8-2018

EQUIPMENT TO ADDRESS INFRASTRUCTURE AND HUMAN RESOURCE CHALLENGES FOR RADIOTHERAPY IN LOW-RESOURCE SETTINGS

Rachel McCarroll

EQUIPMENT TO ADDRESS INFRASTRUCTURE AND HUMAN RESOURCE CHALLENGES FOR RADIOTHERAPY IN LOW-RESOURCE SETTINGS

by

Rachel Elizabeth McCarroll, B.S.

APPROVED:

_____
Laurence E. Court, Ph.D.
Advisory Professor

_____
Peter Balter, Ph.D.

_____
Beth Beadle, M.D., Ph.D.

_____
David Followill, Ph.D.

_____
Christopher Nelson, Ph.D.

_____
Christine Peterson, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

EQUIPMENT TO ADDRESS INFRASTRUCTURE AND HUMAN RESOURCE CHALLENGES FOR RADIOTHERAPY IN LOW-RESOURCE SETTINGS

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Rachel Elizabeth McCarroll, B.S.

Houston, Texas

August 2018

Dedicated to Fahd.

You have changed my life forever and I couldn't be more grateful.

Thank you for your unwavering support and constant encouragement.

May we forever keep laughing and learning.

I love you.

Acknowledgements

I would first like to thank my advisor, Dr. Laurence Court for the guidance, support and opportunities he has provided during my experience in graduate school; our trip to South Africa and the collaboration with physicists abroad has been a highlight of my career and I expect it to be a wonderful launching point. I owe a great deal to you. Your commitment to the graduate school experience does not go unnoticed. The collaborative and supportive atmosphere of our group is due much to your leadership.

Thank you also to my committee, Drs. Peter Balter, Beth Beadle, David Followill, Christopher Nelson, Christine Peterson, and Francesco Stingo. Your support and commitment to my growth was invaluable and your insight crucial to the success of this work.

This work would not have been possible without the tremendous support and efforts of other faculty and staff – Kelly Tharp, Danna Fullen, Rajat Kudchadker, Jinzhong Yang, Joy Zhang, and Ryan Williamson.

I feel lucky to be part of such an amazing research group – Brian Anderson, Carlos Cardenas, Xenia Favè, David Fried, Skylar Gay, Scott Ingram, Rachel Ger, Kelly Kisling, Dennis Mackin, Tucker Netherton, Calli Nguyen, Constance Owens, Ashley Rubinstein, and DJ Rhee. These members of the Court Lab, past and present, have made the past five years an amazing experience and have set a high bar for colleagues to come!

I would like to thank my family for being there every step of the way and giving the foundation I needed to succeed.

Finally, A very special thanks to all the friends who have walked with me along this journey – Bethany Boeschen, Ashley Rubinstein, Kelly Kisling, Xenia Favè and Travis Ray, David and Erica Fried, Mattie and Joshua Loucks, Olivia and Haden Popnoe, and Scott and Reagan Ingram. May we remain friends for life!

EQUIPMENT TO ADDRESS INFRASTRUCTURE AND HUMAN RESOURCE CHALLENGES FOR RADIOTHERAPY IN LOW-RESOURCE SETTINGS

Rachel Elizabeth McCarroll, B.S.

Advisory Professor: Laurence E. Court, Ph.D.

Millions of people in low- and middle- income countries (LMICs) are without access to radiation therapy and as rate of population growth in these regions increase and lifestyle factors which are indicative of cancer increase; the cancer burden will only rise. There are a multitude of reasons for lack of access but two themes among them are the lack of access to affordable and reliable teletherapy units and insufficient properly trained staff to deliver high quality care. The purpose of this work was to investigate to two proposed efforts to improve access to radiotherapy in low-resource areas; an upright radiotherapy chair (to facilitate low-cost treatment devices) and a fully automated treatment planning strategy.

A fixed-beam patient treatment device would allow for reduced upfront and ongoing cost of teletherapy machines. The enabling technology for such a device is the immobilization chair. A rotating seated patient not only allows for a low-cost fixed treatment machine but also has dosimetric and comfort advantages. We examined the inter- and intra-fraction setup reproducibility, and showed they are less than 3mm, similar to reports for the supine position.

The head-and-neck treatment site, one of the most challenging treatment planning, greatly benefits from the use of advanced treatment planning strategies. These strategies, however, require time consuming normal tissue and target contouring and complex plan optimization strategies. An automated treatment planning approach could reduce the additional number of medical physicists (the primary treatment planners) in LMICs by up to

half. We used in-house algorithms including mutli-atlas contouring and quality assurance checks, combined with tools in the Eclipse Treatment Planning System®, to automate every step of the treatment planning process for head-and-neck cancers. Requiring only the patient CT scan, patient details including dose and fractionation, and contours of the gross tumor volume, high quality treatment plans can be created in less than 40 minutes.

Table of Contents

List of Illustrations

List of Tables

Chapter 1: Introduction

Cancer is a worldwide and growing epidemic; projected to kill nearly 13 million people by 2030. It is a fallacy that cancer is a disease only of the developed world; the most recent statistics showed more than two thirds of cancer-related deaths occur in low- and middle-income countries (LMICs)(1). As the population of LMICs ages and as risk factors such as smoking, poor dietary habits, and sedentary lifestyles increase, the cancer burden will only rise(2).

Though colloquially referred to as a single disease, cancer is a class of diseases which share common traits, and, therefore, the treatment of cancer is necessarily very varied. Treatment options often involve one or more of three common techniques; surgery, chemotherapy, and/or radiation therapy. Surgical intervention has the longest history with records dating back to Greek physicians in the first and second centuries A.D.(3) and may be required for up to 80% of cancer patients(4). Surgery, however, is often insufficient for metastatic disease. Chemotherapy, introduced at the beginning of the 20th century, uses drugs to combat cancer growth and spread at the molecular level and can only be administered if sufficient laboratory facilities are also available (5). Finally, radiation therapy, introduced at the end of the 19th century and available for both the palliative and curative treatment of cancer uses energetic subatomic particles due induce DNA damage and kill cancer cells(6). Radiotherapy is required for the treatment of more than 50% of the cancer population (7).

Radiation and its use in medicine have been intertwined from the start; in late 1895 Wilhelm Röntgen designed an experiment to identify the source of flouresence on a painted cardboard screen and it was only two weeks later he took the famed first radiograph of his wife's hand, which he included in the first manuscript describing this new kind of rays(8). The use of radiation to treat cancers came shortly thereafter, for stomach cancer in 1896, for basal

cell carcinoma in 1899, and many others to follow(9). Radiation therapy has advanced quickly since its introduction and an improved understanding of how to harness the biological effect of radiation on cancer cells, improvements in the safe and accurate delivery of radiation throughout the body, and, more recently, advances in computing power and have led to rapid changes in treatment delivery and improved patient outcomes.

Radiation therapy has been shown to be cost-effective for both palliative and curative cancer treatment in LMICs (10, 11); one study showed that the cost of an entire course of radiation therapy in Senegal is only $300USD (12). However, due in part to the substantial capital investment and the ever-increasing pace of technological advancements, the practice of radiation oncology in developed and less developed regions rapidly is rapidly diverging. In low- and middle-income countries, defined as those countries with a gross national income per capita less than $12,235 USD and where 84% of the world's population (13) and 57% of the cancer population lives(1), there exist only 30% of the world's radiation therapy equipment (14). It is not only the equipment for treatment that is lacking – so too are trained personnel required to safely and effectively deliver treatment. It is estimated that by 2020 LMICs will need an additional 9,169 radiation therapy machines (3.2 times what was available in 2014), 9,915 medical physicists (3.9 times 2014 availability), 12,149 radiation oncologists (2 times 2014 availability), and 29,140 radiation therapists (3.7 times 2014 availability)(15). In order to bridge the gap in cancer care around the world, and to provide safe, effective, and possibility lifesaving treatment to 6.3 million people, urgent and innovative solutions are needed.

In order to address the overwhelming demand for radiotherapy, a multi-thronged approach is necessary. Efforts should seek to utilize technology and ideas at the forefront of the field, should be developed in concert with professionals for which the solutions are aimed, and should utilize and build upon previous efforts. Current initiatives underway to bridge the gap in radiotherapy needs worldwide include efforts to provide necessary equipment and ancillary supplies to less developed regions, programs for the training (initial and ongoing) of

radiotherapy professionals, systems for peer-to-peer collaborations which encourage professional growth and promote high quality care, the development of low-cost and effective solutions, and there are many others. Despite these efforts, there remains a significant challenge to provide radiotherapy services worldwide. The work herein represents two proposed efforts; an upright radiotherapy chair and a fully automated treatment planning strategy.

An Upright Radiotherapy Chair

External beam radiation therapy machines have been in use since the 1950s and early treatments had reports of patients treated in lying, seated, and standing positions (16, 17). The invention of the computed tomography (CT) scanner in 1972 gave physicians the ability to accurately visualize each patient's anatomy and, in the context of radiation therapy, optimize treatment delivery. The incorporation of CT image acquisition for treatment planning represented a breakthrough in radiation therapy and the acquisition of pre-treatment CT scans for use in treatment planning has become routine practice in radiation oncology clinics. The orientation of CT scanners is such that the patient lies, generally supine, on a treatment couch inside a bore with a diameter of up to 90cm (18), and this effectively requires that radiation treatment plans are developed for and delivered to patients in a lying position. Further, it requires that teletherapy machines accommodate this treatment position. Unfortunately, some patients, particularly those with head-and-neck or lung cancers, may develop orthopnea, dyspnea, dysphagia, or other conditions that make lying flat for the duration of treatment difficult or impossible. Further, it has been shown that dosimetric advantages can be had in other position; for example, when patients assume a seated or upright position lung volume and motion are reduced, allowing for sparing of normal tissues and fewer radiation-induced symptoms (19-21).

To accommodate the lying position during treatment delivery radiotherapy machines most commonly feature a rotating gantry around a stationary isocenter and this contributes greatly to the considerable cost of teletherapy units. In an effort to reduce the cost of teletherapy treatment machines and with added benefits of patient comfort and limited dosimetric improvement; we propose the development and validation of a treatment chair for use in radiation therapy. The use of treatment chair which rotates around the patient's axis would allow for a fixed radiation field and tremendously reduce machine cost. In addition, such a treatment paradigm would allow for a reduction in cost due to shielding, set-up, treatment delivery, machine downtime, and other factors of which are currently under investigation (22-24). This lower cost machine is of interest to machine vendors and would be highly applicable in areas with limited resources.

The work herein represents an important step in the assessment of the clinical utility of a radiotherapy treatment chair; set-up reproducibility. As radiation treatment often occurs in small fractions over the course of many weeks and the prescribed treatment is developed from a single pre-treatment image it is critical that the patient be positioned in very similar positions each day; this is known as inter-fraction set-up reproducibility. Further, as radiation treatment, especially advanced techniques, may take more than ten minutes for delivery(25) it is important the patient maintains the same position for the duration of treatment; this is known and intra-fraction reproducibility. Both inter- and intra- fraction reproducibility are important aspects when considering the implementation of new patient positions into clinical practice. The traditional supine position has been shown to have inter- and intra- fraction reproducibility less than 5mm on average for treatment of the breast (26, 27), prostate (28), head and neck (29), and whole brain (30). The main goal of this study was to assess the setup reproducibility of a novel treatment chair design, paving the way for clinical use and supporting further investigation into a fixed-beam low-cost radiotherapy treatment machine.

Automated Treatment Planning

Treatment planning is a mainstay in the field of radiation oncology. During the required treatment planning process a representation of the patient, most often a simulation CT, is used to design or select beam arrangements, shapes, energies, and combinations which will best deliver the prescribed therapeutic dose to the target volume while sparing as much normal tissue possible. The practice of treatment planning has evolved quite dramatically in the past thirty years. In the 1990s the practice of acquiring a computed tomography scan of the patient, on which the target and organs at risk are delineated and the dose delivery is visualized and optimized, became common practice. Compared to conventional techniques, 3D conformal treatment allowed tumor dose to be escalated without a significant increase in normal tissue dose (31), thus improving tumor control (32, 33) while reducing normal tissue complication (34). The use of 3D conformal techniques did, however, also increase the cost, complexity, burden on equipment and personnel, and time needed to plan and deliver radiation therapy (31, 35, 36).

Treatment planning and delivery further improved with the introduction of intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) which rely on the use of non-uniform beam fluences and an inverse approach to the dose optimization problem. Through the use of optimization criteria and a search algorithm beam parameters are optimized in order to meet the constraints set by the user through the minimization (or maximization) of an objective function. The beam fluence variations in IMRT and VMAT offer distinct advantages over simpler beam shaping devices for irregularly shaped targets and those in close proximity to critical organs at risk. The advantages of these advanced techniques is underlined in the treatment of cancers in the head-and-neck. Cancers in this region are in close proximity of up to 25 organs at risk, and advanced planning techniques have been shown to significantly increase planned target coverage and conformality while

decreasing normal tissue complications including xerostomia and fibrosis (37-40). More than two thirds of the head-and-neck cancer population lives in less developed regions(1) where access to these advanced techniques is severely limited and while, for many clinics, the lack of advanced equipment is the primary barrier in the delivery of advanced treatments, there are many others for whom the machinery exists but treatment delivery is limited due to staffing considerations, this is the scenario for our partner clinics in South Africa and the Philippines.

In low-resource areas the time and cost required for treatment planning, especially for these advanced techniques, contribute to the limited number of patients able to receive advanced treatment (41). The cost of personnel, including highly skilled treatment planners, while less than in high-income countries, represent 10% of the total cost of radiotherapy (6). Combined with the well documented “brain drain” on human resources in low-income areas (42) the need for tools which can relieve part of the staffing burden while maintaining a standard of care is tremendous.

The increasing complexity of treatment planning has also brought other challenges. Complex treatments often involve many decisions and tradeoffs during the planning process and it has been shown that every IMRT and VMAT treatment plan may not offer the same benefits; in a study on treatment plan quality investigators found the plan quality was not significantly correlated with the treatment planning system, modality of delivery, plan complexity, education or certification of the planner, planner confidence or experience, the number of beams, or the number of monitor units. Instead, it was concluded, plan quality was most contributable to general “planner skill” (43). An automated planning strategy may also standardize treatment plan quality which are prone to significant variability; this variability represents a significant challenge when comparing patient outcomes (44) and when optimizing plan delivery parameters. Additionally, these advanced IMRT and VMAT

techniques come with a significant increase in technical burden including infrastructure required for plan development, quality assurance checks, and delivery.

We propose a solution which would automate the treatment planning process for head-and-neck cancers; removing a time intensive aspect of the treatment planning process; standardizing treatment across patients, and relieving highly trained staff for other duties, ultimately reducing the tremendous deficit in the numbers of these staff.

Automated methods are emerging throughout the radiotherapy process. From beam commissioning to patient plan checks, researchers and vendors alike are exploring automated methods of improving radiotherapy (45, 46), and this includes treatment planning (47-49). At the forefront of automated plan optimization is knowledge based planning (KBP) which uses information gathered from previously treated patients to generate treatment plans for new patients and was first reported in 1990 for its potential to organize and harness the current state of knowledge of treatment planning in order to improve treatment planning (50). With technological advancement and improved computing power, the interest in KBP has increased dramatically and has found recent success in the planning of breast, prostate and head-and-neck cancer showing an improvement in plan quality and a reduction in plan variability (51-54). Unfortunately, these studies fall short of full plan automation, as often these KBP algorithms require manual inputs. One of the most time intensive requirements of advanced planning techniques is the delineation of normal tissues and targets. The automatic contouring of these structures is an area of ongoing research; technologies ranging from traditional atlas based deformable solutions to newer deep learning pixel-wise classification methods are finding success in the contouring of both normal tissues and targets for a variety of treatment sites (55-57). These approaches, however are limited in scope and are not fully integrated into routine clinical practice. The full validation of automatic contouring for clinical use and oversight of the use of an autocontouring technique in an automated treatment planning system has not yet been presented in the literature. Together automatic plan

optimization and autocontouring represent a majority of the tasks required for treatment planning.

Ideally, a fully automated planning approach would eliminate the need for human intervention, would produce clinically acceptable or superior plans, would include methods to assure the quality and safety of each step, and would do so while reducing the time and human effort needed for planning. Such an automated planning technique would not only be tremendously helpful in low resource areas but would prove an essential tool in all clinics, may improve the standard of care, and would be a reliable tool for comparing planning techniques within and among patients.

Towards realizing the benefits of full automation in treatment planning, we propose a fully automated treatment planning approach for head-and-neck cancer. This system would decrease the time required to generate a high quality plan, would reduce the training needed for plan production, would reduce plan variability, and may facilitate the transition to advanced planning techniques. We envision a system for which the user must only provide an approved CT scan, patient information including information about the prescribed dose and fractionation, and the identification of the primary target through the contouring of either a high dose CTV or gross tumor volume(s) including both the primary and nodal disease as indicated. The system would then use this information to generate a treatment plan through many, validated steps including plan preparation, automatic contouring of normal tissues, automatic contouring of target volumes, plan optimization, dose calculation, and plan finalization. Secondary and independent checks of each step should be implemented for redundancy and to ensure the safe use of an automated system. An overview of the system can be seen in Figure 1. The work herein represents validation and development of three key components, the knowledge based plan optimization, the contouring of normal tissues and the generation of targets volumes.

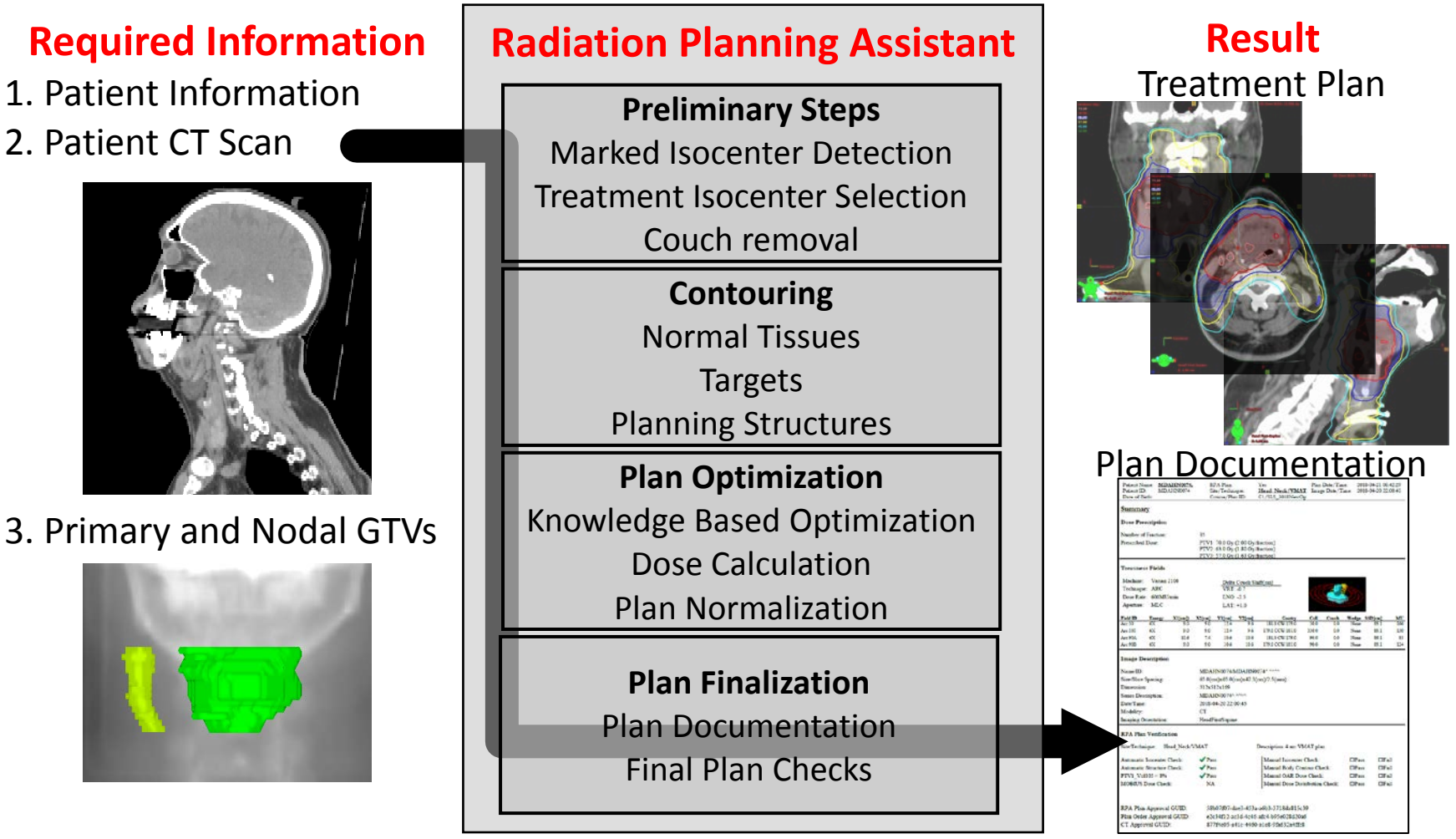


Figure 1. Schematic of the proposed automatic treatment planning strategy. Presented in this work are the development and validation of automatic contouring of both normal tissues and targets and the knowledge based plan optimization technique and associated quality assurance checks.

Chapter 2: Purpose and Central Hypothesis

Central Hypothesis

Equipment can be developed which may partially alleviate the staff and infrastructure burden of radiotherapy in low- resource settings and are feasible and clinically appropriate.

Specific Aim 1: An Upright Radiotherapy Chair

Aim: To preclinically validate the use of an upright radiotherapy chair for head-and-neck patients

Hypothesis: An upright radiotherapy chair has clinically acceptable inter- and intra- fraction reproducibility

Experiment 1.1: Establish the inter- and intra- fraction setup variation of an upright radiotherapy chair

Experiment 1.2: Assess patient experience in an upright radiotherapy chair

Specific Aim 2: A Fully Automated Treatment Planning Strategy

Aim: To develop and validate a single optimization treatment planning strategy for the head-and-neck

Hypothesis: Single optimization head and neck treatment plans perform with equal quality to clinically acceptable plans and 90% are accepted by radiation oncologists for use without edit.

Experiment 2.1: Quantitative and physician review of treatment plan quality

Specific Aim 3: Automatic Delineation of Normal Structures in the Head-and-Neck

Aim: The assess the feasibility of the use of automatically contoured normal structures in the head and neck in a fully automated treatment planning strategy

Hypothesis: Automatically contoured normal structures can be used for treatment planning purposes without significant impact on plan quality.

Experiment 3.1: Evaluation of automatic contouring algorithms for normal structures in the head and neck

Experiment 3.2: Clinical use of an autocontouring algorithm for normal structures in the head and neck

Experiment 3.3: Development of a random forest model for assessment of anatomical errors in autocontours of normal structures in the head and neck

Experiment 3.4: Assessment of dosimetric impact of using autocontoured normal structures for treatment planning in the head and neck

Specific Aim 4: Automatic Delineation of Target Volumes in the Head-and-Neck

Aim: To assess the feasibility of the use of automatically countered intermediate and low dose target volumes in the head and neck

Hypothesis: Automatically contoured clinical target volumes can safely be safely used for treatment planning purposes.

Experiment 4.1: Quantitative and physician review of an automatically contoured clinical target volumes in the head and neck

Experiment 4.2: Assessment of dosimetric impact of using autocontoured clinical target volumes for treatment planning in the head and neck

Experiment 4.3: Development of a method to ensure quality of autocontoured clinical target volumes

Chapter 3: An Upright Radiotherapy Chair

A substantial portion of this chapter is written or based on the following publication:

McCarroll RE, Beadle BM, Fullen D, Balter Pa, Followill DS, Stingo FC, Yang J, Court LE. Reproducibility of patient setup in the seated treatment position: A novel treatment chair design. Journal of Applied Clinical Medical Physics. DOI: 10.1002/acm2.12024 Volume 18, Issue 1, Pages 223-229 © John Wiley & Sons

In this chapter we describe the results for Specific Aim 1 which pertains to the development and clinical validation of an upright treatment approach. Our working hypothesis is that an upright radiotherapy chair has clinically acceptable inter- and intra- fraction reproducibility equal to or less than that in the traditional supine position.

Introduction

The majority of patients treated with radiation therapy are positioned supine on the treatment couch, with a small proportion prone. These positions are supported by decades of experience and are suited for the routine practice of 3-dimensional treatment planning with imaging from computed tomography (CT) scanners which utilize horizontal bores. However, some patients, particularly those with head-and-neck or lung cancers, may develop orthopnea, dyspnea, dysphagia, or other conditions that make lying flat for the duration of treatment difficult or impossible. Further, an upright treatment allows for an increase in lung volume and decrease in lung motion which allow for the sparing of normal tissues and fewer radiation-induced symptoms (19-21). In addition to the comfort and dosimetric advantages of treatment in the upright position, this treatment position could allow for the development of a treatment paradigm centered on a fixed treatment beam and seated rotating radiotherapy

patient. This delivery approach would prove advantageous in the development of a low-cost linear accelerator system, applicable to low- and middle- income countries. Advantages of this approach in terms of cost, shielding, set-up, treatment delivery, machine downtime, and others factors are under investigation (22-24). Interest from vendors in a fixed beam system has further supported this work.

Historically, chairs for radiation therapy are used primarily as an exception for patients unable to tolerate the lying position and have involved temporary replacement of the treatment couch with an upright unit(58-60) Additionally, these previous studies are from an era in which treatment planning was carried out primarily using 2D image acquisition(61-63) and margins which were much more tolerant of positional inaccuracies. The degree of these uncertainties is not well documented in the literature; only one description of an upright system included an assessment of the reproducibility of patient position(61) and found that for all patients shifts of the treatment blocks of at least 5mm were required.

Concerning treatment planning in the upright position, traditional CT scanner geometries do not allow for image acquisition in the seated position. Recent studies, however, have explored the feasibility of acquiring cone beam CT (CBCT) scans of seated patients using the on-board imaging capabilities of modern medical linear accelerators by positioning the gantry at 0° degrees, and then rotating the patient couch instead of the gantry (64, 65). Studies have also demonstrated the feasibility of using cone beam CT images acquired at the treatment unit for treatment planning (64, 66-68). This work supports our expectation that we will soon be able to take CBCT images of patients in an upright position for the purpose of treatment planning, by rotating the treatment couch. It has been reported that acquisition of a field of view of 40 cm x 26 cm at isocenter is possible (64).

Given the above, we have developed a treatment chair suitable for use with standard gantry based linear accelerator geometries for head-and-neck cancer regions, incorporating

measures designed to optimize the reproducibility of inter- and intra- fraction patient setup. Herein we report the details of the chair design, inter- and intra- fraction reproducibility measurements for five head-and-neck cancer patients under simulated treatment scenarios, and patient feedback and discuss considerations for future development.

Preclinical validation of a treatment chair is necessary for the development of a treatment approach centered on treatment chair and paired with a fixed radiation beam; such a treatment paradigm could greatly reduce the upfront and ongoing cost of radiation therapy make the treatment more available in low-resource settings.

Methods

Chair Design

The chair was initially designed by engineering students at Rice University (Houston, TX), with major refinement to improve patient comfort and ease of patient positioning. The general concept was based on a massage chair, as the forward leaning position was expected to give better stability than a regular chair design. Additionally, this forward leaning position is beneficial for patients with an excess accumulation of saliva. The chair was constructed in two major parts: (i) the seat with the back rest was constructed such that it slid onto the end of the treatment couch and was securely fastened to avoid shifts in position, and (ii) a unit consisting of footrests (15 x 30 x 2 cm acrylic), a chest plate (T-shaped acrylic), a face piece, and a wooden support post. Once the patient sat down, the second unit slid into position between the patients' legs and securely tightened into position. Having the chair attach to the couch allowed us to make use of the couch's remote motions to correct patient position based on pre-treatment imaging. Additionally, set-up of the chair fits smoothly into patient treatment workflow, where therapists gather and position any accessories needed for treatment shortly before the patient enters the treatment vault. The chair allows for many positional variations due to patient size, height, and comfort. Figure 2 shows the available

chair adjustments, including adjustment of the seat depth (A), the chest plate height (B), the chest plate angle (C), the face piece angle (D), and the footrest height (E). The chair was manufactured in-house, primarily from wood and acrylic materials. Limiting the use of metal was important to avoid affecting beam or imaging quality. Further, the construction allowed for easy maneuverability into position. For setup and reproducibility, indexing measures including notches and angle identifiers were incorporated.

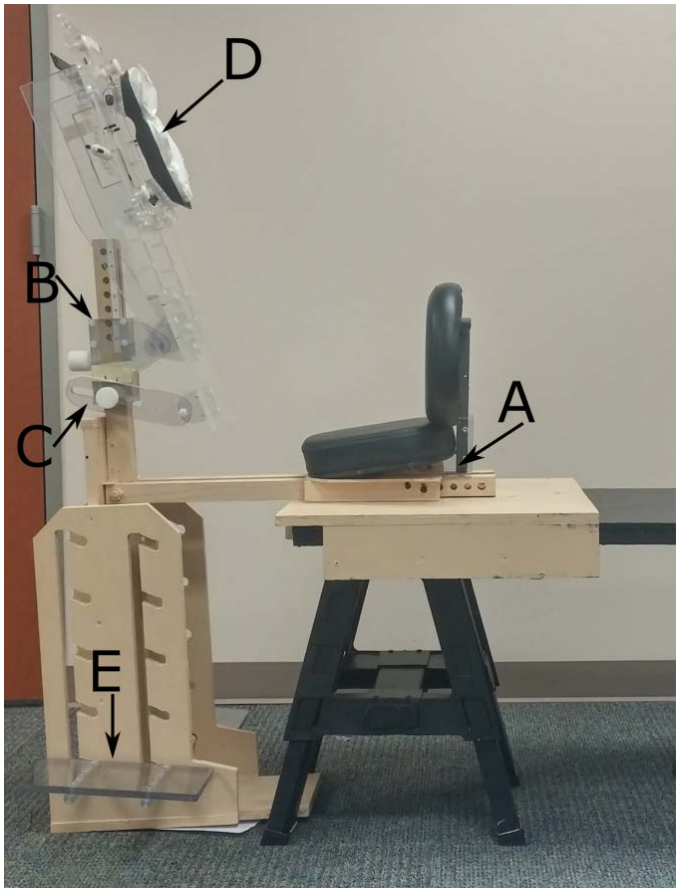


Figure 2. Treatment chair setup. For simulation, a flattop bench was used to mimic the treatment couch in the treatment vault. For image acquisition, the seat was securely fastened to the treatment couch. The setup is adjustable for patient size and comfort including adjustment of the seat depth (A), chest plate height (B), chest plate angle (C), face piece angle (D), and footrest height (E).

Intra- and Inter-Fraction Imaging

Six head-and-neck cancer patients undergoing radiotherapy (in a supine position) were accrued with approval from our institutional review board. Five patients completed the study and are included in the analysis. The seated patients were first set-up in the treatment chair outside of the treatment room using a flattop bench in lieu of the treatment couch

(Figure 2). For setup, the chair position was established and a Vac-Lok cushion (MTVLG35C; Civco Medical Solutions, Coralville, IA) and thermoplastic head mask (MTAPU; Civco Medical Solutions, Coralville, IA) were made. The Vac-Lok cushion was used to fill any space between the patient's chest and the chest piece, to create pseudo arm rests for patient comfort, and to facilitate set-up reproducibility, especially lateral stabilization (Figure 3). The head mask was secured over the back of the patient's head, in contrast to typical head-and-neck cancer treatment for which a thermoplastic mask is generally placed over the patient's face and secured to the treatment table. Additionally, the patients were assessed for the need of additional accessories, including an A-bar for arm and hand positioning and comfort and a pillow behind the back for added support.

Figure 3. Patient setup for lateral image acquisition in the treatment vault , with kV imagers extended and couch positioned at 0°. The Vac-Lok cushion was shaped so as to create armrests for patient comfort, the head mask was secured of the back of the patient's head.

For imaging, the chair position was duplicated in the treatment vault. For two patients, acrylic shims were needed to loosen the thermoplastic mask at the face, after the mask had hardened. A TrueBeam® linear accelerator (Varian Medical Systems, Palo Alto, CA) was used for this study, primarily due to the couch end load limit of 200 kg, which allows for positioning of the treatment seat and patient at the couch's end. With the gantry at 330°, kilovoltage imagers retracted, and the patient couch lowered to the full extent and positioned at 270°, the patient was set up as in simulation. The gantry rotation to 330° was necessary to improve access in this relatively tight space. We also inserted a custom tray into the physical wedge slot to protect the exit window in case of accidental contact.

The longitudinal table position was selected so that the patient's vertebrae were approximately at the beam's isocenter. Using orthogonal lasers, the patient's position was marked on the thermoplastic mask. The gantry was rotated to 0° and kilovoltage imagers were extended outward. The position of kilovoltage imagers varied between patients due to patient size and couch location, to which the chair was attached. The superior-inferior, left-right, and anterior-posterior positions of the imagers relative to the patient ranged 10, 3, and 5cm respectively. The position of the imagers was such that anatomical regions captured in the image were similar between patients. Posterior-anterior images were acquired first (couch at 90°, gantry at 0°, kV imagers extended), the couch was rotated to 0° and then lateral images were acquired. All mechanical motions occurred under supervision inside the treatment vault.

After image acquisition and under supervision, the couch was rotated for 5 minutes to simulate treatment delivery. Two additional images (lateral and P-A) were then acquired. Image registration of these two sets of images was used to calculate the first intra-fraction reproducibility measurement. The patient then got out of the treatment chair and rested for a few minutes, and the process was repeated to acquire 2 more sets of images, providing one inter-fraction reproducibility measurement and one additional intra-fraction measurement.

Upon completion, the patient was asked to complete a questionnaire (see supplemental materials, page 180) regarding both their supine treatments and their experience in the chair.

Image Registration

The head-and-neck region has many degrees of motion, so inter- and intra-fraction alignment was evaluated for several sub-regions of the acquired images. We used a method similar to that used previously to evaluate setup reproducibility in patients with head-and-neck cancer after cone beam CT guidance (69). Sub-regions of interest on kV projection lateral images were cervical vertebrae 1-3 (C1C3), C3C5, the mandible, and the occipital bone. The sub-regions of interest on P-A images were the left temporomandibular joint and the nasal sinuses. These regions were chosen to facilitate accurate evaluation of patient motion, to match those studied previously (69), and to obtain high visibility on the acquired images. The images were processed via histogram normalization and sub-regions were chosen by hand to include the area of interest, see Figure 4. Rigid 3-dimensional image registration (bi-directional translation and rotation) was carried out between the two 2-dimensional kilovoltage images in Matlab (MathWorks, Natick, MA) using the gradient descent method with the mean square error as the registration metric. All registration results were verified visually.

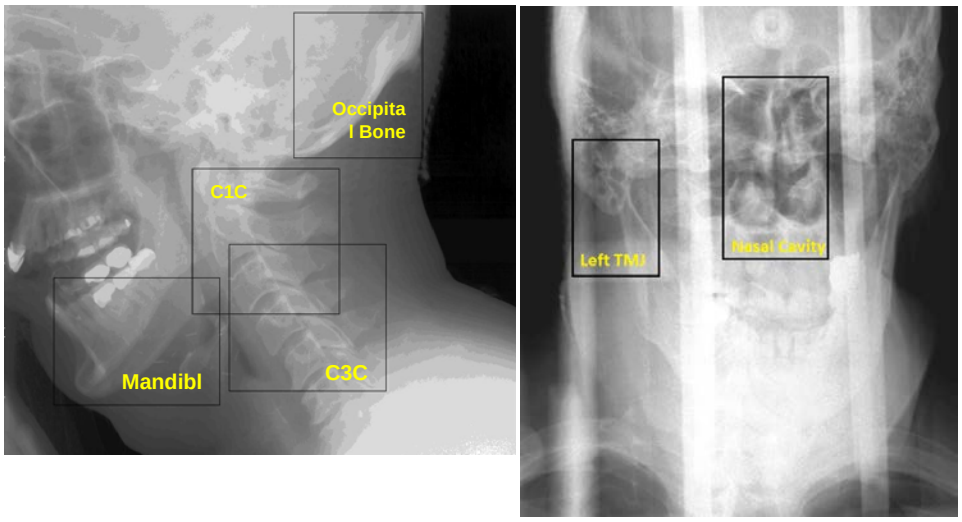


Figure 4. Histogram-normalized kilovoltage image of a representative patient , outlining the sub-regions selected on the lateral image (A) and PA image (B) for registration. C1C3 and C3C5, cervical vertebrae 1–3 and 3–5, respectively.

Simulated Image Guidance

The use of image guidance for patient positioning and tumor localization in head-and-neck cancer patients receiving radiotherapy is a routine procedure in many clinics and is necessary for the delivery of advanced treatment techniques such as IMRT and VMAT. We therefore simulated the use of image guidance in our inter-fraction displacement images, as done by others, which were acquired without patient realignment prior to the simulated treatment delivery. The two images were first registered according to the position of C1C3 (lateral images) or the spinal column (posterior-anterior images). Then, the remaining sub-regions were registered as previous. This approach provided a measure of the residual error in inter-fraction displacement given the use of image guidance.

Measurement of Registration Uncertainty

To best approximate the possible uncertainty in the rigid registration, we placed an Alderson Radiation Therapy phantom (ART-210, Radiology Support Devices, Ramsey, NJ) in the treatment chair, and images were acquired within the range of imaging parameters used to acquire patient images. The chair and phantom were shifted by a known amount and the images were registered. The difference between the registration and the true table positon provides a measure of the uncertainty in our rigid registration technique.

Development of an Updated Chair

Results from this study and feedback from the patient questionnaire indicated several features of the current chair design which could be improved for both patient comfort and radiotherapeutic use. A second treatment chair was developed in-house using primarily acrylic pieces and an increased number of indexing options.

Results

Patient Cohort

The 5 patients in this study were all male, with a median age of 65 years (range: 55–78 years), mean height of 181.1 cm (range: 180–183 cm), and mean weight of 88 kg (range: 76–111 kg). Female subjects ($n > 8$) were positioned in the chair during trial development and found no difficulties in positioning or comfort. On the basis of the feedback from the first two patients imaged, the face piece was changed from the bolus-based chin-and-forehead piece, to a suctionable cushion which conforms to the patient's face (Figure 5). Further, a small piece of loop fastener was applied to the top of the face piece as a barrier between the seam of the plastic and the patient's forehead.

Figure 5. Face piece of the chair before and after the change implemented after feedback from the first two patients. Before the change, the chin and forehead pieces were covered with bolus material for comfort, and the inferior chin piece was arched for anatomical conformity.

Image Registration

Table 1 lists the intra- and inter- fraction displacement for the 6 sub-regions measured. Rotation displacement was found to be small, ranging between -0.2° and 0.7°. The error in the registration, as measured with the phantom measurements was found to be no more than 0.4 mm. Average intra- fraction displacements were less than 2 mm across all patients. Average inter-fraction displacements were less than 3 mm. The largest displacements were seen in the anterior-posterior direction. Image guidance improved inter-fraction patient set-up in the anterior-posterior and left-right directions by an average of 1mm.

Table 1. Intra-fraction and inter-fraction displacements with and without image guidance in the treatment chair

	Mean displacement (mm) ± SE [Range] (n=5)								
	Intra-fraction			Inter-fraction					
				Without IGRT			With IGRT		
	S-I	A-P	L-R	S-I	A-P	L-R	S-I	A-P	L-R
C1C3	0.1±1.2 [-1.8-1.9]	1.2±2.5 [-3.3-4.7]		0.5±2.0 [-1.8-3.3]	-2.5±5.3 [-6.8-6.1]		Used for IGRT	Used for IGRT	
C3C5	-0.1±1.1 [-1.8-1.1]	1.2±3.3 [-4.9-5.9]		0.2±2.4 [-3.2-3.4]	-2.0±5.7 [-6.8-7.1]		-0.3±0.7 [-1.3-0.4]	0.3±0.5 [-0.1-1.0]	
Mandible	0.1±1.1 [-1.1-2.3]	0.5±1.6 [-1.7-3.6]		1.0±1.8 [-1.1-3.7]	-1.2±4.3 [-7.1-4.0]		0.5±1.3 [-1.6-2.0]	1.1±3.6 [-2.1-7.0]	
Occipital Bone	0.2±1.4 [-2.3-1.6]	0.3±2.5 [-5.5-3.5]		-0.3±2.3 [-3.2-3.0]	-2.7±4.3 [-7.4-1.5]		-0.8±0.4 [-1.4--0.3]	-0.4±2.6 [-4.6-2.3]	
Nasal Cavity	0.4±1.8 [-3.2-2.4]		-0.7±1.2 [-3.2-1.0]	0.4±2.0 [-1.9-3.5]		2.1±3.4 [-1.0-7.3]	-1.1±0.9 [-1.9-0.3]		1.7±6.8 [-9.1-7.6]
Left TMJ	0.6±1.6 [-2.5-3.0]		-0.8±1.8 [-4.7-1.2]	0.3±1.9 [-2.1-3.0]		3.0±4.1 [-1.5-8.4]	-1.3±1.1 [-2.4--0.1]		2.6±4.5 [-4.0-7.8]

SE, standard error; S-I, superior-inferior; A-P, anterior-posterior; L-R, left-right; IGRT, image-guided radiation therapy; C1C3 and C3C5, cervical vertebrae 1–3 and 3–5, respectively; TMJ, temporomandibular joint

Table 2. Comparison of inter-fraction displacements in the seated and supine treatment positions for simulated image guidance with respect to cervical vertebrae 1–3

Region of interest	Mean displacement (mm) ± SE		
	Upright position (this study)	Supine position	
		Kapanen et al(70)*	van Kranen et al(69)
Cranial-caudal			
C3C5	-0.3 ± 0.7	1.2*	0.10 ± 0.00
Mandible	0.5 ± 1.3	2.9	1.30 ± 2.50
Occipital bone	-0.8 ± 0.4	1.3	0.60 ± 2.0 0
Anterior-posterior			
C3C5	0.3 ± 0.5	3.1*	0.10 ± 0.50
Mandible	-1.1 ± 3.6	2.2	-0.30 ± 1.20
Occipital bone	-0.4 ± 2.6	1.9	0.30 ± 0.60

*Standard errors (SE) were not reported by Kapanen et al. Additionally, cervical vertebrae 1–2 (C1C2) were used as a reference, and C5C7 data were reported instead of C3C5 data.

Patient Questionnaire

Patients were asked to rate various aspects of their treatment in the supine and seated positions by completing a questionnaire consisting of 15 items. Fourteen of the fifteen resultant comparisons were less than one point apart on a 6 point (0-5) scale. In Figure 6, the questions separated by 4 tenths of a point or more are illustrated. Regarding comfort in the arms during treatment, patients preferred the seated position over the supine position, with a mean score of 4.6, compared with 3.6 for the supine position (a score of 5 corresponded to “perfectly comfortable”). Patients also had the opportunity to provide written and verbal feedback about the treatment experience. Verbal feedback included discomfort at the chin and lips, which was partially alleviated with the change in the face piece, as indicated by fewer verbal reports of discomfort after the change was made. Pressure from the head mask and pressure at the chest from the Vac-Lok cushion were also noted. Several patients

expressed the expected benefit of a deeper seat cushion. One patient requested a strap around the back to help prevent slouching and to remind the patient to relax forward into the chair.

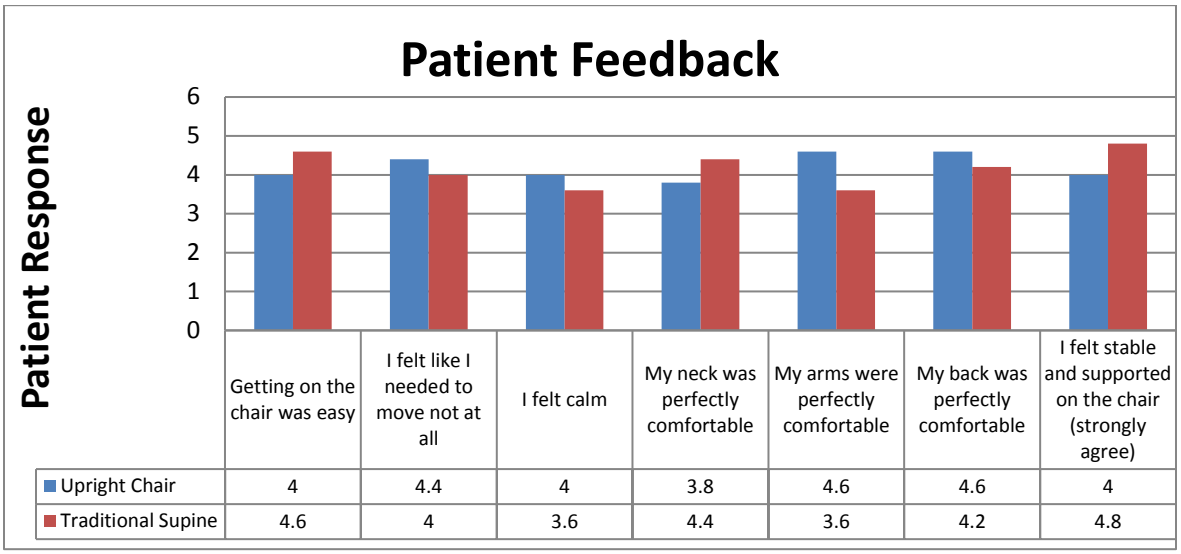


Figure 6. The results of the patient questionnaire. Only questions separated by an average of 0.4 points (5 point scale) or more are shown. The full questionnaire can be found in the supplemental materials. The questionnaire alternated the score assigned to a positive response. For example, a rating of 5 was assigned to answers of “I felt calm” and “Getting on the chair (couch) was difficult”. In this figure all positive responses are correlated to a ratings of 5 and text has been altered for clarity.

Development of an Updated Chair

We modified the chair design to reflect patient feedback and our accumulated experience. Changes include an increase in seat depth (from 23 cm to 45 cm), the footrest/chest/face piece no longer attaches to of the seat portion between the patients’ legs but rather on the outside of one’s hips, and for dosimetric consideration the thick supportive materials at the chest piece were removed. A schematic and picture of the new treatment chair can be found in Figure 7.

.

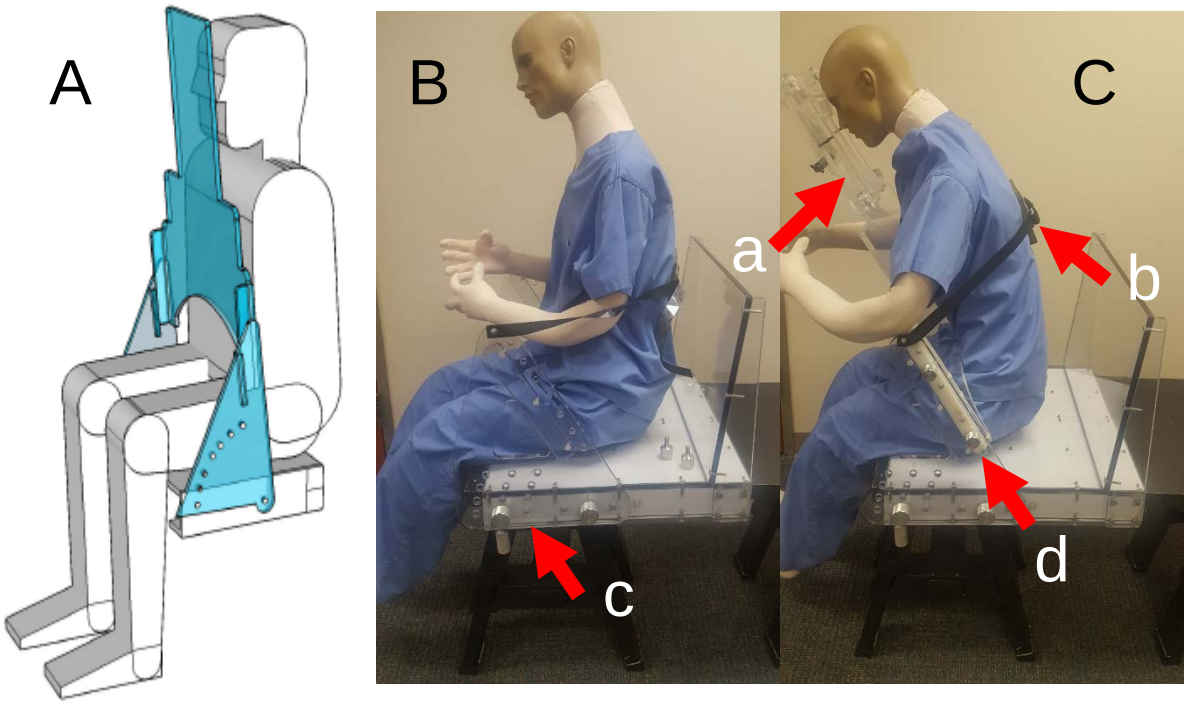


Figure 7. New radiotherapy chair design. Shown are the pre-build schematic (panel A) and photos of the seat only (Panel B) and the full chair (Panel C). The new chair design features adjustments for the angle of the face piece (a), a back strap for patient comfort (b), adjustments for chest plate depth and angle (c), and adjustments for chest plate height (d).

Discussion

As radiation therapy treatment planning has moved almost entirely to 3-dimensional methods, the acquisition of CT scans for planning has become routine in many clinics around the world. The horizontal bore of such scanners is a limiting factor for possible patient positions and therefore nearly all patients are treated lying in a prone or supine position. However, this position may not be suitable for all patients, especially those suffering from orthopnea, dyspnea, or dysphagia. Further, dosimetric considerations may indicate upright or

seated patient positioning. As techniques for image acquisition at the treatment unit continue to advance, for treatment paradigms still reliant on 2-dimensional planning techniques, for patients unable to tolerate a lying treatment position, and for the development of a fixed-beam low-cost system an upright treatment chair may prove optimal for treatment. We have designed a treatment chair compatible with current linear accelerator geometries and have tested patient intra- and inter- fraction displacement for the head and neck region. Patient displacement was on average less than 2 mm in the intra- and 3 mm in the inter- fraction scenarios. These raw inter- fraction measurements prove much better than those found for a previous upright setup for mantle treatments, for which all patients required block shifts of at least 5 mm, and 35% requiring shifts greater than or equal to 1 cm(61).

We also evaluated inter-fraction displacements in a scenario of simulated image guidance. While in clinical scenarios the radiation therapist would typically compare the whole acquired image to a planning image for use in image guidance, we have used only a sub-region of the acquired image to simulate image guidance. This approach is consistent with techniques used previously (69) and both the mean and standard deviation of inter-fraction displacement in the seated position in our study are on the same order as those reported for the traditional supine-position techniques (Table 2).

There are limitations to this technique. One patient was not able to complete the testing, and review of his images before the trial was aborted suggests that he had significant intra-fractional displacement (up to 3.3 cm). This was likely due to the fact that he was falling asleep and not feeling well, resulting in significant positional changes. While this only affected one patient, this may be more widespread; our attempts to create a treatment chair that is better tolerated than the supine position may not be tolerated by some patients. Furthermore, we largely enrolled “healthy” patients who tolerated the supine position quite well, and they also tolerated the upright position quite well. It remains to be seen how patients with significant medical issues (for instance, orthopnea, dyspnea, thick secretions)

and for patient nearing the end of treatment may tolerate the upright treatment, and whether it reflects an improvement over the supine position. These issues will be investigated in future studies.

The tight geometry of the gantry system and the chair tested in our study is partially a result of a minimum vertical height of the treatment couch and seat height; therefore, care must be taken when positioning the patient. We estimate that by the sixth patient setup took approximately 8 minutes including marking of lasers on head mask, similar to that for supine positioning. A complete assessment of the shift accuracy of the used registration algorithm was completed. However, while the rotational displacement of patient images was small, less than 1°, a similar analysis was not completed for the rotational accuracy of the registration algorithm and is potential source of error in this study.

The ability to acquire treatment planning images in the upright position, mirroring that of treatment, is an important aspect of the complete treatment process in the upright position. Using onboard imaging systems or other techniques, the acquisition of planning images in the treatment position is possible. Herein we have explored the set-up reproducibility of the upright treatment position in an in-house built chair. The results show the chair to have inter- and intra- fraction set-up reproducibility similar to current supine techniques. This works support the further investigation of the use of this position in the development of a fixed beam and rotating patient treatment paradigm. Such a position could dramatically reduce the upfront and ongoing cost of radiotherapy machines which may aide in their implementation in low-resource and LMIC settings.

Conclusion

In conclusion, our preliminary tests indicate that it is feasible to create an upright treatment chair with geometry suitable for 3-D imaging (with cone beam CT) and robust reproducible patient position between and within radiotherapy fractions. This work supports

this vision of the a system whereby patients can be simulated and treated in an upright position without degradation of a conformal, modern radiation treatment plan towards a fixed-beam low-cost radiotherapy system.

Chapter 4: A Single Optimization Treatment Planning Strategy in the Head-and-Neck

In this chapter we describe the results for Specific Aim 2 which pertains to the development of a fully automated single optimization head-and-neck treatment planning approach. Our working hypothesis is that a single optimization head-and-neck treatment planning approach produces plans which perform with equal quality to clinically acceptable plans and 90% of which are accepted by radiation oncologists for clinical use without edit.

Introduction

Towards a fully automated treatment planning approach for the head-and-neck an optimized and validated plan optimization approach is the most important aspect. The treatment of head-and-neck tumors; in close proximity of up to 25 organs at risk and divided into multiple target dose levels, represents one of the most challenging sites in treatment planning. Interest in automated plan optimization strategies has increased dramatically and researchers have identified several approaches which have found success for prostate, lung, and the head-and-neck cancers (54, 71-74). These approaches, however are subject many limitations, including their evaluation for only a small subset of patients, the need for manual adjustment or fine-tuning after automatic optimization, and the need for manual input including beam parameters or patient-specific dose parameters.

We seek to develop a fully automated planning strategy in the head-and-neck which produces plans of equal or superior quality compared to clinically treated plans, does not require the user to select beam angles, or modify the final treatment plan. Such a technique could dramatically reduce the human effort needed to generate treatment plans and ultimately reduce the deficient of highly trained radiotherapy staff in LMICs and represents a critical component of a fully automated treatment planning system.

Methods

Patient Cohort

For this analysis, 54 patients treated at The University of Texas MD Anderson Cancer Center selected as to represent seven head-and-neck subsites; larynx, nasopharynx, oral cavity, oropharynx, paranasal sinuses and cavity, and salivary glands and 30 patients from other institutions treated on a clinical trial were retrospectively collected. The clinical trial data was obtained through The Cancer Imaging Archive (75) and was originally derived from the Radiation Therapy Oncology Group 0522 study (76). Selection criteria for both patient cohorts included a maximum of three planned target volume dose levels, the use of IMRT or VMAT treatment delivery, and availability of the treated dose distribution for comparison. Autoplans were generated for each of the patients and compared against clinical plans at clinical dose constraints and relevant endpoints. Additionally, 20 patient CTs, 10 collected from each of 2 partner institutions in South Africa, were used to evaluate the single optimization treatment planning approach but were not compared to the clinically treated plans.

Planning Strategy

The planning strategy mirrors the clinical planning approach but competes each step automatically and without human intervention.

Plan Initialization

Primarily developed by other members of our group, plan initialization steps includes assignment of the target prescription dose levels, removal of the treatment couch, selected of the treatment isocenter, initial setup of beam parameters, and contouring of any structures (planning or otherwise) required for treatment planning but not included in the clinical

treatment plan. Target prescription dose levels were matched to the clinically delivered plan. The treatment isocenter was selected to be at the center of all combined planning target volumes (PTVs). Either three or four 360-degree treatment arcs were selected depending on target size and orientation. The collimator angles for the first two arcs were set to be 30 and 330 degrees and jaws were determined such that the entire target remained in the beam's eye view (BEV) as the gantry rotates through 360 degrees. If the field size exceeded 18 cm in the x-direction (the direction of travel of the multi-leaf collimators (MLCs)) then the jaw was set to be symmetric with a field size of 18 cm. This constraint was imposed on the jaw settings due to mechanical constraints of the MLCs modeled in the treatment planning system (Varian 2100 series linear accelerator, Millennium 120 MLCs) which have a maximum distance of travel of 14.5cm, and in an effort to design a planning approach widely applicable to many machine types which, for example, may not have jaw tracking capabilities. The collimator was then set to 90 degrees and the jaws set such that all PTVs remained in the field of view with a 1cm margin over 360 degrees of rotation. If this required that the x jaw size exceed 18 cm, a fourth field was added such that the two fields cover the PTVs, one from the most superior extent and from the most inferior and both with a maximum size of 18cm. Structures required for plan optimization which were not included in the original treatment plan, including planning and normal structures, were automatically contoured using an in-house multi-atlas deformable image registration technique, details of which are described in Chapter 5.

Dose Optimization and Knowledge Based Planning

A knowledge based treatment planning module, RapidPlan®, has been implemented and is available for clinical use in the Eclipse Treatment Planning system (Varian Associates, Palo Alto, CA). Using a number of previously treated plans, from which both the geometries of the targets and organs at risk as well as the planned dose distributions are extracted, DVH estimation models are used to parameterize a DVH estimation algorithm which sets the

constraints for new patients. The methodology employed by the proprietary RapidPlan® software is based primarily on work from Duke University Medical Center which characterizes inter-patient organ at risk sparing (77). To estimate the DVH for each organ at risk, quantitative metrics are used which include the distance to target histogram, the relative relationship between the OAR and the targets, and additional anatomical features including; the relative volume overlap, the relative out-of-field volume, the absolute OAR volume and the absolute target volume (78). In RapidPlan® the user has the capability to create their own DVH estimation models based on patients selected by the user, though there are two vendor provided DVH estimation models for the head-and-neck – the "*CancerCare Manitoba Head and Neck"* and the "*Washington University Head & Neck* models". The latter served as the basis for this work. The details of the data set used to develop the Washington University Head & Neck Model can be found in the provided model description. In short, the model was trained and tested on plans with targets delineated in the nasopharynx, oral cavity, oropharynx, hypopharynx, larynx, or other unknown head-and-neck primary anatomical locations. The model allows for one, two, or three targets and estimates the DVH curve for the brain, brainstem, upper esophagus, larynx, lips, spinal cord, mandible, middle ear, oral cavity, parotid glands, pharyngeal constrictors, and submandibular glands. Provided in the model description are contouring guidelines for both targets and organs at risk. The model was trained and validated on unilateral and bilateral cases planned as head first, supine, 6-9 field IMRT cases with 6x photons, couch rotation of 0 degrees and fields with allowable gantry angles in increments of 40 degrees from 0 to 320 degrees. The objectives and relative priorities used for clinical training cases can be found in the model description.

For this study we used the model-provided line constraints for normal structures and modified or added new templated constraints for both structures with model provided line constraints and other structures. Through iterative testing structure constraints and priorities were optimized based on quadrative analysis and physician feedback. The final constraints

used for the planning strategy presented here can be found in the Appendix. As is currently mandatory in the Eclipse TPS when using RapidPlan®, the Photon Optimizer (PO) algorithm (version 13.5.35) was used to optimize VMAT plans. Dose was calculated using the anisotropic analytic algorithm (version 13.5.35) implemented in the Eclipse treatment planning system.

To ensure adequate coverage of each target autoplans were normalized such that 95% of the target volume received at least 98% of the prescribed dose. To reduce plan sensitivity to normalization, normalization criteria were implemented. Structures with volume less than 20cc, which can greatly effect plan normalization especially if the structure is in a high gradient area, and structures with more than 20% of the target volume having Hounsfield unit less than -800, in which high doses are not theoretically achievable and the uncertainty in dose calculation algorithms is high, were excluded for normalization purposes, unless it was the only target.

Quantitative Plan Analysis

For evaluation, 4 patient groups were considered; (1) all patients with corresponding clinical plans, (2) only patients treated at MD Anderson, (3) only patients treated at other institutions as part of a clinical trial, and (4) only patients from partner institutions in South Africa for which corresponding clinical plans were not available. The autoplans were evaluated at typical clinical constraints and those outlined in RTOG protocol 1016. Autoplans were assessed as the mean, standard deviation, minimum and maximum values across patients at dosimetric points of interest, and the percentage of plans meeting the clinical constraint.

For plans with corresponding clinically delivered plans, autoplans were also compared against the corresponding clinical plans at many DVH endpoints using the Wilcoxon signed rank test, a nonparametric statistical hypothesis test based on rank orders for two planning groups. The two planning techniques were also compared using the Brown-Forsythe test, a

non-parametric test which tests for the equality of the variances in the two planning groups by examining the absolute distance of each point from the median of the distribution(79). For both tests, significance was established as a p-value of less than 0.05. When comparing plans using the Wilcoxon signed rank test, plans were normalized such that DVHs were matched at the dose received by 95% of the high dose PTV, allowing for a fair comparison of OAR sparing and target coverage.

Physician Review of Autoplans

Throughout the process, specialized head-and-neck radiation oncologists were consulted concerning plan quality. Feedback was used to improve the planning strategy. Once the planning strategy was finalized, plan review documents were created for 40 patients, 20 from the cohort of patients treated at MD Anderson and 20 from the cohort of patients treated on a clinical trial. Review documentation included CT slices with overlaid dose distributions, beam and field information, patient information and DVH curves for target and normal structures. A head-and-neck radiation oncologist was asked to rate the plans on a three point scale, either needing no edit for clinical use, needing minor edit, or needing major edit.

Results

VMAT optimization of plans took 5.5 ± 2.0 minutes (average ± standard deviation) and dose calculation took 9.0 ± 3.3 minutes, an additional 1 minute was required for additional steps including opening the plan and setting the prescription. Other pre-planning activities were not systematically recorded for this study, but for a sample of ten patients automated tasks which include the removal of the treatment couch, detection of the body, selection of the isocenter, determination of the field parameters, contouring of planning structures, data format conversion, and import of DICOM files into Eclipse took 22.5±1.1 minutes, with the

largest time commitment (on average, 18 minutes) coming from contouring. The entire planning process is therefore estimated to require an average of 37 minutes.

Patient Cohort

Patient characteristics can be found in *Table 3*. One of the thirty patients collected and treated on the clinical trial had a CT scan with variable slice spacing, which is not compatible with the necessary autocontouring and was thus excluded from analysis.

Table 3. Characteristics of patients used to validate the treatment planning strategy. In parenthesis are additional statistics of the 20 patients collected from the South African partner institutions

Characteristic	Number
Number of Target Dose Levels	
3	56 (4)
2	18 (9)
1	9 (7)
Gender	
Male	38
Female	16
Unknown	29 (20)
Sub-site	
Hypopharynx	3(1)
Larynx	12(3)
Nasopharynx	5(1)
Oral Cavity	12(3)
Oropharynx	41(2)
Paranasal Sinuses	5
Salivary Glands	5
Unknown	20(10)

Quantitative Plan Analysis

Plans were normalized such that 95% of each of the PTVs received at least 98% of the prescribed dose, excluding structures with volume less than 20cc or for which more than 20% of the PTV had Hounsfield unit less than -800, as previously discussed. Of 250 total

targets, 2 had more than 20% of the volume with Hounsfield unit less than -800 (22% and 26%) and 8 had volume less than 20cc [range 4-19 cc]. All ten of these exceptions were from the MD Anderson patient data set. Of the ten exceptions only four targets received less than 98% of the prescribed dose to 95% of the volumes after plan normalization. The structure chosen for normalization as well as the percentage of dose covering 95% of the target volume can be found for all targets in Figure 8. Forty six of the 103 autoplans were normalized to the high dose PTV, 25 to the intermediate dose PTV, and 32 to the low dose PTV.

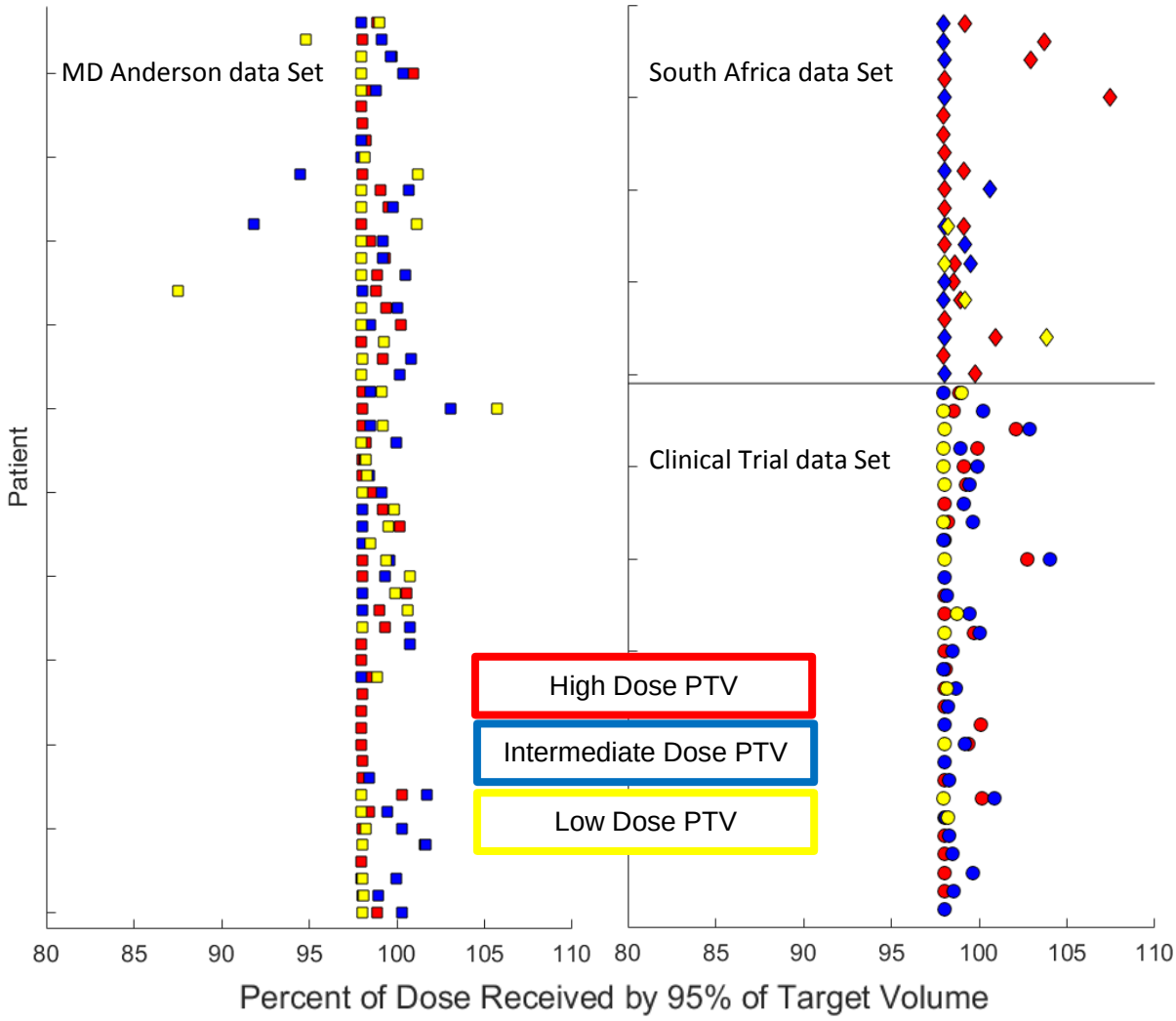


Figure 8. Coverage of 95% of target volumes , used for normalization. High dose PTVs are shown in red, intermediate dose in blue, and low dose in yellow. Patients from the MD Anderson data set in squares, from the clinical trial data set are displayed as circles, and from the South African data set in diamonds. The target receiving 98% of the prescription dose to 95% of the volume was chosen for normalization.

Quantitative metrics describing the autoplan and corresponding clinical plans including the mean, standard deviation, and minimum and maximum value at clinically relevant dosimetric endpoints can be found in Table 4 and the fraction of both clinical and autoplans meeting clinical constraints in Table 5. Clinical constraints were, on average, met in clinical plans more often for patient cohort from MD Anderson (86%) than those treated on a clinical trial (72%).

Constraints least often met were the mean dose less than 39 Gy the submandibular glands, for which of 83 clinical plans 16% and 35% met for ipsilateral and contralateral glands, respectively, and autoplans met less often with rates of 12% and 28%, respectively. Next least often met was the clinical constraint of a mean dose less than 26 Gy to the parotid glands, for which clinical plans met with a rate of 46% and 83% for contralateral and ipsilateral glands respectively. Autoplans met these constraints more often at rates of 51% and 88%.

Of the 103 autoplans, all but two had maximum spinal cord dose less than 45Gy. One plan, from the MD Anderson data set had a spinal cord maximum dose of 45.1Gy and while exceeding a 45Gy constraint, meets a constraint of less than 0.03cc with dose greater than 48Gy which has been reported in the literature(80, 81). The other plan not meeting this constraint had the spinal cord contour within 1.3mm of the high dose PTV, and had a maximum dose of 48.44Gy, this patient, from the cohort of patients from South Africa, did not have a corresponding clinical plan. Similarly, two patients from the South African data set did not meet the clinical constraint of brainstem maximum dose less than 54 Gy, having maximum doses of 60.2 Gy and 57.4 Gy. For these patients the brainstem contour and high dose PTV overlapped with volumes of 0.3 and 0.1cc. These patients failing to meet brainstem and spinal cord constraints can be seen in Figure 9.

Table 4. Quantitative evaluation of autoplans and corresponding clinical plans (when available) at clinically relevant dosimetric points.

	AutoPlans		Clinical Plans		AutoPlans		Clinical Plans	
	mean±std	[min-max]	mean±std	[min-max]	mean±std	[min-max]	mean±std	[min-max]
Structure	**Spinal Cord, Maximum Dose [Gy]**				**Brainstem, Maximum Dose [Gy]**			
All Patients (n=83)	39.2±5.2	[4.8-45.1]	32.2±11.6	[2.8-48.9]	33.6±15.3	[0.0-48.3]	28.6±18.5	[0.0-71.6]
MDACC Patients (n=54)	38.0±6.2	[4.8-45.1]	26.0±9.6	[2.8-48.9]	29.3±16.7	[0.0-48.3]	18.5±14.5	[0.0-49.1]
Clinical Trial Patients (n=29)	41.1±0.8	[40.2-43.7]	43.7±3.0	[38.8-48.9]	41.6±7.3	[16.5-47.7]	47.2±6.9	[33.5-71.6]
South Africa Patients (n=20)	41.0±3.8	[32.0-48.4]			28.7±20.2	[1.7-60.2]		
Structure	**Ipsilateral Parotid, Mean Dose [Gy]**				**Contralateral Parotid, Mean Dose [Gy]**			
All Patients (n=83)	27.6±16.2	[0.1-67.8]	30.4±17.7	[0.0-70.5]	16.5±7.6	[0.1-34.9]	16.1±9.8	[0.0-33.2]
MDACC Patients (n=54)	26.6±18.8	[0.1-67.8]	27.4±19.6	[0.0-70.5]	14.0±7.5	[0.1-34.9]	11.0±7.9	[0.0-23.4]
Clinical Trial Patients (n=29)	29.5±9.9	[16.5-57.3]	36.0±12.1	[23.3-66.0]	21.2±5.1	[12.9-29.2]	25.6±4.6	[15.3-33.2]
South Africa Patients (n=20)	38.5±17.7	[6.0-64.3]			24.3±14.2	[5.6-50.7]		
Structure	**Ipsilateral Parotid, Volume Receiving > 30Gy [%]**				**Contralateral Parotid, Volume Receiving > 30Gy [%]**			
All Patients (n=83)	37.1±29.1	[0.0-100.0]	44.4±31.0	[0.0-100.0]	14.6±13.6	[0.0-52.7]	18.2±14.4	[0.0-49.0]
MDACC Patients (n=54)	35.9±33.5	[0.0-100.0]	40.2±33.3	[0.0-100.0]	10.4±12.1	[0.0-52.7]	11.4±11.3	[0.0-32.6]
Clinical Trial Patients (n=29)	39.2±18.8	[16.5-57.3]	51.9±25.0	[12.7-100.0]	22.5±13.0	[0.8-43.0]	30.5±10.6	[3.7-49.0]
South Africa Patients (n=20)	56.6±29.4	[6.0-64.3]			31.8±27.7	[0.0-86.7]		
Structure	**Ipsilateral Submandibular Gland, Mean Dose [Gy]**				**Contralateral Submandibular Gland, Mean Dose [Gy]**			
All Patients (n=83)	57.9±19.6	[0.5-72.4]	58.2±19.2	[0.5-74.9]	48.6±19.9	[0.4-72.1]	43.8±24.9	[0.5-73.0]
MDACC Patients (n=54)	53.4±22.1	[0.5-72.4]	53.1±20.8	[0.5-72.3]	42.6±21.3	[0.4-72.1]	34.2±24.9	[0.5-72.1]
Clinical Trial Patients (n=29)	66.2±9.9	[20.2-72.3]	67.9±10.5	[17.3-74.9]	59.7±10.3	[23.6-71.9]	61.810.7	[19.7-73.0]
South Africa Patients (n=20)	58.1±15.8	[26.4-75.3]			49.3±18.3	[11.9-73.0]		
Structure	**Cochleae, Maximum Dose [Gy]**				**Brain, Maximum Dose [Gy]**			
All Patients (n=83)	13.1±15.1	[0.0-69.5]	13.6±14.8	[0.0-72.9]	38.2±19.1	[0.0-69.5]	41.019.9	[0.0-71.9]
MDACC Patients (n=54)	14.8±17.2	[0.0-69.5]	10.0±13.5	[0.0-72.9]	35.6±22.1	[0.0-69.5]	34.521.4	[0.0-70.3]
Clinical Trial Patients (n=29)	10.1±9.2	[2.2-44.5]	20.5±14.6	[2.1-57.4]	43.2±9.9	[13.3-57.1]	53.27.5	[37.3-71.9]
South Africa Patients (n=20)	19.6±23.5	[0.6-75.2]			34.2±9.9	[1.5-78.5]		

Table 4. Continued from previous page.

Structure	Optic Chiasm, Maximum Dose [Gy]				Optic Nerves, Maximum Dose [Gy]			
All Patients (n=83)	5.1±10.0	[0.0-62.0]	4.9±10.5	[0.0-52.5]	5.5±11.2	[0.0-61.7]	5.6±12.2	[0.0-57.8]
MDACC Patients (n=54)	6.6±12.2	[0.0-62.0]	6.2±12.0	[0.0-52.5]	7.4±13.5	[0.0-61.7]	7.2±13.8	[0.0-57.8]
Clinical Trial Patients (n=29)	2.1-0.9	[1.0-4.2]	1.4±1.2	[0.0-4.7]	2.1±0.8	[0.9-4.1]	1.2±1.3	[0.0-4.6]
South Africa Patients (n=20)	9.8±16.6	[0.5-64.6]			12.5±22.1	[0.5-71.4]		
Structure	**Lens, Maximum Dose [Gy]**				**Eyes, Maximum Dose [Gy]**			
All Patients (n=83)	2.1±2.5	[0.0-15.6]	1.7±2.5	[0.0-12.2]	5.4±10.8	[0.0-59.7]	4.8±10.0	[0.0-56.9]
MDACC Patients (n=54)	2.4±3.0	[0.0-15.6]	1.9±2.6	[0.0-12.2]	7.1±13.0	[0.0-59.7]	6.3±12.0	[0.0-56.9]
Clinical Trial Patients (n=29)	1.5±0.5	[0.7-3.4]	1.1±1.7	[0.0-7.1]	2.2±1.1	[0.8-7.0]	1.9±2.3	[0.0-12.8]
South Africa Patients (n=20)	3.5±5.0	[0.3-21.4]			8.7±14.6	[0.4-49.0]		
Structure	**High Dose PTV, Dose Received by hottest 1cc [% of Rx]**				**High Dose PTV, Volume Receiving 95%Rx [%]**			
All Patients (n=83)	106.4±2.1	[102.7-113.5]	106.6±2.7	[97.5-114.5]	98.7±1.0	[98.0-102.8]	100.2±1.6	[89.8-102.0]
MDACC Patients (n=54)	106.7±2.3	[102.7-113.5]	105.2±1.5	[97.5-107.6]	98.7±0.9	[98.0-101.6]	100.1±1.9	[89.8-102.0]
Clinical Trial Patients (n=29)	105.9±1.5	[103.9-110.1]	109.1±2.5	[104.1-114.5]	98.8±1.3	[98.0-102.8]	100.3±0.5	[99.0-101.1]
South Africa Patients (n=20)	108.7±3.4	[106.1-119.4]			99.5±2.5	[98.0-107.5]		
Structure	**Intermediate Dose PTV, Volume Receiving 95%Rx [%]**				**Low Dose PTV, Volume Receiving 95%Rx [%]**			
All Patients (n=83)	99.1±1.6	[91.8-104.0]	100.6±1.7	[93.3-106.6]	98.4±2.0	[87.5-105.7]	100.6±1.9	[93.0-107.7]
MDACC Patients (n=54)	99.1±1.8	[91.8-103.1]	100.5±1.5	[93.3-102.7]	98.5±2.3	[87.5-105.7]	100.6±1.6	[93.0-107.7]
Clinical Trial Patients (n=29)	99.1±1.4	[98.0-104.0]	100.7±2.0	[97.5-106.6]	98.2±0.3	[98.0-99.0]	100.5±2.6	[97.2-107.7]
South Africa Patients (n=20)	98.4±0.8	[98.0-100.6]			99.8±2.7	[98.1-103.8]		

Table 5. Fraction of clinical and autoplans meeting clinical dosimetric constraints

		MDACC & CT Patients (n=83)		Clinical Trial (CT) Patients (n=29)		MDACC Patients (n=54)		South Africa Patients (n=20)
Structure	**Constraint**	Clinical Plans	Autoplans	Clinical Plans	Autoplans	Clinical Plans	Autoplans	Autoplans
Spinal Cord	**Dmax<45Gy**	88%	99%	66%	100%	100%	98%	90%
Brainstem	**Dmax<54Gy**	96%	100%	90%	100%	100%	100%	85%
Ipsilateral Parotid	**Dmean<26Gy**	46%	51%	24%	38%	57%	57%	25%
Contralateral Parotid	**Dmean<26Gy**	83%	88%	52%	76%	100%	94%	50%
Ipsilateral Parotid	**V30Gy<50%**	64%	75%	55%	79%	69%	72%	40%
Contralateral Parotid	**V30Gy<50%**	100%	99%	100%	100%	100%	98%	75%
Ipsilateral Submandibular Gland	**Dmean<39Gy**	16%	12%	3%	3%	22%	17%	20%
Contralateral Submandibular Gland	**Dmean<39Gy**	35%	28%	3%	7%	52%	39%	25%
Cochleae	**Dmax<35Gy**	91%	89%	83%	97%	95%	85%	73%
Optic Chiasm	**Dmax<54Gy**	100%	99%	100%	100%	100%	98%	95%
Optic Nerve	**Dmax<54Gy**	99%	99%	100%	100%	98%	98%	89%
Lens	**Dmax<7Gy**	95%	95%	96%	100%	94%	93%	81%
Eyes	**Dmax<35Gy**	98%	98%	100%	100%	97%	97%	88%
Brain	**Dmax<54Gy**	70%	84%	59%	97%	76%	78%	70%
High Dose PTV	**V1cc<110%**	87%	93%	62%	97%	100%	91%	85%
High Dose PTV	**V1cc<117%**	100%	100%	100%	100%	100%	100%	95%
High Dose PTV	**V95%>95%**	74%	24%	72%	17%	76%	29%	100%
Intermediate Dose PTV	**V95%>100%**	100%	100%	100%	100%	100%	100%	8%
Intermediate Dose PTV	**V95>80%**	71%	9%	43%	0%	81%	12%	100%
Low Dose PTV	**V95>100%**	100%	100%	100%	100%	100%	100%	25%
Low Dose PTV	**V95%>78%**	98%	100%	100%	100%	96%	100%	100%

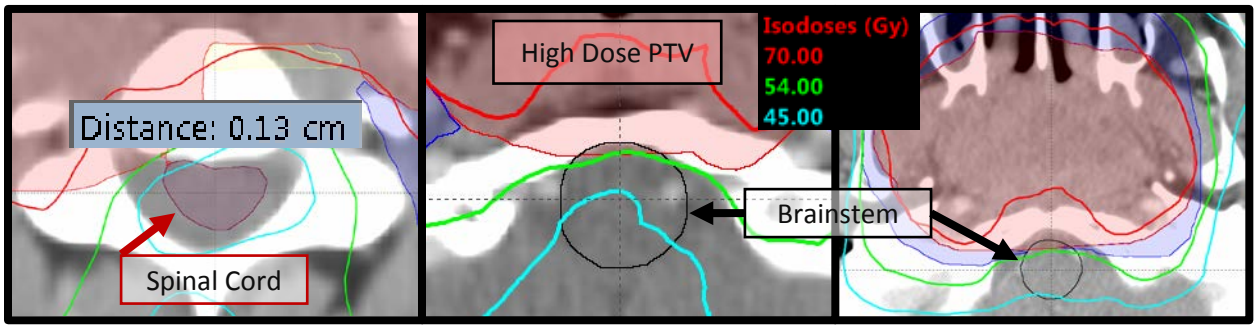


Figure 9. Examples of normal structures not meeting clinical constraints . In only these three patients did spinal and brainstem structures fail to meet clinical maximum dose constraints. In all cases the critical structure in question was in very near proximity or overlapping with the high dose target volume. In panel A, the spinal cord and high dose planning target volume are separated by only 1.3mm, in panels B and C the brainstem overlaps with the high dose planning target volume. These cases underline the need for sanity checks on contours prior to plan optimization.

A paired Wilcoxon signed rank test was used to compare the clinical and autoplans for three groups – all 83 patients with available clinical plans, 54 patients from MD Anderson, and 29 patients treated at on a clinical trial protocol. The p-value of the Wilcoxon signed rank test for each DVH metric can be found in *Table 6*. Considering all 83 patients with corresponding clinical plans, the autoplans significantly outperformed their clinical counterparts considering the volume of spinal cord receiving more than 45 Gy, the maximum dose and the volume receiving more than the clinical threshold of 54 Gy to the brain, the mean dose to the ipsilateral parotid and volume receiving 30 Gy to the ipsilateral and contralateral parotids. The clinical plans performed better considering the maximum dose to the spinal cord, brainstem, optic chiasm, optic nerves, eyes, and lens. However, for all these the plans were not significantly different considering the volume receiving more than the clinical maximum dose threshold. The clinical plans had significantly better sparing of the submandibular glands with

a lower average mean dose and significantly better coverage of the low dose PTVs at 95% of the prescription dose and a lower dose to 1cc of the high dose PTV. Similarly, when considering only the patients from the MDACC data set the clinical plans outperformed the autoplans on all these categories. The autoplans only outperformed the cohort of patients from MD Anderson considering the volume receiving more than 30Gy to the ipsilateral parotid, the mean dose to the contralateral parotid, and the volume of the intermediate dose PTV receiving more than 95% of the prescribed dose.

When considering the cohort of patients treated on a clinical trial, the autoplans significantly outperformed the clinical plans for a majority of evaluated constraints. The autoplans outperformed the clinical plans in terms of maximum dose and volume exceeding the maximum dose to the spinal cord, brainstem, brain, and cochlea, mean dose and V30Gy to both the contralateral and ipsilateral parotids, and dose to the hottest 1cc of the high dose PTV. The only categories in which the 29 clinical plans treated on a clinical trial outperformed the autoplans were the maximum dose to the optic nerves, chiasm, lens, and eyes, though for none of these did the clinical plans outperform the autoplans when considering the volume of the structure *exceeding* the clinical constraints and the maximum dose constraints were met for all patients.

Table 6. Comparison of autoplan performance to clinically treated plans. Shown are p-values of the paired Wilcoxon signed rank test. In green are parameters for which the autoplans outperformed the clinical plans, and in red which the clinical plans outperformed the autoplans.

		p-value, Wilcoxon signed rank		
Structure	Test Point	All Patients (n=83)	Clinical Trial Patients (n=29)	MDACC Patients (n=54)
Spinal Cord	D_max	<0.01	<0.01	<0.00
	V_45Gy	<0.01	<0.01	1.00
Brainstem	D_max	<0.01	<0.01	<0.01
	V_54Gy	0.25	0.25	1.00
Ipsilateral Parotid Gland	D_mean	<0.01	<0.01	0.23
	V_30Gy	<0.01	<0.01	<0.01
Contralateral Parotid Gland	D_mean	0.27	<0.01	<0.01
	V_30Gy	<0.01	<0.01	0.39
Ipsilateral Submandibular Gland	D_mean	0.02	0.15	<0.01
Contralateral Submandibular Gland	D_mean	<0.01	0.28	<0.01
Cochleae	D_max	0.51	<0.01	<0.01
	V_35Gy	0.11	0.01	<0.01
Optic Chiasm	D_max	<0.01	<0.01	<0.01
	V_54Gy	1.00	1.00	1.00
Optic Nerves	D_max	<0.01	<0.01	<0.01
	V_54Gy	0.50	1.00	0.50
Lens	D_max	<0.01	<0.01	<0.01
	V_7Gy	0.13	1.00	0.08
Eyes	D_max	<0.01	<0.01	<0.01
	V_50Gy	0.50	1.00	0.50
Brain	D_max	0.01	<0.01	0.01
	V_54Gy	0.01	<0.01	0.54
High Dose PTV	D_1cc	<0.01	<0.01	<0.01
	V_95%	0.69	1.00	1.00
Intermediate Dose PTV	V_95%	0.06	0.77	0.02
Low Dose PTV	V_95%	0.02	0.07	0.26

Considering the variance between the two planning strategies, the results of the Brown-Forsythe test for all patient with corresponding clinical plans are shown in *Table 7*. A significant p-value, as for maximum dose to the spinal cord and brainstem; and mean doses to the contralateral parotid and submandibular glands, indicates that the two distributions come from distributions with different variances. When considering all 83 patients, the standard deviation across plans was *less* for autoplans than clinical plans, suggesting the autoplanning technique provides plans with decreased inter-patient variability.

Table 7. Brown-Forsyth test of equal variance between autoplans and clinical plans. A significant p-value ($p<0.05$, shaded green) indicate that the distributions of clinically relevant DVH points have different variances. For each tissue the RPA plan distribution had a significantly smaller variation than the clinical plans.

		p value, Brown-Forsythe Test
Structure	Test Point	All Patients (n=83)
Spinal Cord	D_max	<0.01
Brainstem	D_max	<0.01
Ipsilateral Parotid	D_mean	0.40
	V_30Gy	0.61
Contralateral Parotid	D_mean	0.01
	V_30Gy	0.63
Ipsilateral Submandibular Gland	D_mean	0.81
Contralateral Submandibular Gland	D_mean	0.04
Cochleae	D_max	0.73
Optic Chiasm	D_max	0.74
Optic Nerves	D_max	0.50
Lens	D_max	0.37
Eyes	D_max	0.99
Brain	D_max	0.48
High Dose PTV	D_1cc	0.11
	V_95%	0.95
Intermediate Dose PTV	V_95%	0.82
Low Dose PTV	V_95%	0.34

Figure 10 describes the layout of Figure 11 in which the performance of the autoplanning strategy against the corresponding clinical plan for all 83 patients is considered at many clinically relevant dosimetric points. Over 83 patients and 18 structures, 4 bilaterally, for a total of 1719 DVH points considered (36 target volumes did not exist, and 71 structures were outside of the dose calculation region in the clinical plan), 610 (35%) had improved DVH metric for the autoplans, but the improvement did not change the whether the plans met clinical constraints. For 989 structures (58%), the clinical plan performed better at the endpoint, but the difference was not beyond the clinical threshold. However, for 76 structures (4%) the autoplan outperformed the clinical plan and this difference was beyond clinical thresholds, this is compared to 44 structures (3%) of endpoints for which the clinical plan performed better and was beyond clinical limits. Of the 4% of endpoints for which the autoplan provided an improvement at the dosimetric endpoint and this resulted in the plan meeting clinical constraints whereas the clinical plan had not met that constraint, 91% came from the population of patients treated on the clinical trial.

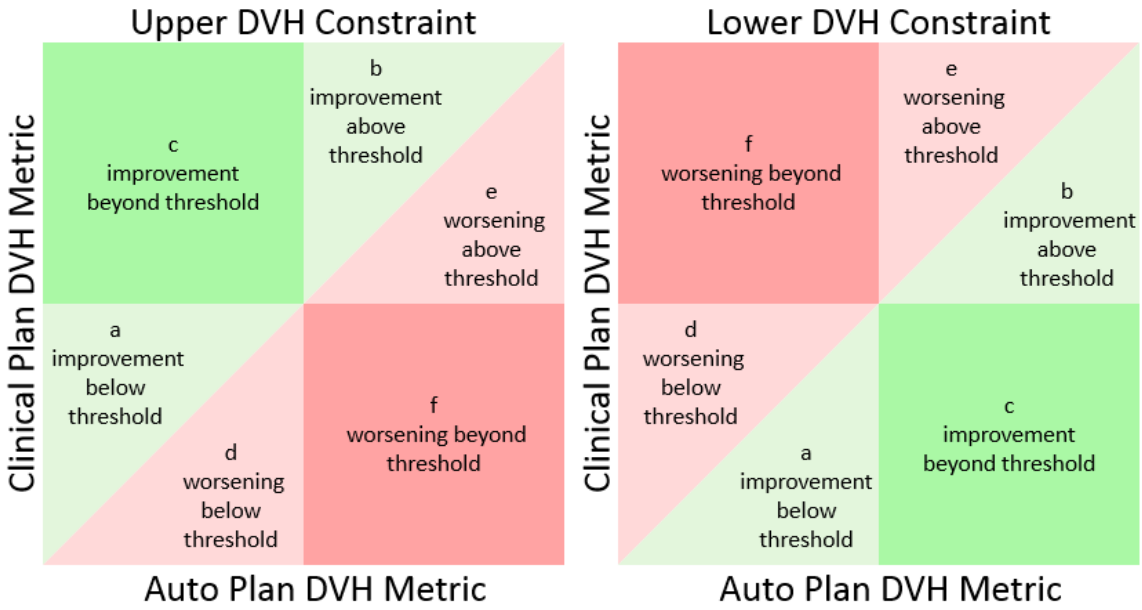


Figure 10. Explanation of clinical and autoplan comparison. For structures with an upper DVH constraint (i.e. spinal cord maximum dose less than 45Gy) (right), plan quality was improved if the DVH metric of the autoplan was less than the DVH metric of the clinical plan, these plans will be reflected in the shaded green regions. Plans for which an improvement was seen but both plans met the clinical constraint (e.g. spinal cord maximum dose of 43 Gy in the clinical plan and 42 Gy in the autoplan) are reflected in a, when both fail to meet the clinical constraint (e.g. spinal cord maximum dose of 47 Gy in the clinical plan and 46 Gy in the autoplan) in panel b. In the darker green region, panel c, the autoplan met the clinical constraint but the clinical plan did not (e.g. spinal cord maximum dose of 46 Gy in the clinical plan and 44 Gy in the autoplan). Similarly if the clinical plan outperformed the autoplan, this is reflected in the red shaded regions. If both plans met the constraint in panel d and if both missed the constraint in panel e. Importantly, if the clinical plan met the clinical constraint but the autoplan did not, the plan will be reflected in panel f. On the left, the same schema is used but for structures with a lower DVH constraint (e.g. volume of the PTV receiving 95% of the prescribed dose). An improvement in plan quality, shaded green, are seen when the autoplan has a higher value than the clinical plan. Color and letter definition are as previously described.

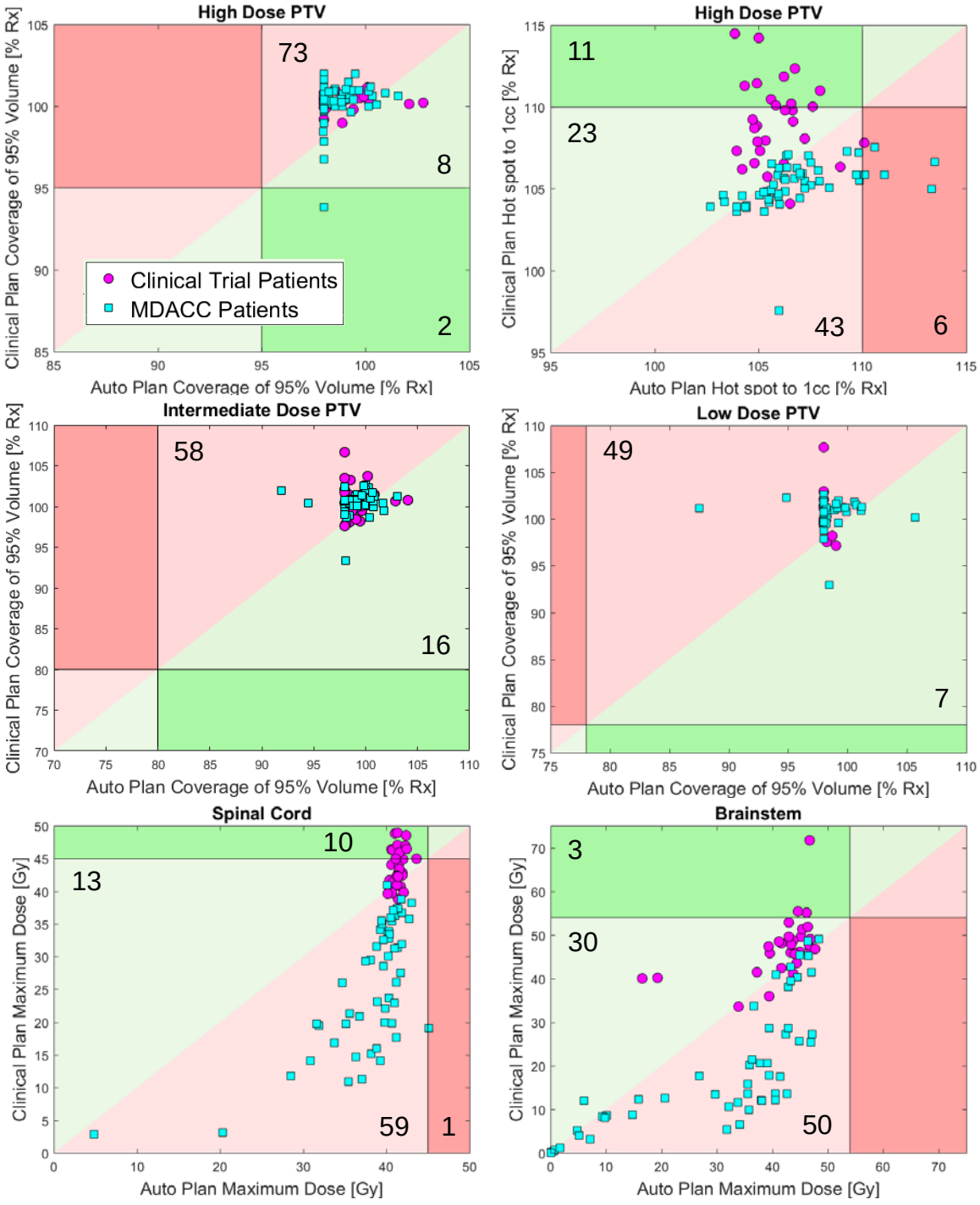


Figure 11. Distribution of autoplan and clinical plan performance at relevant dosimetric points. Autoplan values are shown along the x-axis and corresponding clinical plan value along the y-axis. Shades of green indicate improvement in the dosimetric descriptor for the autoplan, shades of red indicate poorer performance due to autoplanning. Darker shades delineate

common clinical thresholds for each dosimetric descriptor, with darker green indicating an improvement beyond the clinical threshold and darker red indicating a poorer performance beyond the clinical threshold. Twenty nine patients treated on a clinical trial are shown in magenta circles, and 54 patients from our institution shown in cyan squares. The number of patients in each group are shown (continued of following pages).

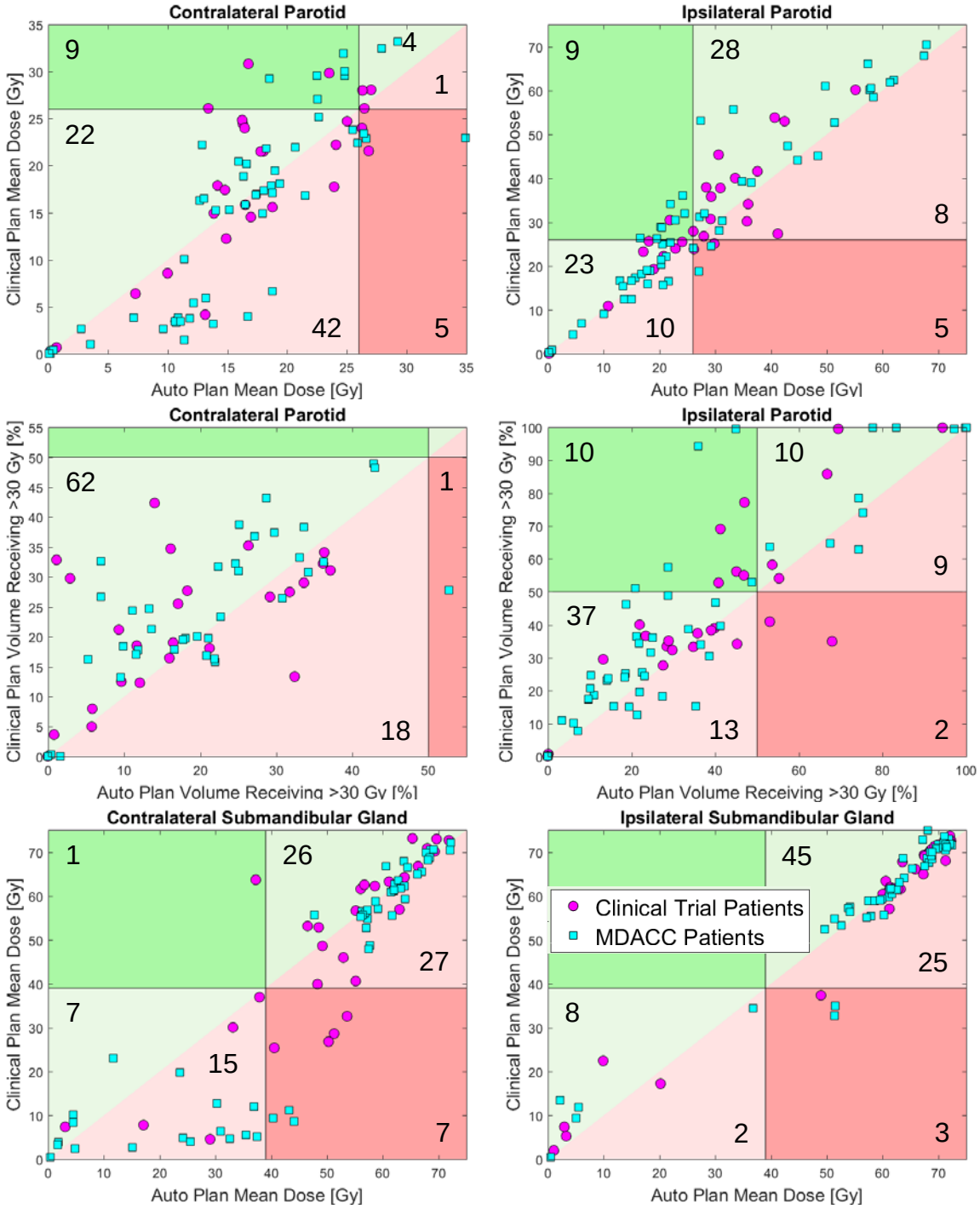


Figure 11, continued from previous page.

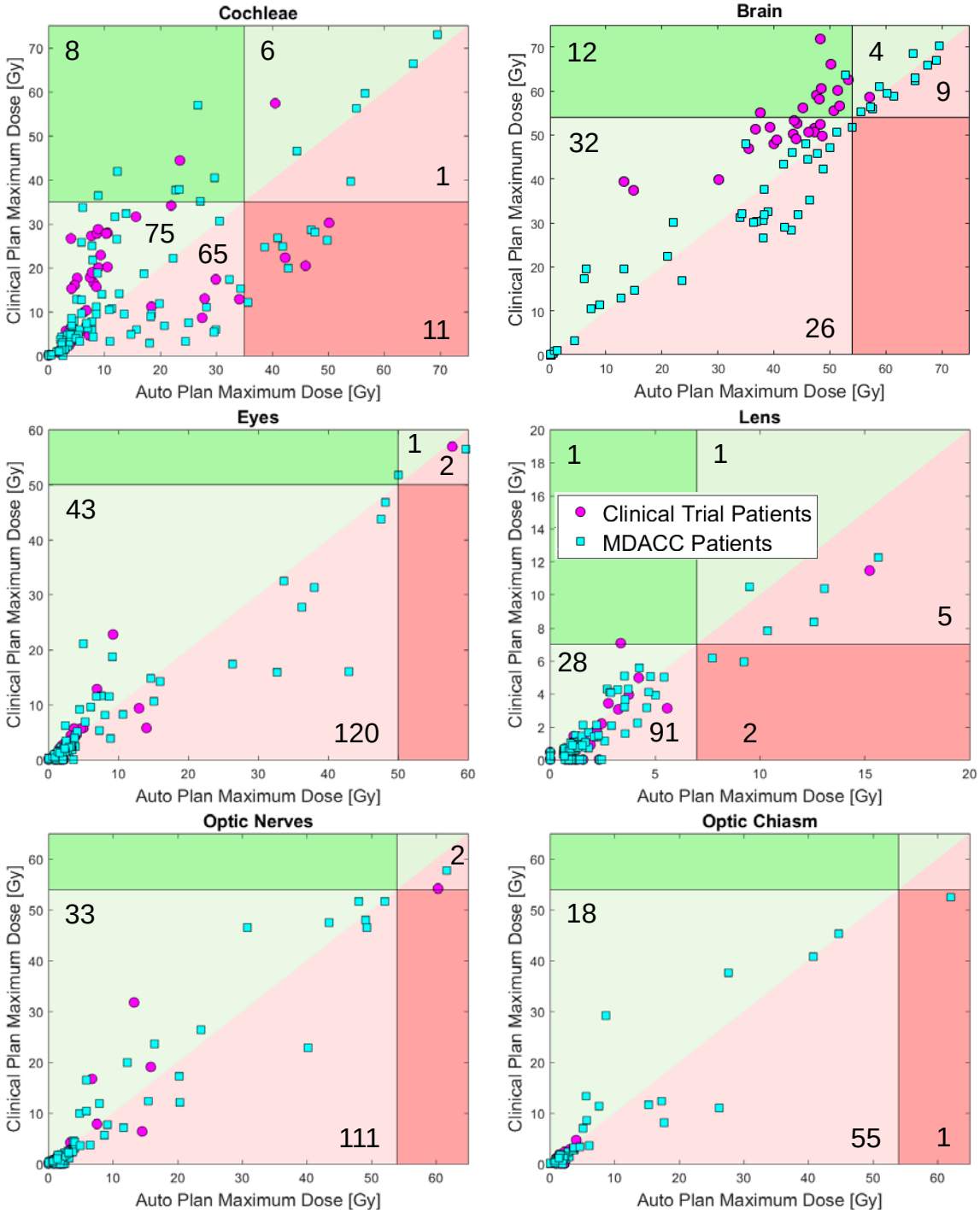


Figure 11, continued from previous page.

Finally, when considering DVHs curves of target and organs at risk, the autoplans show significant improvement over clinical plans treated on a clinical trial. Shown in Figure 12 are average DVH curves for the two planning techniques and the negative of the log of the p-value of the Wilcoxon signed rank test which is significant for values greater than or equal to 3. For all structures, the autoplans significantly outperformed the clinical plans in DVH analysis.

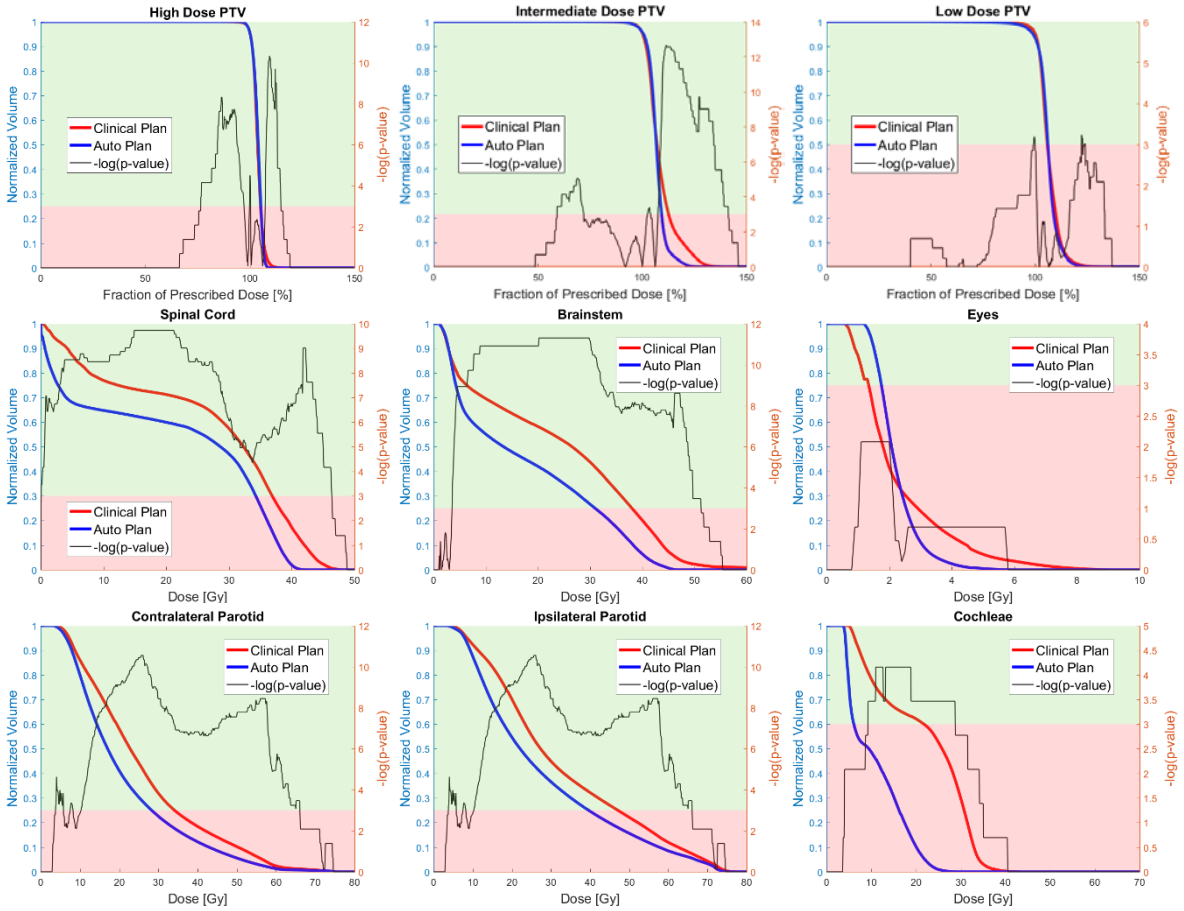


Figure 12. DVH comparison of autoplans and clinical plans treated as part of a clinical trial . The average DVH over all 20 patients is shown for both the clinical plan (red) and the autoplan (blue). Overlaid is negative of the log of the p value of the Wilcoxon signed sum test, significant in the green shaded region at values greater than or equal to 3. Autoplans routinely outperform clinical plans.

Physician Review of Autoplans

During the development of the planning approach, four head-and-neck site specific radiation oncologists review the autoplans and assessed them for clinical utility. When presented with 8 plans, 2 radiation oncologists from South Africa approved 100% of the autoplans. Upon presentation of the autoplans for 20 patients treated on a clinical trial XX were approved for use without and one required only minor edits. However, of 20 patients treated at MD Anderson Cancer Center only 7 (35%) were approved for use without edit and one required only minor edits. These patients included patients from subsites which differ in treatment strategy from those primarily used for strategy development including patients with disease of the oral cavity in which dose prescription levels (typically 60 ,57 and 54 Gy) are much closer than, for example, for oropharynx cases (70, 63 and 57Gy or 66, 60 and 54 gy). Further, the strategy was tested for patients with primary disease of the larynx for which, in scenarios of low-resource, advanced techniques are unlikely to be used. While for this cohort of patients, clinical dose constraints were met at rates similar to clinical plans as seen in *Table 5*, other features of the plans including dose heterogeneity and relative location of hotspots with underlying anatomy result in major modification needed for treatment planning. This suggests that, while the strategy is high performing for some subsites (particularly disease of the oropharynx), further investigation is likely needed for other subsites.

Discussion

In summary, we have developed an automated single-optimization approach for VMAT planning in the head-and-neck. Using a full automated approach based on in-house algorithms and commercial solutions, head-and-neck VMAT plans are created without human intervention. After manual contouring of targets and organs at risk, algorithms are used to set beam parameters including the jaw and collimator, identify the isocenter, and automatically contour planning structures. Finally, using a validated RapidPlan ® model in combination with

added population optimization objectives, VMAT plans were optimized and dose was calculated automatically. The entre process takes, on average, less than 40 minutes.

This autoplanning technique represents the first fully automated approach which requires no human intervention aside from the contouring of targets and organs at risk. DVH analysis of 29 patients treated on a clinical trial showed considerable improvement to the dose to organs as risk with limited impact of dose coverage and as rated by a head-and-neck specific radiation oncologist, XX% of plans were clinically acceptable without edit. A decrease in plan variability at clinically relevant dosimetric endpoints between plans was seen for four organs at risk when the autoplanning approach was used. A reduction in plan variability is often cited as desired for quality management purposes (82), though a reduction in variability should not come with a reduction in quality. The plans are generated in less than 40 minutes and this represents time during which the user does not need to intervene or be present and during which other tasks (i.e. contouring, plan approval, etc.) may occur. This time includes that needed for the autocontouring of structures not included in the treatment plan, which if supplied by the attending physician would not be required. Built on clinically implemented tools including the Eclipse Treatment Planning system (VMAT optimizer, dose calculation engine) and autocontouring algorithms the autoplanning technique can easily be expanded for use with other treatment machines and at other institutions. Evaluated on 20 patients from international partner institutions, the autoplanner performed well and radiation oncologists from these institutions found the plans clinically acceptable.

This autoplanning approach does, however, have limitations. First, when compared against clinically delivered plans from our institution, the clinical plans outperformed the autoplans for 10 of 26 DVH metrics analyzed, including having a lower maximum dose to the spinal cord, brainstem, optic chiasm, nerves, eyes and lens, better coverage at the 95% dose level of the low dose PTV and a lower hotspot to 1cc of the high dose PTV. The autoplans only outperformed for 6/26 including volume receiving 30Gy to the ipsilateral and contralateral

parotid glands, mean dose to the ipsilateral parotid gland, and maximum dose to the brain. It is known, and a fundamental characteristic of knowledge based planning, that the autoplans will represent the quality of the knowledge base on which the approach was designed. Also, as rated by a head-and-neck specific radiation oncologist, all rated plans for disease of the oral cavity and salivary glands needed major edit for clinical use, it is likely a subs-site specific optimization strategy is needed. A model built based wholly or partly on the sample of the 54 MD Anderson patients herein, may represent an improvement in plan quality over the current knowledge base and therefore allow for an improvement in the autoplans.

Similarly, while the current knowledge based schema does consider the spatial geometry of the patients' targets and normal tissues the use of a single model for the planning of many head-and-neck subsite may limit to performance of the autoplanner. Investigation in subsite specific knowledge based planning techniques for the head-and-neck may allow for incremental improvement in plan quality. Finally, the autoplanner performance was noticeably poor for patients which met certain criteria including very large tumor volumes, tumor volumes which include large volumes with very low density, and targets in very close proximity to critical normal structures. Many times, in these cases, compromises in plan quality must be identified by highly trained staff and therefore are not suitable to a fully automated approach. This could be implemented through a feedback system which analyzes spatial geometry of the plan prior to optimization.

In addition to general performance, the autoplanning schema used here has fundamental limitations. The limitation of 3 discrete target volume prescriptions which, while acceptable for the majority of patients, may exclude some patients who have disease very close to the brain or brachial plexus and generally require a fourth prescriptions volume to shape the dose around these critical structures. Additionally, as seen for one patient, the use of the current autocontouring technique requires consistent CT slice spacing, a requirement which does not apply to manual contouring or traditional treatment planning approaches.

As with any automated approach rule base objectives and tradeoffs must be implemented which may need revising over time. It is critical that characteristics of the incoming data are known and systems, which warn the user or may halt the process when a result is outside of standard values are implemented. For example, herein we implemented a normalization approach meant to ensure adequate coverage of all target volumes, which was realized by normalizing the final plan such that at least 98% of each target volume received 95% of the prescribed dose. However for some patients this resulted in very small normalization values and unacceptable plans. Evaluation of the characteristics of the target volumes used for normalization revealed the normalization to very small or very low density target volumes could drastically effect the plan quality. Towards this, pre-normalization rules were implemented to avoid these scenarios and over time new scenarios may arise which require the revaluation of the assumptions on which the algorithms were developed.

There are many opportunities for further study concerning this automated treatment planning approach. As mentioned, refinement of the knowledge based model may improve plan quality or may allow for tailoring to each individual site for implementation. Further, an iterative approach which may involve re-optimization with added constraints on hot or cold spots has seen success (83) and could be implemented without requiring human intervention. Additionally, though the current use of 3-4 arcs is not desired as the use of additional arcs increases the time needed for delivery, dose calculation, and plan quality assurance. Optimization of the planning approach including jaw and collimator selection algorithms may allow for the use of two arcs. Further refinement of the normalization algorithm to accommodate additional head-and-neck subsites, and not yet encountered circumstances, and in order to satisfy individual treating physicians may be required. The time needed for plan development may also be reduced by increasing processing power or using improved dose calculation algorithms including Acuros® (84) which is available in the Eclipse TPS but requires additional investigation including how to handle dental artifacts. Other considerations

for planning/overrise structures such as dental artifacts or bridges may be required and result in an improvement plan quality. Finally, and most importantly, the clinical use of the autoplanning approach will reveal how the system will be integrated into a radiation oncology clinic, will likely reveal the need for further refinement, and may prove to be an essential tool to help reduce the human infrastructure burden of radiotherapy in low-resource settings.

Conclusion

In conclusion, a single optimization treatment planner along with automatic field settings can generate clinically acceptable autoplans in under 40 minutes without the need for human intervention.

Chapter 5: Automatic Contours of Normal Structures in the Head-and-Neck

A substantial portion of this chapter is written or based on the following publications:

Placeholder – Paper under 2nd round revision.

In this chapter we describe the results for Specific Aim 3 in which, through four experiments, we evaluate the feasibility of automatic delineation of normal structures for treatment planning in the head-and-neck. Our working hypothesis is automatically contoured normal structures can be used for treatment planning purposes without significant impact on plan quality.

Introduction

Advanced techniques including 3D conformal radiotherapy, IMRT, and VMAT which have been widespread in high-resource clinics since the 1990s, have made the process of contouring an essential step in the treatment planning process. Contouring, however, is known to be time consuming, plagued by considerable inter-physician variability (85-89), and the component of radiation therapy treatment planning which introduces the most error (90, 91). For head-and-neck cancer, the contouring of 3 or more target volumes and as many as 25 normal structures may be required (92). For any clinic, but particularly for those in low-resource clinics the enormous time, human resource, and training burden of contouring may prohibit the transition to these advanced techniques, which are particularly critical for the treatment of head-and-neck cancer.

In the context of an automated treatment planning approach for head-and-neck cancer, automatic contouring would increase the time savings, reduce plan variability, and may eliminate some of the barriers prohibiting the transition to advanced delivery techniques. Many efforts to automate the contouring process for normal structures have been reported – techniques include image registration atlas based segmentation, machine learning, and shape modelling (57). Findings indicate that for head-and-neck treatment plans, automatic segmentation of normal structures can significantly reduce inter-observer variability and contouring time (93-95). Authors have reported on the limited implementation of automated contouring for small structure sets (e.g., brachial plexus(96), heart chambers(97),etc.) and for other anatomical sites (e.g. the prostate(98)). However, use of autocontouring for only a subset of structures requires a deviation in workflow, if contours require manual editing the time saving advantages are partially lost, and importantly, and the long term clinical use of autocontoured normal tissues has not been reported.

Further, reports on the clinical implementation of automatic contouring methods note that automatic contours should be carefully reviewed and edited by the physician (99), which may take up to 60 minutes (100). The use of automatically contoured normal structures for treatment planning without edit would offer increased time savings and would be interest not only for a fully automated treatment planning approach but also for adaptive planning in which treatment plans are made under tight time constraints and for general clinical practice. However, using unedited automatically contoured normal structures should be considered with great caution and after comprehensive analysis.

In this chapter we describe four experiments which assess the viability of using unedited normal structures in the head-and-neck for treatment planning. First, several autocontouring algorithms were retrospectively assessed for accuracy compared to physician drawn structures and were qualitatively rated by head-and-neck radiation oncologists. Then, the best performing algorithm was implemented into our head-and-neck clinic with careful

physician review and editing for clinical use. Next, a technique to ensure safe use in an automated process was developed in order to detect both gross errors in autocontours and autocontours which required physician edit for clinical use. Finally, treatment plans were created on autocontours without edit and DVH analysis revealed the impact planning on these structures, as opposed to physician derived structure, had on the treatment plan.

Methods

Analysis of Autocontouring Algorithms

In this first experiment we sought to evaluate automatic contouring algorithms for normal structures in the head-and-neck.

Patient Cohort

For this study, we collected 128, the latest in our database, treated for head-and-neck cancer at MD Anderson Cancer Center, and who had physician approved contours. Normal structure contours were identified manually by naming convention and visual assessment. From the 128 patients, the 8 contours most often included in the final treatment plan, and thus included in this analysis, were the brain, brainstem, cochleae, eyes, lungs, mandible, parotid glands, and spinal cord. Limited analysis of the autocontouring of others structures, including lens, optic chiasm, optic nerves, and submandibular glands was also performed.

A subset of 10 patients was randomly selected for initial review of four autocontouring techniques. For the highest performing algorithm on initial review the 118 remaining patients were selected to further analyze the autocontouring algorithm's performance.

Autocontouring Algorithms

Of the four autocontouring algorithms used for this study, three are based on two commercially available autocontouring algorithms available in the Eclipse Treatment Planning system; Smart Detection® and Smart Segmentation®(101). Smart Detection® is a heuristic

solution available for a select number of structures and requires only the selection of desired contours. Available structures included brain, eyes, lungs, mandible, and spinal canal. The second algorithm, Smart Segmentation®, is a deformable image registration based technique which, in Eclipse TPS version 13.2, requires the user to select one of many available expert cases from which the autocontours will be propagated. According to the manufacturer:

> "Users match their patient case to one of the hundreds of expert cases in the library, and through a sequence of co-registration and proprietary deformation algorithms, the contours of an expert case are deformed to fit the CT images of the patients"(101)

The Smart Segmentation software offers guidance on which expert case should be selected through a 5 star rating system. Expert cases for each structure for each patient were selected in order of star rating, with up to 14 expert cases selected for each structure for each patient, though not all expert cases were available for all 8 normal structures considered. Later versions of the Eclipse TPS software allows the selection of multiple atlas and fuses the contours using a majority voting algorithm. To simulate this and based on physician feedback, up to 12 individual contours were used as inputs to an in-house majority voting algorithm outside of the treatment planning system, resulting in a single fused contour per structure. The fused contours represent the third autocontouring approach.

The fourth autocontouring method was based on an in-house multi atlas deformable image registration technique termed multi-atlas contouring service (MACS) (96, 97), was previously developed (102, 103) and consists of three distinct steps. First, rigid registration is performed between the test patient's simulation CT and the CT of each of twelve atlas patients, using 2D sagittal and coronal projections. Second, the test patient is deformably registered to each of the atlas patients using dual-force Demons deformable registration (104). Using the resultant deformation vector fields, the contours from each atlas patient are mapped to the test patient(105); resulting in a number of individual contours equal to the number of atlas patients. Finally, the STAPLE algorithm with a built-in tissue appearance

model(106) is used to combine the individual segmentations, generating a fusion contour approximating a true segmentation. Central to the MACS algorithm is the building of an atlas of patient CTs, which is representative of the patients for which the algorithm will be used. To build the atlas used in this work, 12 patients recently treated at our institution for head-and-neck cancer were selected. The contours used for atlas building were either extracted from patient treatment plans (reviewed prior to treatment by head-and-neck quality assurance peer review clinic (107)) or created using thresholding tools (brain, lung, and mandible). All contours were carefully reviewed by a medical dosimetrist and head-and-neck radiation oncologist with 13 and 8 years of experience, respectively. The autocontouring process of multi-atlas DIR based methods can be seen in Figure 13.

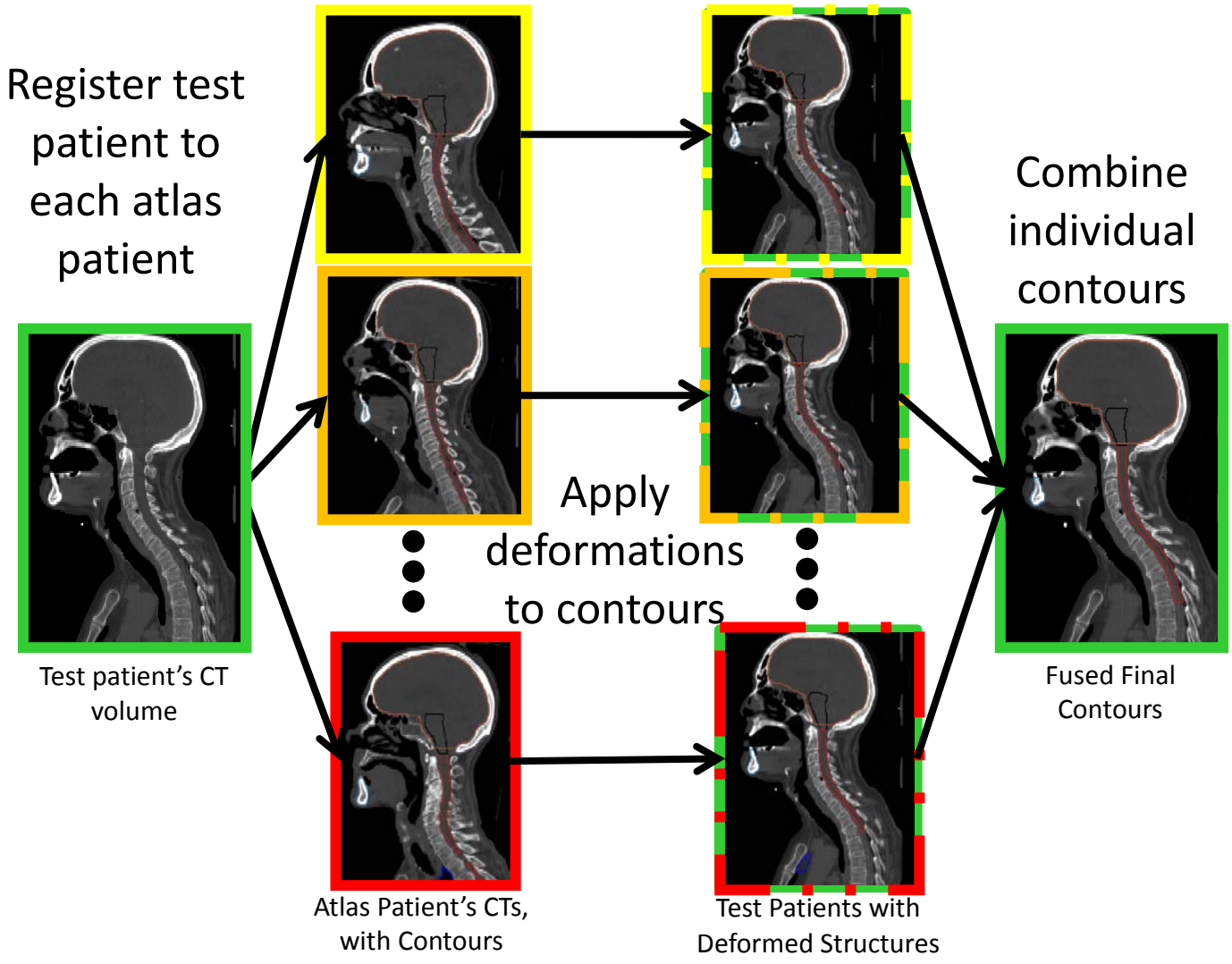

Register test patient to each atlas patient
Apply deformations to contours
Combine individual contours
Test patient's CT volume
Atlas Patient's CTs, with Contours
Test Patients with Deformed Structures
Fused Final Contours

Figure 13. Schematic of the multi-atlas deformable image registration autocontouring process. For each new patient, a deformable image registration algorithm is used to link the test patient to each of the chosen atlas patients, which have validated contours already drawn. The deformation vector field is used to map the contours from each of the atlas patients to the test patients, which results in a number of contours for each structure equal to the number of atlas patients. A fusion algorithm is used to combine the contours into a single contour. The in-house contouring algorithm uses 12 atlas patients, a deamons dual force DIR algorithms and STAPLE as the fusion algorithm. For the fused Smart Segmentation contours a proprietary DIR and contours from up to 12 patients were used with a majority voting algorithm. A third algorithm, discussed in the next sections, used a Deeds deformable image registration algorithm, 10 atlas patients, and a majority voting algorithm.

Quantitative Contour Assessment

Contours of the 8 normal structures from each of the 4 autocontouring techniques were quantitatively compared to independently physician drawn contours, when available. The Dice similarity coefficient, mean surface distance, and Hausdorff distance were measured to assess contour accuracy. The Dice similarity coefficient (DSC) measures the volume overlap of the physician drawn contour, P, with the autocontour, A, as a ratio to their total volume, with a minimum value of 0 when the contours have no overlap and a maximum value of 1 when the contours agree perfectly, as in Equation 1.

Equation 1 – Dice Similarity Coefficient

$$DSC(P,A) = 2\frac{|P \cap A|}{|P| + |A|}.$$

The mean surface distance was calculated as a symmetric 3D mean surface distance between two volumes (P and A) and has a minimum value of 0 when the contours agree completely and no maximum value, as in Equation 2.

Equation 2 – Mean Surface Distance

$$MSD(P,A) = \frac{1}{2}\left[\frac{1}{|P|}\sum_{p\in P}\min(d(p,a)) + \frac{1}{|A|}\sum_{a\in A}\min(d(a,p))\right].$$

The Hausdorff distance (HD) was used to measure the maximum Euclidean distance from the points both contours to the nearest point in the other and has a minimum value of 0 when the contours agree completely and no maximum value, as in Equation 3.

Equation 3 – Hausdorff Distance

$$HD(P,A) = max\left\{\max_{p\in P}\min_{a\in A} d(p,a), \max_{a\in A}\min_{p\in P} d(a,p)\right\}.$$

For two of the eight normal structures (lungs and spinal cord) we performed the quantitative analysis using a modified structure which included only slices of the contours of the autocontour between the most superior and most inferior contours drawn by the physician. This analysis better represents the contouring accuracy of the algorithm (compared to whole-structure quantitative analysis) as it eliminates errors that arise owing to differences in CT scan extent both between the test and atlas patients and among the atlas patients, and takes into consideration incomplete contouring of structures at distance far from the treated volumes, see Figure 14.

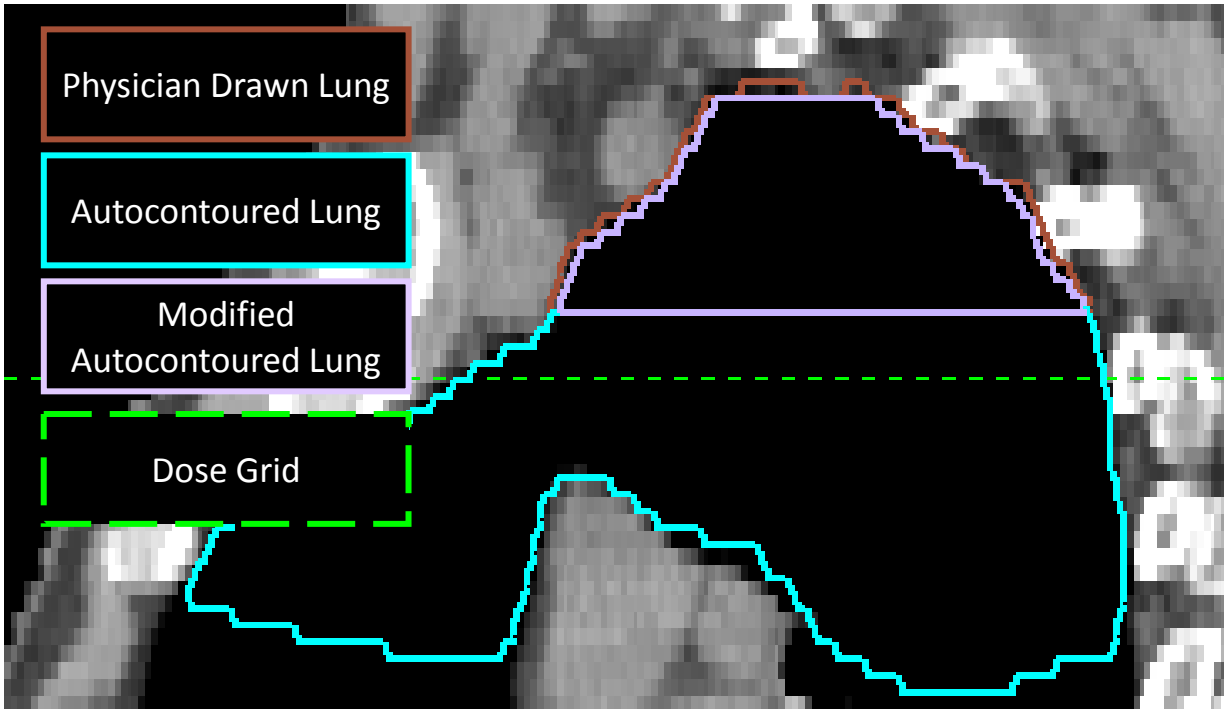


Figure 14 - Modified structures considered for analysis of autocontour performance. For spinal cord and lung (shown), a modified autocontour was created for analysis. This structure only includes slices of the original autocontour between the most superior and most inferior slices of the corresponding physician drawn structures. In this example, the physician drawn lung structure (brown) was only contoured for part of the true lung. The original autocontoured lung (cyan) extended for many slices below the physician lung, when modified (lavender) the slices extended below the most inferior slices of the physician drawn lung were removed. This is appropriate for many reasons including, as shown in this example, often the autocontour will extend further even than the dose grid (green dash) and is therefore irrelevant in the treatment planning process.

Physician Review of Autocontours

A radiation oncologist was asked to rate the normal structures on a five-point scale, Table 8. For contours receiving a score indicating minor or major edit would be needed for use in dose volume histogram (DVH)-based planning (scores 1-3), the physician was asked

to note the main failure mode. This allows for analysis of potential improvement of the algorithm.

Table 8. Five point scale for physician rating of normal structures. For structures receiving a score of 1, 2, or 3, which indicates either minor or major edit is needed for use in DVH based treatment planning, the physician was asked to cite the main failure mode.

Rating	Description
5	Perfect; indistinguishable from physician-drawn contours for DVH based planning purposes
4	Within acceptable inter-physician variation for planning purposes, as described above
3	Good; needs minor edits if normal structure is near a target
2	Fair; needs significant edits to be used for the planning purposes described above
1	Poor; large areas need minor or major edit. Is unusable for planning purposes described above

For the top performing algorithm the remaining 118 patients were rated on the same scale and, to assess the possibility of rater bias, contours from this top performing algorithm for 10 patients were reviewed by five additional radiation oncologists from four international institutions. Physician agreement was assessed by grouping each pair of ratings (one rating from the primary physician and one from an outside physician) into one of three categories. Category I agreement includes instances when the primary and outside physician agreed as to the degree of edit needed, Category II agreement indicates that the physicians agreed that the contour required either no more than minor edits or required major edits. The final category, Category III agreement, includes those contours where the physicians disagreed on the acceptability of the contour, with one physician indicating that the contour needed major edit with the other indicating no or minor edit for use. Additionally, inter-physician variability in ratings was assessed using a Wilcoxon signed rank test.

Additional Contour Analysis

Exploratory analysis of the autocontouring of additional structures including optic chiasm, optic nerves, submandibular glands, esophagus, and lens using the previously described algorithms and others including a commercially licensed algorithm (108) using (1) a proprietary atlas (2) the same atlas used in MACS and (3) the proprietary atlas in combination with a shape model supplied by the vendor (only available for some structures) were also performed results of which can be found in the appendix.

The results of this first experiment are found on page 100.

Clinical Use of Normal Tissue Autocontours

In the second experiment we investigated the clinical use of an autocontouring algorithm for normal structures in the head-and-neck.

Clinical Implementation

After retrospective validation of the contouring algorithms, we began a limited introduction of the top performing algorithm, MACS, into our head-and-neck clinic. The automated contouring algorithm was already in use for a limited number of sites and structures in our clinic prior to the implementation of this atlas for normal structures in the head-and-neck. The software based on the MACS algorithm is accessible via a script in the Pinnacle^3 Treatment Planning System (*Philips Medical Systems*, Milpitas, *CA*). Initially, use of the software was limited to a select number of radiation oncologists for review, which prompted standardization of the contour color to match that already in clinical practice, and inclusion of a script to overwrite empty structures already in the treatment plan (our clinical workflow involves the population of a standard set of empty structures through a script). Additionally, due to algorithm limitation as discussed, the inferior portion of the lung was often either not contoured or was contoured in many segments per slice, which prompted renaming of the lung structure to “lung_avoid,” indicating its use for treatment planning purposes rather

than anatomical definition. After the automated contouring workflow was finalized, the software was released for use by all attending physicians. Typically, the clinical workflow of automated contouring involves the initialization of the algorithm by a dosimetrist using a script in the treatment planning system, import of the structures into the treatment plan, followed by review and any needed editing of the contours by the attending radiation oncologist. All final contours were reviewed and edited by the attending physician, see Figure 15, in the same way that initial resident contours would be reviewed. The final contours reflected approval by the physician, with or without editing as deemed appropriate.

Quantitative Analysis of Autocontour Edit

To assess the degree of edits made of the autocontours for clinical use, the contours generated by the algorithm were compared to the contours edited for treatment planning by the physician using the DSC (Equation 1), MSD (Equation 2), and HD (Equation 3). In addition to analysis of the eight clinically implemented normal structures, we also quantitatively compared edits of two modified contours, as previously described.

Additionally, in order to elucidate possible planning margins needed if autocontours are used without edits, we determined the minimum uniform expansion to the autocontour needed to cover 95% or 100% of the edited contour for 90% or 95% of the population. These margins may reveal that for some structures a planning at risk volume (PRV) may be added to the autocontoured normal tissue in order to compensate for potential errors in contouring. This expansion may ensure the safety of using unedited structures in treatment planning. This may be especially true for structures with a maximum dose constraint for which, if the modified structure meets constraint, any structure completely encompassed would also meet constraints.

Finally, to assess the possible motivation for contour edits outside of anatomical OAR definition, we used a one-sided *t* test to assess the association between automatic contour edits and the minimum distance to target volume.

The results of this second experiment are found on page 107.

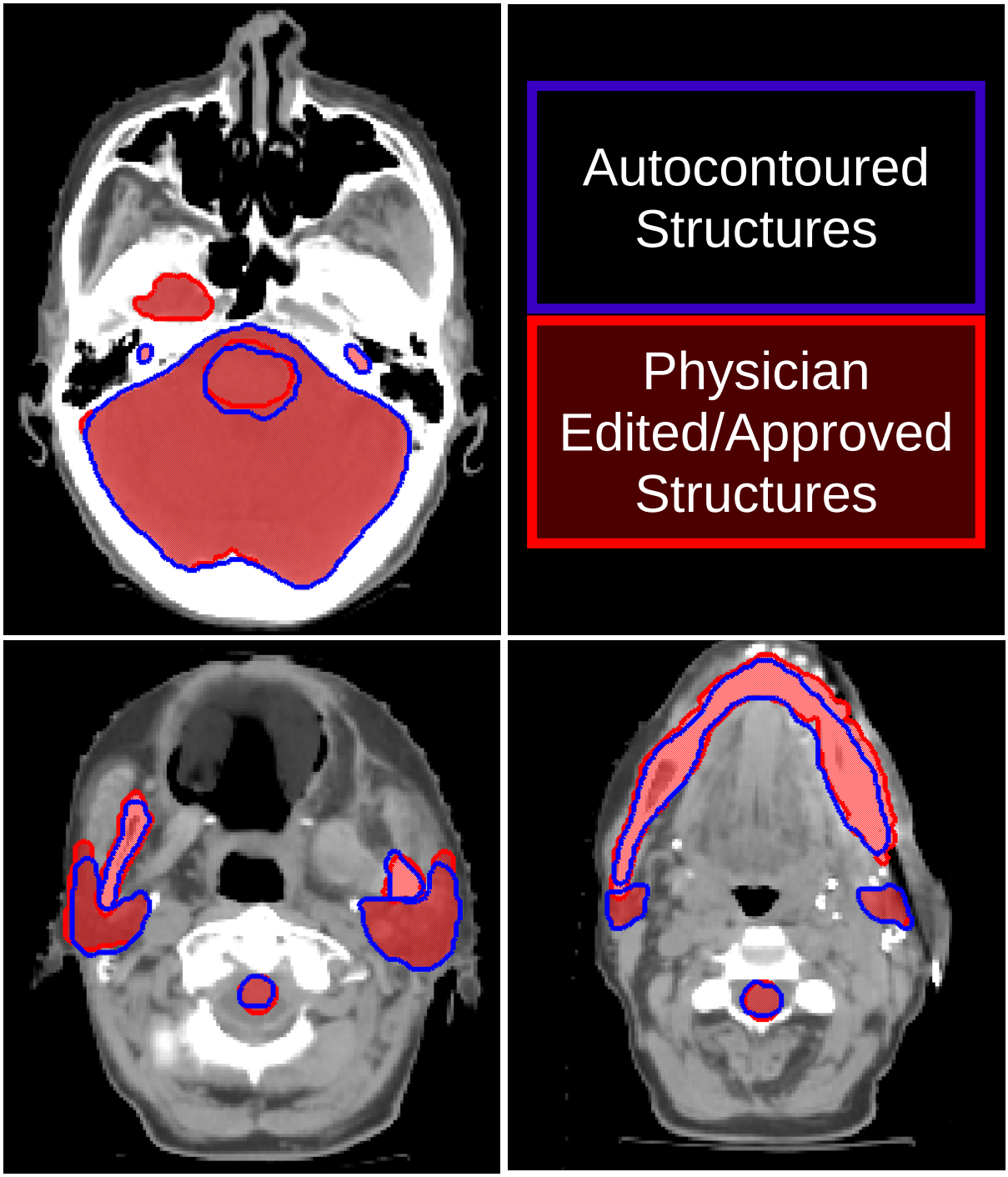


Figure 15 - Example of physician edits to clinical autocontours. Shown for one patient are autocontours of the brain, brainstem and cochleae (panel A) and spinal cord mandible and parotid glands (panels B and C) in blue and the physician edits and/or approved structures which remained in the patient's treatment plan in red. For this patient, no edits were made to the cochlea structures, and some boundaries of other structures remained unedited.

Machine Learning Model for Prediction of Autocontour Errors

In the third experiment we sought to develop of a random forest model for assessment of anatomical errors in autocontours of normal structures in the head-and-neck

Contour QA technique

Necessary for implementation as part of a fully automated treatment planning approach is a method to detect autocontouring errors. Detection of gross and simulated errors represents a check of safety and catching of significant autocontouring errors. The detection of smaller, potentially necessary clinical edits to autocontours would allow for the possibility to flag to the user of an automated treatment planning approach that autocontours presented may require editing. Other investigations of contour QA have been reported using historical data with heuristically selected metrics and thresholds to predict simulated contouring errors (109), mislabeled contours, and the effects of noise (110). These studies introduce the important topic of contour QA and demonstrate the feasibility of automated techniques in detecting some errors. However a comprehensive or optimized prediction model was not used to differentiate cases of automatic contouring failure and manual detection of errors is still heavily relied upon.

To predict contour errors, we implemented a random forest (RF) method in Matlab (Mathworks, Natick Massachusetts). RFs are machine-learning models proposed by Breiman (111) that do not require a priori information about any relationship between input metrics and output predictions (unlike, for example, regression models, which typically assume linearity) and provide a measure of predictor importance, and have low susceptibility to overfitting. Random forest models have proven successful in other contour QA approaches (49, 110) as well as non-linear approaches such as radiation toxicity (112-115).

Three RF models were trained for each of the eight normal structures, one model for each of the three classes of automatic contouring errors (simulated errors, and two degrees of

true errors), as described below. Each model was then tested on an independent dataset. The RF models were trained for 1,000 decision trees with a minimum leaf node size of three. The number of patients and error types for each training and testing dataset are detailed in the following sections.

Engineered predictive metrics, developed to assess the contour accuracy, were extracted from each contour and corresponding CT dataset. A summary of the metrics can be found in Table 9. The predictive metrics included volumetric (e.g., Dice similarity coefficient) and distance agreement metrics (e.g., mean surface distance) with contours generated from three independent contouring techniques. For the first technique, a commercially licensed algorithm (108) (Varian Medical Systems, Palo Alto, CA) was implemented in a multi-atlas deformable image registration contouring approach, whereby contours from 11 atlas patients were deformably registered to the test patient and the resultant contours were fused using a majority voting algorithm, a schematic is shown in Figure 13. For the second technique, contours were generated using the heuristic algorithm Smart Detection®, as implemented in the Eclipse treatment planning software (Varian Medical Systems). Owing to algorithm limitations, only the brain, eyes, lungs, and mandible contours were generated and thus compared with the corresponding contour from the test patient. For the third technique, 11 atlas patients were rigidly registered to the test patient and the most accurate registration (as measured by the mutual information between atlas and test patient data set over the whole image) was used for contour propagation.

Other predictive metrics included the volume of the contour, metrics extracted from the Hounsfield unit (HU) distribution within the contour, and the spatial relationship of the contour with other automatically contoured structures (i.e., other normal structures). Additionally, we analyzed the highest order coefficient of first-, second-, and third-order polynomial fits to inter-slice distance measures, both between adjacent slices and between bony tissues and the inter-slice area Figure 16. Because several of the available metrics were derived from the

same primary data (e.g., metrics from the HU distribution, agreement with secondary techniques, and polynomial fits to slice measures), a univariate selection was performed by choosing the metric within the group with the lowest average misclassification probability in a small (10-tree) forest using only the predictor in question. The full list of predictive metrics can be found in Table 9.

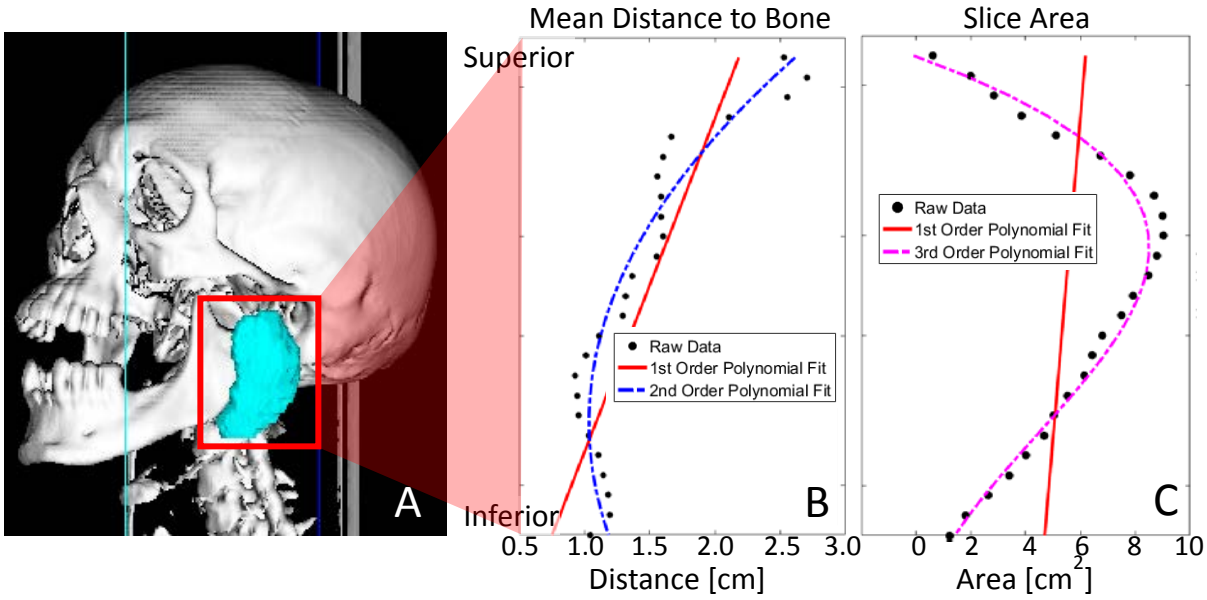


Figure 16. Inter-slice metrics for autocontour error detection. Contour features for a representative patient are shown in panel A and include polynomial fits to inter-slice metrics such as the mean distance to bone (in panel B) and the slice area (panel C).

Table 9. Engineered metrics used for the prediction of errors to automatic contours.

Shape metrics		Size/Hounsfield unit (HU)-derived metrics		Positional metrics		Metrics showing agreement with independent automatic contours	
Metric/metric group	**Definition**	**Metric/metric group**	**Definition**	**Metric/metric group**	**Definition**	**Metric/metric group**	**Definition**
Inter-slice slice area statistics	Mean, maximum, minimum, and standard deviation	Volume	(no. of voxels in contour) × [voxel size (cm^3)]	Inter-slice minimum distance to bone statistics	Mean, maximum, minimum, and standard deviation	Independent deformable atlas approach, volumetric agreement*	Volume ratio, true positive ratio, false positive ratio
Inter-slice slice area polynomial fits*	Highest order coefficient of first-, second-, and third-degree polynomial fits	HU statistics	Mean, standard deviation of HU of pixels within contour	Inter-slice minimum distance polynomial fits*	Highest order coefficient of first-, second-, and third-degree polynomial fits	Independent deformable atlas approach, distance agreement*	Hausdorff distance, mean surface distance
Inter-slice 3D Hausdorff distance	Mean, maximum, minimum, and standard deviation	Lower extreme of HU distribution within contour*	HU at 1%, 2%, 5%, and 10% cumulative probability	Inter-slice maximum distance to bone statistics	Mean, maximum, minimum, and standard deviation	Independent heuristic approach, volumetric agreement*	Volume ratio, true positive ratio, false positive ratio
Inter-slice 3D Hausdorff distance*	Highest order coefficient of first-, second, and third-degree polynomial fits	Upper extreme of HU distribution within contour*	HU at 90%, 95%, 98%, and 99% cumulative probability	Inter-slice maximum distance polynomial fits*	Highest order coefficient of first-, second-, and third degree polynomial fits	Independent heuristic approach, distance agreement*	Hausdorff distance, mean surface distance
Inter-slice 3D mean surface distance	Mean, maximum, minimum, and standard deviation	% of contour with HU below given HU values	-500, -100, 0, 40, 300, 500	Separation of contour centroids in X, Y, and Z directions	Seven other automatic contours	Independent single-patient rigid approach, volumetric agreement*	Volume ratio, true positive ratio, false positive ratio
Inter-slice 3D mean surface distance*	Highest order coefficient of first-, second-, and third-degree polynomial fits					Independent single-patient rigid approach, distance agreement*	Hausdorff distance, mean surface distance

*Univariate selection was performed within the metric group.

Simulated failure models

In the RF models for simulated errors, prediction classes included unedited clinical automatic contours and simulated failures, including automatic contours that had been shifted or expanded by amounts that varied by structure. Contours were expanded uniformly in three dimensions or shifted in a random direction by a known amount. The size of the shifts and expansions varied by structure according to the average volume of the structure. Contours for small structures (eyes and cochleae) were shifted and expanded to a lesser extent than those for medium-sized structures (brainstem, mandible, and parotid glands), which were shifted and expanded less than that for large structures (brain, lungs, and spinal cord). For RF training, eight shifted structures from each of four size shifts and 10 expanded contours from each of three size expansions were combined with 62 unedited clinical contours, providing a total of 124 contours. See Table 10 for details of the training data-set, including the sizes of shifts and expansions. Predictor importance was assessed by summing the change in the mean squared error due to splits on that predictor divided by the number of branch nodes after the split.

For testing of the simulated error RFs, 50 contours from each of the aforementioned error types (shifts and expansions) were combined with contours propagated on 13 patient scans of other anatomical locations (seven patients with cervical cancer and six patients with breast cancer), and contours from nine patients in nonstandard positions for which the automatic contouring algorithm was grossly incorrect, as well as contours from one patient who had a very large primary tumor volume that impinged on nearby normal structures. Figure 17 shows representative examples of contours with simulated errors and for a patient not in a similar position to the atlas patients.

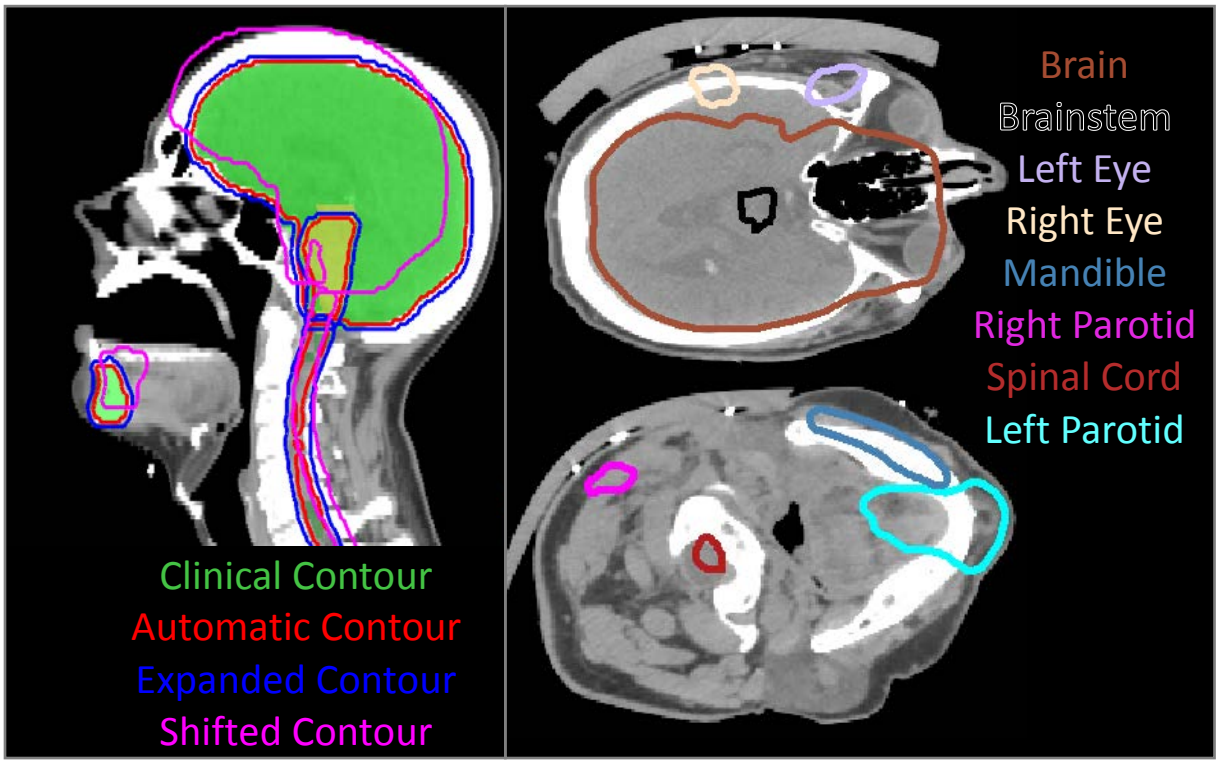


Figure 17. Contouring errors for building and testing of QA models. Simulated errors included autocontours which had been shifted and expanded (left) and autocontours propagated on patients for which the atlas based technique is not suitable (right). The atlas used for the current study included patients positioned supine on the treatment table; when autocontours are propagated on patients in significantly different positons automatic contouring errors arise.

Table 10. Number of type of simulated failure modes used for building of the random forest models. In the prediction of simulated errors, four sizes of shifted contours and three sizes of expanded contours were used for each structures. The sizes of the shifts and expansions varied by structure in accordance with the structure size, classified into three groups. Contours for smaller structures (eyes, cochleae) were shifted and expanded less than those for medium-sized structures (brainstem, mandible, parotid gland), which were shifted and expanded less than those for large structures (brain, lung, spinal cord).

	Shifted, size 1		**Shifted, size 2**		**Shifted, size 3**		**Shifted, size 4**		**Expanded, size 1**		**Expanded, size 2**		**Expanded, size 3**	
Structure	**Size, cm**	**No.**	**Size, cm**	**No.**	**Size, cm**	**No.**	**Size, cm**	**No.**	**Size, cm**	**No.**	**Size, cm**	**No.**	**Size, cm**	**No.**
Brain	1	8	2	8	3	8	5	8	3	10	5	10	8	10
Brainstem	0.5	8	1	8	2	8	3	8	2	10	3	10	5	10
Cochleae	0.2	8	0.3	8	0.5	8	1	8	1	10	2	10	3	10
Eyes	0.2	8	0.3	8	0.5	8	1	8	1	10	2	10	3	10
Lungs	1	8	2	8	3	8	5	8	3	10	5	10	8	10
Mandible	0.5	8	1	8	2	8	3	8	2	10	3	10	5	10
Parotid glands	0.5	8	1	8	2	8	3	8	2	10	3	10	5	10
Spinal cord	1	8	2	8	3	8	5	8	3	10	5	10	8	10

Prediction of clinical edits

For the prediction of true errors, a group of RF models (two for each of the eight OARs) were trained and tested on autocontours that were reviewed, and, if necessary, edited for clinical use. RF models were trained to differentiate contours that had been edited clinically from those that had not been edited. Two thresholds were chosen to determine whether the contour was edited: first, a Hausdorff distance (see Equation 3) between the edited contour and the contour in the patient's final treatment plan greater than 0 (i.e., any clinical edit) and second, a Hausdorff distance between the edited contour and the contour in the patient's treatment plan greater than 5 mm, classified as a "significant" clinical edit.

To obtain the initial RF model, we used data collected from the first 104 patients during clinical implementation. Following the initial modeling, and as the automatic contouring software was used clinically, the results were collected and a prediction was obtained for each structure. After the prediction for 10 structures from at least five patients (because bilateral structures, if all are left in the treatment plan, accrue 10 structures from five patients), predictive metrics were calculated and the RF models were retrained to include the new clinical results. This iterative machine learning mechanism was implemented to determine whether model performance improved over time as the models learned from new failure modes; the process of model building and rebuilding is illustrated in *Figure 18*. The final analysis contained RF models tested on 120 independent patients. The number of autocontours used for training and testing varied by structure because predictions were generated only for structures that remained in the patients' treatment plans upon treatment. Predictor importance was assessed by summing the change in the mean square error due to splits on that predictor divided by the number of branch nodes after the split.

The results of this third experiment are found on page 112.

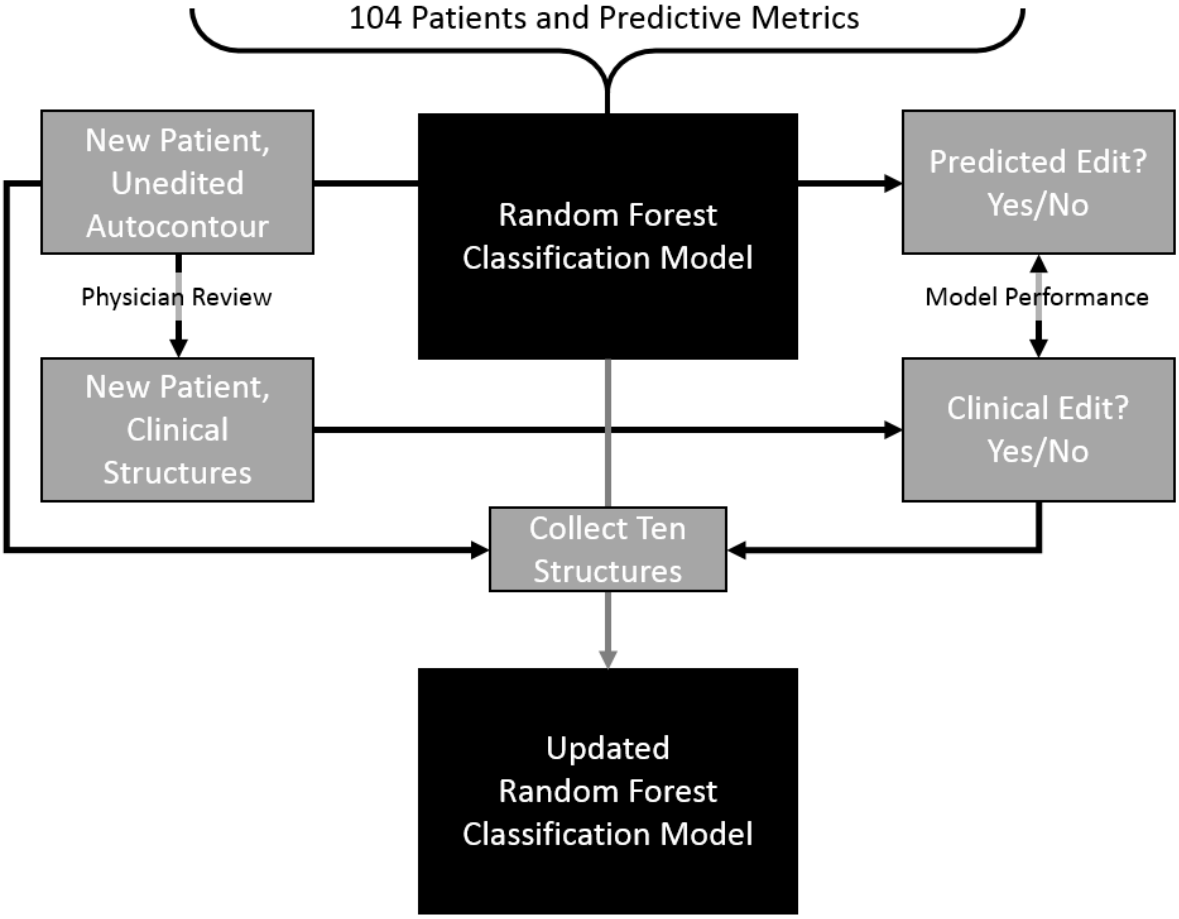


Figure 18. Workflow of the machine learning framework for the prediction of clinical edits to automatic contours (i.e., true errors). After the initial model building (104 patients), the model was used to predict clinical edits, and after prediction on 10 successive structures, the new data were added and the models were updated. Data from 120 patients were used for this prospective model testing and model updating.

Dosimetric Impact of Normal Tissue Autocontours for Treatment Planning

In an ideal scenario, the autocontouring of normal structures would produce contours which do not need edit for treatment planning, as they would be anatomically correct. While we do not expect that we have achieved this ideal, we sought to investigate the possibility of using unedited autocontours from the MACS autocontouring algorithm for treatment planning and to assess the effect this may have on the treatment plan. Other investigations have shown that for salivary glands the use of autocontoured or even simplified geometric structures may be sufficient for treatment planning purposes(116) suggesting that the same may be true for other normal structures.

Patient Cohort

For this analysis, two cohorts of patients were considered. Patients were selected as to represent seven head-and-neck subsites; larynx, nasopharynx, oral cavity, oropharynx, paranasal sinuses and cavity, and salivary glands. Selection criteria included a maximum of three physician-drawn planning target volume dose levels and availability of the physician approved contours.

In the first cohort, 54 patients treated at The University of Texas MD Anderson Cancer Center had clinical normal structures which were derived from autocontours and either edited to satisfaction or used with approval by the attending physician. The autocontouring of eight normal structures (brain, brainstem, cochlea, eyes, lung, mandible, parotid glands, and spinal cord) was implemented clinically and thus were considered in this analysis.

In the second cohort, normal tissue contours were drawn independently by the attending physician who was blind to the autocontour. In this group, structures not included in the clinical autocontouring atlas for 54 patients treated at The University of Texas MD Anderson Cancer Center and for all structures from 30 patients from other institutions treated on a clinical trial were collected. The clinical trial data were retrospectively collected through

The Cancer Imaging Archive (75) and was originally derived from the Radiation Therapy Oncology Group 0522 study (76).

Using a two sided t-test the distribution of differences of clinical structures compared to autocontours was compared to the larger group of clinically edited patients, as presented in the second aim (Page 43) to determine how they compare to a population of patients with clinically edited contours.

Dosimetric Evaluation

Using the planning strategy presented in Chapter 4 - A Single Optimization Treatment Planning Strategy in the Head-and-Neck – two treatment plans were created for each patient. The first treatment plan was developed using the clinical normal structures and the second plan was created using unedited autocontours of normal structures (including the brain, brainstem, cochleae, eyes, lens, lungs, mandible, optic chiasm, optic nerves, parotid glands, submandibular glands, spinal cord). For all patients, physician drawn PTVs were used for treatment planning. Additional autocontoured planning structures were created as needed per the treatment planning strategy. Treatment plans were normalized as in Chapter 4. A schematic describing this investigation can be seen in Figure 19.

For both plans, dosimetric analysis at clinically relevant DVH points was performed using the clinical normal structures, regardless of which structure (clinical structure or unedited autocontour) was used for treatment planning. Comparisons were made only for structures included in the clinical treatment plan. A Wilcoxon signed rank test was used to compare the dosimetric impact at clinically relevant dosimetric endpoints.

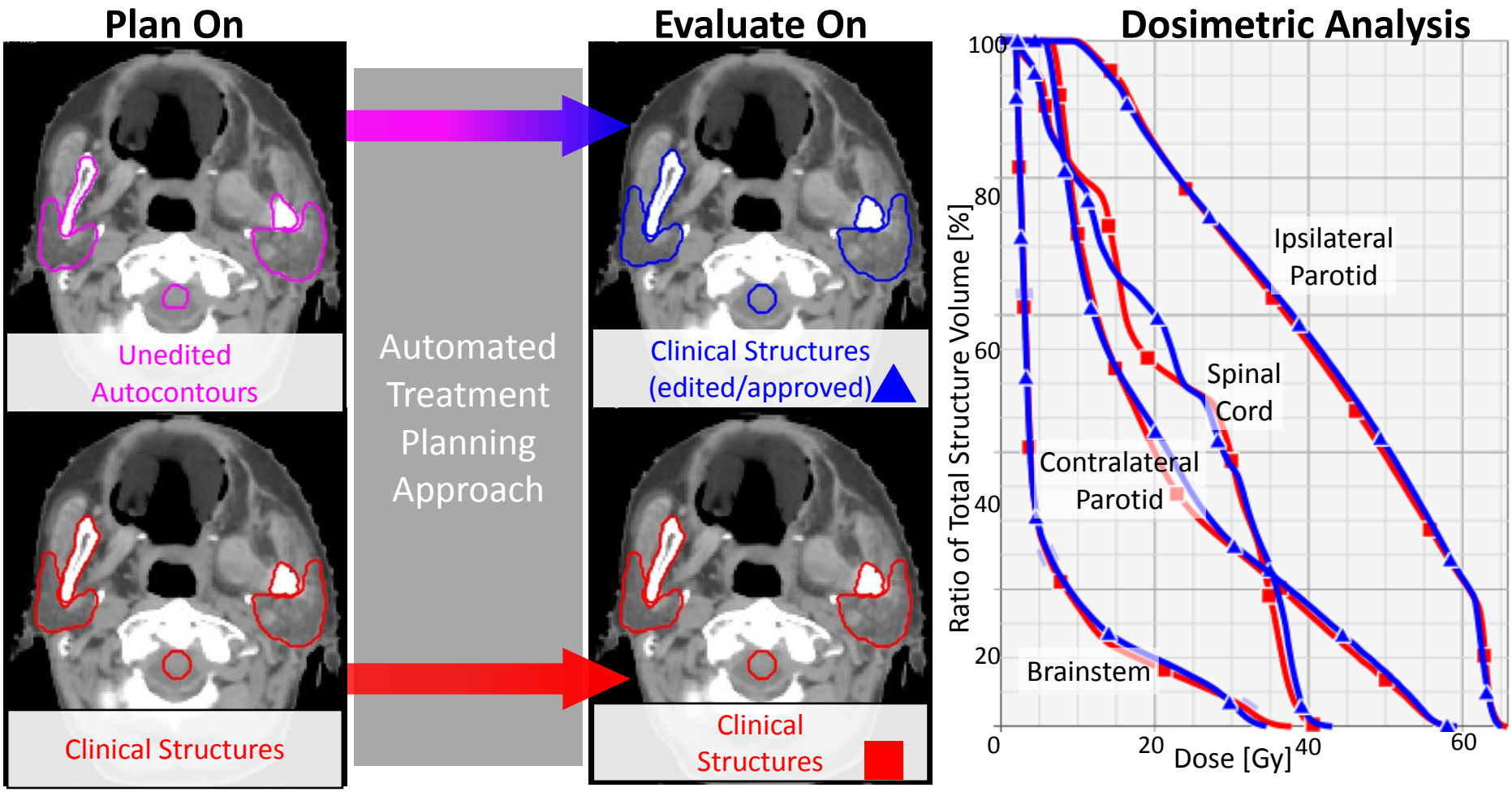


Figure 19. Experiment to identify the dosimetric impact of using autocontoured normal tissues for treatment planning. Using both the unedited autocontours (top) and clinical structures (bottom) the planning strategy described in Chapter 4 was used to create high quality head-and-neck treatment plans. Evaluation of the plans was carried out on the "true" physician edited/approved structures, any differences in the DVH curves identifiy the impact the chioce of planning structure may have on th treatment plan.

Correlation Dosimetric Impact of Normal Tissue Autocontouring to Quantitative Predictors

Given the expected distribution of clinical edits and the knowledge of their impact on treatment plan quality we wished to investigate the possibility of predicting, prior to plan optimization and dose calculation, the potential impact of edits to normal tissue autocontours for four key normal structures in the head-and-neck; the spinal cord, the brainstem, and the contralateral and ipsilateral parotid glands. It may be expected that the relative importance of edits to normal tissue contours is related to patient specific features. For example, if the brainstem is close to the target volume, than edits to this autocontour may be more impactful than if the brainstem was at a distance from the targets. The correlation of the patient and structure features to the dosimetric impact of planning on these structures could provide the user a pre-assessment of potential autocontour quality.

To investigate this, three dosimetric metrics including; the absolute dose to the clinical structure, the absolute difference in dose to the clinical structure when planned on either unedited autocontour or the clinical structure, and the relative dose difference when planned on the two structures, were evaluated for correlation with several features of the structures.

Predictive metrics included the clinical edits to the structure, the prescription dose levels of the (up to three) PTVs, the minimum distance between the autocontour and the PTVs, the maximum distance of the closest 10% of points of the autocontour and the PTVs, and the Dice Similarity coefficient between the autocontours and the PTVs analyzed in 3 ways (1) the autocontour and the original physician drawn PTVs (2) the autocontours and the PTVs with a uniform expansion of 0.5cm and 1cm and (3) the convex hulls of the autocontours and the PTVs. Examples of these structures and metrics can be found in Figure 20. The convex hull of the normal tissues are the smallest convex shape which encompasses the contour. They, for example, may provide insight when a target is within the concave region of a parotid contour.

The results of this fourth experiment are found on page 119.

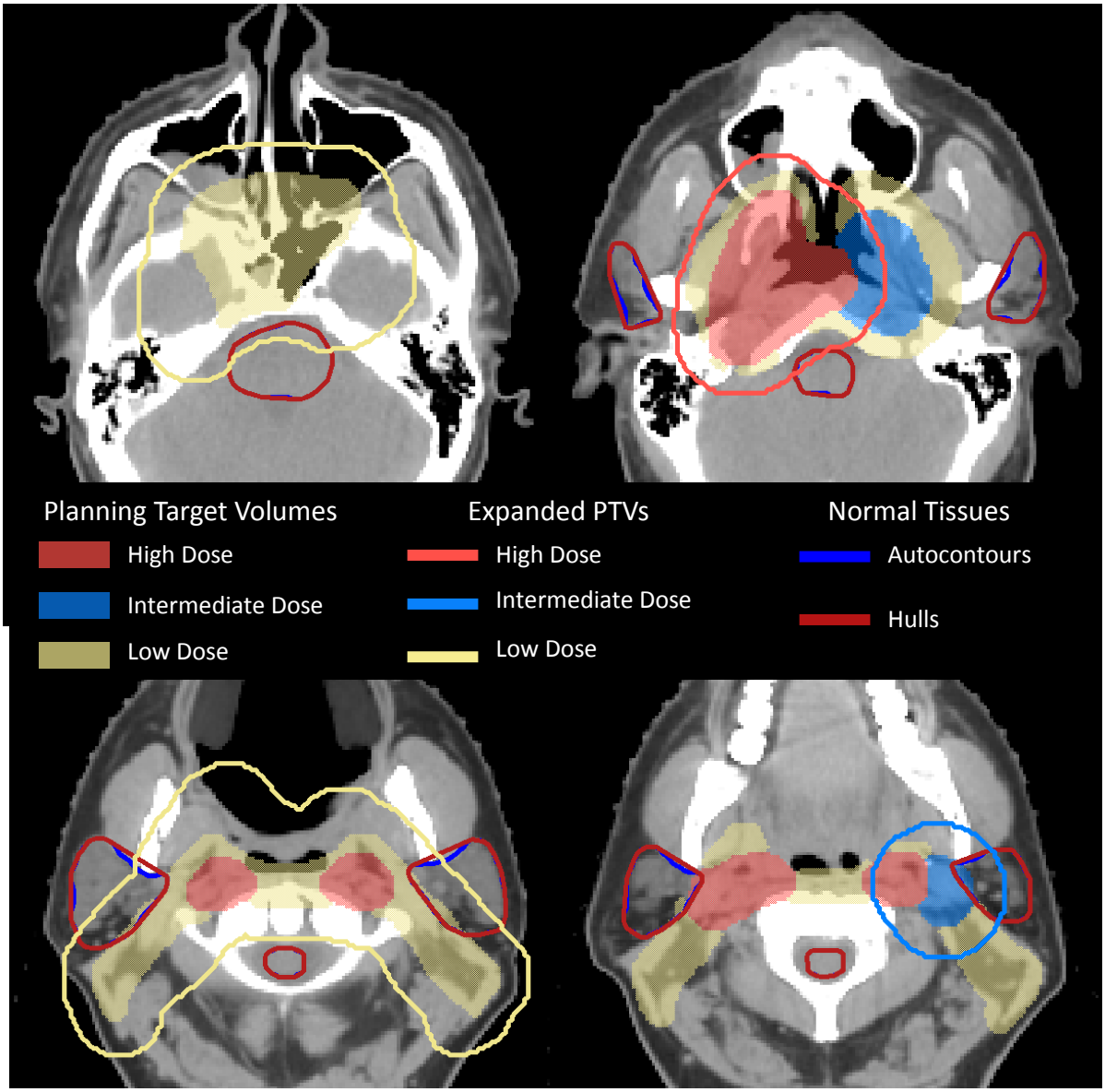


Figure 20. Patient specific metrics for correlation with the effect ofthe use of autocontoured targets for treatment planning. For four normal structures (brainstem, parotid glands, and spinal cord), in blue, metrics which quantify the spatial relationship to each of the patients' target volumes were calculated and included the minimum distance and minimum 10% of distances to each PTV (in colorwash, by dose level), the DSC and MSD with the PTV and with expanded PTVs (solid line, corresponding color), shown here for 1 cm uniform expansions, and the DSC and MSD of the normal tissue hulls (in maroon) with the PTVs.

Results

Analysis of Autocontouring Algorithms

Pilot Test of 4 Contouring Algorithms

Considering the 10 initial test patients, the in-house MACS contouring algorithm was best performing when considering all 8 normal structures. For some structures, the Smart Segmentation® approach with and/or without fusion was not investigated owing to the success of the other autocontouring techniques (see Table 11). For lung contours the Smart Detection® algorithm performed the best with an average rating of 4.6 and a minimum of 3, indicating minor edits. The lung was the worst performing structure for the in-house algorithm, with an average rating of 3.5 and a minimum rating of 1. The main failure mode noted was in the inferior portion of the lung, where, due to the deformation fields, the combining of several contours in the STAPLE algorithm and the varying scan extent (and thus portion of the lung) in each of the atlas patients, the contours often have small slices inferiorly or occasionally islands. For many patients this part of the lung is distant from the treated volume, and given that the whole lung is seldom included in the simulation scan and therefore whole lung DVH metrics cannot be quantitated may not be included in the dose calculation region. For these ten patients, the most inferior portion of the PTV was at an average distance of 11.5 cm from the slices of inferior lung with noticeable errors, and for the patient with the closest PTV-to-inferior-lung distance the dose grid was 3cm superior to the lung region with errors, see Figure 21. Given these things, we feel that errors on these bottom few slices are acceptable and may either be ignored or removed on post processing. For all other structures including; brainstem, cochleae, eyes, mandible, parotid glands, and the spinal cord, the MACS algorithm had an average rating better than all other algorithms.

Quantitative comparison of the autocontours from each of these 4 autocontouring algorithm on the same ten patients in shown in Table 12. For all structures MACS had a

higher average Dice similarity coefficient and lower average Mean surface distance and Hausdorff distance indicating that these structure agree more closely with independently drawn physician structures.

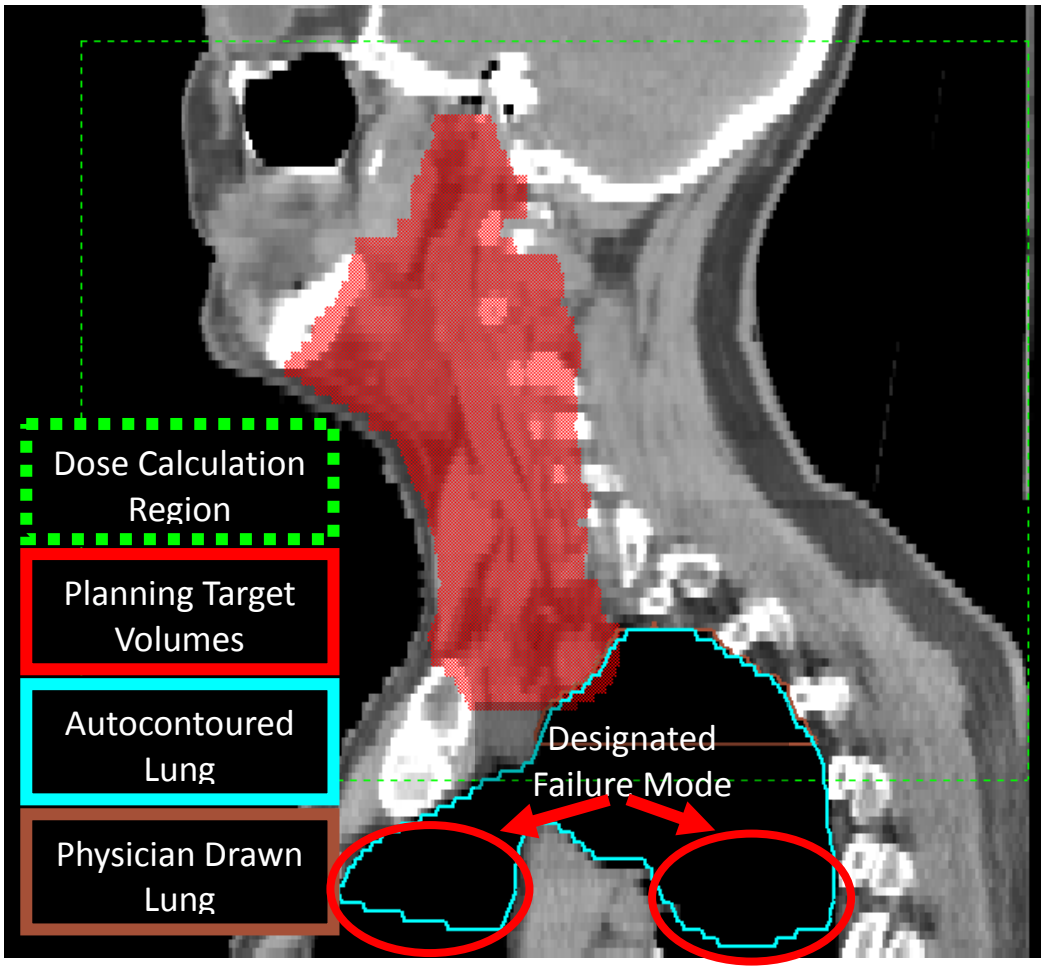


Figure 21. Example of autocontouring errors at the inferior structure extent. This patient, with the lowest physician score had large areas for which the autocontoured lung (cyan) failed to contour, as shown in the red circles. However, the physician drawn/approved structure (brown) also had large areas of incomplete contour. However, these missed areas are at the most inferior portion, or outside of, the dose grid (dashed green) and far from the contoured planning target volumes (red color wash).

A reoccurring failure mode of the smart detection algorithm, as cited by the rating physician, was structures that the structures were too large or were offset. This feedback prompted the decision to provide a third contouring approach in which these contours were fused using a majority voting algorithm. Unfortunately, many of the fused contours were also reported as being too large or offset, indicating that neither method was suitable for automatic contouring purposes. For example, the average volume across the ten patients of the physician drawn brainstem was 25.0 cc, the average volume of the MACS autocontoured brainstems was 23.85 cc, the average volume of the single and fused Smart Segmentation brainstem contours was 30.72 cc and 31.26 cc, respectively.

Table 11. Physician ratings of four autocontouring algorithms on ten patients. The mean, standard deviation, and minimum and maximum are shown for four autocontouring algorithms for ten patients. Contours were rated on a five point scale by a single physician. For brain, spinal cord and lungs, due to the success of other algorithms, Smart Segmentation® with and/or without fusion were not analyzed. The smart detection algorithm was not available for the structures with no reported statistics.

	Physician Ratings of Autocontouring Algorithms											
	Smart Detection			Smart Segmentation			Smart Segmentation + Fusion			Multi Atlas Contouring Service		
Structure	N	mean ± std	[min-max]	N	mean ± std	[min-max]	N	mean ± std	[min-max]	N	mean ± std	[min-max]
Brain	10	4.9 ± 0.3	[4 - 5]							10	5.0 ± 0.0	[5 - 5]
Brainstem				91	3.1 ± 0.6	[2 - 5]	10	3.1 ± 0.7	[2 - 4]	10	4.6 ± 0.5	[4 - 5]
Cochleae				95	3.0 ± 0.4	[2 - 4]	20	2.8 ± 0.5	[1 - 3]	20	5.0 ± 0.2	[4 - 5]
Eyes	20	3.4 ± 0.9	[2 - 5]	185	3.3 ± 0.9	[2 - 5]	20	3.6 ± 0.6	[2 - 4]	20	3.6 ± 0.6	[3 - 5]
Lungs	10	4.6 ± 0.7	[3 - 5]	93	3.0 ± 1.0	[1 - 5]				10	3.5 ± 1.1	[1 - 5]
Mandible	10	2.8 ± 1.5	[1 - 5]	101	2.8 ± 0.8	[2 - 5]	10	2.9 ± 0.7	[2 - 4]	10	4.7 ± 0.5	[4 - 5]
Parotid Glands				172	2.9 ± 0.8	[1 - 5]	20	3.2 ± 0.9	[2 - 4]	20	4.5 ± 1.1	[1 - 5]
Spinal Cord*	10	4.2 ± 0.4	[4 - 5]	93	4.1 ± 0.4	[3 - 5]				10	5.0 ± 0.0	[5 - 5]

*Smart detection only allowed for the contouring of spinal canal

Table 12. Quantitative comparison of four autocontouring algorithms with independently drawn physician normal tissue contours. The Dice similarity coefficient, the mean surface distance, and the Hausdorff distance were used to compare the normal tissue autocontours for 8 structures generated on the same ten patients rated by the physician. The in-house multi-atlas contouring algorithms most closely matched the physician drawn structures with the highest average DSC and lowest average MSD and HD.

	Dice Similarity Coefficient							
	Smart Detection		Smart Segmentation		Smart Segmentation + Fusion		Multi Atlas Contouring Service	
Structure	N	mean ± std	N	mean ± std	N	mean ± std	N	mean ± std
Brain	10	0.98 ± 0.00					10	0.98 ± 0.00
Brainstem			91	0.74 ± 0.05	10	0.79 ± 0.04	10	0.88 ± 0.04
Cochleae			95	0.38 ± 0.09	20	0.38 ± 0.07	20	0.65 ± 0.09
Eyes	20	0.75 ± 0.05	185	0.74 ± 0.08	20	0.78 ± 0.06	20	0.87 ± 0.03
Lungs	10	0.96 ± 0.01	93	0.92 ± 0.06			10	0.92 ± 0.02
Mandible	10	0.67 ± 0.07	101	0.68 ± 0.08	10	0.71 ± 0.06	10	0.90 ± 0.03
Parotid Glands			172	0.67 ± 0.10	20	0.72 ± 0.10	20	0.84 ± 0.05
Spinal Cord*	10	0.57 ± 0.05	93	0.67 ± 0.07			10	0.81 ± 0.03
	Hausdorff Distance (cm)							
	Smart Detection		Smart Segmentation		Smart Segmentation + Fusion		Multi Atlas Contouring Service	
Structure	N	mean ± std	N	mean ± std	N	mean ± std	N	mean ± std
Brain	10	1.13 ± 0.24					10	0.80 ± 0.29
Brainstem			91	0.98 ± 0.27	10	0.83 ± 0.17	10	0.68 ± 0.24
Cochleae			95	0.84 ± 0.21	20	0.69 ± 0.19	20	0.30 ± 0.07
Eyes	20	0.58 ± 0.11	185	0.70 ± 0.27	20	0.57 ± 0.09	20	0.44 ± 0.12
Lungs	10	2.35 ± 0.89	93	2.30 ± 0.88			10	1.90 ± 0.66
Mandible	10	2.40 ± 1.06	101	1.71 ± 0.48	10	1.13 ± 0.27	10	0.63 ± 0.19
Parotid Glands			172	2.00 ± 0.76	20	2.00 ± 0.85	20	1.25 ± 0.75
Spinal Cord*	10	5.27 ± 1.98	93	2.90 ± 2.04			10	1.48 ± 1.18
	Mean Surface Distance (cm)							
	Smart Detection		Smart Segmentation		Smart Segmentation + Fusion		Multi Atlas Contouring Service	
Structure	N	mean ± std	N	mean ± std	N	mean ± std	N	mean ± std
Brain	10	0.08 ± 0.02					10	0.07 ± 0.02
Brainstem			91	0.28 ± 0.05	10	0.24 ± 0.05	10	0.14 ± 0.04
Cochleae			95	0.25 ± 0.05	20	0.23 ± 0.04	20	0.09 ± 0.03
Eyes	20	0.20 ± 0.04	185	0.23 ± 0.07	20	0.20 ± 0.05	20	0.11 ± 0.02
Lungs	10	0.10 ± 0.02	93	0.21 ± 0.22			10	0.25 ± 0.14
Mandible	10	0.39 ± 0.18	101	0.29 ± 0.09	10	0.24 ± 0.06	10	0.08 ± 0.02
Parotid Glands			172	0.38 ± 0.14	20	0.34 ± 0.14	20	0.18 ± 0.06
Spinal Cord*	10	0.53 ± 0.17	93	0.31 ± 0.13			10	0.14 ± 0.06

*Smart detection only allowed for the contouring of spinal canal

Further testing of in-house multi-atlas contouring service (MACS)

The MACS contouring algorithm was best performing on initial review and was therefore selected for comprehensive analysis. Central to the MACS algorithms is the atlas of patients from which contours are propagated. Of the 12 patients in the contouring atlas, nine were male and 11 had primary oropharynx disease, one patient had unknown primary. The mean age was 72 years. Ten were AJCC 7th edition clinical stage IVa, one IVb and one stage III; all were treated with curative intent.

For all 128 patients the distribution of physician ratings of the 8 normal structures on the same 5-point scale can be seen in Figure 22. One patient had a surgically removed parotid and for five patients the lungs were not visible in the patient CT, thus no rating was recorded for these structures. Of the eight normal structures, six were, on average, indicated as clinically acceptable for use without edits in DVH-based planning, scoring either a four or a five. The remaining two structures had ratings indicating the need for minor edits for use in DVH-based planning, depending on their spatial relation to the target volume. For all normal structures, 87% received ratings which indicate no need for edit for use in treatment planning (a score of 4 or five). Furthermore, 97% of normal structures received a rating which indicates that, at most, only minor editing is needed for use in DVH based planning.

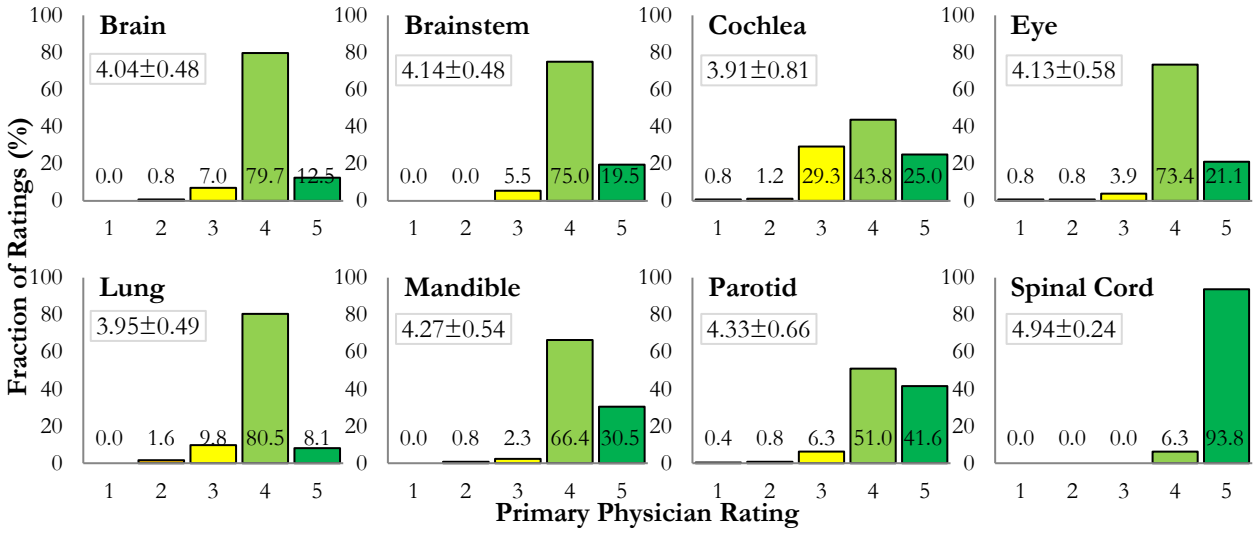


Figure 22. Distribution of the primary physician ratings of the in-house MACS algorithm. Contours of eight normal structures generated from the MACS algorithms were reviewed by a physician on a five point scale for 128 patients. The mean (± standard deviation) physician ratings are displayed in the graphs and the % receiving each rating overlaid on the bars

Inter-observer variability

To assess inter-observer variability, a subset of 10 randomly selected patients was reviewed by five additional radiation oncologists from four outside institutions. The radiation oncologists, per a self-reported questionnaire, had an average of 8.25 years of experience (range, 3.0-12.5 years) and contour and/or review an average of seven patients per week (range, 2-15 patients), spending an average of 95 minutes per patient on contouring (range, 45-180 minutes). For all structures except parotid glands, the scores assigned to the 10 patients differed significantly (Wilcoxon signed rank test, $p<0.05$) between the primary physician and at least one outside physician. However, no structure was significantly different between primary and all of the outside physicians.

For all structures, 45% (245/547) of the ratings by the outside physicians matched those of the primary physician and were classified as Category I agreements. Considering

Category II agreements, the physicians assigned an additional 48% (262/547) of the contours to same group, either as needing no or minor edit for use (48%) or as needing major edit for use (0%). Finally, only 7% of contours received scores indicating the need for major edits by one physician while needing no or minor edits by the other physician. In Table 13 the percentage of contours classified into each of the three agreement categories can be found for the eight normal structures assessed.

Table 13. Percentage of normal tissue rating pairs in three agreement categories. Category I indicates the scores by the two physician matched, Category II indicates that the ratings did not match but the contours were rated into the same "group" (either as needing no or minor edit or as needing major edit) by both physicians. Category III represents disagreement between two revewing physicians. Due to rounding, not all rows sum to 100.

	% of agreements in Category		
Structure	I	II	III
Brain	22	76	2
Brainstem	48	32	20
Cochleae	46	49	5
Eyes	47	42	11
Lungs	60	36	4
Mandible	40	58	2
Parotid gland	57	38	5
Spinal Cord	22	68	10
Total	45	48	7

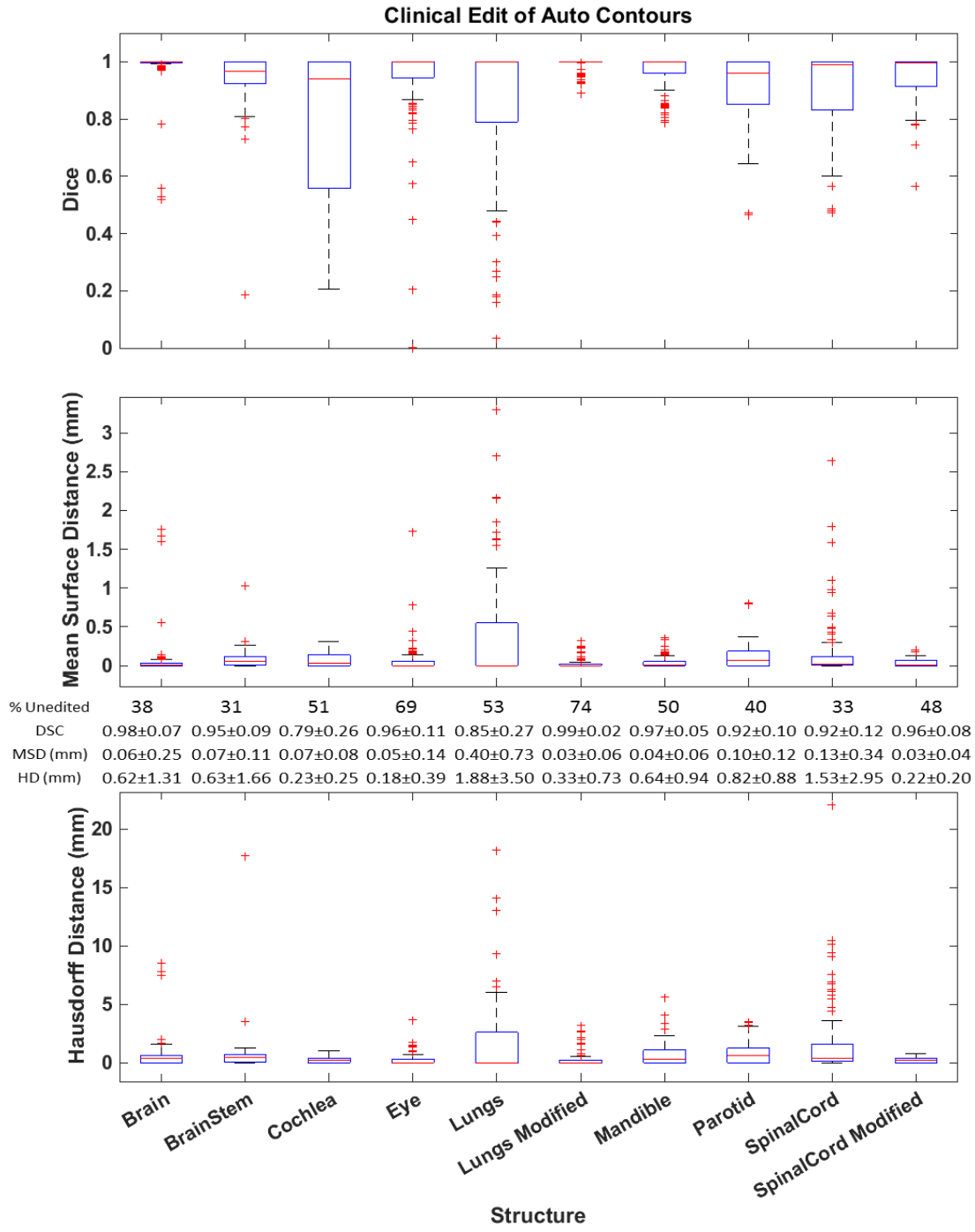
Clinical Use of Normal Tissue Autocontours

Due to its success on retrospective evaluation, the MACS autocontouring algorithm was implemented into the head-and-neck clinic at MD Anderson. During 10 months of clinical implementation, 22 radiation oncologists used the automated contouring software to generate

normal structure contours for 166 patients treated at our institution. Inclusion criteria included availability of the approved treatment plan and record of the requesting of the autocontours through the script in the treatment planning system. The 7 attending physicians who used the tool the most accounted for 23, 15, 14, 9, 7, 7 and 5% of the total use. The mean (± standard deviation) time required for generation of the autocontours was 11.5 ± 3.1 minutes when run on a Windows 2012-based PC with an 8-core Xeon E5-2697 v3 2.6-GHz CPU and 16 GB of memory. Multithread computing was enabled in the deformable registration algorithm, and 2 registration tasks were allowed to be run simultaneously on the server. This time does not require oversight by a physician or dosimetrist and can therefore occur simultaneous to other required treatment planning tasks. The distribution of autocontour edits, as measured by the Dice similarity coefficient, mean surface distance, and Hausdorff distance is shown in Figure 23. We measured edits only for structures which remained in the treatment plan at the time of treatment with the same naming convention as the automatically contoured structures.

Notably, radiation oncologists did not edit 49.8% of the contours for treatment planning. As shown in Figure 2, 31%, 40%, and 48% of automatically contoured brainstems, parotid glands, and modified spinal cords were not edited for clinical use. The structures edited least often were the eyes (69% were not edited) and modified lungs (74% were not edited). The Dice similarity coefficient was lowest for the cochlea (0.79±0.26). The maximum mean surface distance edits were seen for the unmodified lung avoidance structure (3.29 mm) and unmodified spinal cord (2.65 mm), due primarily to inferior CT scan extent discrepancy. However, considering edit to the modified lung avoidance and spinal cord structures, the maximum mean surface distances were decreased for lungs and spinal cord to 0.33mm and 0.20mm, respectively. This reduction is shown in Figure 23 as a reduction in both the medians and interquartile ranges of edits to the modified lungs and spinal cord structures relative to those edits of the unmodified versions.

A non-paired, one-sided *t* test showed that for all structures, contours that were edited for clinical use were significantly closer to the target volume than contours that remained unedited in the treatment plan ($p < 0.003$).



	Brain	BrainStem	Cochlea	Eye	Lungs	Lungs Modified	Mandible	Parotid	SpinalCord	SpinalCord Modified
% Unedited	38	31	51	69	53	74	50	40	33	48
DSC	0.98±0.07	0.95±0.09	0.79±0.26	0.96±0.11	0.85±0.27	0.99±0.02	0.97±0.05	0.92±0.10	0.92±0.12	0.96±0.08
MSD (mm)	0.06±0.25	0.07±0.11	0.07±0.08	0.05±0.14	0.40±0.73	0.03±0.06	0.04±0.06	0.10±0.12	0.13±0.34	0.03±0.04
HD (mm)	0.62±1.31	0.63±1.66	0.23±0.25	0.18±0.39	1.88±3.50	0.33±0.73	0.64±0.94	0.82±0.88	1.53±2.95	0.22±0.20

Figure 23. Distribution of the clinical edits to autocontours. The red line within the box plot represents the median and box edges represent the 25th and 75th percentile. Outliers are indicated by red crosses and are values outside the 25th or 75th percentile by more than 1.5 times the interquartile range. Between the boxplots are the percent of unedited contours and

the means (± standard deviations) of the Dice similarity coefficients and the mean surface distances for each for the automatically contoured or modified structures. Abbreviations: DSC = Dice Similarity Coefficient, MSD = Mean Surface Distance, HD = Hausdorff distance.

Given these clinical edits, we sought to identify the minimum uniform expansion to the autocontours needed to encompass either 95% or 100% of the clinically edited contour for 90% or 95% of the patients. Table 14 lists the minimum uniform expansions needed for each of the eight normal structures examined in our study, including the inferiorly modified lung and spinal cord contours. If autocontours are used without edit, our findings show that a uniform expansion between 0 and 5 mm would provide coverage of 95% of the physician edited structure for 90% of the population. To cover 100% of the physician edited structure for 95% of the population the minimum uniform expansion needed ranges from 4mm (cochlea) to more than 15 mm (lungs, mandible, parotid).

Table 14. Minimum uniform expansion to autocontoured normal tissues needed to cover a given fraction of the structure for a given fraction of the population.

	Patient population			
	90%		95%	
	Contour coverage			
Structure	95%	100%	95%	100%
Brain	0.00	7.53	0.00	10.64
Brainstem	3.55	7.22	4.51	8.88
Cochlea	2.28	3.02	3.42	4.00
Eye	1.64	3.34	2.06	5.28
Lung (modified)	0.64	>15	4.20	>15
Mandible	1.74	12.80	3.59	>15
Parotid gland	4.67	>15	9.05	>15
Spinal cord (modified)	0.98	3.74	1.63	4.81

Machine Learning Model for Prediction of Autocontour Errors

As necessary for the implementation of autocontouring as part of a fully automated treatment planning approach a method to detect autocontouring errors was investigated. Detection of gross and simulated errors represents a check of safety and catching of significant autocontouring errors. The detection of smaller clinical edits represents the possibility to flag to the user of an automated treatment planning approach that autocontours presented may require editing.

Simulated failure models

The performance of the RF models to predict simulated and gross errors is summarized in Table 15. The overall sensitivity and specificity of the models to gross and simulated error prediction were greater than 0.85 for all structures. When considering simulated errors due to small shifts and expansions, we found that the sensitivity was greater than 0.9, and this improved with the size of the shifts and expansions.

When considering patients in non-standard positions, we observed that the model sensitivity was slightly lower for some structures; a minimum sensitivity of 0.5 was seen for the cochlea, lungs, and mandible, although the size of the test set was small (2-14 patients). For one patient, a large primary gross tumor volume (637 cm^3) resulted in substantial automatic contouring errors for the surrounding contours, and for contours with a Dice similarity coefficient of less than 0.8 compared with the clinically edited structure (right eye and mandible), the contouring errors were successfully detected by their respective RF models. When considering the test patients in nonstandard positions plus the patients with other scan sites, we observed that at least half of the structures were predicted to have some contour errors in all patients. This suggests that if automatic contouring is predicted to fail in more than half of the structures, the patient is not a good candidate for this in-house autocontouring technique.

Table 15. Random forest model performance for the prediction of simulated and gross errors. Sizes of shifts and expansions can be found in Table 10.

Structure	Overall model performance*		Shifted, size 1		Shifted, size 2		Shifted, size 3		Shifted, size 4		Expanded, size 1		Expanded, size 2		Expanded, size 3		Nonstandard positions		Other scan locations	
	Sens	Spec	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens	N, test set	Sens
Brain	0.88	0.95	50	1.00	50	1.00	50	1.00	50	1.00	50	0.94	50	1.00	50	1.00	7	0.71	13	0.46
Brainstem	0.90	1.00	50	0.96	50	0.98	50	1.00	50	1.00	50	0.74	50	1.00	50	1.00	5	0.80	13	0.77
Cochleae	0.87	1.00	50	1.00	50	1.00	50	1.00	50	1.00	50	0.84	50	1.00	50	1.00	14	0.50	26	0.38
Eyes	0.89	0.94	50	0.98	50	1.00	50	1.00	50	1.00	50	0.84	50	0.98	50	1.00	16	0.94	26	0.77
Lungs	0.88	1.00	50	0.98	50	0.98	50	1.00	50	0.96	50	1.00	50	0.98	50	1.00	2	0.50	13	0.54
Mandible	0.90	1.00	50	1.00	50	1.00	50	1.00	50	1.00	50	1.00	50	1.00	50	1.00	8	0.50	13	0.77
Parotid glands	0.85	0.98	50	0.92	50	0.96	50	1.00	50	0.98	50	0.96	50	1.00	50	1.00	15	0.60	26	0.69
Spinal cord	0.91	1.00	50	0.96	50	0.98	50	0.96	50	0.98	50	1.00	50	1.00	50	1.00	3	0.67	13	0.77

*Sens, sensitivity; spec, specificity.

For all structures a metric derived from the HU distribution or an agreement metric with the independent deformable image registration contouring technique was the most important predictor. Other important predictor classes were derived from the autocontour distance to bony structures and the quantitative comparison to an independently and heuristically derived autocontour.

Prediction of clinical edits

Table 16 shows results of the overall performance of the RF models for the prediction of clinical edits to automatic contours (i.e., true errors), in which the thresholds of Hausdorff distances greater than zero (any clinical edit) or greater than 5 mm (significant clinical edit) were considered for contour errors. On average, 40% of contours (including modified structures) were edited for clinical use and 26% were edited such that the Hausdorff distance was greater than 5 mm. Models were not trained to predict significant (≥5-mm Hausdorff distance) edits to the brain, lung, or spinal cord contours owing to the very small proportion of contours edited and available for model building and testing purposes.

Table 16. Random forest model performance for the prediction of clinical edits to autocontours. Two random forest models were trained for each structure, in which automatic contour error was defined as a Hausdorff distance (compared with the edited automatic contour) of greater than 0 (i.e., any clinical edit) or greater than 5 mm (i.e., significant clinical edit).

		Any clinical edit				Significant clinical edit			
Structure	**No.**	**No. edited**	**Accuracy**	**Sensitivity**	**Specificity**	**No. edited**	**Accuracy**	**Sensitivity**	**Specificity**
Brain	113	63	0.56	0.71	0.36				
Brainstem	109	60	0.55	0.92	0.10	39	0.56	0.54	0.57
Cochleae	220	79	0.57	0.39	0.67	31	0.57	0.48	0.58
Eyes	205	58	0.68	0.07	0.92	16	0.73	0.75	0.72
Lungs	71	14	0.79	0.00	0.98				
Mandible	113	36	0.62	0.36	0.74	29	0.57	0.59	0.56
Parotid gland	214	102	0.61	0.80	0.44	96	0.63	0.63	0.63
Spinal cord	108	38	0.53	0.68	0.44				

The model performance over time, both for any clinical edit and for large clinical edit can be seen in Figure 24 and Figure 25, respectively. Little model improvement was seen over time and performance plateaued for most structures after 10 rounds of testing. Generally, the RF models to predict any autocontour edits did not perform as well as expected, with an average accuracy of 0.61. The RF model for the prediction of edits to the lung contours did not correctly identify any edited contours, though only 14 of 71 were edited and had Dice values with an average and standard deviation of 0.96±0.04.

The RF models to predict significant clinical edits performed, on average, better than those to predict any clinical edit; the best performing model for the prediction of significant clinical edits was for the eyes. Model performance improved when only contours that had significant clinical edits were considered. The RF models for the prediction of significant clinical edits were 100% sensitive to edits with Hausdorff distance greater 12 mm to the

brainstem, 7 mm to the cochlea, 6 mm to the eyes, and 22 mm to the mandible contours and were 76% sensitive to edits with Hausdorff distance greater than 2.2 cm to parotid contours. This indicates that the size of edit to which the models are sensitive may be larger than for edit size for which it was built.

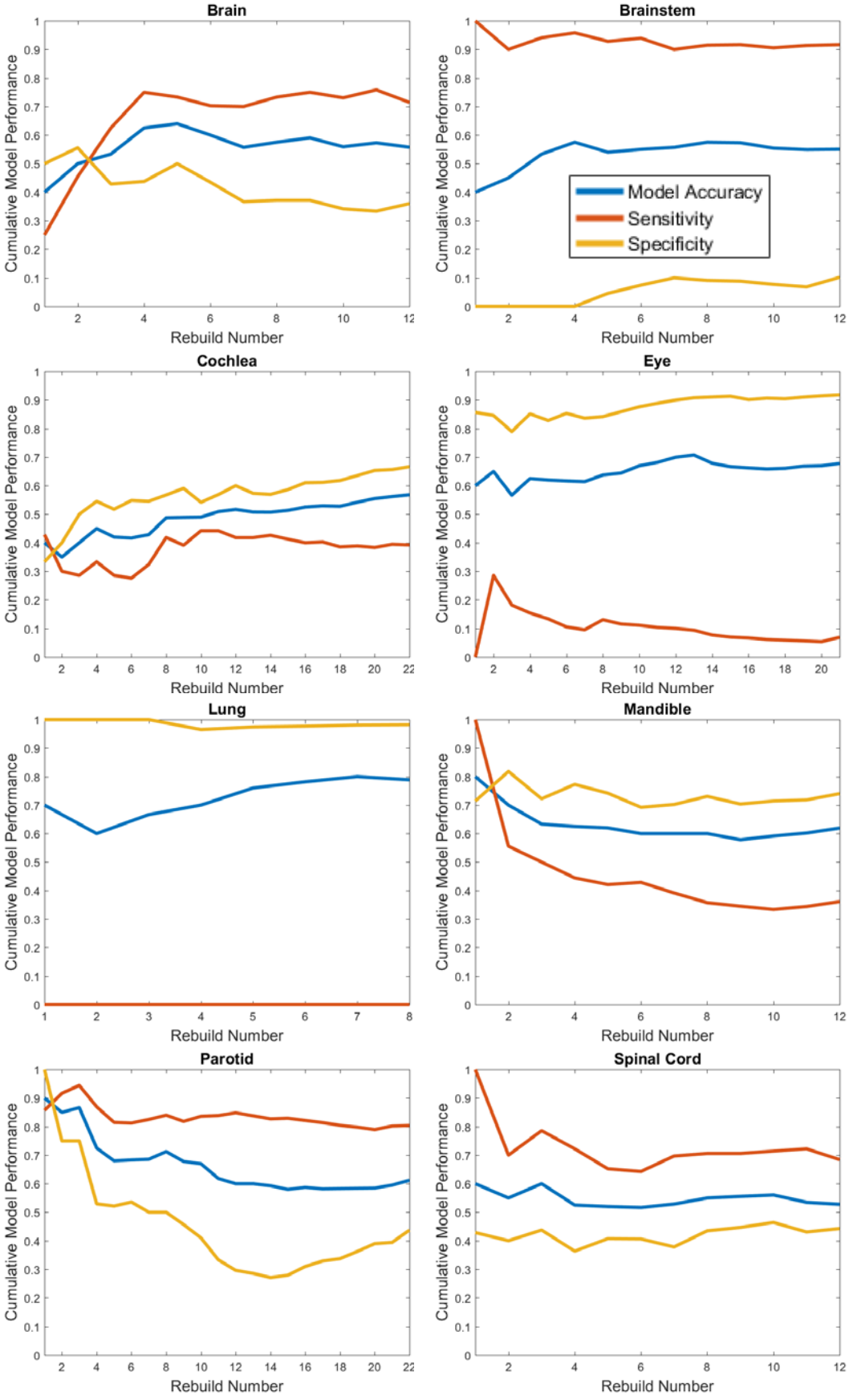

Brain
Brainstem
Cochlea
Eye
Lung
Mandible
Parotid
Spinal Cord
Cumulative Model Performance
Rebuild Number
Model Accuracy
Sensitivity
Specificity

Figure 24. Performance of QA models to detect any clinical edit. Models with high sensitivity (red) (e.g. brainstem, parotid) generally had low specificity (yellow). The accuracy (blue) of the detection of clinical edits to normal tissue structures was generally low, less than 0.6 for 4 of 8 structures. (Previous page)

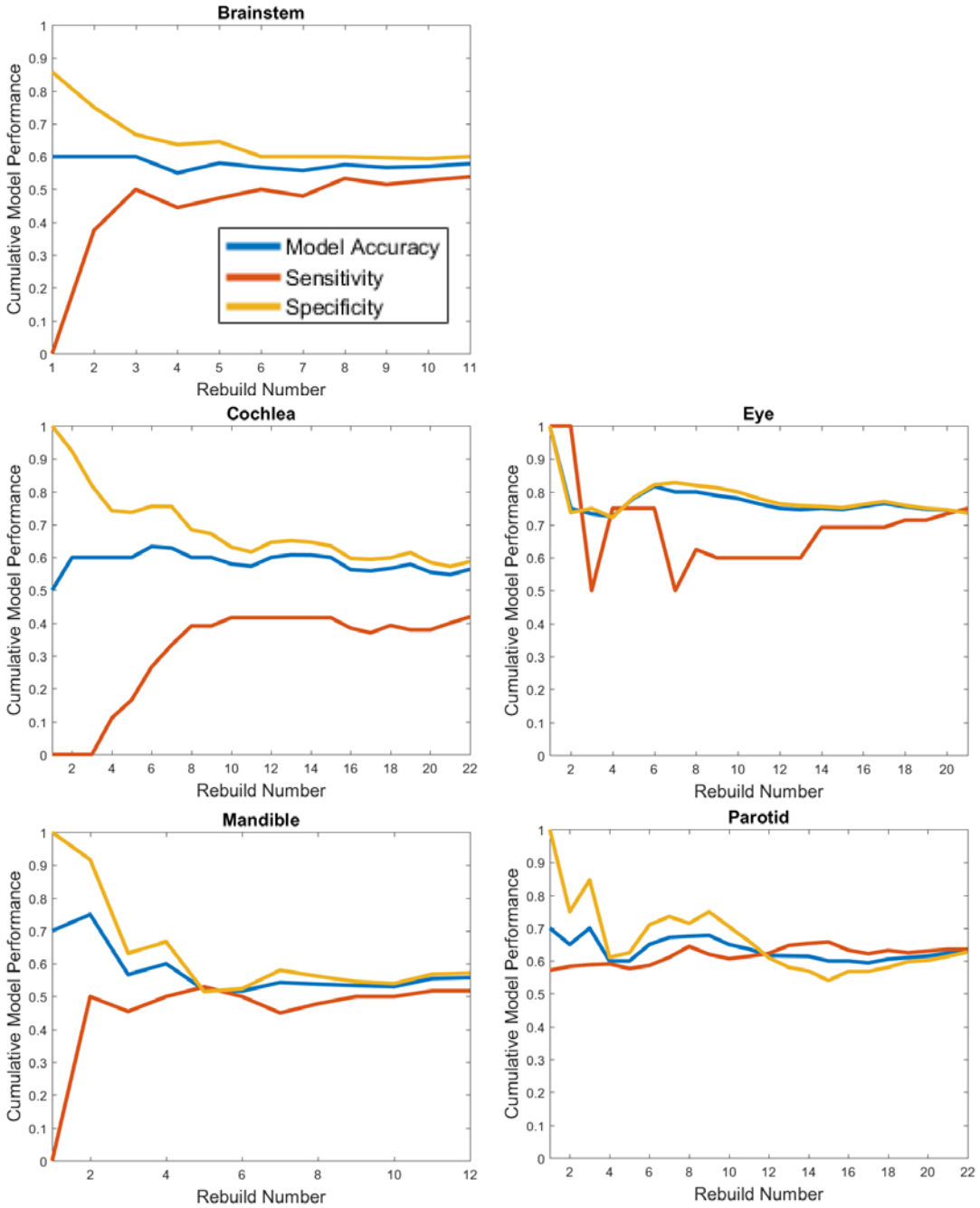


Figure 25. Performance of QA models to detect large clinical edits. Models were rebuilt after ten patients. Model accuracy (blue) was above 0.55 for all structure, with the model to detect errors in the eyes having the highest accuracy at 0.73. Model performance plateaued over time.(Previous page)

Among the most important predictors were a quantitative comparison to the independent deformable image registration contouring technique and metrics from the HU distribution. After the initial RF was built, the models were updated after every 10 structures. The accuracy, sensitivity, and specificity of the models were tracked over time, although an improvement in model performance was not observed.

Insight into model performance may be had by understanding the reasons for which contours were edited. A non-paired, one-sided *t* test showed that for all structures, contours that were edited for clinical use were significantly closer to the target volume than contours that remained unedited in the treatment plan ($p < 0.003$). For the RF models least sensitive to contour editing (lungs, eyes, and mandible), the edited contours were on average 1.9 cm (lungs), 1.9 cm (eyes), and 0.4 cm (mandible) closer to the target volume than contours that were not edited for clinical use. This result may indicate that reasons other than anatomical accuracy (e.g. potential dosimetric effect) are used for contour edit decision making.

Dosimetric Impact of Normal Tissue Autocontours for Treatment Planning

Patients Cohort

For the eight structures implemented for autocontouring into our clinical practice, the distribution of clinical edits seen clinically were compared using a two sided t-test to both the group of 54 patients selected from this larger cohort and to the group of 29 patients treated on a clinical trial. The results are shown in Table 17. For all structures, except eye, we found no significant difference in the mean of the distribution of dice similarity coefficients or Hausdorff

distances between the cohort of 54 patients and all patients with clinically edited normal structures. For eye, the mean Dice similarity coefficient for 289 clinical edited eye structures was 0.96±0.11, for this cohort the edits were slightly less with a mean and standard deviation of DSC of 0.98±0.04.

Considering the 29 patients treated on a clinical trial and when normal tissues were drawn independently there was a significant difference in the distribution of clinical and autocontour agreement. The difference between autocontours and physician structures was *larger* for this group of patients than for those with clinical edited autocontours indicating that the use of these for planning would represent a worst case scenario, as clinical edits would be expected to be smaller.

Table 17. Comparison of contour disagreement in the dosimetric cohorts with clinical edits to autocontours. A two sided t-test was used to test the hypothesis that the distribution of edits sizes for both the cohort of 54 patients with clinical edited autocontours and the cohort of patient with independently contours normal structures had the same mean as all patients with clinically edited normal structures, as described in the previous section.

	p-value, two sided t test			
	Independently Drawn Contours (n=29)		Edited contours (n=54)	
Structure	Dice	Hausdorff Distance	Dice	Hausdorff Distance
Brain	0.60	<0.01	0.28	0.38
Brainstem	<0.01	<0.01	0.43	0.16
Cochleae	<0.01	<0.01	0.48	0.68
Eye	0.01	0.06	0.03	0.01
Lungs	No Structures		0.94	0.99
Mandible	<0.01	<0.01	0.56	0.23
Parotid Glands	<0.01	<0.01	0.87	0.83
Spinal Cord	<0.01	<0.01	0.33	0.89

Dosimetric Impact

A paired Wilcoxon signed rank test was used to compare the treatment plans using the physician drawn/edited structures with those created on unedited normal tissue autocontours. The comparison was done independently for the two patient groups (1) clinical structures derived from autocontours and (2) structures drawn independently from the autocontours.

Considering the group of patients from which physician structures were derived from autocontours, DVH metrics which showed a significantly w*orsening* effect on the DVH metric included the mean doses to both the ipsilateral and contralateral parotid glands and the volume of the ipsilateral parotid gland receiving more than 30Gy (p=0.01). The p-values of the Wilcoxon signed rank test for all structures can be found in *Table 18*. These results suggest that for most structures using unedited autocontours generated using the method described in Chapter 4, results in no significant dosimetric effect on the treatment plan. The exception to this is for the parotid glands which are often in very close proximity to targets in the head-and-neck.

Table 18. Comparison of DVH metrics to true structures when planned on unedited structures and clinically edited structures for 54 patients with clinical edited normal structure autocontours. Shown are p-values of the paired Wilcoxon signed rank test. In green are parameters for which having planned on unedited autocontoured normal structures showed an improvement in the DVH metric compared to having planned on the true physician edited/approved structure. In red are metric for which planning on the unedited structures caused a significantly poorer performance at the clinical DVH point. All plans were evaluated on physician edited/approved structures.

Clinically Edited /Approved Structures			
Structure	N	Test Point	p-value, paired Wilcoxon rank sum
Spinal Cord	54	D_max	0.08
		V_45Gy	1.00
Brainstem	53	D_max	0.68
		V_54Gy	1.00
Ipsilateral Parotid	53	D_mean	0.01
		V_30Gy	0.01
Contralateral Parotid	50	D_mean	0.06
		V_30Gy	0.01
Cochleae	106	D_max	0.49
		V_35Gy	0.68
Eyes	88	D_max	1.00
		V_50Gy	0.13
Brain	50	D_max	0.16
		V_54Gy	0.95
High Dose PTV	54	D_1cc	0.18
		V_95%	0.68
Intermediate Dose PTV	45	V_95%	0.98
Low Dose PTV	42	V_95%	0.09
Detriment to DVH Metric		Improvement in DVH Metric	

Considering the group of patients treated on a clinical trial, and for which normal tissue contours were contoured independently, there was also limited dosimetric impact due to the use of unedited autocontours compared to physician drawn contours, *Table 19*. A significantly worsening effect due to the use of the unedited autocontoured normal structures was seen for the maximum dose to the brainstem ($p<0.01$) and both the maximum dose ($p<0.01$) and the volume receiving more than the maximally allowed dose of 45 Gy ($p<0.01$) for the spinal cord. However, for some patients, there is a clear discrepancy in the naming of structure and its contoured anatomy, this is especially easy to detect for the spinal cord contour. Common were two errors, either the physician drawn spinal cord either more closely represents a spinal canal contour, or is clearly *not* a contour of the spinal cord (i.e. includes part of the spinal column or brainstem. Examples of these errors can be seen in Figure 26. This analysis underlines the importance of assumptions which are made about the development of treatment plans and their retrospective analysis.

Table 19. Comparison of DVH metrics to true structures when planned on unedited structures and independently drawn physician contours. Shown are p-values of the paired Wilcoxon signed rank test. In red are metric for which planning on the unedited structures caused a significantly poorer performance at the clinical DVH point. All plans were evaluated on physician drawn structures.

Independently Physician Drawn Structures			
Structure	N	Test Point	p-value, paired Wilcoxon rank sum
Spinal Cord	29	D_max	<0.01
		V_45Gy	<0.01
Brainstem	27	D_max	<0.01
		V_54Gy	1.00
Ipsilateral Parotid	29	D_mean	0.55
		V_30Gy	0.77
Contralateral Parotid	26	D_mean	0.12
		V_30Gy	0.08
Ipsilateral Submandibular Gland	16	D_mean	0.21
Contralateral Submandibular Gland	21	D_mean	0.57
Cochleae	8	D_max	0.84
		V_35Gy	1.00
Optic Chiasm	21	D_max	0.14
		V_54Gy	1.00
Optic Nerves	46	D_max	0.44
		V_54Gy	0.50
Lens	56	D_max	0.34
		V_7Gy	0.46
Eyes	6	D_max	0.44
		V_50Gy	1.00
Brain	1	D_max	1.00
		V_54Gy	1.00
High Dose PTV	29	D_1cc	0.38
		V_95%	0.87
Intermediate Dose PTV	29	V_95%	0.17
Low Dose PTV	14	V_95%	0.95
Detriment to DVH Metric		Improvement in DVH Metric	

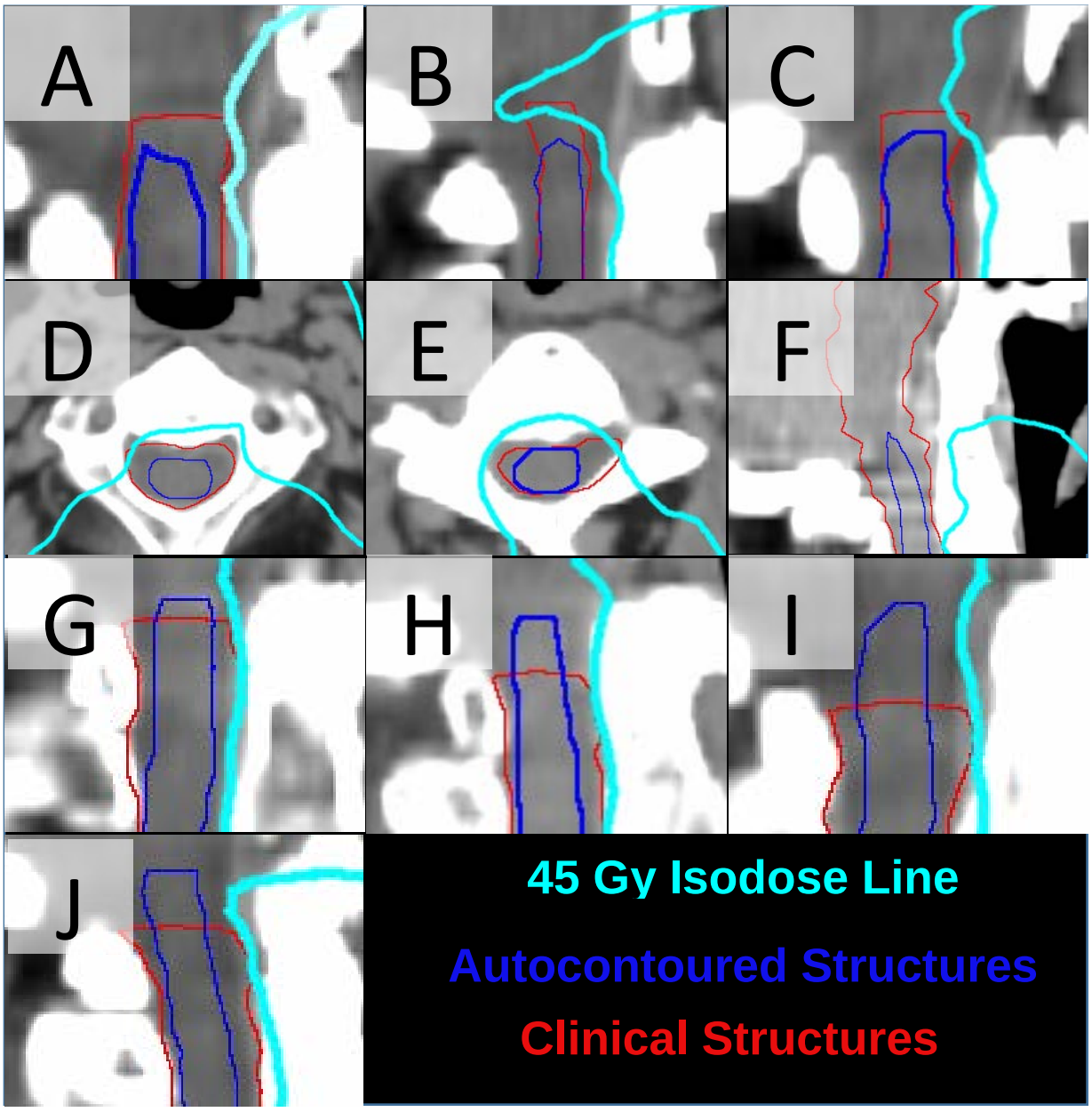


Figure 26. Ten patients with clinical spinal cords exceeding clinical constraints when planned using autocontours. For these ten patients, the maximum dose to the clinical spinal cord (red) exceeded 45Gy (cyan) when the autocontoured spinal cord (blue) was used for treatment planning; the maximum dose when the physician drawn spinal cord was used was less than 45Gy. In panels A-C, the clinical contour exceeds more superiorly than the autocontoured spinal cord. In panels D-J, the clinical spinal cord more resembles a spinal canal. In panel F, the clinical spinal cord contour extends very far superior to include the brainstem, which was not included as a separate structure for this patient.

In Figure 27, the dosimetric impact of the choice of planning structure can be seen. For 898 physician drawn normal structures, 474 showed a decrease in plan quality and 424 showed a dosimetric improvement due to planning on unedited autocontours. For spinal cord, 13 of 83 plans (12 of which were drawn independently from autocontours) had a spinal cord which exceeded clinical dose constraints when planning was done on autocontoured spinal cords. Though, as discussed, and shown in Figure 26, all of these structures have physician edited/approved/drawn spinal cords which are not reflective of their naming. Considering all other normal structures, in only five cases (three parotid gland, one optic nerve, and one eye) did planning on an unedited autocontour result in an exceedance of the corresponding clinical threshold. The relative percentage difference of these five structures can be found in *Table 20* and were all less than 15%, indicating that while the structure did exceed clinical thresholds, the relative difference as small.

The results above should be considered with the understanding that comparisons were made *only* on structures which were included in the original treatment plan. It is likely that structures not included in the clinical treatment plan were not of significant dosimetric interest because, for example, the structure may be far from the target. In these cases using an unedited automatic structure in lieu of a clinically approved structure is likely to have very limited dosimetric impact and therefore results herein represent the expected dosimetric impact for a subset of patients for which accurate contouring is expected to be of consequence in the treatment plan and therefore these results may overestimate the impact on the population as a whole.

Table 20. Dosimetric Impact of structure which exceed clinical structure when autocontours are used for treatment planning. Of 815 structure s 5 exceeding corresponding clinical constraints when autocontours, instead of clinical structures, were used for treatment planning.

Structure	Dosimetric metric	Clinical Threshold	Dosimetric Value when planned on		Relative Difference (%)
			Clinical Structures	Automatic Structures	
Parotid Gland	Mean Dose	26 Gy	25.9 Gy	27.5 Gy	6.3
			24.7 Gy	26.0 Gy	5.5
Parotid Gland	V_30Gy	50%	47.0%	53.6%	14
Eyes	Max Dose	50 Gy	48.2 Gy	51.4 Gy	6.6
Optic Nerve	Max Dose	54 Gy	52.1 Gy	54.4 Gy	4.4

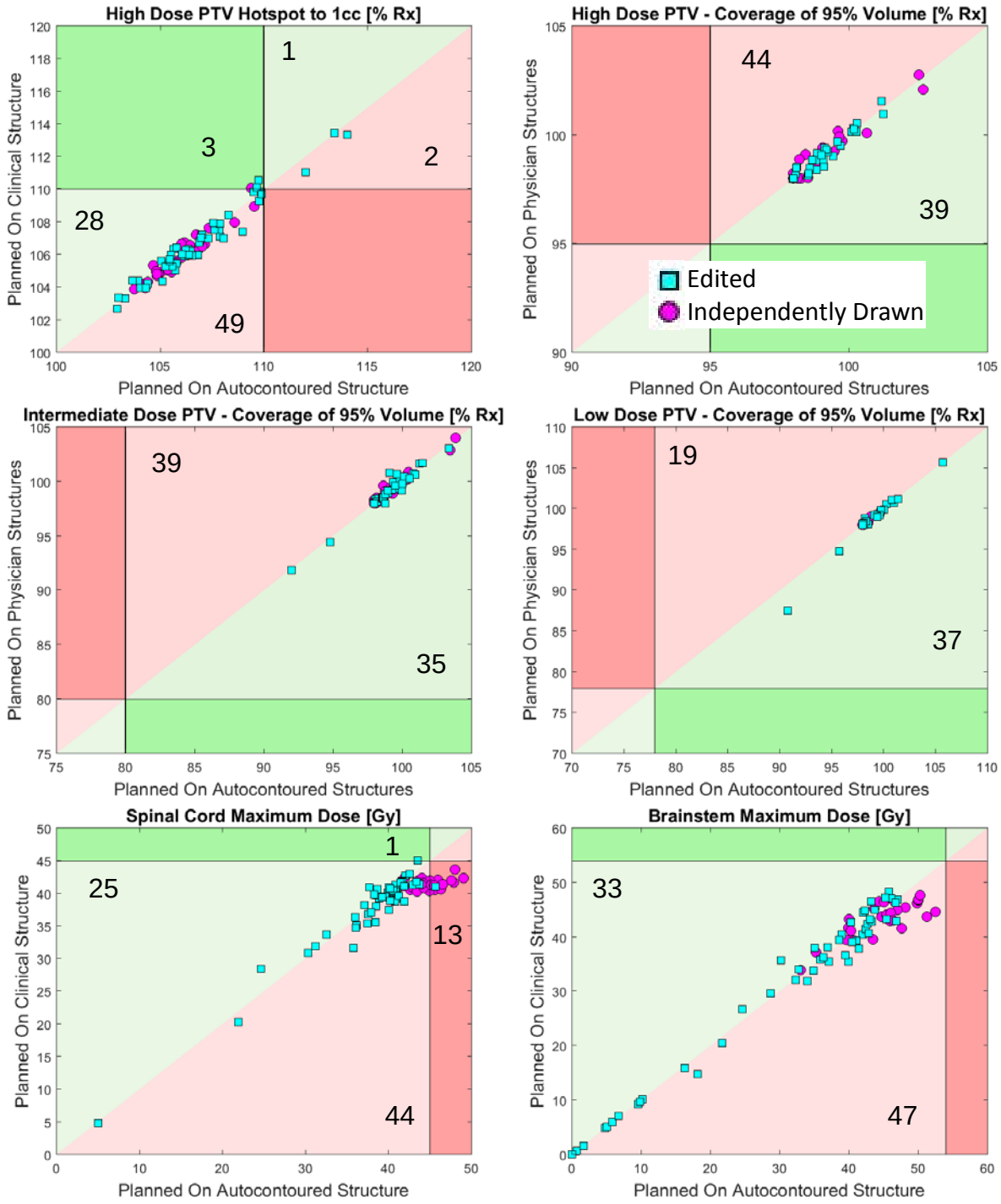


Figure 27. Distribution of relevant dosimetric points when planned on clinical structures and autocontoured structures. Along the x-axis, the DVH metric to true physician approved structures when planning on autocontoured structures and along the y-axis for the same patients the DVH metric when planned on the true clinical structure. Shades of green indicate

improvement in the dosimetric descriptor when planned on the autocontours, shades of red indicate poorer performance due to use of the autocontours for planning. Darker shades delineate common clinical thresholds for each dosimetric descriptor, with darker green indicating an improvement beyond the clinical threshold and darker red indicating a poorer performance beyond the clinical threshold. In cyan squares, for 54 patients, clinical structure originated from autocontours with physician approval/editing for clinical use. In magenta circles, other normal structures were independently drawn for clinical use (continued on following pages).

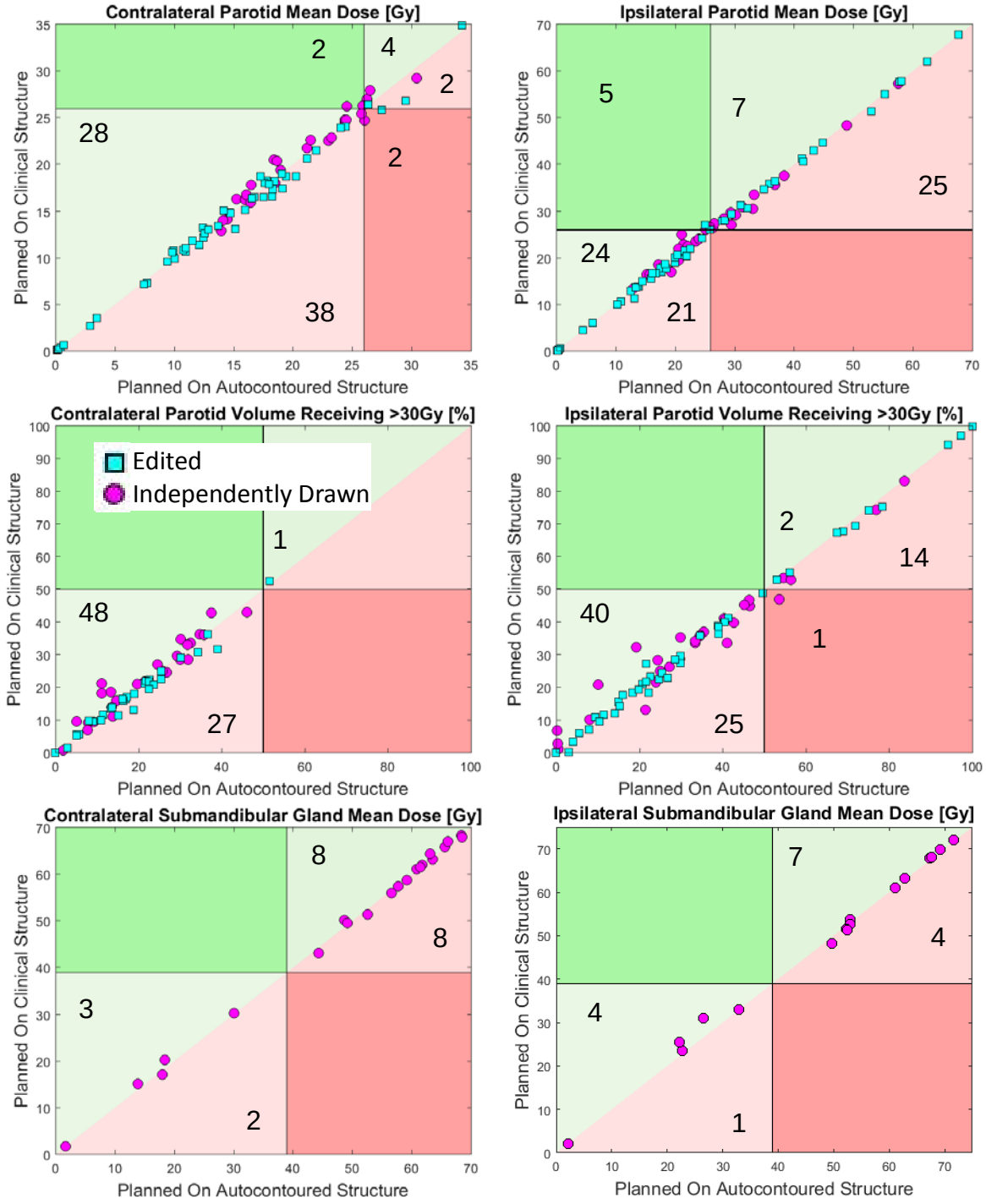


Figure 27. (Continued from previous pages)

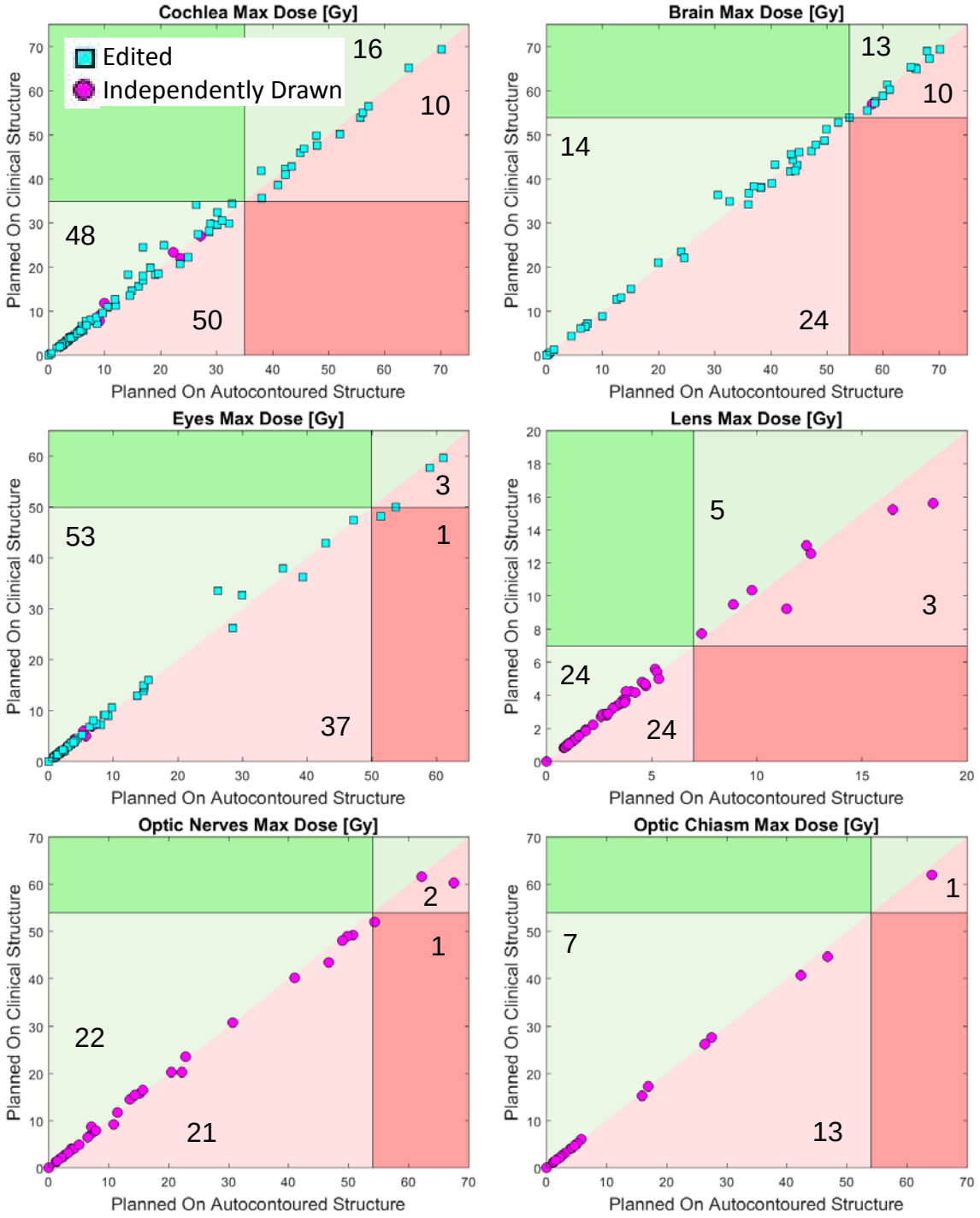


Figure 27. (Continued from previous pages)

Correlation to Patient Features

Considering four structures of particular importance to the treatment of head-and-neck cancer (spinal cord, brainstem, and contralateral and ipsilateral parotid glands), there were few examples when the use of unedited autocontoured structures impacted plan quality or dose delivered to true clinical structure as measured by clinically used dose metrics. However, variations in impact including positive and negative impact as well as a limited number of cases for which clinical constraints were exceeded in autocontours were used for planning indicate a possibility of detecting these patients for the warning to the user of a fully automated system or for further investigation. Three dosimetric metrics; the absolute dose to the clinical structure, the absolute difference in dose to the clinical structure when planned on either unedited autocontour or the clinical structure, and the relative dose difference when planned on the two structures, were evaluated for correlation with several features of the plans, as shown in Table 21.

For the spinal cord, brainstem, and contralateral parotid the absolute dose to the clinical structure when planned on the unedited autocontour was significantly correlated to the size of clinical edit, quantified as the Dice similarity agreement, Hausdorff distance, and mean surface distance. This indicates that the degree of clinical edit is correlated to its relative dose region. Unfortunately this data, the size of clinical edit, would not be available for incoming patients for which the need for clinical edits partly motivates the model prediction itself. The dosimetric impact for these three structures was also positively correlated to the prescription dose level of all three PTVs, and both the minimum distance and the distance of the 10% closest points to one or more PTVs. These correlations indicate that the higher a prescribed dose and the closer a normal tissue is to the PTVs, the higher the dose to the true normal structure will be.

The dosimetric impact due to the use of unedited autocontoured structures for treatment planning was assessed using the two remaining metrics; the absolute and relative dose differences. Significant correlation between features of the treatment plan and the dosimetric impact of the choice of planning structure was found for a limited number of plan features. For example, the absolute difference in the mean dose to the contralateral parotid was significantly correlation to both the minimum distance and the distance of the 10% closest points to the low dose PTV, this may be expected given that edits to the parotid may be more impactful if the target is near the structure. Considering the very limited impact the choice of planning structure had on the resultant plans, as presented in the previous section, strong correlations would not be expected.

Table 21. Correlation between dosimetric impact of using unedited autocontoured normal structures and patient and plan features. P-values of the hypothesis that there exist no relationship between the dosimetric impact (rows) and plan features (columns), significant established at <0.05 and shown in green. The absolute dose to the clinical structure of four key normal structures in the head-and-neck were significantly correlated with the size of clinical edit (although not available for prediction of new patients), the PTV prescribed dose level and the minimum distance to the target. The absolute and relative dose difference were less often correlated to plan features. No dosimetric features were correlated with the DSC of the targets or target variations.

p-values of Correlation Coefficient		Clinical Edit			Prescription Dose Level			Minimum Distance			Max Distance of 10% closest points		
	Structure	Dice	HD	MSD	PTV1	PTV2	PTV3	PTV1	PTV2	PTV3	PTV1	PTV2	PTV3
Absolute Dose	Spinal Cord	<0.01	0.01	0.02	<0.01	0.03	<0.01	<0.01	0.01	0.02	<0.01	0.01	0.01
	Brainstem	<0.01	<0.01	<0.01	0.01	0.03	<0.01	<0.01	0.05	0.07	<0.01	0.04	0.05
	Ipsilateral Parotid	0.68	0.72	0.74	0.35	0.71	0.86	0.98	0.66	<0.01	0.68	0.70	<0.01
	Contralateral Parotid	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	0.24	0.09	<0.01	0.23	0.07
Absolute Dose Difference	Spinal Cord	<0.01	0.01	0.04	0.04	0.19	0.08	0.01	<0.01	0.09	0.01	<0.01	0.05
	Brainstem	<0.01	<0.01	<0.01	0.61	0.61	0.45	0.52	0.77	0.42	0.58	0.78	0.41
	Ipsilateral Parotid	0.23	0.55	0.43	0.97	0.54	0.46	0.36	0.09	0.81	0.34	0.09	0.90
	Contralateral Parotid	0.20	0.33	0.35	0.28	0.75	0.30	0.56	0.18	0.68	0.66	0.16	0.73
Relative Dose Difference	Spinal Cord	<0.01	0.03	0.07	0.29	0.25	0.10	0.04	<0.01	0.10	0.06	<0.01	0.07
	Brainstem	<0.01	<0.01	<0.01	0.48	0.87	0.28	0.47	0.76	0.64	0.53	0.75	0.65
	Ipsilateral Parotid	0.37	0.94	0.76	0.82	0.34	0.68	0.17	0.04	0.90	0.15	0.04	0.83
	Contralateral Parotid	0.17	0.21	0.23	0.07	0.38	0.15	0.46	0.24	0.87	0.56	0.25	0.92
		Dice with PTVs			Dice of Structure Hull with PTVs			Dice with 0.5 cm expanded targets			Dice with 1cm expanded targets		
	Structure	PTV1	PTV2	PTV3	PTV1	PTV2	PTV3	PTV1	PTV2	PTV3	PTV1	PTV2	PTV3
Absolute Dose	Spinal Cord								0.81	0.49	0.23		
	Brainstem							0.22	0.09	0.49	0.18	0.74	0.36
	Ipsilateral Parotid	0.03	0.22	0.99	0.03	0.18	0.98	0.03	0.04	0.15	0.30	0.54	0.36
	Contralateral Parotid	0.61	0.61	0.08	0.65	0.63	0.07	0.27	0.14	0.86	0.88	0.13	0.17
Absolute Dose Difference	Spinal Cord								0.54	0.94	0.58		
	Brainstem							0.42	0.02	0.17	0.12	0.74	0.36
	Ipsilateral Parotid	0.26	0.19	0.73	0.31	0.25	0.75	0.26	0.26	0.30	0.23	0.94	0.83
	Contralateral Parotid	0.23	0.05	0.04	0.23	0.06	0.04	0.44	0.72	0.02	0.03	0.09	0.15
Relative Dose Difference	Spinal Cord								0.53	0.97	0.62		
	Brainstem							0.45	0.05	0.24	0.17	0.69	0.44
	Ipsilateral Parotid	0.22	0.15	0.54	0.25	0.18	0.55	0.21	0.21	0.18	0.12	0.79	0.95
	Contralateral Parotid	0.14	0.13	0.06	0.14	0.13	0.06	0.39	0.72	0.05	0.06	0.14	0.26

Discussion

We have investigated the use autocontoured normal structures in the head-and-neck for treatment planning purposes. We first evaluated retrospective and prospective algorithm performance including the clinical implementation of the 8 normal structures into our head-and-neck clinic. In an effort to ensure the safe use of autocontoured for automated treatment planning, which may or may not include the careful review of contours, we developed an approach to QA autocontours in order to flag them to the user. Finally, we investigated the potential use of unedited autocontours for treatment planning. The results show that autocontouring is a viable method to save time, reduce required infrastructure, while maintaining a high standard of care, and that the use of unedited autocontours has limited dosimetric impact.

In an analysis of four autocontouring algorithms, an in-house methods, MACS, was highest performing. Two commercial algorithms were also analyzed –Smart Segmentation® a deformable image registration technique which performance was not found to be adequate for clinical use and Smart Detection® a heuristic approach which worked well but is only available for a limited number of structures, and Smart Segmentation® a deformable image registration approach which performed poorly with or without the use of multiple atlas patients. While the heuristic approach slightly outperformed the in-house technique for a few structures, the advantages of using a single algorithm which include a reduction in the time needed to contour, a simplification in workflow, and the availability to modify the contours which is not possible using the commercial approach the in-house method alone was selected for further analysis. Interest in other automatic contouring methods are increasing, including the use of edge detection, image gradient, and voxel intensities, which do not require prior information or model building as well as shape model and machine learning techniques which use the contours of prior patients to inform the contour of new patients and have performed

well on select patient data sets (99). These techniques may offer advantages and the validation and assessment of these algorithms both in the clinic and for planning purposes may be evaluated using the principles presented here.

Given the known inter-physician variability in contouring(88, 117) and as the atlas for contour generation was derived from patients from the primary rating physician we sought to present autocontours from MACS to 5 physicians and, as expected, the acceptability of the contours, especially for use without edit, varied among physicians, however only 7% of contours received contours which indicated physician disagreement (i.e. outside of their own opinion of acceptable inter-physician variability). We expect similar results for a larger group of physicians.

We successfully implemented into routine clinical the autocontouring of 8 normal structures (11 total contours as three were bilateral). These 8 structures were chosen as they were the most often contoured structures in a sample of head-and-neck cancer patients. This clinical implementation provides ongoing data at the size of edits of autocontours and possible trends into normal tissue contouring practices. Other structures are also required and include the submandibular glands, optic structures, esophagus, and others (92). A limited analysis of these structures, presented in the appendix, show less success for their contouring compared to independently drawn physician structures and therefore these structures may warrant further analysis.

While no formal analysis of time savings or systematic method for physician feedback was developed, we believe that, in line with previous findings, autocontouring saves time (118-120) and that the continued use of the software indicates that it has been well received, with over 1000 patients having had these contours requested through the script in the treatment planning system. Upon implementation a description of the tool as well as guidance for its use was sent to the attending physician, however due to software limitations we were

unable to track approval and/or editing of the structures on a patient-by-patient basis and it was assumed that if normal structures were requested for the patient and image set of the final treatment plan and if the contours remained in the final treatment plan with the same naming as the autocontours (which is the same as our clinical naming) then the contours were edited and/or approved for clinical use. These assumptions likely have an impact of the fidelity of the data, and a limited number of instances which highlight the impact they may have were encountered. In one such example, for the patient and image set in question, the contours were requested through the script in the treatment planning system and remained in the final treatment plan, however, upon analysis of the cochlea structure it was found that the naming was inconsistent with the structure location (i.e. the left cochlea was on the right side of the body, and the right cochlea on the left). These structures were removed from analysis but other, less obvious examples, may have remained. Generally, clinical implementation was considered a success as nearly 50% of structures were not edited for clinical use and physicians appreciate and use the tool.

In the third experiment, a method to ensure the safe use of autocontours in automated treatment planning was investigated. Random forest models were developed such that a patient with grossly incorrect autocontours would not be allowed to continue in the automated treatment planning process. Additionally, the possibility of warning the user to suspected contour inaccuracies was investigated. Due to the success of the autocontouring algrotihm, there were few “true” autocontouring failures and simulated failures were therefore used to supplement. It is unlikely that these simulated errors (e.g. shifts and expansions) accurately represent the potential failure modes of an autocontouring algorithm. However, models to detect gross errors successfully identified patients in non-standard positions, for a patient with a very large gross tumor, and for a limited number of other gross contouring errors. For all patients in nonstandard positions and of other disease sites, at least half of the normal structures were flagged as failing by their respective models and therefore we suggest that if

more than half of the structures are flagged as failing by the gross error models then the patients should not continue in the autoplanning process.

Unfortunately, efforts to detect smaller contour edits found only limited success, with average accuracy just over 60%. The reason for the poor performance is likely multifaceted, though a driving reason is likely the (mis)alignment of the data collected with underlying assumptions. During clinical implementation it was assumed that contours which remain in the treatment plan represent the anatomical structure which corresponds to their name. An extension of this is the idea that edit of an autocontour indicates autocontouring error and also the inverse; lack of contour edit indicates perfect anatomical agreement. However, numerous examples which do not support this idea were identified. Further, the significance of the correlation between contour edit and distance to target suggest that contours are not edited solely for the purpose of anatomical accuracy. If a model to detect anatomical accuracy of contours is desired then vetting and careful development of a data set which matches this desire should be curated. While strict adherence to anatomical boundaries of normal structures for treatment planning purposes is ideal, contours for the purpose of treatment planning may serve other purposes which are not tracked, noted, or otherwise identifiable.

To investigate the impact that the use of unedited autocontours may have on the treatment plan we developed treatment plans on both clinically used structures (both edited and/or approved autocontours and independently drawn structures) and unedited versions. The results of this experiment indicate that the automated treatment planning strategy as presented in Chapter 4 is robust to edits on the size of those seen clinically and even larger differences seen when the contour is drawn independently and that using unedited structures for treatment planning, which may save up to an hour of editing time, may be a feasible option. Through this, again, we saw the impact of structures which do not align with their naming (121); for at least seven of ten patients with a spinal cord dose which exceeded

clinical constraint when the treatment plan was developed on unedited structures there was some error in the spinal cord contour.

There are several limitations and opportunities for further study. Generally, the curating of a data set which explicitly meets the needs of the hypothesis is desired, though the use of real clinical data exposes the purposes and methods with which a clinical tool may be utilized and therefore, should be considered. The structure set implemented clinically included only 8 normal structures, and though others were investigated including their impact of plan dosimetry, further investigation is needed. The study of the dosimetric impact of autocontours was completed for a cohort of only 54 patients, from 6 sub-sites in the head and neck, and the accrual of additional patients may reinforce the conclusions or may identify specific scenarios (e.g. groups of patients) for which the current conclusions may not apply. Finally, if we are to suggest that true anatomical structures are not required for high-quality treatment planning we must be cognizant of the impact or doing so. First, the naming of such structures should always be in line with their physician description (121). Second, normal tissue contours routinely serve purposes beyond that for treatment planning (e.g. for the analysis of normal tissue complications, in the case of retreatment, etc.), and if anatomically defined contours are not routinely generated there is a risk of losing new and clinically reliable data which may be used.

Over all, we believe this work represents the first comprehensive assessment of automatic contouring for a large set of normal structures in the head-and-neck, including over 2 years of use. Further, we have developed a technique to identify contours with gross errors which may jeopardize the safety of a fully automated treatment planning approach. Finally, this is the first study to evaluate the potential of using unedited autocontours of normal structures for head-and-neck treatment planning, and the results show that unedited autocontours do not significantly impact plan quality at clinically relevant dosimetric endpoints, and compared to clinical edited or approved autocontours. The use of autocontouring in the

head-and-neck, with or without editing, could dramatically reduce the time needed to head-and-neck treatment plan development, possibly allowing for the transition to advanced techniques in low-resource settings.

Conclusion

The automatic contouring of normal structures in the head-and-neck is a promising avenue, producing accurate contours which are suitable for use without edits for treatment planning.

Chapter 6: Automatic Contours of Intermediate and Low Dose Clinical Target Volumes and Their Use in an Automated Planning System

In this chapter we describe the results for Specific Aim 4 in which, through three experiments we evaluate the feasibility of the use of automatically countered intermediate and low dose target volumes in the head and neck. Our working hypothesis is that automatically contoured clinical target volumes can safely be used for treatment planning purposes.

Introduction

Towards a fully automated treatment planning system for the head-and-neck, we have presented the validation of the planning technique and optimization algorithms in Chapter 4 and the use of fully automated normal structure contours in Chapter 5. The next step towards full automation is the implementation of automatically contoured target volumes. Intermediate- and low- dose clinical target volumes are good candidates for automatic contouring as they are often based on anatomically defined tissues rather than, as for the high dose treatment volume, the cancerous tumor.

In the treatment of head-and-neck cancer, the irradiation of subclinical disease in the lymphatic system of the neck to between 45 and 60 Gy has long been shown to improve patient outcomes (1, 2). Historical records of patient reoccurrence reveal likely patterns of disease spread of each sub-site in the head-and-neck (3-6) and advise physicians on the nodal levels which would benefit from prophylactic radiation coverage. While irradiated nodal levels and prescribed doses may vary between patients and physicians (7, 8) there is some consensus on which to base standardized nodal selection according to stage and disease sub-site (9, 10). An atlas of nodal levels on a contrast enhanced CT of one patients is

available on the RTOG website (https://www.rtog.org/CoreLab/ContouringAtlases/HNAtlases.aspx).

Once the intended nodal volumes for irradiation have been selected, they must be delineated. Contouring of these elective targets in the head-and-neck faces the same burdens as does the contouring of normal structures; contouring is time consuming and even with published guidelines and publically available atlases of nodal volumes there exists significant inter-physician variability(11-14). Further, target definition represents a major contributing factor to geometric inaccuracy in radiotherapy (15). To reduce cntour variability as well as save time and standardize treatment, methods for the automatic delineation of these anatomically based nodal volumes have been investigated and are promising techniques (16-19), though small patient cohorts and limited or patient specific nodal volume selections limit their applicability more broadly.

We seek to provide a set of atlas based automatically contoured nodal volumes which, with post processing, can provide clinically usable clinical target volumes for elective neck coverage of many head-and-neck subsites and stages. In the context of an automated treatment planning system for the head-and-neck, we envision a process by which the attending physician is required to provide information about the location of gross disease (by identifying the head-and-neck subsite) and the location of gross nodal disease, if present. This information along with the physician drawn primary and nodal gross tumor volume (GTV) contours will be used to provide a preliminary clinical target volume, divided by dose level, which the physician will then have the opportunity for edit for treatment. This process can be seen in Figure 28.

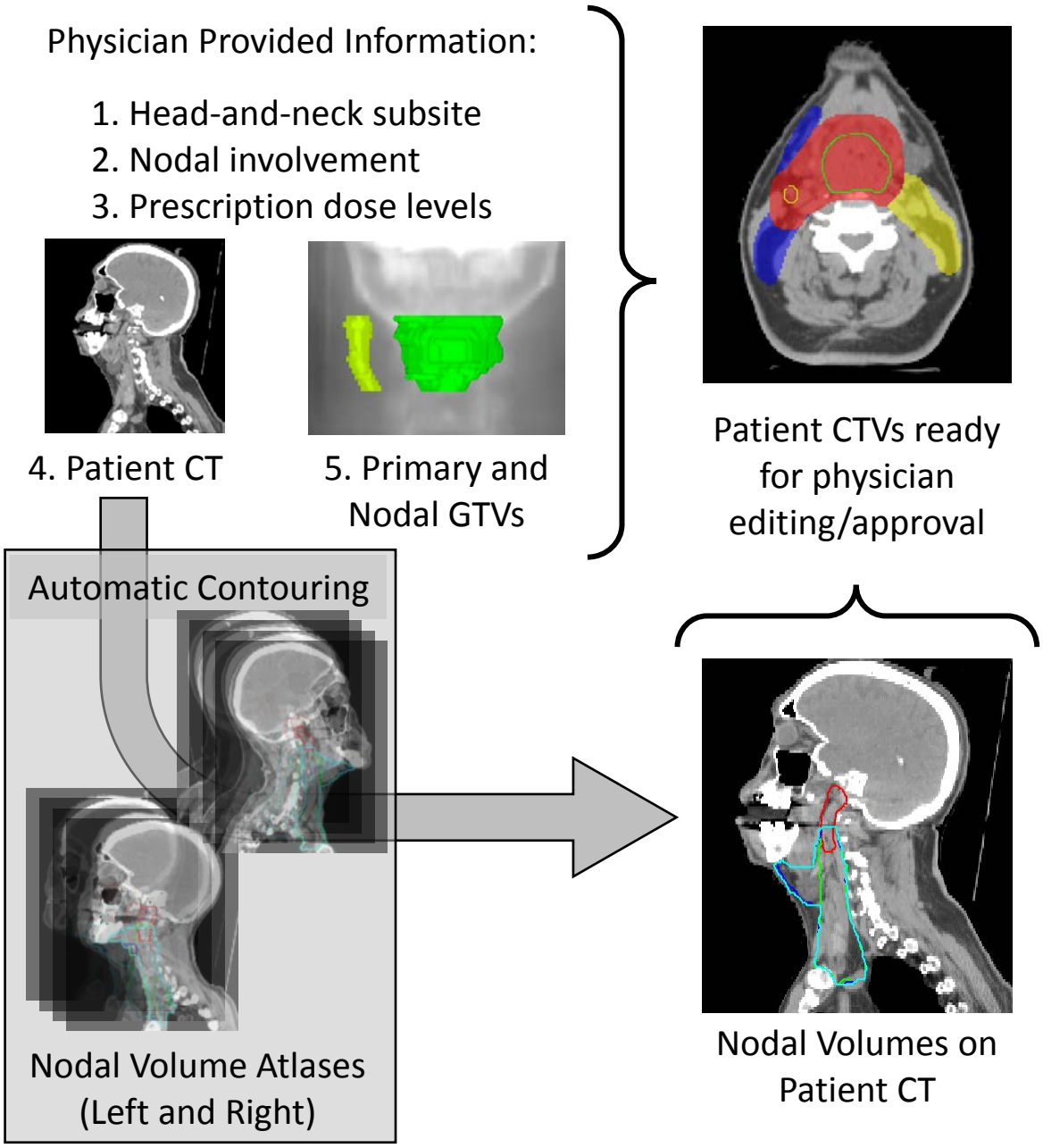


Figure 28. Workflow of the propagation of automatically generated CTVs. Using 5 user provided inputs, patient specific CTVs are created. The patient CT is used in a multi-atlas autocontouring algorithm and information provided about the disease type and stage is used to select the appropriate nodal volumes based on templates. From the patient GTV, the high dose CTV is created and then combined with the nodal volumes to provide patient specific CTV contours. In a contouring workspace the physician then has the opportunity to edit and the obligation to approve the contours for clinical use.

Methods

Autocontouring Algorithm

Autocontours of four anatomically defined nodal level groupings, as described in Table 22 and Figure 29, were created using the in-house multi-atlas deformable image registration approach known as MACS as described in Chapter 5. Two separate atlases were developed, one for left sided nodal volumes and one for right sided nodal volumes. The left and right sided atlases were comprised of 10 and 12 patients without gross disease on the corresponding side and who were previously treated patients at The University of Texas MD Anderson Cancer Center. Contours of the four target volumes were collected from the patients' treatment plans or drawn by a resident radiation oncologist; all contours used in the atlas were reviewed and approved for use by an attending radiation oncologist with 13 years of experience.

Retrospective Autocontouring Performance

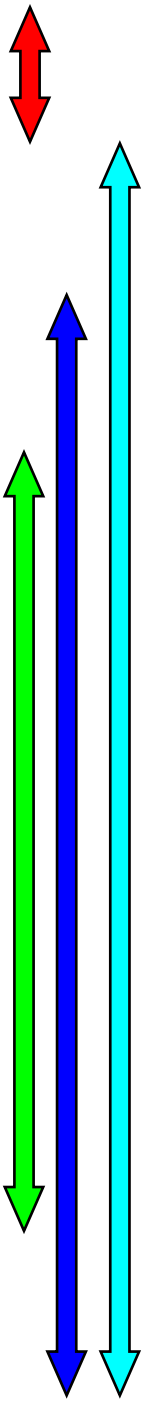
Physician Review

For this study, 115 patients were selected from the most recent 128 patients stored in the database of patients treated at our institution for head-and-neck cancer; the remaining 13 patients were used for atlas development. A radiation oncologist was asked to rate the nodal level contours as seen in Figure 29 on a five-point scale, *Table 8*, with a score of 1 indicating major edits are needed and a score of 5 indicating a perfect autocontour. To assess the possibility of rater bias, nodal level contours for 10 randomly selected patients were reviewed by five additional radiation oncologists from four international institutions. Physician agreement was assessed by grouping each pair of ratings (one rating from the primary physician and one from an outside physician) into one of three categories. Category I agreement includes instances when the primary and outside physician agreed as to the degree of edit needed, Category II agreement indicates that the physicians agreed that the

contour required either no more than minor edits or major edits. The final category, Category III agreement, includes those contours where the physicians disagreed on the acceptability of the contour, with one physician indicating that the contour needed major edit with the other indicating no or minor edit for use. Additionally, inter-physician variability in ratings was assessed using a Wilcoxon signed rank test.

Table 22. Anatomical definitions of individual nodal levels used in this study. Individual nodal levels were grouped allowing for four atlas derived volumes; the retropharyngeal nodes (red arrow), nodal levels II-IV (green), levels Ib-V (dark blue), and levels Ia-V (light blue). This table was adapted from published guidelines, Gregoire et al(9)

	Anatomical Boundaries					
Nodal Level	**Cranial**	**Caudal**	**Anterior**	**Posterior**	**Lateral**	**Medial**
Retropha-ryngeal Nodes	Base of skull	Cranial edge of the body of hyoid bone	Fascia under the pharyngeal mucosa	Prevertebral m. (longus colli, longus capitis)	Medial edge of the internal carotid artery	Midline
Ia	Geniohyoid m., plane tangent to basilar edge of mandible	Plane tangent to body of hyoid bone	Symphysis menti, platysma m.	Body of hyoid bone	Medial edge of ant. belly of digastric m.	n.a.
Ib	Mylohyoid m., cranial edge of submandibular gland	Plane through central part of hyoid bone	Symphysis menti, platysma m.	Posterior edge of submandibular gland	Basilar edge/inner side of mandible, platysma m., skin	Lateral edge of ant. belly of digastric m.
IIa	Caudal edge of lateral process of C1	Caudal edge of the body of hyoid bone	Post. edge of sub-mandibular gland; ant. edge of int. carotid artery; post. edge of post. belly of digastric m.	Post. border of int. jugular vein	Medial edge of sternocleidomastoid	Medial edge of int. carotid artery, paraspinal (levator scapulae) m.
IIb	Caudal edge of lateral process of C1	Caudal edge of the body of hyoid bone	Post. border of int. jugular vein	Post. border of the sternocleidomastoid m.	Medial edge of sternoclei-domastoid	Medial edge of int. carotid artery, paraspinal (levator scapulae) m.
III	Caudal edge of the body of hyoid bone	Caudal edge of cricoid cartilage	Postero-lateral edge of the sternohyoid m.; ant. edge of sternocleidomastoid m.	Post. edge of the sternocleidomastoid m.	Medial edge of sternoclei-domastoid	Int. edge of carotid artery, paraspinal (scalenius) m.
IV	Caudal edge of cricoid cartilage	2 cm cranial to sternoclavicular joint	Anteromedial edge of sternocleidomastoid m.	Post. edge of the sternocleidomastoid m.	Medial edge of sternocleidomastoid	Medial edge of internal carotid artery, paraspinal (scalenius) m.
V	Cranial edge of body of hyoid bone	CT slice encompassing the transverse cervical vessels	Post. edge of the sternocleidomastoid m.	Ant-lateral border of the trapezius m.	Platysma m., skin	Paraspinal (levator scapulae, splenius capitis) m.



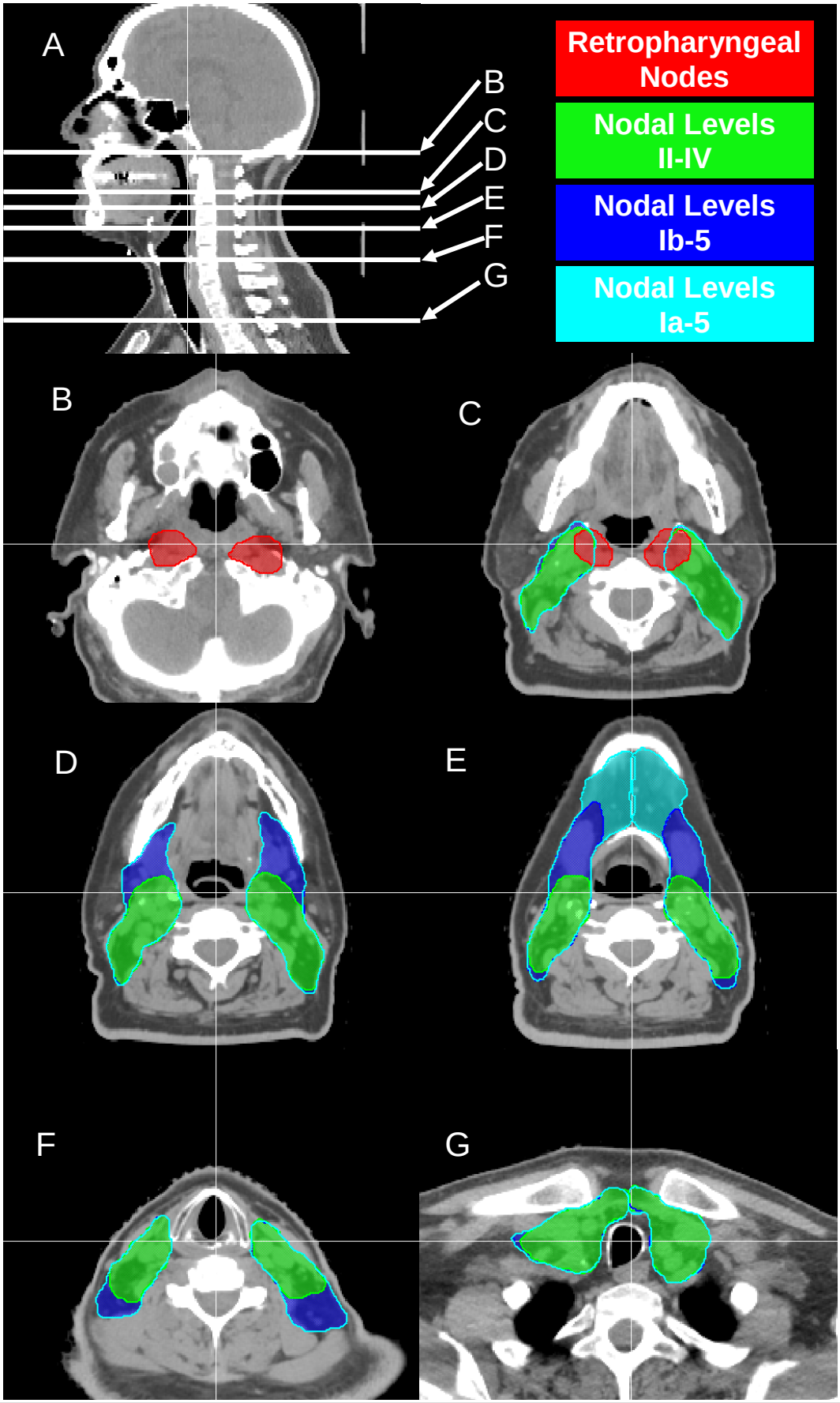

A
B
C
D
E
F
G
Retropharyngeal Nodes
Nodal Levels II-IV
Nodal Levels Ib-5
Nodal Levels Ia-5
B
C
D
E
F
G

Figure 29. Examples of automatically contoured nodal level group. Shown for a representative patient are the retropharyngeal nodes (red), nodal levels II-IV (green), levels Ib-V (dark blue), and levels Ia-V (light blue), on five slices (Panels B-G) at locations as seen in Panel A.

Quantitative Contour Performance

From the two patient cohorts as previously discussed, [1) 115 retrospectively collected patients with reviewed nodal levels and 2) 54 patients from various head-and-neck subsites and clinically edited normal tissues autocontours], inclusion criteria included clinical nodes which visually resembled anatomically defined nodal regions, and were treated for one of the 7 head-and-neck subsites. This left 55 of the 115 retrospective patients and 24 of the prospective patient cohort, for a total of 79 patients included in this analysis.

Physician drawn CTVs served as the gold standard in this study and were created by combining physician drawn CTVs of all dose levels into a single “clinical CTV”. This was necessary as often the division of clinical target volumes among many prescription doses can vary dramatically among physicians and were not standardized according to our automated planning approach. Atlas derived CTVs were generated by a visual matching of the intermediate and low dose clinical CTV with one or more of the four autocontoured nodal level groups, Figure 29, or one of three additional nodal level groups (including nodal level groups Ia-IV, Ib-IV, or II-V), which were not included in the autocontoured structure set and were only used for 2, 2, and 4, patients respectively. The selected nodal levels were combined with the clinical high dose CTV which was copied directly from the clinical treatment plan to mimic the anticipated automated planning approach to create the “atlas derived CTVs”. Two quantitative comparisons of the clinical CTVs and atlas derived CTVs were assessed using the Dice similarity coefficient, mean surface distance, and Hausdorff distance, as previously described. First, the clinical CTV and atlas derived CTV were compared directly, as seen in Figure 30. Second, to eliminate the bias introduced by including the physician drawn high dose CTV in

the atlas derived CTV, the physician drawn high dose CTV was removed from both the clinical CTV and the atlas derived CTV to create atlas derived and clinical “nodal CTVs”, see Figure 31.

The results of this experiment can be found on Page 156.

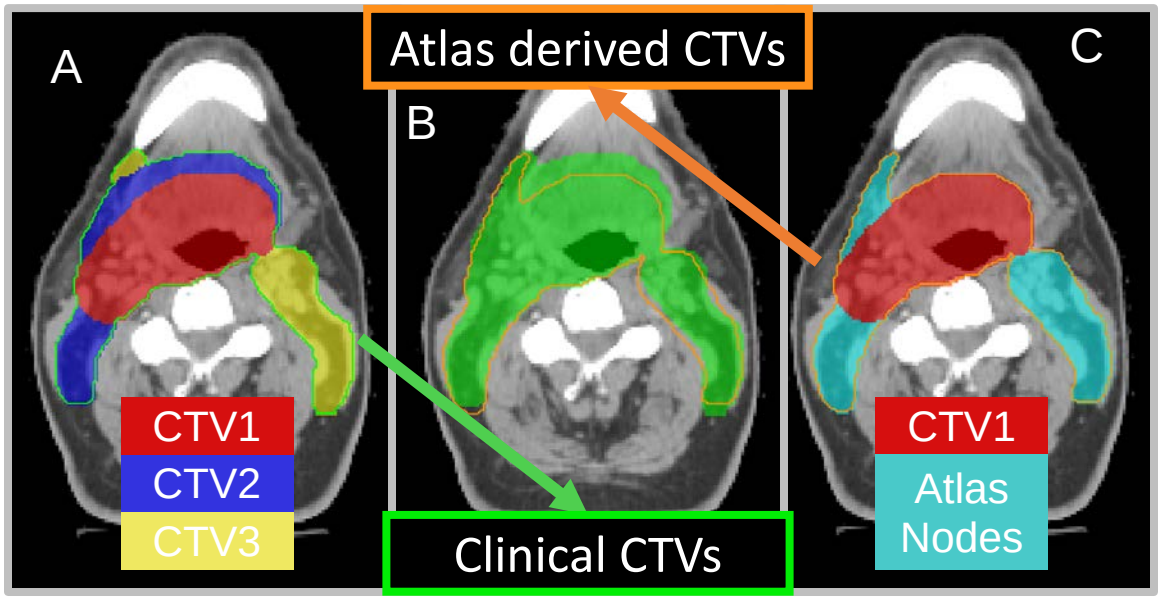


Figure 30. Quantitative comparison of clinical and atlas derived CTVs. In panel A, all clinical CTVs (here three dose levels indicated in red blue and yellow) were combined to create a single clinical CTV (green). IN panel C, atlas derived nodal volumes (light blue) were combined with the physician drawn high dose CTV (red) to create atlas derived CTVs (orange). In panel C, these two volumes were compared using the DSC, MSD, and HD.

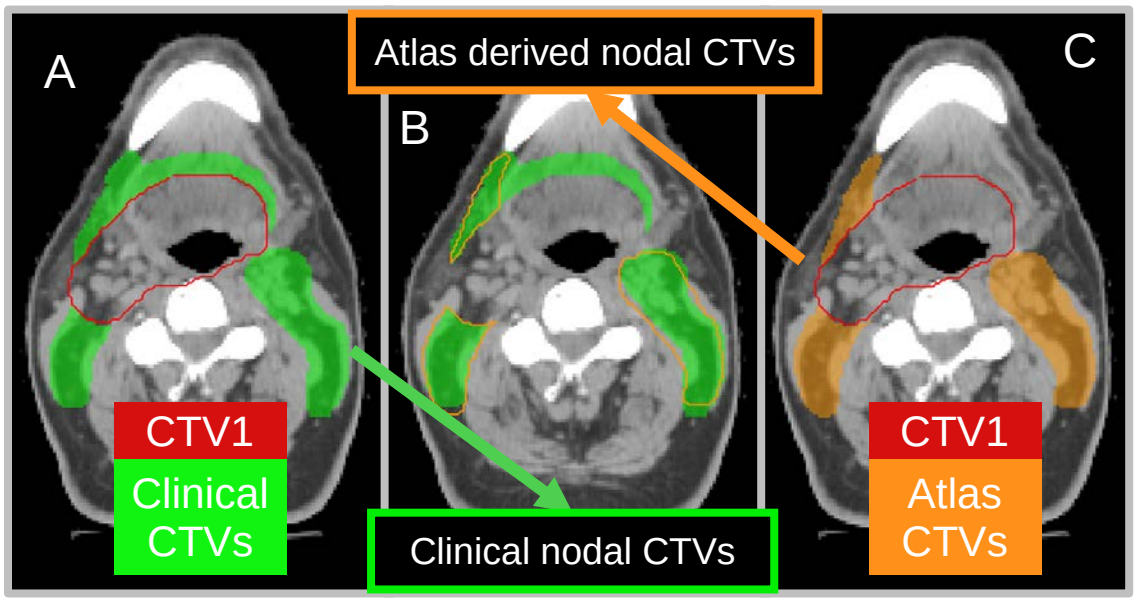


Figure 31. Quantitative comparison of clinical and atlas derived nodal CTVs. The high dose physician drawn CTV was removed from the clinical CTVs (panel A) and the atlas derived CTVs (panel C). The two volumes were compared the DSC, MSD, and HD.

Dosimetric Impact of Planning on Autocontoured Target Volumes

For this study, of the 54 patients used for evaluation of the treatment planning strategy and impact of normal tissue autocontours, 40 were selected because they met the inclusion criteria of at least 2 physician drawn CTVs and or PTVs, bilateral treatment, and intermediate and/or low dose CTVs based on anatomical nodal levels. For comparison, two treatment plans were developed. In the first strategy, plans were optimized to deliver dose to two physician drawn planning target volumes. Clinical PTVs were reduced to two such that the high dose physician drawn target was copied form the original treatment plan and any lower dose PTV was combined into one target volume. For planning purposes the prescription to this volume was chosen to be either equal to the original volume (if there was only an intermediate dose target) or between the two remaining target volumes (if there were both intermediate and low dose target volumes). See Figure 32 for an example of the reduction of

target volumes from three to two dose levels for two patients. In the second strategy, plans were optimized to deliver dose the atlas derived PTVs which were generated from atlas derived CTVs with either 3 or 5 millimeter margin. The atlas derived high dose CTV was a copy of the physician drawn high dose CTV. The second atlas derived CTV was derived by combining the selected nodal levels which best matched the physician drawn nodes and subtracting the high dose CTV, this volume was prescribed the same dose as the intermediate dose level in the first strategy. Atlas derived CTVs were expanded by a margin of either 3 mm or 5mm to create atlas derived PTVs. Both planning strategies used physician drawn normal tissues, when present, for optimization. If physician drawn normal structures were not included in the clinical plan MACS was used to generate these structures. For optimization, planning structures were also generated using the MACS software as needed for the planning strategy.

The dose delivered to the clinical CTVs and PTVs was evaluated when the two planning strategies were used. In the first strategy, the clinical PTVs were the targets for which the plan was optimized, the second strategy aimed to deliver dose a different target, the one that would be used if atlas derived nodal volumes were used as target volumes without edit. Coverage at the 95%, 98% and 100% isodose levels were evaluated.

The results of this experiment can be found on Page 161.

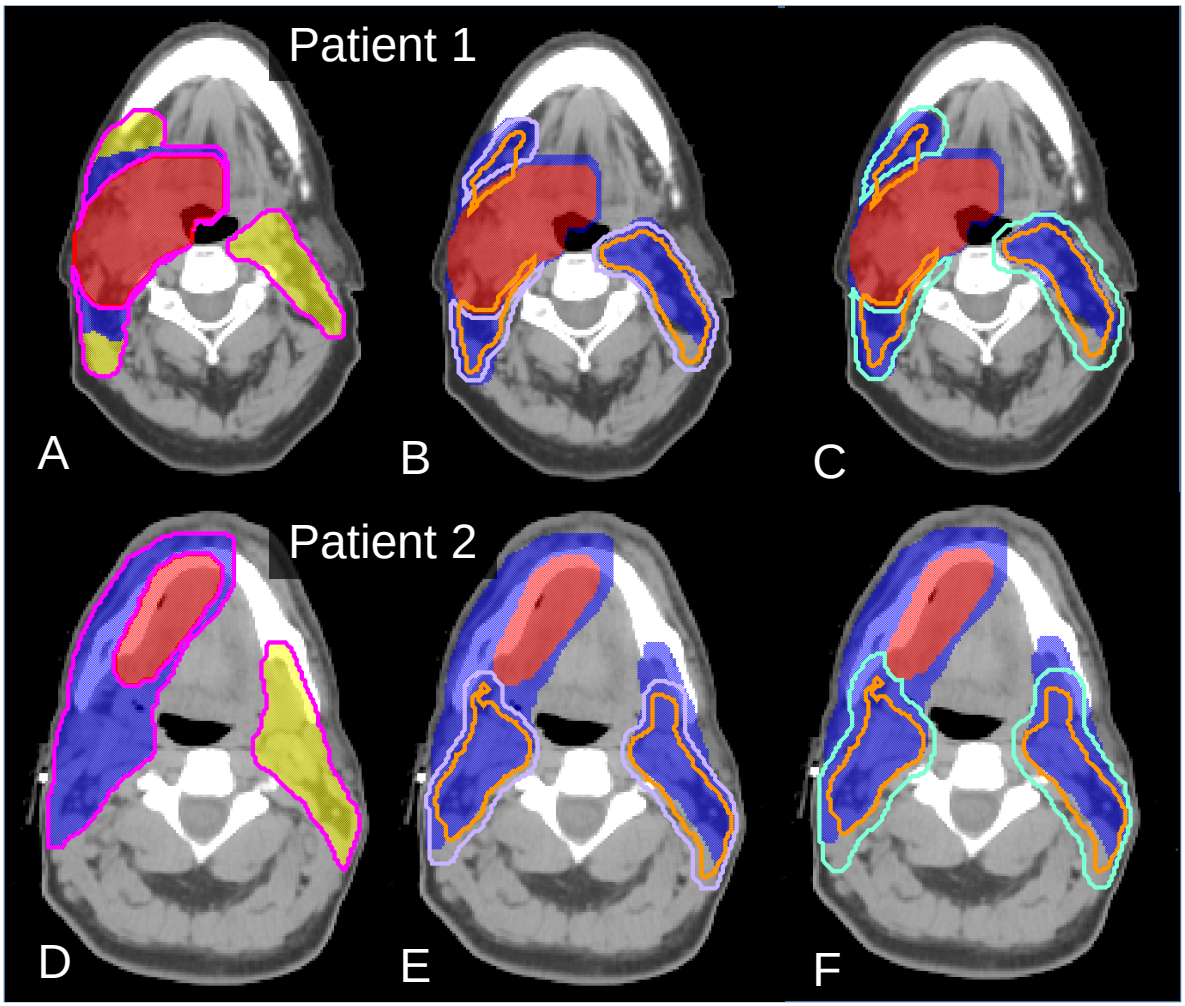


Figure 32.Clinical and atlas derived targets used for treatment planning. In panels A (patient 1) and C (patient 2) intermediate and low dose clinical PTVs (blue and yellow colorwash) were combined to create a single secondary PTV (pink line). The two physician derived PTVs were used for treatment planning and represent the true planning scenario. Atlas derived CTVs (orange lines panel B,C,E, and F) were used with 3 mm (lavender line, panels B and E) and 5 mm (blue line, panels C and F) for comparison.

Evaluation of an Independent Technique for QA of autocontoured Target Volumes

If automatically generated targets are to be used for treatment planning, similar to the strategy proposed for normal structures, it is necessary that automatically generated target volumes be verified independently. Towards this, we assessed the use of a second independent target volume contouring approach as a QA check of atlas derived CTVs.

Second Clinical Target Volume Contouring Approach

As developed by a member of our group, a machine learning approach has been shown to be successful in the contouring of a combined target volume structure in the head-and-neck (CARDENAS – In Submission – Segmentation of Oropharyngeal Clinical Target Volumes using a two-channel 3D U-Net Architecture MICCAI). This approach uses a 3D variant of a two channel U-Net architecture and requires the user to input the patient CT volume, the physician drawn gross tumor volume(s), an external body contour, and two anatomical landmarks, as seen in Figure 33. The anatomical landmarks are used to identify the extent of the area of interest in the machine learning algorithm and include the fusion of sphenoid bone and basilar part of the occipital bone and most cranial extent of the sternum; the landmarks were manually identified for this study but their identification could be easily automated. The model was developed using 210 head-and-neck cancer patients treated for oropharynx disease with bilateral treatment. Assessed on 85 patients not included in the training set, this machine learning approach agreed with independent drawn physician CTVs with an average and standard deviation in the Dice similarity coefficient of 0.78±0.05. In this study, to assess the quality of the CTVs generated using this approach, CTVs for 79 patients were compared with both the clinical CTV and the clinical nodal CTV as previously described and as shown in *Figure 34*

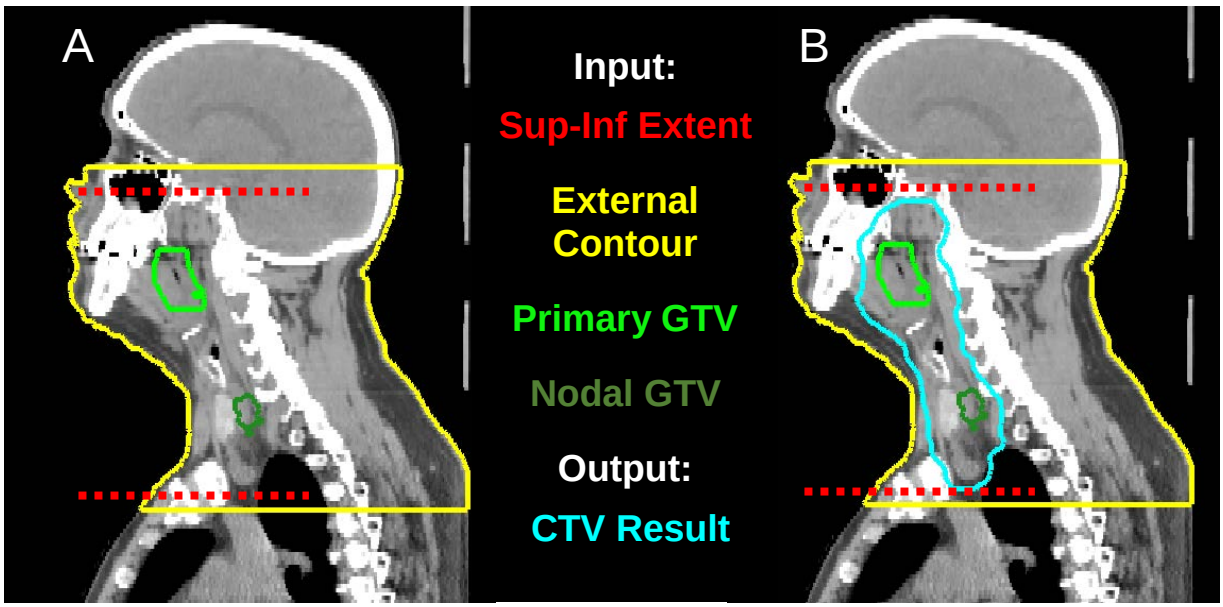


Figure 33. Secondary technique for the generation of CTVs. This approach requires as input (panel A) the primary and nodal GTV, an external contour in the region of interest, and anatomical landmarks which define the superior and inferior region of interest. The technique outputs a combined CTV volume (panel B).

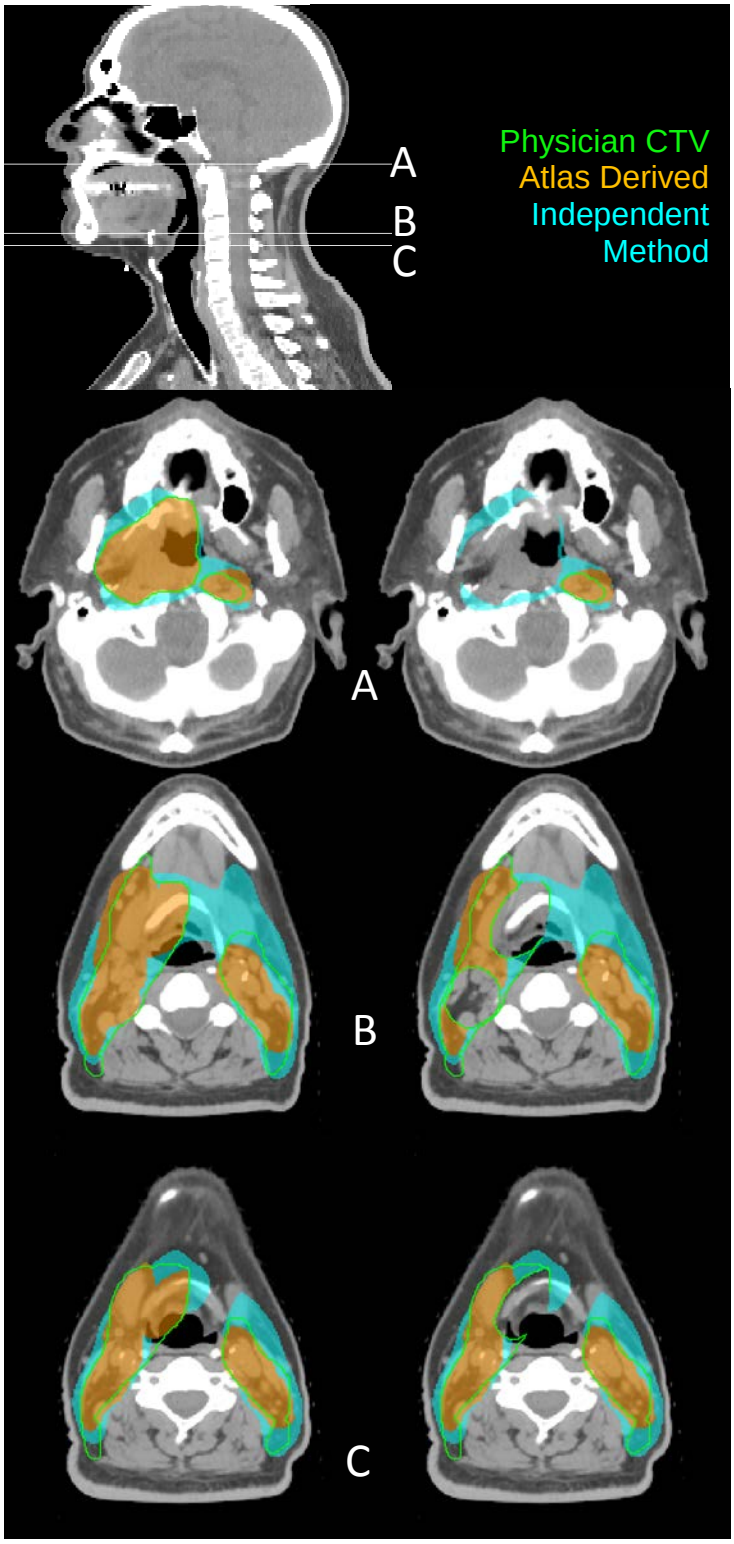


Figure 34. Quantitative comparison of CTVs from the secondary technique to clinical and atlas derived CTVs. Shown for three slices are the CTVs from the independent technique which were compared to both the clinical CTVs (green) and atlas derived CTVs (blue) for both the whole CTV volume (left panels) and the nodal CTV only (right panels).

Independent Clinical Target Volumes as a QA Approach

In order to assess the use of these secondary contours for QA purposes we evaluated the correlation between the quantitative agreement as measured by the DSC, MSD, and HD between 1) the atlas derived CTVs and the CTVs generated using the independent machine learning technique and 2) the atlas derived CTVs and the physician drawn CTVs. Correlation was established using the correlation coefficient with a p-value less than 0.05 establishing significant correlation. If correlated, the agreement of the contours generated for the automated planning approach (the atlas derived CTVs) could be compared with the secondary technique and this result provided to the user of the system to guide the need for edits.

The results of this experiment can be found on Page 163.

Results

Retrospective Autocontouring Performance

Physician Review

For 115 patients with atlas derived nodal volumes, the distribution of physician ratings of the 4 nodal volumes on the 5-point scale according to **Error! Reference source not found.** can be seen in Figure 35. The average rating of the four nodal volumes was between 3.5 and 3.9, and 75% of contours [686/920 contours (4 volumes bilaterally on 115 patients)] were rated as needing no edit for clinical use. One quarter of the nodal contours (230/920) were rated as needing minor edit and only 0.4% (4/920) received a score of a 2 and were indicated as needing major edit for use in treatment planning.

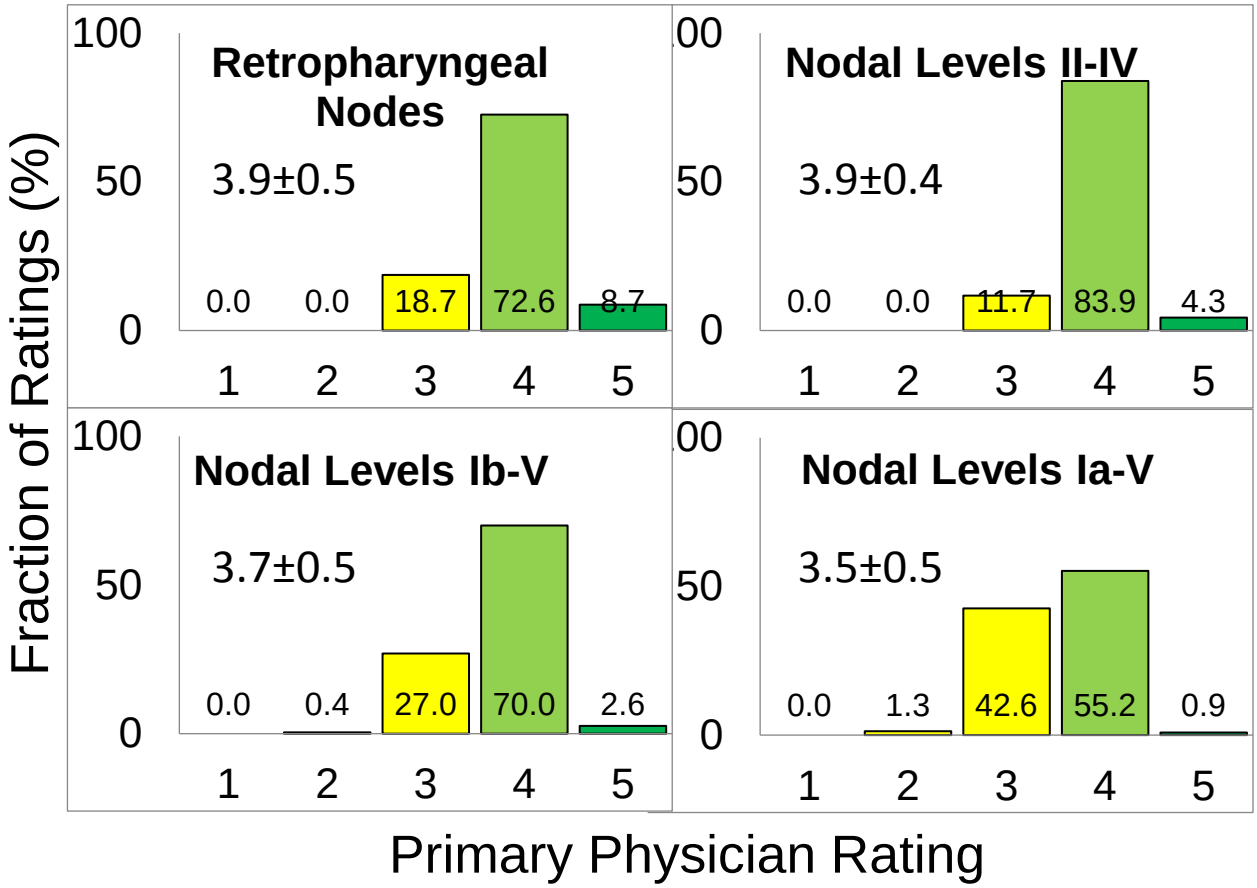


Figure 35. Distribution of physician scores of the atlas based nodal volumes. A score of 4 or 5 indicates no edits are needed for use in treatment planning.

To assess inter-observer variability a subset of 10 randomly selected patients was selected for review by five additional radiation oncologists from four outside institutions. The radiation oncologists, per a self-reported questionnaire, had an average of 8.25 years of experience (range, 3.0-12.5 years).

For all nodal volumes the scores assigned to the 10 patients differed significantly (Wilcoxon signed rank test, $p<0.05$) between the primary physician and at least one outside physician. However, no volume was significantly different between primary and all of the outside physicians. The average score across the subset of 10 patients for each of the four nodal volumes and each of the 5 physicians can be seen in Table 23. Some systematic

differences in contour ratings are seen; for example the third outside physician rated on average, all volumes lower than the primary physician. The fifth outside physician, however, rated, on average, all volume as high as or higher than the primary reviewer. This supports the known presence of inter-physician variability in contouring studies.

Of 399 total ratings by outside physicians (4 bilateral structures on 10 patients by 5 outside physicians with one structure not rated by one physician) 49% (195/399) of the ratings matched those of the primary physician and were classified as Category I agreements. Considering Category II agreements, the physicians assigned an additional 47% (186/399) of the contours to same group, either as needing no or minor edit for use (47%) or as needing major edit for use (0%). Finally, only 5% of contours received scores indicating the need for major edits by one physician while needing no or minor edits by the other physician. In Table *24* the percentage of contours classified into each of the three agreement categories can be found for the eight normal structures assessed.

Table 23. Average scores of the four nodal volumes by the primary and 5 outside physicians. Three physicians rated at least half of the contours, on average, higher than the primary physician. Two physician rated all contours, on average, worse than did the primary physician.

	Average Contour Rating					
Nodal Volume	Primary Physician	Outside Physician 1	Outside Physician 2	Outside Physician 3	Outside Physician 4	Outside Physician 5
Nodal Levels Ia-V	3.7	3.0	4.0	3.1	4.0	4.4
Nodal levels Ib-5	3.9	3.1	4.0	3.1	4.0	4.4
Nodal Levels II-IV	4.0	3.1	3.9	3.1	4.0	4.5
Retropharyngeal Nodes	4.0	2.7	3.7	3.7	2.9	4.0

Table 24. Inter-physician agreement in the rating of automatically contoured nodal levels. Category I indicates the scores by the two physicians matched, Category II indicates that the ratings did not match but the contours were rated into the same “group” (either as needing no or minor edit, or as needing major edit)) by both physicians. Category II represents disagreement between the two physicians. Due to rounding, not all rows sum to 100.

	% of agreements in Category		
Structure	I	II	III
Retropharyngeal Nodes	39	55	6
Nodal Levels II-IV	55	41	4
Nodal Levels Ib-V	54	42	4
Nodal Levels Ia-V	47	48	4
Total	49	47	5

Quantitative Contour Performance

Compared to clinical CTVs, atlas derived CTVs had an average (± standard deviation) DSC of 0.81±0.05. When considering only the nodal part of the CTVs (i.e. with the high dose CTV removed) the dice similarity coefficient decreaseD to 0.63±0.10. The results of both analyses are found in Table 25 and the distribution of value for all 79 patients are found in Figure 36. The analysis without the high dose CTV better represents the quality of the atlas contours in mimicking the clinical CTV contours, but the whole CTV analysis more closely represents the true treatment scenario in the RPA whereby the final treatment volume would include the physician drawn/edited high dose CTV. The results should be taken with some consideration given that chosen atlas derived nodal volumes may not have accurately reflected the attending physician’s intention when contouring the targets for each patient. Generally, the atlas contours agree well with physician drawn contours compared to previously published whole CTV agreement metrics (11, 19, 20).

Table 25. Quantitative agreement between atlas derived CTVs and clinical CTVs. Comparisons were made for both the whole CTV and the nodal CTV only, for which the high dose CTV was removed.

	Dice Similarity Coefficient	Hausdorff Distance (mm)	Mean Surface Distance (mm)
	Mean ± std [min - max]	Mean ± std [min - max]	Mean ± std [min - max]
Nodal CTVs	0.63 ± 0.10 [0.34 - 0.85]	4.52 ± 1.56 [1.57 - 10.33]	0.55 ± 0.27 [0.04 - 1.53]
Whole CTVs	0.81 ± 0.05 [0.69 - 0.91]	2.33 ± 0.82 [1.06 - 4.80]	0.30 ± 0.07 [0.14 - 0.46]

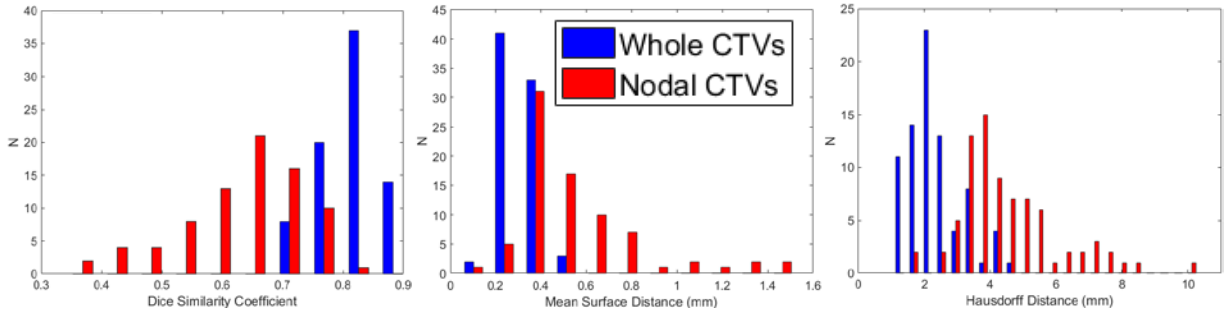


Figure 36. Distribution of the quantitative agreement between clinical and atlas derived CTVs. Shown are the distributions of the Dice similarity coefficient (left) the mean surface distance (middle) and the Hausdorff distance (right) for both the whole CTVs (in blue) and the nodal CTVs only (red).Whole CTVs has better agreement due to the use of the original high dose CTV from the clinical plan in the autoplan.

Dosimetric Impact of Planning on Autocontoured Target Volumes

Coverage of the clinical CTVs and PTVs was poorer if atlas derived PTVs were used for treatment planning (i.e. optimization and normalization) than if the clinical PTVs were used for planning. Coverage was improved, however, if 5mm margins were used when compared to 3mm margins for the expansion of atlas derived CTVs to PTVs. RTOG trial 1016, in which the investigators studied the benefit of adding a chemotherapy drug cetuximab to a radiotherapy treatment strategy for patients with oropharynx cancer, required that for the intermediate dose target, prescribed 56Gy in the trial, volume should receive at least 45Gy (80% of the prescribed dose) to 95% of the target volume(21). Considering this threshold of 80% of the prescribed dose to 95% volume, if 3 and 5mm margins are used 32% and 54% of the 40 treatment plans considered here meet that constraint, respectively. The average and standard deviation of the volume of both the clinical CTVs and clinical PTVs receiving 95%, 98% and 100% of the prescribed dose for each of the two margin values are shown in Table 26. Coverage of clinical PTVs was significantly poorer as measured using a Wilcoxon sign-rank test when comparing the volume receiving 95%, 98%, and 100% ($p<0.001$) of the prescribed dose when 3mm margins were used as compared to 5mm margins. In Figure 37 are the average DVH curves to the high dose and intermediate dose CTVs and PTVs given the two margins While the curves are similar for doses above 100% of the prescribed dose, the DVH curve of the true physician drawn intermediate dose PTV has a noticeably rounder shoulder when using either a 3 or 5 mm margin as compared to when the true structure was used for planning.

Table 26. Average coverage of the clinical CTVs and PTVs when atlas derived PTVS were used for planning.

		3 mm Margin		5 mm Margin	
		mean±std	[min-max]	mean±std	[min-max]
PTV	V95%(%)	81.1±13.4	[45.7-98.9]	86.9±11.1	[56.4-99.8]
	V98%(%)	78.0±14.0	[41.7-98.1]	84.7±12.0	[52.1-99.5]
	V100%(%)	74.9±14.1	[38.9-96.7]	82.3±12.5	[49.3-98.5]
CTV	V95%(%)	86.6±10.8	[58.4-99.2]	91.3±8.4	[67.1-99.8]
	V98%(%)	84.1±11.8	[56.0-98.4]	89.6±9.3	[64.8-99.5]
	V100%(%)	81.7±12.3	[53.7-97.1]	87.9±9.9	[63.0-99.1]

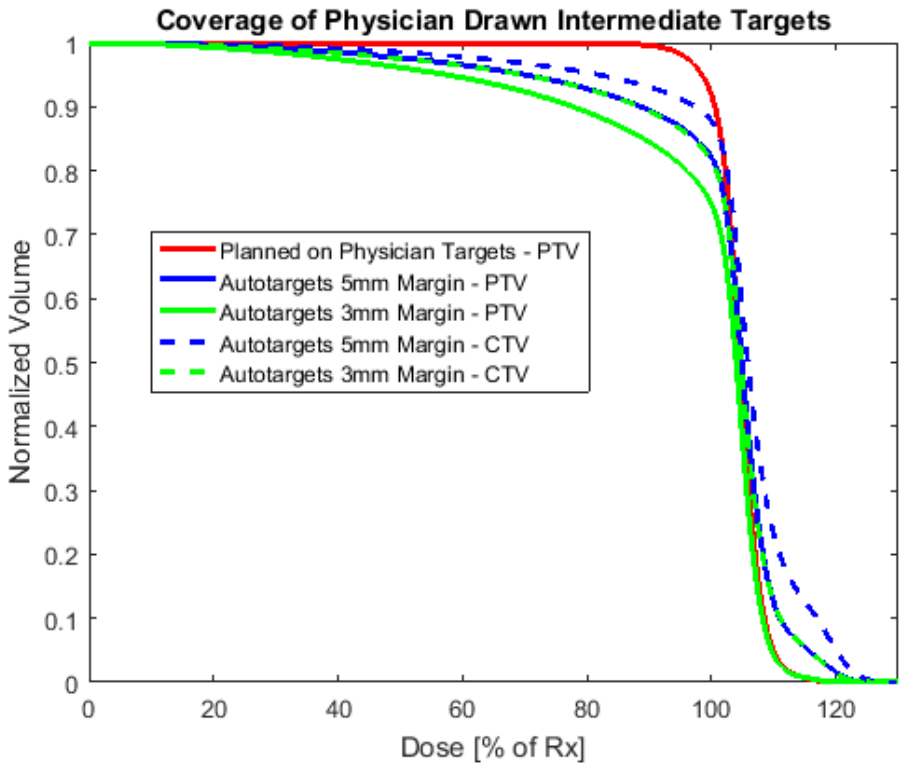


Figure 37. Average DVH curves to clinical PTVS when using atlas derived PTVs for planning. Atlas derived CTVs with 3 mm (green line) and 5 mm (blue line) were used for expansion to the PTVs. Coverage of clinical PTVs is noticeably poorer when clinical PTVs are not used for treatment planning.

Evaluation of an Independent Technique for QA of Autocontoured Target Volumes

Quality of the Second Independent Target Contouring Technique

The second independent target contouring technique performed well in the contouring of CTVs for 79 patients, see Table 27. The contours agreed with a mean (± standard deviation) Dice similarity coefficient of 0.80±0.07 when considering the whole target volume and with an average of 0.66±0.11 when considering only the nodal CTV. Using a Wilcoxon signed rank text, the atlas based technique had a significantly better DSC ($p=0.02$), MSD ($p<0.01$), and HD ($p<0.01$) when considering the whole target volume.

Table 27. Quantitative agreement of independently derived CTVs with clinical CTVs. Whole and nodal only CTVs were compared using the Dice similarity coefficient, mean surface distance, and Hausdorff distance.

	Dice Similarity Coefficient	Hausdorff Distance (mm)	Mean Surface Distance (mm)
	Mean ± std [min - max]	Mean ± std [min - max]	Mean ± std [min - max]
Nodal CTVs	0.66 ± 0.11 [0.35 - 0.89]	3.07 ± 1.25 [0.75 - 6.37]	0.35 ± 0.16 [0.09 - 0.83]
Whole CTVs	0.80 ± 0.07 [0.64 - 0.94]	2.64 ± 1.10 [1.01 - 6.37]	0.35 ± 0.11 [0.11 - 0.63]

Independent Clinical target Volumes as a QA Approach

In order for the independent target contours generated using a machine learning technique to be useful as a QA check of the atlas based targets we first sought to establish if the agreement between the two contours are correlated to the agreement of the atlas based targets to the physician target. In Table 28 are the correlation coefficient and corresponding p values for 3 quantitative metrics between the atlas based and physician targets and the atlas based and independent targets. Significant correlation was established for all metrics indicating that agreement between the two contouring techniques is correlated to the ultimate agreement of the atlas based technique to the physician technique. The distribution of values can be seen in *Figure 38*.

Table 28. Correlation of the agreement between the atlas derived CTVs with the clinical CTVs and the independent CTVs. Correlation coefficients and p-values are shown for the agreement between both the whole volume CTVs and the nodal CTV only for three quantitative metrics. All metrics were significant, with the strongest correlation between for the nodal CTVs using the mean surface distance.

	Correlation of Atlas and Physician and Atlas and Independent Targets					
	Whole CTV			Nodal CTV		
	Dice	HD	MSD	Dice	HD	MSD
Correlation Coefficient	0.44	0.21	0.27	0.60	0.78	0.83
p-value	6.0E-05	6.8E-02	1.6E-02	5.7E-09	3.1E-17	1.3E-21

The positive correlation coefficients indicate that the larger the disagreement between the two contouring techniques the larger the disagreement between the atlas-based technique and the physician CTVs is expected to be. This correlation was stronger for the nodal CTVs which is expected as the subtraction of the high dose CTV removes the biased introduced because the atlas based technique used this contour directly while the independent technique does not. As seen in *Figure 38*, the identification of thresholds in the agreement of the two contouring techniques may be selected in order to identify potential disagreement with the physician drawn CTVs. For example, a threshold of 6 mm Hausdorff distance between the two contouring techniques would identify all atlas based contours with a Hausdorff distance of greater than 8mm and 82% of contours with HD greater than 6mm compared to the physician drawn CTVs. These threshold represent this data set only and should be further evaluated. The significant correlation supports the use of this second independent technique for use a QA check of atlas based CTVs when , for example in Figure 39, when the atlas based CTVs do not agree well with clinical contours nor do they agree with CTVs from the independent technique and which doesn't agree well with the clinical CTVs either.

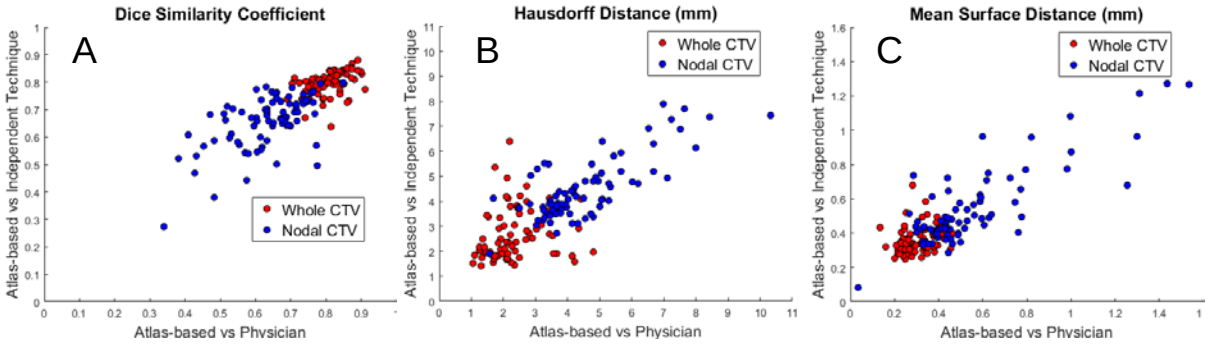


Figure 38. Distribution of the CTV quantitative agreement. Shown are the distributions of the dice similarity coefficient (left), mean surface distance (middle) and the Hausdorff distance (right) between the whole volume CTVs (red) and nodal CTVs (blue) between the atlas derived CTVs with both the clinical CTVs (x-axis) and the independent technique (y-axis). Clear trends are seen with the whole volume CTVs having better agreement.

.

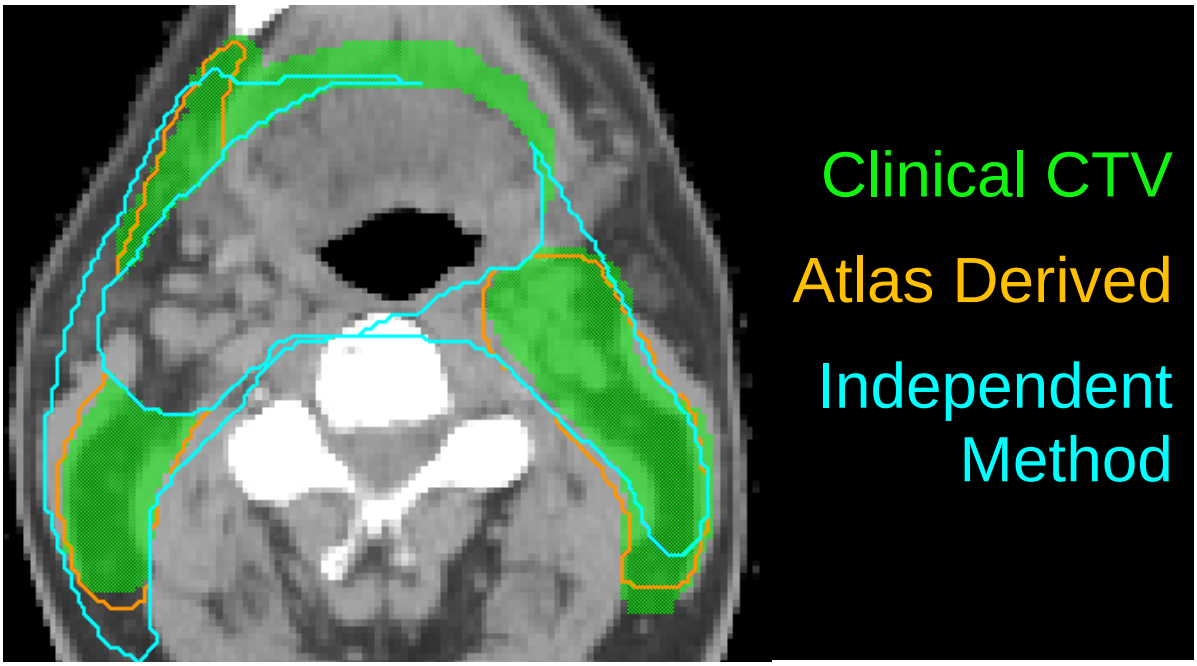


Figure 39. Example of disagreement between atlas derived CTVs, clinical CTVs, and CTVs from the independent technique. This indicates that use of the independent contouring of the CTVs may be useful as a QA tool in cases, like the one here, where all three contours disagree

Discussion

In this work we have investigated the use of atlas based automatically contoured nodal levels for the definition of intermediate and low dose target volumes in the head-and-neck. Upon physician review we found that the four nodal volumes considered were well received by head-and-neck radiation oncologists with nearly half reviewed as needing no edit. In an inter-physician analysis we found that there exists inter-physician variability in the acceptability of these contours, though for only between 0.4% (as rated by the primary physician) and 4% (as rated by outside physicians) would be expected to need major edits for use in treatment planning.

Compared to clinical CTVs, we found good agreement with an average Dice value of 0.81, which is comparable to the agreement found in other studies (11, 19, 20). These results, however, should be taken with the understanding that the physician drawn CTVs used for comparison were copied from clinical plans directly, without consideration of the intent of the physician when drawing the target. There are known exceptions to the general guidelines for the treatment of elective nodal volumes in head-and-neck cancer and it is likely that for some of the plans considered here the choice of atlas-derived nodal CTVs simply did not match with what the physician delineated. Given this, this high DSC indicates that the atlas based technique for the delineation of intermediate and low dose target volumes in the head-and-neck performed very well.

Further, we sought to assess what impact planning on these atlas derived targets would have on the coverage of the clinical targets. As expected, if the target of interest (in this case, the clinical target) is not used for treatment plan optimization, then coverage suffers. We investigated the use of both 3 and 5 mm margins for CTV to PTV expansion in order to, in part, compensate for possible contouring errors. The CTV to PTV margin as used to compensate for uncertainties in CTV contouring on the initial treatment scan is often not cited

as a primary use, but, as it is to compensate for geometric uncertainties(22), if contouring uncertainties are expected it may be wise to err on the side of safety and use a larger margin. When a 5 mm CTV to PTV margin was used for atlas-derived intermediate dose targets, slightly more than half of plans met the RTOG 1016 coverage constraint to the intermediate dose physician drawn PTV.

Finally, if automatically delineated target volumes are to be used for fully automated treatment planning, it is advisable to assess and, ideally, ensure contour quality before presentation to the user. Towards this, we investigated the use of an independent technique for the delineation of target volumes and the agreement between the atlas-based targets and this secondary technique to the agreement between the atlas-based targets and the physician targets. The two were found to be strongly correlated indicating that in a fully automated system, both techniques could be used to delineate the CTV contours, their agreement assessed and this result provided to the user as guidance for the need for editing. However, as before, the motivation behind the delineation of the physician drawn targets is not known and this may influence the results found here. There are a few clear examples of disagreement between all three contours, which would indicate promise for the use of this independent technique as a QA tool, however further investigation into the usefulness of this strategy should be conducted.

There are general limitations of this analysis and potential utility of automatically delineated targets more broadly. First, given the prominence of extensive inter-physician variability in the delineation of target volumes it may be ambitious to expect very high agreement between any automatic contouring technique and single physician drawn volumes, especially, as in this case, if (1) the physician intent in target delineation is unknown or ambiguous and (2) the atlas derived volumes originated from contours approved by a single physician. A detailed record of physician intent in the delineation of CTVs and consensus contours which match that intent would reveal the true ability of automated techniques for

CTV contouring. Further, for multi-atlas autocontouring techniques, the presentation of the atlas to multiple head-and-neck physicians or the development of a consensus atlas may further reduce inter-physician variability in the reviews of contours propagated from such an atlas and, ultimately, edits to such contours. Second, in the current strategy nodal levels are limited to the four discussed here, though upon retrospective assessment approximately 10% of patients required nodal levels not included in the currently available set of nodal volumes. The need for additional volumes in a larger cohort of patients should be evaluated and, if needed, the additional nodal level groups could be added to the atlas patients, but must be accompanied by comprehensive analysis. Third, while the current autoplanning strategy is limited to the use of three dose volumes, and the autocontouring supports this with the intermediate dose level assigned to the ipsilateral neck and the lower dose assigned to the contralateral neck, there exist examples where a fourth dose level may be required due to the location of the disease, including when the disease is close to the temporal lobe or brachial plexus. This division of the targets into multiple dose levels is likely to vary depending on patient factors and physician experience. Therefore, a check of the target location should be performed and the user made aware if an additional dose level is likely to be required. Fourth, the use of the presented secondary technique, while proven to be correlated with the agreement between the atlas-based targets and physician targets, needs adjustment in order to be used in a fully automated system. In order to generate the contours a method to detect the superior and inferior landmarks must be developed. Additionally, the current model was intended and tested exclusively for patients with bilateral disease of the oropharynx by the original authors and would require further development for use for many sub-sites in the head-and-neck which systematically differ in the selection of nodal levels. Further, while the correlation between atlas-physicians CTVs and atlas-secondary CTVs was established here, the true ability of this technique to detect and warn physicians of possible inaccuracies in atlas-based nodal level contours and or their use in CTV contours needs further development.

Finally, for use in an automated treatment planning strategy it is our recommendation that atlas based nodal levels be generated using the presented autocontouring technique for the initiation of CTVs but are used with physician approval and with the the allowance of possible editing. The volumes, both prior to and after possible edits should be compared with the secondary independent technique in order to further examine the use of this strategy for QA purposes. Only true clinical implementation will reveal the extent to which these contours are edited clinically and the extent to which QA is possible.

Conclusion

We have examined the use of automatically contoured nodal volumes for use as intermediate and low-dose CTVs in the treatment of head-and-neck cancer with radiotherapy. The contours were well received by head-and-neck radiation oncologists and agreed well with independently drawn physician contours. An independent target contouring method proved promising in the identification of contours requiring edits. We suggest the careful use of these automatically derived nodal volumes as CTVs in a fully automated treatment planning system with physician approval and editing as needed.

Chapter 7: Discussion

The goal of this work as a whole was to investigate two initiatives which may help alleviate the burden of radiotherapy in low resource settings. The dire need for radiation therapy in low-resource settings, including for 135 of 139 LMICs whose population have inadequate access to radiation therapy and for 55 countries in which no radiotherapy facilities exist (1), demands attention. A multifaceted approach is necessary and collaboration with professionals in these regions is central to success.

An Upright Radiotherapy Chair

The first method aimed to make more accessible the necessary high energy external beam radiotherapy machines required for safe radiotherapy delivery. The concept involves the use of a rotating patient treatment chair which, when paired with a fixed treatment beam offers the same degrees of freedom as traditional treatment machines.

In the first aim, we have investigated inter- and intra- fraction setup reproducibility of an upright radiotherapy chair with the hypothesis that *an upright radiotherapy chair has clinically acceptable inter- and intra- fraction reproducibility*.

The use of such a chair, combined with a fixed treatment beam could greatly reduce the upfront and ongoing cost of teletherapy machines and may be of interest in low-resource settings or LMICs. While treatments in this seated position are not new to the field, the use of treatment chairs has declined and are currently very rare due to the routine acquisition of treatment planning images from CT scanners which, except for in a few rare exceptions, only allow for horizontal or nearly horizontal acquisition. As technology improves and other methods for the acquisition of planning images emerge (2, 3) the possibility of other treatment positions is becoming a reality. If treatment chairs are to be reconsidered their acceptability in the context of current treatment delivery must be re-examined. Accurate setup

reproducibility is an important part of a robust treatment position. We found that the inter- and intra-fraction reproducibility in the treatment chair was, on average, less than 3mm and this is comparable to that in the traditional supine position. Given patient feedback, we manufactured a second treatment chair which incorporates additional patient indexing measures, increases the depth of the seat, and is built to limit the potential dosimetric impact of inter and intra- fraction motion.

Limitations of this study include the use of a small patient set, all of whom were relatively healthy and tolerated both the seated and supine positions well. Further, the assessment of set-up reproducibility under conditions of image guidance were simulated using pre-alignment of the images using select anatomy rather than true adjustment based on pre-treatment images as is clinical practice. The treatment chair was only evaluated for head-and-neck patients, due primarily to the tight geometry of the set-up using a traditional linear accelerator.

The ultimate goal of this work is to support the development of a fixed beam system, which would eliminate some of these challenges. Treatment of other sites in the seated position, for example the cervix or prostate, may bring other challenges due to a difference in patient anatomy in this seated positon. The current study, as it was limited to the head and neck, would be expected to have little impact of internal anatomy between the supine and upright positons. If a fixed beam-rotating patient system is to be developed, the impact of other aspects of the radiotherapy process must be evaluated, including treatment delivery, machine shielding, and patient throughput. Additionally, the compatibility of this treatment positon including the import of planning images acquired in this position with treatment planning systems and in concordance with DICOM standards must be evaluated. Finally, end-to end testing of the entire treatment process must be completed both at the start and at reoccurring intervals. Clinical implementation should be accompanied by appropriate clinical protocols, FMEA and root cause analysis studies.

Towards the development of a fixed-beam and rotating seated patient treatment paradigm which could greatly reduce the cost of the teletherapy machinery necessary for the delivery of both life-saving curative and symptom alleviating palliative radiotherapy in low-resource settings, we have shown herein that the set-up reproducibility of the treatment chair is comparable to that in the supine position. Members of our group have previously showed the capability of the onboard KV imagers to capture and reconstruct CBCT images (4), which as reported by other investigators, can be used for treatment planning. This technology, however, is not released for clinical use and represents the next step in the efforts to realize this treatment paradigm. Other investigators are moving forward with this type of treatment working towards commercialization of a fixed-beam upright system (LEO Cancer Care, http://leocancercare.com).

Automated Treatment Planning

Our second proposal for the improvement of radiotherapy in low-resource settings was that of a fully automated treatment planning process for head-and-neck cancers. Given the extreme shortage of and growing need for trained radiotherapy personnel including medical physicists in LMICs, methods and tools to reduce this burden are desperately needed. Technological improvements have the ability to both improve the quality of care while simultaneously reducing the required human involvement. However, the introduction of these advanced technologies has the risk of further dividing those who have the resources and infrastructure to encourage advancement and those who do not and therefore risk falling further behind. In the case of advanced treatment planning techniques, the introduction of IMRT and VMAT greatly improved outcomes for head-and-neck cancer patients, however the introduction of these advanced treatment techniques requires improved equipment (MLCs, availability of service personnel for complicated machinery, adequate dose rate to compensate for beam modulation, etc.) and staff capable of implementing the techniques

which include advanced treatment planning. Without the required infrastructure the gap in treatment techniques only increases.

The aim of the second aspect of this work was to partially reduce the burden on staff by automating the treatment planning process. Automated methods are proving beneficial in many areas of radiation therapy and the automation of the treatment planning process would be applicable in all clinics but may be especially useful in low-resource settings.

In Specific Aim 2, Chapter 4, we sought to develop and validate a treatment planning strategy for the head-and-neck with the hypothesis that *single optimization head and neck treatment plans perform with equal quality to clinically acceptable plans and 90% are accepted by radiation oncologists for use without edit.* We found that a fully automated approach can produce plans that perform as well as, and in some cases better than, those plans as treated on a clinical trial with XX% rated as acceptable without edit by a dedicated head-and-neck radiation oncologist. The automated approach utilizes the RapidPlan® tools available in the Eclipse Treatment Planning System (Varian Medical Systems, Palo Alto, CA) in conjunction with in-house algorithms and requires no user input outside of the patient CT and treatment details including prescription and fractionation. The results were based on 54 patients treated for cancers of various head-and-neck subsites at The University of Texas MD Anderson Cancer Center and 30 patients treated on a clinical trial.

In Specific Aim 3, Chapter 5, we sought to assess the feasibility of the use of automatically contoured normal structures in the head-and-neck in a fully automated treatment planning strategy with the hypothesis that *automatically contoured normal structures can safely be used for treatment planning purposes without significant impact on plan quality*. In an investigation of the use of four autocontouring algorithms we found that in-house multi-atlas contouring approach performed well and upon its clinical implementation half of contours were not edited for treatment planning. Further, we found that unedited

autocontours can be used without edit for treatment planning with very limited impact on the treatment plan. While our hypothesis was confirmed, the use of contours without oversight is far reaching and represents a significant change in practice and should be of further investigation. Therefore, we conclude that automatically contoured normal tissues should be reviewed by the radiation oncologist and approved for use in treatment planning though the initialization of such structures will greatly reduce the manual input required and increase the time savings offered of an automatic system.

Finally, in Specific Aim 4, Chapter 6, we sought to assess the feasibility of the use of automatically countered intermediate and low dose target volumes in the head and neck with the hypothesis that *automatically contoured clinical target volumes can be safely used for treatment planning purposes*. We investigated the use of automatic contours for the delineation of nodal levels which were then used as intermediate and low-risk clinical target volumes. We found that, generally, the automatically contoured nodal volumes, generated using the same in-house algorithm as the normal tissue autocontours, were well received by radiation oncologists from five institutions. Quantitatively, intermediate and low-dose CTVs agreed well with physician drawn targets. If automatic targets are used for treatment planning than the coverage of physician targets is reduced, the results however suffer from a fundamental difference in the intent of target delineation, for which the physician intent was unknown, and thus the analysis of these results is limited. Finally, we found significant correlation in the agreement between the atlas-based target contours with an independent machine learning contouring technique and the physician contours suggesting that this independent method may be a useful QA tool to aide users of the automated system in the review of automatically contoured targets. We feel that the use of fully automatically contoured intermediate and low-dose CTVs is not yet suitable for safe use and should be of further investigation. We concluded that CTVs generated from the multi-atlas based

automatically contoured nodal volumes can be used in treatment planning but should be accompanied by physician editing and approval as deemed necessary.

Together, through these three aims we sought to investigate the extent to which the development of head-and-neck treatment plans can be automated. We found, similar to other investigators, that knowledge based planning approaches can be utilized successfully for automatic treatment planning. Our group took automation further through the identification of algorithms for the determination of treatment isocenter, collimator angles, and jaw settings and through the use of the Varian API, the combination of all steps from the import of patient DICOM files through to dose calculation without human intervention. We further automated the process by automatically contouring normal tissues and anatomically based intermediate and low dose clinical target volumes in the head-and-neck. We have showed that through automation we can reduce the time needed for treatment planning to 40 minutes, most of which time does not require supervision. The automation of normal structures represents an important step in the treatment planning process and the algorithm described herein was successful in the contouring structures which can be used without edit for treatment planning. It is unlikely that clinicians are currently amenable to the use of unedited autocontours but it has been shown that the review and editing of autocontours represents a significant time savings. Together, the strategy is ready for limited clinical use and should be done so under close supervision and with extensive data collection as to its impact on clinical workflow and patient treatments.

The success of this work should be considered in the context of its limitations. First, we found that the planning approach outperformed clinical plans for a cohort of patient treated on a clinical trial, but, in general, did not perform as well as plans optimized by highly trained dosimetrists at our institution or for plans of all head-and-neck subsites. The availability of highly trained staff is limited to select clinics, however, and as the approach outperformed plans treated on a clinical trial, we believe the quality of plans is sufficient for patient

treatment. Improvement upon the clinical knowledge base used for treatment plan generation including the optimization of the templated constraints could improve resultant plan quality and should be of routine consideration. The development of multiple knowledge bases, for multiple treatment subsites, may be considered and plans for which the strategy did not outperform clinical plans may provide useful for the development of such a knowledge base.

Second, considering the automatic delineation of normal tissues, while a strategy was presented which allowed for use of unedited contours in treatment planning it is not standard practice to use unapproved contours for treatment planning and the consideration of such use should be carefully considered, both in the context of the treatment of the individual patient and if questions are to be asked of the dataset in which the patient may be included. As big data and deep learning find use in radiation oncology it is essential that the data from which these approaches are built can be traced back to thoughtful implementation. As an example, unapproved contours should be tagged or named to indicate their use to future researchers. We found many examples of contour and naming mismatches which, while likely irrelevant in the individual patient treatment plan had an effect on the work herein.

Finally, the use of automatically contoured targets represents the aspect of this work most in need of further investigation. There exist enormous variability in not only the contouring of targets among physician but also in the choice of target anatomy when given patient disease characteristics. These challenges will be underlined in a system for which a simplified approach is desired. It is likely that additional nodal level groupings may be required, the atlases may need further development for patients with gross disease, and/or for patients from other demographics. The use of a second contouring technique as a check of the atlas based contours represents a novel contribution but requires further investigations, specifically the extension of the secondary technique for additional head-and-neck subsites and stages.

As a system, we envision the use of the automated treatment planning strategy as follows; a user is required to submit an approved patient CT and patient plan order which includes information about the patient's acceptability for treatment (pregnancy status, prior radiation, implants, etc) and treatment goals (prescribed dose and fractionation). This information is then used to generate automatic normal tissues and nodal level groupings. The process is then suspended for the manual delineation of primary disease. Then on a single button click clinical and planning target volumes are generated using rules and the atlas generated nodal volumes; the targets are then reviewed by the attending physician. At this time the physician will also be presented with the results of the results of the random forest models for the detection of normal tissue contour errors and should use this to aide in the approval and editing of these structures as deemed appropriate. The automatic process then resumes with the automatic identification of the marked and treatment isocenter, the selection of beam parameters, plan optimization, dose calculation and production of accompanying plan documentation. The whole process is estimated to take less than an hour, much of which time is unsupervised.

Data concerning the editing of normal structures, targets, the results of the error prediction models, the timing of each step, the acceptability of the final plan, and others should be collected for all patients. This will allow in depth retrospect analysis of the performance of the automated planning technique and improvement of the system.

We believe this system has the ability to greatly reduce the human effort needed for one of the most technical and tedious aspects of radiation oncology. Automation of this process, however, should be considered for use in low-resource settings in the context of radiation therapy as a whole. Notably, while advanced techniques have been shown to outperform older, simpler techniques for the treatment of head-and-neck cancer there exist downstream effects due to this transition. Quality assurance process including that of individual treatment plans as well as of the machines and even of the treatment planning

process itself must reflect the transition to these advanced techniques. Further, the tasks completed by an automated approach must be able to be accomplished by, ideally, a number of trained staff; in the case of head-and-neck treatment this requires knowledge not only of plan optimization but also of the contouring of normal tissues and targets. In circumstances where the autoplanner may not be appropriate for patient treatment plan development, when the system is not functional due to planned or unplanned circumstances, and for routine quality assurance checks it is essential that treatment planning skills not be lost, or never developed, in centers where autoplanning is used.

Conclusions

The work presented here represent efforts towards two solutions to reduce the great and growing disparity of radiation therapy around the world. In LMICs, where 84% of the world's population lives and two thirds of the cancer population lives, there are only 30% of the worlds radiation facilities; 11 countries of more than a million people have not a single teletherapy unit. Improving access to radiation therapy is not a simple process, it requires innovation approaches, collaborative efforts, and consistent reassessment of the needs and status in the areas of interest. The approaches presented here represents efforts both to reduce the upfront and ongoing cost of radiotherapy machines through the introduction of a fixed-beam rotating patient paradigm and the automation of one of the most time-intensive and technologically complex aspects of advanced radiotherapy – treatment planning. We believe these efforts could benefit clinics in low-resource areas and partly address the growing need for radiotherapy around the world.

Appendix

Appendix A – Specific Aim 1

The following questionnaire was administered to patients after their participation in the inter- and intra- fraction setup study in the treatment chair. Participants were also encouraged to give verbal feedback.

This questionnaire evaluates your degree of comfort when in a treatment position.

The first 15 questions refer to treatments in a seated position

1. Getting on the chair was: Easy → Difficult (0-5)

2. When initially positioned on the chair, I felt: Relaxed → Tense (0-5)

DURING THE SIMULATED TREATMENT (ON THE CHAIR)…

3. …I felt like I needed to move. Not at all → Constantly (0-5)

4. …I felt: Restless → Calm (0-5)

5. …My body felt tense. Strongly Disagree→Strongly Agree (0-5)

6. …My breathing felt fluid and easy. Strongly Disagree→ Strongly Agree (0-5)

7. …My neck was: Not at all→Perfectly Comfortable (0-5)

8. …My arms were: Not at all→Perfectly Comfortable (0-5)

9. …My back was: Not at all→Perfectly Comfortable (0-5)

10. …My legs were: Not at all→Perfectly Comfortable (0-5)

11. …Overall my body was: Not at all→Perfectly Comfortable (0-5)

12. …I could have fallen asleep. Strongly Disagree→Strongly Agree (0-5)

13. …I had discomfort due to the chair I was positioned on. Strongly Disagree→ Strongly Agree (0-5)

OVERALL (ON THE CHAIR)…

14. I felt stable and supported on the chair. Strongly Disagree→ Strongly Agree (0-5)

15. I felt absolutely safe on the chair. Strongly Disagree→ Strongly Agree (0-5)

The next 15 questions are about you regular actual treatments (lying down)

16. Getting on the patient couch was: Easy → Difficult (0-5)

17. When initially positioned on the couch, I felt: Relaxed → Tense (0-5)

DURING TREATMENT (LYING DOWN, ON THE COUCH)…

18. …I felt like I needed to move. Not at all → Constantly (0-5)

19. …I felt: Restless → Calm (0-5)

20. …My body felt tense. Strongly Disagree→Strongly Agree (0-5)

21. …My breathing felt fluid and easy. Strongly Disagree→ Strongly Agree (0-5)

22. …My neck was: Not at all→Perfectly Comfortable (0-5)

23. …My arms were: Not at all→Perfectly Comfortable (0-5)

24. …My back was: Not at all→Perfectly Comfortable (0-5)

25. …My legs were: Not at all→Perfectly Comfortable (0-5)

26. …Overall my body was: Not at all→Perfectly Comfortable (0-5)

27. …I could have fallen asleep. Strongly Disagree→Strongly Agree (0-5)

28. …I had discomfort due to the chair I was positioned on. Strongly Disagree→ Strongly Agree (0-5)

OVERALL (LYING DOWN, ON THE COUCH)…

29. I felt stable and supported on the couch. Strongly Disagree→ Strongly Agree (0-5)

30. I felt absolutely safe on the couch. Strongly Disagree→ Strongly Agree (0-5)

FINAL: COMPARISON OF SEATED AND LYING DOWN POSITIONS

31. I prefer sitting up to lying down for treatment: strongly agree→strongly disagree (0-5)

Please note any comments or feedback below.

_____

_____

_____

_____

Thank you for completing this survey; we appreciate your feedback.

Appendix B – Specific Aim 2

The contouring of small structures is often more difficult, as indicated by lower Dice similarity coefficients and larger mean surface and Hausdorff distances, compared to larger structures. Cochlea, the smallest structures implemented into the clinic had the largest disagreement with the clinically edited structures and it was therefore decided that an "uncertainty margin" of 5mm, similar to a PRV, would be added to the autocontoured cochlea for treatment planning; In Table 14, we showed that a margin of 5mm would be sufficient to cover 100% of the edited cochlea for more than 95% of the population.

Other smalls structures required for treatment planning but not implemented clinically include the optic chiasm, optic nerves, and lens. The accurate contouring of these structures is of particular importance with the tumor is close to the optic structures and because they are serial structures (chiasm and nerves) or have particularly low clinical dose constraints (lens, D_{max}<7Gy). As data of the clinical edit to these autocontours was not available, we examined the agreement between autocontoured structures using the MACS algorithm and independently drawn physician contours for 37 patients. Occasionally, the MACS algorithm would result in an empty structure due to insufficient overlap of individual contours from the 12 atlas patients, in this case automatic optimization could not continue and in the context of the automated planning approach manual delineation of these structures would be required. To compensate, for this we investigated the possibility of using structures which have been expanded on the atlas patients in order to increase their volume and increase the likelihood of overlapping anatomy as input into the staple algorithm. In *Table 29* are the average volume, Dice Similarity Coefficient, true positive fraction and false positive fraction as well as the percentage of structure with true positive fraction of 100% and greater than 95% given pre-MACS expansion of 0-2mm in 1mm increments and with post macs (PRV-style) expansions

of 1 to 5mm in 1mm increments. It was decided that structures with 1mm pre-MACs expansion would be used with 5mm post MACS expansions for treatment planning.

Details of the final autoplanning strategy can be found in Table 30

Table 29. Agreement of autocontoured optical structures with pre- and post- contouring expansions.

		Volume (cc)			True Positive Fraction			False Positive Fraction		
Pre-MACS Expansion (mm)	Post-MACS Expansion (mm)	Optic Chiasm	Lens	Optic Nerves	Optic Chiasm	Lens	Optic Nerves	Optic Chiasm	Lens	Optic Nerves
Physician Drawn		1.25	0.37	1.16						
0	0	0.70	0.12	0.50	0.27	0.27	0.37	0.52	0.24	0.17
0	1	1.12	0.27	0.91	0.39	0.47	0.52	0.57	0.39	0.34
0	2	2.41	0.66	2.27	0.58	0.75	0.81	0.70	0.57	0.58
0	3	3.12	0.98	2.98	0.66	0.83	0.85	0.73	0.68	0.67
0	4	5.25	1.81	5.29	0.82	0.95	0.90	0.80	0.80	0.80
0	5	6.56	2.48	6.62	0.87	0.96	0.91	0.83	0.85	0.84
1	0	1.26	0.34	1.10	0.45	0.56	0.59	0.56	0.44	0.38
1	1	1.83	0.62	1.72	0.56	0.69	0.69	0.61	0.61	0.54
1	2	3.62	1.26	3.94	0.72	0.85	0.93	0.75	0.75	0.73
1	3	4.50	1.74	4.92	0.77	0.87	0.95	0.78	0.81	0.78
1	4	7.26	2.97	8.29	0.90	0.95	0.98	0.84	0.88	0.86
1	5	8.86	3.87	10.11	0.93	0.95	0.98	0.86	0.91	0.89
2	0	2.99	1.08	3.63	0.66	0.85	0.92	0.72	0.70	0.71
2	1	4.06	1.70	5.04	0.75	0.90	0.95	0.77	0.80	0.78
2	2	6.58	2.81	8.72	0.82	0.93	0.99	0.84	0.87	0.87
2	3	7.93	3.65	10.38	0.85	0.93	0.99	0.86	0.90	0.89
2	4	11.63	5.57	15.44	0.95	0.99	1.00	0.89	0.93	0.93
2	5	13.84	6.96	18.28	0.96	0.99	1.00	0.91	0.95	0.94

Table 29. Continued from previous page.

		Dice			% with TPF > 100%			% with TPF >=95%		
Pre-MACS Expansion (mm)	Post-MACS Expansion (mm)	Optic Chiasm	Lens	Optic Nerves	Optic Chiasm	Lens	Optic Nerves	Optic Chiasm	Lens	Optic Nerves
0	0	0.34	0.38	0.50	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.39	0.49	0.57	0.00	0.00	0.00	0.00	2.63	0.00
0	2	0.39	0.51	0.54	0.00	18.42	0.00	0.00	28.95	13.89
0	3	0.37	0.44	0.47	0.00	34.21	0.00	3.45	52.63	25.00
0	4	0.32	0.32	0.32	11.11	69.44	13.89	22.22	83.33	44.44
0	5	0.28	0.25	0.27	14.81	86.11	22.22	33.33	88.89	50.00
1	0	0.43	0.53	0.59	0.00	2.56	0.00	0.00	5.13	0.00
1	1	0.44	0.48	0.54	0.00	20.51	0.00	0.00	23.08	2.78
1	2	0.36	0.37	0.42	0.00	53.85	11.11	6.90	64.10	58.33
1	3	0.33	0.30	0.36	6.90	69.23	30.56	20.69	76.92	66.67
1	4	0.27	0.21	0.24	18.52	83.78	50.00	51.85	86.49	86.11
1	5	0.23	0.17	0.20	29.63	86.49	72.22	66.67	89.19	86.11
2	0	0.38	0.43	0.44	0.00	35.90	11.11	3.45	53.85	52.78
2	1	0.35	0.32	0.35	3.45	58.97	38.89	13.79	79.49	66.67
2	2	0.26	0.22	0.23	13.79	76.92	61.11	27.59	84.62	97.22
2	3	0.23	0.17	0.20	17.24	84.62	86.11	51.72	89.74	97.22
2	4	0.19	0.13	0.14	29.63	97.30	97.22	66.67	97.30	100.00
2	5	0.16	0.10	0.12	48.15	97.30	100.00	77.78	97.30	100.00

Table 30. Planning constraints of automated planning strategy. The origin of the contours in the RPA, in the evaluation of the planning strategy and the constraints (if used for optimization) are listed. *Optic chiasm, nerves and lens are expanded by 1mm on the atlas patients prior to MACS contouring. *

XXXXcGy indicates that the prescribed dose of the corresponding structure is embedded in the structure name and varies between patients.

Structure	Structure Name	Origin (Prospective RPA Use)	Origin	Type	Volume	Dose	Priority
Primary and Nodal GTV	GTVp, GTVn	Physician Drawn	Physician Drawn				
High Dose CTV	zCTV1_XXXXcGy	Physician Drawn or Edited and Approved GTVp+1cm and GTVn+0.5cm	Physician Drawn				
Intermediate Dose CTV	zCTV2_XXXXcGy	Physician Drawn or Edited and Approved atlas derived nodes	Physician Drawn				
Low Dose CTV	zCTV3_XXXXcGy	Physician Drawn or Edited and Approved atlas derived nodes	Physician Drawn				
High Dose PTV	zPTV1_XXXXcGy	High Dose CTV + 0.5cm, within 3mm Contracted BodyAuto	Physician Drawn				
Intermediate Dose PTV	zPTV2_XXXXcGy	Intermediate Dose CTV + 0.5cm, within 3mm Contracted BodyAuto	Physician Drawn				
Low Dose PTV	zPTV3_XXXXcGy	Low Dose CTV + 0.5cm, within 3mm Contracted BodyAuto	Physician Drawn				

High Dose Planning PTV	zpPTV1_XXXXcGy	High Dose PTV		Upper	0	108%	400
				Upper	20	103%	400
				Lower	100	96%	400
				Lower	96		400
High Dose PTV Ring	zrPTV1_XXXXcGy	PTV1 - PTV1 contracted 5mm		Upper	0	103%	300
5mm Ring 1 mm from PTV1	zWallPTV1_05	PTV1+6mm - PTV1+1mm		Upper	0	100%	120
PTV1 Minimum Dose assessment structure	z8bodyPTV1	PTV1 within body contracted 8mm					
Intermediate Dose Planning PTV	zpPTV2_XXXXcGy	Intermediate Dose PTV pulled 5mm from pPTV1		Upper	0	108%+1Gy	190
				Upper	0	103%+1Gy	190
				Lower	100	97%+1.5Gy	190
				Lower	97	100%+1.5Gy	250
Low Dose Planning PTV	zpPTV3_XXXXcGy	Low Dose PTV pulled 5mm from pPTV1 and pPTV2		Upper	0	108%+1Gy	190
				Upper	0	103%+1Gy	190
				Lower	100	97%+1.5Gy	190
				Lower	97	100%+1.5Gy	250
Body	BodyAuto	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	106	900
Normal Tissue Avoidance	zNT_avoid	BodyAuto outside all PTVs on slices within 1cm of PTV1		Upper	0	45	70
				Upper	10	35	70
				Upper	20	30	70
Non Target Tissue	zNonPTVs	BodyAuto outside all PTVs					
Brain	Brain	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	45	130
				Line	Model		75
Brainstem	BrainStem	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	45	200
				Upper	1	Model	35

				Line	Model		75
Brainstem 7mm PRV	zBrainStem_07	BrainStem + 7mm		Upper	0	50	300
Brainstem 5mm PRV	zBrainStem_05	BrainStem + 5mm					
Left Cochlea	Cochlea_L	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Cochlea	Cochlea_R	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Expanded Left Cochlea	zCochlea_L_05	Cochlea_L + 5mm		Upper	0	45	75
Expanded Right Cochlea	zCochlea_R_05	Cochlea_R + 5mm		Upper	0	45	75
Esophagus	Esophagus	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Esophagus Avoidance Structure	zEsophagus_avoid	Esophagus + 5mm outside of all PTVs		Line	Model		75
Left Eye	Eye_L	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	40	130
Right Eye	Eye_R	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	40	130
Larynx	Larynx	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Larynx Avoidance	zLarynx_avoid	Larynx + 5mm outside all PTVs		Line	Model		70
Left Lens*	Lens_L_A1	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Lens*	Lens_R_A1	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Lens Avoidance	zLens_L_A1B2	Lens_R + 5mm		Upper	0	45	130
Left Lens Avoidance	zLens_R_A1B2	Lens_R + 5mm		Upper	0	5	130

Lung Avoidance	Lungs_Avoid	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Mandible	Mandible	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	2	105%	150
				Lines	Model		75
Left Optic Nerve*	OpticNrv_L_A1	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Optic Nerve*	OpticNrv_R_A1	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Optic Chiasm*	Chiasm_A1	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Optical Nerve Avoidance	zOpticNrv_R_A1B2	OpticNrv_R + 2mm		Upper	0	45	130
Left Optical Nerve Avoidance	zOpticNrv_L_A1B2	OpticNrv_L + 2mm		Upper	0	45	130
Chiasm Avoidance	zChiasm_A1B2	Chiasm + 2mm		Upper	0	45	130
Oral Cavity	OralCavity	Automatically Contoured					
Oral Cavity Avoidance	zOralCavity_plan	OralCavity outside all PTVs		Line	Model		100
Left Parotid Gland	Parotid_L	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Right Parotid Gland	Parotid_R	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
Left Parotid Gland Planning structure	zParotid_L_sub	Parotid_L outside of all PTVs		Line	Model		120
Right Parotid Gland Planning Structure	zParotid_R_sub	Parotid_R outside of all PTVs		Line	Model		120
Posterior Neck	fsPostAvoid	Automatically Contoured					

Posterior Neck Avoidance	zPostNeck_avoid	fsPostAvoid outside of all PTVs		Upper	0	35	120
Shoulders	Shoulders	Automatically Contoured		Upper	0	25	40
Spinal Cord	SpinalCord	Automatically Contoured	Clinical Plan (if DNE, autocontoured)	Upper	0	40	300
				Upper	1	Model	35
				Line	Model		75
Spinal Cord 5mm PRV	zSpinalCord_05	SpinalCord + 5mm					
Spinal Cord 7mm PRV	zSpinalCord_07	SpinalCord + 7mm		Upper	0	50	300
Spinal Canal	Spinal Canal	Automatically Contoured					
Left Submandibular Gland	Glnd_Submand_L	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
				Upper	10	35	70
				Upper	20	30	70
Right Submandibular Gland	Glnd_Submand_R	Automatically Contoured	Clinical Plan (if DNE, autocontoured)				
				Upper	10	35	70
				Upper	20	30	70
Vertebral Column	VeterbalColumn	Automatically Contoured					

Appendix C – Specific Aim 3

Additional analysis of autocontour algorithm performance was performed and is presented here. The deformable image registration known as “Deeds” was presented in 2013 and is currently licensed by Varian Medical Systems. This algorithm is used as the independent method for normal tissue contouring used in the machine learning models for the detection of autocontouring errors as presented in Chapter 5. In one version of the algorithm active shape models have been implemented for improved contouring(1). A physician review of the this algorithm, which includes the shape model and uses a Varian provided atlas, for ten patients was conducted and the results can be seen in Figure 40. The algorithm was also investigated for use with the same atlas as used in the MACS algorithm and with a second atlas generated from 11 patients treated at MD Anderson Cancer Center, this independent atlas was used for contour propagation for the machine learning models. The agreement to independently drawn physician contours of these three algorithms based on the Deeds DIR as well as for MACS can be seen in Table 31.

In the smart segmentation software, as described, there are vendor provided ratings of the expected agreement between the test patient in question and each of the available atlas patients. We sought to investigate if there existed trends in these ratings relative to physician ratings or quantitative agreement with physician contours – no apparent trends were observed, the results can be seen in Table 32. We also sought to investigate if the physician ratings produced noticeable trends compared to both the average star rating and quantitative agreement with physician drawn structures – no apparent trend was observed, the results can be seen in Table 33.

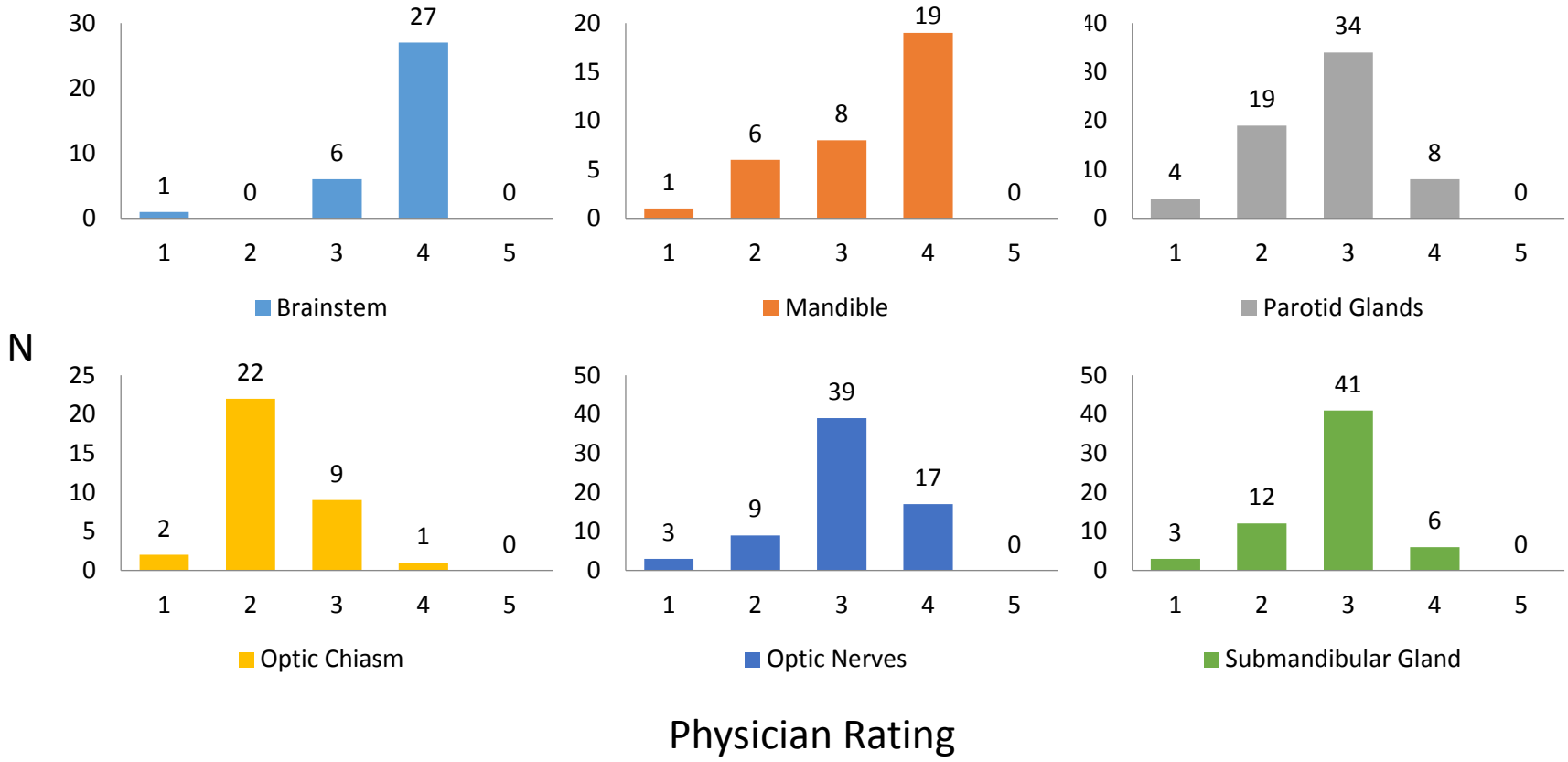


Figure 40. Distribution of physician scores of Deeds based multi-atlas algorithm with shape models.

Table 31. Quantitative analysis of additional normal structure autocontouring algorithms.

Contouring Method	Structure	N	Dice Similarity Coefficient				Hausdorff Distance (cm)				Mean Surface Distance (cm)			
			Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
Deeds MDA Independent Atlas vs Phys	Brainstem	75	0.81	0.12	0.24	0.90	0.92	0.62	0.35	4.06	2.24	1.53	1.16	10.70
	Mandible	39	0.80	0.13	0.29	0.92	1.89	1.36	0.47	5.71	2.36	2.21	0.89	12.23
	Parotid Glands	140	0.72	0.10	0.33	0.90	1.74	0.55	0.70	3.68	3.03	0.97	1.46	6.74
	Brain	26	0.97	0.04	0.76	0.98	1.53	1.21	0.49	4.54	1.36	1.49	0.77	8.19
	Cochlea	94	0.59	0.17	0.11	0.85	0.45	0.22	0.16	1.27	1.46	0.63	0.66	3.95
	Esophagus	29	0.51	0.16	0.10	0.80	3.59	1.74	0.76	7.20	5.90	4.40	1.33	25.68
	Eye	58	0.79	0.08	0.63	0.93	0.46	0.11	0.22	0.72	1.75	0.57	0.83	3.03
	Lungs	12	0.88	0.09	0.74	0.97	3.75	2.42	1.39	7.94	4.33	4.15	1.01	13.04
	Spinal Cord	70	0.71	0.12	0.24	0.88	5.37	3.92	0.25	17.33	5.83	6.01	0.82	33.05
Deeds MDA Original Atlas vs Phys	Brainstem	75	0.79	0.13	0.24	0.90	0.97	0.69	0.43	4.06	2.47	1.65	1.19	10.46
	Mandible	39	0.81	0.13	0.27	0.91	1.92	1.29	0.59	5.62	2.28	2.11	1.00	12.76
	Parotid Glands	140	0.74	0.09	0.34	0.88	1.65	0.47	0.80	2.94	2.86	0.84	1.61	6.71
	Brain	26	0.98	0.01	0.96	0.98	1.37	0.94	0.50	3.81	1.01	0.17	0.80	1.51
	Cochlea	94	0.57	0.16	0.13	0.83	0.47	0.22	0.17	1.27	1.49	0.58	0.72	3.85
	Esophagus	29	0.51	0.14	0.12	0.76	3.54	1.46	0.94	5.96	5.45	3.50	1.98	20.07
	Eye	58	0.79	0.07	0.65	0.93	0.49	0.11	0.25	0.73	1.70	0.45	0.76	2.66
	Lungs	12	0.84	0.14	0.51	0.97	3.80	3.47	0.97	11.50	6.34	7.05	0.85	24.20
	Spinal Cord	70	0.74	0.11	0.28	0.87	3.90	3.23	0.37	13.47	4.29	5.32	0.90	28.89
MACS vs Phys	Brainstem	75	0.80	0.12	0.25	0.91	0.98	0.62	0.41	3.70	2.37	1.55	1.10	9.73
	Mandible	39	0.85	0.06	0.64	0.93	1.83	1.12	0.41	5.56	1.66	0.89	0.76	4.62
	Parotid Glands	140	0.79	0.07	0.44	0.89	1.47	0.64	0.59	3.60	2.37	0.76	1.30	6.16
	Brain	26	0.98	0.00	0.97	0.99	1.33	0.86	0.48	3.75	1.06	0.17	0.80	1.43
	Cochlea	94	0.50	0.17	0.06	0.88	0.48	0.19	0.14	1.41	1.61	0.68	0.42	4.27
	Esophagus	29	0.64	0.12	0.37	0.83	2.32	1.21	0.65	5.42	3.16	1.57	1.15	8.02
	Eye	58	0.84	0.07	0.56	0.93	0.47	0.15	0.25	1.06	1.42	0.39	0.72	2.36
	Lungs	12	0.76	0.21	0.39	0.94	4.56	4.23	1.24	14.08	9.09	9.47	1.49	29.89
	Spinal Cord	70	0.73	0.09	0.30	0.86	3.75	2.91	0.35	16.31	3.78	3.90	0.94	22.39

Deeds Varian Shape vs Phys	**Brainstem**	75	0.78	0.13	0.23	0.90	1.09	0.71	0.49	4.26	2.59	1.73	1.26	10.54
	Mandible	39	0.80	0.14	0.27	0.92	2.16	1.31	0.52	5.63	2.51	2.40	0.87	12.30
	Optic Chiasm	21	0.47	0.13	0.13	0.62	0.88	0.34	0.47	2.00	2.16	0.73	1.27	3.74
	Optic Nerve	48	0.63	0.10	0.38	0.86	0.81	0.45	0.17	1.87	1.48	0.57	0.72	3.11
	Submandibular Glands	53	0.72	0.13	0.27	0.88	0.97	0.52	0.42	3.18	2.33	1.62	1.16	9.83
	Parotid Glands	140	0.74	0.08	0.38	0.87	1.94	0.60	0.68	3.62	2.95	0.85	1.52	6.73

Table 32. Autocontour performance and Eclipse Star Rating. Mean and standard deviation of physician scores and agreement with physician contours divided by Eclipse Smart Segmentation star rating.

Structure	Star	N	Physician Rating		Dice		Hausdorff Distance (cm)		Mean Surface Distance (cm)	
			Mean	Std	Mean	Std	Mean	Std	Mean	Std
Brainstem	3	6	2.83	0.41	0.73	0.05	0.89	0.11	0.28	0.04
	4	77	3.17	0.59	0.74	0.05	0.96	0.25	0.29	0.05
	5	8	3.13	0.35	0.76	0.03	1.20	0.47	0.27	0.05
Cochlea	3	7	2.71	0.49	0.26	0.06	0.96	0.13	0.30	0.03
	4	80	3.04	0.37	0.40	0.08	0.81	0.20	0.23	0.04
	5	8	2.88	0.35	0.29	0.06	1.07	0.18	0.31	0.04
Eye	3	11	3.73	1.01	0.77	0.07	0.77	0.39	0.20	0.06
	4	158	3.28	0.90	0.74	0.08	0.69	0.27	0.23	0.07
	5	16	3.00	0.73	0.72	0.06	0.76	0.16	0.24	0.05
Lung	3	6	3.17	0.75	0.92	0.03	2.46	0.51	0.16	0.03
	4	79	2.95	0.99	0.92	0.05	2.28	0.87	0.20	0.19
	5	8	3.13	1.46	0.90	0.11	2.42	1.26	0.30	0.45
Mandible	3	6	2.50	0.84	0.61	0.05	1.69	0.21	0.37	0.06
	4	87	2.85	0.83	0.68	0.08	1.71	0.50	0.29	0.09
	5	8	2.88	1.13	0.69	0.09	1.74	0.41	0.29	0.11
Parotid	3	8	2.25	0.46	0.46	0.12	3.39	0.60	0.74	0.17
	4	148	2.91	0.82	0.68	0.09	1.94	0.72	0.36	0.12
	5	16	2.63	0.96	0.67	0.06	1.94	0.48	0.36	0.09
Spinal Cord	3	6	4.33	0.52	0.72	0.01	1.30	0.12	0.18	0.01
	4	79	4.04	0.41	0.67	0.07	2.95	2.13	0.31	0.13
	5	8	4.25	0.46	0.65	0.06	3.51	1.17	0.33	0.09

Table 33. Autocontour performance and physician rating. Mean and standard deviation of Eclipse Stars and agreement with physician contours divided by Physician Rating.

Structure	Rating	N	Star		Dice		Hausdorff Distance (cm)		Mean Surface Distance (cm)	
			Mean	Std	Mean	Std	Mean	Std	Mean	Std
Brainstem	2	8	3.88	0.35	0.70	0.07	0.94	0.08	0.32	0.05
	3	63	4.03	0.44	0.74	0.04	1.02	0.31	0.29	0.05
	4	19	4.05	0.23	0.77	0.03	0.85	0.16	0.25	0.03
	5	1	4.00	0.00	0.79	0.00	1.00	0.00	0.24	0.00
Cochlea	2	7	3.86	0.69	0.33	0.06	0.99	0.17	0.28	0.03
	3	81	4.02	0.39	0.39	0.09	0.83	0.21	0.24	0.05
	4	7	4.00	0.00	0.40	0.07	0.83	0.11	0.23	0.03
Eye	2	41	4.07	0.35	0.69	0.11	0.90	0.44	0.27	0.11
	3	66	4.06	0.43	0.74	0.05	0.68	0.19	0.23	0.05
	4	63	4.02	0.34	0.76	0.06	0.62	0.12	0.21	0.05
	5	15	3.80	0.41	0.78	0.07	0.62	0.14	0.20	0.06
Lung	1	6	4.17	0.41	0.78	0.10	3.82	0.98	0.77	0.50
	2	25	4.04	0.35	0.91	0.06	2.72	0.51	0.23	0.17
	3	32	3.97	0.40	0.93	0.03	2.27	0.85	0.16	0.08
	4	25	3.96	0.35	0.94	0.02	1.74	0.58	0.14	0.06
	5	5	4.40	0.55	0.95	0.01	1.44	0.32	0.09	0.01
Mandible	2	41	4.00	0.45	0.61	0.07	1.92	0.52	0.36	0.09
	3	41	4.02	0.27	0.72	0.05	1.52	0.36	0.24	0.06
	4	14	4.00	0.39	0.71	0.06	1.67	0.36	0.27	0.08
	5	5	4.20	0.45	0.76	0.02	1.67	0.73	0.21	0.02
Parotid	1	1	4.00	0.00	0.29	0.00	2.57	0.00	0.68	0.00
	2	63	4.06	0.50	0.60	0.10	2.37	0.63	0.47	0.15
	3	75	4.01	0.26	0.70	0.07	1.86	0.75	0.34	0.10
	4	26	4.08	0.27	0.75	0.04	1.68	0.69	0.27	0.07
	5	7	4.14	0.38	0.76	0.08	1.39	1.01	0.26	0.14
Spinal Cord	3	5	4.00	0.00	0.56	0.05	3.47	2.98	0.42	0.19

	4	76	4.03	0.36	0.67	0.06	3.10	2.00	0.31	0.12
	5	12	4.00	0.60	0.73	0.04	1.37	1.23	0.21	0.08

References

1. J Ferlay, I. S., M Ervik, R Dikshit, S Eser, C Mathers, M Rebelo, DM Parkin, D Forman, F Bray. 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 International Agency for Research on Cancer, Lyon, France.
2. Thun, M. J., J. O. DeLancey, M. M. Center, A. Jemal, and E. M. Ward. 2010. The global burden of cancer: priorities for prevention. *Carcinogenesis* 31: 100-110.
3. Papavramidou, N., T. Papavramidis, and T. Demetriou. 2010. Ancient Greek and Greco–Roman Methods in Modern Surgical Treatment of Cancer. *Annals of Surgical Oncology* 17: 665-667.
4. Teresiak, M., and J. Busza. 1997. The role of surgery in combinated treatment of cancers. *Reports of Practical Oncology* 2: 60.
5. DeVita, V. T., and E. Chu. 2008. A History of Cancer Chemotherapy. *Cancer Research* 68: 8643-8653.
6. Atun, R., D. A. Jaffray, M. B. Barton, F. Bray, M. Baumann, B. Vikram, T. P. Hanna, F. M. Knaul, Y. Lievens, T. Y. M. Lui, M. Milosevic, B. O'Sullivan, D. L. Rodin, E. Rosenblatt, J. Van Dyk, M. L. Yap, E. Zubizarreta, and M. Gospodarowicz. 2015. Expanding global access to radiotherapy. *The Lancet Oncology* 16: 1153-1186.
7. Delaney, G., S. Jacob, C. Featherstone, and M. Barton. 2005. The role of radiotherapy in cancer treatment. *Cancer* 104: 1129-1137.
8. Rontgen, W. C. 1896. On a New Kind of Rays. *Science* 3: 227-231.
9. Holsti, L. R. 1995. Development Of Clinical Radiotherapy Since 1896. *Acta Oncologica* 34: 995-1003.
10. Barton, M. B., M. Frommer, and J. Shafiq. 2006. Role of radiotherapy in cancer control in low-income and middle-income countries. *The Lancet Oncology* 7: 584-595.

11. Van Der Giessen, P.-H., J. Alert, C. Badri, M. Bistrovic, D. Deshpande, D. Kardamakis, D. Van Der Merwe, N. Da Motta, L. Pinillos, R. Sajjad, Y. Tian, and V. Levin. 2004. Multinational assessment of some operational costs of teletherapy. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 71: 347-355.
12. Fisher, B. J., L. C. Daugherty, J. P. Einck, G. Suneja, M. M. Shah, L. K. Dad, R. W. Mutter, J. B. Wilkinson, and A. J. Mundt. 2014. Radiation oncology in Africa: improving access to cancer care on the African continent. *International journal of radiation oncology, biology, physics* 89: 458-461.
13. Bank, W. 2018. Data.
14. IAEA. n.d. Directory of Radiotherapy Centers. Vienna, Austria.
15. Datta, N. R., M. Samiei, and S. Bodis. 2014. Radiation Therapy Infrastructure and Human Resources in Low- and Middle-Income Countries: Present Status and Projections for 2020. *International Journal of Radiation Oncology • Biology • Physics* 89: 448-457.
16. Lyman, J. T., and C. Y. Chong. 1974. Isah: A versatile treatment positioner for external radiation therapy. *Cancer* 34: 12-16.
17. Miller, R. W., A. A. Raubitschek, F. S. Harrington, J. Van de Geijn Ovadia, J. Ovadia, and E. Glatstein. 1991. An isocentric chair for the simulation and treatment of radiation therapy patients. *International Journal of Radiation Oncology • Biology • Physics* 21: 469-473.
18. Zhang, D. 2015. SU-E-I-20: Comprehensive Quality Assurance Test of Second Generation Toshiba Aquilion Large Bore CT Simulator Based On AAPM TG-66 Recommendations. *Medical physics* 42: 3246-3246.
19. Yang, J., D. Chu, L. Dong, and L. E. Court. 2014. Advantages of simulating thoracic cancer patients in an upright position. *Practical Radiation Oncology* 4: e53-e58.

20. Dellamonica, J., N. Lerolle, C. Sargentini, S. Hubert, G. Beduneau, F. Di Marco, A. Mercat, J. L. Diehl, J. C. Richard, G. Bernardin, and L. Brochard. 2013. Effect of different seated positions on lung volume and oxygenation in acute respiratory distress syndrome. *Intensive care medicine* 39: 1121-1127.

21. Graham, M. V., J. A. Purdy, B. Emami, W. Harms, W. Bosch, M. A. Lockett, and C. A. Perez. 1999. Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). *International Journal of Radiation Oncology*Biology*Physics* 45: 323-329.

22. Hsieh, M., P. Balter, B. Beadle, P. Chi, F. Stingo, and L. Court. 2014. SU-E-T-273: Radiation Shielding for a Fixed Horizontal-Beam Linac in a Shipping Container and a Conventional Treatment Vault. *Medical physics* 41: 286-286.

23. Court, L., J. Yang, D. Fullen, N. Han, J. Ko, S. Mason, K. Nguyen, S. Stein, X. Fave, M. Hsieh, S. Kuruvila, E. Hillebrandt, J. Palmer, B. Beadle, B. Dabaja, H. Skinner, G. Ibbott, and P. Balter. 2013. SU-E-T-359: Patients Could (and Should) Be Treated in An Upright Position. *Medical Physics* 40: 287-287.

24. Eslick, E. M., and P. J. Keall. 2015. The Nano-X Linear Accelerator: A Compact and Economical Cancer Radiotherapy System Incorporating Patient Rotation. *Technology in cancer research & treatment* 14: 565-572.

25. McCarroll, R., B. Youssef, B. Beadle, M. Bojador, R. Cardan, R. Famiglietti, D. Followill, G. Ibbott, A. Jhingran, C. Trauernicht, P. Balter, and L. Court. 2017. Model for Estimating Power and Downtime Effects on Teletherapy Units in Low-Resource Settings. *Journal of Global Oncology* 3: 563-571.

26. Michalski, A., J. Atyeo, J. Cox, and M. Rinks. 2012. Inter- and intra-fraction motion during radiation therapy to the whole breast in the supine position: A systematic review. *Journal of Medical Imaging and Radiation Oncology* 56: 499-509.

27. Jain, P., T. Marchant, M. Green, G. Watkins, J. Davies, C. McCarthy, J. Loncaster, A. Stewart, B. Magee, C. Moore, and P. Price. 2009. Inter-fraction motion and dosimetric consequences during breast intensity-modulated radiotherapy (IMRT). *Radiotherapy and Oncology* 90: 93-98.

28. Wu, J., T. Haycocks, H. Alasti, G. Ottewell, N. Middlemiss, M. Abdolell, P. Warde, A. Toi, and C. Catton. 2001. Positioning errors and prostate motion during conformal prostate radiotherapy using on-line isocentre set-up verification and implanted prostate markers. *Radiotherapy and Oncology* 61: 127-133.

29. Hurkmans, C. W., P. Remeijer, J. V. Lebesque, and B. J. Mijnheer. 2001. Set-up verification using portal imaging; review of current clinical practice. *Radiotherapy and Oncology* 58: 105-120.

30. Rubinstein, A. E., W. S. Ingram, B. M. Anderson, S. S. Gay, X. J. Fave, R. B. Ger, R. E. McCarroll, C. A. Owens, T. J. Netherton, K. D. Kisling, L. E. Court, J. Yang, Y. Li, J. Lee, D. S. Mackin, and C. E. Cardenas. 2017. Cost-effective immobilization for whole brain radiation therapy. *Journal of applied clinical medical physics* 18: 116-122.

31. Perez, C. A., J. A. Purdy, W. Harms, R. Gerber, M. V. Graham, J. W. Matthews, W. Bosch, R. Drzymala, B. Emami, S. Fox, and et al. 1995. Three-dimensional treatment planning and conformal radiation therapy: preliminary evaluation. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 36: 32-43.

32. Zelefsky, M. J., Y. Yamada, Z. Fuks, Z. Zhang, M. Hunt, O. Cahlon, J. Park, and A. Shippy. 2008. Long-term results of conformal radiotherapy for prostate cancer: impact of dose escalation on biochemical tumor control and distant metastases-free survival outcomes. *International journal of radiation oncology, biology, physics* 71: 1028-1033.

33. Bradley, J., M. V. Graham, K. Winter, J. A. Purdy, R. Komaki, W. H. Roa, J. K. Ryu, W. Bosch, and B. Emami. 2005. Toxicity and outcome results of RTOG 9311: a phase

I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma. *International journal of radiation oncology, biology, physics* 61: 318-328.

34. Wolden, S. L., W. C. Chen, D. G. Pfister, D. H. Kraus, S. L. Berry, and M. J. Zelefsky. 2006. Intensity-modulated radiation therapy (IMRT) for nasopharynx cancer: update of the Memorial Sloan-Kettering experience. *International journal of radiation oncology, biology, physics* 64: 57-62.
35. Austin-Seymour, M., G. T. Chen, J. Rosenman, J. Michalski, K. Lindsley, and M. Goitein. 1995. Tumor and target delineation: current research and future challenges. *International journal of radiation oncology, biology, physics* 33: 1041-1052.
36. Rudat, V., M. Flentje, D. Oetzel, M. Menke, W. Schlegel, and M. Wannenmacher. 1994. Influence of the positioning error on 3D conformal dose distributions during fractionated radiotherapy. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 33: 56-63.
37. Pirzkall, A., M. Carol, F. Lohr, A. Hoss, M. Wannenmacher, and J. Debus. 2000. Comparison of intensity-modulated radiotherapy with conventional conformal radiotherapy for complex-shaped tumors. *International journal of radiation oncology, biology, physics* 48: 1371-1380.
38. Gupta, T., J. Agarwal, S. Jain, R. Phurailatpam, S. Kannan, S. Ghosh-Laskar, V. Murthy, A. Budrukkar, K. Dinshaw, K. Prabhash, P. Chaturvedi, and A. D'Cruz. Three-dimensional conformal radiotherapy (3D-CRT) versus intensity modulated radiation therapy (IMRT) in squamous cell carcinoma of the head and neck: A randomized controlled trial. *Radiotherapy and Oncology* 104: 343-348.
39. Tribius, S., and C. Bergelt. 2011. Intensity-modulated radiotherapy versus conventional and 3D conformal radiotherapy in patients with head and neck cancer: is there a worthwhile quality of life gain? *Cancer treatment reviews* 37: 511-519.

40. Verbakel, W. F., J. P. Cuijpers, D. Hoffmans, M. Bieker, B. J. Slotman, and S. Senan. 2009. Volumetric intensity-modulated arc therapy vs. conventional IMRT in head-and-neck cancer: a comparative planning and dosimetric study. *International journal of radiation oncology, biology, physics* 74: 252-259.

41. Van Dyk, J., E. Zubizarreta, and Y. Lievens. Cost evaluation to optimise radiation therapy implementation in different income settings: A time-driven activity-based analysis. *Radiotherapy and Oncology* 125: 178-185.

42. Abdel-Wahab, M., E. Zubizarreta, A. Polo, and A. Meghzifene. Improving Quality and Access to Radiation Therapy—An IAEA Perspective. *Seminars in Radiation Oncology* 27: 109-117.

43. Nelms, B. E., G. Robinson, J. Markham, K. Velasco, S. Boyd, S. Narayan, J. Wheeler, and M. L. Sobczak. 2012. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. *Practical Radiation Oncology* 2: 296-305.

44. Das, I. J., C.-W. Cheng, K. L. Chopra, R. K. Mitra, S. P. Srivastava, and E. Glatstein. 2008. Intensity-Modulated Radiation Therapy Dose Prescription, Recording, and Delivery: Patterns of Variability Among Institutions and Treatment Planning Systems. *JNCI: Journal of the National Cancer Institute* 100: 300-307.

45. Wexler, A., B. Gu, S. Goddu, M. Mutic, S. Yaddanapudi, L. Olsen, T. Harry, C. Noel, T. Pawlicki, S. Mutic, and B. Cai. 2017. FMEA of manual and automated methods for commissioning a radiotherapy treatment planning system. *Medical Physics* 44: 4415-4425.

46. 2016. Varian Medical Systems 2016 Annual Report. Varian Medical Systems, Palo Alto, CA. 142.

47. Purdie, T. G., R. E. Dinniwell, A. Fyles, and M. B. Sharpe. Automation and Intensity Modulated Radiation Therapy for Individualized High-Quality Tangent Breast

Treatment Plans. *International Journal of Radiation Oncology • Biology • Physics* 90: 688-695.

48. Laberta, V. 2017. Benefits of automation in radiation oncology. LWW. 1-9.

49. McIntosh, C., M. Welch, A. McNiven, D. A. Jaffray, and T. G. Purdie. 2017. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. *Physics in medicine and biology* 62: 5926-5944.

50. Kalet, I. J., and W. Paluszynski. 1990. Knowledge-based computer systems for radiotherapy planning. *Am J Clin Oncol* 13: 344-351.

51. Wang, J., W. Hu, Z. Yang, X. Chen, Z. Wu, X. Yu, X. Guo, S. Lu, K. Li, and G. Yu. 2017. Is it possible for knowledge-based planning to improve intensity modulated radiation therapy plan quality for planners with different planning experiences in left-sided breast cancer patients? *Radiation Oncology* 12: 85.

52. Good, D., J. Lo, W. R. Lee, Q. J. Wu, F.-F. Yin, and S. K. Das. A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning. *International Journal of Radiation Oncology • Biology • Physics* 87: 176-181.

53. Tol, J. P., A. R. Delaney, M. Dahele, B. J. Slotman, and W. F. Verbakel. 2015. Evaluation of a knowledge-based planning solution for head and neck cancer. *International journal of radiation oncology, biology, physics* 91: 612-620.

54. Chanyavanich, V., S. K. Das, W. R. Lee, and J. Y. Lo. 2011. Knowledge-based IMRT treatment planning for prostate cancer. *Medical Physics* 38: 2515-2522.

55. Lim, J. Y., and M. Leech. 2016. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. *Acta Oncologica* 55: 799-806.

56. Gambacorta, M. A., C. Valentini, N. Dinapoli, L. Boldrini, N. Caria, M. C. Barba, G. C. Mattiucci, D. Pasini, B. Minsky, and V. Valentini. 2013. Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumors using auto-segmentation computed system. *Acta Oncologica* 52: 1676-1681.

57. Lustberg, T., J. van Soest, M. Gooding, D. Peressutti, P. Aljabar, J. van der Stoep, W. van Elmpt, and A. Dekker. 2018. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiotherapy and Oncology* 126: 312-317.

58. Shah, A. P., J. B. Strauss, M. C. Kirk, S. S. Chen, T. K. Kroc, and T. W. Zusag. 2009. Upright 3D Treatment Planning Using a Vertical CT. *Medical Dosimetry* 34: 82-86.

59. Duisters, C., H. Beurskens, S. Nijsten, M. Starmans, S. Wanders, T. Verschueren, P. Lambin, A. Minken, and D. D. Ruysscher. 2006. Palliative chest irradiation in sitting position in patients with bulky advanced lung cancer. *Radiotherapy and Oncology* 79: 285-287.

60. Boag, J. W., and H. J. Hodt. 1971. Adjustable chair for radiotherapy of head and neck cancer. *The British Journal of Radiology* 44: 316-317.

61. Marcus, K. C., G. Svensson, L. P. Rhodes, and P. M. Mauch. 1992. Mantle irradiation in the upright position: a technique to reduce the volume of lung irradiated in patients with bulky mediastinal Hodgkin's disease. *International journal of radiation oncology, biology, physics* 23: 443-447.

62. Miller, R. W., A. A. Raubitschek, F. S. Harrington, J. Van de Geijn Ovadia, J. Ovadia, and E. Glatstein. 1991. An isocentric chair for the simulation and treatment of radiation therapy patients. *International Journal of Radiation Oncology*Biology*Physics* 21: 469-473.

63. Klein, E. E., T. Wasserman, and B. Ermer. 1995. Clinical introduction of a commercial treatment chair to facilitate thorax irradiation. *Medical dosimetry : official journal of the American Association of Medical Dosimetrists* 20: 171-176.

64. Fave, X., J. Yang, L. Carvalho, R. Martin, T. Pan, P. Balter, and L. Court. 2014. Upright cone beam CT imaging using the onboard imager. *Med Phys* 41: 061906.

65. Feain, I., C.-C. Shieh, P. White, R. O'Brien, S. Fisher, W. Counter, P. Lazarakis, D. Stewart, S. Downes, M. Jackson, S. Baxi, B. Whelan, K. Makhija, C.-Y. Huang, M. Barton, and P. Keall. 2016. Functional imaging equivalence and proof of concept for image-guided adaptive radiotherapy with fixed gantry and rotating couch. *Advances in Radiation Oncology* 1: 365-372.

66. Yoo, S., and F.-F. Yin. 2006. Dosimetric feasibility of cone-beam CT-based treatment planning compared to CT-based treatment planning. *International Journal of Radiation Oncology*Biology*Physics* 66: 1553-1561.

67. Yoo, S., and F. Yin. 2007. Feasibility of Cone-beam CT based treatment planning. In *World Congress on Medical Physics and Biomedical Engineering 2006*. R. Magjarevic, and J. H. Nagel, eds. Springer Berlin Heidelberg. 1831-1834.

68. Ding, G. X., D. M. Duggan, C. W. Coffey, M. Deeley, D. E. Hallahan, A. Cmelak, and A. Malcolm. 2007. A study on adaptive IMRT treatment planning using kV cone-beam CT. *Radiotherapy and Oncology* 85: 116-125.

69. van Kranen, S., S. van Beek, C. Rasch, M. van Herk, and J. J. Sonke. 2009. Setup uncertainties of anatomical sub-regions in head-and-neck cancer patients after offline CBCT guidance. *International journal of radiation oncology, biology, physics* 73: 1566-1573.

70. Kapanen, M., M. Laaksomaa, T. Tulijoki, S. Peltola, T. Wigren, S. Hyödynmaa, and P.-L. Kellokumpu-Lehtinen. 2013. Estimation of adequate setup margins and threshold for position errors requiring immediate attention in head and neck cancer radiotherapy based on 2D image guidance. *Radiation oncology (London, England)* 8: 212-212.

71. Nawa, K., A. Haga, A. Nomoto, R. A. Sarmiento, K. Shiraishi, H. Yamashita, and K. Nakagawa. 2017. Evaluation of a commercial automatic treatment planning system for

prostate cancers. *Medical dosimetry : official journal of the American Association of Medical Dosimetrists* 42: 203-209.

72. Zhang, X., X. Li, E. M. Quan, X. Pan, and Y. Li. 2011. A methodology for automatic intensity-modulated radiation treatment planning for lung cancer. *Physics in Medicine & Biology* 56: 3873.

73. Yong, Y., and X. Lei. 2004. Clinical knowledge-based inverse treatment planning. *Physics in Medicine & Biology* 49: 5101.

74. Masoud, Z., L. Troy, L. Nan, T. Zhen, R. H. Edwin, J. Xun, and J. S. B. 2014. A DVH-guided IMRT optimization algorithm for automatic treatment planning and adaptive radiotherapy replanning. *Medical Physics* 41: 061711.

75. 2018. The Cancer Imaging Archive.

76. Ang, K. K., Q. Zhang, D. I. Rosenthal, P. F. Nguyen-Tan, E. J. Sherman, R. S. Weber, J. M. Galvin, J. A. Bonner, J. Harris, A. K. El-Naggar, M. L. Gillison, R. C. Jordan, A. A. Konski, W. L. Thorstad, A. Trotti, J. J. Beitler, A. S. Garden, W. J. Spanos, S. S. Yom, and R. S. Axelrod. 2014. Randomized Phase III Trial of Concurrent Accelerated Radiation Plus Cisplatin With or Without Cetuximab for Stage III to IV Head and Neck Carcinoma: RTOG 0522. *Journal of Clinical Oncology* 32: 2940-2950.

77. Yuan, L., Y. Ge, W. R. Lee, F. F. Yin, J. P. Kirkpatrick, and Q. J. Wu. 2012. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. *Med Phys* 39: 6868-6878.

78. Systems, V. M. 2017. RapidPlan Knowledge-Based Planning Frequently Asked Questions. In *Varian Medical Systems*, Palo Atlo, CA. 4.

79. Brown, M. B., and A. B. Forsythe. 1974. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association* 69: 364-367.

80. Peszynska-Piorun, M., J. Malicki, and W. Golusinski. 2012. Doses in organs at risk during head and neck radiotherapy using IMRT and 3D-CRT. *Radiology and Oncology* 46: 328-336.

81. Rodrigues, G., V. Velker, and L. Best. 2013. *Radiation Oncology Primer and Review Essential Concepts and Protocols*. Demos Medical Publishing.

82. Moore, K. L., G. C. Kagadis, T. R. McNutt, V. Moiseenko, and S. Mutic. 2014. Vision 20/20: Automation and advanced computing in clinical radiation oncology. *Medical Physics* 41.

83. Hansen, C. R., A. Bertelsen, I. Hazell, R. Zukauskaite, N. Gyldenkerne, J. Johansen, J. G. Eriksen, and C. Brink. 2016. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. *Clinical and Translational Radiation Oncology* 1: 2-8.

84. Yan, C., A. G. Combine, G. Bednarz, R. J. Lalonde, B. Hu, K. Dickens, R. Wynn, D. C. Pavord, and M. Saiful Huq. 2017. Clinical implementation and evaluation of the Acuros dose calculation algorithm. *Journal of applied clinical medical physics* 18: 195-209.

85. Horan, G., T. W. Roques, J. Curtin, and A. Barrett. 2006. "Two are better than one": a pilot study of how radiologist and oncologists can collaborate in target volume definition. *Cancer imaging : the official publication of the International Cancer Imaging Society* 6: 16-19.

86. Jeanneret-Sozzi, W., R. Moeckli, J. F. Valley, A. Zouhair, E. M. Ozsahin, and R. O. Mirimanoff. 2006. The reasons for discrepancies in target volume delineation : a SASRO study on head-and-neck and prostate cancers. *Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft ... [et al]* 182: 450-457.

87. Brouwer, C. L., R. J. Steenbakkers, E. van den Heuvel, J. C. Duppen, A. Navran, H. P. Bijl, O. Chouvalova, F. R. Burlage, H. Meertens, J. A. Langendijk, and A. A. van 't

Veld. 2012. 3D Variation in delineation of head and neck organs at risk. *Radiation Oncology* 7: 32.

88. Nelms, B. E., W. A. Tomé, G. Robinson, and J. Wheeler. 2012. Variations in the Contouring of Organs at Risk: Test Case From a Patient With Oropharyngeal Cancer. *International Journal of Radiation Oncology*Biology*Physics* 82: 368-378.

89. Bortfeld, T., and R. Jeraj. 2011. The physical basis and future of radiation therapy. *The British journal of radiology* 84.

90. Jameson, M. G., L. C. Holloway, P. J. Vial, S. K. Vinod, and P. E. Metcalfe. 2010. A review of methods of analysis in contouring studies for radiation oncology. *Journal of medical imaging and radiation oncology* 54: 401-410.

91. Njeh, C. F. 2008. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *Journal of medical physics / Association of Medical Physicists of India* 33: 136-140.

92. Brouwer, C. L., R. J. Steenbakkers, J. Bourhis, W. Budach, C. Grau, V. Gregoire, M. van Herk, A. Lee, P. Maingon, C. Nutting, B. O'Sullivan, S. V. Porceddu, D. I. Rosenthal, N. M. Sijtsema, and J. A. Langendijk. 2015. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 117: 83-90.

93. Hu, K., A. Lin, A. Young, G. Kubicek, J. W. Piper, A. S. Nelson, J. Dolan, R. Masino, and M. Machtay. Timesavings for Contour Generation in Head and Neck IMRT: Multi-institutional Experience with an Atlas-based Segmentation Method. *International Journal of Radiation Oncology • Biology • Physics* 72: S391.

94. Stapleford, L. J., J. D. Lawson, C. Perkins, S. Edelman, L. Davis, M. W. McDonald, A. Waller, E. Schreibmann, and T. Fox. Evaluation of Automatic Atlas-Based Lymph

Node Segmentation for Head-and-Neck Cancer. *International Journal of Radiation Oncology • Biology • Physics* 77: 959-966.

95. Gambacorta, M. A., C. Valentini, N. Dinapoli, L. Boldrini, N. Caria, M. C. Barba, G. C. Mattiucci, D. Pasini, B. Minsky, and V. Valentini. 2013. Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumors using auto-segmentation computed system. *Acta oncologica (Stockholm, Sweden)* 52: 1676-1681.

96. Yang, J., A. Amini, R. Williamson, L. Zhang, Y. Zhang, R. Komaki, Z. Liao, J. Cox, J. Welsh, L. Court, and L. Dong. 2013. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiotherapy. *Practical radiation oncology* 3: 3139-e3147.

97. Zhou, R., Z. Liao, T. Pan, S. A. Milgrom, C. C. Pinnix, A. Shi, L. Tang, J. Yang, Y. Liu, D. Gomez, Q.-N. Nguyen, B. S. Dabaja, L. Court, and J. Yang. 2017. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiotherapy and Oncology* 122: 66-71.

98. Gambacorta, M. A., L. Boldrini, C. Valentini, N. Dinapoli, G. C. Mattiucci, G. Chiloiro, D. Pasini, S. Manfrida, N. Caria, B. D. Minsky, and V. Valentini. 2016. Automatic segmentation software in locally advanced rectal cancer: READY (REsearch program in Auto Delineation sYstem)-RECTAL 02: prospective study. *Oncotarget* 7: 42579-42584.

99. Sharp, G., K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan, and J. Yang. 2014. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys* 41: 050902.

100. Teguh, D. N., P. C. Levendag, P. W. Voet, A. Al-Mamgani, X. Han, T. K. Wolf, L. S. Hibbard, P. Nowak, H. Akhiat, M. L. Dirkx, B. J. Heijmen, and M. S. Hoogeman. 2011. Clinical validation of atlas-based auto-segmentation of multiple target volumes and

normal tissue (swallowing/mastication) structures in the head and neck. *International journal of radiation oncology, biology, physics* 81: 950-957.

101. Systems, V. M. 2012. Clinical Evaluation of an Automated Segmentation Module. 1-8.

102. Wang, H., L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung. 2005. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 50: 2887-2905.

103. Warfield, S. K., K. H. Zou, and W. M. Wells. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging* 23: 903-921.

104. Wang, H., L. Dong, M. F. Lii, A. L. Lee, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung. 2005. Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. *International journal of radiation oncology, biology, physics* 61: 725-735.

105. Yang, J., Y. Zhang, L. Zhang, and L. Dong. 2010. Automatic Segmentation of Parotids from CT Scans Using Multiple Atlases. In *Medical Image Analysis for the Clinic: A Grand Challenge*. B. vanGinneken, K. Murphy, T. Heimann, V. Pekar, and X. Deng, eds. 323-330.

106. Yang, J., A. Garden, Y. Zhang, L. Zhang, L. Court, and L. Dong. 2012. WE-E-213CD-09: Multi-Atlas Fusion Using a Tissue Appearance Model. *Med Phys* 39: 3961.

107. Cardenas, C. E., A. S. Mohamed, R. Tao, A. J. Wong, M. J. Awan, S. Kuruvila, M. Aristophanous, G. B. Gunn, J. Phan, B. M. Beadle, S. J. Frank, A. S. Garden, W. H. Morrison, C. D. Fuller, and D. I. Rosenthal. 2016. Prospective Qualitative and Quantitative Analysis of Real-Time Peer Review Quality Assurance Rounds Incorporating Direct Physical Examination for Head and Neck Cancer Radiation Therapy. *International journal of radiation oncology, biology, physics*.

108. Heinrich, M. P., M. Jenkinson, M. Brady, and J. A. Schnabel. 2013. MRF-Based Deformable Registration and Ventilation Estimation of Lung CT. *IEEE Transactions on Medical Imaging* 32: 1239-1248.

109. Altman, M. B., J. A. Kavanaugh, H. O. Wooten, O. L. Green, T. A. DeWees, H. Gay, W. L. Thorstad, H. Li, and S. Mutic. 2015. A framework for automated contour quality assurance in radiation therapy including adaptive techniques. *Phys Med Biol* 60: 5199-5209.

110. Beasley, W. J., A. McWilliam, N. J. Slevin, R. I. Mackay, and M. van Herk. 2016. An automated workflow for patient-specific quality control of contour propagation. *Phys Med Biol* 61: 8577-8586.

111. Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5-32.

112. Ospina, J. D., J. Zhu, C. Chira, A. Bossi, J. B. Delobel, V. Beckendorf, B. Dubray, J. L. Lagrange, J. C. Correa, A. Simon, O. Acosta, and R. de Crevoisier. 2014. Random forests to predict rectal toxicity following prostate cancer radiation therapy. *International journal of radiation oncology, biology, physics* 89: 1024-1031.

113. Valdes, G., T. D. Solberg, M. Heskel, L. Ungar, and C. B. Simone, 2nd. 2016. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. *Physics in medicine and biology* 61: 6105-6120.

114. Valdes, G., A. T. Berman, C. Hann-Hsiang, T. D. Solberg, and C. B. Simone, II. 2016. Machine Learning Analysis Can Identify Clinical and Dosimetric Factors Predictive of Chest Wall Syndrome in Patients With Stage I Non-Small Cell Lung Cancer Treated With Stereotactic Body Radiation Therapy. *International Journal of Radiation Oncology • Biology • Physics* 96: S172.

115. Yu, W., C. Tang, B. P. Hobbs, X. Li, E. J. Koay, I. I. Wistuba, B. Sepesi, C. Behrens, J. Rodriguez Canales, E. R. Parra Cuentas, J. J. Erasmus, L. E. Court, and J. Y. Chang.

Development and Validation of a Predictive Radiomics Model for Clinical Outcomes in Stage I Non-small Cell Lung Cancer. *International Journal of Radiation Oncology • Biology • Physics*.

116. Delaney, A. R., M. Dahele, B. J. Slotman, and W. F. A. R. Verbakel. 2018. Is accurate contouring of salivary and swallowing structures necessary to spare them in head and neck VMAT plans? *Radiotherapy and Oncology*.

117. Brouwer, C. L., R. J. Steenbakkers, E. van den Heuvel, J. C. Duppen, A. Navran, H. P. Bijl, O. Chouvalova, F. R. Burlage, H. Meertens, J. A. Langendijk, and A. A. van 't Veld. 2012. 3D Variation in delineation of head and neck organs at risk. *Radiation oncology (London, England)* 7: 32.

118. La Macchia, M., F. Fellin, M. Amichetti, M. Cianchetti, S. Gianolini, V. Paola, A. J. Lomax, and L. Widesott. 2012. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiation oncology (London, England)* 7: 160-160.

119. Daisne, J.-F., and A. Blumhofer. 2013. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiation oncology (London, England)* 8: 154-154.

120. Thomson, D., C. Boylan, T. Liptrot, A. Aitkenhead, L. Lee, B. Yap, A. Sykes, C. Rowbottom, and N. Slevin. 2014. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. *Radiation Oncology* 9: 173.

121. Mayo, C. S., J. M. Moran, W. Bosch, Y. Xiao, T. McNutt, R. Popple, J. Michalski, M. Feng, L. B. Marks, C. D. Fuller, E. Yorke, J. Palta, P. E. Gabriel, A. Molineu, M. M. Matuszak, E. Covington, K. Masi, S. L. Richardson, T. Ritter, T. Morgas, S. Flampouri, L. Santanam, J. A. Moore, T. G. Purdie, R. C. Miller, C. Hurkmans, J. Adams, Q.-R. Jackie Wu, C. J. Fox, R. A. Siochi, N. L. Brown, W. Verbakel, Y. Archambault, S. J. Chmura, A. L. Dekker, D. G. Eagle, T. J. Fitzgerald, T. Hong, R. Kapoor, B. Lansing,

S. Jolly, M. E. Napolitano, J. Percy, M. S. Rose, S. Siddiqui, C. Schadt, W. E. Simon, W. L. Straube, S. T. St. James, K. Ulin, S. S. Yom, and T. I. Yock. 2018. American Association of Physicists in Medicine Task Group 263: Standardizing Nomenclatures in Radiation Oncology. *International Journal of Radiation Oncology*Biology*Physics* 100: 1057-1066.

Vita

Rachel Elizabeth McCarroll was born in Liberty, Missouri on January 31, 1991, the daughter of Sara Anne McCarroll and Scott Roger McCarroll. After completing her work at Excelsior Springs High School, Excelsior Springs, MO in 2009, she entered Truman State University in Kirksville, Missouri. She received the degree of Bachelor of Science with a major in Physics and a minor in Mathematics from Truman in May, 2013. In August of 2013 she entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.