Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics

Fanny Orlhac, Frédérique Frouin, Christophe Nioche, Nicholas Ayache, Irène Buvat

▶ To cite this version:

HAL Id: hal-02401340

https://hal.archives-ouvertes.fr/hal-02401340

Submitted on 10 Dec 2019

Validation of a Method to Compensate Multicenter Effects Affecting CT Radiomics

Fanny Orlhac, PhD • *Frédérique Frouin, PhD* • *Christophe Nioche, PhD* • *Nicholas Ayache, PhD* • *Irène Buvat, PhD*

From the UCA, Inria Sophia Antipolis–Méditerranée, Epione, 2004 route des Lucioles–BP 93, 06 902 Sophia Antipolis Cedex, France (F.O., N.A.); and Imagerie Molécu-laire In Vivo, CEA-SHFJ, Inserm, CNRS, Université Paris-Sud, Université Paris-Saclay, Orsay, France (F.F., C.N., I.B.). Received August 30, 2018; revision requested October 8; revision received December 6; accepted December 12. **Address correspondence to** F.O. (e-mail: *orlhacf@gmail.com*).

Conflicts of interest are listed at the end of this article.

See also the editorial by Steiger and Sood in this issue.

Background: Radiomics extracts features from medical images more precisely and more accurately than visual assessment. However, radiomics features are affected by CT scanner parameters such as reconstruction kernel or section thickness, thus obscuring under-lying biologically important texture features.

Purpose: To investigate whether a compensation method could correct for the variations of radiomic feature values caused by using different CT protocols.

Materials and Methods: Phantom data involving 10 texture patterns and 74 patients in cohorts 1 (19 men; 42 patients; mean age, 60.4 years; September–October 2013) and 2 (16 men; 32 patients; mean age, 62.1 years; January–September 2007) scanned by using different CT protocols were retrospectively included. For any radiomic feature, the compensation approach identified a pro-tocol-specific transformation to express all data in a common space that were devoid of protocol effects. The differences in statistical distributions between protocols were assessed by using Friedman tests before and after compensation. Principal component analyses were performed on the phantom data to evaluate the ability to distinguish between texture patterns after compensation.

Results: In the phantom data, the statistical distributions of features were different between protocols for all radiomic features and texture patterns ($P < .05$). After compensation, the protocol effect was no longer detectable ($P > .05$). Principal component analysis demonstrated that each texture pattern was no longer displayed as different clusters corresponding to different imaging pro-tocols, unlike what was observed before compensation. The correction for scanner effect was confirmed in patient data with 100% (10 of 10 features for cohort 1) and 98% (87 of 89 features for cohort 2) of P values less than .05 before compensation, compared with 30% (three of 10) and 15% (13 of 89) after compensation.

Conclusion: Image compensation successfully realigned feature distributions computed from different CT imaging protocols and should facilitate multicenter radiomic studies.

© RSNA, 2019

Online supplemental material is available for this article.

Since 2012, the use of radiomics has expanded (1) in on-cology with the objective to characterize tumor hetero-geneity from medical images. Radiomics extracts features from medical images that quantify tumor shape, intensity histogram, and texture of the lesions more precisely and more accurately than visual assessment by a radiologist to build models that involve features to assist patient treat-ment. In particular, texture analysis from CT images has led to promising results to distinguish between tumor le-sions with different histopathologic characteristics and to predict treatment response or patient survival (2). How-ever, several studies highlighted the sensitivity of radiomic features to CT acquisition and reconstruction parameters by using phantoms (3–8) or patient data (9–12). Indeed, feature values are affected by section thickness, pixel size, reconstruction kernel, tube voltage, tube current, and contrast agent enhancement. They also differ between dif-ferent scanners with the same settings (8). Moreover, the effect of imaging protocols varies according to the texture pattern and the radiomic feature (4).

One of the most widely cited studies (13) in radiomics, which included 1019 patients, used different CT imaging protocols involving different CT scanners, pixel size, and section thickness with or without intravenous contrast en-hancement, and without accounting for this variability in the data analysis. To reduce that variability, it was proposed to resample images with a fixed voxel size, to filter the im-ages (5), or to change the definition of features (6,11). These approaches require a modification of the CT images or are not applicable to all radiomic features.

The same issue is encountered in PET imaging, in which radiomic features are sensitive to the acquisition protocol and reconstruction algorithm (14). A compen-sation method was initially described in genomics (15), where the so-called batch effect is the source of variations in measurements caused by the handling of samples by dif-ferent laboratories, different technicians, and on different days. The batch effect is conceptually similar to variations induced by the scanner or the protocol effects in radiomics. The compensation method identifies a batch-specific

Summary

Nonbiological differences related to CT scanner type can be removed from radiomic feature values, allowing radiomics features to be combined in multicenter or multivendor studies.

Key points

- Radiomic feature values obtained by using different CT imaging protocols or scanners can be corrected for the protocol or scanner effect by using the proposed compensation method.
- The use of realigned features will enable multicentric studies to pool data from different sites to build reliable radiomic models based on large databases.
- The proposed compensation method is easily available, fast, and requires neither phantom acquisition nor feature recalculation.

Table 1: Patient Characteristics

Parameter	Patient Cohort 1	Patient Cohort 2
Sex		
Men	19 (45)	16 (50)
Women	23 (55)	16 (50)
Mean age (y)*	60.4 (31–81)	62.1 (29–82)
No. of primary lung lesions	8 (19)	29 (91)
No. of metastatic lesions	34 (81)	3 (9)

Note.—Unless otherwise indicated, data are the number of patients and data in parentheses are percentages.

* Data in parentheses are range.

transformation to express all data in a common space devoid of batch effects. It has been shown (16) to be effective in PET to realign the radiomic feature distributions between three different protocols for healthy liver tissue and breast lesions, without altering the biologic information. The purpose of our study was therefore to determine whether this compensation method could also correct for the CT protocol effect by using phantom and patient data.

Materials and Methods

All patient data were anonymized and are publicly available in the supplemental data of Kim et al (9) and Lu et al (10). All authors had control of the data and information submitted for publication.

Phantom Experiments

The phantom data used in our study have been produced by Mackin et al (4) and are publicly available in the associated supplemental data (4). The Credence Cartridge Radiomics phantom consists of 10 layers with different materials corresponding to different texture patterns. This phantom was scanned by using 17 different imaging protocols from four medical institutes involving various reconstruction kernels, scan types, section thickness, pixel spacing, spiral pitch factor, and effective milliamperage. Additional information on phantom and acquisition characteristics are provided in Tables E1 and E2 (online). For each layer, 16 non-overlapping volumes of interest with an average cubic volume of 8 cm³ (range, 7.6–9 cm³; corresponding to 2708–14332 voxels depending on the imaging protocols) are also made available in Dicom-RTstruct format. For each volume of interest and each imaging protocol, we (F.O. and C.N., with 7 years and 20 years of research experience in medical imaging, respectively) computed 40 radiomic features by using the LIFEx freeware (17) (Inserm, Orsay, France, *www.lifexsoft.org*; Table E3 [online]) with a fixed bin size (18) set to 10 HU between —1000 HU and 3000 HU without any spatial resampling. We performed the radiomic analysis for 16 of 17 imaging protocols because of a reading issue with acquisition Credence Cartridge Radiomics 1-GE2.

Patients

Publicly available radiomic features from two patient databases (cohort 1 and cohort 2) were used in our study. The first set

of features was derived from cohort 1 (42 patients with lung cancer; mean age, 60.4 years; age range, 31–81 years; Table 1) between September and October 2013 (9), including 19 men. All patients underwent a CT scan with the same machine and protocol (Table E4 [online]), and CT images were reconstructed by using three algorithms: filtered back projection and Sinogram Affirmed Iterative Reconstruction (Siemens Healthcare, Forchheim, Germany) with a noise reduction strength of level 3 (hereafter, referred to as S3) and 5 (hereafter, referred to as S5). For each patient, the dominant tumor lesion was segmented manually three times: twice by a radiologist and once by a technologist. For each of the three volumes of interest per patient and each reconstruction, 15 radiomic features were calculated. Five geometric features (volume, diameter, surface, sphericity, and compactness) were excluded from the analysis (F.O.) because they mostly depend on the segmentation.

A second set of features was obtained from cohort 2 (32 patients; mean age, 62.1 years; age range, 29–82 years; Table 1) between January and September 2007 (16 men) with a lung cancer who underwent two CT scans (Table E4 [online]) within 15 minutes (10). This dataset was originally collected in a clinical trial (NCT00579852) to evaluate the reproducibility of tumor volume and diameter measurements and is part of the Reference Image Database to Evaluate Therapy Response project (19). The CT images were reconstructed by using six protocols that combined two reconstruction algorithms (lung and standard) and three section thicknesses (1.25 mm, 2.5 mm, and 5 mm) (10). For one lesion per patient (29 primary and three metastatic lesions), a tumor volume of interest was obtained from a consensus among the manual segmentations by three radiologists. After resampling the volume-of-interest voxels to 0.5 × 0.5 × 0.5 mm³ by using a trilinear interpolation, 89 radiomic features were calculated for the six imaging protocols (two reconstructions × three section thicknesses) and for each of the 64 scans (32 patients with two scans each).

Compensation Method

To correct for differences in features caused by the various imaging protocols, we used the ComBat function (*https://github.com/Jfortin1/ComBatHarmonization*) compensation method (15). This method has been used for cortical thickness measurements from MR images (20) and for radiomic features from different PET protocols (16). It is a data-driven method

that identifies the protocol effect assuming that the value of each feature, y, measured in volume of interest, j, with imaging protocol, i, can be written as

$$y_{ij} = \alpha + \gamma_i + \delta_i \varepsilon_{ij} \qquad (1),$$

where α is the average value for feature y_{ij}, γ_i is an additive protocol effect, and δ_i is a multiplicative protocol effect affected by an error term (ε_{ij}). The compensation consists in estimating the model parameters α, γ_i, and δ_i by using a maximum likelihood approach on the basis of the set of available observations:

$$y_{ij}^{ComBat} = \frac{y_{ij} - \hat{\alpha} - \widehat{\gamma_i}}{\widehat{\delta_i}} + \hat{\alpha} \qquad (2),$$

in which $\hat{\alpha}$, $\widehat{\gamma_i}$ and $\widehat{\delta_i}$ are estimators of α, γ_i, and δ_i.

We used the nonparametric form of the model in which no assumptions are made regarding the laws followed by the parameters. In this setting, ComBat determines a transformation for each feature separately. For each texture pattern of phantom data and of each patient dataset, we used the ComBat function in R (version 3.4.2; R Foundation for Statistical Computing, Vienna, Austria) to identify the transformation parameters.

Statistical Analysis

Statistical analysis was performed by using software (R; R foundation for Statistical Computing). To determine whether the protocol setting (independent variable i in Equation [1]) affected the distributions of radiomic feature values (dependent variables y_{ij} in Equation [1]), we (F.O. and F.F., with 30 years of experience) performed two-sided Friedman tests before and after ComBat compensation for each feature as summarized in Table E5 (online). The null hypothesis is that there is no difference between the distributions. Benjamini-Hochberg procedure was used to control the false discovery rate (21). P values less than .05 indicated statistical significance. Because the goal of the compensation is to realign the distributions in terms of mean and standard deviation, a P value of the Friedman test greater than .05 indicated that the realignment was successful.

For the phantom data, we also performed a principal component analysis of the 2560 samples (16 volume of interest × 10 texture patterns × 16 imaging protocols) described by 40 variables (radiomic features). Principal component analysis was performed before and after ComBat to view the effect of the compensation method on the distinction between patterns. We also studied whether two textural patterns could be distinguished when pooling data from the three imaging protocols before and after compensation and for balanced and unbalanced groups.

Results

Patient characteristics are shown in Table 1.

Phantom Experiments

In the phantom data, 399 of 400 P values of the Friedman tests performed for all features on the basis of 16 imaging protocols

Table 2: Significant Friedman Tests without and with Compensation in the Phantom and Clinical Cohorts

Parameter	Without Realignment	With Realignment
Phantom		
Pattern 1, 20% filled ABS	40/40 (100)	0/40 (0)
Pattern 2, 30% filled ABS	40/40 (100)	0/40 (0)
Pattern 3, 40% filled ABS	40/40 (100)	0/40 (0)
Pattern 4, 50% filled ABS	40/40 (100)	0/40 (0)
Pattern 5, acrylic	40/40 (100)	0/40 (0)
Pattern 6, cork	40/40 (100)	0/40 (0)
Pattern 7, dense cork	39/40 (98)	0/40 (0)
Pattern 8, plaster resin	40/40 (100)	0/40 (0)
Pattern 9, rubber particles	40/40 (100)	0/40 (0)
Pattern 10, wood	40/40 (100)	0/40 (0)
Patient cohort 1	10/10 (100)	3/10 (30)
Patient cohort 2	87/89 (98)	13/89 (15)

Note.—Data are numerator/denominator; data in parentheses are percentages. For the phantoms, there were 40 tests for each pattern. There were 10 tests for patient cohort 1 and 89 tests for patient cohort 2. ABS = acrylonitrile butadiene styrene.

and 10 texture patterns were less than .05 before compensation (Tables 2, E6 [online]). Only one P value for skewness was greater than .05 for pattern 7 (dense cork; $P = .46$). After compensation, all P values of Friedman tests were greater than .05, which indicated that the protocol effect was no longer detectable.

These results were visually confirmed by using the projection of the data in the space spanned by the first two principal components of principal component analysis. Figure 1 shows an overlap between textural patterns before ComBat because of the large variability of radiomic feature values computed from 16 different CT protocols. For each textural pattern (each color), several clusters corresponding to different CT protocols could be identified. After ComBat, textural patterns could be clearly distinguished and were no longer composed of different clusters, demonstrating that the compensation method properly corrected for the scanner effect while retaining the specific characteristics of each texture pattern. The variance explained by the first two components was higher after ComBat (65.6% vs 53.2%), with approximately the same features contributing to the first two principal components before and after compensation (data not shown).

On the basis of three CT acquisitions (GE1, P2, and S2; Table E2 [online]), Figure 2 shows that when data were pooled without realignment, the sensitivity for distinguishing cork from dense cork was 67% (32 of 48 volumes of interest) with a specificity of 98% (47 of 48 volumes of interest) by using the cutoff maximizing the Youden index. After ComBat, both sensitivity and specificity were 100% (48 of 48 volumes of interest). For unbalanced groups, Figure E1 (online) shows that the compensation method also yielded a perfect distinction between these two patterns.

Patient Data

For patient cohorts 1 and 2, 100% (10 of 10) and 98% (87 of 89), respectively, of Friedman tests had P values less than
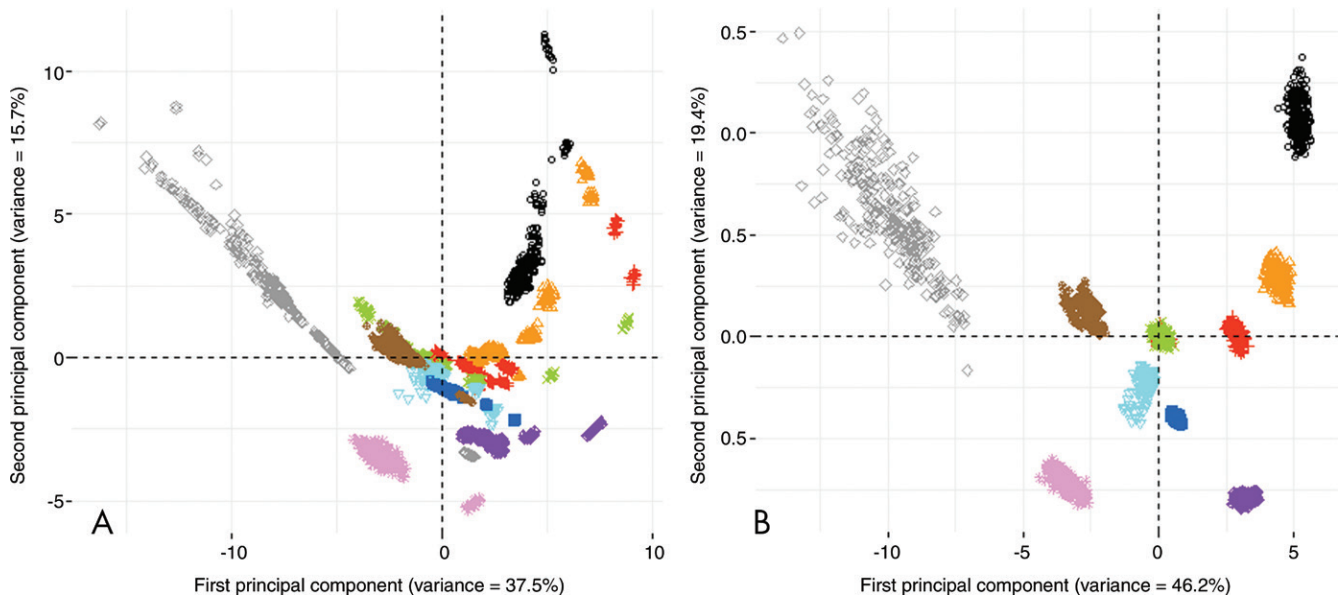
Figure 1: Phantom data: principal component scores for 2560 samples corresponding to 16 volume of interest × 10 texture patterns (represented by 10 different colors in the figure) × 16 imaging protocols described by 40 radiomic features on the first two principal components before, *A*, and after, *B*, ComBat (*https://github.com/Jfortin1/ComBatHarmonization*). After compensation, each texture pattern (ie, each color) was no longer composed of several separate and different clusters. Instead, the 10 texture patterns are nonoverlapping, demonstrating that the effect of the CT scanner protocol has been removed.
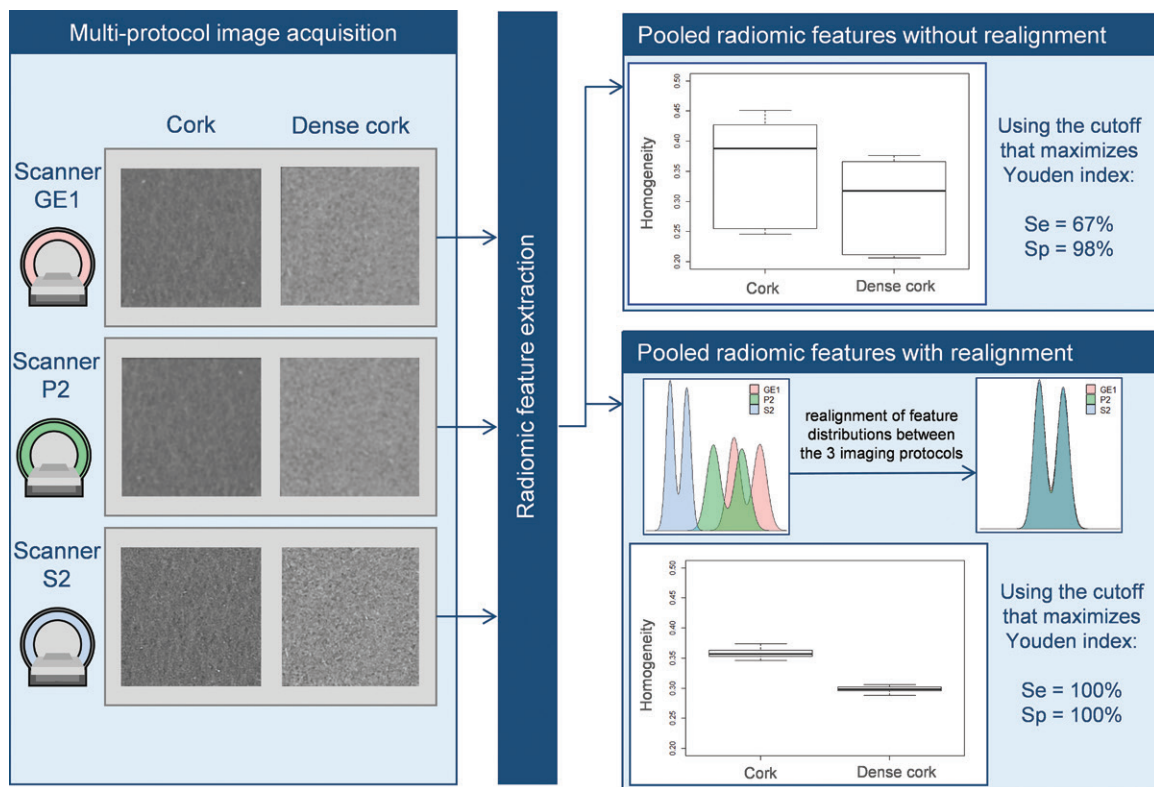

Figure 2: Example of ComBat application in phantom experiments. Two texture patterns (cork and dense cork) were scanned by using three different imaging protocols with 16 volumes of interest in each case (GE1, P2, S2; Table E2 [online]). When pooling all radiomic feature values, the optimal cutoff could not perfectly distinguish the patterns; a perfect distinction was observed after compensation of scanner effects. *Se* = sensitivity, *Sp* = specificity.

.05 between imaging protocols before ComBat (Tables 2, E7, E8 [online]). After ComBat, 30% (three of 10) of *P* values for cohort 1 and 15% (13 of 89) of *P* values for cohort 2 were less than .05. Visual inspection of the radiomic feature value dis-

tributions when Friedman tests remained statistically significant after ComBat showed that the residual difference between protocols was always small and that the protocol effect was reduced (Fig E2 [online]), demonstrating the effectiveness of
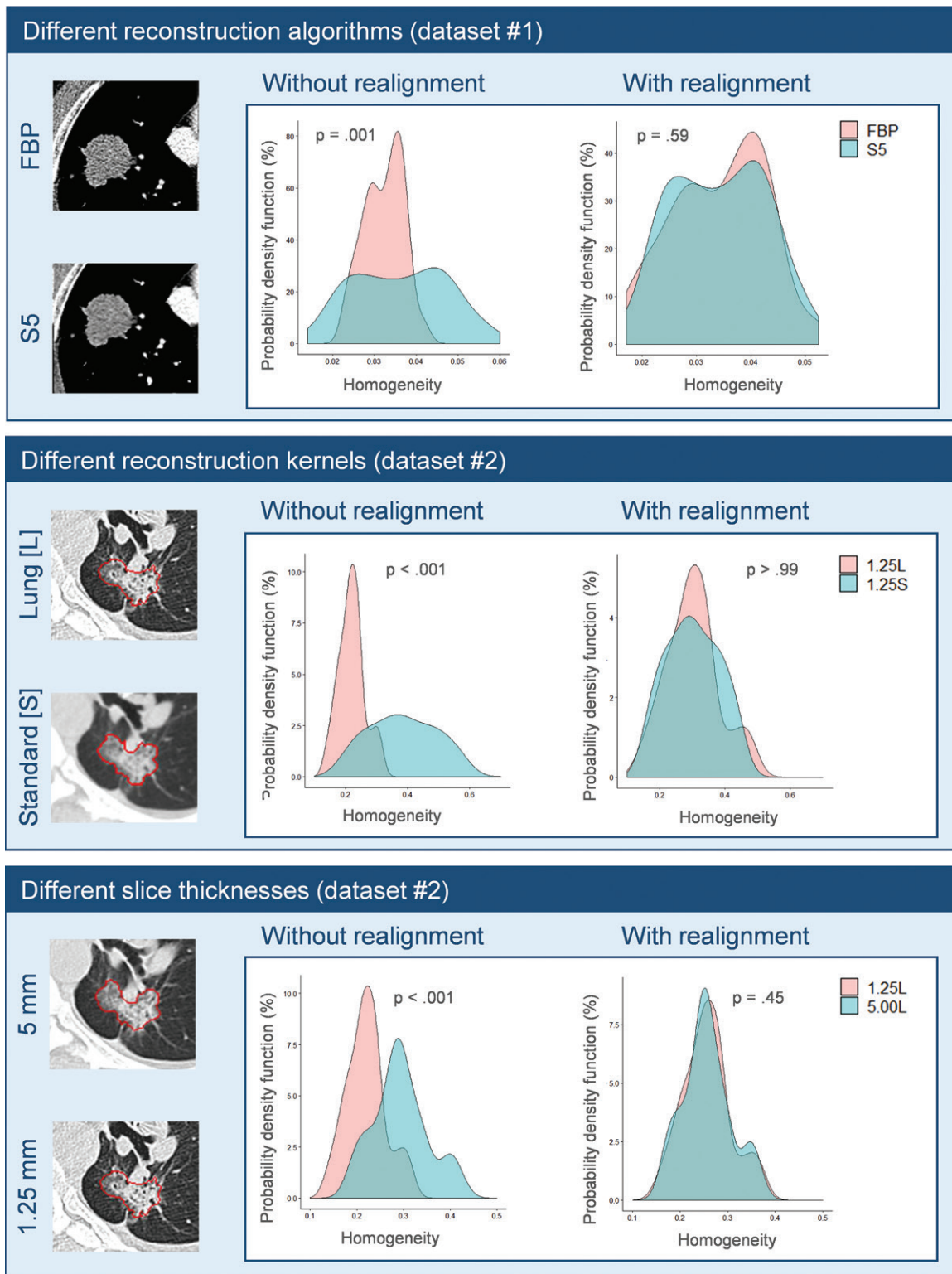
Figure 3: Probability density function of homogeneity before (without realignment) and after (with realignment) ComBat in patient data by using two CT reconstruction algorithms, two reconstruction kernels, and two voxel thicknesses. *P* values are for Friedman tests. FBP = filtered back projection. CT images reprinted, under a CC BY license, from references 9 and 10.

the compensation. ComBat corrected the protocol effect with a realignment of feature values among the three protocols for cohort 1 and among the six protocols for cohort 2 (homogeneity feature in Fig E3 [online]). For example, before ComBat,

the plot shows a shift in distribution with greater homogeneity values for reconstruction algorithm S5 than for S3 and filtered back projection. This was expected because reconstruction S5 involved higher noise reduction. After compensation, the dis-

tributions between the three reconstructions better overlapped. Figure 3 shows three examples of realignment of features between different reconstruction algorithms, reconstruction kernels, and section thicknesses.

Discussion

Radiomic features are sensitive to the acquisition and reconstruction parameters of CT images. Feature values are therefore not directly comparable between different imaging protocols, which limits their use in multicenter studies. We demonstrated that the ComBat method can realign radiomic features computed from different CT imaging protocols. By using phantom data, we showed that ComBat removed the scanner and protocol effect while preserving the differences between texture patterns. The correction for the scanner effect was confirmed by using patient images reconstructed with different imaging protocols.

The use of this compensation method should facilitate multicentric radiomic analyses that are needed to demonstrate the practical usefulness of radiomic features for patient care. Data standardization is a subject of interest in the international imaging community, with increasing awareness of the need to reduce the variability in image quality between centers and machines (22,23). ComBat offers a solution to realign radiomic features with several advantages. ComBat is easily available to all and fast (a function available for free in R software; R Project for Statistical Computing). The transformations are estimated on the basis of the measured feature values, without the need to go back to images or to perform phantom experiments. No learning set is needed. ComBat does not change the feature definitions (6,11) and can therefore be used with all software and algorithms and with any radiomic features. This is illustrated in three data sets by using three different implementations for the radiomic feature calculation (Table E5 [online]). ComBat does not require spatial resampling of the CT images to a single pixel size and/or image filtering (5). It is applicable when radiomic features values are available while images are not available. ComBat can account for covariates of interest if the patients scanned with different imaging protocols do not have the same characteristics (eg, different age distributions). ComBat can model these covariates in the compensation process, as illustrated in PET imaging with different proportions of cancer subtypes in different departments (16), if enough patients with the same characteristics are available.

As demonstrated by using the phantom data (Table 2), before compensation the values of radiomic features were significantly different between imaging protocols for a given pattern. Ignoring this effect in multicentric studies might bias the findings and reduce statistical power. A recent study (24) showed that the four features selected in Aerts et al (13) to build the radiomic signature were correlated with tumor volume, which might explain why the model remained robust on data from different centers. The use of the ComBat method might help to determine whether radiomic features reflecting the lesion biologic heterogeneity but affected by the center effect more than the lesion volume also have some predictive value.

Our study had limitations. Our findings should be confirmed by using other cancer types for other imaging protocols

and scanners, and the actual effect on diagnostic performance on clinical data needs to be demonstrated. Independent multicenter validation of radiomic models is also essential for them to become mainstream (1,25,26).

In conclusion, ComBat makes it possible to pool radiomic features from different CT protocols. This method appears promising to address the center effect in multicenter radiomic studies and to possibly raise the statistical power of those studies. ComBat is data driven, which means that the transformations identified by ComBat to set all data in a common space should be estimated for each study involving data from different centers and protocols. Our analysis was on the basis of less than 50 patients for each acquisition protocol, which demonstrated the efficiency of the method even for small patient cohorts. By using simulations in which we gradually removed patient data (results not shown), we found satisfactory results with as few as 20 patients per imaging protocol. The minimum number of patients required per imaging protocol to successfully apply ComBat remains to be comprehensively investigated.

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–577.
2. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: definitions, applications, biologic correlates, and challenges. RadioGraphics 2017;37(5):1483–1503.
3. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in ct characterization of tumors: a preliminary phantom study. Transl Oncol 2014;7(1):88–93.
4. Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015;50(11):757–765.
5. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. PLoS One 2017;12(9):e0178524 [Published correction appears in PLoS One 2018;13(1):e0191597.].
6. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44(3):1050–1062.
7. Caramella C, Allorant A, Orlhac F, et al. Can we trust the calculation of texture indices of CT images? A phantom study. Med Phys 2018;45(4):1529–1536.
8. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 2018;288(2):407–415.
9. Kim H, Park CM, Lee M, et al. Impact of reconstruction algorithms on CT radiomic features of pulmonary tumors: analysis of intra- and inter-reader variability and inter-reconstruction algorithm variability. PLoS One 2016;11(10):e0164924.
10. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing agreement between radiomic features computed for multiple CT imaging settings. PLoS One 2016;11(12):e0166550.
11. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. Sci Rep 2018;8(1):10545.
12. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 2016;6(1):34921.
13. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5(1):4006.
14. Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med 2015;56(11):1667–1673.
15. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118–127.

16. Orlhac F, Boughdad S, Philippe C, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. J Nucl Med 2018;59(8):1321–1328.

17. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res 2018;78(16):4786–4789.

18. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. PLoS One 2015;10(12):e0145063.

19. Armato SG 3rd, Meyer CR, Mcnitt-Gray MF, et al. The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. Clin Pharmacol Ther 2008;84(4):448–456.

20. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 2018;167:104–120.

21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289–300.

22. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. Radiology 2015;277(3):813–825.

23. O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14(3):169–186.

24. Vallieres M, Visvikis D, Hatt M. Dependency of a validated radiomics signature on tumor volume and potential corrections. J Nucl Med 2018;59(Suppl 1):640 http://jnm.snmjournals.org/content/59/supplement_1/640.

25. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? J Nucl Med 2015;56(11):1642–1644.

26. Reuzé S, Schernberg A, Orlhac F, et al. Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges. Int J Radiat Oncol Biol Phys 2018;102(4):1117–1142.