Author's Post-print version

Publisher version: Int J Radiation Oncol Biol Phys. (2016);

http://dx.doi.org/10.1016/j.ijrobp.2016.04.033

A voxel-based approach to explore local dose differences associated with radiation-induced lung damage

Giuseppe Palma, Ph.D.,[1] Serena Monti, M.Sc.[2], Vittoria D'Avino,M.Sc.[1], Manuel Conson, M.D.[1,3], Raffaele Liuzzi, Ph.D.,[1] Maria Cristina Pressello, Ph.D.,[4] Vittorio Donato, M.D.[5], Joseph O Deasy, Ph.D.,[6] Mario Quarantelli, M.D.[1], Roberto Pacelli, M.D.[1,3]*, Laura Cella, Ph.D.,[1]*

[1]Institute of Biostructures and Bioimaging, National Research Council (CNR), Naples, Italy

[2]IRCCS SDN, Naples, Italy

[3]Department of Advanced Biomedical Sciences, Federico II University School of Medicine, Naples, Italy

[4]Department of Health Physics, S. Camillo-Forlanini Hospital, Roma, Italy

[5]Department of Radiation Oncology, S Camillo-Forlanini Hospital, Rome, Italy

[6]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York, USA

*Co-senior authors

Corresponding author: Laura Cella, Institute of Biostructures and Bioimaging, National Council of Research (CNR), Via T De Amicis, 80145 Naples, Italy. Tel: +39-081-2203187 ext 227; E-mail: laura.cella@cnr.it

Short title: voxel-based approach to radiation-induced lung damage prediction

Key words: Radiation-induced lung damage, elastic image registration, voxel-based analysis, multiple comparison problem, NTCP.

Conflict of Interest Notification: The authors declare no conflict of interest.

Acknowledgements

This work was partially supported by grants from the Italian Ministry of Education, University and Research (MIUR) in the framework of FIRB (RBFR10Q0PT_001 “DROPS”).

Abstract

Purpose: To apply a voxel-based (VB) approach aimed at exploring local dose differences associated with late radiation-induced lung damage (RILD).

Materials and Methods: An inter-institutional database of 98 patients of Hodgkin lymphoma (HL) survivors treated with post-chemotherapy supradiaphragmatic radiation therapy was analyzed in the study. Eighteen patients developed late RILD, classified according to the Radiation Therapy Oncology Group (RTOG) scoring system. Each patient's Computed Tomography (CT) was normalized to a single reference case anatomy (common coordinate system-CCS) via a log-diffeomorphic approach. The obtained deformation fields were used to map the dose of each patient into the CCS. The co-registration robustness and the dose mapping accuracy were evaluated by geometric and dose scores. Two different statistical mapping schemes for non-parametric multiple permutation inference on dose maps were applied, and the corresponding $p<0.05$ significance lung sub-regions were generated. A Receiver Operating Characteristic (ROC) based test was performed on the mean dose extracted from each sub-region.

Results: The co-registration process resulted in a geometrically robust and accurate dose warping. A significantly higher dose was consistently delivered to RILD patients in voxel clusters near the peripheral medial-basal portion of the lungs. The area under the ROC curves (AUC) from the mean dose of the voxel clusters was higher than the corresponding AUC derived from the total lung mean dose.

Conclusions: We implemented a framework including a robust registration process and a voxel-based approach accounting for the multiple comparison problem in dose-response modeling, and applied it on a cohort of HL survivors to explore a local dose-RILD relationship in the lungs. Patients presenting RILD received a significantly greater dose in parenchymal regions where low doses (~6

Gy) were delivered. Interestingly, the relation between differences in the high dose range and RILD seems to lack a clear spatial signature.

Summary

The risk estimation of radiation-induced lung damage (RILD) is generally based on lung DVHs disregarding any spatial dose-distribution information. In this study, a robust registration and voxel-based approach were applied to explore lung dosimetric regional differences associated with RILD. We highlighted a local dose-RILD relationship: a significantly higher dose was delivered in the low-dose parenchymal regions, whereas the relation between high dose range differences and RILD seems to lack a clear spatial signature.

Introduction

Radiation induced lung damage (RILD) may present as an acute inflammatory phase (pneumonitis) or a later fibroproductive phase, referred to as lung fibrosis. Regional differences in lung response to radiation have been the subject of several preclinical and clinical studies, which overall suggested that the middle and caudal lung regions are more sensitive to the radiation insult (1). Normal tissue complication probability (NTCP) models, developed to estimate the risk of RILD, generally rely on lung dose-volume histogram (DVH) analyses, which disregard any spatial dose distribution information and possible inhomogeneity in regional organ radio-sensitivity.

Recently, 2D or 3D dose distribution based methods, collectively referred to as voxel-based (VB) methods, which evaluate dose-response relationships and overcome the organ-based philosophy of NTCP modelling, have been proposed as alternative approaches to predict urinary (2), gastrointestinal (3), or rectal toxicity (4) after radiation therapy for prostate cancer.

The current study was devised to apply a VB approach in order to investigate the relationship between local lung dose and late RILD. The lungs are organs with high morphological variability and their accurate matching requires a non-rigid registration strategy (5). We implemented an inter-patient elastic image registration (EIR) framework to map all patient dose distributions into a single reference case anatomy. Thereafter, a voxel-by-voxel analysis was performed to test dosimetric regional differences between patients with different outcomes. We accounted for the massive multiple comparison (MC) problem, which may arise in the analysis of imaging data when the statistical analysis is run separately for each voxel, by applying non-parametric procedures based on randomization/permutation testing (6)). While imaging-based methods have been developed to measure individual patient reaction to lung irradiation (7-10), we are unaware of studies that apply this approach to explore lung dosimetric patterns associated with RILD.

Materials and Methods

Patient database

The retrospectively collected dataset reported in this analysis includes 98 eligible patients from an interinstitutional database of 148 Hodgkin lymphoma (HL) survivors treated with post-chemotherapy supradiaphragmatic involved-field 3D-conformal radiation therapy (CRT). Treatment-planning Computed Tomography (CT) were acquired in free-breathing modality and dose maps were generated with heterogeneity corrections. A median treatment total dose of 30.6 Gy (range: [20.8, 45.0] Gy) in daily fractions of 1.5-1.8 Gy was prescribed (11). Applied selection criteria include a minimum follow-up of 12 months, lack of pre-radiation treatment lung disease, availability of treatment planning CT with associated 3D-dose map, and adequate CT coverage of the lungs. All patients have been monitored for late pulmonary toxicity following the Radiation Therapy Oncology Group (RTOG) scoring system (12). At a median time to event of 13 months (range: [9, 83] months), 18 patients displayed radiological changes on follow-up CT. Four patients were diagnosed with Grade 3 (severe symptomatic fibrosis showing dense radiographic changes), 5 cases developed Grade 2 RILD (two slight radiological changes with severe cough and three moderate symptomatic fibrosis with patchy radiographic appearances). Nine patients were diagnosed with Grade 1 (slight CT radiological changes without symptoms). We considered all CT radiological density changes (i.e. any grade of RILD) as outcome. Time to event was computed from the beginning of RT to the first radiological signs.

A detailed description of patients' and treatment characteristics has been previously reported (13, 14).

The contours of lung tissue and heart were reviewed on planning CTs following RTOG 1106 and heart atlas contouring guidelines (15, 16). The CT matrix size was 512x512 in plane with a slice thickness of 5 mm.

Dataset processing

Individual DICOM RT plans (CT scans, doses and contoured organs) were converted into a Matlab (MathWorks, Natick, MA)-readable format using the CERR (Computational Environment for Radiotherapy Research) software (17).

Unless noted otherwise, all processing steps described below were handled using an in house software developed in Matlab.

Elastic Image Registration

Prior to the EIR, CT scans were pre-processed according to the following steps. A binary mask was extrapolated from the organ-at-risk segmentations of the treatment plan. For each patient, the mask, computed as the union and dilation (spherical structuring element of radius 30 mm) of heart and lung structures, was used to crop the field-of-view and to align the structures of interest according to an affine transformation based only on the mask outer contours. CT images were masked accordingly in order to hide some inter-individual or gender-related anatomical differences of limited interest to our study, and to allow the registration algorithm to work more efficiently on tissue contrast inside the chest.

The dataset corresponding to the patient with the median lung-volume was chosen as reference image for the cohort, defining a study-specific common coordinate system (CSS) for the EIR. The log-diffeomorphic extension (18) of the demons algorithm was used to register each other patient CT on CCS. The obtained deformation fields were then used to map the dose of each patient to the CSS. The chosen EIR algorithm guarantees the estimated deformation fields to be invertible for the warping of the dose matrix.

Given that the prescribed dose to the target has to be an absolute scalar field (i.e. relative scalar field of weight 0) under spatial transformations, no Jacobian intensity modulation was applied to the deformed dose.

To evaluate the performance of inter-patient EIR, the Dice Index (DI) (19) and modified Hausdorff distance (MHD) (20) were calculated. Besides those pure geometrical scores, we also computed the dose-organ overlap (DOO) introduced by Acosta *et al.* (4) in order to evaluate dose warping.

The median Hausdorff distance was used to define the full width at half maximum of a spherical Gaussian kernel used to smooth the co-registered dose maps (21).

Statistical mapping

In order to compare dose maps associated with patients who developed RILD versus those of patients who did not, two different non parametric methods accounting for MC problem were applied.

First, a non-parametric multiple comparisons permutation testing by single maximum threshold was performed following the method proposed by (6) by means of an in-house developed Matlab library. Briefly, at each voxel the average dose difference was normalized to the standard deviation computed over all random samples generated from 1000 permutations (22) on the RILD labels (yes vs. no). The normalized maximum dose difference ($T_{\max}$) was selected as test statistic summarizing the discrepancy between the two RILD groups and therefore avoiding a voxel-wise test and a consequent MC problem. After each permutation *i*, we obtain a distribution of test statistic $T_{\max,i}$, from which the adjusted *p*-value can be computed as the probability of having a $T_{\max,i}$ greater than $T_{\max}$ in the observed sample ($\widetilde{\boldsymbol{T}}_{max}$) and compared with a significance level of 5%. The normalized maximum dose difference value corresponding to the 95 percentile (T^*) possibly determines a voxel region with a statistically significant dose difference.

In addition, we adopted a non-parametric permutation inference (1000 permutations) coupled to the Threshold-Free Cluster Enhancement (TFCE) method (21). The TFCE is a method by which cluster-like structures are enhanced without having to define clusters in a binary way. Permutation

testing with TFCE is implemented in the *randomise* tool (23) available in the FMRIB Software Library v5.0 (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL).

Finally, for comparison with the scheme described by (4), an uncorrected voxel-wise two sample *t*-test was also performed on dose distribution maps of each group.

Eventually, sub-contours corresponding to $p<0.05$ were generated; Dose-Volume Histograms (DVH) computed; and the mean dose (D_{mean}) extracted.

Statistical analysis

The median and the range were used to describe all continuous variables. A non-parametric paired test (Wilcoxon signed rank test) was used to compare each score before and after EIR. The Mann-Whitney *U*-test was employed to test the mean dose difference between RILD and non-RILD patients. Logistic regression and receiver operating characteristic (ROC) analyses were performed to test dosimetric prediction performance. Statistical analysis was performed with SPSS 18.0 and MedCalc statistical software.

Results

The robustness of registration can be assessed by visual inspection (Fig. 1a-b) and by the comparison of the DI, MHD and DOO scores (Fig. 1c-e) computed on the whole population before and after EIR. As shown in Table 1, a significant ($p<0.05$) improvement of concordance metrics is obtained after the registration process.

The non-parametric MC permutation test showed a significant difference between the dose maps belonging to RILD vs non-RILD patients. Indeed, the distribution of $T_{max,i}$ obtained from the 1000 random permutations resulted in a significant adjusted *p*-value of 0.02 (Fig. 2).

Clusters of statistically significant dose difference between the response groups ($p<0.05$) were detected by all the three different statistical mapping schemes (*T*, TFCE and uncorrected *t* tests)

(Fig. 3). In Table 2, the absolute volumes and the associated median doses delivered on average to the three different voxel clusters are reported. As expected, overlapping and increasingly restrictive lung sub-regions S_t, S_{TFCE} and S_T were detected by *t*, TFCE and *T* tests, respectively (Fig 3d-f).

For each lung sub-region, the extracted D_{mean} has been tested as a predictive variable of the RILD outcome. An NTCP model was calculated with a logistic regression using D_{mean} for each of S_t, S_{TFCE} and S_T regions. All the corresponding ROC curves (Fig. 4) resulted in higher AUC values when compared to total lung D_{mean}: AUC(S_t)=0.75, AUC(S_{TFCE})=0.69, AUC(S_T)=0.71 and AUC(total lung)=0.60.

Discussion

In the present study, we devised a comprehensive framework for the application of a voxel-based approach aimed at investigating the relationship between local organ dose and radiation induced toxicity. In particular, we implemented a procedure to explore dosimetric lung regional differences associated to the development of late RILD. The procedure was applied on a cohort of HL survivors treated with post-chemo radiation therapy in the supradiaphragmatic region (13). Thoracic irradiation, also at the relatively low dose range inherent to HL treatments, may be responsible for late-phase subclinical lung radiation-induced injuries such as fibrosis, resulting in radiological density changes detectable on radiographic studies or by computed tomography. Lung fibrosis, even if asymptomatic, is likely to drive a decline in pulmonary function, consequently affecting the long-term quality of life of cancer survivors (1).

Voxel-wise methods with image registration techniques have been proposed as effective tools to identifying critical organ sub-regions strongly correlated with organ toxicity (4). As such, the VB approach seems extremely promising to investigate the response of lungs to radiation. Lungs are characterized by a complex and heterogeneous anatomical architecture with peripheral alveolar-

capillary units appearing far more sensitive to the effects of radiation than the central conducting airways and vessels (1).

The use of a VB analysis strategy, more diffuse in other fields, such as neuroimaging (24), raises the question of the massive multiple comparison problem. The MC problem can be overcome by permutation methods that allow inferences while taking into account the multiplicity of tests, as described by (22) using a single maximum statistic. Alternatively, permutation testing can also be coupled to statistics that combine the spatial extent of signals, such as threshold-free cluster enhancement (21). The TFCE aims to enhance areas of signal that exhibit some spatial contiguity, without relying on hard-threshold-based clustering. In the present study, both statistical mapping schemes were applied.

Another key issue for the VB approach is the choice of the registration strategy and its impact on dose warping accuracy. The accuracy, reproducibility, and computational performance of several deformable image registration algorithms applied to thoracic CT image registration have been evaluated in various studies (5, 25). In the present work we have chosen to use the demons deformable registration that has been shown to provide higher accuracy for landmark matching of masked lungs in serially acquired chest CT scans compared with B-splines, affine, or rigid registration (7). Similarly, the demon algorithm achieved high matching accuracy between phases of 4D-CT scans (5, 7, 25). This is the first dosimetric study using an improved version of the demon approach proposed by (18), which guarantees deformation fields to be diffeomorphic (invertible). The implemented registration process was optimized for the registration of thoracic CT data and resulted geometrically robust and accurate in dose warping, as shown by the inter-patient match improvement summarized in Table 1.

Using a VB approach and different statistical mapping schemes, we were able to identify a local dose-RILD relationship in the lungs. Interestingly, all the applied schemes (including the simple *t*-

test – which disregards the MC problem) identified overlapping and progressively enlarging sub-regions, namely S_T, S_{TFCE} and S_t (Table 2 and Figure 3). The S_T volume represents the minimal sub-region that contributes to the significance of the dose difference between groups, while the S_{TFCE} defines an extended volume by including neighboring voxels in order to increase sensitivity. In the end, the S_t lung sub-region can be seen as an extra spatial safety margin, despite with a lower specificity in localizing suspicious regions.

In this respect, a reliable *a posteriori* compromise in terms of extent of the suspicious region is provided by the TFCE approach, which exploits the spatial distribution of dose differences to properly recover statistical significance of voxels in the detected clusters and, incidentally, embodies the most sophisticated techniques among the three statistical schemes. Of note, ROC-based testing of different lung sub-regions and corresponding D_{mean} values revealed a similar prediction performance for each of the three models (Figure 4).

Overall, our findings suggest that the irradiation of peripheral parenchymal region in the middle and caudal lung is correlated with RILD. Interestingly, a higher dose was delivered in the low-dose (~6 Gy for S_{TFCE}) parenchymal regions (Fig. 3a-b), in agreement with some recent DVH analyses showing that the lung volume exceeding 5 Gy is consistently more predictive for RILD than other dosimetric variables (13, 26). Although this result seems to be also in agreement with a suspected higher sensitivity to radiation of peripheral alveolar-capillary units, conclusions about cause-effect relationship cannot be drawn. The VB approach can only identify regions that are correlated with toxicity , but not necessarily responsible for it (4). This analysis is hypothesis-generating; however, given the limited current dataset, to obtain more powerful insights on possible local lung radiosensitivity, the method should be further applied to larger databases, evaluating RILD in heterogeneously treated lungs.

The following source of uncertainties that may impact upon the accuracy of the considered approach should also be pointed out: geometric uncertainties intrinsic to deformable image registration process and dosimetric uncertainties associated to breathing motion, which changes not only the volume of the lung, but also the location and the density of local anatomy in a non-rigid fashion. However, even under these caveats, it should be here stressed that such uncertainties are not group-related (patients with RILD vs. patients without RILD) and therefore no bias is expected in the performed analysis.

Conclusion

In conclusion, we implemented a framework including a robust registration process and a VB approach accounting for MC problem intended to investigate local dose-lung toxicity relationship. We illustrated it on a cohort of HL survivors analyzed for RILD development. The two explored statistical mapping schemes identified two nested sub-regions at risk. While the relation between the high dose range of the HL treatment and RILD seems to disregard the spatial location of dose deposition, our method highlighted a spatial signature in the lower dose range (1.5-6 Gy) that is related with lung damage.

References

1. Rubin P. Michael T. Milano and Louis S. Constine. *ALERT• Adverse Late Effects of Cancer Treatment: Volume 2: Normal Tissue Specific Sites and Systems* 2013:189.

2. Fiorino C, Palorini F. SP-0215: Advanced methods for 2D/3D dose map correlation in modelling toxicity. *Radiother Oncol* 2015;115:S108-S109.

3. Wortel RC, Witte MG, van der Heide UA*, et al.* Dose–surface maps identifying local dose–effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer. *Radiother and Oncol* 2015;117(3):515-520.

4. Acosta O, Drean G, Ospina JD*, et al.* Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. *Phys Med Biol* 2013;58(8):2581.

5. Murphy K, Van Ginneken B, Reinhardt JM*, et al.* Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *Medical Imaging, IEEE Transactions on* 2011;30(11):1901-1920.

6. Chen C, Witte M, Heemsbergen W*, et al.* Multiple comparisons permutation test for image based data mining in radiotherapy. *Radiat Oncol* 2013;8(1):1.

7. Cunliffe AR, Al-Hallaq HA, Labby ZE*, et al.* Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration. *Med Phys* 2012;39(8):4679-4690.

8. Cunliffe A, Armato SG, 3rd, Castillo R*, et al.* Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *Int J Radiat Oncol Biol Phys* 2015;91(5):1048-1056.

9. Defraene G, van Elmpt W, Crijns W*, et al.* CT characteristics allow identification of patient-specific susceptibility for radiation-induced lung damage. *Radiother Oncol* 2015;117(1):29-35.

10. Ghobadi G, Wiegman EM, Langendijk JA*, et al.* A new CT-based method to quantify radiation-induced lung damage in patients. *Radiother Oncol* 2015;117(1):4-8.

11. Cella L, Conson M, Pressello MC, et al. Hodgkin's lymphoma emerging radiation treatment techniques: Trade-offs between late radio-induced toxicities and secondary malignant neoplasms. Radiat Oncol 2013;8:22

12. Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Int J Radiat Oncol Biol Phys* 1995;31(5):1341-1346.

13. Cella L, D'Avino V, Palma G, et al. Modeling the risk of radiationinduced lung fibrosis: Irradiated heart tissue is as important as irradiated lung. Radiother Oncol 2015;117:36-43.

14. Cella L, Liuzzi R, D'Avino V, et al. Pulmonary damage in Hodgkin's lymphoma patients treated with sequential chemo-radiotherapy: Predictors of radiation-induced lung injury. Acta Oncol 2014;53: 613-619.

15. Feng M, Moran JM, Koelling T*, et al.* Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *Int J Radiat Oncol Biol Phys* 2011;79(1):10-18.

16. 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(5):1442-1457.

17. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys* 2003;30(5):979-985.

18. Vercauteren T, Pennec X, Perchant A*, et al.* Symmetric log-domain diffeomorphic registration: A demons-based approach. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: Springer; 2008. pp. 754-761.

19. Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26(3):297-302.

20. Dubuisson M-P, Jain AK. A modified Hausdorff distance for object matching. *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proc of the 12th IAPR International Conference*. Vol 1: IEEE; 1994. pp. 566-568.

21. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 2009;44(1):83-98.

22. Holmes AP, Blair R, Watson J*, et al.* Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow & Metabolism* 1996;16(1):7-22.

23. Winkler AM, Ridgway GR, Webster MA*, et al.* Permutation inference for the general linear model. *Neuroimage* 2014;92:381-397.

24. Friston KJ, Ashburner JT, Kiebel SJ*, et al.* Statistical parametric mapping: the analysis of functional brain images: Academic press; 2011.

25. Brock KK, Consortium DRA. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys* 2010;76(2):583-596.

26. Pinnix CC, Smith GL, Milgrom S*, et al.* Predictors of Radiation Pneumonitis in Patients Receiving Intensity Modulated Radiation Therapy for Hodgkin and Non-Hodgkin Lymphoma. *Int J Radiat Oncol Biol Phys* 2015;92(1):175-182.

Table 1. Co-registration scores: Dice Index (DI), Modified Hausdorff Distance (MHD), Dose-Organ Overlap (DOO), for pre- and post- Elastic Image Registration. The *p*-values express the significance of the inter-patient match improvement.

Score	DI		MHD (mm)		DOO	
	pre	post	pre	post	pre	post
Median	0.77	0.95	1.18	0.16	0.59	0.87
Range	[0.55,0.88]	[0.85,0.95]	[0.29,6.74]	[0.09,1.25]	[0.32,0.78]	[0.74,0.92]
p-value	$<10^{-17}$		$<10^{-17}$		$<10^{-17}$	

Table 2. Geometric (region volume) and dosimetric (dose delivered on average to the regions) characterization of the lung regions exhibiting a statistical significant difference between groups (p<0.05) according to the three applied statistical schemes. The p-values express the significance (Mann-Whitney U test) of the mean dose difference between groups of patients.

	Statistical Mapping Scheme					
	T		TFCE		*t*	
S volume (cm^3)	14.7		109		603	
	RILD	Non-RILD	RILD	Non-RILD	RILD	Non-RILD
Median dose (Gy)	2.35	1.45	5.89	3.74	8.21	5.00
Range (Gy)	[0.12,8.69]	[0.03,7.10]	[0.27,13.1]	[0.03,12.4]	[0.93,12.1]	[0.33,14.9]
p-value	0.005		0.011		0.001	

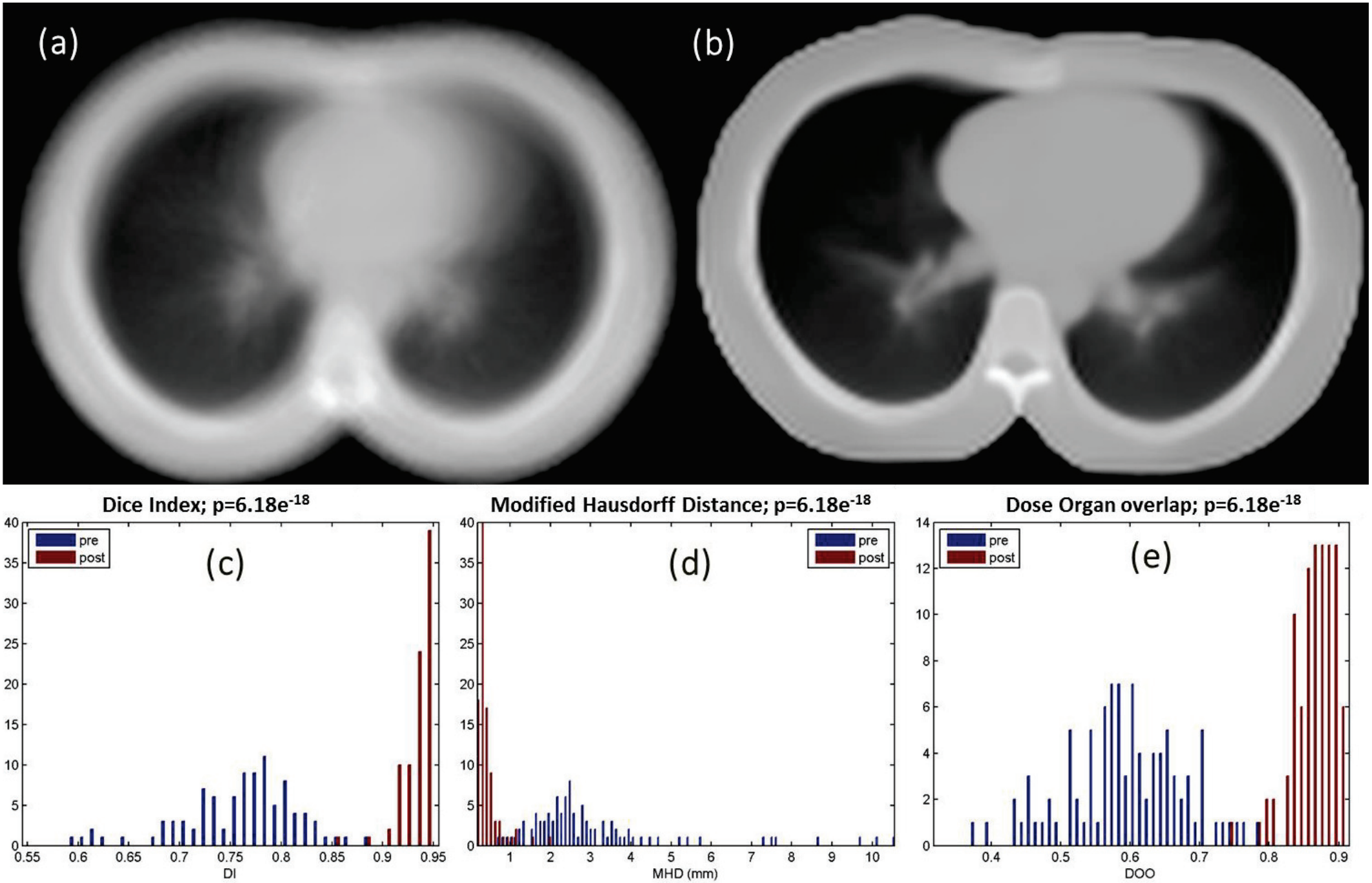
Figure Captions

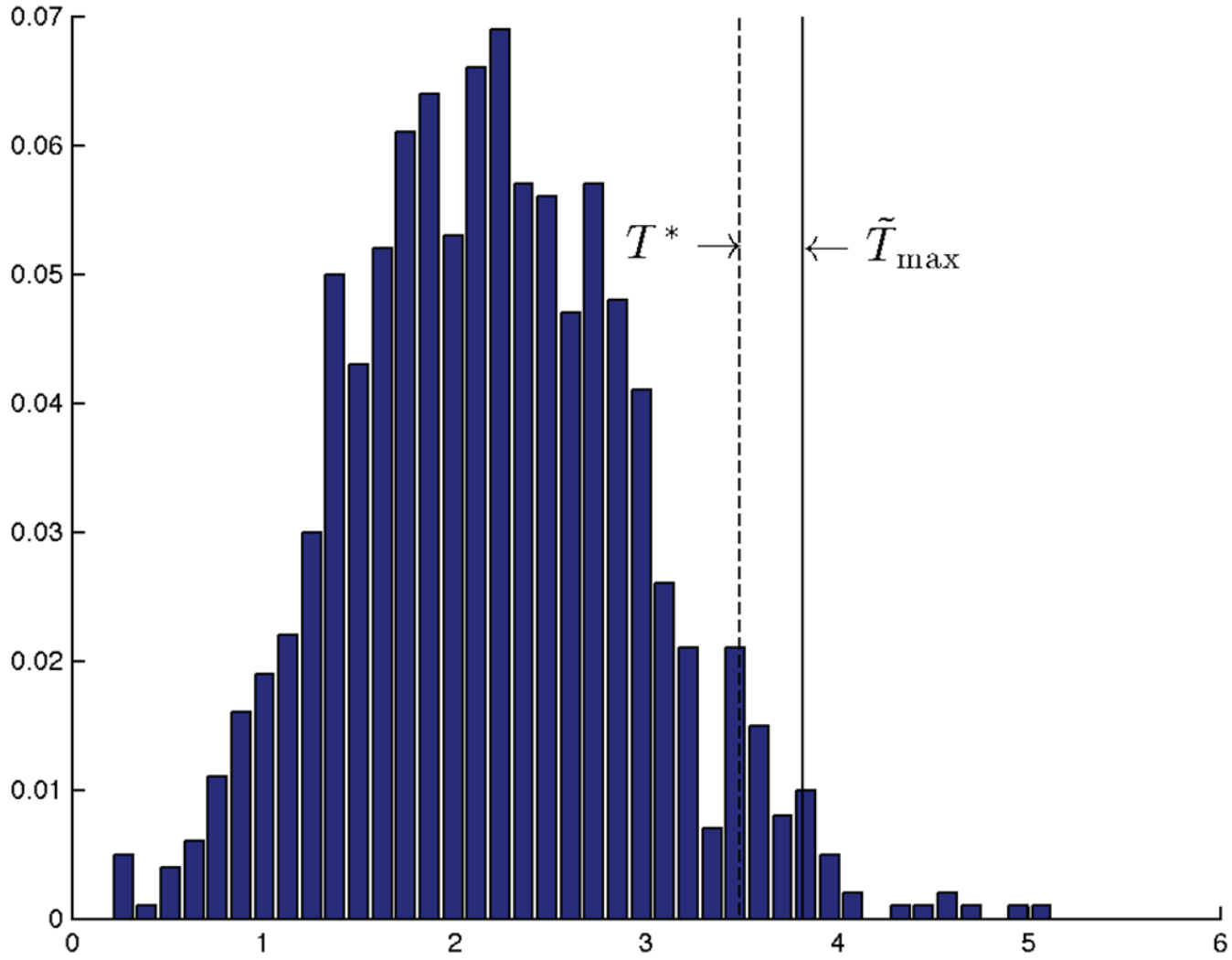
Figure 1. Elastic Image Registration (EIR) evaluation. Upper panel: Average of the patient population CTs before EIR (a) and after EIR (b). Lower panel: Distribution of (c) Dice Index (DI), (d) modified Hausdorff distance (MHD), and (e) dose-organ overlap (DOO) scores before and after EIR.

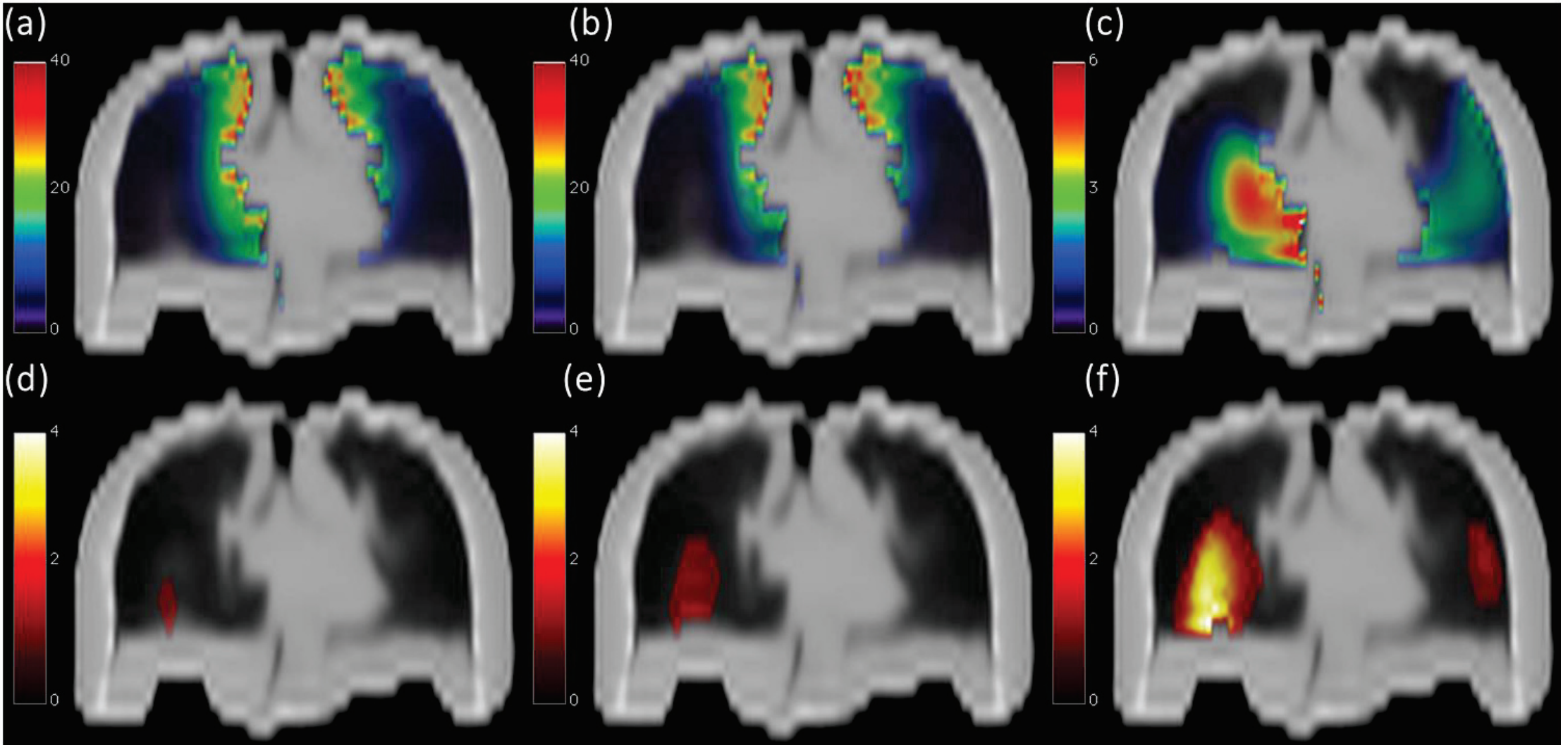
Figure 2. Distribution of normalized maximum dose difference ($T_{max,i}$) obtained from *i*= 1…. 1000 permutations. $\widetilde{T}_{max}$: normalized maximum dose difference in the observed sample; T^*: normalized maximum dose difference value corresponding to the 95 percentile of $T_{max,i}$. A significant adjusted *p*-value of 0.02 is obtained.

Figure 3. Upper panel: lung coronal view of the mean dose map (Gy) for patients (a) who developed radiation induced lung damage and (b) for patients who did not; (c) corresponding dose difference maps (mean dose map a minus mean dose map b). Lower panel: coronal view of lung sub-regions exhibiting a statistical significant dose difference between groups (p<0.05) according to permutation test *T* (d), TFCE test (e) and voxel-wise two sample *t*-test (f). The color map represents –log$p$. Images are displayed using radiological convention (patient's right at observer's left).

Figure 4. Comparison of ROC curves corresponding to normal tissue complication probability (NTCP) model based on the mean dose (D_{mean}) for each of lung sub-region (S_t, S_{TFCE}, S_T) and total lung.


(a)
(b)
Dice Index; p=6.18e-18
pre
post
(c)
DI
Modified Hausdorff Distance; p=6.18e-18
pre
post
(d)
MHD (mm)
Dose Organ overlap; p=6.18e-18
pre
post
(e)
DOO


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
1
2
3
4
5
6
$T^* \rightarrow$
$\leftarrow \tilde{T}_{\max}$


(a)
40
20
0
(b)
40
20
0
(c)
6
3
0
(d)
4
2
0
(e)
4
2
0
(f)
4
2
0