

SCIENTIFIC REPORTS



OPEN

Inter-patient image registration algorithms to disentangle regional dose bioeffects

Serena Monti¹, Roberto Pacelli², Laura Cella³ & Giuseppe Palma³

Radiation therapy (RT) technological advances call for a comprehensive reconsideration of the definition of dose features leading to radiation induced morbidity (RIM). In this context, the voxel-based approach (VBA) to dose distribution analysis in RT offers a radically new philosophy to evaluate local dose response patterns, as an alternative to dose-volume-histograms for identifying dose sensitive regions of normal tissue. The VBA relies on mapping patient dose distributions into a single reference case anatomy which serves as anchor for local dosimetric evaluations. The inter-patient elastic image registrations (EIRs) of the planning CTs provide the deformation fields necessary for the actual warp of dose distributions. In this study we assessed the impact of EIR on the VBA results in thoracic patients by identifying two state-of-the-art EIR algorithms (Demons and B-Spline). Our analysis demonstrated that both the EIR algorithms may be successfully used to highlight subregions with dose differences associated with RIM that substantially overlap. Furthermore, the inclusion for the first time of covariates within a dosimetric statistical model that faces the multiple comparison problem expands the potential of VBA, thus paving the way to a reliable voxel-based analysis of RIM in datasets with strong correlation of the outcome with non-dosimetric variables.

Elastic Image Registration (EIR) has recently gained momentum in the field of Radiation Therapy (RT). In particular, intra-patient EIR plays an integral role in modern Treatment Planning (TP) strategies, such as the Image Guided RT, which exploits the acquisition of several patient imaging datasets during the course of RT. In this context, intra-patient EIR has paved the way to real-time plan re-optimization¹ thanks to its capability to automatically track the contoured structures over time or to perform a dose warping. Intra-patient EIR was also interestingly used to study the RT-induced texture changes in serial computed tomography (CT) scans² or for 4D CT dose accumulation at different respiratory phases³.

Besides the relatively common application of intra-patient EIR to TP, the inter-patients version of EIR has been more recently exploited in population analysis to correlate local dose and radiation-induced morbidity (RIM). RIM has historically been estimated by condensing the 3D dose distribution into a monodimensional dose-volume histogram (DVH) that disregards spatial information of the dose. Increasing evidences, however, suggest that a voxel-based approach (VBA) allows to identify unprecedented correlations between RIM and local dose release. In particular, VBAs were applied to identify local dose effects for rectal⁴, gastrointestinal⁵ and urinary toxicity^{6,7} in prostate cancer patients. In addition, VBA was demonstrated suitable for unveiling potential spatial signatures of radiation sensitivity in inhomogeneous organs – such as the lungs⁸ – or in composite regions – such as the head and neck district⁹. Of note, so far none of the previous VBAs took into account non-dosimetric variables, such as patients' or treatment-related characteristics.

A key issue of a fully 3D VBA is represented by the EIR to a common anatomical reference (the so called *spatial normalization*) of the analyzed cohort of patients. The schemes for EIR most commonly used in medical imaging or RT applications are the Demons and B-spline algorithms^{10–14}.

In the context of RIM analyses, the Demons approach has been shown to guarantee excellent performance in inter-patient EIR, with a robust match of anatomical structures both from a pure geometrical point of view and in terms of dose-organ overlap^{4,8,9}. On the other hand, the B-spline parameterization approach was exploited in the spatial normalization of a cohort of lung cancer patients included in a survival analysis¹⁵.

¹IRCCS SDN, Napoli, Italy. ²Department of Advanced Biomedical Sciences, "Federico II" University School of Medicine, Napoli, Italy. ³Institute of Biostructures and Bioimaging, National Research Council, Napoli, Italy. Correspondence and requests for materials should be addressed to L.C. (email: laura.cella@cnr.it)

Structure		DI			MHD (mm)			DOO			RMSE (Hounsfield Units)					
		pre	B-spline	Demons	pre	B-spline	Demons	pre	B-spline	Demons	Pre	B-spline	Demons			
Lungs	Median	0.77	0.96	0.94	2.55	0.15	0.33	0.59	0.90	0.87	357	96	114			
	Range	[0.55,0.88]	[0.93,0.97]	[0.85,0.95]	[0.73,10.72]	[0.07,3.62]	[0.21,2.05]	[0.32,0.78]	[0.83,0.93]	[0.74,0.92]	[223,485]	[73,193]	[93,194]			
	<i>p</i> -value [§]		<10 ⁻¹⁶				<10 ⁻¹⁵				<10 ⁻¹⁶				<10 ⁻¹²	
Heart	Median	0.71	0.89	0.82	4.05	0.88	3.46	0.57	0.79	0.70	209	35	39			
	Range	[0.22,0.87]	[0.69,0.93]	[0.70,0.89]	[0.97,24.31]	[0.30,4.48]	[1.00,7.60]	[0.06,0.85]	[0.26,0.88]	[0.23,0.88]	[38,534]	[29,71]	[33,78]			
	<i>p</i> -value [§]		<10 ⁻¹⁵				<10 ⁻¹⁵				<10 ⁻⁸				<10 ⁻¹⁴	

Table 1. Registration scores pre- and post-elastic image registration by B-spline and Demons algorithms. DI = Dice Index, MHD = Modified Hausdorff Distance, DOO = Dose-Organ Overlap, RMSE = Root Mean Squared Error. [§]*p* values refer to the comparisons between B-spline and Demons algorithms.

To the best of our knowledge, none of the previous studies explored the influence of the EIR algorithm on highlighting the regional dose effects on RIM.

Given this background, the main focus of the present study was a comparative analysis of the regional dose effects on RIM when the Demons and B-spline algorithms are used within the VBA framework. The method was applied to a cohort of patients followed up for late lung toxicity after thoracic irradiation. We tested the null hypothesis of group differences (patients with RIM versus patients without RIM) on both physical dose and biologically effective dose (BED)¹⁶. Moreover, non-dosimetric variables were tested for significant correlation with RIM, and for the first time the VBA was designed to take into account their potential influence.

Methods and Materials

Patients' dataset. For the present analysis, we have considered a cohort of $N = 98$ patients treated for Hodgkin lymphoma (HL) with post-chemotherapy supradiaphragmatic involved-field 3D-conformal RT. 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. All patients were followed up for late pulmonary toxicity according to the Radiation Therapy Oncology Group scoring system¹⁷. At a median time to event of 13 months (range: [9, 83] months), 18 patients displayed radiation-induced CT radiological density changes (the RIM actually considered in the present analysis). All participants gave written informed consent and the patient data were analyzed anonymously. This retrospective study was approved by the local Ethics Committee (Comitato Etico per le Attività Biomediche, Università Federico II, Napoli, n. 222–10). All experimental protocols and procedures were performed in accordance with the guidelines of the Università Federico II, Napoli. Details on patients' and treatment characteristics (Supplementary Table) were previously reported¹⁸.

The contours of lung tissue and heart were reviewed on planning CTs following RTOG 1106¹⁷ and heart atlas contouring guidelines^{19,20}. The CT matrix size was 512×512 in plane with a slice thickness of 5 mm.

Voxel-based approach. 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) software²¹. BED maps were voxelwise extracted from dose maps for late toxicity effects ($\alpha/\beta = 3$ Gy) according to²². All processing steps described below were handled using in house software developed in Matlab.

The VBA consists of two main processes^{8,9}:

1. Spatial normalization of the patients' cohort, whose core part consists in the EIR of the planning CTs to a common anatomical reference and the consistent warp of the associated doses and BEDs;
2. Statistical analysis of regional dose differences between patients with and without RIM.

The patient with the median lung volume was chosen as the common anatomical reference (*i.e.* the common coordinate system – hereafter CCS) for the spatial normalization of the cohort. Before running the actual EIR of all other patients on the CCS, in order to enhance both the robustness and the efficiency of a potentially cumbersome registration task, CT scans were pre-processed as follows. For each patient, a binary mask of the region of interest was computed as the union and dilation (by a spherical structuring element of radius 30 mm) of heart and lung TP structures. The field-of-view was cropped accordingly and a coarse alignment of the structures of interest was obtained by an affine registration based on the mask boundary. CT images were also masked, in order to hide some inter-individual or gender-related anatomical differences of limited interest to our study, and to have the EIR algorithms work more effectively on tissue contrast inside the chest.

Then, the actual EIR was performed by open source implementations of two different algorithms: the B-spline Elastix registration²³ with Adaptive Stochastic Gradient Descent (ASGD) optimization algorithm, and the log diffeomorphic extension of the Demons registration²⁴; both methods were implemented with a multi-resolution strategy and adopted the mutual information as loss function. First, the patients' CTs were registered to the CCS, taking advantage of the anatomical details provided by the structural imaging; then, the obtained deformation fields were used to warp the dose and BED maps into the CCS as well. It should be emphasized that, since spatial normalization only aims at anchoring the next statistical analysis to a known and patient-independent anatomy, no Jacobian intensity modulation was applied to the deformed doses and BEDs.

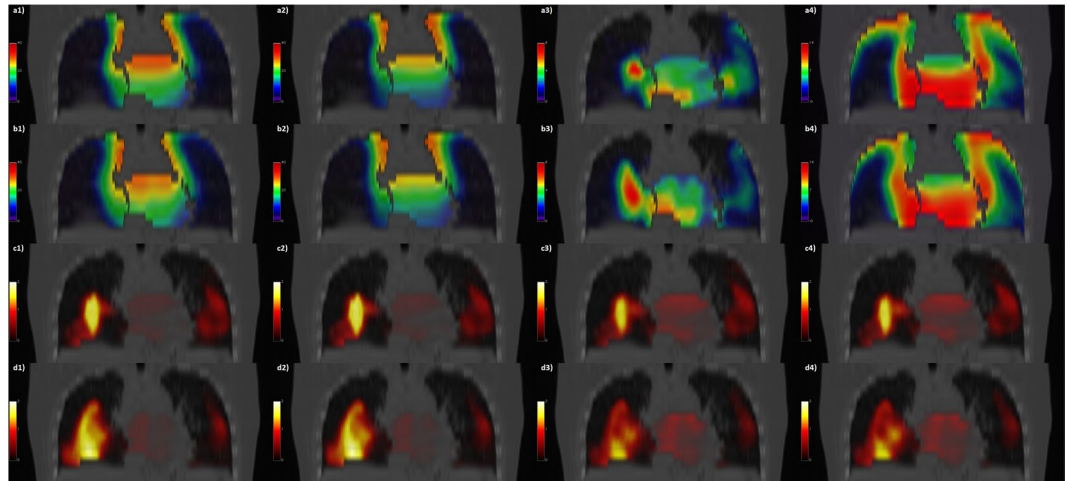


Figure 1. Coronal views of the thorax CT fused with dose-related maps (rows *a* and *b*) and with significance *p*-maps (rows *c* and *d*). Columns in rows *a* (B-spline) and *b* (Demons) show: 1) mean dose maps (Gy) for patients with RIM; 2) mean dose maps (Gy) for patients without RIM; 3) dose difference between 1) and 2); 4) dose standard deviation map within the cohort. Please note that color maps in columns 3) and 4) have different scales of 1) and 2). Columns in rows *c* (B-spline) and *d* (Demons) show the maps of $-\text{Log}(p)$ as derived from tests for: 1) GLM on dose; 2) GLM on BED; 3) GLM on dose and age; 4) GLM on BED and age.

GLM		Dose		BED	
		B-spline	Demons	B-spline	Demons
No age	p_{\min}	0.017	0.011	0.016	0.009
	$f(0.05)$	44.7 cc	118 cc	49.9 cc	93.7 cc
Age	p_{\min}	0.018	0.022	0.017	0.017
	$f(0.05)$	30.8 cc	46.6 cc	37.0 cc	51.1 cc

Table 2. Minimum p value and volume of significant cluster at $p = 0.05$ level for the analyzed configurations. Abbreviations: GLM = General Linear Model, BED = Biologically Effective Dose.

The median Hausdorff distances between the union M of heart and lung structures of the CCS and the corresponding spatially normalized unions of the $N-1$ patients were calculated for each EIR algorithm. These defined the full width at half maximum of the two spherical Gaussian kernels used to smooth the spatially normalized dose and BED maps²⁵ for the following regional statistical analysis.

Next, the statistical analysis constituting the second main process of the VBA was performed according to a test for general linear models (GLMs) counteracting the multiple comparison problem inherent in the task. In particular, a non-parametric permutation test²⁶ based on the Threshold Free Cluster Enhancement²⁵ of a maximum- T statistics allows for an excellent compromise between the sensitivity of the analysis and the control over imagewise Type I errors.

In order to adjust the GLM for possible non-dosimetric covariates, a univariate analysis of their differences among patients with and without RIM was first performed by Pearson's χ^2 - or Fisher's exact test (categorical variables) or Mann-Whitney U -test (continuous variables). A multivariate analysis with backward stepwise selection (based on the Wald statistic) was then performed including any variable having a univariate test p -value ≤ 0.25 ²⁷.

Afterwards, an extension of the former GLM was designed by adding a new regressor for each variable left in the regression model.

The voxel-based statistical analysis was performed by means of the “randomise” tool available in the FMRIB Software Library v5.0 (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL>)²⁸. The output of each voxel-based analysis was a p -map of dose or BED differences between the groups of patients with and without RIM, possibly adjusted for potential covariates.

Data analysis. The accuracy of the two different spatial normalizations was evaluated by computing the Dice Index (DI)²⁹, the modified Hausdorff distance (MHD)³⁰, the dose-organ overlap (DOO) index⁴ and the Root Mean Squared Error (RMSE). Wilcoxon signed rank test was used to compare each score achieved by the two different EIRs.

To evaluate the impact of EIR algorithm on the results of VBA, a Minkowski distance between two p -maps, A and B , was computed as

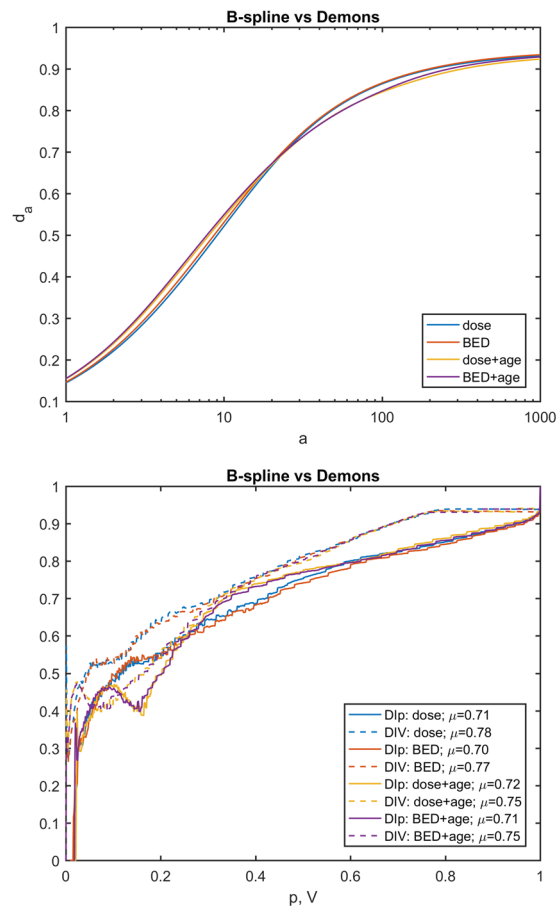


Figure 2. Plots of concordance metrics d_a , DIp and DIV (computed for each of the four pairs of choices C2 and C3) for comparison of p -maps derived from B-spline and Demons registration algorithms. In the legend, μ is the mean of the function over its domain.

$$d_a(A, B) = \left(\frac{1}{\lambda(M)} \int_M |A - B|^a d\lambda \right)^{1/a} \quad (1)$$

($\lambda(M)$ being the measure of the union of lungs and heart), and two further concordance metrics of the p -maps were *ad hoc* devised. The first one aims at evaluating the matching of the significance regions of two p -maps (e.g. the p -maps obtained by the VBA test without adjustment for covariate on the BED maps spatially normalized with the two EIRs) at a given confidence level. Denoting by S_p the sublevel set of a p -map for a given P value, the DI over p (DIp) between two p -maps A and B was defined for each $\min(\min(A), \min(B)) < P \leq 1$ as:

$$DIp[A, B](P) = DI(S_p[A], S_p[B]) \quad (2)$$

The second metrics provide a similar hint, except that the match is evaluated for an equal volume of the significant regions, instead of equal confidence level. Strictly speaking, if $f_X(P)$ is the relative volume of $S_p[X]$, the DI over volume (DIV) was defined for each $0 < P \leq 1$ as:

$$DIV[A, B](V) = DI(S_{f_A^{-1}(V)}[A], S_{f_B^{-1}(V)}[B]) \quad (3)$$

Data availability. The data analyzed in the present study are available at <http://www.ibb.cnr.it/?command=viewcms&id=216>.

Results

The paired comparison analysis of the DI, MHD, DOO and RMSE values (Table 1) computed on the whole cohort for Demons and B-spline showed that significantly better scores were obtained by the B-spline registration algorithm.

Age was the only non-dosimetric variable selected by multivariate regression analysis with backward elimination ($p=0.012$). Median age in patients with RIM was 32 yr (range: [22, 69] yr) vs 27 yr (range: [13, 58] yr) in patients without RIM. Age was therefore included within the GLMs adjusted for possible covariates.

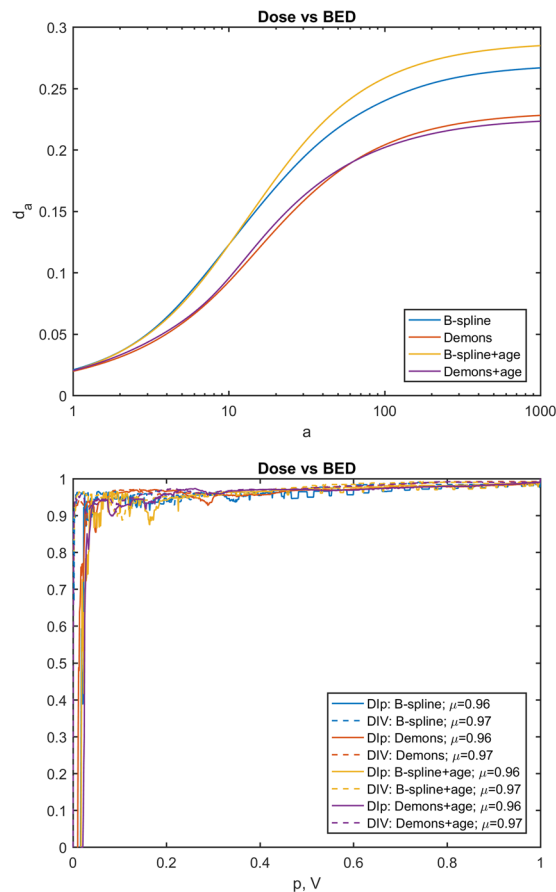


Figure 3. Plots of concordance metrics d_a , DIp and DIV (computed for each of the four pairs of choices C1 and C3) for comparison of p -maps derived for dose and BED. In the legend, μ is the mean of the function over its domain.

One significance map was obtained by the VBA for each of the eight triads (Fig. 1) given by the choice between B-spline or Demons (C1), dose or BED (C2), and adjustment for age or not (C3). All the p -maps⁴ highlight clusters of significant dosimetric differences between the two groups of patients, *i.e.* all $S_{0.05}[X_i]$ were non-empty (Table 2).

A visual inspection suggests that all the maps highlight clusters of high statistical significance in overlapping regions. The highest source of variability among the results is associated with the choice of the EIR algorithm, with a smaller influence exerted by the inclusion of age within the GLM, and an almost negligible effect of switching between dose and BED. This is confirmed by the quantitative analysis of the p -map concordance (Figs 2–4): when switching from C1 to C3 and from C3 to C2, d_a curves tend to decrease, while DIp and DIV show a consistent positive trend in the integral means of the functions.

Discussion

A quite large amount of EIR tools are nowadays available for intra- or inter-patient warping of structural datasets¹⁰. Many of them can be substantially classified as belonging to one of two transformation models: the Demons and the B-spline. Among these, we selected two highly effective algorithms for which an open-source implementation is available: the log diffeomorphic extension of the Demons registration²⁴ and the B-spline Elastix registration²³. Both approaches are known to have pros and cons³¹. Diffeomorphic Demons ensure the invertibility of the deformation fields and allow following complex warps thanks to the high number of degrees of freedom; conversely, the same high number of degrees of freedom leads to a potential sensitivity to noise. On the other hand, B-spline transforms, in spite of their flexibility, need some special care in order to ensure that non-invertible deformations are unlikely to occur³²: proper regularization terms can be added to the cost function to ensure smoothness and avoid singularities in the deformation field, or appropriate optimization algorithms (such as the ASCD adopted in this work) can be chosen to avoid local minima associated to undesirable deformation³³.

This study deals with a problem, namely the VBA, which requires reliability at an inter-patient level of the EIR process. This makes it intrinsically more challenging than tasks resting on an intra-patient registration. Indeed, the need to compare doses released in the same anatomical location to different patients demands the capability to overcome the high variability of examined structures potentially related to individual characteristics (gender, height, weight, etc.).

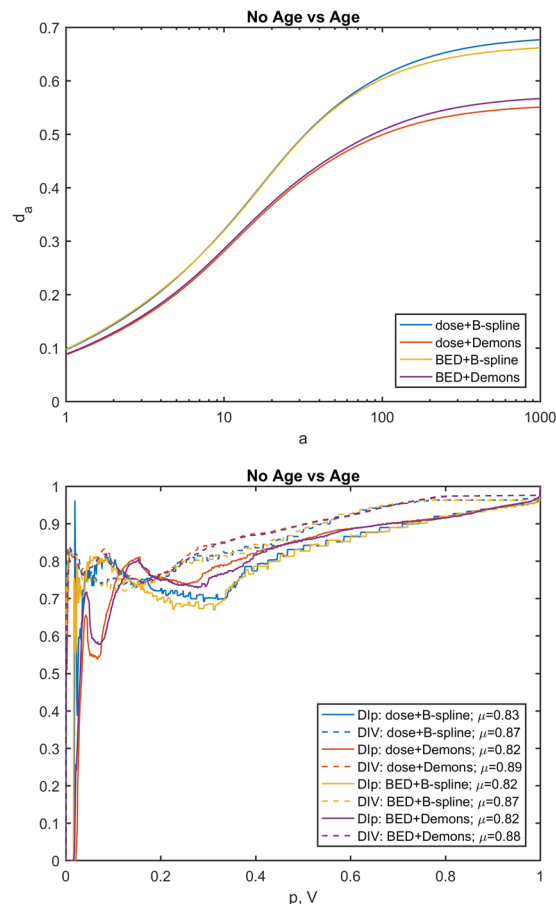


Figure 4. Plots of concordance metrics d_a , DIp and DIV (computed for each of the four pairs of choices C1 and C2) for comparison of p -maps derived from GLM adjusted or not for age. In the legend, μ is the mean of the function over its domain.

The VBA, performed by both EIR algorithms, was able to highlight a local dose–RIM relationship in the lungs, suggesting that the irradiation of peripheral parenchymal region in the middle and caudal lung is correlated with RIM (Fig. 1). In particular, as previously discussed in Palma *et al.*⁸, a higher dose was delivered in the low-dose parenchymal regions, in agreement with some recent DVH analyses showing that the lung volume exceeding 5 Gy is consistently more predictive for RIM than other dosimetric variables^{18,34}.

A large body of literature addressed the issue of quantifying the accuracy of the different registration algorithms under the action of a variety of deformation fields on a ground truth and according to several fidelity metrics¹⁰. Differently, our aim was to evaluate the overall impact of the EIR algorithms on the final output of the VBA pipeline. We believe that a non-trivial test bench can be the thorax region, due to the poor structural CT information content in relatively large scales of lung parenchyma. Moreover, the high inter-individual anatomical variations of bronchi and of related pulmonary vessels³⁵ hinder the exploit of the vessel CT contrast as a reliable landmark for the deformation process.

As a first step, we evaluated the EIR performance in heart and lungs by means of DI and MHD scores, both accounting for pure geometric match; in addition, DOO score was computed in order to weight the DI by the involved doses. It should be emphasized that these scores only give hints on boundary mismatch between structures, but are not able to measure the consistency of the deformation model well within the considered organs. In this respect, the RMSE is designed to take into account the anatomical match inside the organs. Both B-spline and Demons achieved results absolutely satisfactory in comparison with the pre-registration scores. Moreover, B-spline significantly outperformed Demons in all the considered metrics.

When we considered the entire VBA pipeline, the overlap of significant regions of dose-related differences – confirmed by both a visual inspection (Fig. 1) and a quantitative analysis of the p -maps (Figs 2–4) – at first justifies our confidence that results obtained by a VBA are quite stable against the choice of the EIR algorithm. Nonetheless, the adoption of the Demons or the B-spline scheme raises differences in the details of the significance pattern of dosimetric differences associated with the RIM development. This offers food for thought regarding the anatomical detail that can be achieved by the hitherto proposed VBA schemes. Namely, it seems confirmed that VBA allows for confidently going beyond the detail of an organ-based DVH approach; however, there is a limit to the level of substructures that can be identified.

This argument equally applies when we aim at considering the additional effects of BED or the inclusion of non-dosimetric covariates. Indeed, the comparison among the variability induced in the p -maps by C1, C2 and

C3 configurations highlights that effects associated with C2 (dose or BED) and C3 (adjustment for age or not) fall below the influence of the EIR algorithm choice. That being said, we recognize that the extent of observed effects of C2 and C3 may be highly dependent on the tested outcome, the considered patients' cohort and irradiation schemes. For instance, in this case, the low ratio between dose fraction in healthy tissues and α/β is likely to be responsible for the low influence we observed on p -maps when correcting the physical dose for biological effects.

The adoption of a VBA can have a twofold goal: the identification of the patterns of high radiosensitivity within organs and the definition of avoidance regions for sophisticated TP strategies⁹. This provides the rationale behind the definition of DI_p and DIV metrics to compare two p -maps: on the one hand, DI_p measures the concordance between clusters at the same significance level; DIV , instead, given that a tradeoff between the risk of RIM and tumor dose coverage has to be established, measures the match between the volumes at higher RIM risk for a given extension of the avoidance region.

In conclusion, in the present work we performed a comprehensive analysis of the impact of EIR on VBA, demonstrating that both the state-of-the-art EIR algorithms may be successfully used to highlight subregions with dose differences associated with RIM that substantially overlap, with the exception of small scale discrepancies. Furthermore, we expanded the potential of VBA by showing that covariates can be included within a statistical model facing the multiple comparison problem, thus paving the way to a reliable voxel-based analysis of RIM in datasets with strong correlation of the outcome with non-dosimetric variables.

We believe that it should be relevant to the radiation oncology community whether the recent findings of several pioneering studies^{4,8,9,15} are stable against the choice of the tool used in a key image processing step. In particular, as the regional dose patterns in RIM or survival analysis are receiving more and more attention, it is clinically crucial to lay robust methodological foundations for the whole voxel-based approach.

References

- Mohan, R. *et al.* Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes. *Int J Radiat Oncol Biol Phys* **61**, 1258–66 (2005).
- Cunliffe, A. *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* **91**, 1048–56 (2015).
- Rosu, M. *et al.* Dose reconstruction in deforming lung anatomy: dose grid size effects and clinical implications. *Med Phys* **32**, 2487–95 (2005).
- Acosta, O. *et al.* Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. *Phys Med Biol* **58**, 2581–95 (2013).
- Wortel, R. C. *et al.* Dose-surface maps identifying local dose-effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer. *Radiother Oncol* **117**, 515–20 (2015).
- Palorini, F. *et al.* First application of a pixel-wise analysis on bladder dose-surface maps in prostate cancer radiotherapy. *Radiother Oncol* **119**, 123–8 (2016).
- Yahya, N. *et al.* Modeling Urinary Dysfunction After External Beam Radiation Therapy of the Prostate Using Bladder Dose-Surface Maps: Evidence of Spatially Variable Response of the Bladder Surface. *Int J Radiat Oncol Biol Phys* **97**, 420–426 (2017).
- Palma, G. *et al.* A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage. *Int J Radiat Oncol Biol Phys* **96**, 127–33 (2016).
- Monti, S. *et al.* Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients. *Sci Rep* **7**, 7220 (2017).
- Murphy, K. *et al.* Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging* **30**, 1901–20 (2011).
- Mohamed, A. S. *et al.* Quality assurance assessment of diagnostic and radiation therapy-simulation CT image registration for head and neck radiation therapy: anatomic region of interest-based comparison of rigid and deformable algorithms. *Radiology* **274**, 752–63 (2015).
- Nithianathan, S. *et al.* Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. *Med Phys* **36**, 4755–64 (2009).
- Cunliffe, A. R. *et al.* Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration. *Med Phys* **39**, 4679–4690 (2012).
- Rigaud, B. *et al.* Evaluation of deformable image registration methods for dose monitoring in head and neck radiotherapy. *Biomed Res Int* **2015**, 726268 (2015).
- McWilliam, A. *et al.* Radiation dose to heart base linked with poorer survival in lung cancer patients. *Eur J Cancer* **85**, 106–113 (2017).
- Bentzen, S. M. *et al.* Bioeffect modeling and equieffective dose concepts in radiation oncology—terminology, quantities and units. *Radiother Oncol* **105**, 266–8 (2012).
- Cox, J. D., Stetz, J. & Pajak, T. F. 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* **31**, 1341–6 (1995).
- Cella, L. *et al.* Modeling the risk of radiation-induced lung fibrosis: Irradiated heart tissue is as important as irradiated lung. *Radiother Oncol* **117**, 36–43 (2015).
- Feng, M. *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* **79**, 10–8 (2011).
- Kong, F. M. *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* **81**, 1442–57 (2011).
- Deasy, J. O., Blanco, A. I. & Clark, V. H. CERR: a computational environment for radiotherapy research. *Med Phys* **30**, 979–85 (2003).
- Fowler, J. F. The linear-quadratic formula and progress in fractionated radiotherapy. *Br J Radiol* **62**, 679–94 (1989).
- Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* **29**, 196–205 (2010).
- Vercateren, T., Pennec, X., Perchant, A. & Ayache, N. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 754–761 (Springer, 2008).
- Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
- Holmes, A. P., Blair, R., Watson, J. & Ford, I. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow & Metabolism* **16**, 7–22 (1996).
- Bursac, Z., Gauss, C. H., Williams, D. K. & Hosmer, D. W. Purposeful selection of variables in logistic regression. *Source Code Biol Med* **3**, 17 (2008).

28. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *Neuroimage* **62**, 782–90 (2012).
29. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945).
30. Dubuisson, M. P. & Jain, A. K. In Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on 566–568 (IEEE, 1994).
31. Hub, M., Kessler, M. L. & Karger, C. P. A stochastic approach to estimate the uncertainty involved in B-spline image registration. *IEEE Trans Med Imaging* **28**, 1708–16 (2009).
32. Kanai, T. *et al.* Evaluation of accuracy of B-spline transformation-based deformable image registration with different parameter settings for thoracic images. *J Radiat Res* **55**, 1163–70 (2014).
33. Klein, S., Staring, M. & Pluim, J. P. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans Image Process* **16**, 2879–90 (2007).
34. Pinnix, C. C. *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* **92**, 175–82 (2015).
35. Boyden, E. A. Segmental Anatomy of the Lungs. A Study of the Patterns of the Segmental Bronchi and Related Pulmonary Vessels (Blakiston Division, McGraw-Hill, New York, 1955).

Author Contributions

L.C. and G.P. conceived and designed the study. R.P. participated in patient recruitment and collected the data. G.P., S.M., L.C. processed and analyzed the data. G.P., L.C. wrote the paper. All authors have reviewed the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-23327-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018