Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer
van Kranen, S.R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer

Simon van Kranen

The cover represents an image of the head and neck region. The matches refer to the process of image registration, in Dutch called 'matching'. There are 2x40 matches used, which equals the amount of collimator-leafs in a conventional linear accelerator.

© 2017
Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer

Simon Robert van Kranen
The work described in this thesis was performed at The Netherlands Cancer Institute/Antoni van Leeuwenhoek, Amsterdam, The Netherlands.

The work was financially supported by the Dutch Cancer Society (KWF), grant NKI 2005-3378.

ISBN/EAN 978-94-6233-723-7

Cover design: S.R. van Kranen
Printing: GildePrint

Copyright © 2017 S.R. van Kranen

No part of this publications may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval without permission of the holder of the copyright.
Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 29 september 2017, te 11 uur

doors

Simon Robert van Kranen
geboren te Amsterdam
Promotiecommissie:

Promotores:  
Prof. dr. C.R. N Rasch  
Prof. dr. M.B. van Herk  
Universiteit van Amsterdam

Copromotor:  
Dr. ir. J. Sonke  
NKI-AvL

Overige leden:  
Prof. dr. M.W.M. van den Brekel  
Dr. C.D. Fuller  
Prof. dr. B.J.M. Heijmen  
Prof. dr. B.W. Raaymakers  
Prof. dr. ir. G.J. Strijkers  
Dr. A. Bel  
Universiteit van Amsterdam  
MD Anderson Cancer Center Houston TX  
Erasmus Universiteit Rotterdam  
Universiteit Utrecht  
Universiteit van Amsterdam

Faculteit der Geneeskunde

Het beschreven werk werd mede mogelijk gemaakt door een subsidie van KWF Kankerbestrijding (NKI:2005-3378).
Voor Maike
Rosa, Liselotte en Morris
Contents

1 Introduction 1
2 Setup uncertainties of anatomical sub-region in head-and-neck cancer patients after offline CBCT guidance 17
3 Evaluation of tumor shape variability in head-and-neck cancer patients over the course of radiation therapy using implanted gold markers 37
4 Analysis of GTV reduction during radiotherapy for oropharyngeal cancer: implications for adaptive radiotherapy 51
5 Correction strategies to manage deformations in head-and-neck radiotherapy 69
6 First clinical experience with a multiple region of interest registration and correction method in radiotherapy of head-and-neck cancer patients 85
7 Deformable image registration for adaptive radiotherapy of head and neck cancer: accuracy and precision in the presence of tumor changes 97
8 Adaptive radiotherapy with an average anatomy model: evaluation and quantification of residual deformations in head and neck cancer patients 115
9 Head and neck margin reduction with adaptive radiation therapy: robustness of treatment plans against anatomy changes 137
10 General discussion and conclusion 161

Publications 183
Summary 185
Nederlandse samenvatting 189
Dankwoord 193
Curriculum vitae 197
Introduction
Radiotherapy (RT) is an important curative treatment modality for patients with head and neck cancer (HNC), as an alternative to or in conjunction with surgery and/or chemotherapy. RT uses ionizing radiation to destroy tumors cells. Typically, linear accelerators (Linacs) are used to produce beams of high-energy photons. These high energy photons interact with tissue by ionizing atoms which create free electrons that ultimately damage the DNA in cells beyond repair. The process is not selective, killing both malignant and normal cells in the irradiated tissue. Improvements in the selectivity of the treatment can be made through geometrical targeting and fractionation – i.e. increasing the differential damage between tumor and normal cells.

Even with improved selectivity, RT for HNC is a toxic treatment, especially if combined with chemotherapy [1]. Side effects of treatment are categorized into either acute or chronic toxicities. Important acute radiation induced toxicities are: severe mucositis, dermatitis, xerostomia, and the need for feeding tube. Important chronic toxicities are: xerostomia, dysphagia and fatigue [2, 3].

Because RT is not selective in the cells it affects and has many toxicities, the tissue irradiated in the body surrounding the tumor has to be minimized. Over the years, RT has become more precise. First of all, pre-treatment imaging has evolved and nowadays includes magnetic resonance imaging (MRI) and/or positron emission tomography (PET) imaging in addition to conventional computed tomography (CT) imaging. Multi-modal imaging has improved detection, staging, and definition of the target volume and organs at risk (OARs) for RT planning [4–8]. Secondly, introduction of new planning and delivery techniques, such as multi leaf-collimators (MLCs), intensity modulated radiotherapy (IMRT) and volumetric arc therapy (VMAT) have allowed the creation of treatment plans with high spatial accuracy, achieving loco-regional control with superior sparing of organs at risk (OARs) [9], improving the therapeutic ratio.

However, trying to deliver these spatially accurate dose distributions is difficult due to geometrical uncertainties in position, size and shape of the target volume in relation to OARs. Geometrical uncertainties during treatment arise, for example, from limitations in pre-treatment imaging, poorly reproducible target and OARs delineation, patient setup and deformations, and tumor regression. To ensure correct target dosage in the presence of geometrical uncertainties, the target volume is expanded with a safety margin that essentially increases the irradiated volume and therefore will increase radiation induced side effects. Reduction of geometrical errors would allow smaller safety margins and thereby reduce the side effects, or allow dose escalation to the tumor without increasing toxicity.

This thesis investigates the use of in-room imaging with CBCT to quantify and reduce geometrical uncertainties during treatment delivery for HNC. In this chapter, we start out describing the general process of RT of HNC, then point out where geometrical uncertainties occur, how margins are derived and how, in general, uncertainties during RT delivery may be reduced. We finalize with the purpose of this thesis.
The chain of radiotherapy

Radiotherapy of HNC consists of several steps that together form the chain of radiotherapy, see figure 1.1. Two main phases can be distinguished: the preparation and the delivery phase. In the preparation phase, the patient’s anatomy is imaged, the tumor and OARs are delineated and a treatment plan is optimized. The delivery phase consists of setting up the patient on the treatment machine, followed nowadays by image guided setup verification and correction, and finally the delivery of the treatment. As delivery is typically fractioned, the delivery phase is repeatedly performed on consecutive days, until the total prescribed dose is delivered.

**Figure 1.1.** The chain of radiotherapy consists of two phases: the treatment preparation and the treatment delivery phase. The preparation consists of pre-treatment imaging, target and organs at risk (OAR) delineation and treatment planning and is performed once. During treatment delivery, image guided radiotherapy (with cone beam CT imaging, image registration and setup correction) allows to accurately deliver the dose to the target. Treatment delivery is repeatedly performed.
Pre-treatment imaging and target & OAR delineation

The goal of pre-treatment imaging is to visualize the tumor in relation to OARs and to provide the plan optimization process with an anatomy for dose calculation. Typically, a CT scan is acquired and used to delineate the **Gross Tumor Volume (GTV)**. The *International Commission on Radiation Units and Measurements* (ICRU) defines the GTV as the *gross palpable or visible/demonstrable extent and location of the malignant growth* [10–12]. In RT of HNC, the GTV consists of the primary tumor (GTVp) and might contain involved lymph nodes (GTVn), where the primary tumor has spread (figure 1.2).

![Figure 1.2. Delineation of the primary gross target volume (GTVp) and involved lymph nodes (GTVn) (green) in the planning CT (right), based on the PET signal marking areas with high metabolic uptake (left). The GTVs are expanded to a high risk clinical target volumes (CTV1) and planning target volumes (PTV1). CTV2 represents medium risk areas with suspected microscopic involvement. Their PTV2 may be treated to a different dose level.](image)

With mono-modal CT imaging, substantial inter-observer variation exists, i.e., different observers will delineate different target volumes [6, 13, 14]. To reduce the variation amongst observers, pre-treatment CT imaging is often complemented with MRI and/or PET, resulting in more consistent target volumes and less inter-observer variation [4–8]. See figure 1.2 for an example of PET/CT usage for target definition. Physical examination, multidisciplinary review sessions, guidelines and the use of non-axial views have been shown to further decrease observer variation in target and OAR definition [6, 8, 15]. Currently, target definition is one of the largest sources of
uncertainty in RT of HNC. Improving target and OARs definition is subject to ongoing research, however, it falls outside the scope of this thesis.

From the GTV the Clinical Target Volume (CTV) is derived. There are two types of CTV. The first type of CTV is an extensions of the GTV to account for the likely presence of microscopic tumor extensions, i.e. areas with malignant cells that are too small to visualize in scans, that should nevertheless be irradiated to achieve local control. Definition of the CTV from the GTV is unclear and based on general assumptions, relating to patterns of failure in similar patients, and pathological findings in surgery patients, in which the tumor has been physically removed. Typically, GTVs are uniformly expanded into CTVs with a fixed margin of in the order of 10 mm [16], and are subsequently edited to anatomical borders, e.g. at bone or air. According to guidelines from the ICRU [10–12], these volumes are indicated with CTV1, also called primary or boost CTV (figure 1.2). A second type of CTV is needed to describe volumes at distance of the GTV that are suspected of subclinical involvement, i.e. regional lymph node stations (N0). These volumes are considered for treatment based on studies of patterns of failure in similar patients and are indicated with CTV2, also called the elective CTV, see figure 1.2. In RT of HNC, these CTVs can be quite large, extending over the full length of the neck. Note that they are anatomically defined, have a probability of containing malignant cells and are assumed to have a low tumor cell density, which allows a lower dose level to be used compared to the CTV1. Some institutes also define a CTV2 around the CTV1 of the primary tumor in an attempt to improve local control, typically 3 – 5 mm.

Geometrical errors, resulting from target definition, tumor misplacement, tumor deformation or other sources of geometric uncertainty may lead to underdosage of a CTV during treatment. To account for such errors, the ICRU recommends to expand the CTV with a safety margin to form the Planning Target Volume (PTV) [10–12], figure 1.2. Following the distinction into CTV1 and CTV2, we also have PTV1 and PTV2, with different dose descriptions and possibly resulting from different margins. At the start of the work for this thesis, overall uniform PTV margins of 5 mm were used in our institute.

Treatment Planning

To start treatment planning, a dose is prescribed to the PTVs. A frequently used schedule is to prescribe a dose of 70 Gy to PTV1, to be delivered in 35 fractions. Depending on the plan type, the elective dose to PTV2 is delivered simultaneously with the (boost) dose to PTV1, (simultaneous integrated boost) and prescribed at 54.25 Gy, or delivered with a sequentially technique, in 23 fractions at 46 Gy. Modern planning techniques achieve a high spatial accuracy (conformity) by using inverse planning: dose-volume constraints on target volumes and OARs are defined beforehand and provided as input for the treatment planning system (the planning objectives). The TPS will then optimize the beam parameters, such as beam shape, amount of segments and monitor units, to best meet these objectives. The treatment planner will repeatedly try to refine the treatment plan by adjusting the objectives [17]. Either multiple fixed beams (IMRT) or
a rotational technique can be used (VMAT). The improved conformity may be applied to lower the dose to OARs while preserving or improving dose to the tumor. With IMRT most benefits have been found in reduction of toxicity [18–21]. An example of improvements possible with VMAT over the older conformal technique, 2DCRT, is shown in figure 1.3.

**Figure 1.3.** Example dose distributions resulting from 2DCRT(left) and VMAT(right). VMAT achieves highly conformal dose distribution to the PTV, thereby optimally sparing organs at risk, for example the spinal cord and the parotid glands, without compromising on target coverage. Note that this patient has no nodal involvement.

**Treatment Delivery**

Treatment delivery starts by positioning the patient on the treatment table. Patient setup is traditionally performed by aligning external markers, i.e. tattoos on the skin or the immobilization mask, with lasers mounted around the treatment machine. The lasers define a coordinate system that is fixed to both the CT scanner and the treatment machine. The lasers allow linking of the coordinate system of the CT scanner with the LINAC. Because the patient was aligned to the lasers at the CT scanner prior to acquisition of the planning CT, the planning position can be reproduced at the time of treatment. With good patient positioning the tumor is correctly aligned with the treatment beams, and the geometry of the patient during treatment planning is well reproduced.
In order to better reproduce the anatomy as captured with pre-treatment imaging, fixation devices are used that limit patient motion and thereby achieve a higher degree of positioning reproducibility. Typical fixation devices for RT of HNC are (customized) neck rests, mouthpieces with tongue suppressors, vacuum cushions and thermo-plastic masks. Fixation devices are continuously being improved to further minimize residual uncertainty [22–24].

To further improve patient setup, imaging devices are, nowadays, used in the treatment room to re-image the anatomy of the patient prior to treatment delivery. Discrepancies in patient setup are determined by comparing the image with the planning anatomy from the planning CT. With a couch shift the setup error may be corrected. Use of in-room imaging to guide the treatment is known as Image Guided RT (IGRT), and is further discussed below.

Thereafter, the treatment is delivered. Because normal tissues recuperate better from radiation than tumor cells, a differential effect is created when the treatment is fractionated. Typically, for HNC treatment, the total dose is delivered in 35 fractions over a period of 6-7 weeks, as stated in the dose description. However, the classical strategy of planning treatment upfront assumes that there are few anatomical changes in target and OARs over this period. As treatment gets more and more accurate and margins are reduced, anatomy changes may have a larger impact.

**Geometrical uncertainties**

Errors are defined as all discrepancies between treatment preparation and delivery. Errors are an instantiation of uncertainties during treatment. In general, errors and uncertainties for a particular patient are unknown beforehand, but the magnitude of the uncertainty can be estimated based on data from a population of similar patients undergoing a similar treatment. During treatment, some errors will be measured and corrected. We distinguish errors as either systematic or random.

**Systematic errors**

Systematic errors are errors that are identical at each fraction throughout the whole treatment and they are characterized by calculating the average error over all fractions. The physical effect of systematic errors is typically described as a shift of the target volume with respect to the planned dose distribution [25].

At pre-treatment imaging, systematic errors may be introduced by imaging artefacts, for example from patient motion (e.g. swallowing), distortion (MRI) or limited resolution (PET). Pre-treatment imaging artefacts can introduce systematic errors in target and OAR delineation as well as in dose calculation, that will propagate into the entire treatment. Furthermore, the use of complementary modalities in the delineation process, requires image registration that may introduce uncertainties, that result in a systematic error in the position and shape of target and/or OARs.
Chapter 1

Setup errors are defined as the difference, between planning and delivery, in the position of bony structures, which are considered representative of the position of the tumor. If setup errors over a series of fractions do not cancel out, they represent a systematic deviation in the position of the bony anatomy and thus the tumor. Setup errors represent a substantial source of both systematic as well as random errors. With IGRT, setup errors may be corrected; however, the image guidance process itself may introduce uncertainties, e.g. in the mechanical couch corrections or the utilized image registration step.

Moreover, the anatomy captured in the planning CT may not be representative for the anatomy during treatment. A different pose results in a systematic discrepancy between expected and actual (average) position of target volumes and/or OARs during treatment. Finally, over the course of treatment, patients with HNC show considerable progressive changes such as weight loss, tumor regression and/or shrinkage of the parotid glands [26]. These time trends also have a systematic component.

Random Errors

Random errors differ from fraction to fraction and cancel out on average. Random errors are therefore characterized with the standard deviation of the errors over the course of treatment. The effect of random errors is generally described by a blurring of the planned dose as seen by the target volume [25].

Random errors originate mainly from the fact that the daily anatomy and its position varies. As discussed, this results in both systematic as well as random errors. In-room imaging artefacts, registration errors, limitations to setup corrections and/or technical limits of the accelerator further contribute to random errors. Also, anatomy changes may occur on a transient, daily basis, e.g. obstruction of air cavities.

Changes during the treatment session, intra-fraction errors, are usually considered small in RT of HNC (well below 2 mm per direction (1 SD) [27–29]), and therefore not covered in this thesis. A special case of an intra-fraction motion is swallowing, which results in large movements with low frequency and short duration [30].

Margins

Despite IGRT, residual geometrical uncertainties during RT of HNC remain that need to be accounted for with safety margins. The size of safety margins depends on the magnitude of the geometric uncertainties. Margin recipes generally distinguish between systematic and random errors [25]. A popular margin recipe, introduced by van Herk et al. [31], weighs systematic and random errors as follows:

\[ M = 2.5 * \Sigma + 0.7 * \sigma \] (1.1)
This recipe ensures that, if a margin $M$ is used to expand the CTV to the PTV, and given a normal distribution of systematic and random errors with standard deviations $\Sigma$ and $\sigma$, that for 90% of patients the CTV will receive at least 95% of the dose prescribed to the PTV. Assumptions in this recipe are a single, uniform and highly 3D conformal dose distribution around a rigid CTV, delivered with radiation beams with an effective penumbra (dose fall-off) of 3.2 mm. In case of multiple error sources, standard deviations should be added quadratically. Note that systematic errors are considered 3-4 times as important as random errors.

Reduction of uncertainties will allow smaller margins and hence, with modern planning and delivery techniques, reduce the treated volumes. Margin reduction might therefore lead to less toxicity. Alternatively, margin reduction allows dose escalation to the tumor, until the same levels of toxicity are reached as with larger margins. Reduction of uncertainties during treatment could result from improved fixation, better patient setup, or plan adaptation to anatomy changes.

**In-room imaging**

In room imaging visualizes the anatomy of the patient during or prior to treatment. The acquired images are used to verify the position of the target volume in relation to the treatment plan. With radiation detectors opposite to the LINAC-head, the transmission of the photon beam through the patient is measured (portal imaging). Due to the high energies (MeV), most photons pass through the patient un-reflected and scatter only at dense objects. Therefore, for soft-tissue imaging the contrast of MeV photons is usually insufficient. In-room imaging techniques that use kV photons may improve soft tissue contrast. Examples of in-room imaging with kV photons are stereoscopic imaging (2-dimensional) [32], in-room CT imaging [33], or cone beam CT (CBCT) imaging [34–37] (both 3-dimensional). With the CBCT system, a kV x-ray source and detector are mounted perpendicular to the treatment beam to acquire projection images (2D), see figure 1.4. Subsequently, a 3D volume of the patient can be reconstructed from a series of projection images that is acquired while rotating around the patient [38, 39].

**IGRT**

Acquisition of images of the anatomy of the patient prior to treatment has led to IGRT, where the patient’s anatomy is scanned and the tumor, or a surrogate, is localized. The images are registered with the planning CT in order to determine the setup error and to derive a couch correction. See figure 1.5 for an example of CBCT-to-CT rigid registration on a large region of interest with bony anatomy. Daily correction of setup errors just prior to delivery (online corrections) allows for correcting both systematic and random errors. With offline corrections, scans are acquired for the first few fractions to estimate the systematic (average) component of errors, which is the basis of correction for the remainder of the treatment [40].

With online corrections, residual errors also have to be expected. For solid tumors, behaving rigidly, and assuming full correction of rotational errors, near perfect alignment is achievable.
In practice however, especially during radiotherapy of HNC, target areas are large and deform as a result of neck flex or shoulder misplacement, or show changes from regression or weight loss. In those cases, a couch correction can only partially align the target areas. In fact, how to arrive at an optimal couch correction in deforming anatomy was not investigated prior to this thesis. Moreover, rotational errors are often not corrected: only the translation component is corrected. Finally, not only correct placement of the target volume is important, OARs may shift into the path of treatment beams where they receive a higher dose themselves and/or alter the radiobiological path length. In many cases IGRT for RT of HNC will not be capable to correct all misalignments and adaptive approaches are required.

ART

The aim of adaptive radiotherapy (ART), as defined by D. Yan [41], is 'to customize each patient's treatment plan to patient-specific variation by evaluating and characterizing the systematic and random variations through image feedback and including them in adaptive planning'. Typically, after observed anatomical changes, a repeat CT scan is made, and targets and OARs are re-delineated. Next, the treatment plan is re-optimized in the repeat CT scan. The new plan thereby mitigates geometrical errors and optimizes the dose distribution to the new position, volume and shape of target volumes and OARs. ART is frequently applied to prevent deterioration of
Figure 1.5. Example of IGRT process: green/purple overlay of a CBCT (purple) over the planning CT (green). The CBCT is registered with the planning CT on the bony anatomy in the rectangular box. Misalignments after setup correction are present due to patient deformation, as can be seen from the green/purple edges at the vertebrae, hyoid, mandible and base of the skull.

the dose distribution due to deformations, anatomy changes, and tumor regression. In contrast, geometrical discrepancies may also lead to opportunities to improve the dose distribution, especially after tumor regression. However, whether target volumes may be reduced depends on how microscopic extensions move with tumor regression. So far, no studies were performed that investigated deformations of the edges of the CTV.

Adaptive planning represents a substantial amount of work. Generally, the increase in workload of ART in the clinic depends on the amount of patients that have been identified for replanning by an ART protocol (patient selection), timing of repeat CT scanning (ad-hoc or scheduled up front), and frequency of adaptive planning. Deformable image registration (DIR) is an important tool to support adaptive procedures. With DIR, the non-rigid or elastic transformation that brings two scans into alignment is determined. With this transformation, the contours from the initial
Chapter 1

scan can be transferred to the new scan. ART workload could substantially be reduced if in-room imaging with e.g. CBCT could replace the acquisition of the repeat CT. Whether CBCT quality is sufficient to form the basis of adaptive re-planning, remains to be investigated.

This thesis

At the start of the work for this thesis, CBCT was just introduced into the clinic. CBCT acquisition and reconstruction protocols were being developed and optimized for various tumor sites. New patient setup protocols were gradually introduced into clinics that were using CBCT scans. However, setup correction protocols tailored to radiotherapy of HNC were not yet developed. Plan adaptation took place infrequently, after ad hoc observed anatomy changes, mostly to prevent extra OAR dose. Development of DIR to support adaptive procedures was in its early days.

Purpose

The aim of the work in this thesis was to investigate the use of in-room imaging with CBCT to quantify and reduce geometrical uncertainties during treatment delivery for HNC, including tumor regression. In addition, we set out to investigate the potential of ART with CBCT for radiotherapy of HNC. The specific objectives in this thesis were:

- To quantify geometrical errors during radiotherapy of HNC;
- To derive an optimal patient setup in deforming anatomy with IGRT;
- To develop procedures for adapting the treatment plan using CBCT to measure residual errors uncorrected by patient setup.

Outline

Novel use of CBCT for patient positioning in radiotherapy of HNC is subject of chapter 2. Posture changes, where bony anatomy may move with respect to each other, are measured with rigid registration of small regions with individual bony parts (multiple region of interest (mROI) registration). We quantify the local residual errors and compare them to global patient setup. We also discuss to what extent deformations are taken into account in safety margins, and how focusing setup on a single region may alter uncertainties.

In radiotherapy for HNC, tumor response is frequently seen, yet little quantitative description exists. Therefore, in chapter 3, we study tumor shape variability based on the position of helical gold markers that were surgically implanted at the edges of the visible tumor. These markers are identified in the planning CT as well as in daily CBCT after alignment on surrogate bony anatomy.
Residual movement, time trends, and correlations between markers are investigated. Discussion focuses on appropriate margins, and whether markers can identify the type of GTV reduction (elastic or dissolving regression), which is of importance for decisions on how to adapt the target volume.

Contrast in (CB)CT is generally insufficient to directly assess CTV changes during treatment except for clear boundaries like the air-tissue interface. A better modality to assess tumor changes is MRI, in which GTV changes are well visible. In chapter 4 we address the question whether GTV reduction in MRI is in line with CTV reduction based on gold marker behavior observed in CBCT.

The topic of chapter 5 is the development of two patient positioning strategies using mROI registration. In the first strategy, the mean error is minimized, while in the second the maximum error is minimized. We compare both strategies with clinical practice, and investigate how the residuals after correction can be used to monitor systematic deformations to act as a trigger for adaptive interventions.

Chapter 6 describes the clinical introduction and the first experience with the mROI registration and correction method for patient positioning. Training of technicians, ease of use, and workload are reported. A warning system is introduced that alerts users on systematic deformations that allows objective selection of patients that may benefit from adaptive replanning.

Though limited in view and contrast, CBCT is widely available, and contains a vast amount of anatomical information. Deformable image registration may quantify anatomical variation. Validation of a CBCT-to-CT deformable registration algorithm is the topic of chapter 7, with the focus on the registration of tumor changes. Validation of soft tissue alignment is performed on anatomical landmarks. Observer induced variability in landmark identification is filtered out with pair-wise analysis of variance. Accuracy of registration of the tumor is assessed via gold markers.

In chapter 8, we introduce patient anatomy modeling as a method to resolve residual local misalignments with ART. Therefore, non-rigid CBCT to CT registrations are used. With the deformation fields, new anatomies are generated from the planning CT. We explain a new method that allows for correcting systematic local misalignments. Improvements in residual errors are quantified for different adaptive scenarios.

In chapter 9 we investigate how large safety margins should be for deformations in treatment plans with realistic VMAT dose distributions. First, we investigate at which margins treatment plans become sensitive for deformations. Next, by simulating adaptive interventions based on patient anatomy modeling, we investigate if target coverage can be restored.

Chapter 10 summarizes the most important findings from this thesis and discusses the future use of IGRT and ART strategies in radiotherapy of HNC.
References


Introduction


Setup uncertainties of anatomical sub-region in head-and-neck cancer patients after offline CBCT guidance

S. van Kranen, S. van Beek, C. Rasch, M. van Herk, J.-J. Sonke

Department of Radiation Oncology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

E-mail: j.sonke@nki.nl

Published April 2009 in International Journal of Radiation Oncology, Biology and Physics, volume 73(5), pages 1566-1573

dx.doi.org/10.1016/j.ijrobp.2008.11.035
Chapter 2

Abstract

Purpose
To quantify local geometrical uncertainties in anatomical sub-regions during radiotherapy for head-and-neck cancer patients.

Methods & Materials
Local setup accuracy was analyzed for 38 patients, who had received intensity-modulated radiotherapy and were regularly scanned during treatment with cone beam computed tomography (CBCT) for offline patient setup correction. In addition to the clinically used large region of interest (ROI), we defined eight ROIs in the planning CT that contained rigid bony structures: the mandible, larynx, jugular notch, occiput bone, vertebrae C1-C3, C3-C5, and C5-C7, and the vertebrae caudal of C7. By local rigid registration to successive CBCT scans, the local setup accuracy of each ROI was determined and compared with the overall setup error assessed with the large ROI. Deformations were distinguished from rigid body movements by expressing movement relative to a reference ROI (vertebrae C1-C3).

Results
The offline patient setup correction protocol using the large ROI resulted in residual systematic errors (1 SD) within 1.2 mm and random errors within 1.5 mm for each direction. Local setup errors were larger, ranging from 1.1 to 3.4 mm (systematic) and 1.3 to 2.5 mm (random). Systematic deformations ranged from 0.4 mm near the reference C1-C3 to 3.8 mm for the larynx. Random deformations ranged from 0.5 to 3.6 mm.

Conclusions
Head-and-neck cancer patients show considerable local setup variations, exceeding residual global patient setup uncertainty in an offline correction protocol. Current planning target volume margins may be inadequate to account for these uncertainties. We propose registration of multiple ROIs to drive correction protocols and adaptive radiotherapy to reduce the impact of local setup variations.
Introduction

The neck is a complex site for radiotherapy. Often the dose to the target is compromised by the nearby organs at risk (OAR). The introduction of intensity-modulated radiotherapy for this group of patients is highly beneficial \[1\]. The steep dose gradients allow for a more conformal target dosage while optimally sparing OARs.

In practice, however, the benefits of conformal radiation techniques may be limited by the accuracy with which the setup of the patient during treatment delivery can be reproduced. Alignment accuracy depends on the skill and experience of physicians or technicians while at the same time skin marks have a limited ability in predicting internal anatomy. Furthermore, the neck is flexible and needs immobilization with masks, bite blocks, or vacuum pillows \[2–4\]. Apart from patient setup accuracy, head-and-neck cancer patients show substantial anatomical changes associated with tumor regression and weight loss \[5\]. Intrafraction organ motion introduces additional uncertainty \[6\].

As the position and shape of the tumor vary over the course of treatment, safety margins are applied to avoid target underdosage \[7\]. A reduction in geometrical uncertainties allows smaller margins, enabling a higher deliverable dose to the target, a lower dose to OARs, and improves clinical outcome \[8\].

To reduce geometrical inaccuracies many institutes have implemented setup correction protocols based on electronic portal imaging devices \[9, 10\]. Ideally, the position of the tumor and OARs are determined in these images and repositioned to their planned locations immediately before treatment. Because of poor image quality of portal images and overlapping anatomy, identification of the tumor and OARs is generally not possible and nearby bony structures are taken as a surrogate. This is justified by the assumption that movement of soft tissue is highly correlated to movement of nearby bony structures.

Patient setup protocols currently are based on the assumption of rigid body motion: the patient is considered as nondeformable and table shifts and optional rotations are used to correct patient setup errors. Under this assumption, an online patient setup protocol for interfractional movement will result in small margins. In practice, however, patients with head-and-neck cancer show changes in posture and anatomy, giving rise to residual geometrical uncertainties, for which larger margins should be used.

In our institute, isotropic margins of 5 mm are used for patients with head-and-neck cancer in combination with an offline setup correction protocol. Since 2004, cone beam computed tomography (CBCT) image guidance (Elekta Synergy, Elekta Oncology Systems Ltd, Crawley, UK) replaced portal imaging for setup verification. The volumetric imaging capacity of CBCT revealed frequent changes in anatomy and posture that were previously hard to detect with planar imaging. There is no general approach to cope with such changes: deformations are simply ignored or in severe cases an ad-hoc decision for replanning is made. This study aims at
quantifying the occurrence and magnitude of such changes during treatment and to assess the consequences for our currently used margins.

**Methods and Materials**

**Patient group**

In this study, 38 patients with head-and-neck cancer were retrospectively selected irrespective of tumor stage in consecutive order. Patients were immobilized with a five-point thermoplastic fixation mask with shoulder fixation (Civco Medical Solutions, Kolona, USA). To maximize the distance between radiation field and the parotid glands, the head was tilted back as far as possible with the help of a patient-specific neck rest. All patients were given a knee support for stability and comfort. Eight patients used a customized bite block to position the palate outside the radiation field.

**Daily setup errors: rigid body registration**

The planning CT scan was reconstructed with in-slice resolution of $0.8 \times 0.8 \text{mm}^2$ and slice distances of 3 mm over the tumor area and 5 mm outside this region. For routine offline patient setup verification, CBCT scans were acquired immediately before treatment delivery at regular intervals during the full course of treatment. A total of 312 scans were available, on average 8.2 scans per patient. The CBCT scans were reconstructed with a resolution of $1 \times 1 \times 1 \text{mm}^3$. In this article we will use the term *reference scan* when referring to the planning CT scan, whereas the daily CBCT scan will be indicated as *localization scan*.

The global setup error was determined by registration of the bony anatomy from the reference scan to the localization scan using chamfer matching [11]. The scans were registered based on an extensive ROI that included the cervical vertebrae. Inclusion of the jugular notch or mandible depended on the position and size of the tumor.

The registration resulted in a set of transformation parameters commonly expressed by a $4 \times 4$ transformation matrix $E$. In the absence of reflections, shear, and scaling, $E$ described a rigid body transformation, consisting of a rotation over the $x$, $y$, and $z$-axis followed by a translation in the $x$-, $y$-, and $z$-direction. A setup error correction vector was then derived by calculating the shift of a clinically relevant point (the correction-reference point), often the center of the planning target volume.

**Patient setup correction protocol**

For clinical setup verification, we applied a shrinking action level (SAL) setup correction protocol [12]. This protocol tested the mean setup error vector length against a shrinking
Local setup errors in head & neck radiotherapy

action level $\alpha/\sqrt{N}$, with $N$ the number of consecutive measurements ($N_{\text{max}} = 2$) and $\alpha = 5 \text{ mm}$ the initial action level. If the mean vector length exceeds the action level, a correction was made for the subsequent fractions and the protocol was restarted. Otherwise, the setup was checked weekly. Only table corrections exceeding 3 mm were actually performed to limit the workload in the absence of an automatic couch system.

Multiple regions of interest analysis

For each patient, we defined eight rectangular shaped ROIs in the reference scan containing bony structures: the mandible, larynx, jugular notch, occiput bone, neck vertebrae C1-to-C3, C3-to-C5, and C5-to-C7, and the vertebrae caudal of C7. In figure 2.1, these definitions are applied to the reference scan of an example patient. For each ROI, the bony anatomy was automatically segmented using histogram-based thresholds in both CT and CBCT and subsequently registered using chamfer matching [11]. Day-to-day local displacements of the bony structures were calculated with the geometrical center of each ROI as a correction-reference point.

Figure 2.1. Definition of regions of interest (ROIs) in sagittal plane. Clockwise: the occiput bone, the upper, central, and lower part of the neck, the vertebra caudal of C7, the jugular notch, the larynx, and the mandible. The clinically used large ROI (dotted line) contains the cervical vertebrae and is extended posteriorly depending on the position of the tumor.
Chapter 2

Calculation of deformations

In contrast to rigid movements, deformations do not preserve mutual distances and orientations. To quantify local translation and changes in orientation and to facilitate easy comparison between patients, a clearly identifiable reference structure was selected: vertebrae C1-C3. This choice was motivated by the fact that this ROI often takes a central position to both target and OARs. Furthermore, this part of the neck is supported by a neck rest, and its position was expected to be quite reproducible.

The multiple ROI analysis resulted in local registrations, each describing a setup error (including rotations). Knowing the setup error \( E_0 \) for the reference structure, we applied its correction with the inverse of \( E_0 \). In the presence of deformations this correction did not fully align other ROIs. A residual setup error remained due to deformations with respect to the reference structure, mathematically calculated by \( E_0^{-1} \cdot E_2 \) (e.g., for structure 2). The residual setup error vector was found with:

\[
d_2 = (E_0^{-1} \cdot E_2 - I) \cdot s_2
\]

(2.1)

with \( s_2 \) the center of ROI-2 in the reference scan and \( I \) the unity matrix. We refer to this residual setup error vector \( d_2 \) when we use the term deformation. Figure 2.2 schematically shows the process of registration and correction for a reference structure, resulting in a residual setup error vector. Note that deformations are not affected by the choice of setup correction protocol. With this approach, we focused on the change in distance between two objects, whereas we disregarded changes in mutual orientation. We report both the local setup accuracy of bony anatomy in each ROI after setup corrections (including rotations), as well as the motion of the bony structures referenced to the vertebrae C1-C3.

Statistical analysis of deformations

For each patient \( p \), all ROIs were registered in the available localization scans. From the local setup errors, we calculated the deformations for all structures \( s \). The patient systematic deformation \( \Sigma_{p,s} \) for ROI \( s \) is the mean displacement over all fractions. The patient random deformation \( \sigma_{p,s} \) is the standard deviation (SD) of the displacements over all fractions. The group mean error \( M_s \) for a structure \( s \) is found by averaging \( \Sigma_{p,s} \), whereas the systematic error \( \Sigma_s \) follows from the SD of \( \Sigma_{p,s} \). The random error \( \sigma_s \) is calculated as the root-mean-square value over all \( \sigma_{p,s} \). Each axis was treated independently. This approach for quantification of deformations is similar to reporting conventional setup uncertainties [13].

Because of a limited field of view of the CBCT scans in the craniocaudal (CC) direction, the jugular notch was not present in the localization scans for one patient. For a second patient, the larynx could not be accurately determined in the reference scan. Both structures for these patients were not scored in the statistical analysis.
Local setup errors in head & neck radiotherapy

Figure 2.2. Deformations are calculated by correcting the setup error $E_0$ (dotted line) of the reference structure 0. Applying the inverse of $E_0$ to structure 2 (with setup error $E_2$) will in general not realign structure 2 with its original position. The residual error vector $d_2$ is the deformation with respect to the reference structure.

Margins

Because margin recipes for deforming (non-rigid) anatomy have not been derived (see Discussion), we applied the margin recipe for rigid body setup inaccuracies to each ROI as a first-order approximation [14]:

$$m = 2.5\Sigma + 0.7\sigma$$  \hspace{1cm} (2.2)

with $\Sigma$ the systematic setup error and $\sigma$ the random error. In this approximation, the recipe applies to only one single ROI at the time.

Correlations

Not all movements are physically possible because of restrictions imposed on patients by muscles and fixation devices. Therefore, correlations in non-rigid motion between certain structures can be expected, which were quantified by the Pearson product moment correlation coefficient $r$ between pairs of structures.
Chapter 2

Time trends

For time trends in deformations, we determined the linear regression between the position of a ROI and the elapsed time after start of treatment, calculated per patient and for each axis separately. As statistical test a Student t test was used in which the level of significance was corrected for the fact that multiple groups were compared. This is achieved with a Bonferroni correction: the initial significance level $\alpha_0$ is reduced with the number of test groups $k = 38$ to $\alpha_t = 0.05/38(\sim 0.0013)$.

Results

Residual errors after correction protocol

The residual rigid body setup error assessed by registration of the clinically applied ROI is shown in table 2.1. The group mean error of $-0.5$ mm in the anteroposterior (AP) direction was statistically significant ($p = 0.020$), but small compared with the systematic error. The group mean rotations over the left-right (LR) and CC axes were also statistically significant ($p < 0.001$ and $p = 0.006$). Based on 312 registrations, a total of 24 corrections were clinically applied; in the LR, CC, and AP directions, the numbers of corrections were respectively 9, 11, and 12. Some corrections were applied simultaneously. For 8 patients, one correction was required, 2 patients had two corrections, 4 had three corrections.

Table 2.1. Overall patient setup accuracy after corrections by the offline shrinking action level protocol, Group mean $M$, systematic $\Sigma$ and random error $\sigma$ result from 312 measurements for 38 patients.

<table>
<thead>
<tr>
<th>M</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displacements(mm)</td>
<td>0.1</td>
<td>0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Rotations(°)</td>
<td>0.7</td>
<td>0.4</td>
<td>-0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Σ</td>
<td>1.2</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Rotations(°)</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>σ</td>
<td>1.4</td>
<td>1.4</td>
<td>1.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right. $M$ = group mean error; $\Sigma$ = systematic error; $\sigma$ = random error.
The registrations were performed for a large region of interest with various bony structures.

In table 2.2, we report the local setup accuracy of the eight ROIs as measured after shrinking action level offline corrections thus as clinically treated. We visualized setup uncertainties in an error map in figure 2.3 (sagittal view), group mean errors are indicated with vectors and systematic and random errors (1 SD) correspond to the semimajor and semiminor axes. The position of the ROIs in this figure is an approximation. The largest systematic errors were found for the mandible and larynx in the CC direction: $2.2$ mm and $3.4$ mm, respectively, and the occiput bone and the ROI caudal of C7 in the LR direction, $2.2$ mm and $2.7$ mm. After correction, the
position of the upper vertebrae C1-C3 was the most reproducible. The group mean errors of the occiput bone, mandible, jugular notch, and caudal C7 suggest a group mean clockwise rotational error around the LR axis consistent with the rotations observed in clinical ROI registrations.

Figure 2.3. Error map of local regions of interest (ROIs) in the sagittal plane. Group mean errors are indicated with vectors, systematic and random errors with dotted and straight ellipse (semimajor and semiminor axes correspond to 1 SD). The picture in the background serves as a map showing the approximate position of the different ROIs. Errors were scaled by a factor 5.

Changes in orientation of the multiple ROIs are summarized in table 2.2. In line with the larger displacements found for the local ROIs compared with clinical patient setup data, these rotations are also larger, but still small.

In figure 2.4, we scored the cumulative incidence of the patient mean setup error vector length for four selected ROIs before and after corrections. Graphs for other ROIs were similar. Note that the correction protocol improves the setup accuracy for all ROI’s except the larynx (see Discussion), but only the clinically used ROI results in apparent setup errors below 5 mm for all patients.
Table 2.2. Local residual setup errors after corrections by the offline shrinking action level protocol.

<table>
<thead>
<tr>
<th></th>
<th>LR (mm)</th>
<th>Σ (mm)</th>
<th>σ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>AP</td>
<td>CC</td>
</tr>
<tr>
<td>Mandible</td>
<td>0.1</td>
<td>1.9</td>
<td>-1.9</td>
</tr>
<tr>
<td>Larynx</td>
<td>0.4</td>
<td>0.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>1.0</td>
<td>0.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Occiput bone</td>
<td>-0.5</td>
<td>-0.1</td>
<td>-1.6</td>
</tr>
<tr>
<td>C1-C3</td>
<td>-0.4</td>
<td>-0.1</td>
<td>-1.5</td>
</tr>
<tr>
<td>C3-C5</td>
<td>-0.3</td>
<td>0.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>C5-C7</td>
<td>-0.1</td>
<td>-0.4</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LR (mm)</th>
<th>Σ (mm)</th>
<th>σ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>AP</td>
<td>CC</td>
</tr>
<tr>
<td>Mandible</td>
<td>-0.5</td>
<td>0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>Larynx</td>
<td>0.3</td>
<td>-0.1</td>
<td>-0.7</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>-0.7</td>
<td>0.4</td>
<td>-0.1</td>
</tr>
<tr>
<td>Occiput bone</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>C1-C3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>C3-C5</td>
<td>0.4</td>
<td>0.5</td>
<td>-0.4</td>
</tr>
<tr>
<td>C5-C7</td>
<td>0.8</td>
<td>0.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>Caudal C7</td>
<td>0.7</td>
<td>0.3</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right. $M = \text{group mean error}$; $\Sigma = \text{systematic error}$; $\sigma = \text{random error}$; SAL = shrinking action level.

Deformations referenced to vertebrae C1-C3

Deformations are summarized in table 2.3. The residual error for the reference C1-C3 is by definition zero in all directions. Again the systematic error in the CC direction for the mandible and the larynx was substantial: 2.5 mm and 3.8 mm, respectively. Deformations are presented in an error map in figure 2.5.

Table 2.3. Statistics of deformations referenced to ROI with vertebrae C1-C3.

<table>
<thead>
<tr>
<th></th>
<th>LR (mm)</th>
<th>Σ (mm)</th>
<th>σ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>AP</td>
<td>CC</td>
</tr>
<tr>
<td>Mandible</td>
<td>0.0</td>
<td>1.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>Larynx</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>1.0</td>
<td>0.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Occiput bone</td>
<td>0.4</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>C1-C3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>C3-C5</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>C5-C7</td>
<td>0.9</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right. $M = \text{group mean error}$; $\Sigma = \text{systematic error}$; $\sigma = \text{random error}$. 
Local setup errors in head & neck radiotherapy

**Figure 2.4.** Cumulative score of mean setup error vector length before and after offline corrections. The correction protocol reduces the mean setup error to within 5 mm for the region of interest (ROI) by which it is driven (clinical ROI). Other ROIs, for instance the larynx, do not necessarily benefit as much and the amount of patients for which the mean setup error vector exceeds 5 mm is still considerable.

**Margins**

In table 2.4, the local anisotropic margins are summarized that would be required for the separate treatment of the individual ROIs with the clinically applied setup correction protocol. For almost all regions, the clinically applied isotropic margin of 5 mm was too small. The last row shows the margins that result from the data from the clinically measured setup errors, thus based on the large ROI.

**Correlations**

In figure 2.6, the correlation plot for the deformations referenced to C1-C3 is depicted, indicating considerable positive and negative correlations in movement between several anatomical regions. Note that the figure is mirrored over the diagonal. In table 2.5, the correlation coefficients $r$ for the bony structures surrounding the larynx are collected.

**Time trends**

In table 2.6, the number of patients with a statistically significant time trend is scored. Significant time trends for more than one patient were only found in the CC direction, for the larynx, occiput
Figure 2.5. Error map of deformations with respect to the reference region of interest, vertebrae C1-C3, in the sagittal view. Note the increased motion with longer distances from the reference.

bone, vertebrae C5-C7, and caudal C7. The extreme values —0.34 mm/day both for the jugular notch and caudal C7 in the AP direction were found for the same patient who, on retrospective analysis, had shown 10 kg weight loss during treatment. This patient was also scored in two categories for the CC direction.

Discussion

Local setup accuracy

Rigid registration of sub-regions of bony anatomy showed that local setup errors exceed the residual global patient setup errors in an offline correction protocol. Consequently, quantifying the overall setup accuracy overestimates the precision of radiotherapy in head-and-neck cancer patients.

The bony anatomy was registered with chamfer matching using segmented bone in the reference and localization scan. The clinical ROI containing multiple relevant structures underestimated the resulting local setup uncertainties. This implies that within the registration algorithm, movement
Table 2.4. First-order approximation of local anisotropic margins calculated with formula 2.2, required for adequate target coverage based on setup accuracies after SAL offline corrections. Group mean errors were small compared with systematic errors and therefore ignored.

<table>
<thead>
<tr>
<th>Structure</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandible</td>
<td>3.9</td>
<td>6.7</td>
<td>5.5</td>
</tr>
<tr>
<td>Larynx</td>
<td>4.6</td>
<td>10.3</td>
<td>5.1</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>6.3</td>
<td>5.7</td>
<td>6.0</td>
</tr>
<tr>
<td>Occiput bone</td>
<td>7.0</td>
<td>5.5</td>
<td>6.0</td>
</tr>
<tr>
<td>C1-C3</td>
<td>4.7</td>
<td>3.8</td>
<td>4.2</td>
</tr>
<tr>
<td>C3-C5</td>
<td>5.1</td>
<td>4.0</td>
<td>4.8</td>
</tr>
<tr>
<td>C5-C7</td>
<td>5.9</td>
<td>5.4</td>
<td>6.0</td>
</tr>
<tr>
<td>Caudal C7</td>
<td>8.3</td>
<td>6.2</td>
<td>6.7</td>
</tr>
<tr>
<td>Clinical ROI</td>
<td>4.0</td>
<td>3.7</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Abbreviations: AP=anterior-posterior; CC=cranio-caudal; LR=left-right; SAL=shrinking action level.

Table 2.5. Correlation in deformations between larynx and 3 nearby bony structures. Note that these structures cannot completely account for the local position variability of the larynx.

<table>
<thead>
<tr>
<th>Larynx</th>
<th>C3-C5</th>
<th>Mandible</th>
<th>Jugular notch</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.3</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>CC</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>AP</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right.

of a particular structure is "compensated" for by opposite movements of other structures. The resulting shifts and rotations become an average and therefore tend to underestimate true local changes.

Furthermore, bony structures such as the larynx or hyoid are small and therefore contribute very little to the distance criterion in the chamfer matching algorithm: their influence on the registration is therefore limited, which is noticeable in the setup accuracy of the larynx compared with the vertebrae.

Reported geometrical uncertainties always contain a measurement error. To estimate an upper limit for the measurement uncertainty, we evaluated differential motion of C3-C5 relative to C1-C3 assuming that all variability was due to measurement errors (no deformation). The upper limit was: 0.7 mm, 0.5 mm, and 0.6 mm (1 SD) in the LR, CC, and AP directions.
Chapter 2

Figure 2.6. Color-plot of the correlation coefficient for movement of pairs of structures, magnitude is indicated by the color scale. The correlation was calculated from the deformations with reference to the region of interest with C1-C3. The three orthogonal directions were treated as independent variables.

Deformations

The calculated deformations show that bony structures are able to move with respect to each other (i.e., the neck shows flexibility despite immobilization). If all movements are referenced to C1-C3, the residual setup error of nearby structures were almost absent, but increased with distance, which is visible in the error map in figure 2.5. This indicates that movement only correlates to nearby structures. Deformations could exceed residual errors within the correction protocol, probably caused by differences in mutual orientation, for example, when the neck is rotated on the neck-rest, but the mask prevents displacements of the mandible.

Zhang et al. [15] performed a similar study with repeat CT data for 14 patients and three ROIs. They limited their analysis to the reproducibility of patient positioning and did not distinguish between the rigid and non-rigid contribution in local setup accuracy. Comparing their setup data for vertebrae C2 and C6 with our data for vertebrae C1-C3 and C5-C7 shows that their systematic errors were larger, which could be due to the fact that their results were collected before setup corrections were applied.

Correlations

Residual movement is restricted because of muscles between bony structures and the fixation mask, thereby inducing correlations between nearby structures (figure 2.6). For instance, nodding
Table 2.6. Score of the number of patients with a statistically significant time-trend in deformations ($p < 0.05/38$, Bonferroni corrected). For each patient a linear regression was performed over the position of the bony anatomy versus elapsed time after start of treatment. In brackets the range of regression as min/max in mm/day is indicated.

<table>
<thead>
<tr>
<th>Number of patients with time trend</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandible</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Larynx</td>
<td>0</td>
<td>2 (0.09/0.14)</td>
<td>1 (-0.14)</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>0</td>
<td>1 (0.10)</td>
<td>1 (-0.34)</td>
</tr>
<tr>
<td>Occiput bone</td>
<td>0</td>
<td>3 (-0.14/0.11)</td>
<td>0</td>
</tr>
<tr>
<td>C1-C3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C3-C5</td>
<td>0</td>
<td>0</td>
<td>1 (-0.05)</td>
</tr>
<tr>
<td>C5-C7</td>
<td>0</td>
<td>2 (0.08/0.09)</td>
<td>1 (-0.08)</td>
</tr>
<tr>
<td>C7-caudal</td>
<td>1 (-0.19)</td>
<td>2 (0.08/0.22)</td>
<td>1 (-0.34)</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right.

"yes," a movement in the CC direction of the occiput bone, is negatively correlated to the CC movement of the mandible. The same goes for nodding "no" in the LR direction. Other structures were less correlated with their surroundings. From table 2.5, it follows that the mandible accounts for 50% of the variations seen in the CC direction of the larynx, whereas vertebrae C3-C5 and the jugular notch account for only 0% and 20% of the variations. Movement of the larynx in the CC direction is therefore largely independent. Principle component analysis provides an alternative description, as suggested by Birkner [16].

Time trends in deformations

Although time trends in deformations were observed in individual patients, no obvious overall trends were found. The time trends that were considered statistically significant ($p < 0.0013$) were predominantly found in the CC direction, although the incidence was in 3 patients at most. From the modest incidence of time trends, we concluded that local setup errors with head-and-neck cancer patients predominantly occur randomly and cannot generally be attributed to progressive changes.

A study by Barker et al. [5] shows that progressive changes in soft tissue volumes and positions with head-and-neck cancer patients are often seen, especially in the presence of weight loss. In our test group, weight loss and tumor shrinkage were also clearly visible. In figure 2.7, such an example is shown with three CBCT scans taken at the start, halfway, and at the end of treatment. Still, we could not link such cases to obvious time trends in the position of bony anatomy except in one patient, implying that gain or loss of soft tissue cannot be derived from the position of bony anatomy. Consequently, local variability of soft tissue is likely to exceed those of nearby
Figure 2.7. Example of visible progressive anatomical changes in soft tissue: tumor shrinkage/weight loss in the neck area (coronal view). The cone beam computed tomography scans were taken at Day 1, 18, and 55. No significant time trends in bony anatomy displacements could be determined for this patient.

bony anatomy. Deformable registration based on image intensity is more appropriate to analyze soft-tissue deformations.

Margins

The clinically applied safety margins of 5 mm were adequate to cover the global patient setup uncertainty, ignoring other geometric uncertainties such as delineation errors. Results also show, however, that local setup uncertainties are equal to or larger than global patient setup uncertainty. A calculation of anisotropic margins based on local setup accuracies shows that the clinically used isotropic margin of 5 mm is only sufficient for the ROI containing the vertebrae C1-C3 or C3-C5. Because of deformations, larger geometrical uncertainties exist for which current margins are inadequate. This conclusion is supported by the score of patient mean setup error vector length for the four selected ROIs (figure 2.4), which shows that the number of patients with systematic setup errors exceeding 5 mm is considerable.

It is of interest what margins are required to account for the residual geometrical uncertainties present after image guidance radiotherapy including deformations. Published margins recipes, however, assume rigid body motion, and are therefore not applicable to the deformations observed in this study. After all, the probability that several partial independent moving targets receive the prescribed dose is less than the probability of each individual target receiving this dose separately. Calculated margins as in table 2.4 are therefore a lower limit, only valid in the theoretical situation where all ROIs move fully correlated as a single rigid object. In a realistic situation with partially independent displacements, these margins are too small. How to derive adequate margins for deformations is subject to current research and was not part of this study.

The CC positioning accuracy of the larynx was already large within the offline setup correction protocol ($\Sigma = 3.4$ mm, $\sigma = 2.5$ mm), and was not improved by repositioning of the reference structure. In fact, the systematic component increases to 3.8 mm. By calculating the deformations, we retrospectively simulated an online correction protocol based on the setup corrections
Local setup errors in head & neck radiotherapy

(translations + rotations) for the upper vertebrae C1-C3. Using the margin recipe, formula 2.2, and assuming that the position of the larynx is a valid substitute for the center of a possible tumor, this online approach would already require anisotropic margins of 11 mm in the CC direction (and 5 mm LR and 4 mm AP). Whether the underlying uncertainties result in a severe underdosage with our clinically applied 5 mm margins depends on the conformality of the actual dose distribution or boost near the larynx.

**Strategies to cope with deformations**

Performing a rigid registration in the presence of deformations can give ambiguous results because even after a successful registration not all structures will be aligned. This situation is difficult to distinguish from an unsuccessful registration. We therefore clinically implemented a technique in which we register multiple ROIs (of which the registrations are unambiguous) and average the results into a single couch correction.

Second, we propose an adaptive strategy to correct the systematic component of the deformations estimated based on a limited number of fractions. By modifying the original planning CT in accordance to these systematic deformations, a new modified CT is obtained that is in better agreement with the average patient’s anatomy than the original CT. Replanning on this modified CT will provide a plan with improved target dosage and lower dose to OARs over the full course of treatment. This approach, using a single adaptive intervention, is limited to non-progressive systematic changes in posture and anatomy, but can include corrections for global systematic setup inaccuracies. Based on the setup accuracy data from the offline correction protocol, we have estimated a threshold that would require adaptive replanning for 25% of the patients. This threshold to limit maximum systematic displacement in all three directions for all structures is 4.8 mm. Extensive study based on dosimetric criteria are required to refine such thresholds.

**Conclusions**

Head-and-neck cancer patients show large local setup variations, resulting from deformations, exceeding residual global patient setup uncertainty in an offline correction protocol, despite the use of immobilization devices. Margins based on global patient setup accuracy may therefore be unsafe. Both for correction strategies and margins design, deformations should be taken into account.
References


Evaluation of tumor shape variability in head-and-neck cancer patients over the course of radiation therapy using implanted gold markers

O. Hamming-Vrieze¹, S. van Kranen¹, S. van Beek¹, W. Heemsbergen¹, M. van Herk¹, M. van den Brekel²,³, J.-J. Sonke¹, C. Rasch¹

¹Department of Radiotherapy, Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, Netherlands
²Department of Head and Neck Surgery, Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, Netherlands
³Department of Otorhinolaryngology, Amsterdam Medical Centre, Amsterdam, Netherlands
⁴Department of Radiotherapy, Amsterdam Medical Centre, Amsterdam, Netherlands

E-mail: o.vriezegnki.nl

Published October 2012 in International Journal of Radiation Oncology, Biology and Physics, volume 84(2), pages e201-e207

dx.doi.org/10.1016/j.ijrobp.2012.03.014
Chapter 3

Abstract

Purpose
This study quantifies tumor shape variability in head-and-neck cancer patients during radiation therapy using implanted markers.

Methods and Materials
Twenty-seven patients with oropharyngeal tumors treated with (chemo)radiation were included. Helical gold markers (0.35 × 2 mm, 3-10/patient, average 6) were implanted around the tumor. Markers were identified on planning computed tomography (CT) and daily cone beam CT (CBCT). After bony anatomy registration, the daily vector length on CBCT in reference to the planning CT and daily marker movement perpendicular to the gross tumor volume (GTV) surface at planning CT (<d_{normal}>) of each marker were analyzed. Time trends were assessed with linear regression of the <d_{normal}>, markers. In 2 patients, 2 markers were implanted in normal tissue to evaluate migration by measuring intermarker distances.

Results
Marker implantation was feasible without complications. Three-dimensional vectors (4827 measurements, mean 0.23 cm, interquartile ratio 0.24 cm) were highest in base of tongue sublocalization (P < .001) and bulky tumors (vectors exceeded 0.5 cm in 5.7% [0 – 20 mL], 12.0% [21 – 40 mL], and 21.7% [≥41 mL], respectively [P < .001] of measurements). The measured inward time trend in 11/27 patients correlated with the visual observed marker pattern. In patients with an outward trend (5/27) or no trend (11/27), visual observation showed predominantly an inhomogeneous pattern. Remarkably, in 6 patients, outward marker movement was observed in the posterior pharyngeal wall. The difference in distance between normal tissue markers (1 SD) was 0.05 – 0.06 cm without time trend, indicating that implanted markers did not migrate.

Conclusions
During head-and-neck radiation therapy, normal tissue markers remained stable. Changes in position of tumor markers depended on sublocalization and tumor volume. Large differences in marker patterns between patients as well as within patients were observed. Based on our study, the cranial and caudal border in the posterior pharyngeal wall are at highest risk to be covered insufficiently. Furthermore, implanted markers could help identify patients with an actual shrinkage of the GTV who might benefit from mid-radiation therapy redelineation to reduce toxicity.
Introduction

Intensity modulated radiation therapy (IMRT) has become the standard of care in the treatment of head-and-neck cancer patients. This technique enables highly conformal dose distributions around the planning target volume (PTV). The PTV should comprise the clinical target volume (CTV) with a margin to account for geometrical uncertainties such as delineation uncertainties, setup errors, and anatomical variability during radiation therapy [1].

Several strategies have been developed to reduce the margin needed. Observer variation is reduced with contouring guidelines [2] and computed tomography (CT) registration with positron emission tomography (PET) and/or magnetic resonance imaging. To reduce setup errors, patients are immobilized with masks, bite blocks, or vacuum pillows. Furthermore, rigid body setup correction protocols based on image-guided systems such as portal imaging devices or cone beam CT (CBCT) are widely implemented. Because soft-tissue contrast of these image-guided systems is insufficient to visualize tumor variability, these strategies are based on bony anatomy. Changes in tumor position, shape, and volume during treatment often go unnoticed and are generally not considered in the design of margin recipes.

Tumor variability may be related to tumor volume change, weight loss, fluid shift within the body, alteration in muscle mass, and fat distribution. With adaptive radiation therapy, repeated imaging is used to adapt the treatment plan to anatomical changes during treatment. The goal is to improve sparing of organs at risk and maintain complete coverage of the target volume [3]. To design the optimal adaptive radiation therapy strategy, detailed information about the extent and timing of anatomical changes of target volume and organs at risk are needed. Preliminary data on tumor position and shape change during radiation therapy have been reported [4, 5]. In these publications, the volume and position of the center of mass was analyzed with CT and PET imaging during the course of RT at several time points. They observed a gradual loss of gross tumor volume (GTV) and shift of the GTV center of mass [4, 5]. The pattern of local surface deformations resulting in center of mass shift has not been reported on. Furthermore, this raises the question if redelineation of the GTV and corresponding CTVs based on imaging mid-radiation therapy is allowed?

This study quantifies tumor shape variability in head-and-neck cancer patients during radiation therapy using implanted markers. In our hospital, helical-shaped gold markers are routinely implanted at the edge of the GTV and used to aid delineation in patients with head-and-neck tumors. These markers are detectable on daily CBCT and observed variability during radiation therapy was analyzed.
Methods and Materials

Patient cohort

Twenty-seven patients with transorally accessible tumors (15 tonsillar, 6 posterior pharyngeal wall [PPW], 6 base of tongue [BOT]), treated between June 2007 and June 2009, were prospectively selected in consecutive order. All patients had locally advanced disease without distant metastasis. Treatment consisted of radiation therapy (20%) or radiochemotherapy (80%) with curative intent. Radiation therapy dose was 70 Gy given in 35 fractions in 6 or 7 weeks. Concomitant chemotherapy consisted of cisplatinum, either 6 mg/m² daily in the first 5 weeks of treatment or 100 mg/m² on days 1, 22, and 43 of treatment.

Patients were immobilized with a thermoplastic mask with shoulder fixation. IMRT planning (Pinnacle, Philips Medical Systems, Eindhoven, The Netherlands) was performed on a planning CT scan (Somatom Sensation Open, Siemens AG, Erlangen, Germany) reconstructed with a slice distance of 3 mm. Implanted markers were identified on an additional reconstruction with a slice distance of 1 mm. Daily CBCT scans (Elekta Synergy, Elekta Oncology Systems Ltd, Crawley, UK) were acquired immediately before treatment delivery for patient setup, on average 31 per patient.

Marker implantation procedure

Before start of treatment, helical-shaped gold markers (0.35 × 2 mm) were routinely implanted to aid delineation. Implantation was done by a head-and-neck surgeon and a radiation oncologist together during investigation under anesthesia with a preloaded needle (VisicoilTM, RadioMed Corporation, Tyngsboro, MA, US). The markers were positioned at the edge of the tumor GTV at 5 mm depth. On average, 6 markers per patient were implanted (range 3-10). In 2 patients, 2 additional markers were implanted in contralateral normal tissue to assess migration.

Analysis of marker motion

The following analyses on marker motion were performed:

Position variability of each marker

The gold markers were semi-automatically identified on planning CT and daily CBCTs, meaning the center of mass was calculated after automatic segmentation of manually selected markers. Marker positions on CT and CBCTs were expressed in machine coordinates. First, in each patient, nearby bony anatomy registration of planning CT and CBCTs was performed. Subsequently, the residual daily marker motion of each marker was calculated in reference to the position of the
marker on planning CT. This resulted in a 3-dimensional (3D) vector with corresponding motion along the left-right axis, craniocaudal axis, and anteroposterior axis. The statistical package SPSS Statistics version 17.0 was used for data analysis. Differences between subgroups were tested using nonparametric tests (Mann-Whitney) for continuous variables. For categorical outcomes, the Pearson chi-square test was used. \( P < .05 \) was considered to be significant. Subgroup analyses were performed based on the location of the markers, location of the tumor, volume of the tumor, and treatment modality.

**GTV surface deformations**

The amplitude of the surface deformations was determined by decomposing the 3D vector of the residual daily marker motion into a component perpendicular to the original GTV \( (d_{normal}) \) and a parallel component. We reasoned that the perpendicular component is along the dose gradient, whereas the parallel component would not result in changes in dose. The perpendicular direction follows from the original position of the markers (in the planning CT) projected onto the surface of the GTV delineated on the planning CT (after smoothing). Negative values indicate inward movement and positive values indicate outward movement. The \( d_{normal} \) of each marker was calculated on every treatment day. Then, the mean \( d_{normal} \) of each marker was calculated. The maximum mean \( d_{normal} \) in each patient corresponds with the patient-specific systematic error. Furthermore, the mean of \( d_{normals} \) of all implanted markers in each patient on every treatment day \( (<d_{normal}>_{markers}) \) was calculated. Correlation of marker position with time was analyzed by linear regression of \( <d_{normal}>_{markers}, \) \( P < .1 \) was considered to be significant. We performed the statistical analysis of subgroups as in the previous paragraph.

**Patient-specific correlation of markers**

The correlation of markers in each patient was visually assessed. The marker motion pattern in each patient was classified as inward (all markers are visually moving to the center of the tumor), outward (all markers are visually moving away from the center of the tumor), inhomogeneous (different patterns are recognized in a patient), or stable (no pattern is visualized).

**Migration of normal tissue markers (2 patients)**

Helical-shaped markers were selected to minimize migration. To validate the stability of the marker position relative to the nearby tissue, the intermarker distance was measured for markers implanted in normal tissue. The underlying assumption is that 2 markers that are implanted at nearly the same location share the same tissue surrounding. Motion or deformation by edema or weight loss will affect the markers equally. Any residual motion between the markers will therefore indicate migration because the probability of migration in the same direction is small.
Chapter 3

Results

Marker implantation

Marker implantation was feasible without complications. On average, 6 markers per patient were implanted (range 3-10). Markers could easily be detected on CT and CBCT. Ninety-six percent (155) of the 162 implanted markers were identified on planning CT. During treatment, an additional 2 markers were lost, resulting in 153 evaluable markers. Identification of these markers on daily CBCTs (on average 31 per patient) resulted in 4827 measurements to assess daily marker motion.

Normal tissue markers

Migration of markers was negligible: the differences in distance between 2 normal tissue markers (1 SD) in 2 patients were 0.05 and 0.06 cm, respectively, without time trends.

Residual marker motion

The residual daily marker motion, after bony anatomy registration, is described as a 3D vector in table 3.1. The vector increased in time during treatment (Pearson correlation 0.227, $P < .001$). The occurrence of vectors exceeding 0.5 cm correlated significantly with tumor volume: 5.7%, 12.0%, and 21.7% of measurements in patients with tumor volumes of 0 − 20 mL, 21 − 40 mL, and ≥ 41 mL, respectively ($P < .001$). Furthermore, the vector is associated with tumor location. On average, BOT tumors have larger vectors during treatment compared with tonsillar ($P < .001$) and PPW tumors ($P < .001$) (figure 3.1).

Table 3.1. Residual daily marker movement of the 3D vector and on the LR, CC, and AP axis (cm).

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Interquartile range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector</td>
<td>0.01</td>
<td>2.18</td>
<td>0.29</td>
<td>0.23</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>LR</td>
<td>-1.80</td>
<td>0.98</td>
<td>0.01</td>
<td>0.20</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>CC</td>
<td>-1.04</td>
<td>0.99</td>
<td>0.00</td>
<td>0.22</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td>AP</td>
<td>-0.97</td>
<td>1.27</td>
<td>0.03</td>
<td>0.22</td>
<td>0.03</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Abbreviations: 3D = 3-dimensional; AP = anteroposterior; CC = craniocaudal; LR = left-right; SD = standard deviation.
GTV surface deformations

The residual daily marker motion, after bony anatomy registration, perpendicular to the surface of the tumor ($d_{\text{normal}}$) is described in table 3.2 and figure 3.1. $D_{\text{normal}}$ had an average of $-0.05 \text{ cm}$ (SD 0.21 cm) and showed both inward and outward movement (table 3.2).

<table>
<thead>
<tr>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>% of outward/inward measurements exceeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3 cm</td>
</tr>
<tr>
<td>$d_{\text{normal}}$</td>
<td>-1.57</td>
<td>0.72</td>
<td>-0.05</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Abbreviations: 3D = 3-dimensional; AP = anteroposterior; CC = craniocaudal; LR = left-right; min = minimum; max = maximum; SD = standard deviation.

The patient-specific systematic error was calculated for each patient. The margin needed to cover this patient-specific systematic error in 90% of our patients would be 0.23 cm.

On average, BOT tumors showed a wider range of $d_{\text{normal}}$ than tonsillar or PPW tumors (percentage of measurements exceeding 0.5 cm in either direction [inward and outward] were, respectively, 8.0% vs 1.8% and 4.8% ($P < .001$), figure 3.1). However, these differences disappeared when focusing at outward movement only (percentage of measurements exceeding 0.3 and 0.5 cm [outward movement only] were respectively 2.7% and 0.7% [BOT], 2.7% and 0.5% [tonsillar], 2.2% and 0.0% [PPW]).
Chapter 3

Linear regression of $< d_{normal} >_{markers}$ during treatment, showed in 11/27 (41%) patients a significant inward trend, in 5/27 (18%) patients a significant outward trend and in 11/27 (41%) patients no trend.

Correlation of the $< d_{normal} >_{markers}$ in time with tumor sublocalization, tumor volume, and type of treatment, is shown in table 3.3 and figure 3.2. In patients with large tumors, we observed more cases with time trends ($0 - 20 \text{ mL}$: 37%, $21 - 40 \text{ mL}$: 57%, and $\geq 41 \text{ mL}$: 75%; table 3.3) and a larger amplitude of measurements ($d_{normal}$ exceeded 0.5 cm in either direction in, respectively, 1.1% [0 – 20 mL], 2.2% [21 – 40 mL], and 6.8% [≥41 mL] of measurements [$P < .001$]). Shifts exceeding 0.2 cm in either direction were only observed after the second week of treatment.

Table 3.3. Correlation of tumor location, tumor volume, and treatment modality with time trend of $< d_{normal} >_{markers}$.

<table>
<thead>
<tr>
<th>Trend</th>
<th>n</th>
<th>Inward</th>
<th>Number</th>
<th>Outward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base of tongue</td>
<td>6</td>
<td>3 (50%)</td>
<td>2 (33%)</td>
<td>1 (17%)</td>
</tr>
<tr>
<td>Tonsillar region</td>
<td>15</td>
<td>7 (47%)</td>
<td>6 (40%)</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>Posterior pharyngeal wall</td>
<td>6</td>
<td>1 (17%)</td>
<td>3 (50%)</td>
<td>2 (33%)</td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤20 mL</td>
<td>8</td>
<td>1 (12%)</td>
<td>5 (63%)</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>21 – 40 mL</td>
<td>7</td>
<td>4 (57%)</td>
<td>3 (43%)</td>
<td>-</td>
</tr>
<tr>
<td>&gt;40 mL</td>
<td>12</td>
<td>6 (50%)</td>
<td>3 (25%)</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>8</td>
<td>3 (38%)</td>
<td>4 (50%)</td>
<td>1 (12%)</td>
</tr>
<tr>
<td>Chemo-Rt</td>
<td>19</td>
<td>8 (42%)</td>
<td>7 (37%)</td>
<td>4 (21%)</td>
</tr>
</tbody>
</table>

Abbreviations: Chemo-Rt = concomitant chemotherapy and R; RT = radiation therapy.

Patient-specific correlation of markers

The visually observed pattern of marker motion showed an overall inward movement of markers in 13/27 (48%) patients (figure 3.3, example 1), stable markers in 2/27 (7%) patients, and an inhomogeneous pattern in 12/27 (45%) patients. In patients with an inhomogeneous pattern, the majority (8/12) showed mixed inward and outward movement of markers. A remarkable finding was the outward movement in craniocaudal direction of markers in the PPW in 6 patients (figure 3.3, example 2). Furthermore, in 3 patients a clear outward movement followed by inward movement was observed (figure 3.3, example 3).

In table 3.4, for each patient the calculated time trend vs the visually observed pattern of marker motion is depicted together with the tumor location, volume, and treatment. In patients with an inward time trend the visual pattern was accordingly, for example, of the 11 patients with an inward time trend, 10 showed an overall inward movement and 1 partially an inward movement. In the 5 patients with an outward time trend, we observed 4 patients with an inhomogeneous response with local outward movement and 1 patient with stable markers.
Figure 3.2. $D_{\text{normal}}^{\text{markers}}$ plotted per patient as a function of treatment week, sorted by tumor group (first row) and by volume group (second row). Mean position: black dots. Tumor groups: red/asterisk = 0 – 20 mL, dark yellow/square = 21 – 40 mL, blue/triangle = 41 – 90 mL. Volume groups: green/triangle = base of tongue; gray/diamond = tonsillar region; purple/circle = posterior pharyngeal wall.

Discussion

This study found considerable tumor shape variability of head-and-neck carcinomas over the course of radiation therapy with implanted markers.

Markers can routinely be implanted before radiation therapy at the edge of the GTV and are detectable during radiation therapy on CBCT scans. Marker migration is negligible based on the helical-shaped structure and stable intermarker distances in normal tissue in our study and in reports by others [6, 7]. Therefore, we conclude that, after local setup registration, measured position variability of the markers reflects tumor variability.

We observed large differences between patients, but also within patients, of the amplitude and direction of marker motion along the dose gradient (both inward and outward movement perpendicular to the GTV $< d_{\text{normal}}^{\text{normal}} >$). This tumor variability may be related to actual increase or decrease of tumor volume; however, weight loss, fluid shift within the body, alteration in muscle mass, and fat distribution will also play a role. Although the majority of tumor marker motion showed inward movement as expected, some markers revealed outward movement. One would expect asymmetric tumor response with tumors expanding toward cavities and originating near rigid structures such as the jaw or vertebrae. Remarkably, however, outward movement repeatedly was observed in the PPW. An explanation might be that tumor response in the PPW leads to stretching of the wound edge. The low percentage (1/6 patients, 17%) of PPW tumors...
### Table 3.4.
Tumor characteristics, treatment modality, time trend of $d_{\text{normal}}$ markers, and marker motion pattern for each patient.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Tumor location</th>
<th>Volume (mL)</th>
<th>Treatment modalities</th>
<th>Time trend</th>
<th>Marker motion pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PPW</td>
<td>51</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Cranial inward, caudal outward* (figure 3.3, example 2)</td>
</tr>
<tr>
<td>2</td>
<td>PPW</td>
<td>14</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Cranial outward*, caudal inward</td>
</tr>
<tr>
<td>3</td>
<td>PPW</td>
<td>82</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Cranial outward*, lateral initially outward followed by inward medial inward</td>
</tr>
<tr>
<td>4</td>
<td>PPW</td>
<td>42</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Cranial stable, caudal inward</td>
</tr>
<tr>
<td>5</td>
<td>PPW</td>
<td>43</td>
<td>Chemo-RT</td>
<td>Outward</td>
<td>Cranial stable, caudal outward*</td>
</tr>
<tr>
<td>6</td>
<td>PPW</td>
<td>7</td>
<td>Chemo-RT</td>
<td>Outward</td>
<td>Stable</td>
</tr>
<tr>
<td>7</td>
<td>Tonsil</td>
<td>23</td>
<td>RT</td>
<td>No</td>
<td>Stable</td>
</tr>
<tr>
<td>8</td>
<td>Tonsil</td>
<td>17</td>
<td>RT</td>
<td>No</td>
<td>Overall inward</td>
</tr>
<tr>
<td>9</td>
<td>Tonsil</td>
<td>30</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Overall inward</td>
</tr>
<tr>
<td>10</td>
<td>Tonsil</td>
<td>19</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Overall inward</td>
</tr>
<tr>
<td>11</td>
<td>Tonsil</td>
<td>10</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Cranio medial outward*, caudal inward</td>
</tr>
<tr>
<td>12</td>
<td>Tonsil</td>
<td>52</td>
<td>RT</td>
<td>No</td>
<td>Cranial stable, caudal inward</td>
</tr>
<tr>
<td>13</td>
<td>Tonsil</td>
<td>43</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>14</td>
<td>Tonsil</td>
<td>8</td>
<td>RT</td>
<td>Inward</td>
<td>Overall inward (figure 3.3, example 1)</td>
</tr>
<tr>
<td>15</td>
<td>Tonsil</td>
<td>34</td>
<td>RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>16</td>
<td>Tonsil</td>
<td>32</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>17</td>
<td>Tonsil</td>
<td>43</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>18</td>
<td>Tonsil</td>
<td>91</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>19</td>
<td>Tonsil</td>
<td>40</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Initially outward followed by inward</td>
</tr>
<tr>
<td>20</td>
<td>Tonsil</td>
<td>43</td>
<td>Chemo-RT</td>
<td>Outward</td>
<td>Cranial outward*, caudal inward</td>
</tr>
<tr>
<td>21</td>
<td>Tonsil</td>
<td>13</td>
<td>RT</td>
<td>Outward</td>
<td>Cranial stable, caudal inward, lateral outward†</td>
</tr>
<tr>
<td>22</td>
<td>BOT</td>
<td>10</td>
<td>RT</td>
<td>No</td>
<td>Anterior outward†, lateral and caudal inward</td>
</tr>
<tr>
<td>23</td>
<td>BOT</td>
<td>23</td>
<td>Chemo-RT</td>
<td>No</td>
<td>Cranial inward, caudal stable</td>
</tr>
<tr>
<td>24</td>
<td>BOT</td>
<td>53</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>25</td>
<td>BOT</td>
<td>21</td>
<td>RT</td>
<td>Inward</td>
<td>Overall inward</td>
</tr>
<tr>
<td>26</td>
<td>BOT</td>
<td>88</td>
<td>Chemo-RT</td>
<td>Inward</td>
<td>Initially outward followed by inward (figure 3.3, example 3)</td>
</tr>
<tr>
<td>27</td>
<td>BOT</td>
<td>51</td>
<td>Chemo-RT</td>
<td>Outward</td>
<td>Anterior outward†, posterior inward</td>
</tr>
</tbody>
</table>

**Abbreviations:**
- BOT=base of tongue; chemo-RT=concomitant chemotherapy and radiation therapy; PPW=posterior pharyngeal wall; tonsil=tonsilar fossa.
- Time trend: Linear regression of $d_{\text{normal}}$ markers.
- * Outward in posterior pharyngeal wall.
- † Outward in lateral pharyngeal wall.
- ‡ Outward in base of tongue.

showing clear overall inward movement supports this hypothesis. Furthermore, in 3 patients, initial outward movement was followed by inward movement. This could be due to real tumor progression in the first weeks of radiation therapy or from edema; either way, the edge of the tumor including microscopic extension extends beyond its original position.

Only a few studies have described volumetric and positional changes of the GTV with repeat
Tumor shape variability during RT imaging at several time points during the course of radiation therapy [4, 5]. Castedot [5] reported on 10 patients with locally advanced disease, predominantly hypopharyngeal tumors, imaged with weekly CT and PET during chemoradiation treatment. They found a decrease of GTV of 3.2% on CT and 3.9% on PET per treatment day, respectively. The center of mass of GTV changed position over time. Similar findings were reported by Barker on 14 head-and-neck cancer patients treated with radiation therapy [4, 5]. Three weekly imaging was performed with an integrated CT linear accelerator system. They concluded that GTVs decreased throughout the course of radiation therapy at a median rate of 1.8% per treatment day. As in our study, the absolute loss was larger for larger tumors; moreover, they observed a higher rate of loss in patients with larger tumors. They also described a change in position of the center of mass over time, indicating asymmetric GTV losses in concordance with our findings. Our observations, however, describe local surface deformations and cannot be used to adequately describe the GTV position. The implanted markers were distributed nonuniformly over the surface of the GTV (e.g., the deep edge of the tumor was not covered nor the surface of a protruding tumor).

In view of our results and those of others described here, we will discuss the relation of anatomical variability to the margins used in radiation therapy treatment planning and the issue of margin reduction of the high-dose-volume mid-radiation therapy in case of response.

Standard head-and-neck setup margins and correction protocols are generally based on the assumption of rigid body motion, the patient is considered as nondeformable. Several articles [8, 9] reported on local geometrical uncertainties in anatomical subregions during radiation therapy for head-and-neck cancer patients. They showed considerable local setup variations based on nearby bony anatomy registration indicating nonrigid deformations. In this study, substantial nonrigid tumor variability was measured after local bony anatomy registration, although organ motion and residual deformation are difficult to distinguish. Nonetheless, this soft-tissue variability is currently not accounted for in standard head-and-neck setup margins. The aim of margin recipes is to cover the target with 95% of the radiation dose in 90% of patients [1]. Systematic and random errors are taken into account. In general, uncertainties introduced during treatment preparation (systematic errors) have more impact than day-to-day variation during treatment execution (random errors) [1]. The margin needed to cover the patient-specific systematic error in 90% of our patients would be 0.23 cm. However, marker movement was frequently asymmetric as described here, making a margin recipe even more difficult. Based on our study, the cranial and caudal border in the posterior pharyngeal wall are at highest risk to be covered insufficiently during a course of radiation therapy in patients with oropharyngeal tumors.

Recently, adaptive radiation therapy has been able to adjust the treatment plan to anatomical changes. Without replanning, in particular the parotid glands are at risk to get a higher dose than planned because of shrinkage and medial shift toward the high isodose volumes [10, 11]. The question rises if redelineation of the GTV based on anatomical shrinkage and a reduction of the CTV mid-radiation therapy is allowed? To answer this question, we should be able to make a distinction between response with an actual shrinking tumor or a dissolving tumor.
without shrinkage of the borders of the original GTV. In the latter, the tissue volumes from which tumors shrank radiographically are still likely to contain a large number of tumor cells based on radiobiological models [12]. Implanted markers on the interface between GTV and the area with microscopic extensions can help solve this question and change policy to only adjust the GTV when clear anatomical change toward the pharyngeal space is present. If we reason that the microscopic involved areas move together with the GTV, the residual marker motion is a surrogate for movement of the high-dose CTV. In our analyses, time trends were associated with BOT sublocalization and large volume tumors. Unfortunately, multivariate analyses could not be performed because of small numbers to identify independent factors associated with time trends to guide selection. Patients with both a measured inward trend and a visual overall inward movement of markers might be considered for adjustment of the mid-radiation therapy CTV. On the other hand, tumors in patients with an inhomogeneous or no marker pattern most likely are dissolving during radiation therapy without shrinkage of the CTV. As discussed previously, implanted markers do not cover the complete surface of the tumor and further analysis of marker motion in comparison to GTV change on repeat CT and MRI during radiation therapy will help to confirm these findings.

In conclusion, tumor variability can be determined with implanted markers during radiation treatment in head-and-neck cancer patients. Large differences in marker patterns between patients as well as within patients were observed. This soft-tissue variability is not taken into account in standard setup margins and correction protocols based on the assumption of rigid body motion. In our study, the cranial and caudal borders in the PPW are at highest risk to be covered insufficiently during radiation therapy in patients with oropharyngeal tumors. Furthermore, implanted markers could help identify patients with an actual shrinkage of the GTV who might benefit from mid-radiation therapy redelineation to reduce toxicity.

Summary

Tumor shape variability of oropharyngeal tumors during radiation therapy can be determined with implanted markers. Displacement of markers on cone beam computed tomography in reference to the planning computed tomography was analyzed in 27 patients. Large differences in marker patterns (both inward and outward) between patients as well as within patients were observed. Based on our study, the cranial and caudal border in the posterior pharyngeal wall are at highest risk to be covered insufficiently during radiation therapy.
Figure 3.3. Weekly marker displacement during radiation therapy in 3 patients. Example 1: inward pattern; example 2: cranial inward and caudal outward pattern; example 3: initially outward followed by inward pattern.
References


Analysis of GTV reduction during radiotherapy for oropharyngeal cancer: implications for adaptive radiotherapy

O. Hamming-Vrieze¹, S. van Kranen¹, W. Heemsbergen¹, C. Lange², M. van den Brekel³, M. Verheij¹, C. Rasch⁴, J.-J. Sonke¹

¹Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
²Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
³Department of Head and Neck Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands
⁴Department of Radiotherapy, Academic Medical Centre, Amsterdam, The Netherlands

E-mail: o.vriezeg@nki.nl

Published February 2017 in Radiotherapy and Oncology, volume 122(2), pages 224-228

http://dx.doi.org/10.1016/j.radonc.2016.10.012
Chapter 4

Abstract

Background and Purpose
Adaptive field size reduction based on gross tumor volume (GTV) shrinkage imposes risk on coverage. Fiducial markers were used as surrogate for behavior of tissue surrounding the GTV edge to assess this risk by evaluating if GTVs during treatment are dissolving or actually shrinking.

Materials and Methods
Eight patients with oropharyngeal tumors treated with chemo-radiation were included. Before treatment, fiducial markers \((0.035 \times 0.2 \text{ mm}^2, n=40)\) were implanted at the edge of the primary tumor. All patients underwent planning-CT, daily cone beam CT (CBCT) and MRIs (pre-treatment, weeks 3 and 6). Marker displacement on CBCT was compared to local GTV surface displacement on MRIs. Additionally, marker displacement relative to the GTV surfaces during treatment was measured.

Results
GTV surface displacement derived from MRI was larger than derived from fiducial markers (average difference: 0.1 cm in week 3). During treatment, the distance between markers and GTV surface on MRI in week 3 increased in 33\% >0.3 cm and in 10\% >0.5 cm. The MRI–GTV shrank faster than the surrounding tissue represented by the markers, i.e. adapting to GTV shrinkage may cause under-dosage of microscopic disease.

Conclusions
We showed that adapting to primary tumor GTV shrinkage on MRI mid-treatment is potentially not safe since at least part of the GTV is likely to be dissolving. Adjustment to clear anatomical boundaries, however, may be done safely.
Introduction

Intensity modulated radiotherapy is standard of care in organ preserving treatment of head and neck cancer. Dose distributions are optimized based on the anatomy of the pre-treatment planning CT (pCT). During treatment, however, shape and volume changes occur. This variability may be related to change in tumor volume, weight, edema, muscle mass or fat distribution. Several studies have described changes of gross tumor volume (GTV) and organs at risk (OAR) with repeat CT imaging during radiotherapy [1] and [2]. For example, Barker et al. [1] described a decrease of GTV during radiotherapy at a median rate of 1.8%/day. Similarly, the volume of parotid glands (PGs) gradually decreased and their position changed during treatment [2] and [3].

In the presence of anatomical changes, the actual delivered dose will be different from the planned dose. This dosimetric difference is especially significant for PGs, for which the mean dose can increase by 10% in selected patients [4] and [5]. Anatomical changes during treatment can be accounted for with adaptive radiotherapy (ART) where the radiation plan is adjusted during treatment [4]. Moreover, combining adjustment to anatomical changes with field size reduction to account for tumor shrinkage, leads to increased sparing of OAR, particularly PGs [4].

Field size reduction following tumor regression assumes that microscopic disease in the clinical target volume (CTV) behaves congruent with changes in the visible GTV. As CTVs do not have clear anatomical boundaries and microscopic disease cannot be visualized in repeat imaging, this assumption cannot directly be verified. In this study, fiducial markers were used as a surrogate for the behavior of tissues surrounding GTV to examine if these tissues shrink together with his GTV. The displacement of implanted markers was visualized with daily CBCT and compared to GTV changes quantified on repeat MRI. Subsequently, the potential risk of adaptive field size reduction following GTV shrinkage on MRI was evaluated, for instance, such a field size reduction imposes risk of under-dosage of microscopic disease when the GTV is dissolving without shrinkage of the surrounding tissues.

Materials and methods

Patients

Eight patients with locally advanced oropharyngeal carcinoma, treated with curative intent, were included in this study with informed consent (table 4.1). All patients underwent seven weeks of radiotherapy to a total dose of 70 Gy in 35 fractions with administrations of Cisplatin 100 mg/m² on day 1, 22 and 43 of treatment. The median age was 63 years (range 48–70).

Before start of treatment, during investigation under anesthesia, helical shaped gold markers (0.035 ×0.2 cm, in-house modified Visicoil™, RadioMed Corporation, Tyngsboro, MA, US) were implanted to aid delineation. Markers were positioned near the outer edge of GTV at approximately 0.5 cm depth. On average five markers per patient were implanted (range 3–8).
Table 4.1. Tumor location, tumor stage, number of markers and pre-treatment volume.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Tumor location</th>
<th>Stage</th>
<th>N*</th>
<th>Volume (cc)</th>
<th>Absolute volume (cc)</th>
<th>Relative volume (%)</th>
<th>∆#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Week3</td>
<td>Week6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Base of tongue</td>
<td>T4N2b</td>
<td>5</td>
<td>78</td>
<td>40</td>
<td>49</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>Base of tongue</td>
<td>T2N2b</td>
<td>5</td>
<td>16</td>
<td>8</td>
<td>46</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>Base of tongue</td>
<td>T2N2c</td>
<td>8</td>
<td>17</td>
<td>11</td>
<td>34</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>Base of tongue</td>
<td>T4N3</td>
<td>5</td>
<td>99</td>
<td>63</td>
<td>36</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>Post pharyngeal wall</td>
<td>T3N0</td>
<td>6</td>
<td>54</td>
<td>29</td>
<td>45</td>
<td>76</td>
</tr>
<tr>
<td>6</td>
<td>Tonsillar fossa</td>
<td>T3N2c</td>
<td>3</td>
<td>101</td>
<td>53</td>
<td>48</td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>Tonsillar fossa</td>
<td>T4N2c</td>
<td>5</td>
<td>60</td>
<td>41</td>
<td>32</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>Tonsillar fossa</td>
<td>T3N0</td>
<td>3</td>
<td>11</td>
<td>6</td>
<td>41</td>
<td>64</td>
</tr>
</tbody>
</table>

*Number of markers, ∆# Decrease.

Markers were located in the lateral pharyngeal wall (37.5%), posterior pharyngeal wall (20%), base of tongue (35%), palate (5%) and floor of mouth (2.5%).

In all patients a pCT (Siemens-AG, Erlangen, Germany) and MRI (3-Tesla, Philips, Eindhoven, The Netherlands) were acquired with a median of respectively ten and eleven days (range 10–14) pre-treatment. Both scans were made with three mm slice distance, with contrast and in treatment position after fixation with a five point thermoplastic mask. Additionally, MRIs in treatment position were acquired in weeks three and six of treatment. MRIs were made with a large flex coil and consisted of coronal contrast enhanced T1-TSE, transverse STIR and transverse T1-TSE. During treatment, daily CBCTs (Elekta Oncology Systems Ltd, Crawley, UK) were acquired.

Analyses of GTV reduction

GTVs were contoured on each MRI separately at all three time points by one observer (pretreatment: GTV pre, at week 3: GTV w3 and at week 6: GTV w6). To assess the potential gain of mid-treatment CTV adaptation based on the changed GTV w3 and GTV w6, treatment plans were re-optimized with adapted CTVs in three patients (respectively small, intermediate and large GTV). Optimization was started from the initial plan, optimization parameters were adjusted if required to achieve the same target coverage to be able to compare OAR doses or if overlap between PTV and an OAR changed leaving room for dose decrease on the OAR.

Absolute and relative GTV volume changes were calculated. Furthermore, in weeks three and six, local GTV surface displacements compared to the pre-treatment surface were measured in perpendicular direction at all marker positions (Supplementary Fig.4.3a). These linear distances of GTV displacement were compared to marker displacements in the same direction at the corresponding locations.
Analysis of marker displacement

The position (center of mass) of each implanted marker was determined on pCT and daily CBCTs using semi-automatic segmentation. Daily marker displacement, compared to the position on pCT, was calculated for each marker after local alignment on nearby bony anatomy. This displacement vector was subsequently decomposed into a component perpendicular to the pre-treatment GTV surface and a parallel component (Supplementary Fig. 4.3b). The perpendicular component of marker displacement ($d$) was used in this study as a measure for volume changes. Negative values indicate inward movement (shrinkage) and positive values indicate outward movement (volume increase). The direction of marker displacement (stable, inward or outward) was visually checked and a Pearson correlation with fraction number was calculated (week 1–3 versus week 4–7). Furthermore, average marker displacement, compared to the position on pCT, was calculated per week for each marker ($\langle d \rangle_{\text{week}}$). Outward movement was quantified by evaluating threshold violation ($0.1, 0.2, 0.3$ and $0.5$ cm) for markers and patients.

Subgroup analyses were performed based on tumor location and tumor volume ($\leq 30$ cc versus $>30$ cc). Observed differences in $d$ between subgroups were tested using the Mann–Whitney test with a significance level of 0.01 to adjust for multiple testing. Software package SPSS Statistics version 20.0 was used for statistical analysis.

Comparing GTV to marker displacement

To answer whether GTV dissolves or actually shrinks during treatment, we evaluated if tissue surrounding the GTV edge behaves congruent with visible GTV volume changes on MRI. Local GTV displacement on MRI was compared to marker displacement at the corresponding location in weeks three and six. A direct comparison is possible since both values were calculated in perpendicular direction at the position of each marker. Because implanted markers are not visible on MRI, registration of pCT, daily CBCTs and MRIs was performed on nearby bony anatomy representative of the tumor location, to allow this comparison. All matches were checked and approved visually for soft tissue alignment.

Implications for adaptive radiotherapy

To evaluate the impact that adaptive field size reduction following GTV shrinkage would have, we calculated the discrepancy between CBCT marker displacement and MRI surface displacement in weeks three and six. Therefore, we took the marker displacement and subtracted the corresponding perpendicular MRI–GTV surface displacement (Supplementary Fig. 4.3c). The result describes the marker displacement to respectively GTV$_{w3}$ and GTV$_{w6}$ for each marker ($\langle \Delta d \rangle_{\text{week}}$). As an outward marker displacement can result in under-dosage if the field size is reduced based on GTV shrinkage, we counted how often and where $\langle \Delta d \rangle_{\text{week}}$ exceeded respectively $0.3$ cm and $0.5$ cm. Additionally, we counted the number of patients with at least one marker exceeding these thresholds.
Chapter 4

Results

GTV volumes delineated on repeat MRIs consistently decreased during radiotherapy, Fig. 4.1 shows an example of GTV delineations. The average relative shrinkage at weeks three and six was respectively 41% (range: 32-49%) and 70% (64-78%) (Table 4.1, Fig.4.2). The average local GTV displacement at the position of each marker in perpendicular direction was $-0.3 \text{ cm}$ in week three and $-0.5 \text{ cm}$ in week six (Table 4.2). Treatment plans after CTV shrinkage based on GTV$_{w3}$ and GTV$_{w6}$ showed OAR dose reduction (Supplementary Table 4.3 and Supplementary Fig.4.4).

![Figure 4.1. GTV delineations in patient 6 on MRI pre-treatment (a), in week 3 of treatment (b) and in week 6 of treatment (c).](image)

![Table 4.2. Displacement of GTV surface and markers in week 3 and 6 of treatment. $S$ is the surface displacement on repeat MRI in reference to the pre-treatment MRI. $⟨d⟩_{\text{week}}$ is the average marker displacement per week in reference to the marker position on planning CT. $⟨Δd⟩_{\text{week}}$ is the average marker displacement per week in reference to the corresponding GTV surface. All values are in cm. Negative values are inward displacement, positive values are outward displacement.](table)

<table>
<thead>
<tr>
<th>Displacement(cm)</th>
<th>Week 3</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S$</td>
<td>$⟨d⟩_{\text{week}}$</td>
</tr>
<tr>
<td>Minimum</td>
<td>-1.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.3</td>
<td>-0.2</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Implications of target reduction for ART

Forty markers were identified on daily CBCT to assess displacement in reference to the pCT. All evaluated markers were visible on pCT. Two markers were excluded due to loss respectively from the fourth and twentieth fraction onward. On average, markers were positioned 0.3 cm perpendicular to the outer edge of the GTV surface. Average marker displacement per week ($\langle d \rangle_{\text{week}}$), in reference to the position of the marker in the pCT, is described in Supplementary Table 4.4. Variation between the marker position on pCT and on CBCT during the first week of treatment was minimal (Mean $-0.1$ cm, SD 0.1 cm). At three and six weeks, the mean marker displacement was $-0.2$ cm at both time points with a SD of respectively 0.2 cm and 0.4 cm.

Subgroup analyses based on tumor location showed more variability in base of tongue (mean: $-0.3$ cm, SD: 0.4 cm) and tonsillar tumors (mean: $-0.4$ cm, SD: 0.4 cm) compared to posterior pharyngeal wall tumors (mean: $-0.1$ cm, SD: 0.2 cm) ($p < 0.001$). Volume analyses showed more variation in large volume tumors ($< 30$ cc mean: $-0.1$ cm, SD 0.2 cm versus $\geq 30$ cc mean: $-0.3$ cm, SD 0.4 cm, $p < 0.001$). After three weeks, the mean marker motion was $-0.2$ cm for large volume tumors and $-0.1$ cm for small volume tumors ($p < 0.001$).

Supplementary Fig. 4.5 shows examples of marker displacement. Visual verification of the direction of all markers was identical to calculated values. Correlation of markers with time in the first three weeks of treatment showed that 48% (19/40) of markers remained stable, 40% (16/40) of markers moved inward and 12% (5/40) moved outward. From week four onward stability increased, 75% (30/40) of markers remained stable, 18% (7/40) of markers moved inward, and only 7% (3/40) of markers moved outward.

Quantification of outward movement showed that the majority of $\langle d \rangle_{\text{week}}$ values moving outward was limited to 0.1 or 0.2 cm (Supplementary Table 4.4). In weeks four and five, one marker exceeded 0.3 cm, in weeks six and seven three markers exceeded 0.3 cm (Supplementary Table 4.5). The maximum outward movement compared to the pCT was 0.4 cm.

Figure 4.2. Absolute (3a) and relative (3b) GTV volume change on MRI for all patients comparing weeks 3 and 6 during treatment with pre-treatment.
Comparing average $\langle d \rangle_{\text{week}}$ marker displacement to average MRI-GTV displacement at three weeks showed a proportion of $−0.2$ cm versus $−0.3$ cm (Table 4.2). Hence, the change of tissue surrounding the original GTV edge represented by the markers was 56% of the average MRI-GTV displacement. In the second half of treatment, the average marker position stabilized, resulting in marker displacement of $−0.2$ cm in week 6, while the average local GTV displacement on MRI further decreased to $−0.5$ cm. Therefore, in week six, on average only 39% of MRI-GTV displacement behaves congruent with the surrounding tissue.

Marker displacement relative to GTV$_{w3}$ and GTV$_{w6}$ ($\langle \Delta d \rangle_{\text{week}}$), to assess the impact of adaptive radiotherapy following the MRI-GTV, is described in Table 4.2. Although the average $\langle \Delta d \rangle_{\text{week}}$ to GTV$_{w3}$ was only 0.1 cm, the distance between markers and GTV$_{w3}$ increased in one third of the markers with more than 0.3 cm and in 10% with more than 0.5 cm (Supplementary Table 4.5). These markers were present in respectively 63% and 50% of patients and indicate a potential under-dosage in case of mid-treatment field size adjustment based on GTV shrinkage on MRI. Remarkably, three out of four markers $>0.5$ cm in reference to GTV$_{w3}$ were located in the pharyngeal wall. In week six, the percentage of markers with an increased distance $>0.5$ cm relative to GTV$_{w6}$, mounted to 33% (Supplementary Table 4.5).

**Discussion**

In patients with an oropharyngeal carcinoma, adaptive field reduction following GTV shrinkage on MRI during radiotherapy imposes risk on CTV coverage since at least part of the GTV is likely to be dissolving instead of shrinking congruent with surrounding tissues. In this study, we were able to visualize behavior of tissue surrounding the GTV edge with implanted markers. Both GTV displacement on MRI, as well as surrounding tissue displacement represented by the fiducials were predominantly inward (shrinkage). Local comparison at the position of each marker showed a somewhat larger GTV displacement on MRI compared to surrounding tissue displacement. This led in a substantial part of markers to an increased distance between markers and GTV$_{w3}$ in the third week. In week six, the amount of markers with an enlarged distance relative to GTV$_{w6}$ further increased due to continuing shrinkage of the GTV on MRI and relative stabilization of surrounding tissue. Therefore, shrinkage of CTV during treatment following shrinkage of GTV on MRI could lead to under-dosage of microscopic disease and this risk increases toward the end of treatment.

Historically, delineation of GTV is based on clinical examination and CT. In recent years, the use of MRI for target volume definition has increased and is especially useful in the oropharynx due to its superior soft-tissue contrast with a reduction of inter-observer variation [6] and [7]. Despite the use of MRI, uncertainties in GTV delineation will remain. Different anatomical (CT/MRI) and biological (PET) imaging modalities do not correlate completely, both in volume and position [8, 9], and [10]. Delineation based on multimodality imaging has the potential to improve accuracy of GTV delineation and automated multimodality segmentation algorithms are being developed [7] and [9]. However, imaging defined GTVs may not represent the true
Implications of target reduction for ART

pathologic extent of the tumor, i.e., in laryngeal carcinoma, imaging modalities tended to result in an overestimation of GTV [8]. Most likely GTV delineations based on anatomical imaging modalities behave similarly during treatment despite delineation uncertainties. We expect these uncertainties to have only limited impact on the results of this study because we compared relative displacement of markers and GTV surface instead of absolute positions.

GTV shrinkage during treatment has been described by several authors [1, 2] by reporting a rate of volume change. The assumption of a linear relationship may not be optimal, i.e., Barker et al. [1] found a more rapid rate of volume change at the beginning of treatment than near the end. In this study, GTV volume was measured at three time-points thus a rate of volume change was not calculated. However, the relative shrinkage between these time-points was remarkably consistent for all patients.

Delineation of the CTV around the GTV is a balance between the risk of recurrence and the ability to spare OAR. However, no consensus exists on how a CTV should be constructed on the original pCT, let alone how to do this during treatment. Microscopic spread of squamous cell carcinomas follows anatomical compartments and routes. Generally, the CTV is adjusted to anatomical boundaries after expansion of the GTV. The recommended GTV-CTV margin varies widely between 0 and 20 mm[11]. Radiological examinations cannot help solve this problem because of their inability to visualize microscopic disease within the CTV. Moreover, little is known about the behavior of the CTV during treatment.

In recent years, deformable image registration (DIR) is emerging in adaptive re-planning strategies to propagate delineations from the original pCT to a repeat CT [12]. However, DIR may not be sufficient to describe CTV variations as it would typically follow the image contrast of the GTV and thus overestimate CTV reduction. Due to lack of anatomical landmarks representing CTV borders, quantification of the accuracy of DIR is hampered. Head and neck phantoms are under development to validate and compare DIR tools [13] and [14]

In this study, implanted markers were used to visualize behavior of the region containing microscopic disease near the GTV edge during treatment. Helical-shaped markers were selected to minimize migration. Stability of these markers was validated in previous work by measuring the inter-marker distance for markers implanted in normal tissue, this was about 0.05 cm[15]. Analysis of displacement was done after performing local bony anatomy registration between pCT and daily CBCT. Although it is difficult to distinguish organ displacement or rotation from volume and shape change using the marker displacement data itself, visual verification of the registered CBCT scans as well as the consistency of our results confirmed that residual marker displacement was dominated by deformation of tissue. Another limitation of the use of implanted markers is that they cannot cover the total surface of the tumor, for example, markers can only be positioned adequately at the interface of tumor and mucosa. Furthermore, the finite slice thickness and registration of pCT, MRI and CBCT introduces uncertainty. The marker displacement analysis of week 1 (Supplementary 4.4) indicates that the accuracy and precision is about 1 mm. As the algorithm for the pCT–MRI bony registration is similar and the same thermoplastic mask headrest was used for all imaging modalities, the same performance is expected. As the magnitude of
marker and contour displacement is substantially bigger, these uncertainties are expected to have only a limited impact on the results of this study.

Despite these limitations, this study provides unique insight in behavior of the primary target volume during treatment in patients with an oropharyngeal carcinoma. Since GTV reduction during radiotherapy overestimates reduction of the surrounding CTV, at least part of the GTV is likely to be dissolving during treatment. However, other causes of deformation could play a role such as stretching of wound edges or tissue edema [15]. Either way, the edge of the tumor including microscopic extension dose not shrink as fast as the MRI–GTV and mid-treatment CTV delineation remains a challenge. Adjustment to clear anatomical borders such as air cavities, bone or unaffected fasciae may be done safely. Contrary, soft tissue adjustments are not recommended outside clinical trials. Regions from which the tumor shrank radiologically, are still at risk to contain a substantial number of tumor cells that are below the detection threshold [16]. For example, when 99% of tumor cells are eradicated in a 1 mm$^3$ voxel that originally represented the GTV, it can still contain 104 tumor cells, but is unlikely to be visible on any imaging modality [17]. Moreover, radiological regression may not represent the eradication of stem cells that are the most important target of therapy [16]. One could hypothesize however, that these regions need less radiation dose than visible GTV mid-treatment. In the Netherlands Cancer Institute, we aim to implement a clinical trial to test this hypothesis with re-imaging and re-planning during treatment with a dose escalation on the remaining GTV visible on repeat imaging and a moderate dose de-escalation on the tissues from which the GTV shrank, thereby re-distributing the dose in the original target volume. An alternative approach could be to use probabilistic planning techniques that take population statistics of tumor regression patterns into account [18]. However, the number of patients in this study is limited and validation in a larger dataset would be required to determine population statistics of tumor regression patterns.

In conclusion, although CTV reduction mid-treatment could potentially decrease OAR dose, CTV re-delineation remains a challenge. We showed that following primary tumor GTV reduction during treatment on MRI poses a risk of under-dosage of microscopic disease. Adjustment to clear anatomical boundaries, however, may be done safely.
Appendix 4.A  Organs at risk dose at planning and after adaptation

Table 4.3. Average dose at planning CT (pCT) and gain in organ at risk dose compared to pretreatment in week 3 and 6 after CTV adjustment based on respectively GTV$_{w3}$ and GTV$_{w6}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>pCT</th>
<th>Week 3</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brainstem</td>
<td>D1</td>
<td>33 Gy</td>
<td>0%</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>D1</td>
<td>34 Gy</td>
<td>-2%</td>
</tr>
<tr>
<td>Parotid gland left</td>
<td>Dmean</td>
<td>38 Gy</td>
<td>-5%</td>
</tr>
<tr>
<td>Parotid gland right</td>
<td>Dmean</td>
<td>48 Gy</td>
<td>-5%</td>
</tr>
<tr>
<td>Oral cavity</td>
<td>Dmean</td>
<td>48 Gy</td>
<td>-5%</td>
</tr>
<tr>
<td>Constrictor muscle</td>
<td>Dmean</td>
<td>52 Gy</td>
<td>-3%</td>
</tr>
<tr>
<td>Larynx</td>
<td>Dmean</td>
<td>44 Gy</td>
<td>-3%</td>
</tr>
</tbody>
</table>
Chapter 4

Appendix 4.B  Weekly average marker displacement

Table 4.4. Average marker displacement per week ($\langle d \rangle_{\text{week}}$), in reference to the position of each marker in the planning CT.

<table>
<thead>
<tr>
<th>$\langle d \rangle_{\text{week}}$</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 5</th>
<th>Week 6</th>
<th>Week 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min (cm)</td>
<td>-0.4</td>
<td>-0.8</td>
<td>-0.9</td>
<td>-1.1</td>
<td>-1.3</td>
<td>-1.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>Max (cm)</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Mean (cm)</td>
<td>-0.1</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>SD (cm)</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>$&gt;0.1$ cm*</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$&gt;0.2$ cm*</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

*Number of markers exceeding respectively 0.1 or 0.2 cm
Appendix 4.C  Marker displacement during treatment

Table 4.5. Percentage of markers exceeding 0.3 cm respectively 0.5 cm in week three and six of treatment. \( \langle d \rangle_{\text{week}} \) is the average maker displacement per week in reference to the marker position on planning CT. \( \langle \Delta d \rangle_{\text{week}} \) is the average marker displacement per week in reference to the corresponding GTV surface.

<table>
<thead>
<tr>
<th></th>
<th>Week 3</th>
<th></th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \langle d \rangle_{\text{week}} )</td>
<td>( \langle \Delta d \rangle_{\text{week}} )</td>
<td>( \langle d \rangle_{\text{week}} )</td>
</tr>
<tr>
<td>( &gt;0.3 \text{ cm} )</td>
<td>0/40 (0%)</td>
<td>13/40 (33%)</td>
<td>3/40 (8%)</td>
</tr>
<tr>
<td>( &gt;0.5 \text{ cm} )</td>
<td>0/40 (0%)</td>
<td>4/40 (10%)</td>
<td>0/40 (0%)</td>
</tr>
</tbody>
</table>
Appendix 4.D Method of surface displacement measurement and marker displacement calculation.

Figure 4.3. Method of surface displacement measurement and marker displacement calculation.

a): Local GTV surface displacement at the position of each marker was measured along the perpendicular direction of the reference marker. The perpendicular direction follows from the original position of the marker in the planning CT (pCT), which is projected, after smoothing, onto the nearest surface of the GTV delineation.

b): The displacement of a marker on Cone beam CT (CBCT) was calculated in reference to the position of this marker on the pCT. The deformation vector was subsequently decomposed into a component perpendicular to the original GTV surface ($\langle d \rangle$) and a parallel component. The marker displacement along the perpendicular direction is used as a measurement for volume changes in tissues surrounding the GTV edge.

c). Relative marker displacement to GTV surface ($\langle rd \rangle$) was calculated by subtracting surface displacement from marker displacement at corresponding locations and time points in the perpendicular direction.
Appendix 4.E  OAR sparing after replanning

Figure 4.4. Treatment planning dose differences of D1 and Dmean in organs at risk (OAR) compared to pre-treatment for three patients, respectively in week 3 and 6 of treatment, after adjustment of the CTV to the GTV in the corresponding week.

In patient 2 (small GTV, Table 4.1) the dose reduction, depending on OAR for D1 or Dmean (Supplementary Table 4.3), was modest for all delineated OAR. The total dose reduction was larger in patient 1 (intermediate GTV) and patient 6 (large GTV). Dose reductions are present in different OAR when comparing patients, for instance, in patient 1 predominantly a reduction of Dmean dose to the parotid glands was observed, in contrast to Dmean of larynx, oral cavity and constrictor muscle in patient 6.
Figure 4.5. Example of a stable marker (1) and two markers (2+3) with inward displacement during treatment in patient 1 in an axial plane. Marker positions follow from CBCT measurements after alignment on nearby bony anatomy. Weekly average marker positions were used to construct the displayed marker trajectory during treatment, S defines the starting point, E the end point. The trajectories were projected into the axial plane of the center of marker 1 on pCT for visualisation. Marker 2 and 3 were located within 0.5 cm in cranio-caudal direction. The various volumes represent the MRI defined GTVs prior to treatment (Pre-RT), mid treatment (week 3) and at the end of treatment (week 6).
References


Correction strategies to manage deformations in head-and-neck radiotherapy

S. van Kranen, S. van Beek, A. Mencarelli, C. Rasch, M. van Herk, J.-J. Sonke

Department of Radiation Oncology, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

E-mail: j.sonke@nk.nl

Published Februari 2010 in Radiotherapy and Oncology, volume 94, pages 199-205

dx.doi.org/10.1016/j.radonc.2009.12.016
Chapter 5

Abstract

Background and purpose
To optimize couch shifts based on multiple region-of-interest (ROI) registrations and derive criteria for adaptive replanning for management of deformations in head-and-neck (H&N) cancer patients.

Materials and methods
Eight ROIs containing bony structures were defined on the planning-CT and individually registered to daily cone-beam CTs for 19 H&N cancer patients. Online couch shifts were retrospectively optimized to correct the mean setup error over all ROIs (mean correction) or to minimize the maximum error (MiniMax correction). Residual error distributions were analyzed for both methods. The number of measurements before adaptive-intervention and corresponding action-level were optimized.

Results
Overall residual setup errors were smallest for the mean corrections, while MiniMax corrections reduced the largest errors. The percentage of fractions with residual errors >5 mm was 38% versus 19%. Reduction of deformations by single plan adaptation was most effective after eight fractions: systematic deformations reduced from 1.7 to 0.9 mm. Fifty percent of this reduction can already be achieved by replanning 1/3 of the patients.

Conclusion
Two correction methods based on multiple ROI registration were introduced to manage setup errors from deformations that either minimize overall geometrical uncertainties or maximum errors. Moreover, the registrations could be used to select patient with large deformations for replanning.
5.1 Introduction

Intensity-modulated radiotherapy (IMRT) generates highly conformal dose distributions with steep dose gradients. The benefits of IMRT are potentially large for patients with head-and-neck (H&N) cancer where the tumor shape is complex and its location often close to organs at risk (OARs). Setup errors, anatomy, and posture changes, however, cause geometrical uncertainties that may lead to serious degradation of the original plan. To account for these geometrical uncertainties safety margins are applied.

Geometrical uncertainties may be reduced by implementation of image-guided radiotherapy (IGRT). In IGRT the anatomy of the patient is imaged prior to treatment. Often direct identification of the tumor in these images is not possible and bony anatomy is taken as a surrogate position. It is current practice to use a single rigid registration of the bony anatomy with the planning-CT to derive a couch correction to minimize setup errors. Non-rigid components in patient setup, if detected, are usually ignored. However, recent studies with patients with H&N cancer have shown that local misalignments should not be ignored [1–6].

Local misalignments may be detected with multiple region-of-interest (mROI) registration [1, 6]. In mROI registration, a set of sub-regions is rigidly registered. mROI registration has the advantage over single large ROI registration that alignment of bony structures does not lead to confusing results due to deformations. However, because of deformations each ROI will produce a different registration. A single couch correction is therefore always a compromise, leading to residual setup errors.

Currently it is not clear how to optimally combine mROI registrations into a couch correction. However, the measured residual setup errors may be used to select patients with large deformation that might benefit from plan adaptation. Therefore the purpose of this study was to optimize couch shifts based on multiple region-of-interest (ROI) registrations and derive criteria for adaptive replanning for management of deformations in head-and-neck (H&N) cancer patients.

5.2 Materials and methods

Patient data

For this study 19 patients with H&N cancer with daily cone-beam CT (CBCT) scans were included regardless of tumor stage or location. IMRT planning was performed on a planning-CT (pCT) (Somaton Sensation Open, Siemens AG, Erlangen, Germany) with $0.8 \times 0.8$ mm$^2$ in plane resolution and 3 mm slice distance. Patients were positioned with a custom made 5-point thermoplastic fixation mask, a standard neck-rest and a knee-support. The CBCT scans (Elekta Synergy, Elekta Oncology Systems Ltd, Crawley, UK) were reconstructed with a resolution of
A total of 573 CBCT scans were available, on average 30 scans per patient (range: 21–36 scans).

Online correction protocols

Retrospectively we have simulated the following workflow:

1. Patient setup prior to treatment according to skin-tattoos and laser lines.
2. Acquisition of a daily CBCT scan.
3. Registration of multiple regions of interest.
4. Derive and apply an optimal couch correction.
5. Treatment delivery.
6. Evaluation of local residual setup errors (after couch corrections) to selection criteria for replanning.

This protocol differs from conventional protocols by the registration of multiple regions of interest (mROI) (step 3), the derivation of a couch correction based on multiple setup errors (step 4) and assessment of systematic local residual setup errors (step 6). In the following paragraphs these steps will be discussed in more detail.

Multiple ROI registration

In each pCT we defined eight box-shaped ROIs containing the following bony structures: the occiput bone, vertebrae C2, C4, and C6, jugular notch, larynx, hyoid, and mandible. As separate bony structures do not deform, we performed for each ROI planning-CT to daily cone-beam CT (CBCT) local rigid registrations of the bony anatomy using chamfer matching. The registration of the mROIs was a loop over all individual regions and did not appreciably increase the registration time compared to a conventional single large ROI registration. Although chamfer registration generally achieves an accuracy less than 1 mm, occasionally the registration might result in an obvious failure. Therefore each registration was visually verified. To simplify this process we applied a thin plate spline algorithm to create a deformation vector field that contained the misalignments of all box-shaped regions. Subsequently we deformed the CBCT for color-fusion with the original planning-CT: any incorrect registration of the bony anatomy was immediately identifiable and could be corrected. Overlaying a deformed CBCT reduced the visual verification of multiple registrations to a single assessment.
Derivation of a couch correction based on multiple setup errors

A set of local setup errors was converted into a couch correction according to two strategies: mean and MiniMax corrections. Both strategies were compared with a single large ROI correction that was determined during actual treatment delivery. Typically a single large ROI encompassed all individual mROIs.

![Figure 5.1. Example of setup errors before corrections projected in the sagittal view (red arrows). From the individual setup errors a correction is calculated (blue arrow), here the mean error. Setup errors are scaled.](image)

Mean correction strategy

The center of ROI \(i\) at the planning-CT was described with \(p_i\). From the mROI registration we obtained a 4 × 4 transformation matrix \(E_i\) describing the misalignments of region \(i\) (translations and rotations only, homogenous coordinates). The position \(r_i\) of ROI \(i\) in the CBCT therefore followed from:

\[
    r_i = E_i \cdot p_i \tag{5.1}
\]

The local setup error \(d_i\) is the vector difference between \(r_i\) and \(p_i\) (shift of the center of ROI \(i\), see figure 5.1). Couch correction \(C\) represents the global setup error, which, if inverted, displaces ROI \(i\) to \(r'_i\). Since we did not allow corrections for rotations, the correction was simply a vector subtraction \(c\):

\[
    r'_i = C^{-1} \cdot r_i = r_i - c \tag{5.2}
\]
Finally we defined the local residual setup error \( \epsilon_i \) after correction:

\[
\epsilon_i = r'_i - p_i = d_i - c
\]  
(5.3)

For the whole patient group setup errors were expressed per structure in the group mean \( M \), systematic \( \Sigma \) and random \( \sigma \) setup errors per direction [10].

In the mean correction strategy the correction \( c \) is simply the average over all setup errors (figure 5.1). Mathematically, this correction minimizes the sum of squared errors \( \epsilon_i \):

\[
\min_c \left\{ \sum_{i=1}^{N} (d_i - c)^2 \right\}
\]  
(5.4)

MiniMax correction strategy

During RT small setup errors will be covered by the planning target volume (PTV) margin. Large errors, however, might exceed this margin. The objective of the MiniMax correction strategy is to minimize the maximum setup error and thereby reduce the largest errors as much as possible. Mathematically the MiniMax optimization objective is formulated as:

\[
\min_c \left\{ \max_i \{ |d_i - c| \} \right\}
\]  
(5.5)

If all setup error vectors were collected with a common origin, the optimization comes down to finding the minimum enclosing bal (MEB) [11]. The position of the MEB is determined by at least two errors. The center of the MEB represents the correction \( c \) in the MiniMax correction strategy, while the maximum residual errors (two at minimum) have a length equal to the radius.

Evaluation of local residual setup errors for plan adaptation

There are two types of plan adaptation: type one, originally proposed by Yan et al. [12], aims at correcting systematic errors and is based on an average patient model estimated from the first couple of treatment/imaging fractions. The second type takes treatment response into account, for instance tumor regression, and therefore deals with progressive changes. This study focuses on plan adaptation type I, since local misalignments of bony anatomy predominately represent posture and shape changes and are less sensitive to treatment response [1, 6].

For each ROI we calculated the residual systematic setup error at fraction \( N \) after online corrections. Patients with an estimated systematic error larger than \( T \) mm over the first \( N \) fractions were identified as candidates for adaptive replanning. Assuming that it is possible to adapt the original plan based on an average patient model in which the anatomy of the patient is in line with the estimated systematic deformations after \( N \) fractions, we recalculated the residual errors for fraction \( N + 1 \) up to the end of the treatment.
H&N setup correction strategies to manage deformations

local setup errors

Mean, systematic and random  Cumulative probability distribution

![Diagram showing mean corrections for H&N setup with cumulative probability distribution graphs for different ROI correction strategies.](image)

**Figure 5.2.** Positioning accuracy in sagittal view (left column) after corrections according to the three different online correction strategies. Small blue arrows indicate group mean errors; the blue ellipses represent systematic errors while red ellipses represent the random errors. Errors are not to scale and shown against a template patient. The cumulative probability distribution of the deformations is shown in the right column.
5.3 Results

Individual correction strategies

To prevent reporting an excessive number of setup errors and simplify comparison between methods, we visualized the results in error plots. The accuracy for the AP- and CC-direction is depicted in figure 5.2, the LR-direction is not shown but is similar to the AP direction. Note that the ratio of systematic vs. random errors is $1 - 1.5$ for all strategies. The plots in the right column of figure 5.2 show the cumulative distribution of residual setup errors per ROI and represent the probability to exceed a certain vector length.

**Table 5.1.** Accuracy, the average length over patients of the systematic error vector, and incidence of large errors (>5 mm vector length) per region after single ROI, *mean* and *MiniMax* corrections.

<table>
<thead>
<tr>
<th>structure</th>
<th>Mean systematic setup error (mm)</th>
<th>Incidence (%) setup error &gt;5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sROI</td>
<td>mean</td>
</tr>
<tr>
<td>C2</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>C4</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>C6</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Mantibula</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Hyoid</td>
<td>3.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Larynx</td>
<td>3.1</td>
<td>2.5</td>
</tr>
<tr>
<td>Occiput Bone</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Jugular notch</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>Fractions</td>
<td>52.4</td>
<td>38.2</td>
</tr>
</tbody>
</table>

*Abbreviations: sROI = single Region Of Interest.*

The positioning accuracy of individual structures is collected in table 5.1 and expressed in the average systematic setup error vector length over 19 patients. Large errors (>5 mm vector length) are found in the tails of the cumulative distributions, and were scored separately in table 5.1 with the percentage of fractions with large errors in the last row.

Comparison of strategies

The cumulative distribution of all setup errors represents the probability of exceeding a certain setup error vector length (figure 5.3a); lower curves indicate a higher positioning accuracy. The cumulative distribution of the maximum systematic error (over all ROIs) is shown in figure 5.3b. The *MiniMax* positioning strategy manages to reduce the systematic errors to within 5 mm. For the *mean* corrections and single large ROI corrections, this is 6 and 8 mm.

As a figure of merit to describe all variability of the different structures for all patients in a single number we have taken the root mean square (RMS) value over the residual systematic
and random errors (RMS-$\Sigma_d$ and RMS-$\sigma_d$), see table 5.2. Regardless of the correction method, RMS errors are largest in the CC-direction. The mean correction strategy produces the smallest RMS residual setup errors.

**Optimization of intervention criteria**

Since the largest errors were predominantly found in the CC-direction, we optimized the intervention criteria for plan adaptation, i.e., the number of measurements before intervention $N$ and the threshold $T$, for the CC-direction only. In figure 5.4a we plotted RMS-$\Sigma_d$ as a function $N$ and $T$. Online corrections were made according to the mean correction strategy. A minimum for RMS-$\Sigma_d$ was found at about $N = 8$ fractions, independent of $T$. Figure 5.4b represents the
workload in terms of percentage of patients selected for replanning (for $N = 8$) as function of $T$. Clearly, with a lower threshold, precision increases at the cost of a higher workload.

5.4 Conclusions and discussion

In this study we derived optimal couch shifts to manage deformations in H&N cancer with emphasis on minimization of overall geometrical uncertainties (mean corrections) or maximum errors (MiniMax corrections). Local residual errors were quantified with mROI registration, using existing registration methods which are already clinical routine and have proven to be simple, fast and robust. Based on the residual setup errors, selection criteria were optimized to objectively select patients that might benefit from a single plan adaptation. The optimum moment for intervention was found at $N = 8$.

Regardless of which correction protocol was used, large local setup errors were found, in line with a previous study in our institute with a similar set of ROIs [6]. Although the present study was arranged to compare different patient setup protocols based on mROI registration in its ability to manage deformations, all efforts should be taken to avoid deformations in the first place. Improvements in patient support and immobilization [13] could reduce local residual setup errors as observed in this study.

Close inspection of figure 5.2 reveals that online couch corrections with the mean or Minimax method deliver more accurate positioning than single large ROI corrections for regions such as the larynx and hyoid, while at the same time the accuracy for the vertebrae reduces. This is due to the segmentation process of the large ROI: the small bony anatomy, like larynx or hyoid, contributes only very little compared to large structures like the occiput bone or the vertebrae. Consequently, these structures are almost ignored in the chamfer registration, while in the mROI methods they contribute equally.

However, larger errors still remain for the larynx and hyoid after mean corrections. Vertebrae movement is highly correlated, i.e. they share a common rigid component. The residual movement of the larynx and hyoid, on the other hand, is often independent of the vertebrae. Mean corrections

<table>
<thead>
<tr>
<th></th>
<th>RMS-$\Sigma_d$(mm)</th>
<th>RMS-$\sigma_d$(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
<td>CC</td>
</tr>
<tr>
<td>Single large ROI</td>
<td>1.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Mean</td>
<td>1.0</td>
<td>1.7</td>
</tr>
<tr>
<td>MiniMax</td>
<td>1.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Abbreviations: AP = anterior-posterior; CC = cranio-caudal; LR = left-right; $\Sigma$ = systematic error; $\sigma$ = random error.
effectively correct the common rigid component while independent motion is not regarded. Note also that alignment of jugular notch and occiput bone does not improve.

Small misalignments are not handled by the MiniMax correction strategy. This strategy acts on the largest 2 or 3 errors; as long as the residual setup error of a structure does not exceed the largest error it may have any setup error. Therefore, possible improvements in small misalignments depend on the relation with the largest errors. Therefore corrections might have no effect or even worsen the accuracy. Consequently we find that the largest errors, hyoid and larynx, are clearly reduced compared to the mean or single large ROI corrections, but other structures are positioned less accurate. In fact, MiniMax corrections result in nearly equal errors for all structures (see figure 5.2).
Furthermore, the MiniMax positioning strategy is less robust to errors in the registration: an incidental large mismatch might contribute heavily to the derived correction. Therefore our method of choice is the mean correction strategy which we have clinically implemented since January 2009 to drive an offline correction protocol [14].

Relevant setup corrections are only achieved with meaningful ROIs, representative of the location of the tumor and OARs with respect to high dose gradients. For this study we used the same subset of bones for all patients, regardless of the actual location of the tumor and/or dose distributions to enable inter-patient comparison. Clinically, however, the selection of ROIs may be limited to bony structures surrounding the tumor and OARs. The same set of bony structures may be used for similar tumor locations (oropharynx, hypopharynx etc.), although exclusion of specific bony structures still depends on tumor size, lymph nodes involvement, or application (boost area’s). Since each ROI counts equally an overrepresentation of a part of the patient’s anatomy, e.g. seven separate ROIs for the spinal cord dilutes the effect of the remaining ROIs.

Birkner et al. [3] analyzed deformations by registration of featurelets, rectangular 2D structures in portal images with relevant anatomy. As basis for corrections they preferred minimization of maximum errors over mean displacement, as they hypothesized that mean displacements would lead to larger systematic errors for some regions. The featurelets were individually defined, making it impossible to compare patients and therefore methods. As was shown in this study, the MiniMax corrections indeed reduce the maximum systematic errors more than mean corrections but at the cost of the smaller systematic errors. Generally it is impossible to be both precise and minimize large errors. A hybrid approach may find a compromise in these two objectives.

Mean and MiniMax corrections may be also be used for offline positioning protocols. As offline correction protocols only manage the systematic component of geometrical uncertainties, this component first needs to be estimated for each region over the previously delivered fractions. The objective function to calculate the couch correction for the mean correction strategy thus becomes:

$$\min_c \left\{ \sum_{i=1}^{N} (\langle d_i \rangle^f - c)^2 \right\}$$

(5.6)

and

$$\min_c \left\{ \max_i \left\{ |(d_i)^f - c| \right\} \right\}$$

(5.7)

for the MiniMax correction strategy, $\langle \cdots \rangle^f$ with indicating the mean over the delivered fractions. As the MiniMax correction strategy operates on the maxima, it is more sensitive to statistical noise in the estimated systematic error due to limited number of fractions used.

Safety margins are needed to account for residual geometrical uncertainties. Classical margin recipes, such as by van Herk [15], are not valid to handle multiple (or deforming) objects: lack
of correlations in 'movement' might decrease the probability that multiple regions are within margins at the same time [16]. An upper limit to account for only the systematic local setup errors after online corrections is 5 mm for MiniMax, 6 mm for mean, and 8 mm for single ROI corrections (figure 5.3). In practice, smaller margins could be applied to regions with smaller errors but the calculations that would reveal these details were beyond the scope of this study.

The optimal number of measurements $N$ before intervention was found at eight fractions and was independent of the selection threshold $T$. We arrived at these results by optimizing RMS-$\Sigma_d$. According to Bortfeld et al. [17] the optimal number of measurements, based on quadratic errors, to correct systematic errors in patient setup would be $N = 5$. To investigate this discrepancy we simulated a set of simple treatments with only rigid setup errors (30 fractions, different ratios of systematic and random errors). The result is shown in figure 5.5, together with the original findings by Bortfeld et al. We find that optimization of systematic errors leads to a later moment of intervention than quadratic errors. Furthermore, we conclude that correction for systematic deformations is similar to corrections of (rigid) systematic patient setup errors.

**Figure 5.5.** A single intervention during treatment could improve accuracy, but requires an estimate of setup errors. Too little measurement will lead to an inaccurate estimate, while too many will not leave enough fractions to actually benefit from the intervention. If accuracy is measured by the overall quadratic error (left axis in black, normalized), the optimum number of measurements lies earlier than measured by the overall systematic error (right axis in red, normalized).
The threshold parameter $T$ allows selection of patients with large systematic local setup errors. A relevant range for $T$ is $2 - 5$ mm, yielding a possible reduction of systematic errors of 45%. The choice of $T$ depends on the clinical acceptable workload. Inspection of figure 5.4 reveals that half of the possible gain is already achieved with a 32% workload ($T = 4$ mm), i.e. replanning for 1/3 of the patients.

We used an average patient model to retrospectively calculate the consequences of plan adaptation. We thereby assumed that we would have a CT scan available with the patient’s anatomy in close agreement with the average anatomy during treatment. Although a repeat CT may reveal treatment response, e.g. weight loss and/or tumor shrinkage, it may also introduce new systematic errors (freezing random errors in the pCT) which could be of the same magnitude as the ones that we intend to solve. A possible solution may be to modify an existing (repeat) CT scan based on an average deformation vector field derived with non-rigid registration of a set of daily CBCT scans with the original pCT. However, this requires very accurate deformable registration with inaccuracies smaller than the systematic setup errors to which we want to adapt. Currently we are developing and validating the use of intensity-based deformable registration of CT to CBCT scans with deformation vector fields parameterized with BSplines [18, 19].

In H&N cancer, bony anatomy is often used as surrogate for the position of the tumor and OARs because lack of contrast and quality in portal images/CBCT scans seldom allows direct identification based on image intensity. Therefore, progressive anatomy changes, such as weight loss or tumor shrinkage, frequently occurring with H&N cancer patients, are hard to capture [1, 6]. Ultimately image guidance should be based on soft tissue structures. Therefore work should continue to improve the image quality of in-room imaging technology. However, the described methods still apply to other tumor sites where soft tissue registration is less challenging and might reveal progressive changes, requiring repeated interventions with plan adaptation.

A second limitation of this study is that the optimal couch correction is based on measured geometrical misalignments and not on dosimetric consequences. Optimal couch corrections should result in the maximum dose to the clinical target volume (CTV) while dose to OARs is within tolerances. Optimization of couch corrections on dose parameters not only requires repeated online dose calculation, also the CTV and OARs at the time of treatment should be identified. Yue describes this approach (limited to the dose to the CTV) together with some practical indications to reduce the total workload [20]. Several groups propose online replanning to fully cope with all deformations, for example Ahunbay [21]. Despite advances in non-rigid registration, dose calculation and online planning, these approaches are time consuming, difficult to validate, and not expected to be available soon for clinical application. This paper introduces simple geometrical methods that already capture an important part of deformations and can be applied in an online fashion, i.e. within a short time with tools that are already clinically available.
References


First clinical experience with a multiple region of interest registration and correction method in radiotherapy of head-and-neck cancer patients

S. van Beek, S. van Kranen, A. Mencarelli, P. Remeijer, C. Rasch, M. van Herk, J.-J. Sonke

Department of Radiation Oncology, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

E-mail: j.sonke@nk.nl

Published Februari 2010 in Radiotherapy and Oncology, volume 94, pages 213-217

dx.doi.org/10.1016/j.radonc.2009.12.017
Chapter 6

Abstract

Purpose
To discuss the first clinical experience with a multiple region of interest (mROI) registration and correction method for high-precision radiotherapy of head-and-neck cancer patients.

Materials and Methods
12-13 3D rectangular-shaped ROIs were automatically placed around bony structures on the planning CT scans \( n = 50 \) patients) which were individually registered to subsequent CBCT scans. mROI registration was used to quantify global and local setup errors. The time required to perform the mROI registration was compared with that of a previously used single-ROI method. The number of scans with residual local setup error exceeding 5 mm/5° (warnings) was scored together with the frequency ROIs exceeding these limits for three or more consecutive imaging fractions (systematic errors).

Results
In 40% of the CBCT scans, one or more ROI-registrations exceeded the 5 mm/5°. Most warnings were seen in ROI "hyoid", 31% of the rotation warnings and 14% of the translation warnings. Systematic errors lead to 52 consults of the treating physician. The preparation and registration time was similar for both registration methods.

Conclusions
The mROI registration method is easy to use with little extra workload, provides additional information on local setup errors, and helps to select patients for re-planning.
6.1 Introduction

Radiotherapy of head-and-neck (H&N) cancer patients requires accurate patient positioning using appropriate immobilization devices. Generally, small clinical target volume-to-planning target volume margins are applied to account for geometrical uncertainties. With the introduction of intensity-modulated radiotherapy (IMRT), having the ability to produce more conformal dose distribution and hence decrease the irradiated volume, local setup errors became increasingly important. There are several studies describing the setup accuracy in H&N cancer patients using portal imaging [1–3], stereoscopic kV devices [4], cone beam CT (CBCT) [3, 5–8], or in-room CT [9]. CBCT guidance quantifies setup errors by CBCT-to-planning CT registration. All these studies show considerable and frequent shape and posture changes over the course of radiotherapy. Nevertheless, rigid registration of the entire H&N region is the standard of care. Visual validation of the global rigid registration is therefore often difficult, because not all bony structures within the single large region of interest (ROI) can be aligned simultaneously in the presence of shape and posture changes. It is therefore not clear whether a miss registration occurred or a misalignment persisted due to shape or posture changes. Additionally, even in case of small global setup errors, local misalignments of more than 5 mm occur frequently [6], which would go unnoticed using registration of a single (large) ROI leading to possible under dosage of the target and/or irradiation of organs at risk [6]. We therefore developed and clinically implemented a multiple region of interest (mROI) registration and correction method. Clinical requirements for the new method were the ability to reliably quantify global setup errors in the presence of shape and posture changes, the ability to quantify local setup errors to define patients eligible for re-planning, robustness, ease of use, visual validation and efficiency. The purpose of this study was to evaluate the clinical introduction and the first clinical experience of this mROI driven setup correction method.

6.2 Materials and methods

Patient group

Fifty patients treated for H&N cancer with curative intent using IMRT with a total dose of 70 Gy \((n = 45)\), 66 Gy \((n = 1)\) or 60 Gy \((n = 4)\) were included in this study. Twenty six patients were treated for oropharynx cancer, eight for larynx cancer, five for base of tongue cancer, five for cancer in the oral cavity, four for hypopharynx cancer and two for nasopharynx cancer. No selection was made on tumor stage. All patients were positioned with a five point thermoplastic mask (Orfit Industries, Wijnegem, Belgium), a standard head rest and a knee support (Civco Medical Solutions, Kolona, USA) for stability and comfort. The planning CT scan (Somatom Sensation Open, Siemens AG, Erlangen, Germany) consisted of images with a slice distance of 3 mm from the cranium to the sternum. The isotropic margins used for IMRT in patients with H&N cancer in combination with an offline setup correction protocol are 5 mm. A total of 578
CBCT scans ($n = 50$) were evaluated with an average of 11.6 CBCT scans per patient (range 6–37). Seven patients had daily CBCT imaging (based on a trial protocol) with an average of 30 CBCT scans per patient (range 19–37). The other 43 patients, following a regular offline correction protocol, had an average of 8 CBCT scans per patient (range 6–11). In this study we corrected for this difference in scanning protocols.

**Radiation treatment and cone beam CT guidance protocol**

The patients were positioned using localization lines on the mask and skin. The cross of the localization lines represented a reference point that was positioned at the machine iso-center. Next a table shift was performed to move the patients to the planned treatment position. Prior to the beam delivery a CBCT scan was acquired. The patient setup on the linear accelerator (Elekta, Crawley, UK) was corrected using a shrinking action level (SAL) protocol which is employed with imaging at the first two fractions followed by weekly checks [6]. CBCT scans (Synergy 3.5 augmented with in-house developed software) were acquired with 16 mA – 40 ms – 120 kV per projection, while ±300 projections were acquired over 360° gantry rotation in 1 min with a dose of $\sim 1$ cGy [10]. Setup errors were measured using chamfer matching on bony structures of the CBCT to the planning CT [11]. The offline correction protocol was driven by the average setup error over the mROI as described below.

**Multiple region of interest registration**

Twelve 3D rectangular-shaped ROIs (figure 6.1) were defined around individual bones with an automatic segmentation method: First, bony anatomy in the planning CT is segmented. Subsequently, vertebrae around the spinal cord are separated based on projected shape analysis. Remaining structures in the head-neck region are found based on their approximate relative position to the vertebrae. This method was validated on 31 cases [12]. The bony structures consisted of cervical vertebrae C1–C7, mandible, hyoid, jugular notch, occiput bone and larynx. Optionally an additional 13$^{\text{th}}$ ROI was manually placed around the maxilla.

The patients with nasopharynx, hypopharynx, oropharynx, larynx and oral cavity carcinomas were verified with the mROI registration method. The carcinomas not verified with the mROI registration method were the parotid gland, nasal cavities and brain tumors since no nodal irradiation was performed in these cases and the impact of shape variation during radiotherapy was estimated to be low. Twelve ROIs were used for the hypopharynx-, larynx- and tongue-based carcinomas. The extra ROI was included for the nasopharynx, oropharynx and oral cavity carcinomas.

Per fraction, planning CT to CBCT local rigid registrations were performed for each ROI producing a set of 12 or 13 translations and rotations. As rotations in the Synergy system are expressed with respect to the iso-center, these sets were converted into the effective displacement of the geometrical center of each ROI. Subsequently, the average of all displacements was calculated to
Clinical introduction of mROI registration & setup corrections

Figure 6.1. Multiple regions of interest registration method on bony structures. (A) Definition of 13 regions of interest in the sagittal view. (B) Example registration of vertebrae C3 in green/purple overlay, note that other bony structures are not aligned. (C) Single assessment of all registration with a thin-plate-spline-based deformation.

drive the offline decision rule as described above. The residual displacements and the rotations were analyzed by a separate decision rule as described below.

Multiple region of interest clinical workflow

Prior to all treatment fractions, the reference data need to be prepared. This includes importing the planning CT scan together with the iso-center and delineated structures, selecting the registration and correction method and the correction reference point. Registrations (CBCT scan on the planning CT scan) of all 12–13 ROIs were performed automatically (bony anatomy) followed by visual validation of all regions. For a small region it is easy to visually validate the registration even in the presence of (global) shape and posture changes using three orthogonal planes. To get an overall view of the individual registrations a single image was generated using thin-plate-spline deformations [13]. Image fusion of the planning CT scan and deformed CBCT scan using complementary color overlay or cut-view allows visual validation of the registration (figure 6.1a and b). The registrations were always checked twice by different radiotherapy technicians (RTTs). If a certain ROI was not registered accurately it was possible to correct the registration with a different starting point, a different registration algorithm, or a manual registration.

Decision rule for deformations

Residual local setup errors (deformations) were defined as the center-of-mass displacement minus the average displacements of all ROIs. The (residual) rotations are with respect to the center-of-mass and represent changes in orientation and were not corrected in the absence
of a tilt and roll couch. Although such orientation changes do not shift the center-of-mass of the local anatomy, they do indicate considerable change relative to the planning situation and indicate local shifts away from the center-of-mass. The registration software automatically gives a warning for residual local setup errors larger than 5 mm and/or 5° for any orthogonal direction/axis. All warnings were checked and noted on a special form attached to the treatment chart. If any ROI had a local error in one particular direction for three consecutive imaging times, the treating physician was consulted to discuss possible re-planning. The limits of 5 mm and/or 5° are a compromise between accuracy and workload.

**Training RTTs for using the multiple region of interest registration method**

The mROI registration method was initially introduced on only one treatment machine executed by a team of 20 RTTs. After this initial period, the method was gradually introduced on different treatment machines and different RTT teams. After seven months, all treatment machines with CBCT (a total of five linear accelerators) were using the mROI registration method.

The mROI registration method was developed by a translational-research group. After completing the method, the translational-research group developed the clinical protocol, wrote a manual, generated a database for practice and prepared an instructive presentation. First, the dedicated imaging RTTs (consisting of five RTTs who supervise the other RTTs regarding clinical imaging tasks) had a training session with an interactive presentation (20–30 min) pointing out changes in workflow and showing changes that had been made in the software. They then received an individual training session on generating the mROIs and adjusting them if necessary. Second, the first RTT group received the same interactive presentations. Four RTTs of this group also received the individual training session to assist the dedicated imaging RTTs during the introductory period. Later, the other three RTT groups were familiar with the ROI registration method. They followed one training session (given twice) with the same interactive presentation.

One RTT-member of the translational-research group was responsible for the complete clinical implementation and remains available for advice on the generation and registration of mROIs.

**Evaluation**

The evaluation of the clinical implementation of the mROI was performed by (1) comparing the preparation and registration time (consisting of alignment and visual validation) with those of the single large ROI, and (2) its ease of use. Additionally, (3) the number of warnings for residual local setup errors exceeding the 5 mm and/or 5° threshold was scored. (4) Systematic misalignments were evaluated by scoring the number of warnings persisting at least three consecutive imaging fractions. A correction to the data was made for daily imaging vs. non-daily imaging.
6.3 Results

Ease of use

The mROI method was introduced into the clinic in January 2009. In the first three months, the preparation time for the mROI method was about 15 min. During this period, 14 patients were treated with this new method. After three months the preparation took about 10 min, which was similar for the preparation time of a single large ROI procedure.

The actual registration time of the mROIs is approximately 10 s, which is somewhat longer than that for the single-ROI method (1 s).

The validation of the mROI is easier than that for the single large ROI due to the thin-plate-spline deformation visualizing the registration result of all ROIs in a single view (figure 6.1c). Moreover, the mROI method provides accurate registration in the presence of shape and posture changes. Typically, one ROI in every CBCT needed to be re-registered, which takes another 10 s. Re-registrations were mostly necessary in case of ROIs with small bony structures or large rotations. Overall, both single and mROI registration methods take 1–2 min including validation.

Local setup errors

The number of scans with residual local setup error exceeding 5 mm and/or 5° is shown in figure 6.2. In 40% of the CBCT scans, one or more ROI-registrations exceeded the limits. The hyoid exhibited the most frequent warning for rotations (31% of the CBCT scans) as well as for translations (14%). The smallest number of rotational warnings was observed for the maxilla (1%) per CBCT scan. No warnings on the translation were observed for ROI C5. Note that multiple ROIs per scan can exceed the limits simultaneously.

The number of times a ROI exceeded the limits in a certain direction/axis for at least three consecutive imaging fractions is shown in figure 6.3. Three consecutive imaging fractions mean a time span of 3 days in the daily CBCT-group and 3 weeks in the non-daily CBCT-group. The ROIs of the hyoid (13 times) and the larynx (12 times) frequently had persistent local residual setup errors. Structures C1, C5 and C6 never exhibited such systematic local setup errors. In case a ROI exceeded the limits for three consecutive imaging fractions the treating physician was consulted to discuss adaptive re-planning to mitigate the systematic deformation. In total, the treating physician was consulted 52 times (in 20 patients).
Chapter 6

Figure 6.2. Number of warnings larger than 5 mm and/or 5° for each region of interest. Number of warnings in percentage (local setup error >5 mm and/or 5°) per region of interest for a total of 578 CBCT scans in 50 patients divided in translations and rotations. Note that multiple regions of interest can exceed the limits simultaneously.

6.4 Discussion

In this paper we described the clinical introduction of, and first clinical experience with, a multiple region of interest registration and correction method for high-precision radiotherapy of H&N cancer patients.

The new method is easy to use, and gives little extra workload. Only for the first couple of patients the preparation time was somewhat longer. After that the RTTs became accustomed to the method and limited extra workload was observed compared to the preparation time of the single large ROI. The individual training sessions on generating the mROIs were useful for quick answers to questions and improve the ability of the RTTs to identify all relevant bony structures.

While the new method required little more time to operate, it provided additional information on residual local setup errors, as illustrated in figure 6.2. In this paper we only quantified the residual local setup errors in terms of the frequency a ROI exceeded limits larger than 5 mm and/or 5°. Statistical analysis on the actual local setup errors and required margins falls outside this study and has already been discussed in the literature [5, 6, 9, 14]. It is difficult to compare our results quantitatively with earlier studies [1–3, 5, 6, 9, 14]. All used different imaging devices,
Clinical introduction of mROI registration & setup corrections

Figure 6.3. The number of times a region of interest exceeds limits for at least three consecutive imaging fractions divided in translations and rotations.

different analysis methods and different ways to quantify the non-rigid geometrical uncertainties. Qualitatively, however, all studies observed considerable local residual misalignments despite the use of immobilization devices. The main difference between our study and earlier ones is that we prospectively evaluated shape and posture changes while others did a retrospective study. This is the first study describing clinical implementation using mROIs in head-and-neck cancer patients. We report on the first year of clinical experience using our mROI method that was introduced into the clinic in January 2009 to manage local setup errors and in some cases adjusting the treatment plan. In the current clinical protocol, the number of times a ROI exceeded the limits for at least three consecutive imaging fractions for one particular direction was scored. This was a simple way to discriminate systematic from random local setup errors. In case of systematic discrepancies between planning and actual treatment delivery, adaptive re-planning is capable of reducing geometrical uncertainties. Most warnings, however, did not lead to re-planning. In this study the treating physician was consulted 52 times (in 20 patients) one patient was re-planned. This was due to warnings made by ROIs that were not important in that particular case. A ROI might represent the clinical target volume (CTV) in one patient while in another patient it represents an organ-at-risk. For example, ROIs C2–C4 are an accurate surrogate for the spinal cord for a base of tongue carcinoma while also representing the CTV of a hypopharynx carcinoma. It is therefore important to discuss every patient, with a ROI exceeding limits for three consecutive imaging times, with the treating physician. Over the first 50 patients it was observed that the number of actual adaptive plan modifications was very small. Currently we are optimizing the protocol for efficiency and efficacy [15]. In the future, a further refinement on the
use of each ROI will be investigated to adjust the mROI registration method to a specific tumor site.

Note that currently, the mROI method only considers changes in the position of bony structures. While this is an accurate method to monitor posture changes, it will be relatively insensitive to changes in anatomy due to treatment response such as weight loss and tumor regression. To monitor such anatomical changes, more advanced deformable registration methods [16, 17] are required. Such methods, however, currently lack adequate validation preventing large scale clinical implementation. The mROI method, which is based on the registration of bony anatomy with chamfer matching, having a reported sub-millimeter accuracy [11], is very simple and already captures an important component of non-rigid geometrical uncertainties.

Currently RTTs are responsible for noting the warnings given by the software to check the number of times a ROI had exceeded the limits. In future software updates will be aimed at automating this part of the work. This requires the software driving the decision rule to read and average over multiple ROIs. Based on the results of our study we are also optimizing patient fixation, using a combined individual head and shoulder support.

### 6.5 Conclusions

A multiple region of interest registration and correction method for high-precision radiotherapy of H&N cancer patients was clinically implemented and evaluated. The method is easy in use and gives little extra workload. Moreover, the multiple region of interest registration method provides additional information on local setup errors and helps to objectively select patients for adaptive re-planning.
References


Deformable image registration for adaptive radiotherapy of head and neck cancer: accuracy and precision in the presence of tumor changes

S. van Kranen\textsuperscript{1}, A. Mencarelli\textsuperscript{1\ast}, O. Hamming-Vrieze\textsuperscript{1}, S. van Beek\textsuperscript{1}, C. Rasch\textsuperscript{2}, M. van Herk\textsuperscript{1}, J.-J. Sonke\textsuperscript{1}

\textsuperscript{1}Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands
\textsuperscript{2}Department of Radiation Oncology, Amsterdam Medical Centre, Amsterdam, the Netherlands
\textsuperscript{\ast} Equally contributed

E-mail: j.sonke@nki.nl

Published November 2014 in International Journal of Radiation Oncology, Biology and Physics, volume 90(3), pages 680-687

dx.doi.org/10.1016/j.ijrobp.2014.06.045
Abstract

Purpose
To compare deformable image registration (DIR) accuracy and precision for normal and tumor tissues in head and neck cancer patients during the course of radiation therapy (RT).

Methods and Materials
Thirteen patients with oropharyngeal tumors, who underwent submucosal implantation of small gold markers (average 6, range 4-10) around the tumor and were treated with RT were retrospectively selected. Two observers identified 15 anatomical features (landmarks) representative of normal tissues in the planning computed tomography (pCT) scan and in weekly cone beam CTs (CBCTs). Gold markers were digitally removed after semiautomatic identification in pCTs and CBCTs. Subsequently, landmarks and gold markers on pCT were propagated to CBCTs, using a b-spline-based DIR and, for comparison, rigid registration (RR). To account for observer variability, the pair-wise difference analysis of variance method was applied. DIR accuracy (systematic error) and precision (random error) for landmarks and gold markers were quantified. Time trend of the precisions for RR and DIR over the weekly CBCTs were evaluated.

Results
DIR accuracies were submillimeter and similar for normal and tumor tissue. DIR precision (1 SD) on the other hand was significantly different ($P < .01$), with 2.2 mm vector length in normal tissue versus 3.3 mm in tumor tissue. No significant time trend in DIR precision was found for normal tissue, whereas in tumor, DIR precision was significantly ($P < .009$) degraded during the course of treatment by 0.21 mm/week.

Conclusions
DIR for tumor registration proved to be less precise than that for normal tissues due to limited contrast and complex non-elastic tumor response. Caution should therefore be exercised when applying DIR for tumor changes in adaptive procedures.
7.1 Introduction

Over the course of radiation therapy (RT), head and neck cancer patients undergo weight loss, tumor response, and posture changes [1–4]. Rigid registration (RR), optimizing only translations and rotations, cannot account for these geometric uncertainties. On the contrary, deformable image registration (DIR) allows the alignment of datasets in a nonlinear way, providing a voxel-to-voxel mapping between the scans. Therefore, several groups are developing dose accumulation [5] and [6] and adaptive replanning strategies fLu2006a,Chao2008 based on DIR.

However, soft tissue contrast in these image modalities is often insufficient to capture the tumor-normal tissue interface. Accurate and precise quantification of tumor variations might thus be challenging for DIR [7, 8] and [9].

Because the purpose of applying DIR is to support clinical decisions, the quantification of accuracy (systematic error) and precision (random error) in all the representative tissues over the course of treatment is an essential part of the DIR development process and its clinical implementation [10].

However, poor visualization and the fact that the tumor may have considerable geometric variation not congruent with the surrounding anatomical tissues [11] hamper DIR validation of the tumor region.

To evaluate the capability of DIR to adequately describe normal tissues changes and tumor response during treatment, we used implanted gold markers around the tumor bed, representative of the tumor borders; we evaluated the DIR performance in the presence of tumor variation, while using manually placed anatomical features (landmarks) representative of the normal tissues; and we evaluated the performance for the normal tissue changes.

The quality of the validation method is hampered by the ability with which gold markers and landmarks can be identified. To avoid the influence of observer errors, the gold markers were automatically identified, whereas to identify the landmarks, a methodology was applied which separated registration and observer errors [12].

The purpose of this study was to compare DIR accuracy and precision of normal tissues with those of tumor tissues in head and neck cancer patients over the course of RT.
Chapter 7

7.2 Methods and Materials

Patients

Twenty-seven head and neck cancer patients, prospectively selected in consecutive order, with transoral accessible tumors underwent investigation under anesthesia, during which 0.35 × 2 mm pieces of Visicoils™ (RadioMed Corporation, Tyngsboro, MA, US) (gold markers) were implanted submucosally around the tumor bed. To reduce the workload for normal tissues validation, 13 of these patients (5 tonsilar, 4 posterior pharyngeal wall, and 4 base of tongue) were retrospectively selected for this study. All these patients presented calcifications at left and right carotids. On average, 6 gold markers (range 4–10) were implanted per patient. The tumor deformations for these patients were assessed in a previous study [11] by an expert radiation oncologist (OH). One patient presented a stable tumor volume during the treatment, 3 presented tumor progression, and 9 had regression.

Patients underwent cone beam computed tomography (CBCT)-guided RT with curative intent. The RT dose was 70 Gy given in 35 fractions in 6 or 7 weeks with concomitant chemotherapy with cisplatinum. Daily CBCT scans were acquired immediately before the treatment for setup verification. On average, 32 scans per patient were available for analysis.

Patients were immobilized in a five-point thermoplastic fixation mask with shoulder fixation (Civco Medical Solutions, Kolona, IA, US) with the head tilted backward by means of a patient-specific neck rest. Planning CT (pCT) scans (Somaton Sensation Open; Siemens AG, Erlangen, Germany) were reconstructed with a voxel size of 0.8 × 0.8 × 3 mm³. The CBCT (Elekta Synergy, Elekta Oncology Systems Ltd, Crawley, UK) scans were acquired with an energy of 120 kV and an isocenter dose of approximately 1 cGy and were reconstructed with a voxel size of 1 × 1 × 1 mm³.

Deformable image registration method

One of the most common DIR methods used in head and neck cancer RT is the B-spline algorithm [13, 14], which is an image intensity-based registration algorithm based on a regular grid of control points. We implemented a cubic B-spline deformable registration algorithm with rigidity and volume constraints driven by a correlation ratio [15] with 128 bins. Gradient descent-based multiresolution optimization [16, 17] was performed, with a final control point spacing (CPS) of 5 mm. The DIR was started after a global RR was performed by a chamfer-matching algorithm and simplex optimizer [18]. A large region of interest including tumor, cervical vertebrae, larynx, hyoid, mandible, and manubrium was defined to perform RR and define the work volume for the deformable registration. The gold markers were digitally removed before performing the DIR. Furthermore, the influence of the control point grid was evaluated, decreasing CPS to 2 mm.
Deformable registration of tumor

Evaluation methods

We selected a total of 102 of the 415 available daily CBCT scans, 7 to 8 per patient, with a rate of 1 for every 4 to 5 fractions (weekly). This selection was made as a compromise between workload and statistical power for the pair-wise difference analysis-of-variance (PWD-ANOVA) method [12]. We evaluated the performance of two human observers to identify the landmarks in the healthy tissues and the performance of registration methods to register the tumor and healthy tissues in terms of their accuracy and precision: defining the accuracy as the closeness of measurements to the true value (systematic error) and precision as the dispersion of the measurements (random error).

Tumor borders

To avoid observer variation, the gold markers were semiautomatically identified on pCT and CBCTs: first, a gold marker position was roughly indicated by a human observer, and then the center of mass was calculated by a threshold based auto-segmentation (figure 7.1a and b).

We evaluated the influence of gold marker migration in a previous study by using the same marker set, concluding that this was negligible (ie, the position variability of the gold markers was largely dominated by surrounding tissues and not by marker migration) [11].

We registered the adjacent bony anatomy on CBCT to pCT to evaluate the gold marker position variability, and consequently, the residual displacement of each gold marker was calculated in reference to the position on pCT [11].

Subsequently, to validate the DIR, pCT gold marker positions were propagated to the CBCT, and the residual misalignments were quantified. Finally, time trends in the residual misalignments following RR and DIR during the course of treatment were evaluated by linear regression.

Normal tissues

Fifteen landmarks distributed in soft tissues 7.1 over a region from the palatine bone to the jugular notch (figure 7.1d) were independently identified by two observers [12]. One expert observer identified the landmarks on the pCT of each patient, and afterwards, the same observer localized the landmarks on the CBCT. The second observer identified the anatomical landmarks on the CBCTs. Subsequently, the landmarks on the pCT were automatically propagated to the CBCT by using RR and DIR. Each observer identified 1530 landmarks.

Analysis of variance

To quantify the precision of the two expert observers and the algorithms, we applied the PWD-ANOVA method [12], which calculates the random variation of an individual observer
Figure 7.1. (a and b) Semiautomatic marker identification. (a) Gold marker numbers 7 and 8 are shown in pCT and CBCT (b). (c) Positions of the markers in green-purple overlay image of deformed pCT (purple) and CBCT (green) are shown. (d) Identified normal tissue landmarks are represented by black dots on pCT. Three landmarks on the opposite side are not visible. CBCT = cone-beam computed tomography; pCT = planning computed tomography.
Deformable registration of tumor

Table 7.1. Anatomical landmarks.

<table>
<thead>
<tr>
<th>IN</th>
<th>Anatomical Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Uvula top</td>
</tr>
<tr>
<td>2</td>
<td>Epiglottis</td>
</tr>
<tr>
<td>3, 4</td>
<td>Left and right external jugular vein bifurcation</td>
</tr>
<tr>
<td>5</td>
<td>Dorsal cricoid cartilage</td>
</tr>
<tr>
<td>6, 7</td>
<td>Left and right middle neck muscle bifurcation</td>
</tr>
<tr>
<td>8, 9</td>
<td>Left and right carotid calcification</td>
</tr>
<tr>
<td>10</td>
<td>Philtrum (upper lip)</td>
</tr>
<tr>
<td>11, 12</td>
<td>Left and right auricula and head contact point</td>
</tr>
<tr>
<td>13</td>
<td>Jugular notch</td>
</tr>
<tr>
<td>14, 15</td>
<td>Skin neck left, right (air/soft tissue interface)</td>
</tr>
</tbody>
</table>

The identification number (IN) of the landmarks is shown in the first column; in the second column is the anatomical location. Skin neck left and right were positioned in the middle between skull and shoulders to evaluate the neck air/soft tissue interface.

by examining the variations in the differences between the measurements of the observers (human or automatic).

The PWD-ANOVA method is sensitive to the presence of outliers [12, 19]. We therefore applied the Mahalanobis distance (with significance level for the critical values equal to 0.01 [20]) to detect the outliers by removing these from the data and counted prior to estimating the variances [12].

The systematic deviations of the algorithm propagations were computed relative to the mean of the measurements of the two human observers [12]. We analyzed the origin of each outlier to evaluate the contribution of the expert disagreement to the outlier generation. If an outlier was found in the human observer pair it was attributed to the human observers. In other cases it was attributed to the registration algorithm [12].

Finally, we evaluated the accuracy and precision of the landmarks on the weekly CBCTs and performed linear regression to detect time trends during treatment.

Statistical analysis and software implementation

Precisions of landmark and the gold marker observations were tested for statistical significance by applying the $\chi^2$ test and computing the 95% confidence interval (CI) with $\chi^2$ distribution [21]. The two-tailed t-test was applied for the accuracy measurements. The F test was applied to evaluate whether the precision of the two samples was significantly different, and 95% CI intervals were computed [21]. The CIs and two-tailed t-tests were computed for the regression slopes. Finally the t-test (two-tailed) for the difference of the DIR regression slopes were computed.

Routines to perform B-spline DIR and to identify landmarks and segment gold markers were implemented in C++ and Delphi software. Matlab version 2009a (MathWorks Inc, Natick, MA) was used for the PWD-ANOVA and regression analyses.
7.3 Results

Normal tissues

DIR calculation time was, on average, 26 minutes (± 7 minutes). The results of RR and DIR precision and accuracy are presented in Table 7.2. The human observer precision ranged from SI0.8 to 1.8 mm. DIR precision ranged from 1.0 to 1.6 mm in each direction. All precision results were statistically significant (P < .0001). The accuracy after RR and DIR was computed with respect to the human observer average. All results were submillimeter.

<table>
<thead>
<tr>
<th>Observer</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
<th>3D</th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>2.0</td>
<td>2.5</td>
<td>2.3</td>
<td>3.9</td>
<td>-0.11</td>
<td>0.05</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>DIR</td>
<td>1.0</td>
<td>1.6</td>
<td>1.2</td>
<td>2.2</td>
<td>0.15</td>
<td>-0.57</td>
<td>-0.02</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table 7.2. Landmarks registration precision and accuracy.

Decreasing the CPS to 2 mm had negligible (<0.02 mm) impact on accuracy and precision (see appendix 7.A). The results indicate that the random variation is considerably larger than the systematic deviation, and subsequently, the performance after DIR and RR can be described by their precision.

Twenty-four outliers were detected, 1.6% of the total number of landmarks. The analysis revealed that the two human observers accounted for 62% of the total outliers and that the DIR algorithm accounted for 38%. The origin of the outliers was due to the limit of registration algorithm to handle drastic changes of the tissues continuity (e.g., seven outliers were generated by attached/detached uvula [landmark 1]) to surrounding soft tissues.

The time trend during the course of treatment is shown in figure 7.2a and b together with the 95% CI for the regression line of the RR and DIR performance. On the x-axis are the weekly selected CBCTs; on the y-axis are the precision vector lengths of the RR and DIR. The slope of the regression line for RR was 0.26 mm/week (95% CI, from 0.16 to 0.37 mm; P = .001), with an $R^2$ value equal to 0.86, indicating that there was a significant time trend due to progressive anatomical changes. The slope of the regression line for the DIR was 0.08 mm/week with $R^2 = 0.32$, $P = .14$, and 95% CI from −0.04 to 0.21 mm/week, indicating no significant time trend.

Tumor borders

During treatment, 2 of the 78 implanted markers were lost before the first CBCT and, consequently, were removed from the analysis. A total of 573 valid measurements were analyzed.
Figure 7.2. Regression line and 95% CI of the soft tissue landmarks’ precision after RR (a) and after DIR (b). On the y-axis, the precision vector length for the soft tissue landmarks analysis. On the x-axis, the weekly selected CBCTs. Regression line and 95% CI for the gold markers after RR (c) and after DIR (d). On the y-axis, the vector length of the residual displacement average. CBCT = cone beam computed tomography; DIR = deformable image registration; RR = rigid registration.

Table 7.3 contains the gold marker precision and accuracy after RR, DIR. The accuracy values in boldface were significantly different from zero ($P < .05$). All the precision values were statically significant ($P < .0001$).

**Table 7.3.** Gold marker registration precision and accuracy.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
<td>CC</td>
</tr>
<tr>
<td>RR</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>DIR</td>
<td>2.0</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Gold marker registration precision (1 SD, mm) and accuracy (mm) after rigid registration (RR) and deformable image registration (DIR) according to axis (LR = left-right; CC = cranial-caudal; AP = anterior-posterior) and vector length (3D). Accuracy values in boldface are significantly different from zero.

RR and DIR performance as a function of time are shown in Figure 7.1 and d, respectively. For the RR, the regression line slope of the average of the residual displacement (vector length) was...
0.49 mm/week with an $R^2$ value equal to 0.98 ($P < .0001$; 95% CI, from 0.42 to 0.56 mm/week). The DIR regression line slope of the average of the residual displacement was 0.21 mm/week with an $R^2$ equal to 0.71 ($P = .009$; 95% CI from 0.07 to 0.34 mm/week) indicating a significant time trend. DIR performance decreased during the course of treatment, with a vector length on average at the end of the treatment for precision of 3.9 mm.

The vector length of the DIR precisions and the regression slopes of landmarks and gold markers were significantly different ($P < .0001$ and $P = .05$, respectively).

Differences in precision and accuracy of $\leq 0.02$ mm were found, decreasing the CPS to 2 mm. Furthermore, digitally removing the implanted markers had negligible impact on the gold marker results with differences in precision and accuracy of $\leq 0.04$ mm and $\leq 0.05$ mm, respectively (see Supplementary material).

Examples of four patients are shown in figure 7.3a-d. Each graph demonstrates the vector length averages of residual displacement after RR and DIR. The horizontal line represents the DIR-validated precision of the normal tissues. Figure 7.3a shows an example of small tumor variation and high registration precision during treatment. Figure 7.3b, c, and d are three examples of increasing gold marker misalignments during treatment after rigid registration. In figure 7.3b, the vector length of the residual displacement average is reduced by the DIR, and its magnitude during the treatment is close to the validated normal tissue precision. In figure 7.3c, the DIR performance decreased during treatment. A worst case scenario is shown in figure 7.3d. Here, the initial performance progressively degraded to a vector length larger than 4 mm at the end of treatment. Notice that the vector length of the residual displacement average after the DIR is similar to the RR vector length.

## 7.4 Discussion

In this study, we showed a significant difference in image-based DIR precision levels between tumor and normal tissues in head and neck cancer patients. Furthermore, although the DIR precision was characterized by a stable-in-time performance in normal tissues, this degraded in the tumor over the course of treatment. Consequently, DIR for dose accumulation and adaptive RT for head and neck cancer patients should be applied carefully.

The gold markers were implanted in the gross tumor volume (GTV)-to-clinical target volume (CTV) margin region, close to the GTV surface, and in general, the acceptance of the error in case of dose deformation is determined by the DIR performance in combination with the dose gradient. This implies that in a uniform dose distribution to the target, the difference between planned and accumulated dose may be less sensitive to registration errors [22]. However, as shown in our previous study [11], some head and neck tumor borders demonstrate considerable position variability during the course of treatment. In non-uniform dose distribution (e.g., dose painting), the deformation error becomes more critical due to the presence of dose gradients either inside or close to the CTV. Furthermore, the capability of the DIR to capture the tumor borders has
Deformable registration of tumor

Figure 7.3. Examples from four patients. Vector length of the residual displacement average as a function of time is shown for all implanted markers after rigid registration (dashed line) and after deformable registration (continuous line). (a) A patient with a (small) tumor variation correctly registered; (b and c) tumor response partially registered; (d) tumor response inadequately registered.

...a critical clinical relevance for the CTV delineation propagation. The degradation of the DIR precision during treatment, resulting in relevant deformation errors, leads to uncertainties in the evaluation of the total tumor volume, the position of the contours, and the subsequent treatment plan.

Image-guided systems for setup verification such as CBCT are widely applied to follow the patient’s anatomical changes. In this context, DIR between CT and CBCT becomes important in order to estimate the delivered dose. However, low soft tissue contrast in changing tumor/tissue is a major challenge for DIR. This implies that the DIR algorithm does not detect the fact that tumor shape variability may not be congruent with the surrounding anatomical tissues nor that it may change in shape as well as in volume.

This also explains the progressive decrease in performance that in some cases (figure 7.3d), presented average residual displacements similar to the RR.

Figure 7.1c illustrates registration around the evolving tumor region. The negligible performance differences between digitally removed and original data sets indicate that small gold markers are typically ignored by the cost function of the B-spline algorithm. That means that the gold
Chapter 7

Marker performance is determined by the surrounding tissues and that the presence of complex tumor changes, such as regression, lead to misalignments.

Algorithms, such as finite element model [23] and the extra-dimensional Demons proposed by Nithiananthan et al [24], in which segmentation and registration are performed simultaneously, account for the disappearing tissues. However, tumor variation is difficult to model, and if there is insufficient contrast, the tumor is difficult to segment.

Results for the normal tissue were in agreement with those in the studies by Nithiananthan et al [25] (using high-resolution surgical C-arm CBCT), Nithiananthan et al [26], Hou et al [27], and Kashani et al [7, 28] (using CT-to-CT on phantom). Results indicate that the applied B-spline algorithm for head and neck is comparable with other image-based implementations found in published reports and that the DIR performance is not limited by image quality.

A limitation of this study is that we were not able to distinguish between the error generated by lack of soft tissue contrast, which hampers DIR performance, and complexity in tumor variability.

The reason for this is due to the fact we were comparing soft tissue landmarks, which had adequate contrast, with tumors with low contrast. However, this only had an impact on the interpretation of the role of the extent of tumor regression compared to the lack of soft tissue contrast, and it did not affect the clinical relevance of this study.

Another limitation is that the systematic deviations of the registration algorithm were measured relative to the mean of the two observers. This implies that the ground truth problem is completely avoided for only the random errors and that the model for systematic deviations does not ensure a correct computation of the DIR algorithm accuracy. This limits the PWD-ANOVA method to measuring only the precision. However, the systematic deviations of the DIR computed with respect to the average of the two observers were submillimeter (i.e., not clinically relevant). These results together with the DIR precision may be used to fully characterize the DIR performance.

The number and location of gold markers were limited by the accessibility of the tumor, and consequently, the coverage of all borders was not possible. This implies that the analysis of tumor volume progression/regression was hampered by the distribution of gold markers on a partial tumor border surface. Analysis of the progression/regression was consequently based on the experience of the radiation oncologist [11]. However, this had an impact only in the interpretation of the tumor displacement and did not affect the validation results.

We only validated a B-spline-based DIR. This choice was arbitrary, as the purpose of this study was to evaluate the accuracy variations of tumor registration with respect to normal tissues and not to validate a specific DIR algorithm. Furthermore, we have demonstrated that the performance of the applied DIR is comparable with the image-based implementations found in published reports. This indicates that our findings have a general impact on dose accumulation and adaptive replanning strategies based on deformable image registration and that DIR must be validated for time trends.
7.5 Conclusions

Deformable image registration in the tumor region of head and neck cancer patients was not capable of capturing the displacement of tumor borders, and precision degraded during the course of the treatment. This occurred despite the fact that high accuracy and precision were characterized by a stable time performance in normal tissues. Caution should therefore be exercised when using such algorithms to evaluate accumulated tumor doses.

7.6 Summary

Deformable image registration (DIR) is anticipated to play an important role in adaptive radiation therapy to quantify and account for complex anatomical changes. Here we compared DIR performance in tumor (using implanted gold makers) and normal tissue (using anatomical landmarks). Precision in the tumor region degraded during the course of treatment, even though high and stable-in-time performance in normal tissues was observed. Caution is therefore recommended when applying DIR to adapt to tumor changes.
Chapter 7

Appendix 7.A  Anatomical landmark definition and identification

Method

To quantify the potential bias introduced by the presence of gold-markers, the DIR was performed using the original data-sets \((DIR_{+5\,mm}^+)\) and digitally removed gold-marker data-sets \((DIR_{-5\,mm}^-)\) with a control point spacing (CPS) of 5 mm.

To evaluate the influence of the B-spline control point grid on accuracy and precision results, the digitally removed gold-marker data-sets were registered with 2 mm control point spacing \((DIR_{-2\,mm}^-)\).

Statistical analysis  Differences in precisions and accuracy between \(DIR_{-5\,mm}^-\) and alternative methods were tested for statistical significance by applying the F test [21] and the two-tailed t-tests [29] respectively.

Results

In table 7.4 are given the precision and accuracy of \(DIR_{+5\,mm}^+\), \(DIR_{-5\,mm}^-\), and \(DIR_{-2\,mm}^-\) in the tumor region.

Table 7.4. Gold-marker registration precision and accuracy with and without digital gold-marker removal.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
<td>CC</td>
</tr>
<tr>
<td>(DIR_{+5,mm}^+)</td>
<td>1.95</td>
<td>2.1</td>
</tr>
<tr>
<td>(DIR_{-5,mm}^-)</td>
<td>1.95</td>
<td>2.14</td>
</tr>
<tr>
<td>(DIR_{-2,mm}^-)</td>
<td>1.96</td>
<td>2.14</td>
</tr>
</tbody>
</table>

Precision (1SD, mm) and accuracy (mm) after 5 mm CPS DIR of data-sets with \((DIR_{+5\,mm}^+)\) and without digitally-removed gold-markers \((DIR_{-5\,mm}^-)\), and 2 mm CPS \((DIR_{-2\,mm}^-)\) per axis (LR = left-right, CC = cranial-caudal, AP = anteriorposterior)and vector length (3D).

Differences in precision and accuracy between \(DIR_{+5\,mm}^+\) and \(DIR_{-5\,mm}^-\) were \(\leq 0.04 \, mm\) \((p \geq 0.5)\) and \(\leq 0.05 \, mm\) \((p \geq 0.5)\), respectively. Decreasing the control point spacing to 2 mm had negligible \((\leq 0.02 \, mm; \ p \geq 0.5)\) impact on the result.

In table 7.5 the precision and accuracy of \(DIR_{-5\,mm}^-\), and \(DIR_{-2\,mm}^-\) in normal tissues are given validated using landmarks.

The results demonstrate that decreasing the control point spacing to 2 mm had negligible \((\leq 0.01 \, mm; \ p \geq 0.4)\) impact on the result. The difference in \(p\) values for the landmarks and gold-markers were attributed to the different sample sizes of the two analysis, 1506 observations for landmarks vs 573 for gold-markers.
Table 7.5. Precision (1SD, mm) and accuracy (mm) of DIR registrations for normal tissue.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
<td>CC</td>
</tr>
<tr>
<td>(DIR_{5\text{mm}})</td>
<td>1.05</td>
<td>1.61</td>
</tr>
<tr>
<td>(DIR_{2\text{mm}})</td>
<td>1.06</td>
<td>1.60</td>
</tr>
</tbody>
</table>

LR = left-right, CC = cranial-caudal, AP = anterior-posterior and 3D vector length. \(DIR_{5\text{mm}}\) is DIR with 5 mm CPS DIR of digitally-removed gold-marker data-sets, and \(DIR_{2\text{mm}}\) is 2 mm CPS DIR.
References


Adaptive radiotherapy with an average anatomy model: evaluation and quantification of residual deformations in head and neck cancer patients

S. van Kranen, A. Mencarelli, S. van Beek, C. Rasch, M. van Herk, J.-J. Sonke

Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

E-mail: j.sonke@nk.nl

Published December 2013 in Radiotherapy and Oncology, volume 109(3), pages 463-468

dx.doi.org/10.1016/j.radonc.2013.08.007
Abstract

Background and purpose
To develop and validate an adaptive intervention strategy for radiotherapy of head-and-neck cancer that accounts for systematic deformations by modifying the planning-CT (pCT) to the average misalignments in daily cone beam CT (CBCT) measured with deformable registration (DR).

Methods and materials
Daily CBCT scans (808 scans) for 25 patients were retrospectively registered to the pCT with B-spline DR. The average deformation vector field ($<DV F>$) was used to deform the pCT for adaptive intervention. Two strategies were simulated: single intervention after 10 fractions and weekly intervention with an $<DV F>$ from the previous week. The model was geometrically validated with the residual misalignment of anatomical landmarks both on bony-anatomy (BA; automatically generated) and soft-tissue (ST; manually identified).

Results
Systematic deformations were 2.5/3.4 mm vector length (BA/ST). Single intervention reduced deformations to 1.5/2.7 mm (BA/ST). Weekly intervention resulted in 1.0/2.2 mm (BA/ST) and accounted better for progressive changes. 15 patients had average systematic deformations $>2$ mm (BA): reductions were 1.1/1.9 mm (single/weekly BA). ST improvements were underestimated due to observer and registration variability.

Conclusions
Adaptive intervention with a pCT modified to the average anatomy during treatment successfully reduces systematic deformations. The improved accuracy could possibly be exploited in margin reduction and/or dose escalation.
ART with an average anatomy model

Introduction

Intensity modulated radiotherapy (IMRT) has been widely adopted for the irradiation of head-and-neck cancer (HNC) patients. The highly conformal dose distributions offer superior sparing of organs-at-risk (OARs) while malignant tissue is optimally irradiated [1]. Moreover, IMRT has been used for dose escalation to clinically demonstrated areas of radioresistant hypoxia [2].

Practically, the benefits of IMRT are limited by the accuracy with which the anatomy of the patient from the planning-CT (pCT) can be reproduced during treatment. Small misalignments, random or systematic, resulting from patient setup, posture or anatomy changes, can significantly influence the position and shape of the dose distribution delivered to the patient. To account for these geometrical uncertainties safety margins are applied [3]. A reduction of geometrical uncertainties allows smaller margins and may increase the therapeutic ratio. With image guided radiation therapy, patient setup errors can be determined and corrected with an opposite shift of the treatment couch. Daily imaging allows for near-perfect correction of patient setup errors, while with a limited amount of imaging the systematic component can be estimated and minimized with an offline correction protocol [4]. Anatomy and posture changes however, are non-rigid and therefore require a different approach.

With adaptive radiotherapy (ART) the treatment plan is adjusted to account for changes in anatomy and posture (deformations). At present, adaptive radiotherapy for HNC mainly deals with treatment response (progressive changes), such as weight loss or tumor shrinkage [5–7]. Therefore a properly timed repeat CT-scan (rCT) during treatment is a suitable basis for plan adaptation to account for treatment response. On the other hand, substantial systematic deformations, up to 3.5 mm, are present with HNC patients, despite extensive immobilization [8–10]. A rCT-scan is a snapshot of the patient's anatomy, subject to random deformations. By freezing an arbitrary pose of the patient, (new) systematic deformations are introduced that cannot be corrected with a couch shift. A rCT-scan is therefore inappropriate to correct systematic posture misalignments [7].

We propose a new adaptive strategy to reduce systematic deformations by modifying the pCT to the average anatomy (AA) observed in a repetitive imaging series during the initial fractions of radiotherapy as estimated with deformable registration (DR): the AA-model. Cone beam CT often provides such a repetitive imaging series as it is routinely acquired for patient position verification. Although CBCT image quality is typically somewhat lower than fan-beam CT, considerable improvements have been achieved in recent years (e.g. scatter correction [11, 12] and reconstruction (e.g. compressed sensing [13])). Moreover, various publications demonstrated that for the head-and-neck region current CBCT image quality is adequate for deformable image registration [14, 15], dose recalculation in CBCT-scans [16, 17], contour propagation to the CBCT [15, 18], dose remapping [19], and adaptive replanning [20]. In this study we evaluated a CBCT based AA-model to account for systematic deformations and quantified the reduction in geometrical uncertainties on anatomical landmarks for various adaptive approaches.
Chapter 8

Materials and Methods

Patient data

Twenty five HNC patients were retrospectively selected. Regular IMRT planning had been performed on a planning CT (Somatom Sensation Open, Siemens AG, Erlangen, Germany) with an in-plane voxel size of $1.0 \times 1.0 \text{mm}^2$ and slice distance of $3 \text{mm}$. Daily CBCT-scans (Elekta Synergy, Elekta Oncology Systems Ltd., Crawley, UK) were available (33 median, range 21-35), reconstructed at a voxel size of $1\text{mm}^3$. Patients were immobilized with a 5 point thermoplastic mask and positioned with a knee support and a standard head-rest (Civco Medical Solutions, Koloma, USA).

Adaptive treatment modification with an AA-model

In the AA-model we describe the local misalignment at position $\vec{r}$ and fraction $f$ by $\vec{u}_f(\vec{r})$. The complete set of displacement vectors for a particular fraction is called the deformation vector field ($DVF$). In this study we determined the $DVF$ with CT-to-CBCT deformable registration (DR).

As with conventional patient setup [3], a series of local misalignments contains a systematic and a random component, quantified by the average and the standard deviation respectively. Consequently we estimated systematic deformations in a series of $N$ fractions with the average local misalignment:

$$<\vec{u}(\vec{r})> = \frac{1}{1 + N} \sum_{f=1}^{N} \vec{u}_f(\vec{r})$$  \hspace{1cm} (8.1)

Extending this to all positions is equal to averaging the deformation vector fields: $<DVF>$. We divided by $N + 1$ since the pCT is also considered a sample of the random position of the anatomy.

For adaptive treatment modification, we applied the $<DVF>$ to the pCT to propagate the local systematic misalignments and thereby generated a sharply defined mCT in which the systematic deformations were eliminated (figure 8.1). Note that the technique to generate a new CT-scan from a set of DVFs is already clinically practiced for treatment planning of lung cancer patients where the mid-position anatomy is derived from all phases of a 4D-CT [21].

Discrepancies in the remainder of the treatment are due to (1) imprecise measurements of the local misalignments, (2) residual uncertainties in the estimation of the systematic local misalignments (limited statistical power in estimating the mean value of a distribution), and/or (3) progressive changes (a non-stationary distribution).
Figure 8.1. Adaptive intervention with an average anatomy model. Daily deformations are calculated with non-rigid registration of CBCT scans to the planning CT. Systematic deformations are estimated with the average over a series of deformation vector fields. Application of the average deformations to the planning CT results in a modified CT for plan adaptation.

Treatment simulation

To quantify the geometrical accuracy with which the patient anatomy is reproduced during treatment, we started by defining a set of atlas-based landmarks in the pCT. Subsequently, these landmarks were identified in the daily CBCT-scans (see below). As baseline geometrical accuracy, we calculated the residual landmark misalignment (CBCT minus pCT position) after online couch shift corrections (referenced further by shift corrections). Next we simulated two possible adaptive approaches with the AA-model: single intervention and weekly intervention. The simulation was performed as follows: upon an intervention we repositioned the landmarks in the pCT according to the displacements from the AA-model. Next, the residual landmark misalignments for the remaining fractions after the intervention were calculated by subtracting this new position from the CBCT position. Simulations did not require actual replanning and were thus independent of a planning technique. With weekly interventions, we hypothesized that we could better follow progressive changes than a single intervention protocol. In addition, we simulated a regular repeat CT intervention by repositioning the pCT landmarks according to the deformations of a single fraction. Finally, to quantify the optimal achievable performance given
the finite accuracy of landmark identification and deformable image registration, we calculated the AA from all fractions and determined the residual systematic misalignments over all fractions relative to this AA (validation series).

Anatomical landmark definition and identification

The proposed protocols were validated with the residual misalignments of clearly identifiable, atlas-based, landmarks on bony anatomy (BA) and soft tissue (ST). Per patient 47 BA landmarks were defined on 12 bony structures (collected in 4 groups) in the pCT. The high contrast of the BA landmarks allowed the use of an automatic method to define these landmarks in the pCT and identify them in all available fractions with sub-millimeter precision [22]. Additionally, 14 ST landmarks were defined on the planning CT for 11 patients by an expert research technologist. Two observers independently identified these landmarks in weekly CBCT-scans. With two observers quantification of observer variability was possible (see below). ST Landmarks describing similar anatomical structures were taken together into 9 subgroups. More details on landmark definition and identification can be found in the appendix 8.A.

Deformable registration

Non-rigid registration was performed with in-house developed software, applying B-spline deformations as described by Rueckert et al. [23], Mattes et al. [24] and Kybic and Unser [25] with correlation ratio [26] as cost function and regularization terms [27] to cope with limited CBCT quality. The registration was performed after setup corrections, in the frame of reference of the pCT with a region of interest encompassing the anatomical landmarks. A 5 step coarse-to-fine multi-resolution approach was applied with a final B-spline control grid spacing of 1 cm. The BSpline control point positions were optimized with a gradient descent method with feedback step size adjustment [25]. Quantifying the precision of our B-spline DR implementation was part of this study.

Observer and registration variability

Observer and registration variability influence the landmark identification and may increase the residual errors in the simulations. We distinguish accuracy (average of repeated measurements) and precision (variation in repeated measurements). We expect that multiple observers will, on average, agree on the location of landmarks, i.e. the expected difference over a series of observations is zero (perfect accuracy). If uncorrelated, the variance (=precision²) of a sum/difference equals the sum of the variances. With two human observers \((O_1,O_2)\) identifying ST landmarks (assuming equal variance), the observer variance \(Var(O)\) becomes:

\[
Var(O) = \frac{Var(O_1) + Var(O_2)}{2} = \frac{Var(O_2 - O_1)}{2}
\]

(8.2)
Observations exceeding ±3 SD were regarded as outliers and omitted. For the variability of the registration algorithm \( T \), we expect over a series of registrations no difference with the human observer average \(<O_1, O_2>\). The variance then becomes:

\[
Var(T) = Var(<O_1, O_2> - T) - \frac{1}{2} Var(O_2) \tag{8.3}
\]

This analysis was performed per direction and per marker. For BA, there is only a single observer (automated method) with sub-millimeter precision [22]; consequently the last term was omitted.

For comparison with other registration algorithms we calculated the fiducial registration error (FRE, sometimes referred to as target registration error TRE), the root mean square distance between corresponding landmarks after registration [28].

### Statistical analysis: residual deformations per anatomical structure

Misalignment was defined as the 3D-difference \( \vec{d} \) in landmark position in the pCT or mCT with the CBCT. Following regular analysis of rigid setup data [3], systematic \( \Sigma \) and random \( \sigma \) misalignments per direction were calculated with:

\[
\begin{align*}
\Sigma_s &= \|AVG_l(SD_p(AVG_f(\vec{d})))\| \\
\sigma_s &= \|AVG_l(RMS_p(SD_f(\vec{d})))\| 
\end{align*}
\tag{8.4}
\]

where \( f \) identifies fractions, \( p \) patients, and \( l_s \) sets of landmarks belonging to structure \( s \). \( AVG, SD \) and \( RMS \) are the average, standard deviation and root-mean-square operators. Systematic and random misalignments were calculated over the whole treatment course and expressed in vector length (veclen).

### Statistical analysis: residual deformations per patient

Improvement of geometrical accuracy was also evaluated for individual patients. To that end, the veclen of the systematic displacement of the center of mass per BA-structure was calculated and grouped by averaging. The effect of the adaptive interventions for individual patients was compared with a conventional intervention using a repeat CT, simulated by modifying the pCT with the deformation field at the intervention fraction.

### Results

Landmarks (47 BA/14 ST) were defined in the pCT and identified in the available CBCTs for all patients. 7% of BA and 20% of ST landmarks showed relevant time trends (>0.1 mm/day,
Chapter 8

$P < 0.05$). In total 808 CBCT-to-CT DRs were performed (~10 min/registration) and visually verified. Per treatment simulation, $<DVF>$s were generated and used to displace the landmarks from the pCT. For several patients we actually generated mCTs, all of CT-like quality without artifacts and normal Hounsfield Units (HU). Propagated contours showed no irregularities, see figure 8.2 for an example.

![Figure 8.2](image)

**Figure 8.2.** Left: planning CT with parotid glands (PG, yellow) and spinal cord (SC, light blue) delineated. Daily contours (orange PG, blue SC, derived from CBCT-to-CT deformable registrations) in week 4 of the treatment show substantial deformations. White contours follow from the average deformation vector field, $<DVF>$, over these fractions. Right: modified CT with the average anatomy and corresponding contours. The planning CT is modified with the average deformations in the $<DVF>$.

**Observer and registration variability**

In table 8.1 shows the observer and DR variability resulting from formula 8.2 and 8.3. Observer variability was 1.7 mm (veclen over all ST landmarks). Generally, registration accuracy quantified on BA was more precise than quantified on ST landmarks (1.2 mm vs. 2.0 mm veclen). Observer and DR variability strongly depended on landmark type (figure 8.3).

**Statistical analysis: residual deformations per anatomical structure**

The veclens of the systematic and random misalignments per structure are given in table 8.2 for the simulated intervention protocols. A single intervention after 10 fractions proved optimal for this patient group (details in appendix 8.A) and reduced systematic misalignments with 40% (BA) and 19% (ST). A further reduction was seen with weekly interventions: 61% (BA) and 33%
Table 8.1. Precision of landmark identification for 2 observers (average) and the B-spline deformable registration (DR) method (1 SD, in mm). Out of 1176 manual soft tissue landmarks identifications, 90 observations were regarded as outliers at a threshold of 3 SD.

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>CC</th>
<th>AP</th>
<th>vector length</th>
<th>FRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer</td>
<td>0.8</td>
<td>1.1</td>
<td>0.9</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>DR soft tissue</td>
<td>1.0</td>
<td>1.4</td>
<td>1.1</td>
<td>2.0</td>
<td>1.9±1.4</td>
</tr>
<tr>
<td>DR bony anatomy</td>
<td>0.5</td>
<td>0.9</td>
<td>0.6</td>
<td>1.2</td>
<td>0.8±0.9</td>
</tr>
</tbody>
</table>

**Abbreviations:** LR = left-right; CC = craniocaudal; AP = anteroposterior; veclen = vector length; FRE = fiducial registration error; DR = deformable registration.

(ST), approaching the retrospective optimum: 65% (BA) and 38% (ST). Notice that interventions to correct systematic deformations increased the random component.

Table 8.2. Residual landmark misalignment after plan adaptation with an average anatomy model. Simulated protocols were: shift corrections (for reference), single intervention after 10 fractions (single), weekly interventions (weekly) and a validation series (validation). Residual systematic and random misalignments (veclen in mm, over the full course of treatment) of separate structures are collected in bony anatomy and soft tissue with the overall average of all misalignments.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Systematic misalignments</th>
<th>Random misalignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shift</td>
<td>Single</td>
</tr>
<tr>
<td>Vertebrae</td>
<td>1.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Palate/skull</td>
<td>2.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Mandible</td>
<td>1.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Hyoid/Larynx</td>
<td>3.8</td>
<td>2.7</td>
</tr>
<tr>
<td>Uvula</td>
<td>6.2</td>
<td>5.4</td>
</tr>
<tr>
<td>Epiglottis</td>
<td>4.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Cricoid</td>
<td>3.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Carotid calcification</td>
<td>3.9</td>
<td>3.2</td>
</tr>
<tr>
<td>Skin neck</td>
<td>2.7</td>
<td>2.2</td>
</tr>
<tr>
<td>Lips</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Ears</td>
<td>2.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Muscle bifurcation</td>
<td>2.3</td>
<td>1.8</td>
</tr>
<tr>
<td>Blood vessel bifurcation</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Bony anatomy</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Soft tissue</td>
<td>3.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Overall</td>
<td>2.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>

In figure 8.3 the residual systematic misalignment (veclen) for all structures is displayed. The weekly intervention protocol reduced BA systematic misalignments to less than 1 mm except for...
the hyoid/larynx. The residual misalignments of ST structures were generally larger, especially for the uvula.
Discussion

We have developed and validated a new method to account for systematic deformations in radiotherapy. Systematic deformations were estimated with CBCT-to-CT DR and successfully used to generate an AA from the pCT of HNC patients. Geometrical validation of the method on anatomical landmarks showed substantial reductions in systematic deformations up to 61% (BA) and 33% (ST) with weekly interventions. Observer and DR variability contributed considerably to the residual deformations for ST. Improvements increased for patients with larger initial deformations. Note that the AA-model is a straightforward extension of the adaptive method proposed by Yan et al. [29] to deal with setup errors.

The application of the AA-model has several practical benefits over conventional plan adaptation approaches such as online replanning [30] or adaptation by rCT [6]. Our method does not require extra time at the CT scanner, since we use CBCT-scans that are already taken for regular patient setup correction. The DR may be initiated automatically and performed offline. Furthermore, since we do not re-optimize treatment directly on CBCT-scans, but on a modified pCT, we do not need accurate HU calibration of the CBCT scanner. Although no impact on dose calculation is expected, quality assurance of the mCT algorithm remains mandatory. In addition, parallel to modifying the pCT, we may propagate the original contours, omitting the need to re-delineate. Since the propagated contours are generated through the same $DVF$ as used for the mCT, no additional uncertainties are introduced (see figure 8.2). Finally, to restrict adaptive replanning to patients with large deformations, we can directly evaluate the $DVF$s in relation to dose and contours to objectively determine if an intervention is required and/or beneficial.

A single intervention with an AA is practically always more accurate than a regular (simulated) rCT (see supplementary figure 8.A). Interestingly, if the $DVF$ at fraction 10 were a sampling of purely random deformations, no average improvement can be expected. The 19% improvement may therefore result from other sources of error (e.g. relaxation after several treatment fractions or progressive changes).

Weekly interventions lead to smaller residual misalignments than a single intervention. This supports our hypothesis that progressive changes were present and are better accounted for with frequent interventions. Indeed, significant time trends in landmark position were frequently observed. Still, with weekly interventions, we generate a patient anatomy with measurements from the previous week. A predictive model, where measurements are extrapolated to later time points, could improve geometrical accuracy of both the single and weekly intervention protocols. Note that weekly interventions are a substantial clinical workload, requiring an efficient patient selection method and a practical clinical workflow.

A prerequisite for generating an AA for replanning is both an accurate and precise DR method. Imprecise registrations may require additional observations to obtain a reliable estimate of an AA (the standard error of the mean equals \(\frac{\text{precision}}{\sqrt{N}}\)). However, inaccuracies in the registration method could result in large residual misalignments after intervention, as is apparent.
for some structures in figure 8.3. A clear example where the B-spline registration method failed to accurately register soft tissue displacements is with the uvula: the uvula is a mobile structure that is often found against the tongue or the opposite oropharyngeal wall. B-spline deformations are inherently smooth and cannot separate these structures, resulting in modeling errors (see figure 8.7 in appendix).

Modest contrast and the presence of artifacts challenge CBCT-to-CT DR. DR of BA landmarks proved precise to within a voxel size, with a FRE of $0.8 \pm 0.9$ mm (1SD). However, for ST landmarks the FRE was larger: $1.9 \pm 1.4$ mm. This compares reasonably with Nithiananthan et al. $0.8 \pm 0.3$ mm [31] (high resolution surgical C-arm CBCT), Nithiananthan and Schafer $2.5 \pm 2.8$ mm [14], or Hou and Guerrer: $2.6 \pm 0.6$ mm [15], although observer variability contributed substantially to the FRE in our landmark set. Nevertheless, a landmark based validation method is a necessary, but not sufficient validation of redefined contours in any adaptive procedure. As an example of an additional validation we included a volume consistency analysis as 8.A supplementary material. Quality of redefined contours should be extensively inspected prior to clinical introduction, and could be part of clinical QA.

Deformable registration of the (nodal) GTV in response to radiation remains questionable, even in high quality CT, due to missing tissue and/or changes of the tissue itself. Promising techniques that incorporate missing tissue are being developed [32]. But even with accurate registrations, field size reduction is controversial in the light of microscopic disease: from CT-scans we simply cannot determine whether the tumor is (elastically) regressing (field size reduction allowed) or dissolving (not allowed). Hamming et al. [33] investigated the behavior of the tumor-normal tissue interface with implanted gold markers, demonstrating substantial uncertainties. For these two reasons, we advocate a conservative approach as it comes to redefinition of the primary CTV: rigidly transferred with, at most, modifications for bony anatomy and/or air. In fact, this is the current adaptive protocol in the ARTFORCE trial [34].

In this study we do not directly demonstrate clinical relevance as in reduced dose to OARs and/or improved target coverage. Wu and et al. [6] investigated the effect of both margin reduction and multiple plan adaptations on dose to CTV and OARs. Their findings show that dose to CTV is usually not compromised, but weekly interventions reduced the mean dose to the parotid glands by 8%, whereas no interventions lead to an increase of 10%. A margin reduction from 5 to 3 and to 0 mm reduced the mean dose by 9% and 22%. By accounting for systematic deformations with the AA-model, we expect that the dose to OARs as the parotid glands may be reduced even further, while improved setup accuracy could justify a margin reduction. Nevertheless, some margin will always be necessary to deal with residual misalignments and delineation uncertainties.
Conclusion

This study is a geometrical validation of adaptive intervention for HNC patients with an AA-model. To our knowledge this is the first study that proposes a method to correct systematic deformations with the use of DR on daily CBCT-scans. With this method we demonstrated that with limited effort geometrical accuracy can be improved substantially during treatment, possibly leading to margin reduction and/or dose escalation.
Chapter 8

Appendix 8.A  Anatomical landmark definition and identification

The proposed method was validated with the residual misalignments of clearly identifiable, atlas based, landmarks. We distinguished bony anatomy (BA) and soft tissue (ST) landmarks.

For BA landmarks we first separated individual bony structures in the pCT with a rectangular bounding box (figure 8.4). Subsequently, we automatically appointed representative points on the surface of these structures by segmenting parts of the structure. The exact subdivision for segmentation depended on the general shape of a structure and typically consisted of left, right, anterior, posterior, superior and/or inferior. Finally, these landmarks were propagated into the CBCT scans according to a local rigid registration based on chamfer matching [35]. In total, 47 landmarks were defined for 12 bony structures. To avoid over-representation of bony structures such as the vertebrae we defined 4 groups: vertebrae, skull/palate, mandible and, hyoid/larynx. The BA landmark identification method barely needed manual adjustments thereby allowing fast and easy inclusion of all available fractions for validation.

For ST validation 14 landmarks were manually defined in the pCT by an expert research technologist and subsequently indentified in one CBCT scan each week for 11 patients by two observers, resulting in 1176 landmark pairs in 84 scans. Landmark definitions were semi-atlas based, thus at similar locations between patients, and were chosen to be clearly identifiable, for example at bifurcations of muscles or blood vessels. Arterial calcifications, if present, were included as surrogate for local ST displacement. Several landmarks were positioned at a ST-air interface: on the neck-skin, ears, lips, trachea, uvula, epiglottis and dorsal cricoid cartilage. ST Landmarks describing similar anatomical structures were taken together, resulting in 9 ST groups.
Figure 8.4. The average anatomy model was validated on bony anatomy by propagating landmarks from 12 separate bony structures in the planning CT according to a local rigid registration of each structure in daily cone beam scans. Encompassing the structures is a large region of interest used for rigid registration prior to deformable registration and couch shift correction.
Appendix 8.B  Timing in a single intervention approach

In the single intervention approach, there is a trade-off between the number of fractions to accurately estimate systematic deformations (the intervention moment) and remaining fractions to benefit from replanning. To optimize the intervention moment, we have plotted in appendix figure 8.5 the overall treatment systematic and random misalignments as calculated with formula 8.4 for the BA structures as function of the intervention moment (simulated over 30 fractions to balance missing imaging fractions for some patients). ST landmarks were not analyzed since these were only available once per week for 11 patients. The optimal moment for a single intervention for systematic errors occurred in week 2 (fraction 6-10), differed only slightly between groups and was not critical to exact timing. Note that an intervention increased random errors.

Figure 8.5. Residual misalignment as function of intervention moment for groups of bony anatomy. Systematic (solid) and random (dotted) misalignments are calculated over the whole course of simulated treatment (30 fractions, 25 patients). A minimum for systematic misalignments exists at a single intervention in week 2 (fraction 6-10). Note that over the whole treatment random errors increase due to the intervention.
With the single intervention protocol there is a trade-off between the number of measurements to reliably estimate systematic deformations and the number of fractions left to benefit from an intervention. In this study we found an optimum at fraction 6-10. Bortfeld [36] and van Kranen [37] have found a theoretical optimum at fractions 5-8 (35 fractions, equal random and systematic deformations). Several factors may influence the optimum. With larger random misalignments more measurements are needed for an accurate estimate of systematic deformations. Furthermore, estimating the systematic component of deformations subject to time trends (progressive changes) will systematically shift the optimum towards a later intervention. Finally, different structures may have different ratios of systematic and random misalignments: averaging such a mixture may lead to shallow optimum.
Appendix 8.C  Volume consistency analysis

To verify the quality of redefined contours from the $<DVF>$s, we compared for the spinal cord (SC), the brainstem (BS) and the parotid glands (PG) the following volumes: 1) as delineated in the pCT, 2) as transferred by a single $DVF$ and averaged over all available fractions and 3) as transferred with the $<DVF>$ from all fractions. For SC and BS volumes during treatment should be constant: indeed, no systematic differences ($<0.2\%$) were found and patient variation was small ($<1.6\%$) and mainly due to the limited field of view of the CBCT (see table 8.3). For the PGs, a treatment average difference of -15% was present, in line with expectations that PGs shrink during treatment. Furthermore, no discrepancies were introduced by averaging over $DVF$s: all volumes from the $<DVF>$s were within 0.3% from the treatment average.

**Table 8.3.** Comparison of relative volumes a) as determined with deformable registration (averaged over treatment) vs. volume at planning, b) as determined with average anatomy model vs. average over treatment, in percent for 25 patients. SC = Spinal Cord, BS = Brainstem, PG = Parotid Gland.

(a) volume treatment average vs planning.  

<table>
<thead>
<tr>
<th></th>
<th>SC</th>
<th>BS</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVG</td>
<td>0.2</td>
<td>-0.2</td>
<td>-15.3</td>
</tr>
<tr>
<td>STD</td>
<td>1.7</td>
<td>1.6</td>
<td>8.4</td>
</tr>
</tbody>
</table>

(b) volume treatment average anatomy vs treatment average.  

<table>
<thead>
<tr>
<th></th>
<th>SC</th>
<th>BS</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVG</td>
<td>-0.3</td>
<td>-0.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>STD</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>
**Appendix 8.D  Additional figures**

![Graph showing residual deformations per patient according to different intervention protocols.](image)

**Figure 8.6.** Residual deformations (systematic misalignment of bony anatomy) per patient according to different intervention protocols. Both the regular rCT intervention and the single intervention were simulated at fraction 10. Patients with large deformations (>2 mm veclen) tend to benefit more from an intervention.
Figure 8.7. B-spline modeling errors for the uvula. Top row CBCT images show the uvula in two possible stable places: against the tongue or at the oropharyngeal wall. The bottom row shows a green purple overlay (with the uvula in the planning CT at the oropharyngeal wall) after deformable registration. The B-spline registration has clearly failed to separate the uvula from the tongue.
References


Head and neck margin reduction with adaptive radiation therapy: robustness of treatment plans against anatomy changes

S. van Kranen¹, O. Hamming-Vrieze¹, A. Wolf¹, E. Damen¹, M. van Herk², J.-J. Sonke¹

¹Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
²Christie Hospital and University of Manchester, Manchester, United Kingdom

E-mail: j.sonke@nki.nl

Published November 2016 in International Journal of Radiation Oncology, Biology and Physics, volume 96(3), pages 653-660

dx.doi.org/10.1016/j.ijrobp.2016.07.011
Abstract

Background and purpose
We set out to investigate loss of target coverage from anatomy changes in head and neck cancer patients as a function of applied safety margins and to verify a cone beam computed tomography (CBCT)–based adaptive strategy with an average patient anatomy to overcome possible target underdosage.

Methods and materials
For 19 oropharyngeal cancer patients, volumetric modulated arc therapy treatment plans (2 arcs; simultaneous integrated boost; 70 and 54.25 Gy; 35 fractions) were automatically optimized with uniform clinical target volume (CTV)–to–planning target volume margins of 5, 3, and 0 mm. We applied b-spline CBCT–to–computed tomography (CT) deformable registration to allow recalculation of the dose on modified CT scans (planning CT deformed to daily CBCT following online positioning) and dose accumulation in the planning CT scan. Patients with deviations in primary or elective CTV coverage >2 Gy were identified as candidates for adaptive replanning. For these patients, a single adaptive intervention was simulated with an average anatomy from the first 10 fractions.

Results
Margin reduction from 5 to 3 to 0 mm generally led to an organ-at-risk (OAR) mean dose ($D_{mean}$) sparing of approximately 1 Gy/mm. CTV shrinkage was mainly seen in the elective volumes (up to 10%), likely related to weight loss. Despite online repositioning, substantial systematic errors were present (>3 mm) in lymph node CTV, the parotid glands, and the larynx. Nevertheless, the average increase in OAR dose was small: maximum of 1.2 Gy (parotid glands, $D_{mean}$) for all applied margins. Loss of CTV coverage >2 Gy was found in 1, 3, and 7 of 73 CTVs, respectively. Adaptive intervention in 0-mm plans substantially improved coverage: in 5 of 7 CTVs (in 6 patients) to <2 Gy of initially planned.

Conclusions
Volumetric modulated arc therapy head and neck cancer treatment plans with 5-mm margins are robust for anatomy changes and show a modest increase in OAR dose. Margin reduction improves OAR sparing with approximately 1 Gy/mm at the expense of target coverage in a subgroup of patients. Patients at risk of CTV underdosage >2 Gy in 0-mm plans may be identified early in treatment using dose accumulation. A single intervention with an average anatomy derived from CBCT effectively mitigates discrepancies.
Introduction

Current planning and delivery techniques for head and neck cancer (HNC), such as intensity modulated radiation therapy and volumetric modulated arc therapy (VMAT), offer dose distributions conformal to the tumor with steep dose gradients, with superior sparing of organs at risk (OARs) [1]. During treatment planning, safety margins around the targets are applied to account for geometric uncertainties such as delineation errors, uncorrected setup errors, and/or anatomy changes during treatment [2]. With large margins, robust plans are created that ensure correct target dosage in case of large errors; however, substantial amounts of normal tissue may be irradiated. A reduction in geometric uncertainties would allow smaller safety margins and thereby increase the therapeutic window.

Image guided radiation therapy reduces geometric uncertainties. In-room imaging (e.g. cone beam computed tomography [CBCT]) allows one to visualize (a surrogate of) the target volume in relation to OARs and derive a couch correction for alignment of the target volume or volumes with the treatment beams. In offline protocols, the systematic component of target positioning errors is estimated and corrected. Online protocols allow for near perfect correction of errors [3].

To further reduce uncertainties arising from nonrigid setup errors and anatomy changes (deformations), adaptive radiation therapy (ART) can be considered [4]. Commonly, ART for HNC is based on a second computed tomography (CT) scan and used to address ad hoc observed treatment response (tumor regression, weight loss) [5, 6]. Recently, however, ART based on an average anatomy model derived from CBCT has been proposed to reduce the impact of systematic deformations [7].

To improve the therapeutic ratio, we propose to use small margins, evaluate delivered dose with CBCT, and use adaptive strategies in case of target underdosage when needed. To test the feasibility of such an approach, we investigated (1) changes in dose to clinical target volumes (CTVs) and OARs from deformations in HNC patients as a function of applied safety margins; and (2) a CBCT-based adaptive strategy with an average anatomy model to overcome loss of coverage in selected patients.

Materials & Methods

Patient group

We retrospectively selected 19 patients with oropharyngeal cancers that were treated with curative intent. Patient data were accessed according to institutional guidelines. The T category distribution was T1 in 2 patients, T2 in 7, T3 in 8, and T4 in 2. Three patients did not have nodal involvement. The planning computed tomography (pCT) scan (Somatom Sensation Open; Siemens, Erlangen, Germany) had a voxel size of $1 \times 1 \times 1 \text{mm}^3$. Daily CBCT scans (Elekta Synergy; Elekta Oncology Systems, Crawley, UK) were acquired ($1 \text{mm}^3$ voxel size). Patients were
immobilized with a 5-point thermoplastic mask and positioned with a knee support and head rest (Civco Medical Solutions, Kalona, IA).

Simulation of treatment and dose accumulation

For all patients, 3 treatment plans with 5-, 3-, and 0-mm CTV–to–planning target volume (PTV) margins (dual-arc VMAT, simultaneous integrated boost [SIB]) were optimized in Pinnacle v9.6 (Philips Medical Systems, Eindhoven, The Netherlands). A boost PTV was defined from the CTV of the primary tumor (CTV_{boost}) and positive lymph nodes (CTVln_{boost}) and prescribed to 70 Gy. An elective PTV was defined from the CTV of suspected lymph nodes (CTVln) and a ring around the CTV_{boost} (CTV) and prescribed to 54.25 Gy (see supplement 9.A). Definitions of CTVs and OARs were taken from clinical treatment plans. To achieve reproducible plans free from planner bias and variation, a previously developed automatic optimization strategy was used [8]. Minimal requirements for the PTVs after optimization were: dose to 99% of the volume ($D_{99\%}$) > 95% and $D_{1\%}$ < 107% of prescribed dose (supplement 9.B). Dose calculation in the first 4 mm from the external contour was considered unreliable and ignored during plan optimization and analysis.

Online patient positioning resulted from alignment of bony anatomy close to the tumor in the CBCT scan. The CBCT scan was shifted accordingly to obtain representative online scans.

Deformable image registration (DIR) was used to map anatomy in the CBCT scan of each fraction to the pCT scan. We used a b-spline deformation model (supplement 9.C), that was previously validated and showed typical 3-dimensional fiducial registration errors (vector length) of 0.8 mm for bony anatomy, 1.9 mm for normal tissue (including anatomically defined CTV borders), and 3.3 mm for the tumor [7, 9]. The registered region was limited to the field of view of the CBCT scan.

To capture the treatment intent of irradiating the original tumor volume in the presence of tumor regression, we penalized deformations of the CTV_{boost} and CTVln_{boost} that were not driven by contrast differences by limiting the curvature of the deformation vector fields (DVFs) in these regions [10]. We compared resulting target volumes with and without this regularization.

The DVFs were used to map the pCT scan to the CBCT scan [11, 12] to obtain the correct Hounsfield units in the anatomy of the day for dose recalculation. The reconstructed CBCT field of view was always smaller than the pCT scan, approximately 25 cm in the craniocaudal direction, but always encompassed all PTVs following isocenter placement in the center of mass of these PTVs. Treatment beams seldom passed through anatomy outside the CBCT scan. Nevertheless, a continuous anatomy in the unregistered parts of the CT scan was obtained by smoothly reducing deformations outside the CBCT scan to zero over a length of 5 cm. Subsequently, dose was mapped back to the pCT scan for accumulation. Dose recalculation and accumulation were performed for all available CBCT scans. Because of maintenance, 14 treatment fractions in 5 patients were delivered without CBCT. In these cases the accumulated dose was normalized to 35 fractions.
Volume changes and systematic deformations

The DVFs were used to quantify volume changes and calculate systematic deformations: for each structure, the average DVF over the whole course of treatment was used to quantify the fraction of volume with residual systematic deformations $>3$ mm.

Dose to targets volumes and OARs

In total, 73 CTVs were available for analysis because in 3 patients no CTV\textsubscript{lnboost} was present. First, we quantified typical sparing of OARs that is achievable with margin reduction. Next, we analyzed accumulated dose to CTVs and OARs. Supplement 9.B describes all dose-volume histogram parameters evaluated.

Adaptive replanning

Retrospectively, patients with a $D_{99\%}$ of the CTV in the accumulated dose $>2$ Gy lower than the $D_{99\%}$ of the CTV in the planned dose were considered candidates for ART. $D_{99\%}$ was used as a practical surrogate for the minimum dose as described in conventional margin recipes [2]. We chose 2 Gy to represent a substantial discrepancy with the planned dose—equivalent to a single fraction to the primary tumor (3% of the total dose) and large enough to represent a serious deviation with respect to dose calculation and mapping uncertainties. ART to avoid extra OAR dose was not considered. In practice, patients eligible for ART should be identified early in treatment. We tested various thresholds and intervention moments ($N$) to select the candidate patients.

Adaptive replanning was performed on a modified pCT scan, representative of the average anatomy of the first $N$ fractions. Therefore, we averaged the first $N$ DVFs (DVF\textsubscript{mean}, each DVF describes how tissue elements in the pCT scan are displaced in the CBCT scan). With progressive changes (weight loss, gland or tumor shrinkage), the most recent CBCT scans are more representative than earlier scans. This was reflected in a weighting factor proportional to the fraction number [7]. The DVF\textsubscript{mean} was used to create an updated CT scan and to transfer the original contours. Plans were reoptimized in the updated pCT scan, and treatment was again simulated as described previously starting at fraction $N + 1$. Recalculated dose was mapped to the initial pCT scan to be added to the dose of the first $N$ fractions.

Results

For the 19 selected patients, plans with 5-, 3-, and 0-mm margins were generated that fulfilled minimal requirements (supplement 9.B). A total of 651 CBCT scans underwent deformable registration to the corresponding pCT scans, and DVFs were visually inspected.
Volume changes and systematic deformations

The primary target (CTV_{boost}) and its elective region (CTV) showed small volume changes (Fig. 9.1), whereas the lymph nodes (CTV_{ln}^{boost} and CTV_{ln}) clearly showed volume reduction up to 10% in week 5. The influence of the additional regularization of the tumor and involved lymph nodes on assessed volume changes was limited: at the end of treatment, CTV regression would sometimes be underestimated to 1 to 2 percentage points (data not shown). As expected, trends were present in parotid gland (PG) volume, with a 4% reduction per week. Remarkably, a small increase in constrictor muscle volume during treatment was found (4%).

Large systematic deformations were quantified by the relative volume with residual systematic errors $>3\, \text{mm}$ ($V_{3\, \text{mm}}$; supplement 9.D). The largest deformations in target volumes occurred with the CTV_{ln} (median $V_{3\, \text{mm}}$, 19.7%; range, 1.4%-63.3%), often because of weight loss and subsequent volume reduction (described earlier) or shoulder movement (visual inspection).

![Relative volumes changes during treatment](image)

**Figure 9.1.** Trends in relative volume during treatment (weekly averages, 19 patients) derived with deformable registration. Error bars, shown for upper and lower curve, indicate variation among patients (1 SD). Marker size is proportional to the standard deviation. Because clinical target volume (CTV) is an expansion of CTV_{boost} (CTV of primary tumor), trend lines nearly coincide. All differences between week 1 and week 7 were statistically different ($P < 0.05$, Wilcoxon signed-rank-test for paired samples). **Abbreviations:** CTV_{ln} = elective CTV of suspected lymph nodes; CTV_{ln}^{boost} = CTV of involved lymph nodes.
H&N margin reduction with ART

contrast, CTV_{boost} and CTV showed relatively small systematic deformations: 0.4% and 0.7%, respectively. Apart from the PGs, the larynx showed substantial systematic deformations (17.7%), in line with reported systematic shifts and rotations of the whole organ [13, 14].

**Dose to targets volumes and OARs**

Evaluation of the planned dose to OARs showed better sparing with smaller margins: approximately 1 Gy mean dose per millimeter margin reduction (Fig. 9.2). The largest differences between accumulated and planned dose (Fig. 9.3) were present in the ipsilateral PGs: on average, 1.2 Gy higher than planned. In 4 patients the ipsilateral PG D_{mean} increased >2 Gy (5-mm plans); in 2 of these patients, margin reduction made no difference. For 13 patients, the accumulated OAR dose was always within 2 Gy of planned, irrespective of margins.

In supplement 9.E, as an example, CTV_{boost} coverage is presented in detail. Table 9.1 reports target coverage for all 73 target areas. Target underdosage (D_{99\%} < 95\% isodose) occurred 1 time, 2 times, and 20 times (27%) in 5-, 3-, and 0-mm plans, respectively, demonstrating that margin reduction reduces coverage mainly in 0-mm plans. Discrepancies >2 Gy occurred in 1, 3, and 7 targets, respectively, mainly within the elective CTVs. Discrepancies with both D_{99\%} < 95\% isodose and \Delta D_{99\%}>2 Gy occurred 1, 2, and 3 times, respectively. Furthermore, most underdosage occurred at the elective lymph node CTVs (CTVln). We regarded the 7 cases with underdosage >2 Gy in 0-mm plans as opportunities to improve treatment with ART.
Figure 9.2. Effect of margin reduction on planned and accumulated organs at risk (OAR) dose (averaged for 19 patients). Maximum dose ($D_{\text{max}}$) to brainstem and spinal cord was never critical. Error-bars, shown for upper and lower curve, indicate variation among patients (1 SD). Marker size is proportional to the standard deviation. Abbreviations: $D_{\text{mean}}$ = mean dose.
Figure 9.3. Boxplot of accumulated (acc) versus planned (plan) organ-at-risk (OAR) dose (19 patients [pts]).

*) $p < 0.05$ with the Wilcoxon signed-rank-test for paired samples.

Abbreviations: constr = constrictor; contra = contralateral; $D_{\text{max}}$ = maximum dose; $D_{\text{mean}}$ = mean dose; ipsi = ipsilateral; PG = parotid gland.
Table 9.1. CTV coverage in planned and accumulated dose ($D_{99\%}$, 19 patients)

<table>
<thead>
<tr>
<th>plan</th>
<th>target</th>
<th>Planned $D_{99%}$, Gy</th>
<th>Accumulated $D_{99%}$, Gy</th>
<th>Accumulated-Planned</th>
<th>Mean ratio of accumulated to planned</th>
<th>$\Delta D_{99%} &lt; -2, Gy$</th>
<th>$D_{99%} &lt; 95% D_0$</th>
<th>$D_{99%} &lt; 95% D_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 mm</td>
<td>CTV$_{boost}$</td>
<td>69.0 0.5</td>
<td>69.6 0.6</td>
<td>0.7* 0.4</td>
<td>1.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln$_{boost}$</td>
<td>68.7 0.7</td>
<td>69.0 1.1</td>
<td>0.3 0.6</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTV</td>
<td>66.7 1.9</td>
<td>66.7 2.1</td>
<td>0.0 0.9</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln</td>
<td>53.2 0.7</td>
<td>53.2 1.4</td>
<td>0.0 1.1</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3 mm</td>
<td>CTV$_{boost}$</td>
<td>68.9 0.4</td>
<td>69.4 0.5</td>
<td>0.5* 0.3</td>
<td>1.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln$_{boost}$</td>
<td>68.6 0.6</td>
<td>68.5 1.3</td>
<td>-0.1 1.1</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CTV</td>
<td>64.0 2.0</td>
<td>63.8 2.1</td>
<td>-0.2 1.2</td>
<td>1.00</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln</td>
<td>53.0 0.6</td>
<td>52.4 2.3</td>
<td>-0.6 2.0</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0 mm</td>
<td>CTV$_{boost}$</td>
<td>67.4 0.3</td>
<td>67.0 0.7</td>
<td>-0.4 0.7</td>
<td>0.99</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln$_{boost}$</td>
<td>67.0 0.3</td>
<td>66.2 2.1</td>
<td>-0.8* 2.0</td>
<td>0.99</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CTV</td>
<td>58.7 1.8</td>
<td>58.2 2.5</td>
<td>-0.6 1.9</td>
<td>0.99</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTVln</td>
<td>52.4 0.6</td>
<td>50.7 3.3</td>
<td>-1.7* 3.1</td>
<td>0.97</td>
<td>3</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

Abbreviations: CTV = clinical target volume; CTV$_{boost}$ = CTV of primary tumor; CTVln = elective CTV of suspected lymph nodes; CTVln$_{boost}$ = CTV of involved lymph nodes; $D_0$ = prescription dose; $D_{99\%}$ = dose to 99% of volume.

The prescription dose ($D_0$) was 70 Gy for the planning target volume of the primary tumor and involved lymph-nodes, and 54.25 Gy for elective regions. CTV coverage was considered at risk when $D_{99\%} < 95\%$ of $D_0$. CTVs with $\Delta D_{99\%} < -2\, Gy$ were considered of interest for adaptive radiation therapy. *$P < 0.05$ with the Wilcoxon signed-rank-test for paired samples.
Adaptive replanning

The 7 CTVs with a dose discrepancy >2 Gy in 0-mm plans were found in 6 patients. Supplement 9.F shows the true- and false-positive rate to select these patients early in treatment as a function of threshold. On the basis of these results, fraction 10 (2 weeks), with a corresponding threshold at −0.6 Gy, was chosen as a practical time point for intervention. Updated pCT scans were generated using the first 10 CBCT scans. Plans were automatically reoptimized, and simulated treatment continued from fraction 11.

Figure 9.4 shows the effect of ART on minimum dose to the CTV. In all patients except one, ART improved $D_{99\%}$ with at least 3 Gy. In 2 patients the final accumulated dose after ART still differed by >2 Gy from the initial plan, indicating possible progressive changes requiring additional adaptive intervention.

Figure 9.4. Improvements in minimum dose (expressed as dose to 99% of volume, $D_{99\%}$) to clinical target volumes (CTVs) from a single adaptive intervention (fraction 10) where CTV coverage was at risk (0-mm plans). Adaptive interventions restored discrepancies to <2 Gy in 5 of 7 CTVs (6 patients). Abbreviations: ART = adaptive radiation therapy; CTVin = elective CTV of suspected lymph nodes; CTVin_boost = elective CTV of involved lymph nodes; P = patient.

Discussion

In this study the potential of ART for anatomy changes during radiation therapy (RT) of HNC was assessed based on accumulated dose in daily CBCT scans. For 73 CTVs in 19 patients, VMAT
treatment plans with 5-mm safety margins were adequate in all but 1 CTV. When margins were reduced, loss of target coverage ($D_{99\%} < 95\%$ prescribed dose) increased to 27% of all CTVs (at 0-mm margins) but discrepancies with planned dose were small. Margin reduction led to sparing of approximately 1 Gy/mm in investigated OARs (PGs, constrictor muscle, oral cavity, larynx). For two-thirds of patients, accumulated OAR dose differed by <2 Gy from planned for all margins evaluated. The simulated CBCT-based adaptive workflow in plans with 0-mm margins, with a single intervention at fraction 10, was found to effectively select candidate ART patients and showed improvements of at least 3 Gy in 6 of 7 target volumes.

With observed residual systematic deformations in excess of 3 mm, mainly for elective lymph node CTVs (supplement 9.D), safety margins of 5 mm could be considered too small according to classical margin recipes [2]. Yet treatment plans with 5-mm or even smaller margins were surprisingly robust for these errors. In plans with 5-, 3-, and 0-mm margins, only 1, 2, and 3 target volumes, respectively, were present with an accumulated $D_{99\%} < 95\%$ isodose and a difference from planned dose >2 Gy. Several factors may explain this robustness against geometric uncertainties: (1) With multiple CTVs, or with concave-shaped target volumes, deformations could shift a CTV surface outside the PTV but inside another/opposite PTV. Currently, margin recipes do not take such effects into account; (2) For the high-dose CTVs, the surrounding elective dose acts as background dose: in case of residual misalignments, minimum dose will reduce to elective levels. McKenzie et al. [15] showed that because of background dose, the margin for random errors, typically approximately 20% of the total margin, can be reduced; (3) In case of tumor regression and/or weight loss, CTV borders may move into the high-dose region. Therefore, during treatment, next to a somewhat higher delivered dose than planned (figure 9.6), the distance between CTVs and the 95% isodose surface increases. This effectively allows for larger errors before coverage is lost; (4) The planned dose distribution does not perfectly conform with the PTV. Consequently, the treated volume (volume within the 95% isodose surface) is typically larger than the PTV, allowing for larger setup errors. Visual inspection showed that distances of 2 to 3 mm were frequently present between the PTVs and corresponding treated volumes. For 6 patients, clinical SIB plans (5-mm margins) were available that had similar discrepancies, showing that loss of conformity was not introduced by the automatic planning algorithm. This additional dosimetric margin around a PTV was described by Gordon et al. [16] for RT of prostate cancer.

In this study, the $D_{99\%}$ of the PTV and CTV was used for planning and evaluation, respectively. However, for superficial tumors, CTV-to-PTV expansions might extend to the buildup region or even to air and challenge treatment plan optimization and the shift invariant assumption underlying the PTV concept [17]. Essentially, the minimum dose and $D_{99\%}$ are surrogates for the tumor control probability and were adopted as such in this article. Ideally, more accurate tumor control probability models would be used to select patients for replanning that might be less sensitive to setup errors and anatomic changes.

Routine adaptive interventions to avoid excess dose to OARs are questionable. Some studies 18, 19, 20 and 21 have reported considerable PG mean dose discrepancies ($\Delta D_{mean} =$
H&N margin reduction with ART

+3.7, +3.6, +2.8, and +2.7 Gy) and advise adaptive interventions. Ahn et al. [18] correlated anatomy changes with dosimetric effects, searching for surrogate intervention criteria. They reported a wide variety of changes with different dosimetric effects. Routine adaptive replanning was beneficial for two-thirds of all patients. Hunter et al. [19] found a median increase of 2.2 Gy in PG dose but could not demonstrate a significant difference in saliva flow and questioned the effect of interventions. In our patient group, extra OAR dose was generally small (maximum of 1.2 Gy, ipsilateral PG) and discrepancies >2 Gy were limited to one-third of patients. To limit clinical workload, decisions on adaptive interventions should be individualized and based on assessment of accumulated dose.

Close monitoring of delivered dose and accurate prediction of the expected final dose to CTVs and OARs are required to identify candidate patients for adaptive intervention. Nie et al. [20] monitored OARs and targets during treatment with DIR and applied principal component analysis and least squares regression to predict future organ geometries for dose estimation. After 3 weeks, the final dose distribution could be predicted to within 2.5% (SD). Focusing on accumulated dose for the PGs, Hunter et al. [19] found that the dose difference from the planned dose on the first treatment day was highly correlated (r=−.92) with the final difference. We were able to select candidate patients with a simple intervention threshold at fraction 10 without false positives. This suggests that dose differences were present from the start of treatment and therefore were more likely related to systematic local misalignments (rotational errors, neck flex, shoulder misplacement) than to progressive changes (e.g., regression or weight loss). Half of all target underdosage occurred at the elective regions, which deformed (among others) from shoulder movement (visually inspected). Our simulated online positioning strategy focuses on alignment of the primary tumor; however, alternative positioning strategies can trade tumor alignment for improved positioning of elective regions [14]. Proper intervention-selection protocols, balancing accurate predictions of delivered dose versus early detection of discrepancies, still have to be developed.

DIR quality is important for correct mapping of the CBCT anatomy in the pCT scan, as well as accurate dose mapping for accumulation and adaptive interventions. To deal with the reduced tissue-to-tissue correspondence in tumor tissue because of regression and lack of contrast with normal tissue [9], we applied a regularization term for the tumor regions. To transfer the CBCT anatomy into the pCT scan for dose recalculation, accurate tissue-to-tissue mapping is less important: the pCT scan should only be modified to reflect the correct Hounsfield units (tissue type). For dose mapping, accurate tissue-to-tissue mapping is most important at dose gradients: in a uniform dose distribution, the mapped dose is insensitive to registration errors [21]. For adaptive purposes, the regularization was also in line with our clinical ART protocol that allows CTV adjustments only for obvious changes [22]. We therefore expect that the larger target registration error in tumor tissue will have had a limited impact on the results of this study. Nonetheless, other registration techniques should be investigated to improve registration of the tumor [23].
A comparable margin reduction study by Wu et al. [24] in intensity modulated radiation therapy SIB plans (5-, 3-, and 0-mm margins) showed increased sparing of OARs in 11 patients. PG dose, evaluated in 6 weekly repeat helical CT scans, significantly increased (+10%), but no CTV underdosage was seen. Routine weekly interventions could prevent extra PG dose. Our study used VMAT as the delivery technique and daily CBCT as the basis for analysis, patient selection, and intervention. Remarkably, both studies showed substantial robustness of treatment plans for deformations. In contrast, we did not find the same extra dose to PGs (+3%).

Conclusions

In this study, the potential of ART for anatomy changes during RT of HNC was assessed based on accumulated dose in daily CBCT scans. For 73 target volumes in 19 patients, VMAT treatment plans with 5-mm safety margins were adequate in all but 1 target volume. Modest extra OAR dose was present. Margin reduction led to an improvement in OAR dose of approximately 1 Gy/mm. With 3- and 0-mm margins, dose to 2 targets and 20 targets, respectively, was compromised, although large discrepancies (>2 Gy) were limited to 6 patients. ART protocols are therefore advised. Care should be taken to include all sources of uncertainty in margin design. A CBCT-based ART workflow in plans with 0-mm margins was found feasible to select candidate patients for ART and substantially restore target coverage.

Summary

The impact of anatomy changes on delivered dose was investigated in head and neck cancer patients as a function of the applied safety margins. Margin reduction reduced organ-at-risk exposure on average by 1 Gy/mm at the expense of target underdosage >2 Gy in 32% of patients. Patients at risk of underdosage could be identified using dose accumulation, and underdosage could be mitigated in 83% of patients by adaptive intervention using an average anatomy model after 10 fractions.
Appendix 9.A  Target definition, dose prescription and margins

The primary tumor and involved lymph nodes are delineated as GTV and N+, lymph node regions suspect to contain tumor cells as N0. The GTV and N+ are expanded with 10 mm to CTV_{boost} and CTV_{ln_{boost}}, which are regarded the primary CTVs. Ln stands for lymph node. As elective CTVs we directly use the delineated lymph node regions (N0) as CTV_{ln} and a region obtained by expanding the CTV_{boost} with 5 mm: CTV. To account for errors in target definition, setup errors, deformations and other sources of geometric uncertainty, the CTVs are expanded to PTVs with a safety margin [25, 26]. The PTV_{boost} is prescribed 70 Gy and the elective PTV 54.25 Gy, over 35 fractions. See figure 9.5 for a graphical overview.

![Figure 9.5. Target definition and prescription dose in head & neck radiotherapy.](image)
Appendix 9.B  Planning objectives and evaluation parameters

Table 9.2 contains the objectives for automatic SIB VMAT plan optimization. Objectives for PTVs, spinal cord and brainstem had to be fulfilled for an acceptable plan. For all OAR-objectives, dose was reduced as much as reasonably achievable, without compromising on target coverage. Objective weights were set according to the dose distribution after the first iteration (of 4) on PTV coverage only, parotid glands were thereby favored over other OARs. For target dose, the CTV dose was evaluated while for OAR dose the same objective parameters were evaluated.

**Table 9.2.** Constraints & objectives for plan optimization with evaluation parameters for accumulated dose.

<table>
<thead>
<tr>
<th>structure</th>
<th>parameter</th>
<th>Dose(Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTVboost</td>
<td>$D_{99%} &gt; 66.5$</td>
<td>i.e.95%*70Gy</td>
</tr>
<tr>
<td></td>
<td>$D_{1%} &lt; 74.9$</td>
<td>i.e.107%*70Gy</td>
</tr>
<tr>
<td>PTV</td>
<td>$D_{99%} &gt; 51.54$</td>
<td>i.e.95%*54.25Gy</td>
</tr>
<tr>
<td></td>
<td>$D_{1%} &lt; 58.05$</td>
<td>i.e.107%*54.25Gy</td>
</tr>
<tr>
<td>spinal cord</td>
<td>$NTD_{max} &lt; 50.0$</td>
<td>($\alpha/\beta = 2Gy$)</td>
</tr>
<tr>
<td>brainstem</td>
<td>$NTD_{max} &lt; 54.0$</td>
<td>($\alpha/\beta = 2Gy$)</td>
</tr>
<tr>
<td>parotid gland</td>
<td>$D_{mean} &lt; 26$</td>
<td></td>
</tr>
<tr>
<td>constrictor muscle</td>
<td>$D_{mean} &lt; 45$</td>
<td></td>
</tr>
<tr>
<td>oral cavity</td>
<td>$D_{mean} &lt; 26$</td>
<td></td>
</tr>
<tr>
<td>base of tongue</td>
<td>$D_{mean} &lt; 45$</td>
<td></td>
</tr>
<tr>
<td>larynx</td>
<td>$D_{mean} &lt; 45$</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTVboost</td>
<td>$D_{99%}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_{1%}$</td>
<td></td>
</tr>
<tr>
<td>CTVlnboost</td>
<td>$D_{99%}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_{1%}$</td>
<td></td>
</tr>
<tr>
<td>CTV</td>
<td>$D_{99%}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_{1%}$</td>
<td></td>
</tr>
<tr>
<td>CTVln</td>
<td>$D_{99%}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_{1%}$</td>
<td></td>
</tr>
<tr>
<td>spinal cord</td>
<td>$NTD_{max}$</td>
<td>($\alpha/\beta = 2Gy$)</td>
</tr>
<tr>
<td>brainstem</td>
<td>$NTD_{max}$</td>
<td>($\alpha/\beta = 2Gy$)</td>
</tr>
<tr>
<td>parotid gland (ipsi-lateral)</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
<tr>
<td>parotid gland (contra-lateral)</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
<tr>
<td>constrictor muscle</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
<tr>
<td>oral cavity</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
<tr>
<td>base of tongue</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
<tr>
<td>larynx</td>
<td>$D_{mean}$</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix 9.C  b-spline registration parameters

<table>
<thead>
<tr>
<th>Deformation model</th>
<th>b-spline 3rd order</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>b-spline 3rd order</td>
</tr>
<tr>
<td>final b-spline control point spacing</td>
<td>1 cm</td>
</tr>
<tr>
<td>Image intensity mapping</td>
<td>3rd order</td>
</tr>
<tr>
<td>b-spline interpolation</td>
<td>3rd order</td>
</tr>
<tr>
<td>Similarity</td>
<td>correlation ratio [27]</td>
</tr>
<tr>
<td>metric</td>
<td></td>
</tr>
<tr>
<td># bins</td>
<td>100</td>
</tr>
<tr>
<td>Regularization (weights) [10]</td>
<td></td>
</tr>
<tr>
<td>global smoothness</td>
<td>0.02</td>
</tr>
<tr>
<td>local rigidity segmented bony anatomy</td>
<td>0.5</td>
</tr>
<tr>
<td>local smoothness primary tumor</td>
<td>0.01</td>
</tr>
<tr>
<td>Optimizer</td>
<td>gradient descent</td>
</tr>
<tr>
<td>optimization method</td>
<td>with feedback step size adjustment</td>
</tr>
<tr>
<td>Multiresolution</td>
<td></td>
</tr>
<tr>
<td># levels</td>
<td>3</td>
</tr>
<tr>
<td>image pyramid</td>
<td>Gaussian</td>
</tr>
<tr>
<td>b-spline control point spacing (cm)</td>
<td>2/2/1</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
</tr>
<tr>
<td>ignore air in pCT</td>
<td></td>
</tr>
<tr>
<td>histogram remapping emphasizing soft tissue</td>
<td>153</td>
</tr>
</tbody>
</table>
Appendix 9.D  Systematic deformations

Systematic deformations in the selected volumes are derived from the average deformations over all fractions. Table 9.4 shows the portion of volume with local systematic deformations larger than 3 mm ($V_{3\,mm}$) for clinical target volumes and organs at risk for 19 patients (expressed in median, range and amount of patients with $V_{3\,mm} > 20\%$).

**Table 9.4.** Systematic volume changes over a course of radiotherapy, averaged for 19 patients.

<table>
<thead>
<tr>
<th>structure</th>
<th>median $V_{3,mm}$(%)</th>
<th>[min, max]</th>
<th>#pts $V_{3,mm} &gt; 20%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTV_boost</td>
<td>0.4</td>
<td>[0.0,61.3]</td>
<td>2</td>
</tr>
<tr>
<td>CTVln_boost</td>
<td>3.7</td>
<td>[0.0,65.8]</td>
<td>5</td>
</tr>
<tr>
<td>CTVln</td>
<td>19.7</td>
<td>[1.4,63.3]</td>
<td>10</td>
</tr>
<tr>
<td>CTV</td>
<td>0.7</td>
<td>[0.0,55.5]</td>
<td>2</td>
</tr>
<tr>
<td>spinal cord</td>
<td>0.9</td>
<td>[0.0,67.9]</td>
<td>4</td>
</tr>
<tr>
<td>brainstem</td>
<td>0.0</td>
<td>[0.0,82.5]</td>
<td>3</td>
</tr>
<tr>
<td>parotid gland (high dose)</td>
<td>13.1</td>
<td>[0.0,76.2]</td>
<td>8</td>
</tr>
<tr>
<td>parotid gland (low dose)</td>
<td>14.4</td>
<td>[0.0,89.5]</td>
<td>9</td>
</tr>
<tr>
<td>oral cavity</td>
<td>2.0</td>
<td>[0.0,77.6]</td>
<td>4</td>
</tr>
<tr>
<td>base of tongue</td>
<td>9.7</td>
<td>[0.0,65.1]</td>
<td>6</td>
</tr>
<tr>
<td>constructor muscle</td>
<td>5.3</td>
<td>[0.0,94.4]</td>
<td>6</td>
</tr>
<tr>
<td>larynx</td>
<td>17.7</td>
<td>[0.0,85.2]</td>
<td>9</td>
</tr>
</tbody>
</table>

Relative volume with systematic deformations $>3$ mm ($V_{3\,mm}$, veclen).
Appendix 9.E  CTV\textsubscript{boost} coverage during treatment in detail

Figure 9.6 shows the CTV\textsubscript{boost} coverage during treatment. With 5 and 3 mm plans, minimum CTV\textsubscript{boost} dose was well above the planned PTV minimum dose criterion ($D_{99\%} > 95\%$ isodose, i.e. 66.5 Gy). However, since CTV and PTV surface coincide at 0 mm, and the minimum dose is typically found at the surface of the PTV, the minimum dose to the CTV is, on average, 1.5 Gy lower than with a margin. Accumulation of dose (over all available fractions, without interventions) showed in general a small increase in minimum dose at 5 and 3 mm margins (0.7 ± 0.4 Gy and 0.5 ± 0.3 Gy), whereas at 0 mm, the minimum dose decreased −0.4 ± 0.7 Gy. Nonetheless, this variation is comparable to the variation in accumulated CTV\textsubscript{boost} $D_{99\%}$ between patients (0.6 / 0.5 / 0.7 Gy, 1 SD). For 5 patients, at 0 mm plans, accumulated $D_{99\%}$ is below 66.5 Gy, but differences between planned and accumulated dose >2 Gy in the CTV\textsubscript{boost} did not occur.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{ctv_boost_d99}
\caption{Example dose analysis: difference between planned and accumulated minimum dose ($D_{99\%}$) to CTV\textsubscript{boost} (19 patients, with 5 / 3 / 0 mm margins).}
\end{figure}

*) p-values < 0.05 with Wilcoxon signed-rank-test for paired samples.

In retrospect 7 target areas in 6 patients showed a difference in an accumulated dose to the CTV larger than 2 Gy with planned ($D_{99\%}$ in 0 mm plans). These patients, were considered candidates for adaptive intervention. In practice these patients should be identified based on a selection threshold during a regular course of treatment. To study the feasibility of patient selection at an earlier moment, say fraction $f$, we performed a Receiver-Operator-Characteristic (ROC) analysis: patients with a final dose difference of 2 Gy in one of the CTVs were defined as the gold standard. Next, the difference between accumulated and planned dose at fraction $f$ was tested against a range of thresholds. For each threshold we scored the correctly and incorrectly selected patients (true and false positives), from which we derived the true positive rate (sensitivity) and false positive rate (1-specificity). From the corresponding ROC curve we calculated the area under the curve (AUC) as a direct measure of how efficient candidates for ART could be selected. This was repeated for various fractions $f$. ROC analysis on intervention thresholds showed that the AUC (see figure 9.7) ranged from 0.73 to 1.00. Errorbars were calculated according to Hanley and McNeil[28]. After 8 fractions all candidates for adaptive intervention were detectable without false positives. However, for simulation of treatment with an adaptive intervention, we used fraction 10, representing end of week 2.
Figure 9.7. ROC analysis of intervention thresholds to detect CTV underdosage $>2\text{ Gy (}D_{99\%}\text{)}. The intervention threshold curve and area represent threshold values on accumulated dose where all candidates for adaptive intervention were selected (no false negatives) with the least false positives. From fraction 8, areas under the curve of 1 were found (no false positives).
References


General discussion and conclusion
This thesis investigates the use of in-room CBCT imaging to quantify and reduce geometrical uncertainties during treatment delivery for HNC. The specific objectives were the following:

- To quantify geometrical errors during RT of HNC;
- To derive an optimal patient setup in deforming anatomy with IGRT;
- To develop procedures to adapt the treatment plan to residual errors uncorrected by patient setup using CBCT.

We met these objectives as follows: 1) Use bony anatomy registration and gold marker tracking to observe geometric uncertainties arising from deforming and changing anatomy as well as tumor regression patterns over the course of treatment. 2) To optimize patient setup accuracy during treatment, we developed and clinically implemented a HNC specific IGRT protocol. 3) For residual errors that cannot be corrected with patient positioning we developed innovative adaptive approaches. In these approaches we use a deformable image registration algorithm that was carefully validated. The adaptive approaches allow highly automated and efficient workflows that where shown to reduce geometrical uncertainties.

In the following sections we will discuss the work presented in this thesis in more detail, including challenges for clinical implementation of ART. To finalize we discuss how the findings in this thesis can be applied in next generation radiotherapy, for instance with linear accelerators with integrated MR imaging or with proton therapy.

### 10.1 Geometrical errors during RT

To quantify geometrical errors during RT, we first measured residual setup uncertainties based on analysis of bony anatomy in sub-volumes of the CBCT (chapter 2). The study showed substantial residual errors even after online setup correction, resulting from non-rigid motion of bony anatomy (posture changes). Second, we measured variations of the tumor surface with respect to bony anatomy. To this end, we inserted gold markers in the GTV-to-CTV margin, representative for the position of the CTV (chapter 3). The required margin to cover the systematic part of such tumor-surface variations was estimated at 2.3 mm. In addition, trends in marker motion were present, and markers were found to move with respect to each other, indicating deformations and tumor regression for which ART could be considered. Since CBCT is not an ideal modality to visualize tumor regression, we also applied magnetic resonance imaging (MRI) to investigate GTV regression (chapter 4). Results were compared against CTV regression based on gold markers in CBCT. Discrepancies between CTV regression and GTV regression were found, and it was concluded that CTV-modifications based on visualized GTV regression in MRI potentially result in lower than prescribed dose to regions still containing microscopic disease and may therefore be incompatible with curative intent.
Summary and general discussion

Several papers have been published on residual local setup errors of bony anatomy after setup correction [1–5]. These studies show residual errors comparable to our study, with some differences based on alignment strategies and/or used fixation devices. These studies also illustrate that deformations occur that cannot be corrected with simple couch shifts and that, as a result, residual uncertainties have to be expected. Use of gold of markers in RT is not new, but so far has mainly been restricted to the pelvic region. To our knowledge, there were only two studies performed previously with gold markers placed in HNC patients. Van Asselen et al. placed two gold markers in the pharyngeal wall in 10 patients [6]. Their goal was to test the feasibility of markers to improve target positioning using portal images. Since markers were explicitly applied at places where regression/migration would be minimal, no conclusions can be drawn from this study on tumor response, i.e., regression and CTV surface motion. Zeidan et al. compared alignment based on bony anatomy versus implanted gold markers in mega-volt CT [7]. Although substantial differences were present in daily alignment, no systematic discrepancies were seen, and the impact on delivered dose was small. Tumor variability was not evaluated.

Unmistakably, markers could be used for IGRT tumor positioning and patient selection for ART. However, the markers that were placed for the studies in chapter 3 and 4 were only placed at well accessible locations, thereby only partially covering the CTV surfaces. Placing the markers is an invasive procedure, and placing them over the entire surface may lead to unacceptable complications. If the entire CTV surface were covered, geometrical variation might be decomposed into CTV motion (center of mass motion with respect to bony anatomy) and residual deformations. The CTV motion would be subject to IGRT corrections, while surface deformations could be addressed with adaptive interventions. Given the margin of 2.3 mm derived for all systematic variability (i.e. tumor motion + surface deformations), alignment on the tumor itself would only give marginal improvements. In addition, involved lymph nodes and regions suspected to contain microscopic disease are also frequently targeted in RT of HNC, and these can move independently of the tumor. Improvements in setup of the tumor could therefore compromise the setup of such other target volumes.

Comparison of marker motion patterns with GTV regression on MRI has led to the conclusion that ART on MRI involves the risk of underdosage of the initial CTV and a conservative approach was advocated (see chapter 4). On the other hand, most recurrences are seen well within the high dose region, not at the boundaries of the tumor [8–12] as might result from a geometrical miss. Therefore, within clinical trials, more aggressive forms of adaptation could be studied, for instance to see if GTV regression allows field size reduction without the loss of local control. Because in general the CTV might be overestimated [12, 13] and elective doses are given to surrounding regions, the actual lower dosage to the regressing regions may be adequate.

A further limitation of our studies related to geometrical uncertainties is the assumption that bony anatomy is a valid surrogate of nearby soft and tumorous tissue. For the tumor, the limits of this assumption were shown in the studies with gold markers. However, anatomy changes in soft tissue organs, such as weight loss or parotid gland shrinkage, frequently seen during RT of HNC [14], are not captured with the registration of bony anatomy. Ideally, with accurate...
Chapter 10

deformable image registration (DIR), all of these geometrical misalignments (bony anatomy, tumor and normal tissue) are quantified. However, accurate DIR is challenging especially with CBCT as will be discussed later.

In summary, substantial geometrical uncertainties during treatment delivery of HNC-RT were found. How to arrive at an optimal patient setup in deforming anatomy with tailored IGRT strategies is subject of the next section. In addition, we investigate if the images acquired during IGRT could be used to identify patients that might benefit from adaptive radiotherapy (ART).

10.2 Improving patient setup in deforming anatomy with IGRT

To derive an optimal patient setup in deforming anatomy we developed an IGRT positioning method based on multiple rigid registrations. Each registration unambiguously determined the local misalignment of bony anatomy. Subsequently a couch correction was derived from the average shift-corrections of the individual sub-volumes (chapter 5). This method was clinically implemented in 2009 (chapter 6) to measure deforming anatomy and to derive a table correction. It was demonstrated that a careful compromise can be reached in positioning individual sub-volumes. In addition, registrations were fast and easily verified. Also a method to objectively identify patients with large deformations was realized.

What has not been described in this thesis, but was also clinically implemented, was an offline decision rule for local deformations [15]. After we had introduced our new IGRT method to position HNC patients in the clinic, we first used a simple and easy procedure to find regions with systematic deformations (exceeding 5 mm/5° for three consecutive measurements, chapter 6). Later we converted this procedure into a decision rule for local residual errors that worked similar to an offline correction protocol with a shrinking action level for regular setup corrections [16]. This deformation decision rule raises alerts if thresholds for local misalignments have been exceeded. The thresholds are typically set to capture systematic misalignments exceeding 3 mm. A physician is then contacted to assess the effect of these deformations, and to advise on follow up. In many cases, these alerts do not trigger any intervention since the deformations occur at places with a low dose gradient (either inside, or far outside a high dose region). As the current deformation decision rule acts on geometrical errors, ignoring the dose distribution, it generates a substantial amount of false positives. Workload and complexity would greatly reduce with decision rules based on dose.

Apart from our work, little progress has been reported on patient positioning in deforming anatomy during RT of HNC. Graff and colleagues [17] investigated the dosimetric consequences of patient setup corrections based on various sub-volumes. They proposed to adjust margins to match residual uncertainties of sub-volumes. Stoiber et al. developed a coverage based optimization strategy that accounts for multiple CTVs and OARs [18]. These volumes were expanded with a special margin, presumably to account for non-setup related uncertainties. Per fraction, coverage was evaluated to select patients for adaptive intervention. Finally, Mencarelli
et al. have described a first step in dose-based setup corrections. This approach accounts for all geometrical uncertainties \([19]\), optimizes tumor control probability (TCP) with minimal normal tissue complication probability (NTCP), and takes residual uncertainties in delineation and intra-fraction motion into account. An estimate of the dose to be delivered in the present anatomy after couch correction is therefore required, as well as segmented target volumes and OARs.

A limitation in our method to derive a couch correction is the use of weights to favor a particular anatomy of the patient. How these weights should be derived is not clear. Some options to set weights are based on the distance of the registered sub-region to either the center of a representative correction reference point, the PTV surface (or the 95% isodose line), or nearby dose gradients, etc. At the time of introduction, binary weights were used, i.e., 0 and 1 to include or exclude structures depending on the extent of the treatment fields. Later, for the boost plans in sequential-boost treatments, we proposed to set weights according to the overlap of the registered regions with the high dose PTV only, since alignment of the elective PTVs in boost plans is less important. The use of weights reflects the current unwieldy approach; we act on geometry, while ultimately dose is the goal.

In summary, we developed and introduced an IGRT method to derive patient setup corrections in deforming anatomy. The proposed method allows for selecting patients for ART based on residual geometrical errors. In the next section, we discuss a more advanced ART method to update the treatment to an altered anatomy, while using accumulated dose for ART patient selection.

### 10.3 Mitigating uncorrected errors with ART procedures

To account for residual errors during RT of HNC, after rigid correction by IGRT protocols, we developed an ART strategy based on CBCT-to-CT deformable image registration (chapter 7). The strategy performs anatomy modeling entirely based on CBCT scans. It did not require acquisition of additional repeat CT scans (chapter 8). The proposed strategy represents a form of offline ART. Geometrical validation showed a 28% reduction in systematic errors with a single intervention, and 44% reduction with weekly interventions. Especially for patients with large systematic residual errors, the proposed strategy resulted in large improvements. In chapter 9, we applied the proposed ART strategy as a technique to correct target underdosage, and studied the dosimetric consequences when treatment plans are used with small safety margins. It was shown that the entire population may benefit from margin reduction (mean dose to OARs decreased with approximately \(1 \text{ Gy/mm margin reduction}\), while the adaptive strategy ensured adequate treatment for those patients with the largest risk of target underdosage. Such patients could be selected early in treatment based on accumulated dose.

The methodology presented in chapter 8 is, as far as we know, the first that effectively corrects systematic anatomical deformations. Moreover, the proposed methodology allows a fully CBCT based workflow. There is a close resemblance with the derivation of the mid-position from
4D-CT in lung with breathing motion: Wolthaus and colleagues [20] used DIR to register different breathing phases to an arbitrary reference phase. Next, the average DVF is calculated and subtracted from the 4D-DVF used to map each phase to the average (the so-called mid-position), followed by averaging the pixel values to create the mid-position CT. In both studies, a CT is created with an anatomy that is expected to be most representative of the anatomy during treatment, but that is never actually recorded. Time scales, however, are quite different: seconds (4D-CT) vs days (CBCT).

There are quite some reports that retrospectively evaluated dosimetric consequences of anatomy changes during treatment of HNC [21–28]. Most hypothesize that ART may be of benefit, although patient selection is required to maximize the clinical impact with minimal workload. In contrast, there are only a few clinical reports on the benefits of adaptive RT for HNC. Chen and colleagues [29] retrospectively looked into a group of 317 patients treated with IMRT (with PTV margins of 3−5 mm), of which 16% was ad-hoc selected for replanning. They found a significant difference in 2-year local control (88% ART vs. 78% non-ART), however, a bias may be present since the study was not randomized. For instance, good responders with considerable tumor shrinkage may have received more adaptive interventions, but also may demonstrate good local control. Ahn et al. [30] have prospectively investigated an adaptive protocol with scheduled repeat CT scanning at fraction 11, 22 and 33. If application of the current treatment plan violated target coverage or OARs-constraints, patients were replanned. In 15 of 23 patients, dosimetric benefits of replanning were present. Schwartz and colleagues [31] performed a prospective study, in which 22 patients were daily imaged with in-room CT and online positioned on cervical vertebra C2 (with PTV margins of 3−4 mm). Upon significant anatomy changes, a dosimetric analysis was performed, and a new plan was optimized without PTV margin, following experience that 3−4 mm margins with daily IGRT were frequently considered too generous for target coverage, and too large to maintain substantial OAR sparing [26]. All patients were replanned at least once, with one third of patients being replanned twice. Dosimetric benefits ranged from 0.6 to 4.1 Gy parotid gland mean dose reduction compared to IGRT only. Integral dose to the entire body also reduced. Local and regional control was 100% and 95% respectively at 2 years. Acute toxicity scores were comparable to conventional IMRT cohorts.

While some studies have shown the feasibility of calculating dose in CBCT [21, 32, 33], little is described on the use of CBCT for adaptive replanning. Yan et al. [34] described a sophisticated method that uses CBCT to construct a probability distribution function to simulate possible patient geometries and thereby estimate the total dose. This estimate was then fed back into the treatment planning system and used to tweak the objectives.

With the average anatomy model, we focused on the correction of the systematic component of the uncorrected errors, using an offline ART approach based on CBCT only. Perfect correction (100%) with offline ART cannot be reached, see chapter 8. To further reduce residual errors, online approaches could be considered, as proposed by Ahunbay et al. [35, 36] and Mohan et al. [37]. Online approaches will lead to a substantial workload: plans are modified for all fractions and all patients, since it is unknown beforehand, whether a patient will have large or small
residual errors. On the other hand, online protocols not only reduce the systematic errors, but also the random errors.

So far, HN specific IGRT protocols have been developed and clinically implemented, and efficient ART protocols have been developed and validated. Still implementation of ART protocols in the clinic lags behind, due to several challenges that are hindering clinical implementation of ART, that we will discuss in the next section.

10.4 Challenges for clinical implementation of ART

Large scale clinical application of ART is challenged by computational and organizational complexities, as well as the increased workload that is associated with repeated imaging, registration and replanning. ART was introduced already in 1997 by Yan et al. [38], and has been demonstrated to improve treatment of HNC in several studies. Still, implementation of ART protocols in the clinic lags behind compared to IGRT. In the following we will discuss some specific challenges that hinder the acceptation of ART in the clinic.

Validation and large scale clinical implementation of deformable registration

DIR is essential in adaptive procedures. Several registration algorithms have been developed, such as Demons, Optical flow, Thin-Plate Splines or B-Splines (see Brock et al. [39] for a comprehensive comparison study). The B-Spline algorithm is often used to register scans from different modalities or quality. B-Splines for medical image registration were introduced by Thevenaz and Unser [40], and further developed by Rueckert [41], Mattes [42], and Loeckx [43], as well as by Klein & Staring for use in radiotherapy [44]. Our implementation is a mixture of methods from these authors, however, the application of correlation ratio [45] as a cost function in (gradient optimizer based) DIR was not reported before. In chapter 7 we investigated the potential of the DIR algorithm to follow tumor regression, using the gold-markers from chapter 3 and 4. To quantify DIR accuracy for normal tissue alignment, anatomical landmarks were manually identified. To correct for observer variation, typically in the same order as the registration accuracy, we applied the pairwise difference ANOVA method [46].

DVFs describe position and shape changes, but cannot account for (dis)appearing objects such as feeding tubes, boluses (soft tissue equivalent material to modify the dose at/near the skin surface), filling of air cavities, etc. Moreover, density changes, as might occur, for instance, from cell kill in the tumor, are also not described by DVFs. In fact, current intensity based registration algorithms are misguided by such changes, and produce incorrect tissue-to-tissue mappings, as we found in chapter 7. It requires novel approaches to register a regressing tumor. Wang et al. [47] modified image intensities of the primary tumor in the planning CT according to the tumor cell survival rate, which was calculated with the linear quadratic (LQ) model and the delivered radiation dose to the tumor. Nithiananthan and colleagues [48], assigned a probability to voxels
Chapter 10

(an extra dimensions) to be excluded from the image. Janssens et al. applied diffeomorphic Morphons that align intensity gradients rather than the intensities themselves [49]. Alderliesten et al. proposed a dual dynamic registration, where the DVF consists of two non-fixed grids. The two grids allow to ‘shrink’ objects away in two directions [50]. Most DIR algorithms are regularized to prevent unrealistic deformations of normal tissue. As a result, neighboring voxels have correlated directional movement. Tissue that can freely move with respect to other nearby tissue (sliding tissue) is therefore difficult to capture in DVFs. Examples of sliding tissue in HN are the epiglottis, the uvula and the tongue (see for example figure 8.7).

DIR is a key tool in adaptive procedures, used to simplify decision support by quantifying anatomy changes and/or accumulating dose, and by facilitating the segmentation process for plan adaptation. For routine clinical use, DIR has to perform automated, fast, reliably, and accurate, along with minimal failures. In recent years, deformable registration algorithms have been refined and implemented on multi-core PC’s and graphics cards, making them fast (<1 min) [51]. For validation, deformable registration methods are, typically, tested on representative, but limited data to quantify failure rate and accuracy, as we did in chapter 7, as well as on artificial data with a ground truth deformation [52, 53]. Phantom studies with realistic head and neck anatomies are rare: Singhrao and colleagues experimented with a phantom from thermoplastic with additives and inserts (figure 10.1), with which two commercially available algorithms were validated for multi-modal image registration with realistic deformations and soft tissue HUs [54]. Large discrepancies with the ground truth were seen and the resulting deformation fields were vastly different. Furthermore, for deformable registration of clinical images made during treatment, quality control currently consists of visual verification. Ideally, each new registration would come with a quality measure. To this end, Bender and colleagues introduced the inverse consistency error (ICE), which quantifies the discrepancies found between the forward and inverse DVF [55]. Essentially, the ICE is a simplification of the more general full circle method that quantifies the error over multiple scans [56]. Varadhan et al. [57] proposed to use intrinsic properties of the DVF, such as the Jacobian, and the harmonic energy, to quantify the preservation of topology and the regularity of the transformation respectively. Schreibmann et al. [58] introduced the detection of vortexes (local rotations) in DVFs as indication of unrealistic registrations. Paganelli et al. used the scale invariant feature transform (SIFT) to identify identical features in planning and repeat CTs in order to quantify the DIR performance [59]. Nonetheless there is currently no consensus on how to assess DVF quality. In addition, tools are required to adjust deformation fields, for instance, by altering them to fit with a re-contoured organ. Such tools are not widely available yet. The American Association of Physicists in Medicine, by means of task group no. 132 (Use of Image Registration and Data Fusion Algorithms and Techniques in Radiotherapy Treatment Planning), acknowledges the importance of DIR for RT by reviewing, discussing, and publishing recommendations on DIR methods, implementation issues, validation techniques, and acceptance testing.
Summary and general discussion

**Figure 10.1.** Deformable phantom for validation of deformable image registration from Singhrao et al. [54] (image reprinted with permission). Based on an actual patient, bone and different tissue types were defined (a). These could physically be created in thermoplastic with additives (c). The result was divided into two halves (d) and holes for inserts were made to define landmarks for validation (visually) and/or to introduce extra contrast e.g. in the brain (b and c). By heating the thermoplastic, deformations like neck flex could by simulated, see arrows in (c). In (e) and (f) kvCT and MV CBCT scans of the phantom are shown.

**Dose accumulation and decision support**

Previously, it was argued that patient selection for adaptive interventions should be based on accumulated dose [21, 32, 33], and prediction models that estimate the resulting dose distributions [34]. For offline patient selection for ART, the delivered dose distribution has to be available quickly after a treatment session. State of the art image registration algorithms allow for the propagation of contours and the correction of Hounsfield Units in CBCT shortly after CBCT acquisition [51]. Developments in dose engines have enabled Monte Carlo based dose calculations within a minute [60]. Ideally, decision support includes prior fractions via dose accumulation, and a prediction of the final dose distribution. Moreover, decisions for plan modification should also include evaluation of the improvements possible in the remaining fractions. However, how accumulated dose should be evaluated still remains unclear. Discrepancies are typically quantified with changes in parameters, such as $D_{\text{mean}}$, $D_{99\%}$, $D_{1\%}$ or $V_{26\text{Gy}}$ and may point at opportunities to improve treatment plans. However, these changes should be evaluated against absolute levels to find clinical relevant improvements. Alternatively TCP and NTCP models could be used, which directly quantify (change in) probability of tumor control and/or complications [61]. Present NTCP models have been constructed from planned dose. How these models change with delivered dose is currently unknown and should be investigated in more detail [62]. Similarly, it is also unclear if and how the dose constraints and objectives from the original treatment plan can be applied in accumulated dose to select patient for adaptive intervention. Furthermore, dose accumulation is not only important for patient selection as the incorporation of dose delivery history in the re-optimization process has been demonstrated to improve the final delivered dose.
distribution [63]. Notably, while for patient selection errors in the accumulated dose may lead to false positives, and therefore unnecessary extra replanning, such errors may lead to systematic under/over dosage, if used as background dose during replanning. Importantly, thresholds for selection of patients for ART directly define the clinical workload. Reliably quantifying the possible benefits of adaptive interventions and balancing them with clinical workload is an area that has not been investigated in great depth.

**Automation of replanning to reduce workload**

Automation of planning alleviates workload and leads to consistent plan choices, necessary for the acceptance of routine offline ART protocols in the clinic. Several techniques are available to reduce the time required for plan adaptation: DIR followed by contour propagation may facilitate target and OAR definition in new CT scan, and starting an new plan optimization with the objectives and beam parameters from the original plan may result in less iterations to reach an optimum. Still, (adaptive) planning requires considerable manual tweaking, and is therefore regarded labor intensive and subjective. Roughly two strategies are seen to further automate the planning process: 1) Knowledge based, which requires a library of high quality plans, from which plans are selected that resemble the current OAR and PTV geometry. The corresponding objectives are then used to drive the optimization process [64]. 2) Using multi-criteria optimization (MCO), in which there are two flavors: a posteriori, where a single plan is selected/interpolated from a pre-generated library of plans with different compromises. Generation of the (patient specific) library may possibly lead to substantial computation times. MCO can also be performed a priori, where optimization starts from a single plan. iCycle is such an a priori plan optimization approach developed by Breedveld and colleagues [65]. iCycle uses wish-lists, containing constraints and prioritized objectives, to traverse solution space, until a satisfying plan is reached. Generally MCO optimization takes place in fluence-space, followed by a conversion to a deliverable plan (sequencing). The conversion for IMRT works well, but is challenging for VMAT. Yet, Voet et al. have succeeded in using wish-lists with iCycle for HNC VMAT [66]. Tol and colleagues have developed and clinically implemented an approach that mimics the choices that are made by experienced planners for RapidArc [67, 68]. In our study in chapter 9, we automatically optimized VMAT plans with a similar approach. Treatment planning systems more and more offer auto-planning algorithms. For example, Krayenbühl et al. validated the auto-planning module in Pinnacle (Philips Radiation Oncology Systems, Fitchburg, Wisconsin, US) for planning HNC radiotherapy [69].

**Administrative burden**

Extra resources for ART are not only required for additional imaging, DIR and its validation, and replanning, but substantial time is spent on administrative handling, despite fully paperless workflows are used nowadays. The radiotherapy process today is essentially one-directional, i.e., first a preparation phase with pre-treatment imaging, target definition, and treatment planning,
followed by an execution phase, with IGRT based patient setup and subsequently dose delivery, see figure 1.1. In adaptive workflows, the loop is closed and the preparation phase is repeatedly entered. Current workflows are not designed for this loop. Plan checks, approval, transfer, and entry in the clinical database are all essential, yet time consuming steps that inhibit smooth integration of adaptive workflows.

10.5 Future directions

New developments are underway that may improve RT of HNC. With biological adaptive RT (BIGART), treatment is adapted to biological response instead of anatomical changes. BIGART can be explored to its full potential with MR integrated radiotherapy machines. Furthermore, proton therapy allows great improvements in the spatial accuracy of dose distributions. In this section we will discuss these developments in relation to the content of this thesis.

Biological ART

Since biological functions of tumor subregions can change over time, a natural extension of ART on anatomical changes is ART based on observed biological response, i.e. BIGART [70]. Biological response may be imaged with PET or MR. For instance, changes in apparent diffusion coefficient maps (ADC maps, derived from diffusion weighted (DW) MRI) have been reported to predict early response, see Kim et al. [71] or Dirix and colleagues [72]. Geets et al. acquired CT, MR and PET scans at several timepoints during treatment to study the impact of different modalities on target volume delineation and replanning [22]. Substantial GTV reduction was present in all modalities, however PET rendered the smallest volumes. Regular replanning based on these volumes did marginally impact OAR dosage, however, opportunities for dose escalation were suggested. These studies show that there is potential for BIGART to adjust target volumes during treatment and to (de-) escalate dose after early response measurements with PET or MRI. However, more research is required to disentangle the sometimes conflicting information that comes from different functional imaging techniques, as underlined by Houweling et al. [73].

BIGART closely relates to dose painting: redistribution and/or escalation of dose to radioresistant parts of the tumor [74]. Where BIGART aims to modify the treatment according to measured response during radiotherapy, and thus requires adaptive workflows, dose painting is part of the treatment preparation phase and uses pre-treatment functional imaging to shape the dose distribution. BIGART tries to distinguish good from bad responders, and may prevent unnecessary toxicity (by dose de-escalation, or switching to alternative treatments) or increase tumor control by dose escalation. Dose painting aims to improve local and locoregional control. For instance, FDG-PET uptake is considered as a surrogate for hypoxia, a microenvironment with increasing radioresistence, and has been shown to correlate with treatment response and loco-regional failure [75–77]. Soto et al. [11] and more recently Due and colleagues [9] have linked the location of recurrences with high uptake of FDG-PET in pre-treatment PET-scans. Several trials
are being performed to find clinical evidence of improved local and locoregional control after
dose escalation to biomarker identified sub-regions. In Belgium, De Neve and colleagues have
started a trial, NCT01341535, where HNC patients are treated with adaptive dose painting by
numbers based on (repetitive) FDG-PET/CT scans. The standard arm uses conventional IMRT. It is
expected that dose painting with BIGART may improve cure rates without increase of radiation
induced toxicity [78–81]. In University Hospital Tübingen, a trial is ongoing (NCT02352792 )
to investigate local control in HNC patients with a 10% dose escalation to hypoxic volumes
determined with FMISO-PET/CT (experimental arm) compared to conventional chemoradiation
(standard arm). The Netherlands Cancer Institute leads a multi-institutional trial (ARTFORCE,
NCT01504815) to investigate improvements in locoregional control by redistributing the dose to
the tumor in order to boost potential radio-resistant regions, as determined from FDG-PET/CT,
while aiming at the same level of toxicity. An adaptive intervention is performed to account
for anatomical changes at the end of the 2nd week of treatment. The standard arm receives
conventional chemoradiation [82]. With these trials, uniform dose distributions are replaced
with non-uniform distributions. Therefore, precise positioning and adaptation to geometrical
changes, as studied in this thesis, becomes even more important.

The integration of functional imaging with the linear accelerator would allow daily assessment
and adaptation to biological response, as well as a reduction of the logistic burden of making
extra scans elsewhere in the hospital. The first step in this direction is the integration of MR
with linear accelerators, which may play a key role in the routine implementation of BIGART, as
they allow to frequently monitor biological function of the tumor and surrounding tissue during
treatment.

MRI integrated radiotherapy machines

Worldwide, several projects are ongoing to integrate MR imaging with radiotherapy machines, see
table 10.1. Compared to X-ray based in-room imaging technologies such as CBCT, MRI delivers
images with superb soft tissue contrast while eliminating imaging dose. The improved image
quality would allow faster and better redefinition of the tumor and other anatomical structures
for replanning, and thus, the adaptation of the treatment to anatomical changes on a daily
basis. Moreover, MRI allows functional imaging and thereby, treatment response monitoring
as required for BIGART. In MR imaging, the magnetic field strength is an important parameter
that determines the signal-to-noise ratio. Yang et al. explored the use of DW-MRI in the 0.35T
Viewray system [83]. Further studies are needed to show how much of the potential of imaging
biological function is affected by the available magnetic field strength. For a further discussion
on the potential and challenges of MRI-guided ART see Kupelian and Sonke [84].

Generally, MRI guided RT has two important challenges to overcome: 1) MRI does not provide
Hounsfield units (electron density) for dose calculation. 2) Image distortion occurs that is inherent
to MRI. A further challenge for adaptive interventions for HNC on the MRlinac was identified in
this thesis: we showed that MRI was unable to properly visualize CTV shrinkage and field size
Table 10.1. Overview of current developments towards irradiation machines with MRI.

<table>
<thead>
<tr>
<th>System/project</th>
<th>Magnetic field strength</th>
<th>Radiation technique</th>
<th>Research group</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRIdian [85]</td>
<td>0.35 T</td>
<td>3 Co-sources or Linac</td>
<td>ViewRay</td>
</tr>
<tr>
<td>Linac-MR [86]</td>
<td>0.5 T</td>
<td>6 MV Linac</td>
<td>Cross Cancer Institute</td>
</tr>
<tr>
<td>Australian MRI-Linac</td>
<td>1.0 T</td>
<td>6 MV Linac</td>
<td>Collaboration of groups in Australia</td>
</tr>
<tr>
<td>MR-LINAC [88]*</td>
<td>1.5 T</td>
<td>6 MV Linac</td>
<td>Elekta/Philips</td>
</tr>
</tbody>
</table>

*) The Elekta/Philips-MR-LINAC (figure 10.2) is being installed at the NKI at the time of writing of this thesis.

Reduction could, therefore, compromise the dosage of microscopic extensions. On the other hand, the original CTV may possibly be treated to a lower dose level after observed GTV shrinkage. However, such experimental protocols should be studied in prospective clinical trials.

Figure 10.2. Artist impression of the MR linac currently being developed by Elekta and Philips after a design by Raaymakers et al.[89]. The superconducting coils have been specially designed to produce a magnetic field for acquiring MR images with diagnostic quality and to allow a compact 6MV linear accelerator to rotate around the patient to accurately deliver radiotherapy. Image reprinted with permission of Elekta.

Both MR and PET are routinely used to detect involved lymph nodes in treatment of HNC. Over the years, image quality and resolution have increased and thereby lowered the detection limit.
of involved lymph nodes [90, 91]. As a result, remaining suspected lymph nodes are smaller and may be treated at lower elective dose levels. A multi-institutional trial has been initiated by the Radboud University in Nijmegen (NCT02442375) to investigate de-escalation of elective radiation dose. Moreover, with the MRlinac, these smaller involved lymph nodes might be visualized during treatment, allowing to target these nodes with less uncertainty [84].

As plan optimization becomes faster, online replanning becomes within reach. MRI provides high quality images of the actual anatomy just prior to treatment for use in plan (re)optimization. By using GPU's for Monte Carlo dose calculations including electron interactions with the magnetic field, it was demonstrated that IMRT plans could be optimized within minutes[92]. Furthermore, Bol et al. [93] showed how online IMRT plan optimization could be integrated in their future clinical workflow for the MRlinac: initially, to allow patient setup corrections by virtual couch shifts (couch corrections are not possible in the MRlinac), and later, to account for all anatomic discrepancies. Moreover, with imaging during treatment delivery and an optimization algorithm, open to inter and intra-fraction anatomy changes, inline adaptation becomes feasible. With inline adaptation, the treatment plan is iteratively optimized during delivery, and a convergence to an ideal dose distribution can be reached even in moving anatomy, as was evaluated for a case of kidney cancer on the MRlinac by Kontaxis et al. [94]. As with replanning with a background dose, errors in DIR may propagate into the estimated delivered dose, and, thereby, incorrectly steer the optimization algorithm. Whether online replanning for interfractional changes, or inline replanning for intrafractional changes will lead to clinically significant improvements in delivered dose for HNC remains to be investigated.

**Proton therapy**

At present many proton centers are opening worldwide. In The Netherlands, three proton centers are under development and these aim to open in 2017-2018. Compared to photon beams, proton beams deposit less dose in their path towards the tumor and leave no exit dose beyond the Bragg-peak, while skin dose (first few mm) may increase [95]. These characteristics allow creating highly conformal dose distributions with reduced dose outside target volumes compared to photons (see for example figure 10.3). For RT of HNC, the benefits of proton therapy are potentially high, since target volumes are usually located near multiple critical organs. Various planning studies have been undertaken to investigate achievable dose distributions of protons vs photons. In an overview by Van de Water et al. [96], generally a lower dose to normal tissue with better target coverage and homogeneity was found for protons compared to photons. Of all examined proton delivery techniques, they considered intensity modulated proton therapy (IMPT) most advantageous. Confirmation of superiority should be studied in properly designed clinical trials. In search of clinical evidence that proton therapy for HNC improves local outcome, Holliday and Frank [97] reviewed literature for prospective and retrospective reports on proton therapy for HNC that reported outcome. Reports were limited (18 in total), but a trend was found that for the tumor sub-sites base of skull, nasal cavity, paranasal sinus, nasopharynx and oropharynx, proton therapy was not only feasible and safe, but also superior in normal tissue sparing, while
either maintaining or improving local control. No reports for other sub-sites were found. At the MD Anderson Cancer Centre, a phase 2/3 prospective randomized trial (NCT01893307) has been initiated to compare the side effects of proton treatments for oropharyngeal cancer versus photon IMRT/VMAT [98].

![Figure 10.3. Example dose distribution of a photon plan (left) and a proton plan (right). Proton plans may achieve a higher conformity of the dose to the planned target volume (purple line), leading to better sparing of organs at risk such as the spinal cord (blue line).](image)

Protons treatments are, however, more sensitive to uncertainties such as range uncertainties from density changes, motion, deformations and anatomy changes, and setup errors than photon treatments, as was demonstrated for IMPT by (for instance) Kraan [99] et al and Barten et al. [100]. In both studies, increasing the number of beams, as first expected to reduce the effect of range-uncertainties, did not improve results. In an overview on proton treatment planning, McGowan et al. [101] concluded that the use of a PTV margin to account for uncertainties with sophisticated IMPT plans did not ensure the delivery of a uniform dose distribution to the CTV. Switching to robust planning, range uncertainties could be mitigated at the cost of conformity. It was remarked that uncertainties largely depended on the IGRT protocols in place, and that ART could restore the delivered dose to the planned dose in the case of unavoidable changes occurring in the patient. These studies emphasize the importance of IGRT and ART protocols tailored to proton therapy of HNC, and most methods described in this thesis would be applicable.

10.6 Conclusion

In conclusion, this thesis investigated the use of in-room imaging with CBCT to quantify and reduce geometrical uncertainties during radiotherapy for HNC. Substantial residual geometric
uncertainties were found, arising from deforming and changing anatomy, as well as complex tumor regression over the course of treatment. With new HNC specific IGRT protocols, these setup uncertainties can be quantified and managed using optimized patient positioning. All HNC patients treated in our clinic are nowadays positioned with such protocols, leading to smaller uncertainties. For residual errors that cannot be corrected with patient positioning, we developed adaptive approaches. To this end, deformable image registration algorithms were developed, implemented and validated. The adaptive approaches enable highly automated, efficient, CBCT based, adaptive workflows that further reduce geometrical uncertainties. Due to the work in this thesis, the precision of RT delivery for HNC has improved, and smaller margins are nowadays used, possibly leading to reduction of radiation induced side effects.
References


Chapter 10


Summary and general discussion


Chapter 10


Published Papers


Markers. *International journal of radiation oncology, biology, physics* 2012;84:e201–e207


**Submitted Papers**

Summary

Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer

Radiotherapy (RT) of head & neck cancer (HNC) is a toxic treatment. Reduction of volumes that receive a high dose leads to opportunities for dose escalation or reduction of toxicity. Pre-treatment multi-modal imaging has allowed to better visualize the tumor with respect to healthy tissue, resulting in precise delineation of the target volume. Introduction of new planning and delivery techniques has lead to spatially highly accurate dose distributions. To account for geometrical uncertainties, arising from patient setup variation and changes in anatomy, safety margins are applied. The introduction of cone beam CT (CBCT) has provided regular imaging in treatment position in which patient setup and anatomy variation can be visualized and quantified. With image guidance RT (IGRT) and adaptive RT (ART) tailored to RT of HNC, geometrical uncertainties can be reduced and hence the irradiated target volumes. Moreover, tumor regression may lead to opportunities to reduce irradiated volumes. In this thesis we investigate the use of in-room imaging with CBCT to quantify and reduce geometrical uncertainties during treatment delivery for HNC, including tumor regression. In addition, we set out to investigate the potential of ART with CBCT for radiotherapy of HNC.

In chapter 2, occurrence and magnitude of deformations during RT of HNC were studied and the consequences on margins were assessed. Multiple region of interest (mROI) registration, applied to individual bony structures, was used to quantify residual errors after clinical, offline setup corrections, which was based on an encompassing large region of interest. It was found that substantial deformations were present, frequently underestimated by the residuals that the offline correction protocol allowed. Furthermore, it was demonstrated that even in a simulated online approach large residual errors would be present. Errors increased with distance from the registered region. Local setup errors often exceeded the uniform 5 mm margin used in clinical practice.

In the study in chapter 3, the tumor shape variations were studied with gold markers implanted near the edges of the tumor. For 27 patients 153 helical shaped gold markers were surgically
implanted at the edge of the GTV (nonuniformly) and tracked during treatment in CBCT scans after alignment on bony anatomy near the tumor in the planning CT. Considerable variation was found: the mean 3D error was 0.3 cm over the whole treatment. Errors increased with treatment time. Large errors occurred more often in bulky tumors. Markers in the base of tongue were more mobile than in the tonsil or posterior pharyngeal wall. For RT, assuming dose distributions conformal to the PTV and thus CTV/GTV, marker motion patterns perpendicular to the original GTV surface are most important. Markers were found that moved inwards (representing opportunities for field size reduction) and outwards (to be covered by the safety margin). A 0.23 cm margin was found to be sufficient to cover the patient specific systematic errors in 90% of patients. Currently, margins recipes do not account for this type of residual tissue motion. In 48% of patients the majority of markers moved inwards, in 7% little marker movement was seen, in 45% markers were present that moved inwards, outwards or both. Especially with tumors in the posterior pharyngeal wall, cranial and caudal borders were found to move outwards. It was discussed that markers are representative of the location of microscopic disease, and that only in case of inwards marker motion CTV reduction could be considered.

CTV regression is difficult to assess not only on CBCT, but also on CT or MR. GTV regression can clearly be visualized in MR, but GTV shrinkage may not correlate to CTV regression. Therefore, in chapter 4, a study is performed to compare GTV regression as seen in MR, with CTV changes as derived from motion patterns in CBCT of implanted gold marker. The study included 8 patients, daily CBCT, 40 markers and repeat MR prior to treatment and in week 3 and 6. It was shown that replanning based on observed GTV reduction in MR at week 3 and week 6 reduced OAR dose. Furthermore, it was found that the distance of the markers to the MR defined GTV surfaces increased during treatment. It was concluded that GTV reduction in MR was larger than expected from CTV reduction in CBCT. It was concluded that solely reducing field sizes according to observed GTV regression in MR is therefore possibly unsafe.

The aim of the study in chapter 5 was develop an IGRT procedure to optimize the couch shift correction for RT of HNC in the presence of deforming anatomy. To this end, mROI registration, previously used to quantify local residual errors, was used. Two strategies were developed, and compared, to convert the mROI registration results into a couch shift correction: minimization of the mean error and minimization of the maximum error. Differences between both methods were small, but the minimization of the mean error was preferred since it led to reduction of errors in nearly all registered regions and was thought to be less sensitive to occasional mis-registrations. Furthermore, the residual errors were used to assess local deformations as basis for patient selection for adaptive intervention. It was found that for a single intervention in a typical treatment schedule with 35 fractions, evaluation at fraction 8 would lead to the best balance between accuracy of estimation of deformations and sufficient remaining fractions to achieve a substantial improvement from adaptive interventions. Furthermore, a relation between threshold for intervention and number of selected patients was derived, allowing to tune the clinical workload.
In chapter 6 we report the clinical introduction of the mROI registration method to drive setup protocols in RT of HNC. The method was applied to 578 CBCT scans in 50 patients and compared to regular single large ROI registration. Application of an automated method to define 12/13 separate regions of interest in HN CT scans resulted in minimal extra preparation time. The mROI registration was also automated and required 10 seconds, which was considered sufficiently fast to use in clinical practice. The validation of the automatic registration was facilitated with a thin plate mapping method, which allowed to assess the registration results visually in one go. A protocol to monitor deformations was developed with warnings if local setup errors after correction exceeded 5 mm/5°. Errors were considered systematic if warnings occurred 3 times in a row. This protocol showed that the hyoid region often deformed, mainly on rotational errors. 52 times, systematic deformations were found and physicians were consulted. It was concluded that objective measures could be used to find patients that might benefit from adaptive interventions.

Deformable image registration (DIR) of the CBCT to the planning CT for HN patients was evaluated (chapter 7) on its accuracy (systematic discrepancies) and precision (random discrepancies) to register normal and tumorous tissue. Discrepancies were quantified by the fiducial registration error, for soft tissue from manually identified unique features (2 observers), for the tumor from automatically identified implanted gold makers. 13 patients were analyzed with on average 6 gold markers and 15 soft tissue features. Analysis of variance on the manually identified features was used to separate registration errors from observer variation. It was found that the accuracy for soft tissue and the tumor was sub-millimeter, while the precision (3D vector length) was 1.8 mm for normal tissue, and 3.3 mm for tumorous tissue. In contrast to normal tissue, a substantial time-trend was present in the registration of the tumor: gold marker precision reduced from 2.5 mm at the start of treatment, to 4 mm at the last week. In the discussion, the use of DIR for dose accumulation was discussed, where it was remarked that registration errors only matter at dose gradients. However, when applying DIR for re-definition of the tumor, additional uncertainties are introduced. DIR for adaptive application should therefore be used with caution.

In chapter 8, we introduced a method to overcome systematic deformations by adapting treatment based on an average of the anatomical deformations observed in successive CBCT scans (average anatomy modeling). Deformable registration is used to quantify how the anatomy differs locally from the planned situation, i.e. the deformations. By averaging the deformations and re-applying them to the anatomy in the planning CT, an anatomy is created without systematic deformations. The study reports a geometrical validation of the method. Therefore two strategies were tested, a single adaptive intervention, and weekly interventions with the average anatomy of the previous week. Residual geometrical errors were quantified based on fiducials derived from segmented bony parts and fiducials manually identified in soft tissue. The bony anatomy and soft tissue systematic errors were reduced from 2.5/3.4 mm to 1.5/2.7 mm with the single intervention strategy, and to 1.0/2.2 mm with weekly interventions. Especially in patients with large systematic deformations benefits were large. Observer variation in identification of soft tissue fiducials may have increased the observed errors. In the discussion, the advantages and shortcomings of anatomy modeling vs. conventional repeat CT imaging are discussed, as well as the need for accurate and precise deformable registration. In addition, a conservative approach
to adaptation to tumor regression is advocated. Finally it is remarked that a planning study is
needed to investigate the dosimetical benefits of average anatomy modeling.

In chapter 9, we investigated the impact of anatomical changes on the delivered dose as function
of applied safety margins (5 /3 /0 mm). Margin reduction generally lead to reduced OAR exposure
by on average 1 Gy/mm. On the other hand, target underdosage occurred more frequently as
margins became smaller, up to 27% of all evaluated CTVs in 0-mm plans. Significant extra dose to
critical organs was infrequently seen. Patients at risk of underdosage >2 Gy could be identified
early in treatment, at fraction 10, with dose accumulation. A single adaptive intervention, using
average anatomy modeling, effectively mitigated loss of coverage in 5 out of 6 patients with
more than 3 Gy.

In the final chapter, we reflect on the findings and limitations of the studies in this thesis in
relation to other studies. We discuss how to bring adaptive protocols to the clinic and how to
take the results of this thesis into the next generation of radiotherapy equipment.
Nederlandse samenvatting

Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer

Bestraling is een essentieel onderdeel van de behandeling voor hoofd/hals (HH) kanker, met goede resultaten tot gevolg. HH kanker patiënten die bestraald worden ondervinden echter vaak ernstige bijwerkingen. Door het bestraalde gebied zo klein mogelijk te maken kunnen deze bijwerkingen worden beperkt. Moderne bestralingstechnieken zijn uiterst nauwkeurig: de rand van het gebied dat een hoge bestralingsdosis ontvangt wordt strak om de tumor gelegd en heeft een snelle dosis afname daarbuiten. Hierdoor worden kritieke organen gespaard en blijven bijwerkingen beperkt. Echter, tijdens de behandeling met dagelijkse bestralingen, die 6 tot 7 weken duurt, wijkt de anatomie van de patiënt vaak af van de geplande situatie onder andere door flexibiliteit van de nek, gewichtsverlies en tumorkrimp. Hierdoor past het bestralingsplan minder goed op de anatomie van de patiënt. Op het moment worden deze geometrische onzekerheden opgevangen met een veiligheidsmarge. Mogelijk zou de patiënt profijt kunnen hebben van een aanpassing van het bestralingsplan tijdens de behandelperiode (Adaptieve RadioTherapie, ART). De meeste bestralingstoestellen zijn tegenwoordig uitgerust met een Cone Beam Computed Tomography (CBCT) scanner. De CBCT scanner levert beelden van de anatomie van de patiënt in de behandelruimte (in-room imaging), in behandelpositie, vlak voordat de bestraling wordt uitgevoerd. Deze beelden worden vergeleken met de geplande situatie om een tafelverschuiving af te leiden die de tumor nauwkeurig op de juiste plaats positioneert. Dit proefschrift onderzoekt de mogelijkheden van in-room imaging met CBCT voor HH kanker om geometrische onzekerheden te kwantificeren en te verkleinen, met en zonder ART. Dit is niet triviaal omdat de beeldkwaliteit van CBCT minder is dan van een reguliere CT scanner.

Ondanks het gebruik van individuele maskers om de patiënt te immobiliseren, kan de hals nog steeds vervormen. We hebben die vervormingen bestudeerd door de plaats van de botten in de nek te bepalen op CBCT scans. Het bleek dat aanzienlijke vervormingen voorkwamen en een veiligheidsmarge van 5 mm mogelijk niet voldoende was (hoofdstuk 2). Botten zijn duidelijk zichtbaar in CBCT scans, maar de tumor is niet goed te zien. Om te meten hoeveel de tumor vervormt (hoofdstuk 3), hebben we kleine gouden markeringen, markers, aangebracht in het
Samenvatting

weefsel rondom de tumor. Deze rand met markers is representatief voor de tumor plus het gebied waar microscopische tumor-extensies verwacht kunnen worden, ook wel het Clinical Target Volume (CTV) genoemd. Wanneer de markers naar binnen bewegen was er sprake van tumorkrimp en zou het te bestralen gebied mogelijk verkleind kunnen worden. Als de markers naar buiten bewegen moet dat worden opgevangen door een grotere veiligheidsmarge. In een groep van 27 patiënten waren bij twee patiënten weinig vervormingen te zien, bij ongeveer de helft was sprake van tumorkrimp en bij het restant was de beweging op verschillende plekken zowel naar binnen als naar buiten. Een extra veiligheidsmarge van 2.3 mm zou voldoende zijn om de naar buiten gerichte bewegingen op te vangen. Zonder de markeringen is CTV regressie niet te zien op CBCT. Een MRI scanner heeft een betere beeldkwaliteit, maar ook daar is de CTV rand niet te zien. Krimp van de tumorbulk (Gross Tumor Volume, GTV) is echter wel goed waar te nemen. Daarom hebben we GTV krimp op MRI vergeleken met eerder gevonden veranderingen op CBCT met markers (hoofdstuk 4). We vonden dat de GTV regressie op MRI groter was dan de CTV-regressie op CBCT. Herplannen nadat GTV regressie op MRI is gezien is daarom mogelijk onveilig omdat een deel van het CTV kan worden gemist.

Het positioneren van patiënten met vervormingen is lastig omdat deze vervormingen niet met één tafelinstelling kunnen worden gecorrigeerd: er blijven altijd residuele fouten over. Hoofdstuk 5 beschrijft daarom de ontwikkeling van een praktische methode om vervormingen te meten en tot een optimale tafelinstelling te komen: de positie van afzonderlijke botten werd bepaald en als optimale tafelverschuiving werd een gemiddelde correctie berekend. De residuele fouten na deze tafelverschuiving werden gebruikt om patiënten te vinden die konden profiteren van ART. De introductie van deze methode in de kliniek staat beschreven in hoofdstuk 6.

Voor ART wordt vaak gebruik gemaakt van zogenaamde vervormbare beeldregistratie (Deformable Image Registration, DIR). Met DIR wordt een scan elastisch vervormd zodat hij past op een andere scan. Met DIR kan ART aanzienlijk worden vereenvoudigd: intekeningen van het doelgebied en kritieke organen worden daarmee automatisch aangepast aan een herhaal CT. In hoofdstuk 7 bepaalden we de nauwkeurigheid van een DIR algoritme op CBCT scans. We beoordeelden de nauwkeurigheid met kenmerkende punten in normale weefsels die handmatig waren aangewezen in zowel de planning CT als de CBCT scans. Met de eerder genoemde markers beoordeelden we de nauwkeurigheid voor de tumor. DIR voor normaal weefsel was nauwkeuriger dan voor de tumor (fout van 1.8 mm vs. 3.3 mm). Bovendien werd de registratie van de tumor minder nauwkeurig naarmate de behandeling vorderde. Als DIR voor ART wordt gebruikt, moet daar rekening mee worden gehouden.

In hoofdstuk 8 gebruikten we DIR om een ART-methode te ontwikkelen die ongevoelig is voor dag tot dag variatie van de patient. Wanneer voor ART een herhaal CT scan wordt gemaakt om de veranderde anatomie vast te leggen, kunnen ook incidentele vervormingen optreden. Hierdoor is de scan minder representatief voor het restant van de behandeling. Met DIR bepaalden we de dag-tot-dag vervormingen in een serie CBCT scans. Door resultaten van een aantal CBCT scans te middelen, berekenden we de zogenaamde systematisch afwijking van de planning CT ten opzichte van de gemiddelde anatomie. Deze vormde vervolgens de basis voor een nieuwe planning. Het
bleek dat de afwijking daarmee 35%-60% verminderde, afhankelijk van de implementatie in een praktisch ART-protocol. Vooral voor patiënten met grote afwijkingen werkte deze methode goed. Deze methode werd ook gebruikt in hoofdstuk 9 waar we onderzochten of mogelijk kleinere veiligheidsmarges kunnen worden gebruikt. Met kleinere marges worden kritieke organen beter gespaard, maar neemt het risico toe dat het CTV te weinig dosis ontvangt. De studie bevestigde dat zonder ART-protocol marges van 5 mm afdoende waren om vervormingen op te vangen. Bij kleinere marges raakte het CTV in toenemende mate ondergedoseerd. Met één enkele ART-interventie vroeg in de behandeling, kon de CTV dosis bij de meeste patiënten worden hersteld.

Tenslotte beschrijft de discussie de uitdagingen die er zijn om adaptieve protocollen in de klinische praktijk toe te passen. Vooruit blikkend op nieuwe bestralingstechnieken, zoals bestraling onder MRI-beeldsturing of met protonen, is gekeken in hoeverre de bevindingen en ontwikkelde methodes in dit proefschrift van toepassing zijn.

In conclusie, het werk in dit proefschrift heeft geresulteerd in een nauwkeurigere bestraling van HH kanker patiënten zodat kleinere veiligheidsmarges kunnen worden gebruikt. We verwachten dat ernstige bijwerkingen hierdoor zullen afnemen.
Eindelijk! Mijn promotie heeft lang geduurd. Tussen de start en dit boekje heeft bijna 12 jaar gezeten. Het was een periode gevuld met trotse momenten, frustraties, afleiding, onderbreking, inzicht, voldoening, en volharding. Velen hebben mij aangespoord, van raad voorzien, van inzicht, discussie, humoristische afleiding, technische ondersteuning of alleen van koffie of bier. Ik kan niet iedereen bij naam danken, maar weet dat ik jullie allemaal zeer waardeer!


Marcel, jouw groep gaf vanaf de start een vertrouwde omgeving waarin jij onconventioneel met creatieve ideeën mijn onderzoek verder bracht. Je bevlogenheid, kennis van natuurkunde, statistiek, beeldverwerking, programmeren en historisch besef is indrukwekkend en enorm motiverend. Manchester mag zich gelukkig prijzen. Je bent een ware mentor, dank dat je mij zover hebt gebracht.

Jan-Jakob, hoe druk kan een mens zijn? Toch: jouw deur staat altijd open voor inhoudelijke discussie! Jouw scherpzinnige blik, en zeker ook gevoel voor Engels, heeft mijn studies en papers altijd op een hoger niveau gebracht. Soms lijkt het dat jij mijn ideeën beter ziet dan ik. Ik heb veel van je geleerd, ook tussen de regels door. Dank voor het vertrouwen dat ik altijd heb gevoeld. Ik was vermoedelijk je eerste student, hoe fijn zal het zijn dat je mij nu kan loslaten! Helemaal nu dat je zelf hoogleraar bent!

Coen, de arts die ook wel natuurkundige had kunnen zijn. Jouw vragen hebben mij altijd gedwongen terug naar de basis te gaan en beter te snappen wat er speelde en dat duidelijk uit te leggen. Ik heb met veel plezier samen gewerkt, ook in de begindagen van de IGART groep. Jouw praktische en pragmatische aanpak botste nog wel eens met mijn wens om de introductie van nieuwe technieken nauwgezet voor te bereiden, toch kwam er veel van de grond. Zelfs vanuit het AMC kwam je nog wel eens aanwaaien om mij op het hart te drukken dat het er echt in zat. Nou, hier is het!

Suzanne van Beek, partner in crime van het eerste uur, je positieve inslag is aanstekelijk! Dankzij jou blijven studies niet op de plank liggen. Hoeveel onzichtbaar werk heb jij wel niet gedaan?
Dankwoord

Jij vertegenwoordigt de link naar de kliniek. En dan naast je research werkzaamheden ook de andere helft van de week gewoon nog in de kliniek staan. Petje af. Maar je bent meer dan een betrokken collega; de ontwikkeling van jou twee kinderen liet gelijk aan mijn eerste twee, tot in Haarlem. We hebben veel ervaringen gedeeld en ik heb alle jaren met enorm veel plezier met je samengewerkt, dank.

**Angelo Mencarelli**, mijn andere partner in crime. Samen hebben we menig studie uitgewerkt. Veel werk had niet mogelijk geweest zonder jouw bijdrage. Maar als Italiaan zijnde heb je mij ook allerlei andere finesses meegegeven, van ulitme beleefdheden tot perfecte koffie. Maestro, jouw moment is aanstaande, nog even doorbijten!

**Olga Hamming-Vrieze**, jij kwam als de KNO radiotherapeut toen Coen naar het AMC vertrok. We hebben samen aan maar liefst 4 artikelen gewerkt, en de pijp is nog niet leeg. Ik werk met veel plezier met je samen, en hoop dat ik nog een verdere bijdrage aan jouw promotie mag leveren, dank voor de samenwerking en support.

**Annelisa Wolf**, ik moet jou zeker danken. Eigenlijk had je geen tijd, maar toch heb je eindeloos scriptjes getweakt om een studie mogelijk te maken. Dank daarvoor!

**De promotiecommissie**, professor van den Brekel, doctor Fuller (thank you for coming to Amsterdam!), professor Heijmen, professor Raaymakers, professor Strijkers, en doctor Bel, ik wil u danken voor het zorgvuldig lezen en beoordelen van mijn scriptie, en het oppositie nemen op 29 september. **Patricia en Diedie**, dank voor het loodsen naar een veilige haven te midden van alle regels en plichtplegingen, jullie zijn het smeermiddel van de afdeling.

Mijn IGART kompanen: **Suzanne van Beek, Rianne de Jong, Nelly Kager, Peter Remeijer, Folkert Koetsveld, Nienke Hoffmans-Holtzer, Lisa Hartgring, Albert Licup, Sareh Nakhaee, Olga Hamming-Vrieze, Monique Bloemers, en alle Imaging laboranten**, jullie zijn de rechtvaardiging van al wat uit de koker van de research komt. Het is plezierig en dankbaar werken met jullie.

Dank ook aan mijn hulplijnen: Emmy Lamers-Kuijper, Eugène Damen, Wilma Heemsbergen, Lennert Ploeger, Lambert Zijp, Joop Duppen, Anke van Mourik, Matthew La-Fontaine, Matthijs Kruis, Barabara Stam, hoe lang was ik nog bezig geweest zonder jullie hulp?

**Monique Smitsmans**, samen met **Jochem, Jasper, Leah en Marnix** was jij aanwezig toen ik startte. Ik kijk met veel plezier terug op die eerste tijd, waarin alles nieuw was en ik veel ontdekte. Jij gaf het goede voorbeeld met jou promotie; dat ik daarbij paranimf mocht zijn heeft me zeker gestimuleerd. Je hebt altijd in mijn hoofd gezeten als ware een geweten (en was ik maar minder eigenwijs). Je hebt me altijd voorgehouden dat er een dag komt dat ook mijn ‘boekje’ af is, dit is de dag. Dank voor je aanmoedigende woorden al die tijd en alle andere gezellige momenten samen.

Lieve familie, **mama en Dé**, ik ben een eigenwijs persoon, ik heb altijd mijn eigen pad gekozen, toch heb ik altijd jullie steun gehad. Het is altijd thuis komen bij jullie. Moeders, je vroeg laatst of je trots kan zijn op andermans prestaties, nou ik zou het maar zijn, en ook namens
Rob! Christiaan, André, Marlies, Jasper, Amanda, Anne Marlein, Renée, Lex, Annie, jullie zijn de fijnste familie die ik me kan wensen, dank voor jullie interesse, geduld en support al die tijd.

Lieve Rosa, Liselotte en Morris, wat is het fijn dat jullie er zijn. Jullie lach, enthousiasme en energie is elke dag weer genieten, wat ik hou ik van jullie!

Mijn lieve Maike, samen waren we mr & ms radio! Zonder jou had ik deze promotie nooit kunnen voltooien. Eén van de dingen die ik zo enorm koester is dat we elkaar altijd stimuleren om te groeien en te ontwikkelen. Elkaar de ruimte te geven om nieuwe dingen uit te proberen, en daar dan getuige van zijn. Deze promotie is daar een voorbeeld van, ik ben je dankbaar dat je naast me hebt gestaan en mij gesteund hebt in deze jaren. Het is veel geweest, maar samen kunnen we de wereld aan! Ik hou van je!
Simon Robert van Kranen werd geboren in 1975 in Amsterdam, zoon van Robert Herman van Kranen en Thea Maria Elisabeth van Kranen-Kalusche, broer van Christiaan. Hij volgde middelbaar onderwijs op het St.Ignatius gymnasium in Amsterdam, waar hij 1993 zijn diploma behaalde. Vervolgens ging hij Technische Natuurkunde studeren in Delft. Daar studeerde hij in 2001 af in de vakgroep Charged Particle Optics op onderzoek naar het gebruik van een grid-lens in een elektronen-lithografiesysteem. Hij werkte aansluitend voor de universiteit aan een aantal wetenschappelijke projecten in de elektronenlithografie en microscopie, waarbij voor het laatste project een periode in Manchester werd gewerkt met een commerciële partner, Shimadzu Research Laboratories. In 2005 begon hij als promovendus bij het Nederlands Kanker Instituut/Antoni van Leeuwenhoek (NKI/AvL) in Amsterdam op onderzoek naar de gevolgen van anatomische veranderingen bij bestraling van hoofd-hals kankerpatiënten. In 2010 startte hij in een postdoc positie op het NKI/AvL in een implementatiegroep die nieuwe beeldgestuurde en adaptieve technieken in de kliniek introduceerde. In 2014 stapte hij over naar de researchafdeling van het NKI/AvL waar hij verder onderzoek deed naar vervormbare registratie van 3D beelden en adaptieve methodes en tijd kreeg om zijn promotie af te ronden. Simon is sinds 1999 samen met Maike van Pelt en samen hebben zij drie kinderen, Rosa(9), Liselotte(6) en Morris (3).