Viscidi. Distinct risk factor profiles for human papillomavirus type 16–positive and human papillomavirus type 16–negative head and neck cancers. *JNCI: Journal of the National Cancer Institute*, 100:407–420, 2008.
- [11] A. F. Olshan. Cancer of the larynx. *Cancer Epidemiology and Prevention*, pages 627–637, 2006.
- [12] D. Luce, M. Gérin, A. Leclerc, J. Morcet, J. Brugère, and M. Goldberg. Sinonasal cancer and occupational exposure to formaldehyde and other substances. *International Journal of Cancer*, 53:224–231, 1993.
- [13] S. Preston-Martin, D. C. Thomas, S. C. White, and D. Cohen. Prior exposure to medical and dental x-rays related to tumors of the parotid gland1. *JNCI Journal of the National Cancer Institute*, 80:943–949, 1988.

- [14] Memorial Sloan Kettering Cancer Center. <https://www.mskcc.org/cancer-care/types/head-neck/diagnosis>, Jan 1970.
- [15] Head and Neck Cancer Alliance. <https://www.headandneck.org/hpv-head-neck-cancer/>, Sep 2020.
- [16] National Cancer Institute. Cancer staging. <https://www.cancer.gov/about-cancer/diagnosis-staging/staging>, Oct 2022.
- [17] D. K. Zanoni, S. G. Patel, and J. P. Shah. Changes in the 8th edition of the american joint committee on cancer (ajcc) staging of head and neck cancer: rationale and implications. *Current Oncology Reports*, 21, 2019.
- [18] J. Heukelom and C. D. Fuller. Head and neck cancer adaptive radiation therapy (art): conceptual considerations for the informed clinician. *Seminars in Radiation Oncology*, 29:258–273, 2019.
- [19] Mayo Clinic Staff. Radiation therapy. <https://www.mayoclinic.org/tests-procedures/radiation-therapy/about/pac-20385162>, Apr 2023.
- [20] A. F. Alfouzan. Radiation therapy in head and neck cancer. *Saudi Medical Journal*, 42:247–254, 2021.
- [21] Douglas Jones. Icru report 50—prescribing, recording and reporting photon beam therapy. *Medical Physics*, 21(6):833–834, 1994.
- [22] John P. Gibbons and Faiz M. Khan. *Khan’s the physics of radiation therapy*. Wolters Kluwer, 2020.
- [23] S. Oh and S. Kim. Deformable image registration in radiation therapy. *Radiation Oncology Journal*, 35:101–111, 2017.
- [24] D. L. G. Hill, P. Batchelor, M. P. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.

- [25] A. Kumarasiri, F. Siddiqui, C. Liu, R. Yechieli, M. Shah, D. Pradhan, H. Zhong, I. J. Chetty, and J. Kim. Deformable image registration based automatic ct-to-ct contour propagation for head and neck adaptive radiotherapy in the routine clinical setting. *Medical Physics*, 41:121712, 2014.
- [26] Varian. <https://www.varian.com/>, 2023.
- [27] K. K. Brock, S. Mutic, T. McNutt, H. Li, and M. L. Kessler. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the aapm radiation therapy committee task group no. 132. *Medical Physics*, 44:e43–e76, 2017.
- [28] Research Cancer Resources from OncoLink — Treatment. Fractionation and radiation. <https://www.oncolink.org/cancer-treatment/radiation/support/fractionation-and-radiation>, Feb 2023.
- [29] G. Raghavan, A. U. Kishan, M. Cao, and A. M. Chen. Anatomic and dosimetric changes in patients with head and neck cancer treated with an integrated mri-tri-60co teletherapy device. *The British Journal of Radiology*, 89:20160624, 2016.
- [30] D. Yan, F. A. Vicini, J. Wong, and Á. Martínez. Adaptive radiation therapy. *Physics in Medicine and Biology*, 42:123–132, 1997.
- [31] J. L. Barker, A. S. Garden, K. K. Ang, J. O’Daniel, H. Wang, L. E. Court, W. H. Morrison, D. I. Rosenthal, K. C. Chao, S. L. Tucker, R. Mohan, and L. Dong. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated ct/linear accelerator system. *International Journal of Radiation Oncology*Biophysics*, 59:960–970, 2004.
- [32] S. Sharma, A. Dewan, H. Srivastava, S. Rawat, A. Kakria, M. Mishra, T. Suresh, and K. Mehrotra. Impact of adaptive radiotherapy on locally advanced head and

- neck cancer - a dosimetric and volumetric study. *Asian Pacific Journal of Cancer Prevention*, 17:985–992, 2016.
- [33] J. Heukelom, O. Hamming, H. Bartelink, F. Hoebbers, J. Giralt, T. Herlestam, M. Verheij, M. W. M. v. d. Brekel, W. V. Vogel, N. J. Slevin, É. Deutsch, J. Sonke, P. Lambin, and C. R. N. Rasch. Adaptive and innovative radiation treatment for improving cancer treatment outcome (artforce); a randomized controlled phase ii trial for individualized treatment of head and neck cancer. *BMC Cancer*, 13, 2013.
- [34] R. Krishnatry, J. Bhatia, V. Murthy, and J. Agarwal. Survey on adaptive radiotherapy practice. *Clinical Oncology*, 30:819, 2018.
- [35] International Atomic Energy Agency. *Record And Verify Systems For Radiation Treatment Of Cancer: Acceptance Testing, Commissioning And Quality Control*. 2013.
- [36] Varian Medical Systems. Velocity. <https://www.varian.com/products/software/information-systems/velocity>, 2021.
- [37] Centers for Disease Control and Prevention. Cost-effectiveness analysis. <https://www.cdc.gov/policy/polaris/economics/cost-effectiveness/index.html>, Oct 2021.