Published in final edited form as:

Pract Radiat Oncol. 2019 ; 9(4): e422–e431. doi:10.1016/j.prro.2019.02.014.

Predictive models to determine clinically relevant deviations in delivered dose for head and neck cancer

Molly M. McCulloch, M.S.^{*,1,2,3}, Choonik Lee, Ph.D.¹, Benjamin S. Rosen, Ph.D.¹, Justin D. Kamp, M.S.¹, Chrissy M. Lockhart, A.S.¹, Jae Y. Lee, M.D., Ph.D.¹, Daniel F. Polan, M.S.E.^{1,2}, Peter G. Hawkin, M.D., Ph.D.¹, Carlos J. R. Anderson, Ph.D.¹, Jolien Heukelom, M.S.⁴, Jan-Jakob Sonke, Ph.D.⁴, Clifton D. Fuller, M.D., Ph.D.⁵, James M. Balter, Ph.D.¹, Randall K. Ten Haken, Ph.D.¹, Avi Eisbruch, M.D.¹, and Kristy K. Brock, Ph.D.³

¹University of Michigan Department of Radiation Oncology, Ann Arbor, MI, USA ²University of

Michigan Department of Nuclear Engineering and Radiological Sciences, Ann Arbor, MI, USA

³The University of Texas MD Anderson Cancer Center, Department of Imaging Physics, Houston, TX, USA ⁴Netherlands Cancer Institute, Amsterdam, The Netherlands ⁵The University of Texas

MD Anderson Cancer Center, Department of Radiation Oncology, Houston, TX, USA

Abstract

Purpose: To improve the understanding of deviations between planned and accumulated dose and establish metrics to predict for clinically significant dosimetric deviations mid-way through treatment to evaluate the potential need to replan during fractionated radiation therapy (RT).

Methods and Materials: 100 head and neck cancer patients were retrospectively evaluated. Contours were mapped from the planning CT to each fraction CBCT via deformable image registration (DIR). The dose was calculated on each CBCT and evaluated based on the mapped contours. The mean dose at each fraction was averaged to approximate the accumulated dose for structures with max dose constraints. A threshold dose for structures with mean dose constraints and the daily max dose was summed to deviation value was calculated to predict for patients needing mid-treatment replanning. This predictive model was applied to 52 patients treated at a separate institution.

Results: Dose was accumulated on 10 organs over 100 patients. To generate a threshold deviation that predicted the need to replan with 100% sensitivity, the submandibular glands required replanning if the delivered dose was at least 3.5Gy higher than planned by fraction 15. This model predicts the need to replan the submandibular glands with 98.7% specificity. On the independent evaluation cohort, this model predicts the need to replan the submandibular glands with 100% sensitivity and 98.0% specificity. The oral cavity, intermediate CTV, left parotid, and inferior constrictor patients each had one patient exceeding the threshold deviation by the end of

*South Campus Research Building, 1881 East Road, Houston, TX, USA, 541-490-6492, mollyma@umich.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

RT. By fraction 15 of 30–35 total fractions, the left parotid gland, inferior constrictor, and intermediate CTV had a dose deviation of 3.1Gy, 5.9Gy, and 4.8Gy, respectively. When a DIR failure was observed, the dose deviation exceeded the threshold for at least one organ, demonstrating that an automated DIR-based dose assessment process could be developed with user evaluation for cases resulting in dose deviations.

Conclusions: A mid-treatment threshold deviation was determined to predict the need to replan for the submandibular glands by fraction 15 of 30–35 total fractions of RT.

Introduction

Approximately 50,000 patients are diagnosed with head and neck (HN) cancer in the United States every year, with the incidence of human papillomavirus (HPV) – related cases rising (1). Standard management includes chemotherapy plus radiation therapy. The combined responds in volume over the course of treatment (2). In addition, due to toxicities seen during These anatomical changes lead to deviations between the prescription (planned) dose and chemoradiation is often so effective, especially for HPV+ patients, that their tumor often treatment (i.e. mucositis), patients often lose weight during the course of chemoradiotherapy. delivered dose, which are well documented for radiation therapy; however the majority of these studies used research software not immediately available for widespread clinical use (3, 4, 5). Deviations call for plan modification during the fractionated treatment course for many head and neck cancer radiation treatment plans in order to ensure proper coverage of the tumor, to avoid overdosing of healthy tissue, and ultimately to ensure that the best treatment is delivered to the patient.

Although some deviation between the planned and delivered dose is expected and not likely clinically significant, deviations that are clinically significant need to be identified early enough in the treatment delivery process to ensure that a plan adaptation can be safely performed to correct for the deviation. The need for adaptive replanning is often left to the discretion of the physician, and a quantitative model could aid in the decision to replan. Currently, there is no existing model to predict the need to replan and uncertainty exists as to what levels of deviation from the planned dose would result in a significant deviation from the initial expectations of tumor control and normal tissue toxicity risk. The goal of this work was to perform a quality control study, using tools commercially available to determine the best estimate of the delivered dose across a large number of patients and then use this data to design a model to identify, by fraction 15 (out of 30–35 total fractions), which patients would complete treatment with a dose deviation exceeding the identified tolerance.

Methods and Materials

Patient Data

Data from 100 HN cancer patients treated with 30–35 fractions between November 2013 and October 2016 were retrospectively evaluated for this quality control IRB exempt evaluation. All patients were planned with volumetric modulated arc therapy (VMAT) using a commercial treatment planning system (Eclipse v13.6, Varian Medical Systems, Palo Alto, CA) and treated on a linear accelerator with daily kilovoltage cone beam CT (CBCT)

imaging for pre-treatment positioning. Patient and treatment planning information is detailed in Table 1a. CT-based treatment planning was performed based on the planning directive structures and goals listed in Table 2. The study focused on soft tissue organs at risk for toxicity that were within the field of view (FOV) of the CBCT.

Image Guided Radiation Therapy

Daily image guided radiation therapy (IGRT) was delivered using automatic registration of the CBCT to the planning CT using a multiple iteration technique, with the initial iterations focusing on automated intensity based alignment of the entire planning CT and the CBCT and the final iterations focusing on the custom region of interest structure that was made from the union of the spinal canal and high-risk planning target volume (PTV). Manual adjustments were permitted following the automatic registration by the treating radiation therapist. The isocenter was defined on the planning CT to be at chin level to ensure the FOV of the CBCT scan covered the anatomy of interest.

Calculation and Evaluation of Delivered Dose

The daily CBCTs were directly accessible in the commercial treatment management software for DIR using a Demon's-based contour propagation (6) and dose calculations (SmartAdapt, Varian Medical Systems, Palo Alto, CA). Figure 1 shows contours on the planning CT and the same contours propagated on the CBCT of a different fraction. Figure 2 shows the planned dose distribution and the recalculated dose on a daily CBCT. The calibrations between the CBCT and planning CT were matched using monthly measurements (QA) of standard materials with known electron density (Catphan®). The delivered dose at each fraction was calculated on the daily CBCT, and the dose was evaluated based on the propagated contours.

The following dose metrics were calculated: *Planned dose* (D_{Plan}) is the clinical prescription planned dose, based on the planning CT and original contours, calculated in the treatment planning software and exported for analysis and comparison with delivered dose.

Delivered (Fraction) dose (D_{Frac}) for each fraction is the calculation of the clinical metric (e.g. mean or max dose) based on the dose calculated on the daily CBCT evaluated for the contours propagated from the planning CT onto the CBCT with DIR.

Accumulated dose (D_{Acc}) is the accumulation of clinical metric (D_{Frac}) over all treatment fractions. As dose accumulation is not currently available in the commercial treatment planning system, the following approximation was made to estimate the accumulated dose. For mean dose metrics, the accumulation is performed by averaging the daily calculation of the mean dose. For max dose metrics, the accumulation is performed by summing the daily calculation of the maximum dose, therefore representing the most conservative estimate (e.g. the highest max dose possible to a voxel).

Model Development

D_{Plan} , D_{Frac} for each fraction, and D_{Acc} were all calculated for each patient's treatment course. D_{Acc} were calculated for each patient for fractions 1–5, 1–10, and 1–15, representing

the accumulated dose to date up to fraction 5, 10, and 15, respectively, but scaled to the total number of fractions. That is, the D_{acc} at 1–5, 1–10, and 1–15 indicate the total dose that would be delivered if the entire treatment course was delivered as indicated by the first 5, 10, or 15 fractions. This scaling allows for direct comparison of dose deviations at each time point. Recently published data (7) suggests that a deviation between planned and delivered dose of less than 3.6Gy (15% of the 24Gy planning constraint) for parotid glands would not have a significant impact on the observed toxicity in a patient population. This 15% deviation criteria was then applied as a threshold for other HN organs with a mean dose constraint. This adaptive triggering model was developed to identify dose deviations of 15% of the planning constraint of each organ at the completion of treatment early enough in treatment to enable corrective action. For CTVs, the dose deviation threshold was conservatively set at 7%, as the ICRU guideline (8) criteria is 95% - 107% relative to the prescribed dose. Table 3 depicts the dose deviation threshold between planned and delivered dose for each organ.

Cases with a completed dose deviation to any structure that exceeded the structure-specific threshold were used to build a model to predict, with 100% sensitivity, whether a patient needs replanning, specifically looking at three time points: at the completion of fraction 5, fraction 10, and fraction 15. Each completed dose was compared to the planned dose to assess if the dose deviation threshold was exceeded for each organ for each patient. For every organ with a D_{Acc} that exceeded the dose deviation threshold, the minimum deviations between the D_{Plan} and D_{Frac} for fraction 1, the average of fractions 1–5, 1–10, and 1–15 were assessed to find the minimum deviation in delivered dose at that time point to indicate that replanning was necessary to avoid completing treatment with a deviation that exceeded the specified tolerance. In some cases, due to proximity of the disease, normal tissue could not be spared at planning, dose constraints were substantially exceeded, and replanning when dose is increased would not improve toxicity risk. Based on 5-year normal tissue complication probability (NTCP) curves for the parotid glands (10), toxicity risk was evaluated at the parotid gland planning constraint (a mean dose of 24Gy), at 34Gy, and 54Gy. A mean dose of 24Gy corresponded to 20% toxicity risk, 34Gy to 30%, and 54Gy to 64%. As a 10Gy increase in mean dose corresponded to a 10% increase in toxicity, the model included organs with planned mean dose within 10Gy above the dose constraint, as we assumed that these patients could still benefit from replanning if the deviation in delivered dose exceeded the planned dose by more than the tolerance.

Data Inclusion

FOV and HU in CBCT was previously investigated (11) demonstrating a 2% accuracy could be achieved with the exclusion of the superior 7mm and tissue in the shoulder region. In the current study, data was included only for organs on CBCT slices inferior to the superior-most 7mm and for organs that were not missing material in the trapezius/shoulder region due to field-of-view limitations. For the spinal canal, where D_{max} was the metric of interest, the dose was evaluated only within the evaluable region of the CBCT.

DIR Evaluation and Validation

All DIR results were qualitatively evaluated using image fusion, evaluation of the deformation vector field, and visual assessment of the propagated structures onto the CBCT anatomy. For cases with DIR results identified as failing due to these qualitative metrics, the images were evaluated for artifacts. Images were re-registered to determine if a successful DIR result could be obtained. If a successful DIR could not be obtained, the treatment fraction was excluded from analysis and the dose accumulation was scaled based on the number of fractions with accurate DIR.

The assessment of DIR accuracy was based on contour propagation and comparison to manually drawn contours, therefore a baseline variation in the manual segmentation of normal tissues was established. Five HN patients were randomly selected for a reproducibility study. Segmentation of targets and critical organs was performed on four of the CBCTs by a radiation oncologist experienced in treating HN cancer and then a repeat segmentation was performed following a greater than two week time period, blinded to the original contours. The Dice similarity coefficient (DSC) was computed for the repeat contours for structures listed in Table 4. The DSC was computed between the CT contours and CBCT contours propagated based on rigid registration alone and DIR.

Independent Model Validation

An independent validation was performed using a cohort of 52 patients treated at the XXXX whose dose was accumulated in a collaboration between the XXXX and the XXXX.

Patient and treatment planning information is detailed in Table 1b. The original treatment plan was recalculated on the daily CT scans for each day of treatment, in order to calculate the delivered dose. The treatment isocenter was adjusted accordingly prior to dose recalculation. Using a deformation vector field (DVF) from DIR (Admire 1.04, Elekta AB, Stockholm, Sweden), the new dose distribution was mapped to the planning CT. The daily dose to each organ was calculated using the original contours on the planning CT. In the case of missing daily CTs, linear interpolation was performed on the CTs of the two closest fractions in order to create the missing fractionated doses. Nearest neighbor extrapolation was performed in the case that the daily CT from the first or last fraction was missing. The predictive model was applied to the delivered dose to 104 submandibular glands at fraction 15. The validation of the predictive model using this external data set was evaluated based on the dose deviation threshold set at 15% of the planning constraint (4.5Gy).

Results

Of the 3291 DIRs performed on 100 patients, 1% failed. These failures were due to metal artifact, beam hardening artifact, shadowing, and lack of contrast on the CBCTs. The maximum number of DIR failures for a single patient was 3. In the cases with DIR failures a deviation from the planned dose of greater than 15% was observed for at least one organ. DIR failures led to the re-registration of images, however that did not result in DIR improvement. Data due to errors in dose calculations and DIR were excluded from further analysis.

Contour Reproducibility and DIR Validation

Table 4 shows the results of the DIR validation and CBCT contour reproducibility. DIR improved DSC over rigid registration for 100% of inferior constrictors, 94% of esophagi and parotid glands, 90% of submandibular glands, and 75% of superior constrictors, spinal canals, and larynges included in the study. Based on a two-tailed paired Student's T-test, improvement in DSC was statistically significant from rigid registration to DIR in all organs. The average DSC over all CBCTs included in the study is included in the table for both the rigid registration and the DIR.

Dose Deviation Threshold

The results of setting a 15% dose deviation threshold at treatment completion are shown in Table 5. There was one patient exceeding the threshold for oral cavity, left parotid gland, intermediate CTV, and inferior constrictor, and the average deviation between D_{plan} and D_{Frac} is reported for the completion of treatment and fraction 15. For the submandibular glands, where 7 patients exceeded the threshold, the maximum deviation at completion of treatment and minimum deviation by fraction 15 are shown.

The left and right submandibular glands were combined into a single model to account for all patients with completed dose deviations exceeding the threshold. Eighty-five total submandibular glands were without DIR failures and in the evaluable region of CBCTs. With a dose deviation threshold set at 15% of the planning constraint (4.5Gy), seven submandibular glands exceeded the threshold. By fraction 15, the minimum deviation of these seven glands, using the D_{acc} at fraction 15 scaled to the total delivered dose, was 3.5Gy (maximum = 7.1Gy, median = 5.1Gy, mean = 5.3Gy). This minimum deviation threshold was set for each time point of treatment, again scaled to the total delivered dose, to evaluate predictive power for final dose deviation. The resulting false positives, sensitivities, specificities, and positive and negative predictive values (PPV and NPV) are shown in Table 6. Out of the 12 patients with evaluable inferior constrictor dose deviations, one had a potentially clinically significant dose deviation of 5.6Gy at completion of treatment. This deviation, and the need to replan, was predicted by fraction 15, with a dose deviation of 5.86Gy.

Out of the 17 patients with evaluable intermediate CTV dose deviations, one had a potentially clinically significant dose deviation of -6.7Gy at the completion of treatment. This deviation, and the need to replan, was predicted by fraction 15, with a dose deviation of -4.8Gy.

Out of the 56 patients with evaluable oral cavity dose deviations, one had a potentially clinically significant deviation of 5.2Gy. However, the early treatment fractions did not show clinically relevant dose deviations, and therefore did not predict the final dose deviation. By fraction 15, the dose deviation was only 0.81Gy.

Out of the 99 patients with evaluable spinal canal dose deviations, one had a potentially clinically significant deviation of 4.3Gy, leading to a D_{acc} of 48.1Gy, which exceeds the maximum planning constraint. Therefore, trends could not be evaluated for this organ.

Independent Model Validation

On the independent validation data set, four submandibular glands exceeded the dose deviation threshold by the end of treatment. By fraction 15, a total of six glands exceeded the 3.5Gy minimum deviation based on the original dataset; four of these glands were those that exceeded the 4.5Gy deviation threshold by the end of treatment. Independent validation predicted the dose deviation threshold by the end of treatment with 100% sensitivity (95% CI: 40%, 100%), 98% specificity (95% CI: 93%, 100%), 67% PPV (95% CI: 22%, 96%) and 100% NPV (95% CI: 96%, 100%).

Discussion

A deviation threshold was developed from a large cohort of patients to aid physicians in the decision to replan HN cancer patients. This replanning strategy has the potential to impact HN cancer patients by automatically identifying cases that could benefit from replanning by the midpoint of treatment. Treatment planning and CBCT image data from 100 HN patients were retrospectively analyzed to build a model to predict clinically relevant deviations between planned and delivered dose. For these 100 patients, 9 HN organs of interest were evaluated over 30–35 fractions of VMAT. If a mean dose deviation threshold of 3.5Gy is set by fraction 15 for the submandibular glands, the need for replanning can be predicted with 98.7% specificity providing 87.5% PPV, and 100% NPV. Applying this model to an independent validation set from a different institution using a different DIR algorithm predicted the need for replanning with 100% sensitivity, 98% specificity, 67% PPV and 100% NPV. The inferior constrictor, intermediate CTV, oral cavity, and left parotid glands only had one patient exceeding the corresponding dose deviation threshold by the end of treatment. With the exception of the oral cavity, all of these patients would be flagged by fraction 15 of treatment using this model. Trends could not be evaluated for spinal canal as only one patient had a substantial deviation. The spinal canal was used for initial alignment of the CBCTs to CTs, likely influencing consistent delivery of planned dose.

As demonstrated in Table 6, the confidence intervals for specificity are relatively narrow (resulting from the majority of subjects not requiring replanning), while the confidence intervals for sensitivity are relatively wide (since the number of positives, i.e., subjects requiring replanning, is small in both cohorts). Although the model achieves 100% sensitivity, it's difficult to make strong statistical claims due to the limited number of cases that require replanning, indicating the need to further validate this model in a prospective clinical trial. Assuming that the true proportion of observations requiring treatment replanning is 8%, a minimum sample size of 100 subjects will be required to achieve 80% power to reject the null hypothesis that the sensitivity is 70% under the alternative that the sensitivity is 98%, using an exact test for a single proportion with one-sided significance level 0.05. This sample size is also sufficient to detect a change in the value of specificity from 85% to 95% (assuming 92 plans do not require replanning), also using a one-sided test with level 0.05.

Although true dose accumulation using an accurate DIR method is ideal, it has been shown (12) that the population differences are small (the difference between the population average of mean PG doses and the population average of DIR mean PG doses was 0.3Gy) and did

not impact the assessment of replanning strategies. This study used manual delineations on CT and used those contours for DIR and to propagate the dose. The current study used DIR between CT and CBCT to propagate the contours, which could potentially increase uncertainties in the estimated delivered dose. The impact of uncertainties in the DIR algorithm on the assessment of delivered dose requires further investigation. Limited data exists in the literature on how uncertainties in DSC translate to uncertainties in dose. A recent study reported on the impact of the accumulated minimum dose as a function of variation in DIR (13). This study demonstrated that, in order to impact the minimum dose, a combination of low contour DSC in a high dose gradient would have to occur. A maximum difference of 2.5 Gy in the minimum dose was observed. The current study reported on the mean dose to each organ and would therefore be less sensitive to low DSC and high dose gradients.

Substantial DIR error occurred in a limited number of cases, 1%, and was attributed to artifact, shadowing effects, beam hardening artifact, shadowing, and lack of contrast on the CBCTs. The potential to develop a completely automated dose accumulation process is limited by the possibility of DIR errors and the potential need to evaluate the results in real time. Dose deviations from all cases with substantial DIR failure were evaluated and all of them were found to lead to reported dose deviations of greater than 15% between planned and delivered dose for at least one organ. This data suggest that it may be possible to develop an automated process and perform DIR evaluation only on those cases that result in a dose deviation greater than 15% (10% of cases). This requires further testing on additional patients.

A previous study (11) determined that FOV differences between CT and CBCT lead to dose calculation error up to 80% on the superior and inferior edge inside the CBCT FOV. For dose calculations to not exceed 2% error, organs in the superior 7mm and the shoulder regions of the CBCTs were excluded in evaluations. Future work must address this limitation to ensure a robust method is available to calculate the daily dose to all structures of interest. Augmenting the missing data from the CT or deforming the CT to the CBCT are potential methods to improve these dose calculations.

The dose deviation threshold was based on research that stated that dose changes of less than 3.6Gy (15% of the planning constraint) to parotid glands would not have a significant impact in a population. As similar data for other soft tissue structures with mean dose constraints does not exist, this 15% was applied as the dose deviation threshold for the other HN organs with meaningful dose deviation has not been reached across the radiation oncology community. mean dose planning constraints evaluated in this study. However, consensus on a clinically.

Conclusion

With the use of this model, HN cases that would benefit from replanning could be identified. For submandibular glands, a dose deviation threshold of 3.5Gy at fraction 15 can predict the need to replan a patient. Therefore, a model was developed to aid physicians in the decision to replan HN patients based on the submandibular glands by fraction 15 of radiation

treatment. At least one organ exceeded the dose deviation threshold for all CBCTs in which a DIR failure was identified, indicating the possibility of an automated process to perform DIR evaluation only for cases that would result in a dose deviation. Additionally, the accuracy of the DIR algorithm in the commercial treatment management software was evaluated using contour propagation.

Acknowledgements –

the authors thank Dr. Matthew Schipper (The University of Michigan) and Dr. Christine Peterson (The University of Texas MD Anderson Cancer Center) for help with statistical analysis.

Dr. Ten Haken reports grants from NIH, during the conduct of the study; non-financial support from Varian Medical Systems, outside the submitted work.

Daniel Polan reports other from Varian Medical Systems, Inc, during the conduct of the study.

Dr. Brock reports a co-development agreement with Varian Medical Systems, Inc, during the duration of the study; In addition, Dr. Brock has a Licensing Agreement with RaySearch Laboratories.

Dr. Sonke reports grants from LSH-TKI Foundation, grants from EU FP7, during the conduct of the study; grants from Elekta Oncology Systems Ltd, other from Elekta Oncology Systems Ltd, outside the submitted work; In addition, Dr. Sonke has a patent Computed Tomography Scanning with royalties paid to Elekta AB, a patent Motion Artefact Reduction in CT Scanning issued to NA, a patent Methods and System for Protecting Critical Structures During Radiation Treatment with royalties paid to Elekta AB, and a patent Radiotherapy and Imaging Methods and Apparatus with royalties paid to Elekta AB.

This research is supported by the Andrew Sabin Family Foundation; Dr. Fuller is a Sabin Family Foundation Fellow. Dr. Fuller receives funding and salary support from the National Institutes of Health (NIH), including: the National Institute for Dental and Craniofacial Research Award (1R01DE025248-01/R56DE025248-01); a National Science Foundation (NSF), Division of Mathematical Sciences, Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant (NSF 1557679); the NIH Big Data to Knowledge (BD2K) Program of the National Cancer Institute (NCI) Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825-01); NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148-01); an NIH/NCI Cancer Center Support Grant (CCSG) Pilot Research Program Award from the UT MD Anderson CCSG Radiation Oncology and Cancer Imaging Program (P30CA016672) and an NIH/NCI Head and Neck Specialized Programs of Research Excellence (SPORE) Developmental Research Program Award (P50 CA097007-10).

Dr. Fuller has received direct industry grant support and travel funding from Elekta AB.

This work was funded in part by NIH R01-CA-184153, Functional MRI utilized to improve local-regional control of poor-prognosis head and neck cancer.

This work was funded in part by P01-CA059827.

References

- (1). Gillison ML, et al. Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma. *Journal of Clinical Oncology*. 2015;33(29).
- (2). Salama JK, Haddad RI, Kies MS, et al. Clinical practice guidance for radiotherapy planning after induction chemotherapy in locoregionally advanced head and neck cancer. *Int J Radiat Oncol Biol Phys* 2009;75(3):725–33. [PubMed: 19362781]
- (3). Risholm P, Balter J, Wells WM. Estimation of delivered dose in radiotherapy: the influence of registration uncertainty. *Med Image Comput Assist Interv*. 2011;14(Pt 1):548–55.
- (4). Vineberg KA, Hunter KU, Schipper MJ, et al. The influence of rotations on delivered dose in oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2012; 84(3) (Suppl)S81.
- (5). Orlandini LC, Coppola M, Fulcheri C, et al. Dose tracking assessment for image-guided radiotherapy of the prostate bed and the impact on clinical workflow. *Radiat Oncol* 2017; 12(1): 78. [PubMed: 28454559]

- (6). Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 2005;50(12):2887–2905. [PubMed: 15930609]
- (7). Hunter KU, Fernandes LL, Vineberg KA, et al. Parotid glands dose-effect relationships based on their actually delivered doses: implications for adaptive replanning in radiation therapy of head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2013;87(4):676–82. [PubMed: 24035328]
- (8). Purdy JA, Current ICRU definitions of volumes: limitations and future directions. *Seminars in Radiation Oncology* 2004; 14(27–40).
- (9). Geng H, et al. Knowledge Engineering-Based Quality Evaluation of NRG Oncology RTOG 0522 Treatment Plans. *Int J Radiat Oncol Biol Phys*. 2017;99(2):S174.
- (10). Braam PM, Roesink JM, Moerland MA, et al. Long term parotid gland function after radiotherapy. *Int J Radiat Oncol Biol Phys*. 2005;62(3):659–64. [PubMed: 15936542]
- (11). XXXX
- (12). Zhang P, Simon A, Rigaud B, et al. Optimal adaptive IMRT strategy to spare the parotid glands in oropharyngeal cancer. *Radiat Oncol*. 2016;120(1):41–47.
- (13). Samavati N, Velec M, Brock K. Effect of deformable registration uncertainty on lung SBRT dose accumulation. *Med Phys*. 2016;43(1):233–240. [PubMed: 26745916]

Figure 1.
Example of contours at planning, fraction 10, and fraction 30 of one patient, showing the deformed structures.

Figure 2.
Example of the planning dose on the planning CT (top) and the dose distribution of the same plan calculated on the fraction 30 CBCT (bottom).

Table 1.

Patient characteristics for the original patient cohort (1a) and independent evaluation cohort (1b). Human papillomavirus is denoted as HPV. Clinical target volume is denoted as CTV. Planning target volume is denoted as PTV. Gross tumor volume is denoted as GTV.

Table 1a. Patient Characteristics - Original patient cohort	
Patients (n)	100
Gender (male/female) (n)	80/20
Diagnosis site (n)	
Pharynx	58
Larynx	21
Lip and Oral Cavity	16
Other	5
HPV status [positive/negative/unrecorded] (n)	59/13/28
CTV [mean (range)] (cm ³)	154 (1–725)
CTV to PTV margin (mm)	3
Prescribed dose [mean (range)] (Gy)	68 (60–72)
Fractions [35/30/other] (n)	66/17/17
Patients with boost (n)	10
Patients with replans (n)	16
Elapsed Tx Time [median (range)] (days)	49 (39–74)
IV contrast used in CT (Y/N)	78/22
10% Weight Loss over entire Tx (n)	25
5% Weight Loss over entire Tx (n)	67

Table 1b. Patient Characteristics - Independent Evaluation set	
Patients (n)	52
Gender (male/female) (n)	36/16
Diagnosis site (n)	
Pharynx	27
Sinonasal	12
Oral Cavity	5
Larynx	1
Other	7
HPV status [positive/negative/unrecorded] (n)	10/10/32
GTV [mean (range)] (cm ³)	45(4–171)
Prescribed dose [mean (range)] (Gy)	67(60–70)
Fractions [35/33/32/30] (n)	14/27/1/10
Patients with boost (n)	13
Patients with replans (n)	13

Table 2.

Planning directives for normal tissue structures used in the initial planning. The mean and max dose refers to the dose metric used for that organ.

Table 2. Normal Tissue Structures and Goals	
Structure	Target Value (Gy)
Left/Right Submandibular Gland	Mean Dose < 30
Larynx	Mean Dose < 20
Inferior Constrictor	Mean Dose < 20
Superior Constrictor	Mean Dose < 50
Oral Cavity	Mean Dose < 30
Left/Right Parotid Gland	Mean Dose < 24
Spinal Canal	Max Dose < 45

Table 3.

Planning and dose deviation thresholds for the organs at risk and the target volumes. Prescription dose is denoted as Rx. The dose deviation threshold is based on 15% of the planning constraint for all tissue with the exception of the target volume and the spinal canal. The threshold for spine is based on the maximum clinical constraint being 48Gy (9).

Organ	Dose Metric	Planning Constraint [Gy]	Dose Deviation Threshold [Gy]
Inferior Constrictor	Mean	20	3
Superior Constrictor	Mean	50	7.5
Spinal Canal	Max	45	3
High risk CTV	D95%	$\pm 7\%$ of Rx	4.10–5.18
Intermediate risk CTV	D95%	$\pm 7\%$ of Rx	2.83–5.17
Larynx	Mean	20	3
Oral Cavity	Mean	30	4.5
Left/Right Parotid glands	Mean	24	3.6
Left/Right Submandibular glands	Mean	30	4.5

Table 4:

Results of DIR validation study and CBCT contour reproducibility (mean and standard deviation). The p-value is based on a two-tailed paired Student's T-test. Manual contour reproducibility was based on results of a resident radiation oncologist manually delineating structures on a CBCT, then repeating at a later date without reference to the original delineation. The submandibular glands are abbreviated to Sub. Glands in this table.

Structure	Registration Evaluation			Manual Contour Reproducibility
	Rigid Registration DSC	DIR DSC	T-Test P-Value	DSC
Spinal Canal	0.81±0.04	0.85±0.03	< 0.01	0.93±0.02
Sub. Glands	0.70±0.08	0.76±0.11	< 0.01	0.87±0.02
Larynx	0.65±0.10	0.74±0.10	< 0.01	0.90±0.03
Parotid Glands	0.64±0.09	0.71±0.08	< 0.01	0.78±0.05
Sup. Constrictor	0.53±0.11	0.61±0.08	< 0.01	0.78±0.03
Esophagus	0.41±0.10	0.58±0.09	< 0.01	0.63±0.18
Inf. Constrictor	0.40±0.19	0.60±0.10	< 0.01	0.64±0.16

Table 5.

Number of organs with clinically relevant dose deviations and values of those deviations. The total number of organs were those that originally had dose calculations done for the study. The organs included in the model were those within the evaluable region of the image and not on images for which DIR failures were observed.

Organ	Planning Constraint [Gy]	Dose Deviation Threshold [Gy]	Total Number of Organs	Organs Included in Model	Organs Exceeding Deviation	Maximum Deviation at Completion of Tx [Gy]	Minimum Deviation by Fx15 [Gy]
Spinal Canal	45	3	99	99	1	3.1	NA
Left/Right SGs	30	4.5	176	85	7	8.22	3.5
Sup. Constrictor	50	7.5	100	60	0		
Oral Cavity	30	4.5	100	56	1	5.18	0.81
High CTV	$\pm 7\% * Rx$	4.10–5.18	103	43	0		
Left Parotid	24	3.6	100	37	1	3.77	3.08
Right Parotid	24	3.6	99	34	0		
Int. CTV	$\pm 7\% * Rx$	2.83–5.17	101	17	1	–6.65	–4.84
Inf. Const	20	3	97	12	1	5.62	5.86

Table 6.

Model based on submandibular glands (SG). The minimum dose deviation for Fx 1–15, 1–10, 1–5, and 1 is based on the minimum dose deviation by fraction 15 that leads to a dose deviation exceeding the dose deviation threshold at completion of treatment. L/R refers to left/right. CI refers to confidence interval.

Organ	Total SGs (N)	Dose Deviation Threshold [Gy]	Organs Exceeding Dose Deviation Threshold	Test (name)	Fx 1–15 Dose Deviation [Gy]	95% CI	Fx 1–10 Dose Deviation [Gy]	Fx 1–5 Dose Deviation [Gy]	Fx 1 Dose Deviation [Gy]
L/R SGs	85	4.5	7	Min. [Gy]	3.5		3.5	3.5	3.5
				False Pos.	1		3	2	10
				Sens.	1.00	0.59, 1.00	100	57.1	28.6
				Spec.	.99	0.93, 1.00	96.2	97.4	87.2
				PPV	.88	0.47, 1.00	70	66.7	16.7
				NPV	1.00	0.95, 1.00	100	96.2	93.2