



Published in final edited form as:

*Int J Radiat Oncol Biol Phys.* 2021 July 01; 110(3): 883–892. doi:10.1016/j.ijrobp.2021.01.007.

## Early prediction of acute esophagitis for adaptive radiotherapy

Sadegh R Alam, PhD<sup>a,\*</sup>, Pengpeng Zhang, PhD<sup>a</sup>, Si-Yuan Zhang, MD<sup>c</sup>, Ishita Chen, MD<sup>b</sup>, Andreas Rimner, MD<sup>b</sup>, Neelam Tyagi, PhD<sup>a</sup>, Yu-Chi Hu, PhD<sup>a</sup>, Wei Lu, PhD<sup>a</sup>, Ellen D Yorke, PhD<sup>a</sup>, Joseph O Deasy, PhD<sup>a</sup>, Maria Thor, PhD<sup>a</sup>

<sup>a</sup>Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, United States

<sup>b</sup>Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, United States

<sup>c</sup>Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing, China

### Abstract

**Purpose:** Acute esophagitis (AE) is a common dose-limiting toxicity in radiotherapy (RT) of locally advanced non-small cell lung cancer (LA-NSCLC). We developed an early AE prediction model from weekly accumulated esophagus dose and its associated local volumetric change.

**Methods:** 51 LA-NSCLC patients were treated via IMRT to 60Gy in 2 Gy-fractions with concurrent chemotherapy and had weekly cone-beam CTs (CBCTs). Twenty-eight patients (55%) developed grade2 AE (AE2) at a median of 4 weeks post-RT start. For early AE2 prediction, esophagus on CBCTs of the first two weeks were deformably registered to the planning CTs and weekly esophagus dose was accumulated. Week1-to-week2 esophagus volume changes including maximum esophagus expansion (MEex%) and volumes with x% local expansions (VEx%; x=5, 10, 15) were calculated from the Jacobian map of deformation vector field gradients. Logistic regression model with 5-fold cross-validation was built using combinations of the accumulated Mean Esophagus Doses (MED) and the esophagus change parameters with the lowest p-value in univariate analysis. The model was validated on additional 18 and 11 patients with weekly CBCTs and MRIs, respectively, and compared to models using only planned mean dose (MED<sub>Plan</sub>). Performance was assessed using AUC and Hosmer-Lemeshow test ( $P_{HL}$ ).

**Results:** Univariately, w1→w2 VE10% ( $p=0.004$ ), VE5% ( $p=0.01$ ) and MEex% ( $p=0.02$ ) significantly predicted AE2. A model combining MED<sub>w2</sub> and w1→w2 VE10% had the best performance (AUC=0.80;  $P_{HL}=0.43$ ), while the MED<sub>Plan</sub> model had a lower accuracy (AUC=0.67;

\*Corresponding author: Sadegh R Alam, PhD, Diagnostic Imaging Physics Resident, Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, 2<sup>nd</sup> floor, New York, NY 10017, Tel: +1929-302-6750, riyahiam@mskcc.org.

**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.

**Conflict of interest:** none

**Disclosure:** Master research agreement between Memorial Sloan Kettering Cancer Center and Varian. NCI/NIH P30 CA 008748 and R01 CA 198121.

**Data Sharing:** Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

$P_{HL}=0.26$ ). The combined model also showed high accuracy in the CBCT (AUC=0.78) and MRI validations (AUC=0.75).

**Conclusion:** A CBCT-based cross-validated and internally validated model on MRI with a combination of accumulated esophagus dose and local volume change from the first two weeks of chemo-RT significantly improved the AE prediction compared to conventional models using only the planned dose. This model could inform plan adaptation early to lower the risk of esophagitis.

## Introduction

Given the proximity between the esophagus and the commonly involved nearby lymph nodes in concurrent chemo-radiotherapy (CRT) for locally-advanced non-small cell lung cancer (LA-NSCLC), moderate to severe acute esophagitis (AE) is the most common significant acute toxicity (1). The development of AE severely impacts patients' quality of life and may even lead to treatment interruption or early treatment termination (2). Identifying predictors of AE prior to its onset may allow early intervention or adaptation to minimize or avoid the development of severe AE (3). Traditional AE predictions assume that the esophagus dose and geometry remain constant during treatment since the planned esophagus dose is used for modeling the prediction of AE and calculating the normal tissue complication probability (4). While such a model using the mean esophagus dose ( $MED_{plan}$ ) has been successfully validated in external data, the associated discrimination, *i.e.* AUC=0.65 (1), is low and limits its wide clinical applicability to better identify patients at AE risk during RT.

Recently frameworks for predicting AE have been suggested using CT (5) or MRI (6). It has been shown that radiation-induced AE (5–8) may cause preclinical esophageal inflammation/edema that can be measured as an expanded esophageal wall (>5mm) (9,10) and this expansion can be accurately quantified using Maximum Esophagus Expansion derived from spatially localized metrics (5,6). The goal of this study was to develop an early AE prediction model within a clinically meaningful time frame that may still allow an adjustment of the treatment plan if indicated, by incorporating esophageal expansion via longitudinal imaging studies with clinically feasible imaging modalities. In addition to accumulated MED, we incorporated locally derived esophagus expansion based on CBCTs acquired during the first two weeks of CRT into a model to perform early prediction of grade 2 AE (AE2). Eventually the proposed model was compared to conventional models using only planned mean dose and validated using internal validation with set aside CBCT and MRI data.

## Materials and Method

Our prediction workflow is illustrated in Fig. 1. The esophagus on CBCTs of the first two weeks were deformably registered to the planning CTs (pCT) and the weekly dose was accumulated to the esophagus on the planning coordinate system ( $MED_{w1,w2}$ ). Subsequently, week2 CBCTs were deformably registered to week1 and  $w1 \rightarrow w2$  local volume change parameters were calculated using the Jacobian map (6). Parameters quantifying esophagus volume changes along with  $MED_{w1,w2}$  were then used to build AE2

prediction model. The process was repeated for the other internally set aside weekly CBCTs and MRIs and the model was validated using those data.

## Dataset

The training set included a cohort of 51 LA-NSCLC patients consecutively treated between June 2016 and March 2019 with intensity-modulated RT in 2Gy daily fractions in a five days/week fractionation (total dose 60Gy) with concurrent platinum-doublet chemotherapy. (hereafter referred to as CBCT training cohort).

The validation cohort consisted of 2 internally set aside datasets: 1)18 LA-NSCLC patients with weekly CBCTs treated with the same criteria as the training cohort (hereafter referred to as CBCT validation cohort) and 2)11 LA-NSCLC patients enrolled in an Institutional Review Board approved study to undergo weekly MRI during RT (Protocol# 15–073). The MRI patients were also treated via intensity-modulated RT in 2Gy daily fractions in a five days/week fractionation (2016–2018, total median dose 60Gy; hereafter referred to as MRI cohort). One MRI patient was treated with preoperative chemo-RT at 1.8Gy/fraction (50.4Gy total). Respiratory triggered (at exhale) T2w MRIs (TR/TE=3000–6000/120ms, 43 slices, NSA=2, FOV=300×222×150mm) were acquired on a Philips Ingenia 3 Tesla scanner (Philips Healthcare, Amsterdam, Netherlands) with vendor geometric distortion corrections and resolution of 0.85×0.85×3.5mm<sup>3</sup>.

Concurrent chemotherapy regimens included taxol/carboplatin (44%), pemetrexed/carboplatin (28%), pemetrexed/cisplatin (20%) and etoposide/cisplatin (8%). All patients had daily kV orthogonal radiographs for daily setup and weekly CBCTs (matched on spine) to monitor positional uncertainties and tumor changes (Varian Linac TrueBeam; Varian Medical Systems, Palo Alto, CA). At simulation, both cohorts were immobilized and a free-breathing pCT scan (Philips Big Bore) was acquired for treatment planning (Eclipse; Varian Medical Systems) and a 4DCT to define Internal Target Volume and evaluate tumor motion. The resolution of the pCTs and CBCTs were 1.17×1.17×3mm<sup>3</sup> and 0.98×0.98×3mm<sup>3</sup>, respectively. A total of 471 weekly image sets (405 CBCT, 66 MRI) were acquired.

The esophagus and Gross Tumor Volume (GTV) on the pCTs, weekly CBCTs and weekly MRIs were contoured (Varian Eclipse) by an experienced radiation oncologist (RO) and represented the ground-truth contour. The esophagus on CBCTs was contoured according to anatomic and contouring atlases of organs at risk (11,12). The esophagus was outlined from the level below the cricoid to the stomach entrance at the GE junction. The internal air cavities and the surrounding mediastinal fat allowed for identifying the esophageal wall. Altering the CT window/level assisted the visualization and delineation of the boundary of the esophagus in continuous coronal and sagittal views. At the locations where distinguishing the esophageal boundary with adjacent mediastinal lymphadenopathy was challenging (especially subcarinal lymph node), the pCT was used as a reference. To assess the reliability/accuracy of the manual contours, the RO blindly re-delineated the contours for a random subset of 10 patients and the intra-observer delineation variation was reported.

Acute esophagitis was assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) v.4.0 (13). A total of 28 (55%), 10 (56%) and 8 patients (73%) developed

Grade 2 (AE2) in the CBCT training, CBCT and MRI validation cohorts, respectively at a median of four weeks after RT start. We focused on AE2, because AE of at least grade 2 is considered a clinically relevant endpoint for LA-NSCLC patients (2), since it is defined as esophagitis requiring medical intervention (13), frequently with strong pain medications such as liquid opioids. Grade 3 AE on the other hand is defined as AE requiring hospitalization or placement of a PEG tube which is clinically quite severe and thus every effort should be made to avoid patients developing such severe AE. In our dataset, only 1 patient experienced grade 3 AE (i.e. 1.6%). Patient characteristics are summarized in Table 1.

### Deformable dose accumulation

In this study, the entire dose accumulation process is performed using the planning coordinate system as the reference frame. The following workflow demonstrates the deformable dose accumulation procedure.

1. Week1 and week2 CBCT (wCBCTs) along with their expanded esophagus contours (delineated by RO as ground-truth) were rigidly registered/propagated ( $Eso_{GT}$ ) to the initial pCT coordinates using the transformations recorded in the clinical offline review system, in which the therapist matched the spine in the CBCT to the pCT.
2. The esophagus dose at each voxel was calculated to  $Eso_{GT}$  using the prescribed planning dose; this represented the ground-truth weekly delivered dose.
3. Then, to calculate the weekly deformable dose at each expanded esophagus voxel, corresponding points between the normal esophagus on pCT and the expanded esophagus on the rigidly registered wCBCT were linked based on voxel-by-voxel correspondence built via Deformable Image Registration (DIR; ref. next section for DIR method) performed between pCT and wCBCT (wCBCT $\rightarrow$ pCT) to account for the weekly esophagus structural changes (e.g. deformation, shift etc.).
4. Eventually, the prescribed dose map in the planning coordinate was scaled to 6 weeks and for each week, the dose from the previous week was added to the current week, where the accumulated deformable dose was calculated (step 3) for each weekly esophagus, allowed deformable dose accumulation over the treatment course.

To perform dosimetric evaluation of DIRs, at each week we generated a DIR derived esophagus contour using the obtained corresponding points (step 3) where the deformable accumulated dose-volume parameters i.e. MED, Max dose and  $D5_{cc}$  were compared to the ground-truth contour  $Eso_{GT}$ .

### Parameters quantifying local volume change of esophagus

Weekly local volumetric esophagus change was quantified using Jacobian maps (J) calculated from a multi-resolution B-spline regularized diffeomorphic registration (6,14) performed between the week2 (deformed) and week1 images (reference) *i.e.* w2CBCT $\rightarrow$ w1CBCT. B-spline mesh size was 32mm at the coarsest level and was reduced

by a factor of two at each sequential level. The optimization step size was set to 0.2 with the number of iterations (100, 70, 30) at each level. (detailed registration method in (6) and (14)).  $J$  was calculated at each voxel as the determinant of the gradient of the deformation vector field that measured the ratio of local volume change, where  $J > 1$  indicates local volume expansion,  $0 < J < 1$  shrinkage and  $J = 1$  no change (15). Consequently, voxel-wise percentage of tissue expansion/shrinkage can be calculated by  $[(J-1) \times 100]$  (e.g.  $J = 0.8$  and  $J = 1.1$  are equal to 20% shrinkage and 10% expansion) and the Jacobian integral defined as  $[(\text{Mean}(J) - 1) \times \text{week1 esophagus volume}]$  measured the net local volume change of esophagus between the two weeks (15,16). Parameters quantifying esophagus volume changes were: the net volume change (VC%), maximum esophagus expansion (MEex%) *i.e.* the average voxel-wise percentage of tissue expansion calculated in a  $3 \times 3 \times 3$  voxel of cuboid region ( $0.1 \text{ cm}^3$ ) centered at the highest  $J$  voxel (6), and volumes with 5%, 10% and 15% local expansions (VE5%, VE10%, VE15%) at week2 with respect to week1 ( $w1 \rightarrow w2$ ).

Registrations were geometrically evaluated by Dice Similarity Coefficient (DSC) and Hausdorff distance (HD) between the deformably registered esophagus contours and the ground-truth contours by RO. To validate volume change parameters and the Jacobian map, we calculated the differences between the esophagus local volume change calculated using Jacobian integral versus the volume change calculated from the ground-truth contours (6,16).

### AE2 model construction (CBCT training cohort)

For an AE2 prediction model to be useful in an adaptive setting, early AE indications are of primary interest. Therefore, only  $\text{MED}_{w1}$ ,  $\text{MED}_{w2}$  (accumulated doses) and esophagus volume change parameters (VC%, MEex% and VE5%, VE10%, VE15%) from the first two weeks ( $w1 \rightarrow w2$ ) in CBCT training cohort were considered candidate features. Univariate  $p$ -values and Area Under the Receiver Operating Characteristic Curve (AUCs) were assessed for each feature using Wilcoxon rank sum test.  $P$ -values were adjusted using Bonferroni correction for multiple feature comparison ( $p < 0.05$  was considered statistically significant). Two features including the accumulated mean dose and  $w1 \rightarrow w2$  volume change parameter with the lowest  $p$ -values in the univariate analysis were sequentially fed into a nested multivariate logistic regression model with 5-fold cross-validation repeated five times to prevent model overfitting and to compare the dose only model against the combined dose plus volume change model. The combined model was considered the final AE2 prediction model. Using the same training process, another cross-validated univariate logistic model was also built incorporating only  $\text{MED}_{\text{plan}}$  from the treatment plan.

The optimal cut-off values to calculate the accuracy was computed using optimal Youden Index (17,18) and 95% confidence interval of AUC (CI) was calculated with 2000 stratified bootstrap replicates (19). The performance of the models was quantified using the median AUC and  $p$ -values from Hosmer-Lemeshow (20) ( $P_{HL}$ ) and Wald tests (21) ( $P_w$ ) to assess model calibration and parameters significance in predicting AE2, respectively. RStudio (22) was used for construction of the prediction model.

## AE2 model validation (CBCT validation and MRI cohorts)

The final model trained on the CBCT training cohort was validated using both CBCT validation and MRI data in which the regression coefficients from the trained model were applied to the corresponding features in the validation data.

## Results

The volumetric change of the esophagus at the end of treatment relative to the start ranged from 40% shrinkage to 82% expansion. Average DSC and maximum (95<sup>th</sup> percentile) HD calculated from the intra-observer variation of contour delineation by the RO on CBCT were  $0.89 \pm 0.02$  and 4.1mm (1.9mm) for the first two weeks, respectively.

Average DSC for w2CBCT→pCT and w2CBCT→w1CBCT registrations were  $0.85 \pm 0.05$  and  $0.81 \pm 0.04$  with maximum (95<sup>th</sup> percentile) HD of 5.9mm (1.7mm) and 6.1mm (1.9mm), respectively. Similarly, the DSC for w2MRI→pCT and w2MRI→w1MRI registrations were  $0.79 \pm 0.05$  and  $0.80 \pm 0.03$  with HD of 6.3mm (3.0mm) and 6.2mm (2.9mm), respectively. (Supplemental Fig.1)

Average dosimetric differences between w2CBCT→pCT registrations vs. dose accumulated in ground-truth contour were: MED= $0.14 \pm 0.4$ Gy (1.6% of population mean MED<sub>w2</sub>=8.7Gy), Max dose= $0.01 \pm 0.1$ Gy and D5<sub>cc</sub>= $0.1 \pm 0.6$ Gy ( $p > 0.2$ ). (Supplemental Fig.1) A reasonable correlation between the Jacobian integral method (i.e. voxel-wise local volume change) and the ground-truth manual segmentation was seen (Supplemental Fig. 2) and the agreement between the two methods evaluated using Bland-Altman plot (23) showed 5.8% (wCBCTs) and 4.2% (wMRIs) mean absolute percentage differences. (Supplemental Fig. 2)

## Prediction of AE2

In univariate analysis, w1→w2 VE10% ( $p=0.004$ ), VE5% ( $p=0.01$ ) and MEex% ( $p=0.02$ ) were statistically significant predictors of AE2 ( $p < 0.05$  after Bonferroni correction; Table 2).

Combining MED<sub>w2</sub> and w1→w2 VE10% in a multivariate model, showed the best performance (Sensitivity=79%, Specificity=83%, AUC=0.80 (95% CI: 0.68–0.92) with good calibration ( $P_{HL}=0.43$ ; Fig.2A).  $P_w(\text{MED}_{w2})$  and  $P_w(\text{VE10}\%)$  were 0.02 and 0.007 respectively, indicating both parameters added impact to the prediction. The performance of using MED<sub>plan</sub> alone was considerably lower (Sensitivity=89%, Specificity=43%, AUC=0.67 (95% CI:0.48–0.80);  $P_{HL}=0.26$ ) (Fig.2B) with  $P_w(\text{MED}_{plan})=0.14$ . When validated on the CBCT validation and MRI data, the multivariate model showed high accuracy (Sensitivity=80%, Specificity=75%, AUC=0.78 (95% CI:0.53–1.0) and (Sensitivity=63%, Specificity=100%, AUC=0.75 (95% CI: 0.46–1.0), respectively. The MED<sub>w2</sub> only model showed similar performance to MED<sub>plan</sub> (AUC=0.69; Supplemental Fig.3 for ROC curves)

As presented in the dose-response curves in Fig.2D, MED<sub>plan</sub> model had high sensitivity (89%), but due to extremely low specificity (43%), wasn't able to demonstrate an accurate

prediction and showed similar steepness (95% CI) to a linear model (0.01–0.17). The proposed model fed with  $MED_{w2} + w1 \rightarrow w2$  VE10% had higher specificity (83%), and better described the AE2 rates, in particular moderate to high risk regions (Fig.2C, AE2[%] 40%) with greater steepness (95% CI) for  $MED_{w2}$  (0.12–0.8) and VE10% (0.3–1.4). According to the proposed model,  $w1 \rightarrow w2$  VE10%  $>2.8\text{cm}^3$  and  $MED_{w2} >6.0$  Gy corresponded to a 50% complication probability of AE2.

Fig.3A–C shows univariate AE2 predictability of ‘dose-volume change’ features where  $w1 \rightarrow w2$  VE10% clearly performed better than the mean dose parameters in both CBCT and MRI cohorts. The population average  $w1 \rightarrow w2$  VE10% and  $MED_{w2}$  for AE2 vs. non-AE2 patients were  $2.3\text{cm}^3$  vs.  $0.7\text{cm}^3$  and  $9.1\text{Gy}$  vs.  $7.7\text{Gy}$ , respectively.

Scatter plots in Fig.3D verifies that the combined model (left) could outperform mean esophageal dose from the initial treatment plan (right) in predicting grade 2 esophagitis. Using  $MED_{plan}$  model, the annotated AE2 outlier case (P1) was wrongly predicted as a non-AE2 due to very low  $MED_{plan}$  (5.5Gy), but by incorporating large expansion ( $w1 \rightarrow w2$  VE10%  $>7\text{cm}^3$ ), the proposed model was able to correctly classify it as an AE2 case. This plot also depicts the correlation between  $MED_{plan}$  and the predictions made using our model that could be used during the planning stage to relate the planned esophagus mean dose with week 2 mean dose and  $w1 \rightarrow w2$  esophagus expansion and act proactively for minimizing the risk of AE2.

### Quantification of $MED_{w2}$ and VE10%

Fig.4 illustrates  $w1 \rightarrow w2$  esophagus expansion for two different AE2 patients from CBCT (top) and MRI (bottom) cohorts. Esophageal wall expansion between week1 and week2 images (red arrows) was captured by the Jacobian maps (displayed on week1 images), from which VE10% (entire red area displayed on week2 images and column 3D) and MEex% (displayed on 3D) were computed.  $MED_{w2}$  for the first and the second patient was 8.3 Gy and 7.3 Gy and  $w1 \rightarrow w2$  VE10% was  $6.4\text{cm}^3$  and  $8.5\text{cm}^3$ , respectively.  $w1 \rightarrow w2$  MEex% for patient 1 was 27% and 44% for patient 2.

### Discussion

In this novel proof-of-concept study, we showed that incorporating the weekly esophagus volume change from the first two weeks of treatment into a prediction model in addition to accumulated mean esophagus dose could significantly improve the prediction of AE2 compared to the commonly reported planned mean dose (AUC:0.67  $\rightarrow$  AUC:0.80) using either CBCT or MRI data. The goal of this study was to develop an early AE prediction model using the longitudinal imaging data before toxicity development (i.e. first two weeks) to be effective for adaptive RT.

Esophagus changes during the first two weeks are expected to be small; thus a sensitive while reliable metric was desirable. This may be the reason that the volume with 10% local expansions (VE10%) showed better performance than the maximum (MEex%) or net volume change (VC%). The  $w1 \rightarrow w2$  MEex% may be affected by noise because of its much

smaller volume ( $\sim 0.1\text{cm}^3$ ) and  $w_1 \rightarrow w_2$  VC% shows small average esophagus changes in CBCT due to lower soft-tissue contrast.

To separately compare prediction performance of VE5% and MEex% against VE10% and also to assess the robustness of the logistic model with respect to these parameters, we separately evaluated ( $\text{MED}_{w_2} + w_1 \rightarrow w_2$  VE5%) and ( $\text{MED}_{w_2} + w_1 \rightarrow w_2$  MEex%) models. Both models showed similar accuracy to ( $\text{MED}_{w_2} + \text{VE10\%}$ ) model. The VE5% model had similar AUC=0.78 but with lower sensitivity (68%) compared to VE10% model while the MEex% model had AUC=0.76 but sensitivity was closer to VE10% model (77%). VE5% had much larger volume thus it captured wider ranges of expansions that may lead to lower false positive classifications. On the contrary, VE10% and MEex% were more sensitive due to smaller volumes. One reason we focused on VE10% for model building was that VE10% was less affected by the delineation uncertainties and undesired deformations caused by registration irregularities compared to VE5% and achieved the best balance between sensitivity and reliability.

We also validated the model of Huang et al. (4) that includes  $\text{MED}_{\text{plan}}$  only and previously introduced as the best AE2 prediction model (1). As expected, the result was similar to our  $\text{MED}_{\text{plan}}$  model (AUC=0.67) with high false positive ratio (57%) in comparison with only 17% using the combined model. While specificity of the  $\text{MED}_{\text{plan}}$  model was considerably lower than that of the proposed model (43% vs. 83%), it showed higher sensitivity (89% vs. 79%) which may be clinically useful. Though not all the potential AE2 patients were identified due to the sensitivity limitation, as demonstrated in Fig.3D, our model can provide a link between  $\text{MED}_{\text{plan}}$  and the early predictions made by the combined  $\text{MED}_{w_2} + w_1 \rightarrow w_2$  VE10% parameters that could be used during the treatment planning stage to relate a given planned esophagus mean dose with the corresponding week 2 mean dose and esophagus expansion and eventually estimate the likelihood of esophagitis, hence act proactively for minimizing the risk of AE2. In addition, our AE2 prediction model works closely and effectively with the workflow of adaptive radiotherapy by incorporating more frequently available weekly data to improve the overall prediction accuracy rather than relying only on treatment plan dose. Since AE2 occurred at week 4 for the majority of patients, the model offers two weeks for clinicians to adapt the treatment strategy, e.g., reoptimizing the treatment plan to lower esophagus dose while maintaining proper tumor coverage to improve the therapeutic ratio.

When the final model was validated using MRI, it showed high accuracy (AUC=0.75) despite the small number of patients. This might be because higher soft tissue contrast in MRI images provided finer visualization of esophagus boundaries and soft tissue textures compared to the CBCT which led to more accurate detection and classification of the changes.

To compare with other works that examined non-dose metrics to predict AE, some previous studies showed that utilizing 18F-fluorodeoxyglucose Positron Emission Tomography (FDG-PET) uptake to quantify inflammatory esophagus response and its correlation to dose-volume parameters improved AE prediction accuracy in NSCLC patients treated with CRT (24–26). In an early prediction study, Mehmood *et al.* (24) reported that change in 95<sup>th</sup>



percentile of FDG-PET Standard Uptake Values (SUV) of the esophagus between week 0→week 2 could better predict esophagitis than V50. In another study using mid-RT FDG-PET, Niedzielski *et al.* (25) investigated prediction performance of spatially localized dose-PET metrics using a logistic regression model and showed that normalized SUV metrics significantly outperformed dose-only models when evaluating AE2. Although FDG-PET scan is not a routine clinical protocol for diagnosing AE and generally prescribed for research purposes, all above studies concluded that planning dose-volume parameters alone were inferior predictors of AE and using other longitudinal metrics that represent change in esophagus improved the predictions, which agreed with our findings.

We thoroughly validated our registration algorithm and Jacobian calculations by comparing against the ground-truth manual segmentations (Results & Supplemental S1), however due to inferior soft-tissue contrast/artifacts in CBCT, the algorithm was susceptible to errors. For this reason, we further validated our registrations using an external publicly available longitudinal dataset containing pCT and weekly CBCTs to find the optimal hyperparameters (Supplemental S1). One limitation in this study was that anatomical imaging e.g. CT with low soft-tissue contrast, cannot confirm that the esophageal expansion measured using the Jacobian map is indeed due to radiation induced inflammation as opposed to daily anatomy variations (e.g. esophageal lumen filling, breathing motion, patient setup etc.) (15). Another limitation was that an accurate CBCT esophagus segmentation was required for the proposed model to be reliably used. To verify reliability, we calculated intra-observer variation of the RO's manual esophagus contours on CBCTs and found that the DSC was 0.89, indicating small uncertainty with little effect on the model building.

Recently, some efforts have been made for a longitudinal segmentation of the esophagus in CBCT using deep learning (27,28) by simulating CBCT artifacts/noise on higher quality pCT to preserve esophagus boundaries. Moreover, more attempts are being made to improve the CBCT image reconstruction algorithms. For instance, the ability to utilize a linear accelerator's kV cone beam CT for treatment planning and dosimetry calculation using the Varian iterative cone beam CT (iCBCT) reconstruction method (Varian Medical Systems, Palo Alto, CA) has showed improved HU accuracy and uniformity as well as better soft tissue visualization than standard reconstruction (29). This work would open the way to taking advantage of this image modality, which is more easily available during treatment than time consuming MRI acquisition or labor-intensive methods that require expert guidance for assessing the RT response in real time rather than only for patient setup purposes.

While adaptive planning is a technical reality, whether re-planning efforts to mitigate AE would certainly benefit patients and improve their treatment requires additional investigation. To explore this, the time course of AE and the extent to which AE can be mitigated once the patient starts treatment may be the key factors to see if adaptive re-planning can help potential LA-NSCLC patients who might experience more difficulties during the treatment.

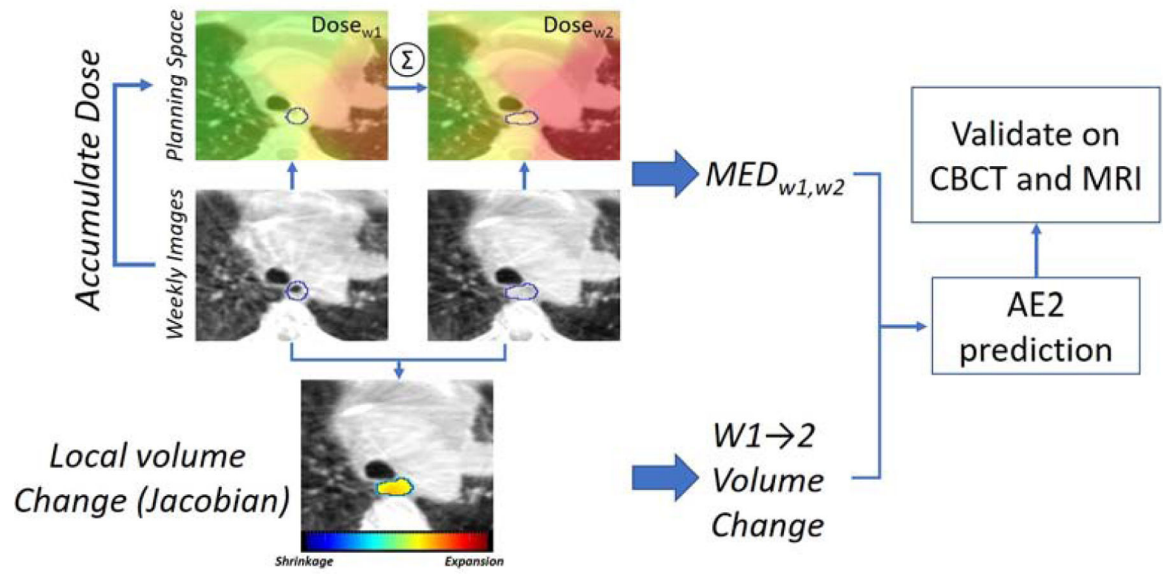
## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

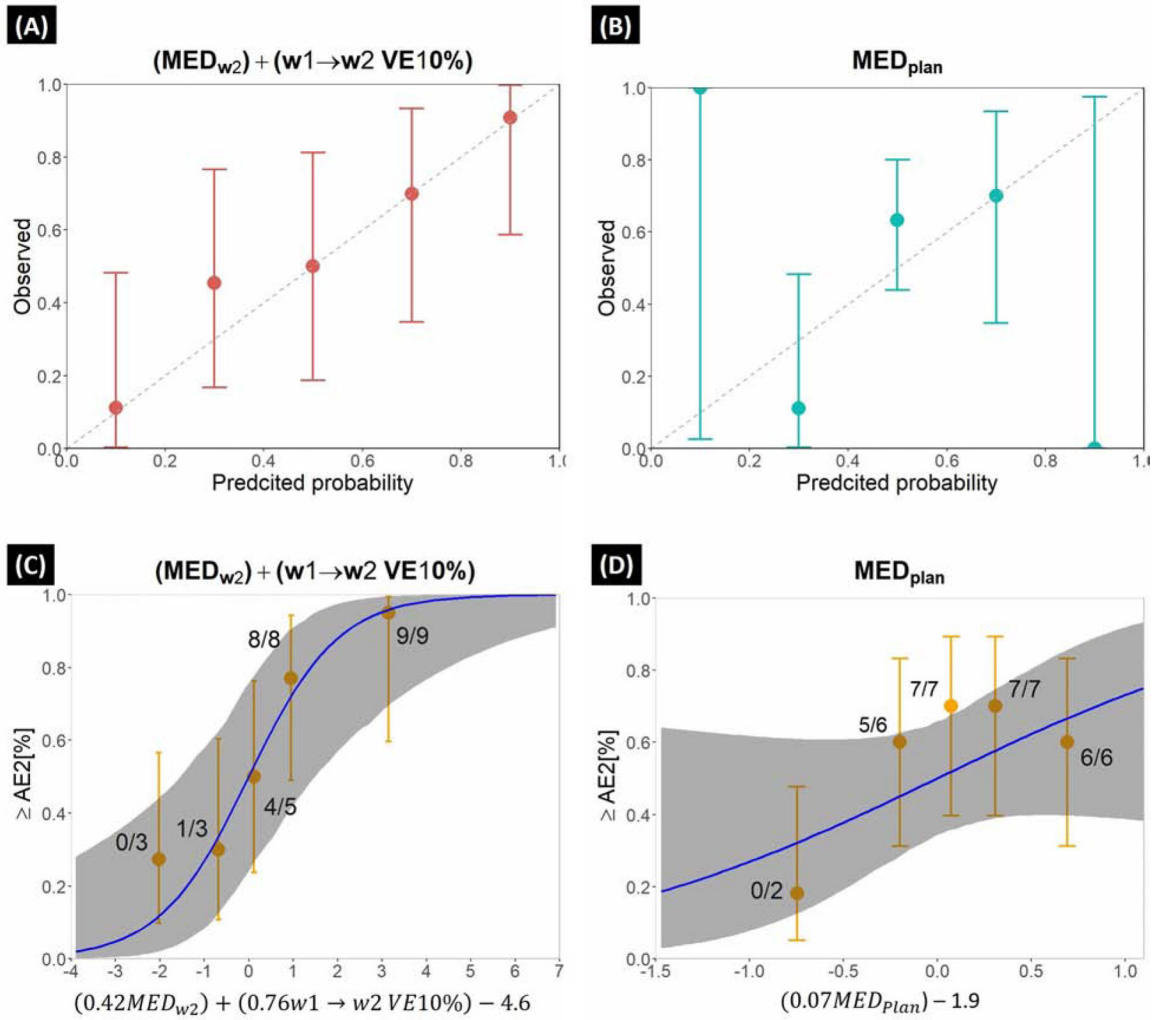
## References

1. Thor M, Deasy J, Iyer A, et al. Towards personalized dose-prescription in locally advanced nonsmall cell lung cancer: Validation of published normal tissue complication probability models. *Radiother Oncol* 2019; 138:45–51. [PubMed: 31146070]
2. Werner-Wasik M, Yorke E, Deasy J, et al. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys* 2010;76:S86–S93. [PubMed: 20171523]
3. Delgado BD, Enguix-Riego MV, Fernández de Bobadilla JC, et al. Association of single nucleotide polymorphisms at HSPB1 rs7459185 and TGFB1 rs11466353 with radiation esophagitis in lung cancer. *Radiother Oncol* 2019;135:161–169. [PubMed: 31015163]
4. uang EX, Bradley JD, El Naqa I, et al. Modeling the risk of radiation-induced acute esophagitis for combined Washington University and RTOG trial 93–11 lung cancer patients. *Int J Radiat Oncol Biol Phys* 2012;82:1674–1679. [PubMed: 21658856]
5. Niedzielski J, Yang J, Stingo F, et al. A novel methodology using CT imaging biomarkers to quantify radiation sensitivity in the esophagus with application to clinical trials. *Sci Rep* 2017;7:6034. [PubMed: 28729729]
6. Alam S, Thor M, Rimner A, et al. Quantification of accumulated dose and associated anatomical changes of esophagus using weekly magnetic resonance imaging acquired during radiotherapy of locally advanced lung cancer. *Physics and Imaging in Radiation Oncology* 2020;13:36–43.
7. Niedzielski JS, Yang J, Stingo F, et al. Objectively quantifying radiation esophagitis with novel computed tomography-based metrics. *Int J Radiat Oncol* 2016;94:385–393.
8. Court LE, Tucker SL, Gomez D, et al. A technique to use CT images for in vivo detection and quantification of the spatial distribution of radiation-induced esophagitis. *J Appl Clin Med Phys* 2013;14: 4195. [PubMed: 23652249]
9. Mesurolle B, Qanadli SD, Merad M, et al. Unusual radiologic findings in the thorax after radiation therapy. *Radiographics* 2000;20:67–81. [PubMed: 10682772]
10. Berkovich GY, Levine MS, Miller WT Jr. CT findings in patients with esophagitis. *AJR Am J Roentgenol* 2000;175:1431–1434. [PubMed: 11044057]
11. Kong FM, Ritter T, Quint DJ, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol Biol Phys* 2011;81:1442–1457. [PubMed: 20934273]
12. Kong FM, Quint L, Machtay M, et al. Atlases for organs at risk (oars) in thoracic radiation therapy. Radiation Therapy Oncology Group (RTOG). <https://www.eviq.org.au/getmedia/a4c012a8-d6a7-4d87-93f7-d3f465b49889/RTOG-heart-contouring-atlas.pdf.aspx?extZ>. Accessed February 4, 2021.
13. National Institutes of Health National Cancer Institute. Common terminology criteria for adverse events (CTCAE) version 4.0. Washington, DC: US Department of Health and Human Services; 2009.
14. Tustison NJ, Avants BB. Explicit B-spline regularization in diffeomorphic image registration. *Front Neuroinform* 2013;7:39. [PubMed: 24409140]
15. Riyahi S, Choi W, Liu C. et al. Quantifying local tumor morphological changes with jacobian map for prediction of pathologic tumor response to chemo-radiotherapy in locally advanced esophageal cancer. *Phys Med Biol* 2018;63:145020. [PubMed: 29911659]
16. Riyahi S, Choi W, Liu C. et al. Quantification of local metabolic tumor volume changes by registering blended pet-ct images for prediction of pathologic tumor response. *Lect Note Comp Scien* 2018;11076:41.
17. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med* 1997;16:2143–2156. [PubMed: 9330425]

18. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction: Oxford Statistical Science Series. 3rd ed. Oxford, UK: Oxford University Press; 2003.
19. Carpenter J, Bithell J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Stat Med* 2000;19: 1141–1164. [PubMed: 10797513]
20. Hosmer DW, Lemeshow S, Surdivant RX. Applied Logistic Regression. 3rd ed. Hoboken, NJ: Wiley; 2000.
21. Norman R, Draper SH. Applied Regression Analysis. New York, NY: John Wiley & Sons; 1998.
22. Team R Rstudio: Integrated Development for R. Boston, MA: RStudio; 2015.
23. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007;17: 571–582. [PubMed: 17613642]
24. Mehmood Q, Sun A, Becker N, et al. Predicting radiation esophagitis using 18F-FDG PET during chemoradiotherapy for locally advanced non-small cell lung cancer. *J Thorac Oncol* 2016;11:213–221. [PubMed: 26718880]
25. Niedzielski JS, Yang J, Liao Z, et al. (18)F-Fluorodeoxyglucose positron emission tomography can quantify and predict esophageal injury during radiation therapy. *Int J Radiat Oncol Biol Phys* 2016;96: 670–678. [PubMed: 27681764]
26. Nijkamp J, Rossi M, Lebesque J, et al. Relating acute esophagitis to radiotherapy dose using FDG-PET in concurrent chemo-radiotherapy for locally advanced non-small cell lung cancer. *Radiother Oncol* 2013;106:118–123. [PubMed: 23219463]
27. Alam SR, Li T, Zhang S-Y, et al. Cross-modality esophagus segmentation using physics-based data augmentation. American Association of Physicists in Medicine (AAPM) Annual meeting. Vancouver, Canada. 2020.
28. Alam SR, Li T, Zhang S-Y, et al. Generalizable cone beam ct esophagus segmentation using in-silico data augmentation. arXivorg 2020.
29. Jarema T, Aland T. Using the iterative kC CBCT reconstruction on the Varian Halcyon linear accelerator for radiation therapy planning for pelvis patients. *Physica Medica* 2019;68:112–116. [PubMed: 31783220]

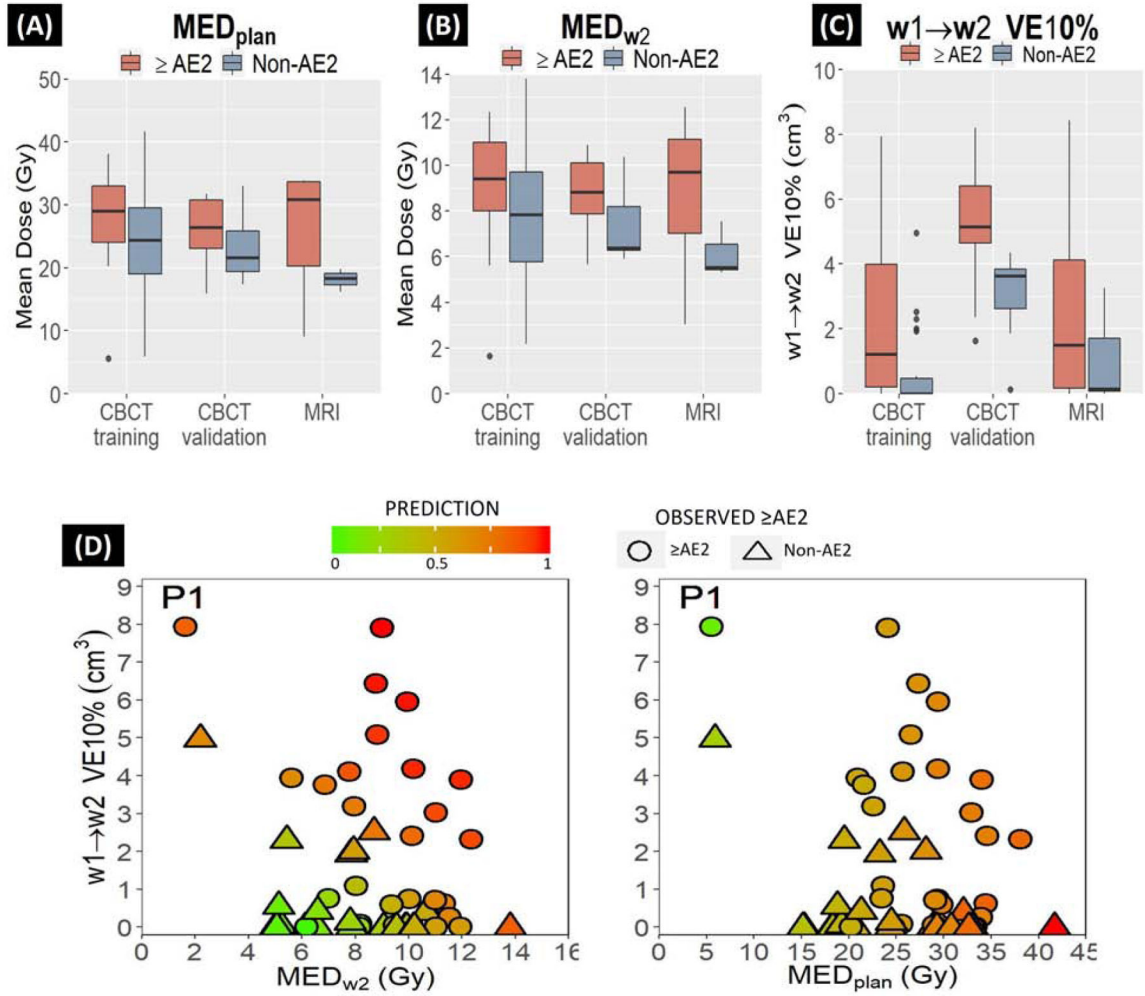


**FIGURE. 1.**  
Main workflow. MED<sub>w1,w2</sub>=week 1 and week 2 accumulated Mean Esophagus Dose.



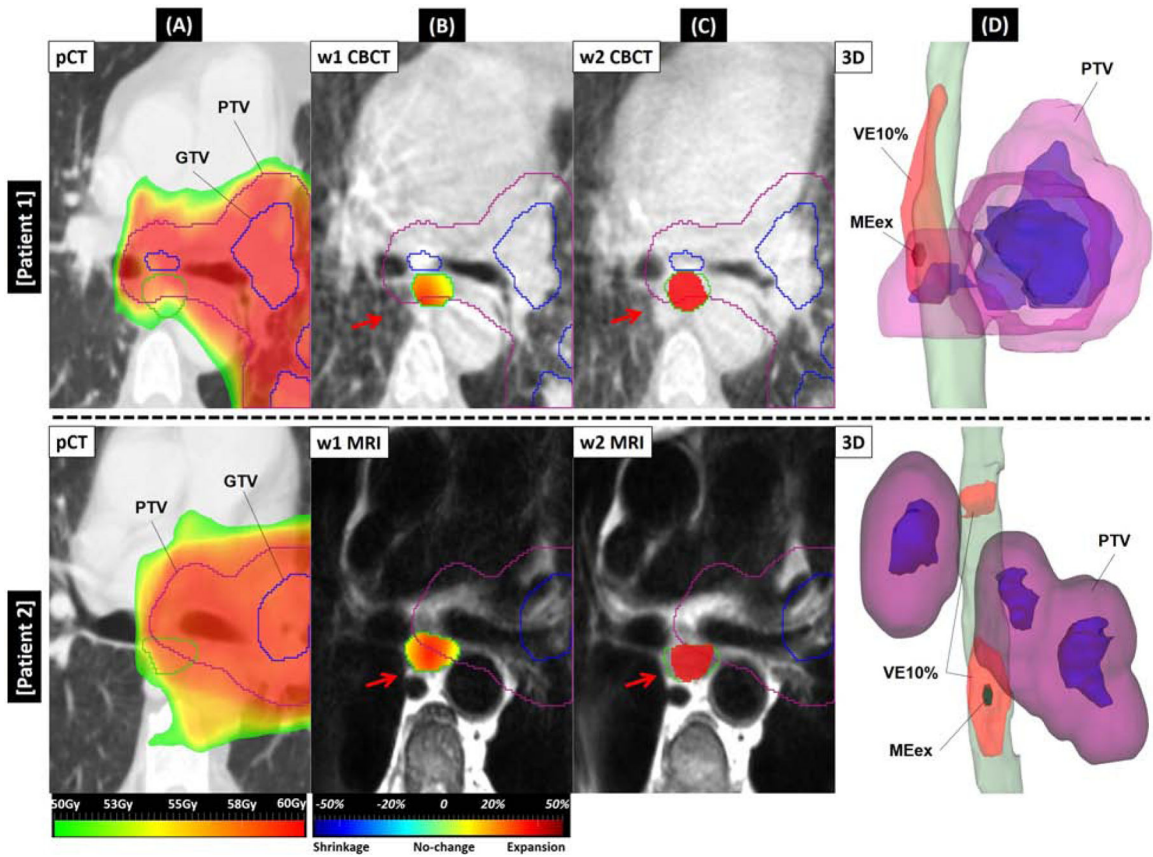
**FIGURE. 2.**

Calibration curves for the (A) proposed and (B)  $MED_{plan}$  model. Error bars denote 95% confidence interval. Dose-response curves for the (C) proposed and (D)  $MED_{plan}$  model where the dose-response predictions for  $\geq AE2$  are given by the blue lines (gray ribbon is 95% CI). Orange quintiles denote observed  $\geq AE2$  risk rate for each quintile (number of observed  $\geq AE2$ /total number of patients) (error bars: 95% binomial confidence intervals); quintile-specific predicted/observed  $\geq AE2$  rates [%] are given above each quintile.  $\geq AE2$  risk rate was calculated using the logistic function of  $[\text{risk} = 1 / (1 + \exp(-x))]$  where  $x$  is the logistic regression argument (log-odds) for each model given below the dose-response curves.



**FIGURE 3.**

AE2 predictability by dose-volume change parameters in CBCT and MRI cohorts for (A) MED<sub>plan</sub> (B) MED<sub>w2</sub> and (C) w1→w2 VE10% (D) scatter plots showing the patient specific predictions using the (left) combination of MED<sub>w2</sub> and w1→w2 VE10% versus (right) only MED<sub>plan</sub>. Symbol shapes represent observed ≥AE2 (circle) vs. Non-AE2 (<AE2, triangle) cases and continuous colors inside the symbols give prediction probabilities for each case. The annotated case (P1) is an outlier observed as an ≥AE2 case.



**FIGURE. 4.**

pCT, week1 and week2 images along with 3D view of w1→w2 expansions for two AE2 cases from CBCT (top) and MRI (bottom) cohorts. (A) prescribed dose map overlaid on pCT. Green contour is the planning esophagus. (B) week1 CBCT/MRI. Green contour is ground-truth esophagus observed on week1. Red arrows depict expansion between week1 and week2 where Jacobian map inside the esophagus contour shows large expansion. (C) week2 CBCT/MRI. Green contour is esophagus on week2. Red filled color wash depicts w1→w2 VE10% inside the esophagus. (D) 3D view showing w1→w2 VE10% volume expanded (red structure) on week1 esophagus (green) and the small dark green structure inside VE10% is maximum esophagus expansion (MEEx%). For each sub-figure, purple and blue contours are PTV and GTV, respectively.

**Table 1.**

Clinical and demographic characteristics of the cohort for CBCT and MRI datasets.

Clinical Characteristic	Number of patients		
	CBCT training	CBCT validation	MRI
Age (years)			
<i>Mean±SD, Range</i>	67±8, 50–86	65±9, 49–85	62±6, 51–74
Sex			
<i>male/female</i>	27/24	11/7	7/4
Histology			
<i>Adenocarcinoma/Squamous/Other</i>	36/14/1	14/4/0	7/2/2
Acute Esophagitis ( AE2)			
<i>AE2/&lt;AE2</i>	28/23	10/8	8/3
Anatomical Stage			
<i>IIB/IIIA/IIIB/IIIC</i>	3/16/32/0	0/4/10/4	1/2/7/1



**Table 2.**

Significant dose-volume change parameters after univariate analysis. P-values for the weekly parameters are adjusted using Bonferroni correction ( $p < 0.05$ ). p-value for  $MED_{plan}$  was calculated individually.

Features	p-value	AUC	Correlation to AE2
w1→w2 VE10%	<i>0.004</i>	0.78	0.50
w1→w2 VE5%	<i>0.01</i>	0.76	0.45
w1→w2 MEx%	<i>0.02</i>	0.74	0.41
w1→w2 VC%	<i>0.1</i>	0.70	0.35
$MED_{w2}$	<i>0.1</i>	0.69	0.33
w1→w2 VE15%	<i>0.2</i>	0.66	0.31
$MED_{w1}$	<i>0.3</i>	0.67	0.29
$MED_{plan}$	<i>0.04</i>	0.67	0.29

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript