Technische Universität München

Klinikum rechts der Isar

Klinik und Poliklinik für Radioonkologie und Strahlentherapie

Direktorin: Prof. Dr. Stephanie E. Combs

Planning of Radiotherapy on Cone-Beam Computerized Tomography –
Inter-Observer Contouring Variability and Hounsfield Unit Uncertainties

Christina Miriam Ertl

Vollständiger Abdruck der von der Fakultät für Medizin der

Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Medizin

genehmigten Dissertation.

Vorsitzender:　　　　　　Prof. Dr. Ernst J. Rummeny

Prüfer der Dissertation:　　　1.　Prof. Dr. Stephanie. E. Combs

　　　　　　　　　　　　　　2.　Priv.-Doz. Dr. Daniela Pfeiffer

Die Dissertation wurde am 25.01.2018 bei der Technischen Universität München eingereicht

und durch die Fakultät für Medizin am 02.01.2019 angenommen.

Planning of Radiotherapy on Cone-Beam Computerized Tomography – Inter-Observer Contouring Variability and Hounsfield Unit Uncertainties

TABLE OF CONTENT

List of Abbreviations

ANOVA	Analysis of variance
CBCT	Kilovolt cone-beam computed tomography
CI_{gen}	Generalized conformity index
COM	Center of mass
CT	Computed tomography
CTV	Clinical target volume
CV	Coefficient of variation
$D_{2\%}$	Near maximum dose received by 2% of the organ volume
D_{mean}	Mean dose covering a defined volume
D_{Tissue}	Dose calculated with heterogeneity correction
D_{Water}	Dose calculated without heterogeneity correction
DVH	Dose volume histogram
DVH-c	Cumulative dose volume histograms
DVH-d	Differential dose volume histogram
ED	Electron density
GTV	Gross tumor volume
H&N	Head and neck
HU	Hounsfield units
ICC	Intraclass correlation coefficient
IGRT	Image guided radiation therapy
IMRT	Intensity modulated radiation therapy
kv-PCT	Kilovolt planning CT
MRI	Magnetic resonance imaging
MU	Monitor units
OAR	Organ at risk
P	Patient in the prostate cancer group
P'	Patient in the head and neck cancer group
PC	Prostate cancer
PCT	Planning CT

PET	Positron emission tomography
PTV	Planning target volume
PG-l	Left parotid gland
PG-r	Right parotid gland
RO	Radiation oncologist
RT	Radiotherapy
SBRT	Stereotactic body radiotherapy
Sp. Cord	Spinal Cord
V_{is}	Intersection volume
V_{un}	Union volume
VMAT	Volumetric-modulated arc therapy
X	Ventro-dorsal direction
Y	Medio-lateral direction
Z	Cranio-caudal direction

1 Introduction

1.1 Overview

Radiotherapy (RT) has developed into an important and recognized pillar in cancer therapy. It has become indispensable in the multidisciplinary state-of-the-art treatment of malignant as well as some non-malignant diseases. The value of RT lies not only in adjuvant treatment or in the palliative care, but it is highly important in the primary treatment of cancer aiming at complete remission and healing from the disease. RT can be conducted in different settings – as pre-surgery RT with the goal of downsizing the tumor to achieve operability and possibly enabling the surgeon to minimize operation margins, as post-operative RT as consolidation therapy and as primary RT as only treatment for the patient. All treatment schemes can be combined with simultaneous chemotherapy. Simultaneous application of chemotherapy can sensitize the tumor cells further to the impact of RT and increase the response rate. Various studies have been carried out to support the feasibility of primary RT with or without simultaneous systemic treatment with equivalent outcome compared to surgery for different tumors, such as prostate cancer (PC) and some types of head and neck (H&N) cancer (Monnier & Simon, 2015; Morgan *et al.*, 2009; Parikh & Sher, 2012; Timme *et al.*, 2015). An example for the use of pre-operative chemo-RT is the treatment of resectable esophageal cancer. Van Hagen *et. al.* demonstrated the superiority of neoadjuvant chemo-RT with carboplatin and paclitaxel over surgery alone (van Hagen *et al.*, 2012). The most common example for the use of post-operative RT, that raises a lot of public awareness, is RT after breast-conserving surgery. Before RT was established, treatment of breast cancer usually implied complete mastectomy, resulting in poor aesthetic outcomes and low psychological acceptance and coping by the patient. With the use of postoperative RT, breast-conserving therapy has become possible (Early Breast Cancer Trialists' Collaborative Group, 2011). Even intra-operative RT is now in routine clinical use, making cancer treatment for patients more convenient, less time consuming and less psychologically stressful (Vaidya *et al.*, 2014). This example emphasizes the importance of the multidisciplinary setting and the constant development and reevaluation of cancer treatment. However, it is only due to recent developments that RT has earned its significant position in cancer treatment, making it safer and more effective at the same time (De Los Santos *et al.*, 2013; Mayyas *et al.*, 2013; Tol *et al.*, 2016; Yu *et al.*, 2016).

1.2   Current State of the Art Radiotherapy

In the last few years, radiation oncology has experienced rapid changes. A lot of new research has been conducted in order to improve setting, safety and outcome of RT. RT has evolved in every aspect - from treatment planning to patient setup and irradiation techniques. Devices for patient immobilization, such as vacuum bags, foils and high precision masks have not only made irradiation of patients safer, but have also made high precision techniques like intensity modulated radiation therapy (IMRT) and stereotactic body radiotherapy (SBRT) techniques possible, where high doses are delivered to a small target volume with maximal protection of the adjacent organs at risk (OAR). Non-invasive methods to immobilize patients like stereotactic masks and vacuum bags minimize patient setup errors and replace more invasive techniques of immobilization like the stereotactic head ring.

IMRT and volumetric-modulated arc therapy (VMAT) have become a standard in RT (Tol *et al.*, 2016; Yu *et al.*, 2016). In IMRT, a multi-leaf collimator is used during RT. By blocking parts of the irradiation beam with a multi-leaf collimator, the beam profile is formed according to the shape of the tumor. This way, the surrounding tissue can be spared even more effectively. In VMAT, the rotation of the gantry is combined with IMRT technique. During the rotation of the gantry around the patient, IMRT is delivered to the tumor from different angles. The collimator leaves are moving at the same time as the gantry is moving around the patient in order to provide the desired homogenous 3D-dose distribution across the tumor. This way the irradiation field can be matched as good as possible even to irregularly shaped tumors, minimizing irradiation of the surrounding tissue (Novaes *et al.*, 2015), however at the cost of a larger low dose area.

One of the most important advances, however, is the introduction of image guided radiotherapy (IGRT) (De Los Santos *et al.*, 2013; Mayyas *et al.*, 2013). To control possible positioning errors or anatomy changes, IGRT is essential. Patient setup errors can be detected and eliminated at the same time. The anatomy of a patient can change in the course of treatment. Examples are tumor shrinkage or swelling of organs because of edema or progress under RT or weight loss. Similarly, the daily physiological change of bowel or urinary bladder filling has an impact on organ movements. These changes can cause adjacent tissue to slide into the irradiation field or tumor cells to escape the full prescribed dose. All this might result in higher toxicity, as well as in local failure. Therefore, these setup controls are essential for treatment safety and success. For this purpose, kilovolt cone-beam CT-scans (CBCTs) are performed on a regular basis. The required technique is integrated into the linear

accelerator. Imaging is performed with the patient in the irradiation position. It is possible to compare the daily CBCT image showing the actual situation to the original kilovolt planning CT (kV-PCT = PCT). Positioning adjustments and treatment table shifts can be performed immediately, which makes irradiation more accurate. Other influence factors, such as difficulties of the patients to hold up a certain bladder filling, become obvious and it can immediately be responded to them adequately. The fusion of PCT and CBCT offers daily reevaluation of the target volume and organ positioning and can help with the decision whether a new PCT with a new irradiation plan is necessary.

The establishment of IGRT also cleared the way for the latest development of tracking and gating systems. The terms "tracking" and "gating" describe a variety of methods with the purpose of adapting RT to moving tumors (Giraud *et al.*, 2006), as it is the case with lung cancer or breast cancer. Tumors of the lung, for example, can be treated using gating techniques. RT is only applied in a certain breathing phase by tracking periodic breathing motions using a marker on the patient and gating the beam to a certain breathing phase. Using the same system for breast cancer reduces the dose to lung and heart (Korreman *et al.*, 2005; Zurl *et al.*, 2010). The tracking system also enables real time adjustments of the beam to the tumor position. The tumor can be tracked in different ways, for example by implanting fiducial markers. Therefore, tumor movement does not have to result in large safety margins any more.

1.3 Treatment Planning and Delivery:

1.3.1 Treatment Preparation

1.3.1.1 Setup

The planning procedure of RT starts with the acquisition of the PCT. The patient is positioned on the CT table the same way he will be positioned during every single treatment session. During the setup, the patient will also receive the skin marks that will ensure the positioning verification via laser systems in the CT room as well as in the treatment room. These marks must be maintained during the whole treatment time. Exemplary, setups and planning for head and neck (H&N) and prostate cancer (PC) are described here. Usually, H&N cancer patients get an immobilization mask to decrease patient movement. A radiation protection rail offers

shielding of the teeth. Knee support or pillows can be installed individually. PC patients receive an inflatable rectal balloon to decrease irradiation of the dorsal rectal wall (Geier *et al.*, 2012). The belly board can help to bring the intestines further away from the target volume, when the patient is lying on the stomach. Here, too, the knee support or extra padding are optional.

1.3.1.2 *Contouring*

After obtaining the PCT, a radiation oncologist (RO) starts to delineate the OARs and defines the gross tumor volume (GTV), the clinical target volume (CTV) and the planning target volume (PTV). The GTV represents the macroscopic, visible tumor, while the CTV includes the area where microscopic spread occurs most likely. The PTV equals the CTV plus an added safety margin to account for setup errors, tissue changes and delineation uncertainties.

Definition of the target volume is not trivial. It forms the basis for further planning and for patients' tumor control and safety. Therefore it has to be as accurate as possible. The reliability of contouring, however, is subject to many influence factors. First of all, delineation depends greatly on the experience of the RO (Dunst, 2016). Due to the often poor soft-tissue contrast in CTs, target definition is often easier on magnetic resonance imaging (MRI) or positron emission tomography (PET) (Wang & Zhe, 2013). PET shows the actual active tumor location by visualizing its metabolism. Depending on the tumor quality, different imaging methods could have advantages in target delineation. These images can be co-registered to the PCT and help with target volume definition. Despite a variety of guidelines for contouring of target volumes and the assistance of other imaging modalities like MRI or PET, a certain contouring uncertainty remains. Everybody is judging images in a slightly different way, according to their subjective views and own experience. This circumstance is called inter-observer contouring variability. Brouwer *et.al.*, for example evaluated inter-observer contouring variability in the head and neck region for six patients (Brouwer *et al.*, 2012). Five ROs delineated the spinal cord, the parotid and submandibular glands, the thyroid cartilage and the glottic larynx. For all OARs, except the glottic larynx, inter-observer contouring variability, assessed via intraclass correlation coefficient, concordance index (=CI) and a 3D measure of variation, was moderate. The glottic larynx showed larger variation. The inter-observer contouring variability in the larynx was largest in the cranial, the caudal and the medial parts. The experience and image quality certainly influences inter-observer contouring variability. Tissue contours can be obscured resulting for example in underestimation of the organ volumes. This is often due to movements, such as breathing motions (Padmanaban *et*

al., 2010). Artifacts deriving from implants and prostheses can block out whole structures in their surroundings. Moreover, there is also influence of intra-observer contouring variability. Intra-observer contouring variability describes the fact that even the same person contouring the same structure on different time points will result in varying contours. Evaluation of inter- and intra- observer variability has become a standard to assess reliability of contouring and image quality.

Fig. 1 and Fig. 2 depict two examples of inter-observer contouring variability on cone-beam CT (CBCT) for the left parotid gland and the prostate.
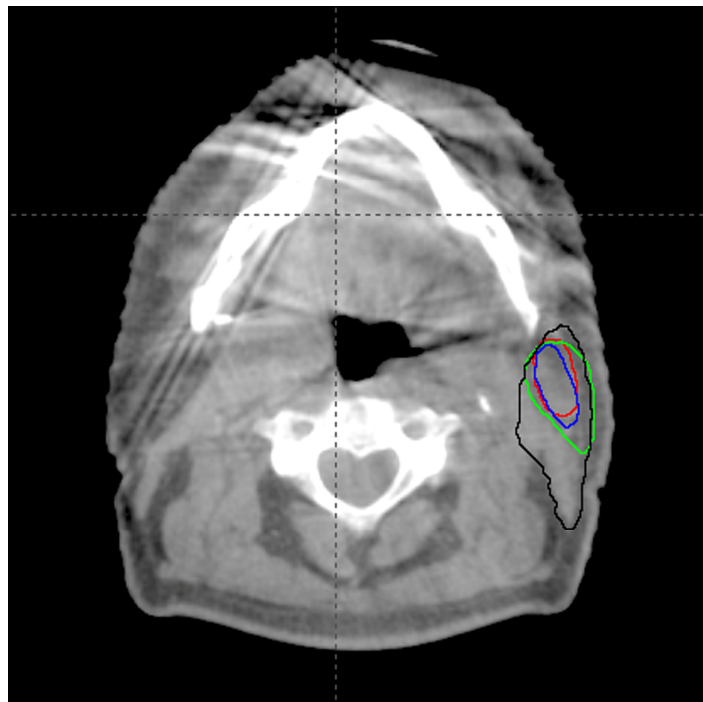

Fig. 1 Example of inter-observer contouring variability of the left parotid gland on CBCT.
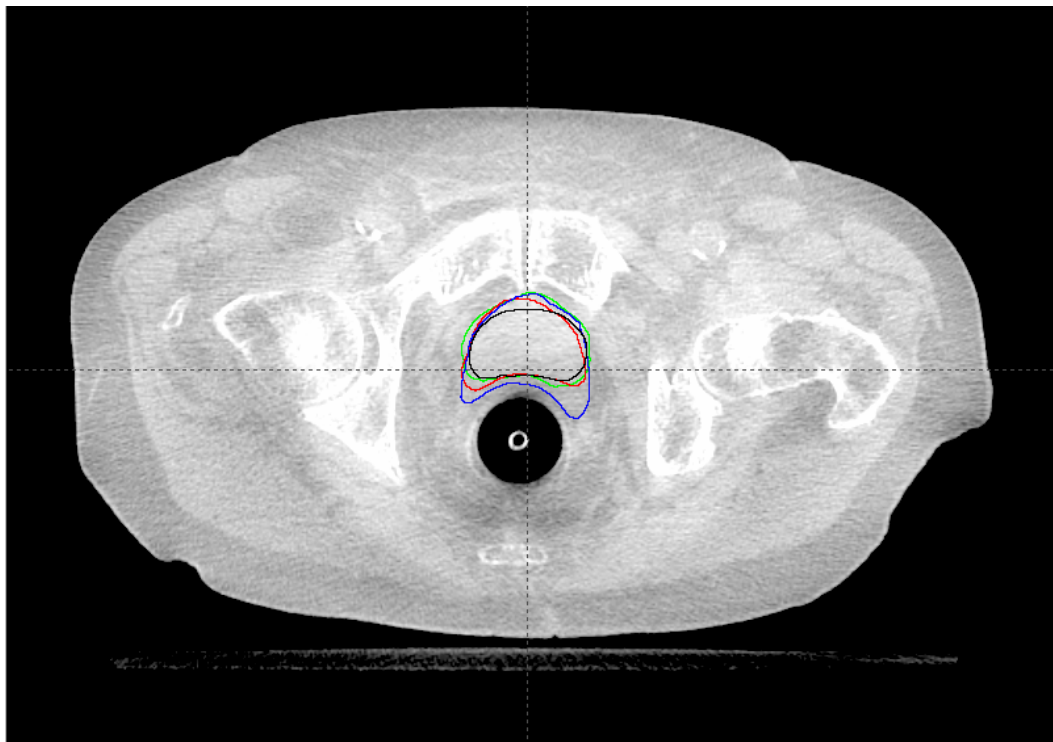
Fig. 2 Example of inter-observer contouring variability of the prostate on CBCT.

1.3.1.3 Dose Calculation

When all volumes are defined, an irradiation dose plan is calculated by a medical physicist. Feasibility of RT is always determined by the balance between tumor control and acceptable toxicity to adjacent organs. Side effects on normal tissues and OARs are limiting dose application. Side effects that must be expected depend mainly on the location of the tumor, the characteristics of the surrounding organs, their radiosensitivity and on the applied dose. Each tissue has its own degree of radiosensitivity. The consequential dose restrictions must be respected in any case. Application of dose to the surrounding tissue must be low to keep toxicity to a minimum. In the immediate proximity of the target volume, steep dose gradients are needed to ensure full coverage of the tumor and best possible protection of the surroundings at the same time. The irradiation plan has to be optimized so the tumor receives the full prescribed dose and the OARs are spared as consequently as possible. As mentioned above, there are empirical critical dose limits for each organ that have to be maintained and that help the physician and the physicist to decide whether a plan is acceptable and can be used for treatment. CT-based dose-calculation algorithms for accurate radiation treatment planning rely on correct assignment of Hounsfield units (HU) to the electron density (ED). HU are values on a scale which represent the degree of X-ray attenuation in different tissues. Per definition, pure water has an HU value of 0, air of -1000 and bone of +1000. Different X-

ray attenuation coefficients of tissues represented by different HU manifest themselves on CTs by different grey shades. Phantoms with known electron densities are used to generate HU to ED calibration curves that are used for dose calculation.

1.3.2 Treatment Delivery – CBCT

In the beginning of each treatment session, the patient has to settle down on the treatment couch in the fixed position as indicated by the skin marks, verified using the laser system. The mask or the rectal balloon as well as all other setup aids such as belly board or knee support have to be installed the same way as for the PCT. Before the actual irradiation can be started, a CBCT is acquired to verify the patient position. During the following treatment sessions, CBCTs are conducted in intervals or in certain critical cases before every treatment session for verification of the target volume as part of IGRT. CBCTs are useful to verify patient setup and organ positions. Moreover, patient anatomy changes or changes in organ filling can be verified. These changes can make replanning necessary. CBCTs help with the decision whether the treatment accuracy is still sustained or whether adjustments and replanning have become necessary.

1.4 Future Aspects: Online Adaptive Planning

Up to now there are no well established algorithms for planning on other imaging modes than the PCT. CT imaging is still the gold standard for planning irradiation. Replanning on CBCTs would be beneficial, as they are regularly conducted in the course of RT. They are therefore easily and routinely available. Currently, CBCTs are used for assessing setup errors and soft tissue changes, but not for planning. If a major change of anatomy is detected, or if the patient cannot follow instructions for example due to the inability to uphold a certain bladder filling because of irradiation toxicity causing urge incontinence, the patient has to endure the whole process of treatment planning again. Planning starts again with a new PCT. If major deviations from the initial plan become obvious on the control CBCT, it would be of great advantage if the CBCT could be used for online plan readjustments. Avoiding the necessity to acquire another PCT would significantly reduce replanning time. The treatment of the already correctly positioned patient could be continued without time loss, being advantageous for treatment goals and the patient's psychology.

However, there are some negative characteristics of CBCTs that have to be taken into account. The width and the shape of the cone beam cause more scattering, especially at the edges (Choi *et al.*, 2011; Eskandarloo *et al.*, 2012; White *et al.*, 2009). CBCTs therefore

contain more artifacts and the image quality is poorer compared to the PCT. Contouring on CBCTs might be more difficult and inter-observer contouring variability larger. Delineation of target volumes and OARs, however, is the fundament of RT. Reliability in identification of size, shape and position is essential to be able to calculate reliable dose plans. Additionally, HU resulting from CBCTs are less reliable than those from PCTs (Yoo & Yin, 2006) and cannot be determined as exactly on CBCTs as on PCTs due to different reasons, for example due to increased scatter. HU to electron density curves are also not as reliable (Eskandarloo *et al.*, 2012; Hatton *et al.*, 2009). There are several studies with the goal to find influence factors on dose calculation on CBCT and possible ways to deal with them. These approaches include HU mapping or the calibration of HU to electron density (ED) curves (Guan & Dong, 2009; Hatton *et al.*, 2009). In the process of HU mapping, empirical or population based HU of CTs are assigned to different kinds of tissue on other imaging modalities, such as CBCTs, but also MRI. These are then used for dose calculations with the goal of making irradiation planning on these other imaging modalities possible (Dunlop *et al.*, 2015; Fotina, Hopfgartner*, et al.*, 2012).

In this study we summarize all influence factors on HU-derivation on CBCTs, which are due to various technical settings and reconstruction problems, as "HU uncertainties".

1.5   Prostate Cancer and Head and Neck Cancer

PC as well as H&N cancer are two prevalent cancer types that frequently pose indications for RT. RT can be given as an adjuvant or additive post-operative therapy, but also oftentimes as primary treatment.

1.5.1   General Principles and Side Effects of Prostate Cancer Radiotherapy

For PC, primary RT with neoadjuvant hormone therapy is an equivalent alternative to surgery in early local or locally advanced stages without metastases (Parikh & Sher, 2012). The equality in outcome of the treatment modalities and their inherent advantages and disadvantages and risk profiles make a multidisciplinary setting even more inevitable to ensure that patients can make well-informed decisions (Guy *et al.*, 2016). The choice of the appropriate treatment is influenced by several factors, including tumor stage, the existence of distant metastases, technical operability, co-morbidities, age and life expectancy as well as the patient's personal preference (Scherr *et al.*, 2003). RT provides not only curative treatment, but can also be used in the palliative setting, for example treating symptomatic bone metastases (Scherr *et al.*, 2003).

Each body region poses its own particular challenges during irradiation planning and treatment. For the PC patients, it is mostly the intrinsic organ changes like varying filling of the hollow organs, such as bladder and bowels. These organs can move in and out of the irradiation field and also affect the position of surrounding tissue and the tumor. To avoid an increase of toxicity or local failure, organ filling must be controlled and kept constant as accurately as possible. Patients should be instructed to empty the rectum and fill the bladder before every treatment session. Irradiation aids like the rectal balloon can be applied and imaging controls should be conducted on a regular basis. Replanning might become necessary in the course of the treatment with the increase of toxicity, as the individual patient can have more difficulties following the instructions to fill the bladder and hold the urine.

 In the course of RT of PC, bladder and bowels are at the highest risk. Acute symptoms can be similar to symptoms of a urinary tract or enteric infection (Chennupati *et al.*, 2014; Zelefsky *et al.*, 2002). This includes alguria, pollakisuria, the sensation of urinary retention and diarrhea, meteorism and bleeding, respectively. Incontinence is a highly impairing side effect with low social acceptance. Late effects include shrinkage of the bladder, urinary tract obstructions by fibrosis as well as ileus and late rectal bleeding. Chronic infections as well as incontinence can become a persisting problem. Due to the close proximity of the responsible nerves, prostate irradiation can result in sexual dysfunction (Talcott *et al.*, 1998). All these possible side effects of RT illustrate the high importance of the protection of the OARs (Schaake *et al.*, 2016). Toxicity of prostate RT can lead to loss of quality of life (Schaake *et al.*, 2014). However, these side effects are mostly mild and can be further decreased by using modern techniques and strict dose limitations to OARs (Chennupati *et al.*, 2014).

1.5.2 General Principles and Side Effects of Head and Neck Cancer Radiotherapy

In early stages of H&N cancer, RT delivers equivalent results to surgery. It can be suggested as a possible alternative according to patients' preference or other factors such as inoperability of the tumor (Monnier & Simon, 2015). Inoperability can be due to technical challenges but also due to co-morbidities and age of the patient. The alternative is primary RT of the tumor with or without combination with simultaneous chemotherapy, such as cisplatin. During the process of decision-making, the expected functional outcomes and the preservation of the affected organ must be thoroughly considered (Karlsson *et al.*, 2014). Palliative RT for symptom control is also possible for advanced H&N cancer.

In the H&N region, changes in anatomy pose a common problem during RT. Weight loss during treatment, swelling due to edema and inflammation, as well as tumor shrinkage can result in a complete change of the local situation. Regular imaging is essential and replanning must be considered when significant changes occur. In the H&N region, symptoms of toxicity to surrounding tissue play a major role as they are especially impairing. In this region, there are many organs in a very enclosed area. Toxicity to these organs affects quality of life significantly (Kakoei *et al.*, 2012). This is where adaptive online planning will play an important role in the future to save time when adjustments of the current plan become necessary to avoid toxicity or local failure.

Side effects of RT in the H&N region are various. During external beam RT, the skin is always at risk. The acute effects on the skin can range from light irritation and redness to open moist and bleeding wounds that tend to get super-infected. Chronic effects include pigmentation disorders, dryness of skin, atrophy as well as fibrosis of the subcutaneous tissue (Bray *et al.*, 2016). Healing can be impaired after irradiation. This is especially important if an operation is planned after RT, but also oftentimes during adjuvant treatment and the operative treatment of relapse after RT. Skin care therefore is an important part of RT. Optimized skin care cannot prevent irradiation effects. However, it can help to attenuate acute toxicity and long-term effects. Late effects have to be distinguished from recurrent disease, if necessary by biopsy (Bray *et al.*, 2016).

There are a number of different organs and sensory organs in the H&N region in a rather narrow space. Effects on salivary glands, such as the parotid or submandibular glands, can cause acute and late xerostomia (Jensen *et al.*, 2010). Xerostomia also aggravates side effects like swallowing impairment and predestines for oral infections, e.g. with candida, which has large impact on the quality of life as well (Jensen *et al.*, 2010). Saliva with its antibodies is a major part of the local immune defense. Damage of the teeth by direct irradiation effects on the dentin and enamel, but also secondary damage by caries can be a likely consequence (Lieshout & Bots, 2014). Caries is favored by xerostomia and the irradiation-caused teeth damage. Loss of taste often leads to loss of appetite while swallowing has already become difficult due to irradiation mucositis with or without super-infection. Dysphagia can also be caused by direct damage of the swallowing apparatus (Agarwal *et al.*, 2011). All together, this can result in severe weight loss. Parenteral nutrition or nutrition through a percutaneous endoscopic gastrostomy can become inevitable. Great care must be taken of the spinal cord. Acute effects can be edema that can usually be treated well, e.g. with dexamethason. Chronic irradiation damages can lead to myelopathy (Kirkpatrick *et al.*, 2010). Damage to the spinal

cord can range from merely radiographic changes to serious functional impairment. There is a possibility that these symptoms improve over time after RT (Kirkpatrick *et al.*, 2010). Dose limits for OARs should be maintained strictly. Accurate irradiation planning and the use of modern techniques are the only way to minimize toxicity and make RT a valuable pillar of cancer treatment.

1.6 Goals of the Present Study

In summary, to maintain a high standard in RT concerning success of the treatment and security of the patient with low toxicity, high precision and flexibility in plan adjustments are crucial. Patients should benefit from all resources and modern techniques, such as IMRT and IGRT described above. The basis of high quality RT and the most important criteria of fulfilling all its requirements are precise delineation of target and risk volumes and reliable dose calculations. To date, treatment planning on PCTs remains the gold standard. Prompt response to anatomy changes or organ filling is required. Up to date, these readjustments are conducted offline on the basis of a new PCT. To improve the work flow and the convenience of the patient, it would be helpful to use CBCTs directly for replanning. However, the influence of inter-observer contouring variability and HU-based dose calculations on CBCTs is not yet well investigated. Particularly with regard to efforts to use CBCT for irradiation planning, inter-observer contouring variability on PCTs and CBCTs, its impact on results of dose calculation, as well as the reliability of dose calculation on CBCTs must be assessed, compared and put into relation to other influence factors.

The goal of the present study was to assess inter-observer contouring variability and dose calculation on PCT in comparison to CBCT for PC patients and H&N cancer patients for selected OARs in these areas. The impact of HU on dose calculation was assessed.

It was investigated whether inter-observer contouring variability or HU have the greater impact on planning outcome.

2 Materials and Methods

Five PC and five H&N cancer patients undergoing IGRT were evaluated. The PCT and two corresponding CBCTs, one obtained at the beginning (CBCT1) of RT, between one and two weeks after the PCT, and one at the end (CBCT2), about 1.5 to 2 months after obtaining the PCT were assessed. Four experienced radiation oncologists (RO1, RO2, RO3 and RO4) were asked to delineate the OARs, namely bladder, rectum and prostate for PC and spinal cord, left parotid gland and right parotid gland for H&N on each CT – the PCT, CBCT1 and CBCT2. The different CT-types with the different contouring formed the basis for the evaluation of inter-observer contouring variability and the assessment of HU and their influence on dose calculation. It was hypothesized that it might be more difficult to contour on CBCTs that were obtained at the end of treatment, due to soft tissue changes, such as edema, developing during RT. This can worsen the subjective image quality and blur the edges of the OARs. For each patient, an irradiation plan was calculated on the PCT by a medical physicist. Subsequently, the initial plan of the PCT of a patient with the original parameters was copied to the two corresponding CBCTs and then the dose distribution was recalculated for the structures as delineated on each CBCT.

2.1 Inter-Observer Contouring Variability

The extent of inter-observer contouring variability is a measure of accuracy of delineation on CTs. The larger the inter-observer contouring variability, the more uncertain is the definition of target volumes or regions at risk. However, all further planning is dependent on confident delineation, especially homogenous dose calculations delivering the prescribed dose to the tumor while protecting surrounding tissue. As mentioned in the introduction, there are different reasons for high insecurity of delineated contours. One cause of error is the inexperienced radiation oncologist (Dunst, 2016). To investigate the influence of human error, the delineations of the four ROs were used to assess the reliability of contouring of target volumes and OARs.

The other reason for high inter-observer variability in contouring is an impaired image quality. The image quality of CBCTs is worse than the image quality of planning or diagnostic CTs due to increased scatter and technical reconstruction problems, that are not subject of this study (Siewerdsen & Jaffray, 2001).

Another influence on inter-observer contouring variability is the alteration of the tissue in the course of irradiation. As the tissue reacts to the photons, inflammation is caused and edema

and swelling develop. These acute tissue reactions might make it more difficult to define the true outlines of a structure and differentiate it from its surroundings. This will be investigated by comparing CBCT1 to CBCT2.

It is not possible to compare delineation of OARs between the three CTs directly. Organ position and size can change significantly in between the treatment fractions. The before-mentioned bladder and bowel filling can vary greatly not only depending on patients' compliance towards the instructions. Also, tumor shrinkage or tissue swelling can change the positioning of all OARs. For this reason, inter-observer contouring variability was calculated for each CT separately and then compared to each other. To get more reliable results for the spinal cord, the distal ends of the contouring were cut to end at the same CT layer, as there have been no guide lines of how far to the distal end the delineation should be performed. The proximal parts were not cut, as the different delineations of the transition into the medulla oblongata were considered an important aspect.

The parameters used to describe inter-observer contouring variability are various and not standardized. On the basis of a meta-analysis, Fotina et al. assessed a number of evaluation parameters used in the literature for the description of inter-observer contouring variability and recommended a combination of overlap measure, descriptive statistics and statistical measures of agreement or reliability analysis (Fotina *et al.*, 2012).

Considering this analysis and to allow for comparison with other studies, three evaluation parameters were chosen for characterization of inter-observer contouring variability of the structure volume size and position as deduced from the observers' delineations: generalized conformity index (CI_{gen}) as proposed by Kouwenhoven et al. for any number of observers (Kouwenhoven *et al.*, 2009), coefficient of variation (CV) and intraclass correlation coefficient (ICC) (Shrout & Fleiss, 1979). These evaluation parameters make up a reliable description of inter-observer contouring variability, as they describe different aspects of conformity. It should be noted that CV and ICC only assess conformity in the volume size determination and provide no information about positioning reliability. This is not sufficient, of course, for assessment of the reliability of irradiation plans. Only the CI_{gen} as defined below gives information about both, concordance in volume size and position.

Additional useful information about the 3D-positioning of the different OARs in the body is provided by the calculation of the geometric center and the distances between the geometric center of each individual OAR and the mean geometric center as averaged over all four

delineations of that OAR. In the case of organ delineation, where tissue density is assumed to be homogenous, the center of gravity equals the center of mass (COM). For compatibility with the literature, in the following, the geometric center will be referred to as COM and mean COM (COM_{mean}).

(1) Generalized conformity index: CI_{gen}

A generalized conformity index for any number of delineations, unbiased with respect to their number, was developed by Kouwenhoven et al. (Kouwenhoven *et al.*, 2009):

$$CI_{gen} = \frac{\sum_{i=1}^{k} i(i-1) V_i'}{2(k-1) \sum_{i=1}^{k} i V_i' - \sum_{i=1}^{k} V_i'}$$

where V_i' is part of the intersection of i different volumes, forming $i(i-1)/2$ pairs delineated by a total of k ROs.

This expression can be simplified to (Kouwenhoven *et al.*, 2009)

$$CI_{gen} = \frac{\sum_{pairs} |V_i \cap V_j|}{\sum_{pairs} |V_i \cup V_j|}$$

CI_{gen} is the ratio between the sum of all pairwise intersections and the sum of all pair wise unions of the analyzed structures V_i delineated by the different ROs, and provides a measure of their overlapping. A CI_{gen} of 1 indicates complete overlapping of the delineations, whereas a CI_{gen} of 0 indicates that there is no intersection of the volumes at all. The magnitude of overlapping is determined by size differences of the different delineations as well as their relative position to each other. Fig. 3 demonstrates for the simple case of two observers that the same value of the conformity index can result from better conformity of position compared to conformity of volume, and vice versa. In order to describe the position of the structures towards each other more specifically than the more general information provided by the CI_{gen}, the COM is an informative additional quantity.
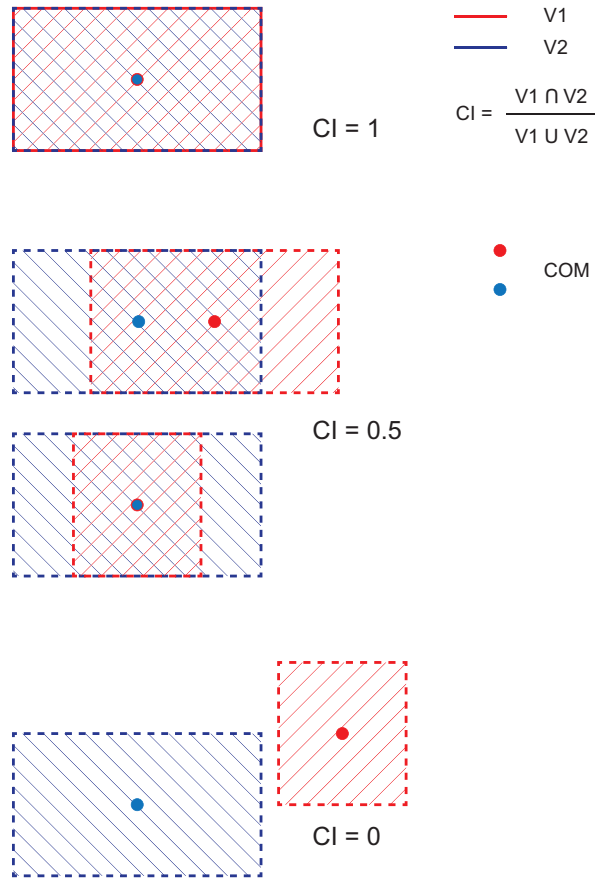
Fig. 3 Illustration of conformity index (CI_{gen}) for two observers. The illustration shows that CI_{gen} evaluates both size and position conformities of the structures. More precise information about the spatial positions and their relationship to each other is provided by evaluation of the *COM* coordinates.

(2) Coefficient of variation: *CV*

CV, a common quantity in descriptive statistics, represents the ratio of the standard deviation (σ) and the mean of a set of data, in the present case of the different volumes of the same organ as delineated by the different ROs.

$$CV = \frac{\sigma}{V_{mean}}$$

where σ is the standard deviation of the volumes as determined by the ROs for one patient and V_{mean} is the mean volume.

As a relative dispersion measure this dimensionless quantity allows for a conclusive comparison of variations of volumes relative to their size, independent of different scales. A smaller value of the *CV* indicates a smaller inter-observer variability in volume size determination.

(3) Intra-class correlation coefficient: *ICC*

The *ICC* completes this description by quantifying the consistency or conformity of studies of multiple raters related to various objects (Shrout & Fleiss, 1979).

$$ICC(2,1) = \frac{MS_{pat} - MS_{err}}{MS_{pat} + (k-1) \times MS_{err} + \dfrac{k \times (MS_{RO} - MS_{err})}{n}}$$

where MS_{pat} is the mean square between patient cases and MS_{err} the mean square of noise term. MS_{RO} is the mean square between raters (ROs) while k is the number of raters (RO1-4) and n is the number of patients. The first parameter 2 in *ICC(2,1)* means all k raters are chosen at random from a population of raters and these k raters rate all n subjects, the second parameter 1 stands for calculation of reliability of a single rater.

Calculation of *ICC* is based on two factor analysis of variances (ANOVA), comparing the variance between different ratings on the same measuring quantity with the total variance generated by all ratings and measuring objects. Therefore, a measurement is more reliable, when the ratings on the same object by different raters are similar, while the ratings on different objects by the same rater are differing. The *ICC* provides a tool for reliability judgement of observations made by different raters. The *ICC* in this study only concerns the reliability of volume size determination. It provides no information about positioning reliability.

According to a scale proposed by Landis and Koch (Landis & Koch, 1977), *ICC* > 0.8 suggests excellent reliability, *ICC* > 0.6 substantial reliability, *ICC* > 0.4 suggests moderate reliability and an *ICC* < 0.4 stands for poor reliability.

(4) Centre of mass: *COM*

Information about the degree of agreement between the ROs on the position of an OAR can be judged by comparison of their computed *COM*.

Combination of all these parameters, *CI_gen*, *CV*, *ICC* and *COM*, enables a reliable assessment of accuracy of volume as well as position of delineated OARs. *CI_gen* and *COM* deliver spatial information, while *CI_gen*, *ICC* and *CV* add information about the agreement in volume size. Comparison of *CI_gen* and *ICC* can identify the source of inter-observer contouring variability.

2.2 Dose Comparison

Results of dose calculation on CBCTs are affected by different factors. Two aspects were assessed in this study: the impact of the inter-observer contouring variability and the influence of HU uncertainties. The results were compared against each other.

The original plan on the PCT for each patient was normalized to the median prescribed dose and copied onto each CBCT. The dose distribution was recalculated for all structure sets defined by the different ROs. For all recalculations for the same patient, the identical number of monitor units (MU) was used. For dose calculations on CBCTs, two different CBCT-calibration curves – one for the H&N region and one for the pelvic region – were applied. Stability of calibration curves based on CBCT is vulnerable to different influences, for example to the patient constitution (Yoo & Yin, 2006). On the other hand, Yang *et al.* found that CBCT calibration curves were stable over time and doses computed based on PCT and CBCT agreed within 1% for phantom cases and prostate cancer patients (Yang *et al.*, 2007). Moving phantoms and lung patient cases resulted in larger dose discrepancies. By copying the original plan onto the new CT, the OARs may have a different position in relation to the beam, due to the discrepancies in organ positioning and sizes. For this reason doses on PCT and CBCTs cannot be compared directly to each other. As an acceptable alternative dose differences due to HU or inter-observer contouring variability were analyzed on PCT and each CBCT separately and then relatively compared between PCT and CBCTs.

The standard procedure of dose calculation for irradiation plans on PCTs takes the different HU of structures into account (D_{Tissue}, dose calculation with heterogeneity correction). To estimate the impact of HU on the results of dose calculation on PCT and CBCTs, dose distributions on CTs calculated with and without heterogeneity correction were evaluated. For dose calculation without heterogeneity correction all CT density values were equalized to water (HU=0). Since it is difficult to visualize and handle three-dimensional dose-distribution data in a feasible way, one-dimensional dose-volume histograms (DVH) are commonly used in radiation therapy planning in order to evaluate dose plans. Depending on the issue, either differential or cumulative DVHs are calculated. The dose on the horizontal axis as well as the volume on the vertical axis are displayed either in absolute values (Gy, cm^3), or in percent relative to the prescribed dose and the defined volume (%).

In the differential histogram (DVH-d) the volume of the organ that receives a dose within a defined dose bin (width of the column) is displayed. It is mainly used to illustrate minimum and maximum dose and dose variations within the structure.

In the cumulative histogram (DVH-c) the volume part of a structure that receives a given dose

or more is plotted. Therefore the graph starts at 100% volume for dose 0 and more, and ends at 0% volume for the individual maximum dose.

A drawback of both methods is that no spatial information is provided.

An optimal DVH-c for a target volume shows that the prescribed dose homogenously covers the whole defined target volume, if the embracing isodose is not defined differently. 100% of the prescribed dose should cover 100% of the defined target volume. At the edges of the PTV, there should be a very steep dose gradient towards the surrounding tissue and OARs. In the case of the OARs, the affected volume as well as the applied dose should be minimized. Since the OARs cannot be spared completely, this requirement is represented by a quick decline of the corresponding DVH-c graph at a low dose.

In daily routine DVH-c are more commonly used than DVH-d.

Calculated doses delivered to the different OARs on CTs with and without heterogeneity correction (D_{Tissue}, D_{Water}) were compared by determining their ratios (D_{Tissue}/D_{Water}). The mean doses (D_{mean}) and the near maximum dose received by 2% of the structure volume ($D_{2\%}$) were used for all calculations except for the spinal cord, where only $D_{2\%}$ was assessed. A ratio of 1 indicates that there is no difference in D_{mean} or $D_{2\%}$ between both calculation modes. This way, good conformities or major discrepancies in dose originating from HU can be identified.

To visualize the influence factors on the results of dose calculation and directly compare them, exemplary DVH for each structure set on PCT, CBCT1 and CBCT2, calculated with and without heterogeneity correction, were computed. This way, it can be illustrated whether the impact of HU or the impact of contouring-variability on dose calculation is larger - whether DVH vary more between D_{Tissue} and D_{Water} or between the different contouring of the ROs. As each structure was delineated by four ROs, and dose calculated twice (with and without heterogeneity correction), eight graphs are visualized in each diagram. DVH of the OARs can also be compared in order to see whether the results are consistent for all structures or if there are other possible influence factors that can result in larger DVH-divergence.

3 Results

3.1 Inter-Observer Contouring Variability

Results of the assessment of inter-observer contouring variability by direct volume size comparison are illustrated in Fig. 4.
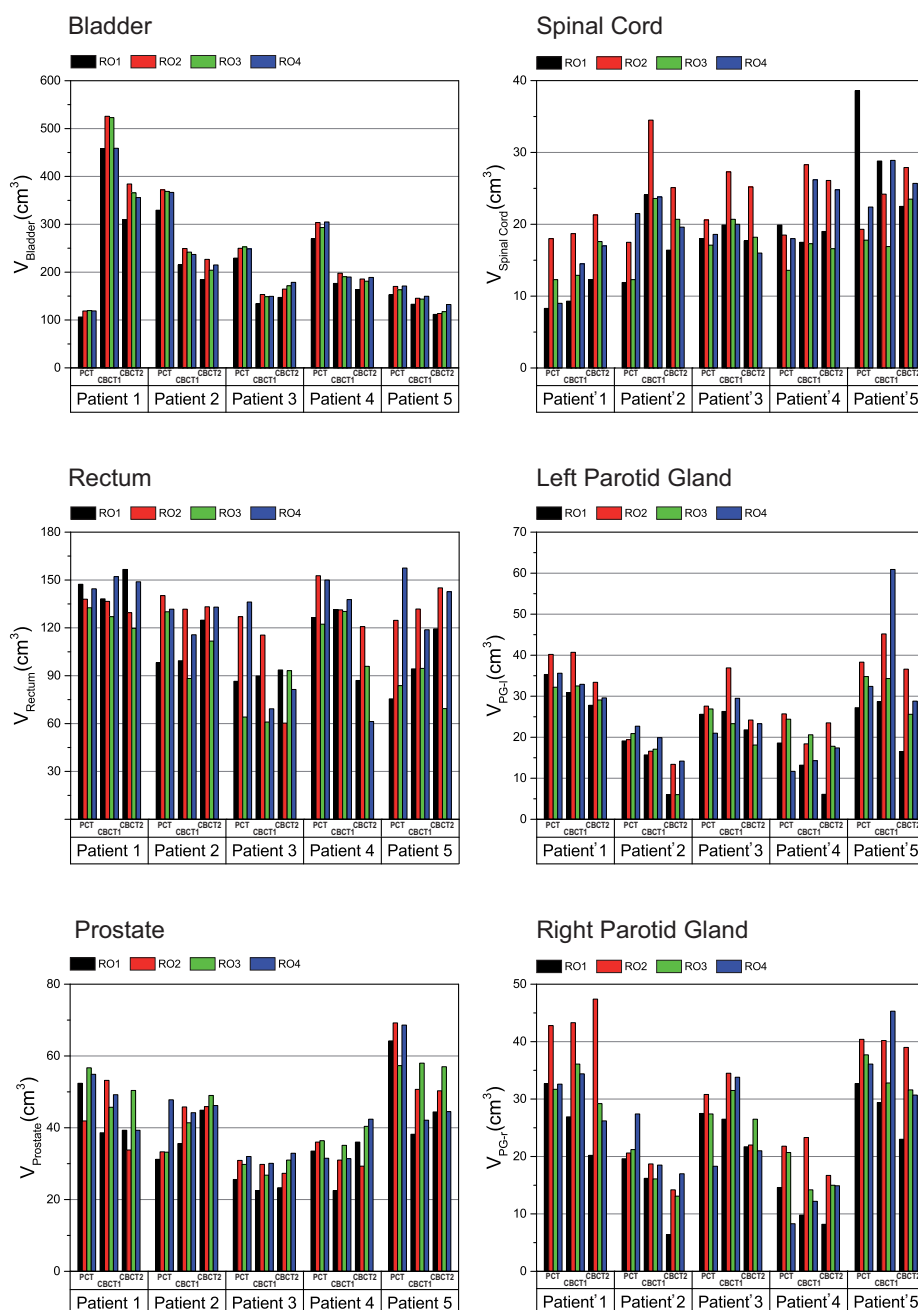

Fig. 4 Volumes of the organs at risks delineated by all radiation oncologists for each patient on each CT.

The bladder displays the best agreement between the four ROs on the PCT as well as on both CBCTs. The differences in bladder volume between the different CT-types are to be expected, as it is difficult to maintain a constant filling of the bladder over the course of RT.

Ratios of the smallest and largest delineated volume of a patient reach up to a factor of 2.20, 2.12 and 2.37 for the left parotid gland and up to 2.63, 2.38 and 2.66 for the right parotid gland for PCT, CBCT1 and CBCT2, respectively. Factors for the spinal cord reach values of 2.17, 2.01 and 1.73 for PCT, CBCT1 and CBCT2, respectively. For the bladder, these factors are 1.13, 1.16 and 1.24. For the rectum the values were 2.09, 1.49 and 1.39. The prostate displayed factors of 1.53, 1.56 and 1.49. The disagreement does not continually increase from PCT to CBCT2. Fig. 5 illustrates the accuracy of delineation when the individual contourings of the ROs are compared to the median. Most volumes differ between 20% to 40% from the median volume, a few outliers differ up to 85%, while the bladder again displays the best agreement. Spinal cord and the parotid glands display the largest deviation from the median.
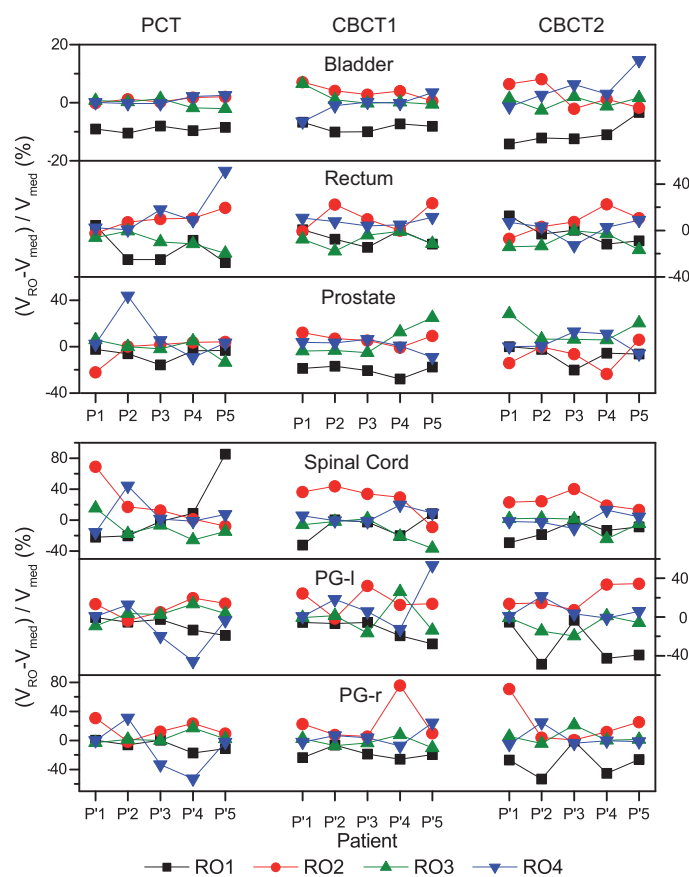

Fig. 5 Relative deviation of the volume sizes as delineated by the individual ROs from the median.

To better understand the impact of the differences in contouring, the intersecting volumes of all RO-determined volumes were compared to the overall union volumes (Table 1).

		PCT			CBCT1			CBCT2		
		V_{is}	V_{un}	V_{is}/V_{un}	V_{is}	V_{un}	V_{is}/V_{un}	V_{is}	V_{un}	V_{is}/V_{un}
		(cm^3)	(cm^3)	(%)	(cm^3)	(cm^3)	(%)	(cm^3)	(cm^3)	(%)
P1	Bladder	100.3	132.5	75.7	435.2	546.3	79.7	300.5	393.9	76.3
	Rectum	112.2	168.3	66.7	106.3	171.7	61.9	104.0	184.9	56.2
	Prostate	33.7	70.6	47.7	29.5	65.3	45.2	26.6	59.0	45.1
P2	Bladder	319.2	395.1	80.8	206.4	270.3	76.4	179.2	245.4	73.0
	Rectum	80.7	175.5	46.0	80.2	153.2	52.3	90.3	164.9	54.8
	Prostate	22.2	53.1	41.8	27.5	57.2	48.1	29.4	62.8	46.8
P3	Bladder	221.0	271.1	81.5	130.0	163.2	79.7	144.2	188.4	76.5
	Rectum	76.4	151.2	50.5	81.2	131.8	61.6	75.6	110.9	68.2
	Prostate	20.7	39.5	52.4	19.7	35.5	55.5	20.1	38.5	52.2
P4	Bladder	260.5	322.2	80.9	170.5	208.1	81.9	156.8	202.1	77.6
	Rectum	111.7	171.0	65.3	107.6	162.7	66.1	79.5	130.7	60.8
	Prostate	24.2	44.1	54.9	18.4	41.1	44.8	22.4	53.9	41.6
P5	Bladder	141.4	188.1	75.2	121.4	164.3	73.9	92.7	145.1	63.9
	Rectum	63.0	170.1	37.0	74.7	143.7	52.0	96.4	165.4	58.3
	Prostate	46.8	85.8	54.5	31.1	66.3	46.9	25.3	76.2	33.2

		PCT			CBCT1			CBCT2		
		V_{is}	V_{un}	V_{is}/V_{un}	V_{is}	V_{un}	V_{is}/V_{un}	V_{is}	V_{un}	V_{is}/V_{un}
		(cm^3)	(cm^3)	(%)	(cm^3)	(cm^3)	(%)	(cm^3)	(cm^3)	(%)
P1'	Sp. Cord	6.0	20.9	28.7	6.8	22.6	30.1	9.5	25.9	36.7
	PG-l	27.9	45.3	61.6	24.8	45.0	55.1	22.1	39.4	56.1
	PG-r	27.2	45.8	59.4	23.4	48.7	48.0	17.6	49.4	35.6
P2'	Sp. Cord	7.9	27.1	29.2	16.4	38.6	42.5	10.9	31.4	34.7
	PG-l	12.9	30.3	42.6	9.3	27.6	33.7	3.3	21.6	15.3
	PG-r	15.1	32.0	47.2	9.9	26.9	36.8	3.9	21.6	18.1
P3'	Sp. Cord	10.8	27.9	38.7	14.6	32.1	45.5	11.3	29.9	37.8
	-l	14.1	37.2	37.9	16.6	44.5	37.3	7.8	41.1	19.0
	PG-r	13.8	37.7	36.6	20.5	44.8	45.8	8.6	41.7	20.6
P4'	Sp. Cord	10.0	28.7	34.8	12.9	33.3	38.7	13.0	30.4	42.8
	PG-l	7.7	32.7	23.5	7.8	25.5	30.6	6.9	28.3	24.4
	PG-r	5.7	27.1	21.0	5.9	26.3	22.4	3.1	27.0	11.5
P5'	Sp. Cord	12.1	46.9	25.8	13.6	37.4	36.4	17.4	34.0	51.2
	PG-l	19.8	47.9	41.3	22.5	74.0	30.4	11.8	42.4	27.8
	PG-r	24.8	48.7	50.9	20.5	56.6	36.2	20.3	44.3	45.8

Table 1 Ratios of intersection volumes (V_{is}) and union volumes (V_{un}) of the organs as contoured by all ROs.

The common volume V_{is} equals the intersecting set of all four volumes defined by the ROs, while the total volume V_{un} equals their set union. The best agreement would be given, if the common volume equals the total volume. The agreement varies largely between good agreement of up to >80% (bladder) and poor agreement of down to 11% (right parotid gland).

The results of CI_{gen}, CV, ICC and the distance of the COM of the various OARs from the mean COM are presented in Table 2.

	PCT	CBCT1	CBCT2
Mean CI_{gen} [Range]			
Bladder	0.88 [0.86 - 0.89]	0.87 [0.85 - 0.89]	0.84 [0.78 - 0.87]
Rectum	0.70 [0.56 - 0.80]	0.74 [0.69 - 0.80]	0.75 [0.72 - 0.81]
Prostate	0.69 [0.62 - 0.72]	0.68 [0.65 - 0.73]	0.64 [0.55 - 0.70]
Spinal Cord	0.53 [0.45 - 0.60]	0.60 [0.53 - 0.65]	0.61 [0.57 - 0.69]
Left Parotid Gland	0.62 [0.50 - 0.76]	0.58 [0.51 - 0.71]	0.50 [0.38 - 0.72]
Right Parotid Gland	0.64 [0.47 - 0.75]	0.59 [0.47 - 0.67]	0.49 [0.35 - 0.65]
Mean CV [Range]			
Bladder	0.05 [0.04 - 0.06]	0.06 [0.05 - 0.08]	0.08 [0.06 - 0.09]
Rectum	0.17 [0.05 - 0.34]	0.11 [0.03 - 0.17]	0.11 [0.08 - 0.14]
Prostate	0.12 [0.07 - 0.21]	0.15 [0.11 - 0.19]	0.13 [0.04 - 0.17]
Spinal Cord	0.25 [0.08 - 0.39]	0.23 [0.16 - 0.28]	0.18 [0.10 - 0.22]
Left Parotid Gland	0.15 [0.08 - 0.32]	0.20 [0.10 - 0.34]	0.24 [0.08 - 0.34]
Right Parotid Gland	0.20 [0.09 - 0.38]	0.20 [0.08 - 0.40]	0.27 [0.11 - 0.38]
$ICC(2,1)$			
Bladder	0.982	0.982	0.958
Rectum	0.194	0.541	0.646
Prostate	0.875	0.703	0.689
Spinal Cord	0.268	0.452	0.346
Left Parotid Gland	0.742	0.672	0.686
Right Parotid Gland	0.746	0.779	0.622

Distance COM-COM_{mean} (cm)

	x	y	z	3d	x	Y	Z	3d	x	Y	z	3d
Bladder	0.02	0.03	0.02	0.04	0.03	0.05	0.09	0.13	0.02	0.07	0.05	0.10
Rectum	0.04	0.12	0.41	0.44	0.03	0.09	0.30	0.33	0.02	0.06	0.13	0.15
Prostate	0.03	0.08	0.14	0.19	0.03	0.08	0.13	0.18	0.04	0.12	0.19	0.25
Spinal Cord	0.05	0.13	0.83	0.85	0.03	0.09	0.47	0.49	0.05	0.06	0.41	0.43
Left Parotid Gland	0.13	0.06	0.12	0.21	0.11	0.06	0.27	0.33	0.14	0.10	0.20	0.29
Right Parotid Gland	0.11	0.07	0.11	0.20	0.10	0.10	0.18	0.27	0.12	0.17	0.19	0.32

Table 2 Description parameters of inter-observer contouring variability: generalized conformity index (CI_{gen}), coefficient of variation (CV), intra-class correlation coefficient ($ICC(2,1)$) and deviation of the center of mass (COM) from the mean center of mass (COM_{mean}). ICC was calculated using "Real Statistics Resource Pack for Excel 2007" (Zaiontz, 2014).

Mean CI_{gen} of the OARs for the five patients in the PC group all reach values >0.6, the bladder being the best with mean CI_{gen} >0.8. Mean CI_{gen} for the rectum is high, with a value of ≥0.7. Inter-observer contouring variability for the H&N region is larger. Mean CI_{gen} for the parotid glands and the spinal cord reached values <0.6, for the right parotid gland on CBCT2 even <0.5. Thus, the CI_{gen} are highest for bladder and lowest for the parotid glands.

Fig. 6 depicts the dispersion of the CI_{gen} averaged over all structures as well as for each single organ on the three different types of CTs for the PC and H&N group. The box plots show no trend towards less reliability on CBCTs. Agreement between the different ROs is best for the bladder. There is no trend indicating a decrease in agreement between the ROs on CBCT1 and CBCT2 compared to the PCT.
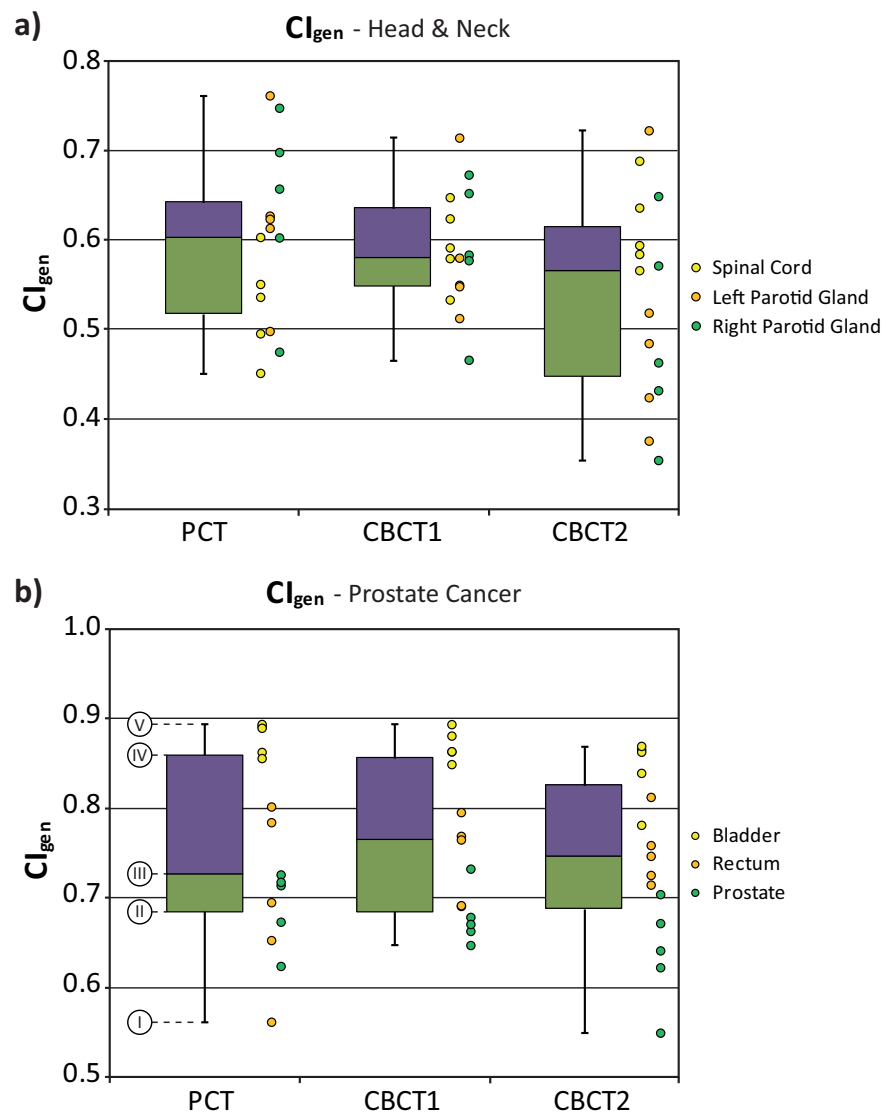
Fig. 6 Box plots depicting the dispersion of the generalized conformity index (CI_{gen}) for head and neck (a) and prostate cancer (b) patients. Here: I/V is minimal/maximal value, II/IV is lower/upper quartile and III is the median.

Fig. 7 shows CI_{gen} for each patient and each organ. CI_{gen} are not generally smaller on CBCT2 than on CBCT1 and PCT. The trend between PCT, CBCT1 and CBCT2 is rather unpredictable and variable for each patient. Only for the bladder and parotid glands, most patient CI_{gen} are smaller on CBCT2 than on PCT. However, COM and ICC do not support this finding.
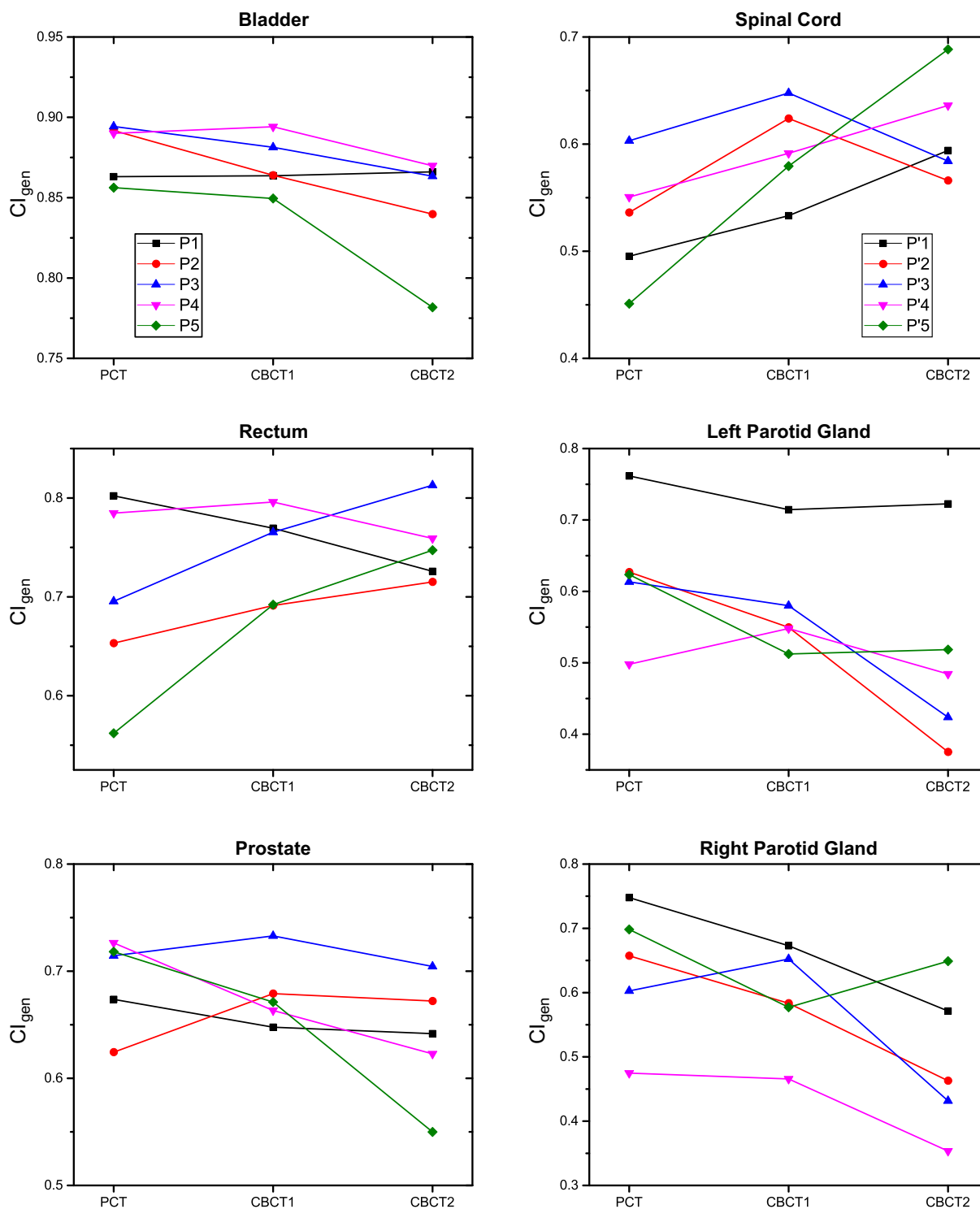
Fig. 7 CI_{gen} calculated for each OAR and for each patient separately. Lines only serve to guide the eye.

Direct volume comparison and CI_{gen} suggest larger contouring variability for the parotid glands. Similar results are displayed by the *CV*. *CV* of the parotid glands and spinal cord are considerably higher than for all other organs, while *CV* of the bladder is very small on all CTs. As explained above, *CV*s only consider volume sizes of OARs, not their position.

All calculated *ICC* reach values >0.6 on all types of CTs, except for the rectum with *ICC* values <0.2 on PCT and the spinal cord, where the PCT and CBCT2 display values <0.4. The best *ICC* values have been identified for the bladder, where the *ICC* is >0.9 on all three CTs. Like the CV, the *ICC* only interprets reliability of volume size delineation, not of positioning. The *ICC* should be interpreted cautiously, due to the small number of patients. This is visualized in Fig. 8 by the large confidence interval bars.
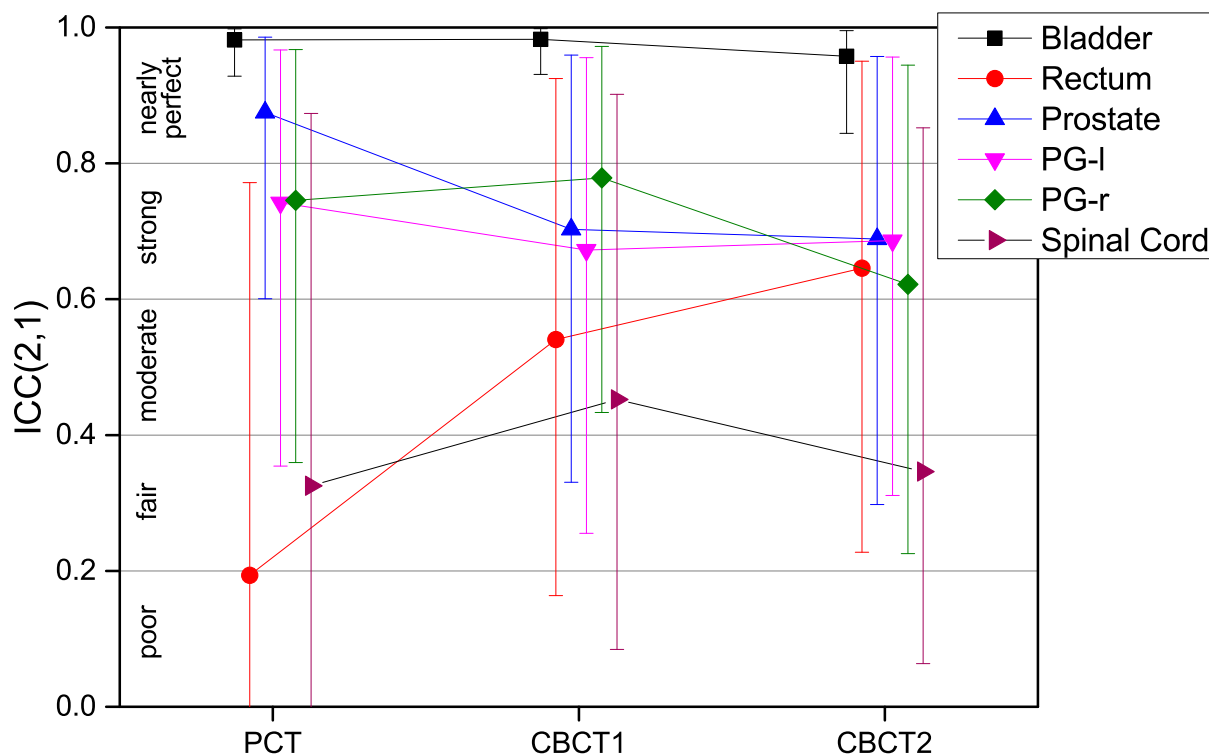
Fig. 8 *ICC* with confidence interval bars. *ICC* must be interpreted with caution due to the small patient number (the lines joining the data points only serve to guide the eye).

As stated above, in addition to the other three parameters, the *COM* was used for description of the three-dimensional positioning of the OARs in the body. To demonstrate the degree of agreement between the different observers, the distance of the individual *COM* of the various OARs from the mean *COM* is illustrated in Table 2. The mean *COM* is the average of all

COM for one OAR as delineated by all the ROs. It can be assumed as the position that is closest to the real center. The farther away each individual *COM* is from the mean *COM*, the farther away the position of the structure is from the true position. Again the distance values were averaged over all patients for each OAR. In order to identify the parts of the largest insecurities, the distances of *COM* to the mean were calculated for every direction separately: medio-lateral, ventro-dorsal and cranio-caudal (x-,y-,z-). In this case, a slightly larger contouring insecurity can be observed for the rectum, the spinal cord and the prostate in the cranio-caudal axis.

In accordance to the results of the other parameters, rectum, spinal cord as well as left and right parotid glands display larger deviations of the *COM* from the mean *COM* than bladder and prostate with values for PCT, CBCT1 and CBCT2 of 0.04 cm/0.13 cm/0.10 cm for the bladder, 0.19 cm/0.18 cm/0.25 cm for the prostate, 0.44 cm/0.33 cm/0.15 cm for the rectum, 0.21 cm/0.33 cm/0.29 cm for the left parotid gland, 0.20 cm/0.27 cm/0.32 cm for the right parotid gland and 0.85 cm/0.49 cm/0.43 cm for the spinal cord, respectively.

It is further noticeable that there is a systematic deviation between the contouring by the four ROs (Fig. 9). In most cases, RO1 is defining a smaller volume and RO2 a larger volume than the other ROs.
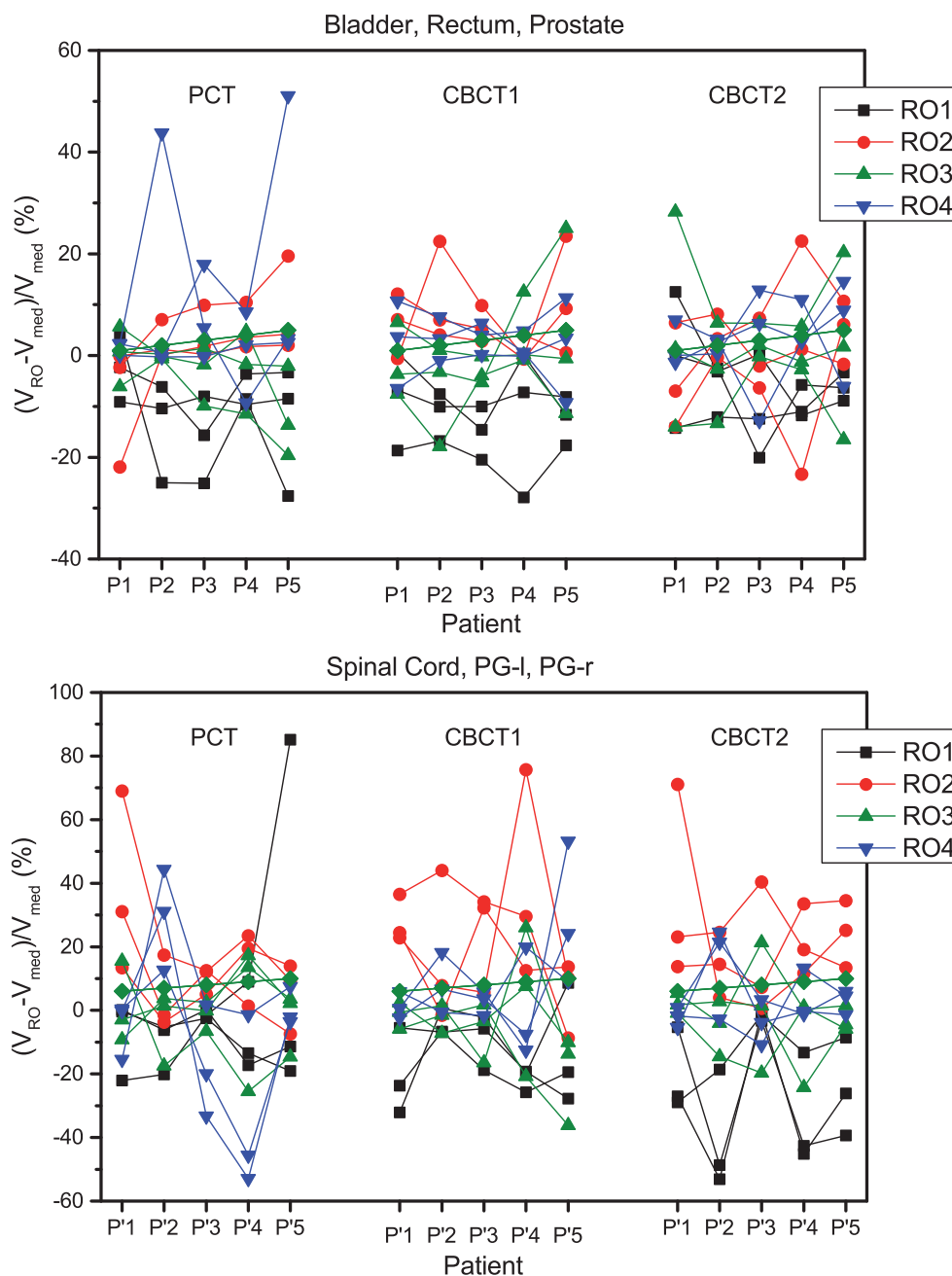
Fig. 9 Systematic contouring deviation and volume deviation from the median.

Fig. 10 shows how often each RO delineated the smallest, medium or largest volume. All in all, each RO contoured 90 structures. RO1 defined 58 times out of 90 the smallest volume, RO2 only 3 times. RO2 defined 51 times out of ninety the largest volumes, RO1 only 4 times. RO3 and RO4 lie in the middle field in the majority of cases.
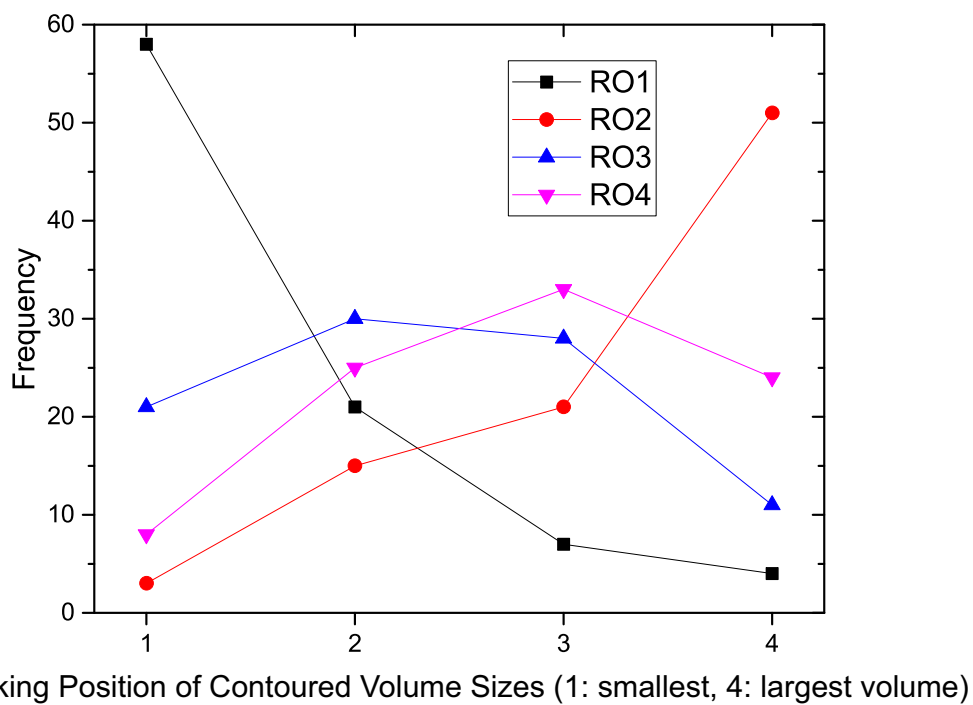
Fig. 10 Frequency of the size-ranking position of each RO's volume contouring accumulated over all organs delineated.
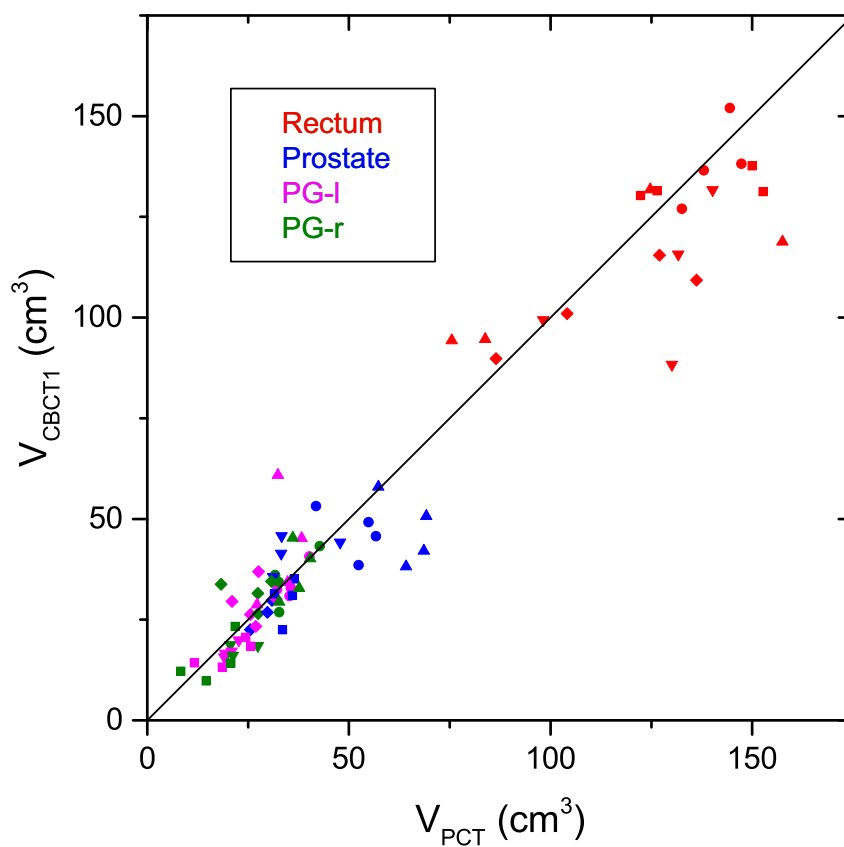
Fig. 11 Comparison of volume sizes delineated on PCT and CBCT1. Different symbols within a colour group represent different patients.

33

In Table 3 and in Fig. 11, volume sizes of the OARs as delineated by the same RO on PCT and CBCT1 are compared. CBCT1 was chosen for comparison to PCT because the time difference to PCT is smaller and therefore the changes in anatomy can be assumed to be small. Bladder and spinal cord are not considered, the bladder because of its varying filling and the spinal cord because the distal parts were cut in different levels on each CT.

Patient	Radiooncologist	V_{PCT}/V_{CBCT1}			
		Rectum	Prostate	PG-l	PG-r
P1 (P1')	RO1	1.07	1.36	1.14	1.22
	RO2	1.01	0.79	0.99	0.99
	RO3	1.04	1.24	0.99	0.88
	RO4	0.95	1.12	1.08	0.95
P2 (P2')	RO1	0.99	0.88	1.22	1.21
	RO2	1.06	0.73	1.17	1.10
	RO3	1.47	0.80	1.22	1.32
	RO4	1.14	1.08	1.14	1.48
P3 (P3')	RO1	0.96	1.14	0.97	1.04
	RO2	1.10	1.04	0.75	0.89
	RO3	1.03	1.11	1.15	0.87
	RO4	1.25	1.06	0.71	0.54
P4 (P4')	RO1	0.96	1.49	1.41	1.49
	RO2	1.16	1.16	1.40	0.94
	RO3	0.94	1.04	1.18	1.46
	RO4	1.09	1.00	0.82	0.68
P5 (P5')	RO1	0.80	1.68	0.95	1.11
	RO2	0.95	1.36	0.85	1.00
	RO3	0.89	0.99	1.01	1.15
	RO4	1.33	1.63	0.53	0.80

Table 3 Ratio of volume sizes derived from organs contoured on PCT and CBCT1 for all patients and ROs (green: <0 .95, black: 0.95-1.05, red >1.05).

The data show that the structures are more often contoured smaller on CBCTs compared to PCTs. There is no evidence that this volume difference correlates with particular patients or ROs in any of the organs contoured.

The different symbols within a colour group (organ) in Fig.11 represent different patients. Equal symbols in a group correspond to the volumes delineated for the same patient by different ROs. Besides the illustration of the differences between volume sizes determined on PCT and CBCT1 the variability in the ROs' contouring of identical structures is visualized.

In clinical routine, only one RO is contouring the prostate. It is important that the derived PTV is definitely covering the whole prostate, the microscopic spread and the positioning uncertainties. The contours delineated by different ROs would result in different PTVs. The delineation preferences of the responsible RO are unknown and it is obviously not clear, which of the ROs produces the real contour most correctly. In our study with four different ROs it is interesting to check if the PTV derived from the smallest CTV including a typical safety margin covers at least the CTVs of the other ROs to provide a degree of certainty. For this purpose, a safety margin of 0.7 cm was created around the prostate with the smallest volume, which is a realistic safety margin in clinical routine. Table 4 depicts the percentage of the prostate volume as delineated by each RO that is included within the volume of the smallest prostate with 0.7 cm safety margin. If less than 100% of the original volume by each other RO is included within the 0.7 cm safety margin, the size of the safety margin might not be large enough and the risk of local failure is increased.

Pat.	RO	V_{RO} (cm³)	$(V_{RO})_{min} \cup V_{SM}$	$\frac{V_{RO}}{(V_{RO})_{min} \cup V_{SM}}$ (%)	RO	V_{RO} (cm³)	$(V_{RO})_{min} \cup V_{SM}$	$\frac{V_{RO}}{(V_{RO})_{min} \cup V_{SM}}$ (%)	RO	V_{RO} (cm³)	$(V_{RO})_{min} \cup V_{SM}$	$\frac{V_{RO}}{(V_{RO})_{min} \cup V_{SM}}$ (%)
		PCT				**CBCT1**				**CBCT2**		
P1	RO1	52.4		92	**RO1**	**38.6**	**111.9**	**96**	RO1	39.3		97
	RO2	**41.9**	**112.8**	**98**	RO2	53.2		98	**RO2**	**33.8**	**99.3**	**99**
	RO3	56.7		89	RO3	45.7		96	RO3	50.4		91
	RO4	54.9		91	RO4	49.2		96	RO4	39.3		97
P2	**RO1**	**31.2**	**92.6**	**96**	**RO1**	**35.6**	**102.0**	**97**	**RO1**	**44.9**	**120.1**	**98**
	RO2	33.3		97	RO2	45.8		98	RO2	45.9		98
	RO3	33.2		96	RO3	41.4		97	RO3	49.0		96
	RO4	47.8		93	RO4	44.2		96	RO4	46.2		97
P3	**RO1**	**25.6**	**74.4**	**96**	**RO1**	**22.5**	**71.5**	**96**	**RO1**	**23.3**	**71.5**	**97**
	RO2	30.9		96	RO2	29.8		99	RO2	27.3		99
	RO3	29.8		87	RO3	26.8		96	RO3	31.0		97
	RO4	32.0		96	RO4	30.1		96	RO4	32.9		96
P4	RO1	33.5		97	**RO1**	**22.5**	**90.0**	**99**	RO1	36.0		89
	RO2	36.0		98	RO2	31.0		99	**RO2**	**29.3**	**89.0**	**99**
	RO3	36.4		96	RO3	35.1		98	RO3	40.4		94
	RO4	**31.5**	**89.2**	**97**	RO4	31.4		98	RO4	42.4		95
P5	RO1	64.2		97	**RO1**	**38.2**	**112.0**	**97**	**RO1**	**44.4**	**123.6**	**100**
	RO2	69.2		98	RO2	50.7		99	RO2	50.3		99
	RO3	**57.3**	**147.0**	**97**	RO3	58.0		96	RO3	57.0		96
	RO4	68.6		96	RO4	42.1		97	RO4	44.5		96

Table 4 Percentage of each RO's delineated prostate volume contained in the smallest delineated volume plus safety margin.

As depicted in Table 4 the intersection between the original volume by an RO and the smallest volume with safety margin is not 100%. The percentage of the volume included in the intersection volume ranges between 87% and 98% for PCT, 96% and 99% for CBCT1 and 89% and 99% for CBCT2, respectively. The disagreement was mostly observed in the apex and dorsal region of the prostate (Fig. 12). The accuracy of the automatic calculation of intersection volumes with the boolean operator function of the eclipse planning software was estimated by comparing the intersection volume of the smallest prostate with the smallest prostate plus 0.7 cm margin. Consequently, the calculated intersection volume should be 100% of the smallest volume. The deviation of the calculated values from this expected value demonstrates that the error of the automatic evaluation software is up to 4%.
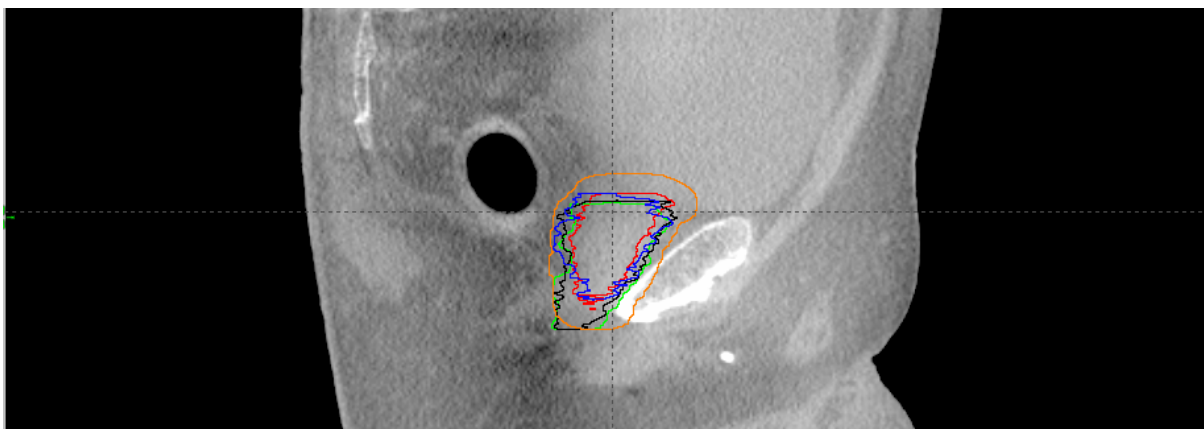
Fig. 12   0.7 cm safety margin (orange) around the smallest prostate contour: Its volume plus safety margin does not include all larger contours.

3.2   Dose Comparison

As described above, doses with and without heterogeneity correction were calculated for all OARs for each RO on PCT, CBCT1 and CBCT2 to assess the influence of HU.

The ratios of D_{Tissue} and D_{Water} are depicted in Table 5. When comparing the calculated D_{Tissue} and D_{Water} on PCT, CBCT1 and CBCT2, rather small differences were found for D_{mean}. The averaged ratios of doses calculated with heterogeneity correction and without heterogeneity correction (D_{Tissue}/D_{Water}) were close to 1. Looking at the single values, mean dose ratios maximally range from 0.95 to 1.04. Looking at $D_{2\%}$, all values are similarly close to 1, ranging from 0.95 to 1.05. Ratios calculated for PCTs are similar to the ratios for CBCT1 and CBCT2 (one patient of the PC group was excluded from dose calculation on CBCT2, due to the usage of a false filter).

	PCT	CBCT1	CBCT2
Inter-observer contouring variability induced scattering of D_{mean} (%) (ref. mean D_{mean}, for D_{Tissue})			
PC			
Bladder	3.4 [0.8 - 7.9]	6.7 [1.3 - 22.4]	7.8 [1.6 - 17.2]
Rectum	13.9 [2.6 - 34.9]	9.6 [1.6 - 24.0]	11.5 [0.9 - 18.9]
Prostate	0.3 [0.0 - 0.7]	0.2 [0.0 - 0.5]	0.7 [0.2 - 1.1]
H&N			
Left Parotid Gland	17.4 [11.5 - 26.8]	24.7 [4.7 - 60.8]	31.0 [14.6 - 66.9]
Right Parotid Gland	17.2 [5.5 - 25.5]	21.5 [4.1 - 48.3]	19.9 [12.5 - 32.9]
Inter-observer contouring variability induced scattering of $D_{2\%}$ (%) (ref. mean $D_{2\%}$, for D_{Tissue})			
PC			
Bladder	0.4 [0.0 - 0.5]	0.5 [0.0 - 1.8]	0.6 [0.0 - 1.3]
Rectum	2.2 [0.4 - 3.7]	0.9 [0.6 - 1.8]	3.0 [1.6 - 4.9]
Prostate	0.1 [0.0 - 0.3]	0.1 [0.0 - 0.2]	0.1 [0.0 - 0.2]
H&N			
Spinal Cord	3.1 [0.8 - 6.5]	5.0 [2.1 - 8.0]	4.4 [1.0 - 8.1]
Left Parotid Gland	7.9 [1.9 - 26.9]	5.6 [1.6 - 15.2]	10.0 [1.5 - 21.6]
Right Parotid Gland	12.4 [0.9 - 33.5]	15.2 [1.2 - 24.5]	7.8 [0.6 - 13.1]
Ratio D_{Tissue}/D_{Water} (for D_{mean})			
PC			
Bladder	1.00±0.00 [0.99 - 1.00]	0.99±0.01 [0.97 - 1.00]	0.99±0.01 [0.98 - 1.00]
Rectum	1.02±0.01 [1.00 - 1.04]	1.01±0.01 [0.99 - 1.03]	1.02±0.01 [1.00 - 1.04]
Prostate	0.99±0.01 [0.99 - 1.00]	0.99±0.01 [0.97 - 1.00]	0.99±0.01 [0.98 - 1.00]
H&N			
Left Parotid Gland	0.99±0.00 [0.99 - 1.00]	0.99±0.01 [0.98 - 1.00]	0.98±0.01 [0.95 - 1.00]
Right Parotid Gland	1.00±0.01 [0.99 - 1.03]	0.99±0.00 [0.98 - 0.99]	0.98±0.01 [0.95 - 0.99]
Ratio D_{Tissue}/D_{Water} (for $D_{2\%}$)			
PC			
Bladder	0.99±0.01 [0.99 - 1.00]	0.99±0.01 [0.97 - 1.00]	0.99±0.01 [0.98 - 1.00]
Rectum	0.98±0.01 [0.95 - 1.00]	0.98±0.02 [0.95 – 0.99]	0.98±0.01 [0.97 - 0.99]
Prostate	0.99±0.00 [0.99 - 1.00]	0.99±0.01 [0.97 - 1.00]	0.99±0.01 [0.98 - 1.00]
H&N			
Spinal Cord	0.99±0.01 [0.98 - 1.00]	1.01±0.02 [0.98 - 1.04]	1.00±0.02 [0.96 - 1.03]
Left Parotid Gland	0.99±0.01 [0.98 - 1.00]	0.99±0.01 [0.97 - 1.00]	0.98±0.01 [0.96 - 1.00]
Right Parotid Gland	1.00±0.02 [0.98 - 1.05]	0.99±0.01 [0.98 - 1.00]	0.98±0.01 [0.96 - 0.99]

Table 5 Influence of inter-observer contouring variability and HU uncertainties on dose. Dose calculation results based on HU with and without heterogeneity correction (D_{Water} and D_{Tissue}) match well.

Exemplary DVH-c of CBCT1 for all organs delineated by each RO and calculated with and without heterogeneity correction are shown in Fig. 13. The DVH-c representing D_{Tissue} and D_{Water} for the same contouring nearly match, while the graphs representing the different ROs vary noticeably. The rectum is of particular interest because of the air filled rectal balloon (Geier *et al.*, 2012). Thus, the influence of air (rectal balloon) on the results of dose calculation can be observed. DVH-d (Fig. 14) show the differences derived from inter-

observer contouring variability and HU differences more distinct. They show shifts of the graphs on the horizontal axis when doses were calculated without heterogeneity correction.
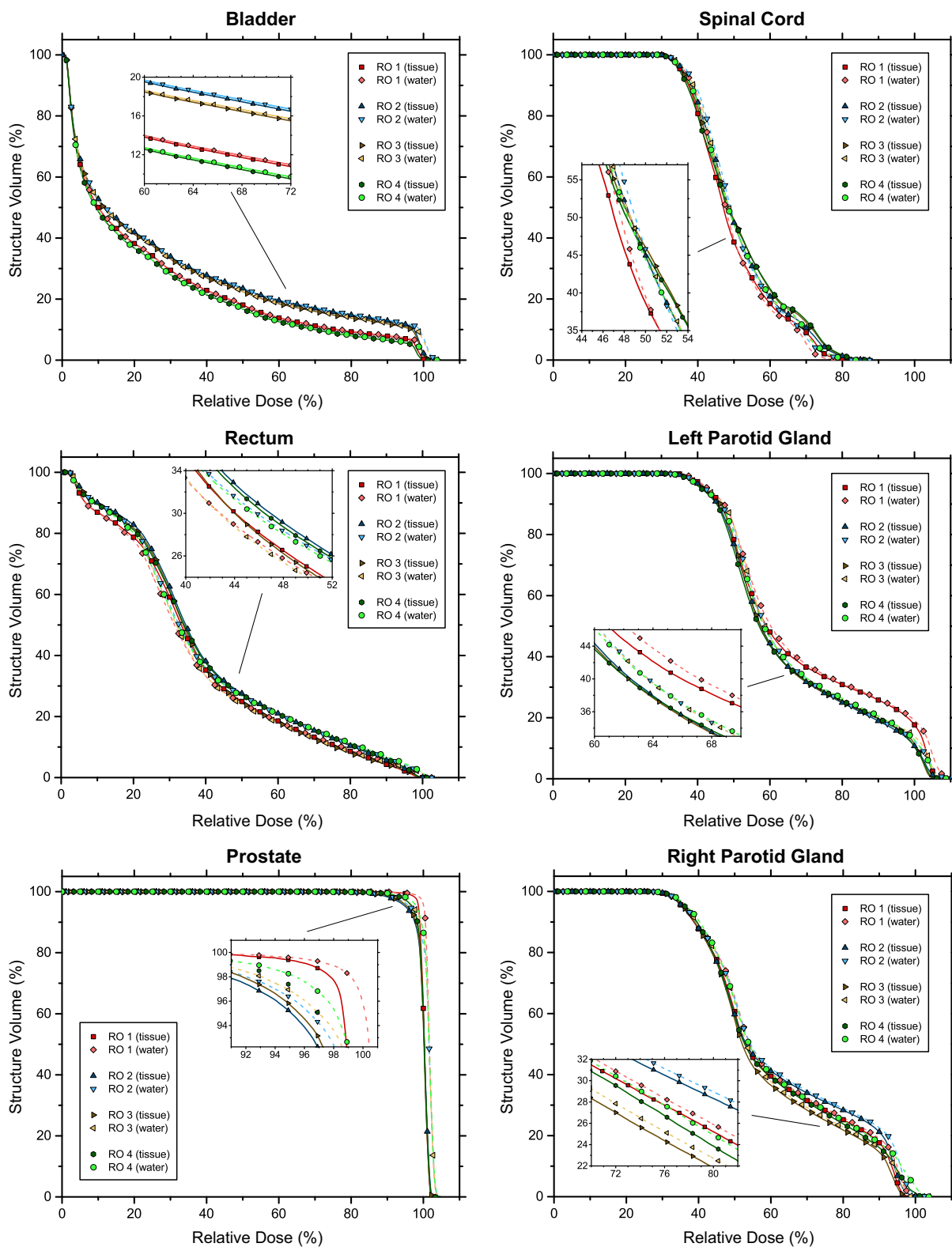

Fig. 13 Exemplary cumulative dose-volume-histograms (DVH-c) for all OARs on CBCT1. For better visibility only part of the calculated data columns are plotted. The divergence between the ROs due to inter-observer contouring variability exceeds the divergence due to Hounsfield units.
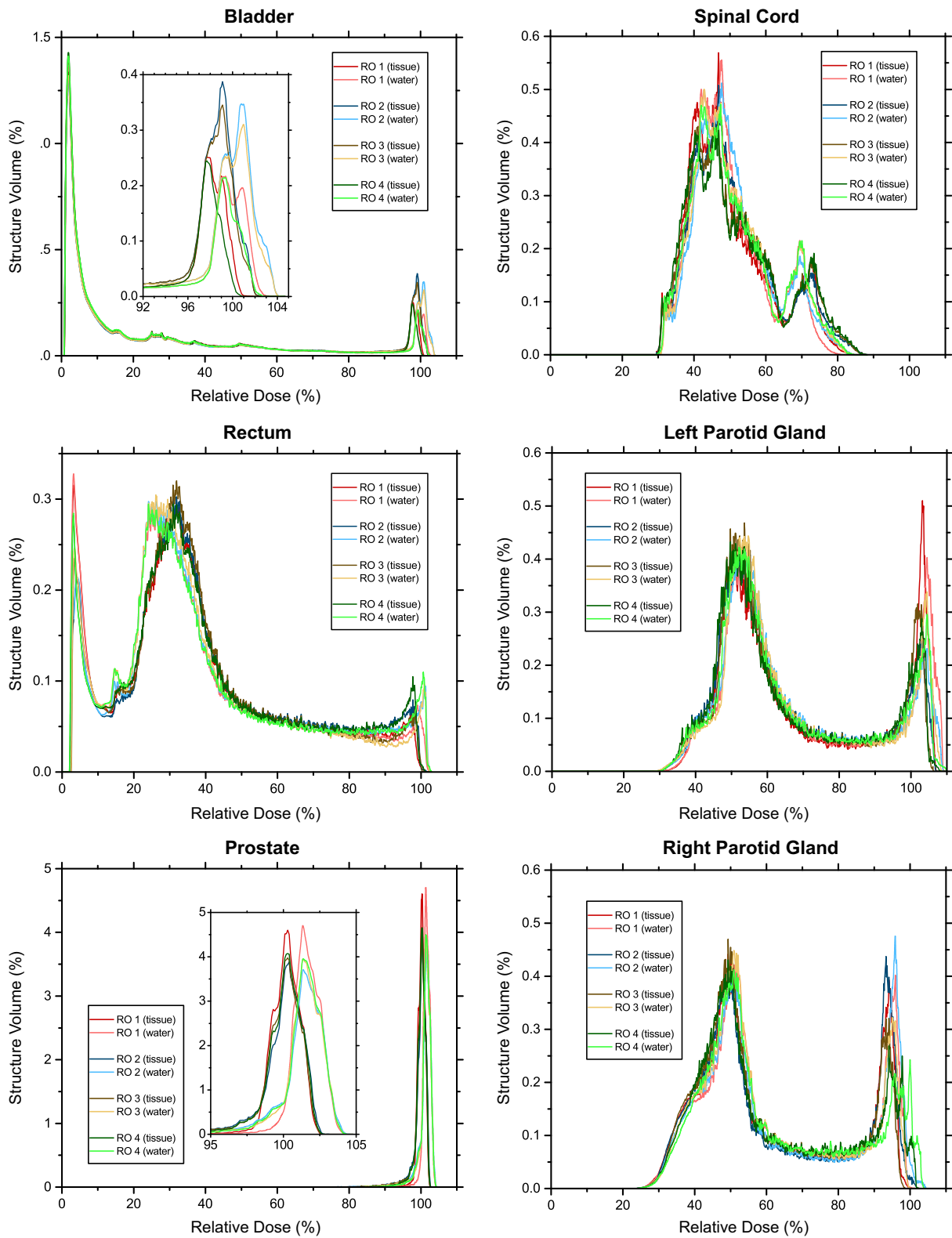
Fig. 14 Differential dose-volume histograms (DVH-d) for the same data set as in Fig. 13. The vertical scale is based on a width of the relative dose bin of 0.1%.

In order to judge the influence of inter-observer contouring variability against the influence of HU on the calculated dose, the largest range of dose values due to the different contouring by the ROs was compared to the largest deviation generated by the different HU.

Inter-observer contouring variability results in dose deviations from the mean value between the RO with the highest and the RO with the smallest calculated dose of up to 66.9% (left parotid gland, CBCT2) for D_{mean}. The largest dose deviation for $D_{2\%}$ is 33.5% (right parotid gland, PCT). In comparison to that, the largest dose difference between calculation with and without heterogeneity correction is rather small. Here, the largest dose deviation is 5% (parotid glands, CBCT2) for D_{mean} and also 5% (rectum, PCT and CBCT1 and right parotid gland, PCT) for $D_{2\%}$.

The calculated dose ratios for all patients, ROs and OARs show that the resulting doses are smaller when calculated with heterogeneity correction than calculated without heterogeneity correction with only few exceptions.

4 Discussion

4.1 Inter-Observer Contouring Variability

In this study, the possibility of RT planning on CBCTs in terms of inter-observer contouring variability and dose calculation in the H&N and the pelvic region was assessed. The H&N and the pelvic region are of special interest, as there are many OARs in these body regions lying close to each other, making organ definition and correct dose application even more significant. Additionally, replanning often becomes necessary during treatment in these body regions. In the pelvic region, intrinsic changes of organ positioning by physiological motion due to varying organ filling occur. In the H&N region, weight loss and sometimes quick tumor shrinkage plays the major role resulting in a different constitution of a patient. Both changes can become significant, making replanning mandatory (Gai *et al.*, 2017).

Contradictory data in the literature are available on inter-observer contouring variability on PCT and CBCT.

Lütgendorf-Caucig *et al.* conducted a study, where seven radiation oncologists delineated the prostate and the seminal vesicles in eight patients with prostate cancer on CT, CBCT and MRI (Lütgendorf-Caucig *et al.*, 2011). Inter-observer contouring variability was judged by the same parameters that were used in the present study, the CI_{gen}, CV and the COM. They reported an increased inter-observer contouring variability in patients with prostate cancer on CBCT compared to CT and MRI for the prostate. The increase in inter-observer contouring variability on CBCT was statistically significant compared to the CT. For seminal vesicles, the contouring on CBCT was less reliable, too. In their case only the differences in CI_{gen} and CV on CBCT compared to CT and MRI reached statistical significance. Differences in volume only reached statistical relevance on CBCT compared to MRI. Delineation of bladder and rectum were considered satisfactory, although differences in CI_{gen} and CV between CT and CBCT were statistically significant. They conclude that planning on CBCTs might only be possible by applying additional safety margins, especially in the superior-inferior direction. Choi et al. assessed inter-observer contouring variability for prostate delineation on PCT and CBCT under offline MRI guidance (Choi *et al.*, 2011). Three radiation oncologists contoured the prostate in ten patients after implantation of fiducial markers. There were no significant differences in prostate volume, while they found significant differences between the inter-observer contouring variability on PCT and CBCT when looking at the three-dimensional

distances of each *COM* to the average *COM* and the *CI$_{gen}$*. The largest deviation was found in the anterior-posterior and the superior-inferior direction. They suggest further guideline protocols for delineation of the prostate on CBCT. White *et al.* found good reliability of the position of *COM* for the prostate, but high variation of the volumes of up to 16%, especially in the superior-inferior axis (White *et al.*, 2009). Foroudi *et al.* assessed inter-observer contouring variability in patients with bladder cancer on PCT in comparison to CBCT. They assessed the bladder and the rectum. They found that there are no significant differences in inter-observer contouring variability between the CTs, even though the quality of CBCT was subjectively judged considerably worse (Foroudi *et al.*, 2009). Nishioka *et al.* assessed bladder and prostate delineation on CBCT for patients with bladder cancer (Nishioka *et al.*, 2013). They also found a small non-significant inter-observer contouring variability in terms of *COM*, *CI$_{gen}$* and *CV* for both bladder and prostate, with *CI$_{gen}$* of 0.81 for the bladder and *CI$_{gen}$* of 0.66 for the prostate. There were no data available for inter-observer contouring variability on CBCT in the H&N region.

The purpose of the direct organ volume comparison in this study was to assess a possible systematic deviation in contouring between the four ROs, i.e. if one RO is consistently contouring a larger or smaller volume. Additionally, it is possible to see the extent of volume deviations and on which type of CT they are most obvious. The best image quality is displayed by the PCT, so inter-observer contouring variability should be small. As CBCT1 was acquired in the beginning of RT, no irradiation effects were expected. CBCT2 was obtained at the end of RT, so irradiation effects such as inflammation and edema might impair image quality and contouring additionally. Therefore, the organ contours of the OARs might be discriminated more easily on CBCT1 than on CBCT2, hypothetically resulting in larger inter-observer contouring variability on CBCT2. It is important to use all parameters proposed above to describe inter-observer contouring variability, as they describe different aspects. The *CI$_{gen}$* together with the *COM* describes the volume conformity and positioning conformity of the organ within the body. The *CV* gives an impression of the agreement of the volume sizes, while the *ICC*, if reviewed carefully, provides an estimation of the consistency of contouring performed by different ROs.

Only combined, these parameters deliver a satisfactory description of inter-observer contouring variability – reliability of volume-size definition and positioning. Additional parameters, such as the systematic error, deviation from the median and the overall agreement

of intersection and union volumes, were assessed to further highlight the various aspects of inter-observer contouring variability.

In the present study, inter-observer contouring variability for all regions investigated is rather large, including size differences between the largest and the smallest contours of more than a factor 2. The distribution of the volume deviation from the median confirms these results. The largest confidence in delineation is found for the bladder, where the deviation from the median lies within 10%. Most volume deviations lie in the range of 20% to 40%. The spinal cord and the parotid glands display the largest deviation from the median with outliers of up to >80%. Inter-observer contouring variability is similar on all CTs. There is no trend of increasing organ definition insecurity on CBCT2 as compared to CBCT1 and both compared to PCT. Only for the parotid glands and the bladder a slight increase of inter-observer contouring variability can be observed on CBCT2. Therefore, irradiation effects do not seem to have the expected impact on inter-observer contouring variability. A systematic error has been found, suggesting that RO2 tends to define larger volumes and RO1 smaller volumes, while RO3 and RO4 are usually placed in the middle field.

The factor of the ratio of the intersection volume of all delineations and the union volumes of all delineations (Table 1) is dependent on both differences in the sizes of the volumes delineated, as well as their position towards each other. The variation between intersection and union volume are surprisingly large. As was already the case with the other description parameters, bladder displays the best results with an agreement of up to 82%. Volume agreement for organs delineated in the H&N region is often less than 40%, for some cases even less than 20%. Results for rectum and prostate vary around 50%. Comparing the volume discrepancies with the *COM* differences and the *CI$_{gen}$* results, it is obvious that the dominant part of inter-observer contouring variability is due to volume discrepancies, rather than positioning differences.

The best certainty in delineation is shown for the bladder. Inter-observer contouring variability for parotid glands is larger than for prostate or bladder. These results suggest that contouring in the H&N region is more demanding.

These findings are consistent in all three CTs (PCT, CBCT1 and CBCT2). Inter-observer contouring variability remains a general problem during the planning of RT. In this study, it is, however, not generally larger on CBCTs and it does not increase in the course of irradiation.

The rectum, the spinal cord and the parotid glands posed the most problems during contouring for different reasons: The low level of agreement on the delineation of the rectum and the spinal cord are mainly caused by the different delineation in the longitudinal direction. The guideline for contouring of the rectum states that it ranges from anus to the beginning of the sigmoid colon. This region, however, is difficult to define and therefore a cause of disagreement. The deviations of *COM* were assessed in the superior-inferior, anterior-posterior and medio-lateral direction separately. While all other deviations are small, it is noticeable that there is a larger deviation for the rectum in the longitudinal direction, with *COM* of 0.41 cm in the z-direction compared to 0.04 cm in the x- and 0.12 cm in the y-direction for PCT. This is also shown for CBCT1 and CBCT2 with x-/y-/z- values of 0.03 cm/0.09 cm/0.30 cm and 0.02 cm/0.06 cm/0.13 cm, respectively. Fig. 15 depicts an example of inter-observer contouring variability of the rectum.
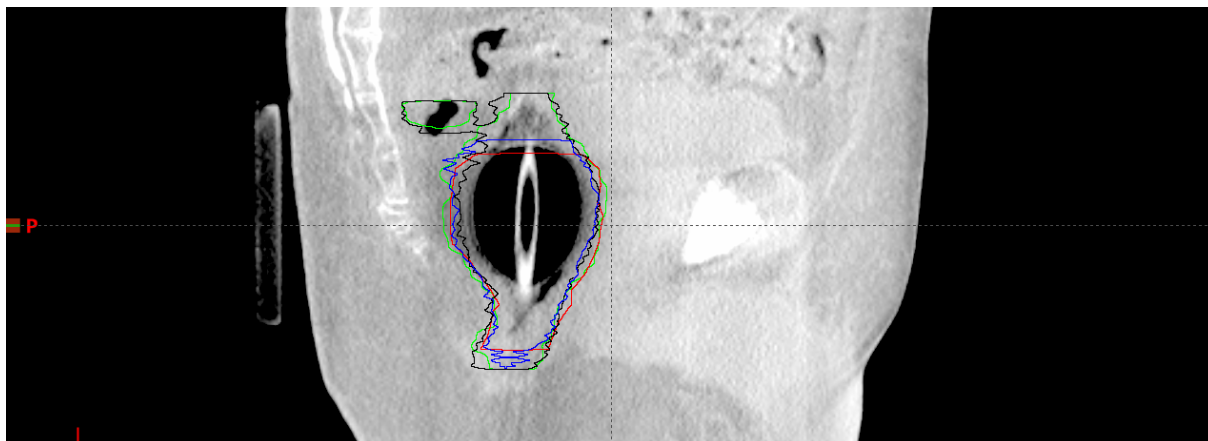
Fig. 15 Inter-observer contouring variability in the longitudinal direction of the rectum at the transition into the sigmoid colon on CBCT1.

There is a similar problem concerning the spinal cord. To judge inter-observer contouring variability more adequately, the spinal cord was cut at the distal part to end at the same CT level, as the distal parts were not defined. Especially in the PCT, where a larger part of the body was imaged, some ROs continued contouring to the end of the image and some stopped contouring at a level they considered relevant for treatment planning. The proximal border is the part where the spinal cord turns into the medulla oblongata. The delineation differences in this region are of interest and often a cause of error. This, too, becomes more obvious by looking at the deviation from the mean *COM* in each direction separately, where the largest deviations are observed in the z-direction with x-/y-/z- values for PCT, CBCT1 and CBCT2

of 0.05 cm/0.13 cm/0.83 cm, 0.03 cm/0.09 cm/0.47 cm and 0.05 cm/0.06 cm/0.41 cm, respectively. This is illustrated in Fig. 16.
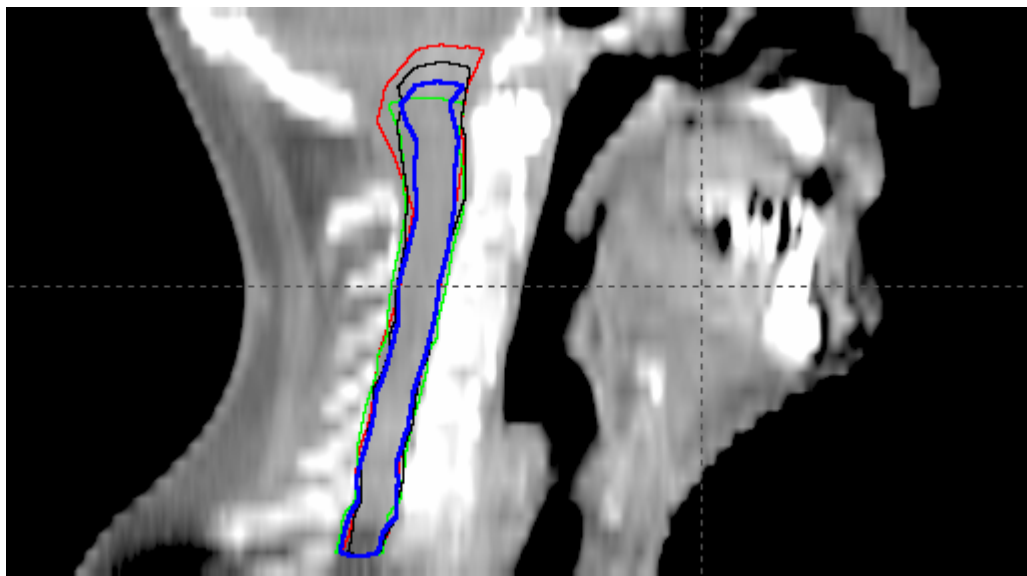
Fig. 16 Inter-observer contouring variability in the longitudinal direction of the spinal cord.

The relatively large inter-observer contouring variability of the parotid glands might be due to the inherent anatomical conditions in the H&N region with many small adjacent structures with similar densities. Therefore, distinguishing the parotid glands from the surrounding structures, such as lymph nodes, blood vessels and other soft tissue, can sometimes be difficult. For the parotid glands, these uncertainties in defining the contours are aggravated by the limitations of CBCT image quality and also seem to be influenced by irradiation artifacts and irradiation effects, such as edema, showing in the development of CI_{gen} and CV between CBCT1 and CBCT2. However, correct organ definition is one of the most crucial steps in planning irradiation. Especially in the H&N region, where contouring looks most difficult, precision is very important, as the close proximity of OARs and PTVs put healthy tissue to greater risk. The local narrow situation implies that OARs are closer to the PTV and therefore to the high dose region.

The larger dispersion in volume size of the prostate between the ROs is mostly due to disagreement on defining the apex of the prostate (Fig. 17).
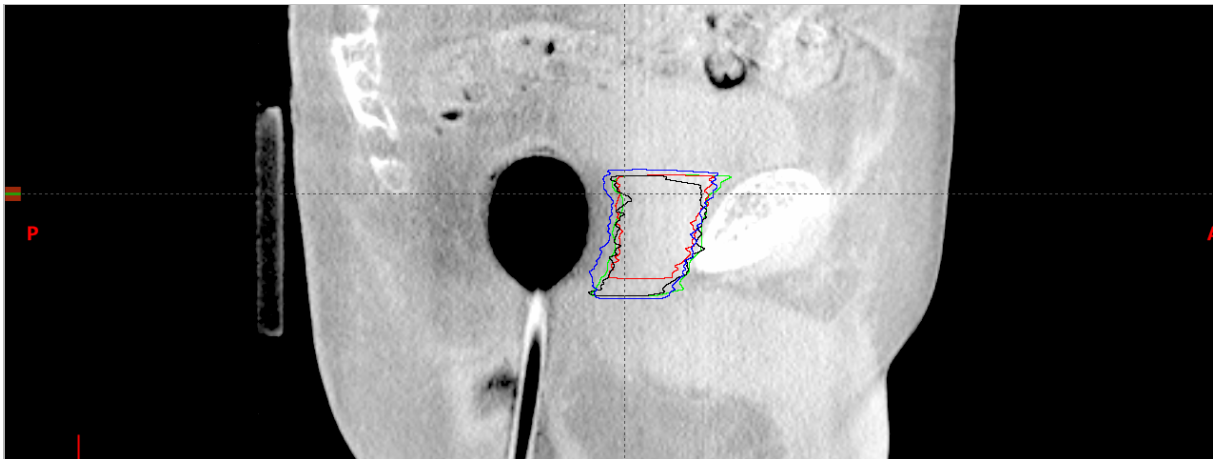
Fig. 17 Inter-observer contouring variability of the prostate, illustrating the high variability in the apex region.

As Fig. 11 illustrates, in the majority of cases volume sizes tend to be delineated smaller on CBCT compared to PCT. Fig. 11 also shows that the largest dispersion of volume sizes can be observed for the rectum and the prostate. There are two possible explanations for the dispersion in rectum size. First of all, the previously described difficulties in defining the transition from the rectum to the sigmoid colon can be a reason for these discrepancies. Moreover, the rectum is an organ with varying filling and varying size. This influence factor could be minimized by the use of the rectal balloon with a standardized inflation volume. The prostate is also more often underestimated on CBCT1. This is probably even more important as it represents a target volume, and emphasizes again the importance of the right choice of the safety margin size.

Reliability evaluation values for application in RT have to be judged more critical than in many other fields. A rating like "moderate reliability" would not be a sufficient attribute. Due to the far-reaching consequences it is important to aim at best possible confidence. In order to handle unavoidable error sources, the introduction of safety margins around target organs is essential in planning. The assessment of the safety margin shows that the apex of the prostate is in danger of not being included within the safety margin. Therefore, this region might be at risk of local failure. This is also the region with the highest inter-observer contouring variability as suggested by the deviation of the *COM* from the mean *COM* in the z-direction. It must be discussed to increase safety margins in the longitudinal direction.

To increase the reliability of contouring, a few procedures have already been suggested by several authors. Especially for the H&N region, it might be useful to co-register the CBCTs

with pretreatment PCTs and especially MRIs. That way, soft tissue can be judged more easily and reliably. For the pelvic region, PET-CTs could be helpful. As Choi *et al.* already suggested for the prostate, additional guidelines could be helpful for delineation, not only on CBCTs but also for PCTs (Choi *et al.*, 2011). However, there is already a number of contouring protocols and house standards and still the inter-observer contouring variability is high. Also the insertion of larger safety margins as proposed by Lütgendorf-Caucig *et. al.* (Lütgendorf-Caucig *et al.*, 2011) must be discussed controversially. The added safety margin decreases the risk of local failure, but possibly at the cost of increased local toxicity. Therefore the introduction of larger safety margins has to be judged carefully.

4.2 Dose Comparison

Several studies report an unreliability of HU measurement on CBCT. Hatton *et al.* investigated HU measurement on CBCT, using phantoms of different radial and longitudinal diameters simulating different patients' constitutions (Hatton *et al.*, 2009). They found that HU vary significantly with larger diameter, especially in the radial circumference, resulting in errors in dose calculation by HU to electron density (ED) calibration. Eskandarloo *et al.* scanned parts of a mandibular bone embedded in three different density materials. They varied the position of the phantoms and used three different scanners.They found significant HU variation on CBCT due to the position of the bone on the irradiation desk (Eskandarloo *et al.*, 2012). There have been a few approaches to calculate dose on CBCT, most of them using phantoms. Guan *et al.* calibrated HU to ED curves for different phantoms of the pelvic region (Guan & Dong, 2009). They found that on CBCTs, different calibration curves are needed for different technical settings and different patients' constitutions. Yoo & Yin found differences in HU measurements on CT and CBCT of up to 200 HU values; in patients' back muscles of up to 300 HU (Yoo & Yin, 2006). They found dosimetric differences of up to 2-3% in MU/cGy for CBCTs and CTs using an inhomogeneous phantom (at the lung or the edges of the field of view) and good agreement between DVH and isodose, with higher discrepancies in the high dose regions. Ding *et al.* assessed HU discrepancies and dose calculation on CBCT on phantom and patient cases for the H&N region and PC (Ding *et al.*, 2007). While in some areas HU differ, they do not influence dose calculation significantly, resulting in nearly matching DVH.

To assess how large the influence of HU uncertainties on dose calculation really is in relation to other input factors (here: inter-observer contouring variability) two scenarios were

compared to each other – dose calculation on the basis of the CT derived HU and dose calculation on the basis of HU that were all equalized to water.

In this study, differences between D_{Water} and D_{Tissue} are rather small. D_{mean}-ratios of the rectum show a slightly larger discrepancy on all CTs. This might be due to the influence of air in the rectal balloon. Other studies found that inaccuracies in HU calculation vary a lot depending on the tissue. As deduced from the other studies, larger dose differences can be expected for certain regions, such as the back muscles close to the spine (Yoo & Yin, 2006). Air and regions with high density gradients might have far bigger influence on HU and dose calculation. The influence of air must be assessed more closely in other studies. It must be noted, that the present study only considers organs and dose calculation in the pelvic and H&N region. It is possible that a dose assessment in the thoracic region would show significant differences.

Calculated DVH allow for an easy direct comparison of the influence factors inter-observer contouring variability and HU. Looking at the DVH, some interesting details become obvious. The eight graphs (Fig. 13), one for each RO calculated for each organ with and without heterogeneity correction, can be judged in regard of their conformity concerning inter-observer contouring variability and different calculation modes. The difference in HU does not have great influence on dose calculation. The graphs representing the DVH of the two calculation modes – with and without heterogeneity correction – match rather accurately in most cases. The rectum poses an exception (Fig. 12). As explained above, the illustrated examples for DVH of the rectum were chosen to observe the influence of air on dose calculation, as the rectal balloon is filled with air (Geier *et al.*, 2012). In the pelvic and H&N regions, there are mostly tissues with similar densities and therefore, the calculated DVH representing the calculated dose with and without heterogeneity correction match satisfactorily. Especially the bladder shows the most accurate conformity, most likely due to its water-like density. However, results for regions containing more air, such as thorax or the sinuses, or regions with steep density gradients might look very different. The graphs representing the different ROs, however, show a much larger discrepancy. The influence of inter-observer contouring variability clearly exceeds the influence of HU. There are larger differences between D_{Tissue} and D_{Water} in the high dose regions (see DVH-c).

The results show that doses calculated with heterogeneity correction (D_{Tissue}) are usually smaller than doses calculated without heterogeneity correction (D_{Water}). This seems logical, as denser tissues, such as bones, absorb part of the photons before the target is reached.

Equalizing bone density to water density while using the same irradiation plan with the exact same beam angles and MU, the absorption by bone or other denser tissue is not considered and this must result in a higher calculated dose in the target. DVH-d also illustrate the shifts on the horizontal axis. Here, also the exceptions are visible, where this ratio is reversed. Especially the rectum shows this reversed ratio in nearly all patients, possibly because of the increased HU-uncertainties in air.

4.3   Impact of Inter-Observer Contouring Variability Versus Impact of Hounsfield Unit Uncertainties

The comparison of the influence factors inter-observer contouring variability and HU suggests a larger impact of inter-observer contouring variability. Only in few exceptions, the influence of HU exceeded the influence of inter-observer contouring variability. The prostate is the only OAR, for which the influence of HU continually exceeds the influence of inter-observer contouring variability. This might be due to the fact, that the prostate is not only an OAR in our study, but it also poses a target volume. The dose is calculated for the PTV, i.e. the prostate with a safety margin around it. The prostate as an organ is therefore homogenously covered by a larger high dose region. Thus, differences in contouring do not result in significant dose differences as all contours are covered by the high dose region. Dose differences due to inter-observer contouring variability are therefore exceeded by the dose differences due to HU.

Nevertheless, as suggested by the results for the rectum, it remains important to consider the tissue densities in the calculation of treatment plans. Although dose calculation with and without heterogeneity correction results in small differences in the calculated doses in this study, there are indications that this might not apply for other body regions. Also, as suggested by the DVH, doses calculated with or without heterogeneity correction differ more in the high dose region, which can become significant considering hot spots in sensitive OARs. As it was described above, the DVH-d imply that in some cases, differences due to HU are not much smaller than dose differences due to inter-observer variability, but they diverge in a different way. The graphs representing the ROs display a shift on the vertical axis, i.e. the same percentage of the prescribed dose covers a larger or smaller part of the volume, while the graphs representing D_{Tissue} and D_{Water} display a shift on the horizontal axis, i.e. the same percentage of the volume is covered by a higher or smaller percentage of the prescribed dose.

The broad consistency of the resulting doses for the different delineations of the target structure (prostate) in our study demonstrates overall satisfactory masking of the target region by the security margins. One should keep in mind, however, that in the DVHs (Fig. 13 & 14) the volume axis is specified in relative units, with 100% representing the total delineated volume. The applied dose integrated over the target organ scales with the volumes delineated by the different ROs. It should be noted that the total irradiation exposure of a patient during RT depends on the delineated volume used as input for dose planning. A smaller delineated volume implies a smaller irradiation exposure for the patient, but it can be at the cost of a geographic miss. A larger volume, even more enlarged by the security margin, however, results in a larger total irradiation exposure of the patient, which emphasizes the importance of the "right" security margin, but even more of the "right" contouring. As a whole, inter-observer contouring variability might have a more profound impact on the whole patient.

Although all calculations suggest that inter-observer contouring variability is a more relevant influence factor than HU, none of them should be neglected in order to ensure most precise dose calculation.

It is difficult to agree on an acceptable dose deviation for PTVs and OARs. Until now, it is not well defined, what a relevant inter-observer contouring variability or dose deviation from the prescribed dose is and which level of variability can be tolerated. Also, these influence factors are only two of many. For example, inter-fractional patient setup errors are also source of inaccuracies.

To increase delineation conformity, co-registering of the PCT with other imaging modalities and additional delineation guidelines can be helpful. However, inter-observer contouring variability remains large despite the already existing protocols. Brouwer *et.al.* suggest in addition to further delineation guidelines and co-registering of other imaging modalities regular joint delineation review sessions (Brouwer *et al.*, 2012). As already discussed above, larger security margins to account for PTV definition uncertainties can increase toxicity to surrounding normal tissue and must therefore be applied carefully.

5 Limitations of the Study

The most important limitation of the present study is the small number of patients and observers. Thus error bars of the calculated evaluation parameters are large and it is difficult to make a valid judgement about clinical significance. Additionally, it would be of interest to analyze the influence of intra-observer contouring variability more closely and to put it into relation to inter-observer contouring variability. The H&N and the pelvic region chosen for this study are two body regions with mostly homogeneous tissue densities, i.e. densities that are close to water. This might be a reason for the similar results of dose calculation with and without heterogeneity correction. It is not possible to draw general conclusions for other regions of the body. As mentioned above, it is also not easy to judge these findings in means of their relevance. It is not trivial to estimate, whether a certain dose deviation has large or small impact on the clinical outcome.

Nevertheless, the clinical time and personnel requirements for this kind of investigations are high and may be too extensive for a single study. Compilation of all available data is necessary, and therefore each contribution is valuable.

6 Conclusion

In this study, inter-observer contouring variability is found comparable between PCT and CBCT for bladder, rectum and prostate for PC patients and for left and right parotid gland and the spinal cord in the H&N region. This suggests that contouring on CBCTs is not generally less accurate than contouring on PCT and that the influence of impaired image quality on CBCTs is manageable. However, there are large differences in inter-observer contouring variability between the different organs. This concerns contouring on all CT types. It seems to be more difficult to define correct organ contours in the H&N region. This might be due to the inherent anatomical situation in this area, where many small structures with similar tissue densities lie close to each other and are sometimes difficult to differentiate. Anatomical landmarks, for example parts of the pelvic bone or the base of the skull, can help to define organ boundaries. Generally, inter-observer contouring variability is large on all CTs. Inter-observer contouring variability is a human error that can never be ruled out completely. To increase accuracy, it can be recommendable to additionally rely on guidance of other imaging modalities, such as MRI or PET, which can be co-registered and fused with the PCT and CBCT. In operated patients, it might be helpful to consult the surgeon to discuss the post-surgery changes in anatomy. This emphasizes the importance of the multidisciplinary setting, even in the course of irradiation planning.

Comparing dose calculation with and without heterogeneity correction, there are only small discrepancies. The impact of HU is low for the chosen body areas in this study. However, as the exemplary dose calculations for the rectum suggest, it must be assumed that results for other regions of the body might vary greatly. This is especially true for body regions containing air, like the thorax, or where tissues with different densities lie close to each other, resulting in steep density gradients between the OAR and its surroundings. It could also be demonstrated that dose calculation without heterogeneity correction with HU equalized to water overestimates doses to the OARs in most cases.

This study shows that for the specific regions chosen, namely the H&N and the pelvic region, contouring on CBCTs results in similar inter-observer contouring variability as contouring on PCT. CBCTs provide all in all reliable dose approximations. However, due to the limitations of this study, such as the small patient number, further studies are essentially necessary to confirm these results and provide assessments for other body regions.

7 Summary

Purpose:

Adaptive online planning is an ambitious goal in modern radiotherapy concepts. Until now, kilovolt cone beam CTs (CBCTs) are used for image guidance during radiotherapy but not for planning due to impaired image quality. However, it would be of advantage for the work flow and the patients, if CBCTs could directly be used for adaptive replanning. The aim of this study was to assess inter-observer contouring variability and the influence of Hounsfield unit (HU) uncertainties on dose calculation based on CBCTs in comparison to planning CTs (PCTs).

Materials/Methods:

Data sets of five head and neck cancer patients and five prostate cancer patients were evaluated. Four radiation oncologists delineated the left parotid gland, the right parotid gland and the spinal cord for the head and neck cancer group. For prostate cancer patients, the prostate, bladder and rectum were contoured, each on the PCT and two CBCTs. Inter-observer contouring variability was judged by the parameters generalized conformity index, coefficient of variation, intra-class correlation coefficient and the deviation of the center of mass. To assess the reliability of dose calculation on CBCTs based on CBCT-derived HU, the original irradiation plans from the PCT were copied to both CBCTs. The doses were recalculated using the same plan parameters and monitor units as in the PCTs, but applying CBCT calibration curves. All doses were computed with heterogeneity correction as well as without heterogeneity correction (HU=0).

Results:

Inter-observer contouring variability was similar on all types of CTs, left and right parotid glands, spinal cord and rectum with the largest variability. Evaluating the results of dose calculation with and without heterogeneity correction, there were only minor differences in mean and maximum doses. Comparing dose volume histograms, the influence of inter-observer contouring variability exceeds the influence of HU on the results of dose calculation. Differences in contouring mainly result in a shift of differential DVH on the vertical axis, while differences in HU mainly result in a shift of the differential DVH on the horizontal axis.

Conclusion:

Inter-observer contouring variability is a general challenge in radiation oncology, which is not unique to CBCTs. Only for certain organs a slight increase in inter-observer contouring variability was observed on CBCT compared to PCT. Inter-observer contouring variability on all CTs have larger impact on the results of dose calculation than HU, but the resulting dose differences were overall not larger on CBCT than on PCT. It must be noted, that these results only concern dose calculation for the pelvic and the head and neck region. More studies are essentially necessary to assess dose calculation in other body regions.

References

Agarwal, J., Palwe, V., Dutta, D., Gupta, T., Laskar, S. G., Budrukkar, A., Murthy, V., Chaturvedi, P., Pai, P., Chaukar, D., D'Cruz, A. K., Kulkarni, S., Kulkarni, A., Baccher, G., and Shrivastava, S. K. (2011). Objective assessment of swallowing function after definitive concurrent (chemo)radiotherapy in patients with head and neck cancer. *Dysphagia,* **26**(4), 399-406. doi: 10.1007/s00455-011-9326-4

Bray, F. N., Simmons, B. J., Wolfson, A. H., and Nouri, K. (2016). Acute and Chronic Cutaneous Reactions to Ionizing Radiation Therapy. *Dermatol Ther (Heidelb).* doi: 10.1007/s13555-016-0120-y

Brouwer, C. L., Steenbakkers, R. J., van den Heuvel, E., Duppen, J. C., Navran, A., Bijl, H. P., Chouvalova, O., Burlage, F. R., Meertens, H., Langendijk, J. A., and van 't Veld, A. A. (2012). 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol,* **7**, 32. doi: 10.1186/1748-717X-7-32

Chennupati, S. K., Pelizzari, C. A., Kunnavakkam, R., and Liauw, S. L. (2014). Late toxicity and quality of life after definitive treatment of prostate cancer: redefining optimal rectal sparing constraints for intensity-modulated radiation therapy. *Cancer Med,* **3**(4), 954-961. doi: 10.1002/cam4.261

Choi, H. J., Kim, Y. S., Lee, S. H., Lee, Y. S., Park, G., Jung, J. H., Cho, B. C., Park, S. H., Ahn, H., Kim, C. S., Yi, S. Y., and Ahn, S. D. (2011). Inter- and intra-observer variability in contouring of the prostate gland on planning computed tomography and cone beam computed tomography. *Acta Oncol,* **50**(4), 539-546. doi: Doi 10.3109/0284186x.2011.562916

De Los Santos, J., Popple, R., Agazaryan, N., Bayouth, J. E., Bissonnette, J. P., Bucci, M. K., Dieterich, S., Dong, L., Forster, K. M., Indelicato, D., Langen, K., Lehmann, J., Mayr, N., Parsai, I., Salter, W., Tomblyn, M., Yuh, W. T., and Chetty, I. J. (2013). Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. *Int J Radiat Oncol Biol Phys,* **87**(1), 33-45. doi: 10.1016/j.ijrobp.2013.02.021

Ding, G. X., Duggan, D. M., Coffey, C. W., Deeley, M., Hallahan, D. E., Cmelak, A., and Malcolm, A. (2007). A study on adaptive IMRT treatment planning using kV cone-beam CT. *Radiother Oncol,* **85**(1), 116-125. doi: 10.1016/j.radonc.2007.06.015

Dunlop, A., McQuaid, D., Nill, S., Murray, J., Poludniowski, G., Hansen, V. N., Bhide, S., Nutting, C., Harrington, K., Newbold, K., and Oelfke, U. (2015). Comparison of CT number calibration techniques for CBCT-based dose calculation. *Strahlenther Onkol,* **191**(12), 970-978. doi: 10.1007/s00066-015-0890-7

Dunst, J. (2016). [IMRT experience of radiation oncologists and prognosis in head-and-neck cancer]. *Strahlenther Onkol*. doi: 10.1007/s00066-016-0976-x

Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby, S., McGale, P., Correa, C., Taylor, C., Arriagada, R., Clarke, M., Cutter, D., Davies, C., Ewertz, M., Godwin, J., Gray, R., Pierce, L., Whelan, T., Wang, Y., and Peto, R. (2011). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet,* **378**(9804), 1707-1716. doi: 10.1016/S0140-6736(11)61629-2

Eskandarloo, A., Abdinian, M., Salemi, F., Hashemzadeh, Z., and Safaei, M. (2012). Effect of object location on the density measurement in cone-beam computed tomography versus multislice computed tomography. *Dent Res J (Isfahan),* **9**(Suppl 1), S81-87.

Foroudi, F., Haworth, A., Pangehel, A., Wong, J., Roxby, P., Duchesne, G., Williams, S., and Tai, K. H. (2009). Inter-observer variability of clinical target volume delineation for bladder cancer using CT and cone beam CT. *J Med Imaging Radiat Oncol,* **53**(1), 100-106. doi: 10.1111/j.1754-9485.2009.02044.x

Fotina, I., Hopfgartner, J., Stock, M., Steininger, T., Lütgendorf-Caucig, C., and Georg, D. (2012). Feasibility of CBCT-based dose calculation: comparative analysis of HU adjustment techniques. *Radiother Oncol,* **104**(2), 249-256. doi: 10.1016/j.radonc.2012.06.007

Fotina, I., Lütgendorf-Caucig, C., Stock, M., Potter, R., and Georg, D. (2012). Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlenther Onkol,* **188**(2), 160-167. doi: 10.1007/s00066-011-0027-6

Gai, X., Wei, Y., Tao, H., Zhu, J., and Li, B. (2017). Clinical study of the time of repeated computed tomography and replanning for patients with nasopharyngeal carcinoma. *Oncotarget,* **8**(16), 27529-27540. doi: 10.18632/oncotarget.16770

Geier, M., Astner, S. T., Duma, M. N., Jacob, V., Nieder, C., Putzhammer, J., Winkler, C., Molls, M., and Geinitz, H. (2012). Dose-escalated simultaneous integrated-boost treatment of prostate cancer patients via helical tomotherapy. *Strahlenther Onkol,* **188**(5), 410-416. doi: 10.1007/s00066-012-0081-8

Giraud, P., Yorke, E., Jiang, S., Simon, L., Rosenzweig, K., and Mageras, G. (2006). Reduction of organ motion effects in IMRT and conformal 3D radiation delivery by using gating and tracking techniques. *Cancer Radiother,* **10**(5), 269-282. doi: 10.1016/j.canrad.2006.05.009

Guan, H., and Dong, H. (2009). Dose calculation accuracy using cone-beam CT (CBCT) for pelvic adaptive radiotherapy. *Phys Med Biol,* **54**(20), 6239-6250. doi: 10.1088/0031-9155/54/20/013

Guy, D., Ghanem, G., Loblaw, A., Buckley, R., Persaud, B., Cheung, P., Chung, H., Danjoux, C., Morton, G., Noakes, J., Spevack, L., Hajek, D., and Flax, S. (2016). Diagnosis, referral, and primary treatment decisions in newly diagnosed prostate cancer patients in a multidisciplinary diagnostic assessment program. *Can Urol Assoc J,* **10**(3-4), 120-125. doi: 10.5489/cuaj.3510

Hatton, J., McCurdy, B., and Greer, P. B. (2009). Cone beam computerized tomography: the effect of calibration of the Hounsfield unit number to electron density on dose calculation accuracy for adaptive radiation therapy. *Phys Med Biol,* **54**(15), N329-346. doi: 10.1088/0031-9155/54/15/N01

Jensen, S. B., Pedersen, A. M., Vissink, A., Andersen, E., Brown, C. G., Davies, A. N., Dutilh, J., Fulton, J. S., Jankovic, L., Lopes, N. N., Mello, A. L., Muniz, L. V., Murdoch-Kinch, C. A., Nair, R. G., Napenas, J. J., Nogueira-Rodrigues, A., Saunders, D., Stirling, B., von Bultzingslowen, I., Weikel, D. S., Elting, L. S., Spijkervet, F. K., Brennan, M. T., and Salivary Gland Hypofunction/Xerostomia Section, O.C.S.G, MASCC/Int., ISOO. (2010). A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: prevalence, severity and impact on quality of life. *Support Care Cancer,* **18**(8), 1039-1060. doi: 10.1007/s00520-010-0827-8

Kakoei, S., Haghdoost, A. A., Rad, M., Mohammadalizadeh, S., Pourdamghan, N., Nakhaei, M., and Bahador, M. (2012). Xerostomia after radiotherapy and its effect on quality of life in head and neck cancer patients. *Arch Iran Med,* **15**(4), 214-218. doi: 012154/AIM.008

Karlsson, T. R., Al-Azzawe, M., Aziz, L., Hurman, D., and Finizia, C. (2014). Survival outcome depending on different treatment strategies in advanced stages III and IV laryngeal cancers: an audit of data from two European centres. *Eur Arch Otorhinolaryngol,* **271**(3), 547-554. doi: 10.1007/s00405-013-2657-z

Kirkpatrick, J. P., van der Kogel, A. J., and Schultheiss, T. E. (2010). Radiation dose-volume effects in the spinal cord. *Int J Radiat Oncol Biol Phys,* **76**(3 Suppl), S42-49. doi: 10.1016/j.ijrobp.2009.04.095

Korreman, S. S., Pedersen, A. N., Nottrup, T. J., Specht, L., and Nystrom, H. (2005). Breathing adapted radiotherapy for breast cancer: comparison of free breathing gating with the breath-hold technique. *Radiother Oncol,* **76**(3), 311-318. doi: 10.1016/j.radonc.2005.07.009

Kouwenhoven, E., Giezen, M., and Struikmans, H. (2009). Measuring the similarity of target volume delineations independent of the number of observers. *Phys Med Biol,* **54**(9), 2863-2873. doi: 10.1088/0031-9155/54/9/018

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* **33**(1), 159-174.

Lieshout, H. F., and Bots, C. P. (2014). The effect of radiotherapy on dental hard tissue--a systematic review. *Clin Oral Investig,* **18**(1), 17-24. doi: 10.1007/s00784-013-1034-z

Lütgendorf-Caucig, C., Fotina, I., Stock, M., Potter, R., Goldner, G., and Georg, D. (2011). Feasibility of CBCT-based target and normal structure delineation in prostate cancer radiotherapy: multi-observer and image multi-modality study. *Radiother Oncol,* **98**(2), 154-161. doi: 10.1016/j.radonc.2010.11.016

Mayyas, E., Chetty, I. J., Chetvertkov, M., Wen, N., Neicu, T., Nurushev, T., Ren, L., Lu, M., Stricker, H., Pradhan, D., Movsas, B., and Elshaikh, M. A. (2013). Evaluation of multiple image-based modalities for image-guided radiation therapy (IGRT) of prostate carcinoma: a prospective study. *Med Phys,* **40**(4), 041707. doi: 10.1118/1.4794502

Monnier, Y., and Simon, C. (2015). Surgery Versus Radiotherapy for Early Oropharyngeal Tumors: a Never-Ending Debate. *Curr Treat Options Oncol,* **16**(9), 42. doi: 10.1007/s11864-015-0362-4

Morgan, M. A., Lewis, W. G., Casbard, A., Roberts, S. A., Adams, R., Clark, G. W., Havard, T. J., and Crosby, T. D. (2009). Stage-for-stage comparison of definitive chemoradiotherapy, surgery alone and neoadjuvant chemotherapy for oesophageal carcinoma. *Br J Surg,* **96**(11), 1300-1307. doi: 10.1002/bjs.6705

Nishioka, K., Shimizu, S., Kinoshita, R., Inoue, T., Onodera, S., Yasuda, K., Harada, K., Nishikawa, Y., Onimaru, R., and Shirato, H. (2013). Evaluation of inter-observer variability of bladder boundary delineation on cone-beam CT. *Radiat Oncol,* **8**, 185. doi: 10.1186/1748-717X-8-185

Novaes, P., Mottas, R. T., Lundgren, M., and Brazilian Society of, R. (2015). Treatment of prostate cancer with intensity modulated radiation therapy (IMRT). *Rev Assoc Med Bras,* **61**(1), 8-16. doi: 10.1590/1806-9282.61.01.008

Padmanaban, S., Boopathy, R., Kunjithapatham, B., Sukumar, P., and Nagarajan, V. (2010). A phantom study on the effects of target motion in non-gated kV-CBCT imaging. *Australas Phys Eng Sci Med,* **33**(1), 59-64. doi: 10.1007/s13246-010-0010-z

Parikh, R., and Sher, D. J. (2012). Primary radiotherapy versus radical prostatectomy for high-risk prostate cancer: a decision analysis. *Cancer,* **118**(1), 258-267. doi: 10.1002/cncr.26272

Schaake, W., van der Schaaf, A., van Dijk, L. V., Bongaerts, A. H., van den Bergh, A. C., and Langendijk, J. A. (2016). Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer patients. *Radiother Oncol*. doi: 10.1016/j.radonc.2016.04.005

Schaake, W., Wiegman, E. M., de Groot, M., van der Laan, H. P., van der Schans, C. P., van den Bergh, A. C. M., and Langendijk, J. A. (2014). The impact of gastrointestinal and genitourinary toxicity on health related quality of life among irradiated prostate cancer patients. *Radiotherapy and Oncology,* **110**(2), 284-290. doi: 10.1016/j.radonc.2013.11.011

Scherr, D., Swindle, P. W., Scardino, P. T., and National Comprehensive Cancer, N. (2003). National Comprehensive Cancer Network guidelines for the management of prostate cancer. *Urology,* **61**(2 Suppl 1), 14-24.

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull,* **86**(2), 420-428.

Siewerdsen, J. H., and Jaffray, D. A. (2001). Cone-beam computed tomography with a flat-panel imager: magnitude and effects of x-ray scatter. *Med Phys,* **28**(2), 220-231. doi: 10.1118/1.1339879

Talcott, J. A., Rieker, P., Clark, J. A., Propert, K. J., Weeks, J. C., Beard, C. J., Wishnow, K. I., Kaplan, I., Loughlin, K. R., Richie, J. P., and Kantoff, P. W. (1998). Patient-reported symptoms after primary therapy for early prostate cancer: results of a prospective cohort study. *J Clin Oncol,* **16**(1), 275-283. doi: 10.1200/JCO.1998.16.1.275

Timme, D. W., Jonnalagadda, S., Patel, R., Rao, K., and Robbins, K. T. (2015). Treatment Selection for T3/T4a Laryngeal Cancer: Chemoradiation Versus Primary Surgery. *Ann Otol Rhinol Laryngol,* **124**(11), 845-851. doi: 10.1177/0003489415588130

Tol, J. P., Doornaert, P., Witte, B. I., Dahele, M., Slotman, B. J., and Verbakel, W. F. (2016). A longitudinal evaluation of improvements in radiotherapy treatment plan quality for head and neck cancer patients. *Radiother Oncol*. doi: 10.1016/j.radonc.2016.04.011

Vaidya, J. S., Wenz, F., Bulsara, M., Tobias, J. S., Joseph, D. J., Keshtgar, M., Flyger, H. L., Massarut, S., Alvarado, M., Saunders, C., Eiermann, W., Metaxas, M., Sperk, E., Sutterlin, M., Brown, D., Esserman, L., Roncadin, M., Thompson, A., Dewar, J. A., Holtveg, H. M., Pigorsch, S., Falzon, M., Harris, E., Matthews, A., Brew-Graves, C., Potyka, I., Corica, T., Williams, N. R., Baum, M., and group, T. t. (2014). Risk-adapted targeted intraoperative radiotherapy versus whole-breast radiotherapy for breast cancer: 5-year results for local control and overall survival from the TARGIT-A randomised trial. *Lancet,* **383**(9917), 603-613. doi: 10.1016/S0140-6736(13)61950-9

van Hagen, P., Hulshof, M. C., van Lanschot, J. J., Steyerberg, E. W., van Berge Henegouwen, M. I., Wijnhoven, B. P., Richel, D. J., Nieuwenhuijzen, G. A., Hospers, G. A., Bonenkamp, J. J., Cuesta, M. A., Blaisse, R. J., Busch, O. R., ten Kate, F. J., Creemers, G. J., Punt, C. J., Plukker, J. T., Verheul, H. M., Spillenaar Bilgen, E. J., van Dekken, H., van der Sangen, M. J., Rozema, T., Biermann, K., Beukema, J. C., Piet, A. H., van Rij, C. M., Reinders, J. G., Tilanus, H. W., van der Gaast, A., and Group, C. (2012). Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med,* **366**(22), 2074-2084. doi: 10.1056/NEJMoa1112088

Wang, Y. Y., and Zhe, H. (2013). Clinical application of multimodality imaging in radiotherapy treatment planning for rectal cancer. *Cancer Imaging,* **13**(4), 495-501. doi: 10.1102/1470-7330.2013.0046

White, E. A., Brock, K. K., Jaffray, D. A., and Catton, C. N. (2009). Inter-observer variability of prostate delineation on cone beam computerised tomography images. *Clin Oncol (R Coll Radiol),* **21**(1), 32-38. doi: 10.1016/j.clon.2008.11.007

Yang, Y., Schreibmann, E., Li, T., Wang, C., and Xing, L. (2007). Evaluation of on-board kV cone beam CT (CBCT)-based dose calculation. *Phys Med Biol,* **52**(3), 685-705. doi: 10.1088/0031-9155/52/3/011

Yoo, S., and Yin, F. F. (2006). Dosimetric feasibility of cone-beam CT-based treatment planning compared to CT-based treatment planning. *Int J Radiat Oncol Biol Phys,* **66**(5), 1553-1561. doi: 10.1016/j.ijrobp.2006.08.031

Yu, T., Zhang, Q., Zheng, T., Shi, H., Liu, Y., Feng, S., Hao, M., Ye, L., Wu, X., and Yang, C. (2016). The Effectiveness of Intensity Modulated Radiation Therapy versus Three-Dimensional Radiation Therapy in Prostate Cancer: A Meta-Analysis of the Literatures. *PLoS One,* **11**(5), e0154499. doi: 10.1371/journal.pone.0154499

Zaiontz, C. (2014). (downloaded 2014) Real Statistics Using Excel. *URL: http://www.real-statistics.com/free-download/real-statistics-resource-pack/real-statistics-resource-pack-excel-2007/*

Zelefsky, M. J., Fuks, Z., Hunt, M., Yamada, Y., Marion, C., Ling, C. C., Amols, H., Venkatraman, E. S., and Leibel, S. A. (2002). High-dose intensity modulated radiation therapy for prostate cancer: early toxicity and biochemical outcome in 772 patients. *Int J Radiat Oncol Biol Phys,* **53**(5), 1111-1116.

Zurl, B., Stranzl, H., Winkler, P., and Kapp, K. S. (2010). Quantitative assessment of irradiated lung volume and lung mass in breast cancer patients treated with tangential fields in combination with deep inspiration breath hold (DIBH). Strahlenther Onkol, 186(3), 157-162. doi: 10.1007/s00066-010-2064-y

List of Figures

List of Tables

Acknowledgement

At this point I would like to thank all those who contributed with their valuable support to the completion of this work.

My first thanks goes to my supervisor PD Dr. M. N. Duma, who introduced me into the field of radiooncology and radiotherapy, which is still fascinating me. She suggested the topic of this work and gave me a lot of freedom in the course of its realization.

I am especially thankful to Professor Dr. S. E. Combs for her scientific advice and her support during completion of this thesis. Her encouragement helped me to find new enthusiasm and strength.

I am grateful to all radiooncologists who dedicated their time and efforts to the contouring work.

Much appreciation goes to Dr. Markus Oechsner, whom I could visit in his office any time with all my questions and who also contributed a huge amount of work, advice and motivation.

Many thanks also to Professor Dr. Wilkens, who gave me valuable comments and input.

I would like to thank my parents and sister for their constant and patient help and motivation. Without them, this work would have never been finished. They kept cheering me up and carrying me through times of crisis. Special thanks to my father who has been my lecturer, advisor and analyst.

Finally my profound gratitude to all who helped me through the sometimes difficult and tenacious process, who supported me in different ways, above all encouraging me to keep going and putting all their trust in me - Hamid Mahmoodnia, my friends and colleagues.