Universiteit
Leiden
The Netherlands

Predicting and evaluating side effects of radiotherapy in cervical cancer

Corbeau, A.

Citation
Corbeau, A. (2026, April 2). *Predicting and evaluating side effects of radiotherapy in cervical cancer*. Retrieved from https://hdl.handle.net/1887/4300428

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4300428

Note: To cite this publication please use the final published version (if applicable).

Chapter 6

Accuracy, repeatability, and reproducibility of water-fat magnetic resonance imaging in a phantom and healthy volunteer

Anouk Corbeau, Pien van Gastel, Piotr A. Wielopolski, Nick de Jong, Carien L. Creutzberg, Uulke A. van der Heide, Stephanie M. de Boer, Eleftheria Astreinidou

Physics and Imaging in Radiation Oncology, 2024, 32:100651.

ABSTRACT

Bone marrow (BM) damage due to chemoradiotherapy can increase BM fat in cervical cancer patients. Water-fat magnetic resonance (MR) scans were performed on a phantom and a healthy female volunteer to validate proton density fat fraction accuracy, reproducibility, and repeatability across different vendors, field strengths and protocols. Phantom measurements showed a high accuracy, high repeatability, and excellent reproducibility. Volunteer measurements had an excellent intra- and interreader reliability, good repeatability, and moderate to good reproducibility. Water-fat MRI show potential for quantification of longitudinal vertebral BM fat changes. Further studies are needed to validate and extend these findings for broader clinical applicability.

INTRODUCTION

Bone marrow (BM) is composed of red BM, which is actively involved in generating blood cells, and inactive yellow BM, which stores fat and has reduced hematopoietic activity [1]. Chemoradiation of pelvic cancers depletes the BM population and allows mesenchymal stem cells to differentiate towards adipocytes, leading to an increase in the fatty yellow marrow [2, 3]. Chemoradiation-induced BM damage causes hematologic toxicity (HT), defined as a reduced number of circulating blood cells.

Primary chemoradiation is the standard treatment for women with locally advanced cervical cancer [4, 5]. HT, especially lymphopenia, due to this treatment can lead to the discontinuation of chemotherapy, the need for blood transfusions, and decreased overall survival [6-8]. Water-fat MRI can generate proton density fat fraction (PDFF) maps. Multiple studies used these PDFF maps to detect BM fat changes in the lumbar vertebrae of gynecological cancer patients treated with chemoradiotherapy, which was related to the number of circulating blood cells [9-11]. Water-fat MRI could identify locations of active and radiation-induced damaged BM and facilitate the development of BM sparing radiotherapy treatment planning techniques and the evaluation of treatment.

However, the studies evaluating BM PDFF in patients with gynecologic cancer have several limitations [9-11]. Studies were not performed in a multicenter or multivendor setting, which decreases the external validity of the outcomes, or do not compare PDFF of vertebrae inside and outside the radiation field, which limits the quantification of radiotherapy impact specifically. Therefore, a prospective, multicenter study was designed to longitudinally assess PDFF changes in BM of the whole vertebral column in women with locally advanced cervical cancer treated with BM sparing chemoradiotherapy [12].

The Radiological Society of North America Quantitative Imaging Biomarkers Alliance (QIBA) recommends to validate protocols under study conditions to determine its inherent reliability to measure the endpoint [13]. The aim of this study is to measure the accuracy, repeatability, and reproducibility of water-fat MRI in a multivendor setting for the assessment of PDFF to use in a longitudinal, prospective study.

METHODS

Data acquisition

The Calimetrix PDFF phantom (Fat Fraction Phantom, Model 300, Calimetrix, USA) consists of a spherical acrylic housing containing twelve gel-based vials with known

reference PDFF values in a doped water bath. The nominal values of the vials are 0, 2.7, 5.3, 7.8, 10.1, 15.5, 20.4, 23.6, 30.2, 39.9, 50.1, and 100%, as indicated in Figure S1. More details on the PDFF phantom can be found in the study by Hu et al (2021) [14]. The volunteer was a healthy, 52-year old woman. The study was approved by the Medical Ethics Review Committee Erasmus Medical Center and Medical Ethics Review Committee Leiden The Hague Delft. Written informed consent was obtained prior to examination.

Clinical 1.5T and 3T MR scanners by the vendors Philips (Philips Healthcare, the Netherlands) (P1.5T and P3.0T) and General Electric (GE Healthcare, USA) (G1.5T and G3.0T) were used. Scanner parameters and settings are listed in Supplementary Table S1. As the vendor-specific quantitative 3D chemical shift-encoding based scanning technologies were developed for hepatic fat quantification, the resolution of the protocols was optimized for the evaluation of the (smaller) vertebrae [15, 16]. For P1.5T and P3.0T, both vendor-specific protocols (v1) and reoptimized protocols with improved resolution (v2) were assessed. For G1.5T and G3.0T, only the optimized protocols were assessed. Vendors' commercial software for online PDFF map reconstruction was used to obtain the PDFF maps. This software includes multipeak fat spectral modeling, T2* confounding effect correction, and T1 bias reduction.

The phantom acclimated at least one hour to the room temperature before scanning, according to the vendor's instructions. The volunteer was scanned in supine position with a field of view extending from the third cervical vertebra to the lesser trochanter of the femur using three to four overlapped axial stacks with three to four table positions. The coils used are specified in the Supplementary Material. The stacks were combined with MR MobiView (Philips Healthcare, The Netherlands) for the Philips system and with MR Autobind (GE Healthcare, USA) for the GE system. Repositioning took place between test and retest scans to imitate interscan position differences. All scans took place within a time window of 30 days and were acquired at a temperature of 19-20°C.

Data analysis

Methods for placing regions-of-interest (ROI) on the PDFF map are detailed in the Supplementary Material. The mean PDFF, including standard deviation, minimum, and maximum value, was calculated over all voxels per ROI and expressed as a percentage [%].

Statistical analysis

Intrareader variation was assessed using two-way mixed-effects intraclass correlation coefficients (ICCs) [17]. Additionally, interreader agreement was assessed with the two-way random-effects ICCs. The level of reliability was classified according to Koo et al (2016) [17].

Pearson correlation (*r*) and Bland-Altman analysis were used to determine the accuracy of the first (test) scanning session of the phantom [18]. Bias and limits of agreement (LoA) were computed with the Bland-Altman analysis according to the difference between measured and reference PDFF as a function of the mean of these two values and have the same unit as the absolute PDFF [PDFF%].

Paired t-tests were performed to compare test and retest volunteer scans per vertebra and reference tissue. The repeatability coefficient (RC) represents the least significant difference between two repeated measurements taken under identical conditions in absolute PDFF [PDFF%] (see Supplementary Material) and was determined per protocol [19].

Reproducibility across all protocols was evaluated using Lin's concordance correlation coefficient (CCC) (see Supplementary Material) and classified according to Carbonell et al. [20] . Furthermore, Bland-Altman analysis was performed to assess bias and LoA across protocol versions, field strengths, and manufacturers for both phantom and volunteer measurements.

Statistical analysis was performed using R for Statistical Computing (version 4.3.1) and epiR package was used for equation 2. Statistical significance level was defined as $p<0.05$.

RESULTS

Intra- and interreader variability

The intrareader ICC for the PDFF measurements across the six protocols was 0.97 [95% CI, 0.96 – 0.98]. The interreader ICC for the PDFF measurements across the six protocols was 0.99 [95% CI, 0.98 – 0.99]. Both the intra- and interreader ICC were >0.9 and indicated an excellent reliability of the PDFF measurements in the volunteer scans.

Accuracy

Figure S2 shows that PDFF measurements were highly correlated to the reference values of the phantom. Absolute differences between measured PDFF and reference values were between -2.81 PDFF% and 6.98 PDFF%. Pearson correlation coefficients of the PDFF measurements were between 0.9967 and 0.9998. Pooled PDFF measurements from the Philips MR systems, GE MR systems, 1.5 T systems, 3 T systems, and all protocols combined likewise showed high correlation with the reference values, all having $r = 1.00$ [95% CI, 1.00 – 1.00]. The mean bias of the phantom measurements ranged between -0.57 PDFF% and 1.09 PDFF% [95% LoA, -4.6, 5.78 PDFF%], showing a slight underestimation of the true PDFF with P3.0T-v1, P3.0T-v2, and G3.0T, and a slight overestimation with P1.5T-v1, P1.5T-v2, and G1.5T.

Repeatability

Figure 1A. and 1B. demonstrate the within-examination (test-retest) PDFF maps for the healthy volunteer with the G3.0T protocol. Qualitative inspection of the PDFF maps identified some breathing artefacts mainly around the organs such as the heart and the liver that did not affect the vertebral column. Figure 1C. visualizes the PDFF and corresponding standard deviations per vertebra on the within-examination scans. The adipose tissue and muscle PDFF ranged from respectively 78.0% to 93.7% and 2.30% to 6.12% (Figure S3). Paired t-tests in Table S2 showed no significant differences between test and retest scans, except for the first sacral vertebra ($p = 0.03$).

The RC ranged from 0.61 PDFF% to 1.34 PDFF% for the phantom measurements, from 4.14 PDFF% to 6.04 PDFF% for the volunteer BM measurements, which means that 95% of the PDFF measurements repeated under the same conditions were below these values.

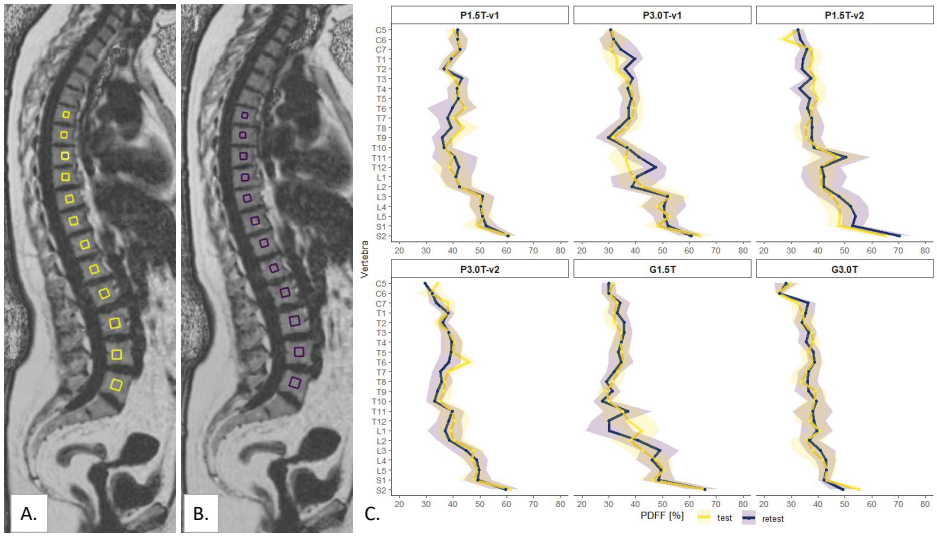

Figure 1: *Repeated PDFF maps of the healthy volunteer acquired with the 3 T General Electric Signa Premier (General Electric, USA) in test (A) and retest (B) setting. In this image, yellow (test) or purple (retest) box-shaped regions-of-interest (ROIs) are positioned in thoracic vertebrae six to twelve and lumbar vertebrae one to five. Note that for the measurements the ROIs were placed in all vertebrae between the fifth cervical vertebra and the second sacral vertebra, but this is not visualized in this figure. The ROIs did not include the bone cortex and, if present, physiological or pathological deviations. The mean PDFF and corresponding standard deviation per vertebra for repeatability measurements per protocol are visualized in C. The images in A. and B. correspond to G3.0 T in Figure C. PDFF = proton density fat fraction, P1.5T = Philips 1.5 T, P3.0T = Philips 3 T, G1.5T = General Electric 1.5 T, G3.0T = General Electric 3 T, v1 = version 1, v2 = version 2, C = cervical, T = thoracic, L = lumbar, S = sacral.*

Reproducibility

The Bland-Altman plots in Figure 2 show good agreement of PDFF measurements across protocol versions, field strengths, and manufactures for both phantom and volunteer measurements. Lower BM PDFF values were measured with protocol versions 2 compared to versions 1, 3T scanners compared to 1.5T scanners, and with GE scanners compared to Philips scanners. The bias for the muscle ranged from -1.20 PDFF% [LoA -3.36, 0.95 PDFF%] to 2.13 PDFF% [LoA 0.95, 3.32 PDFF%] and for adipose tissue from -9.43 PDFF% [LoA -26.66, 7.8 PDFF%] to 2.76 PDFF% [LoA -1.60, 7.10 PDFF%] (results not shown in Figure 2). Furthermore, the CCC for the phantom measurements were excellent, with the highest CCC for P1.5T-v1 v.s. P1.5T-v2 (0.9997) and the lowest CCC for Philips v.s. GE (0.9982). For the volunteer the CCC was moderate to good with the highest CCC for P3.0T-v1 v.s. P3.0T-v2 (0.8463) and the lowest CCC for G1.5T v.s. G3.0T (0.6701).

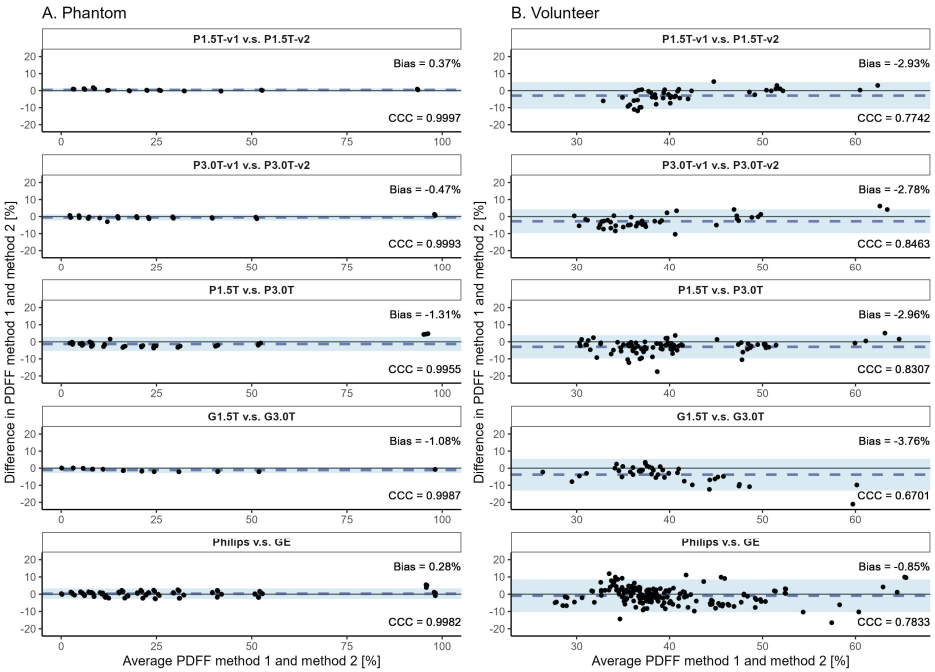

Figure 2: *Reproducibility of phantom (A.) and volunteer (B.) fat fraction measurements expressed with the concordance correlation coefficient (CCC) and Bland-Altman analysis (bias and limits of agreement (LoA) [PDFF%]). PDFF = proton density fat fraction, P1.5T = Philips 1.5 T, P3.0T = Philips 3 T, G1.5T = General Electric 1.5 T, G3.0T = General Electric 3 T, v1 = version 1, v2 = version 2.*

DISCUSSION

In this study, the use of quantitative water-fat MRI in a multivendor setting was validated with measurements with a PDFF phantom and a healthy, female volunteer. The phantom measurements showed a high correlation with the reference values, were highly repeatable, and had an excellent reproducibility when comparing vendors, field strengths, and scan protocols. Water-fat scans of the volunteer had limited artefacts in vertebral BM, an excellent intra- and interreader reliability for vertebral BM, and a good repeatability for both vertebral BM and reference tissues. The reproducibility was lower than that of the phantom measurements, but nevertheless showed moderate to good agreement across manufacturers, field strengths, and protocols.

The phantom measurements showed excellent results, whereas the volunteer measurements had a lower repeatability, a lower CCC, and a higher mean bias for reproducibility. Various biological factors, such as regional inhomogeneities in the vertebrae, and MR scanning

characteristics could contribute to these differences. It is important to inspect the scans on potential artefacts to identify factors that could influence the measured PDFF. The detected biases are, however, clinically less relevant as the fat fraction of BM within the irradiation field, including the sacral and lumbar vertebrae, could increase from 43–50% towards 63-74% [9-11].

Our study serves as an example of a robust methodology to assess the accuracy, repeatability, and reproducibility of PDFF measurements. This method can be applied before the implementation of quantitative MRI techniques in the clinic or for longitudinal clinical studies and should be repeated after software or hardware upgrades of the MR scanner.

Our study had several limitations. The commercially available phantom used had a limited PDFF range, so PDFF between 50% and 100% could not be validated optimally. The accuracy could also not be evaluated for the volunteer, since there was no reference standard used such as ^{1}H-MR spectroscopy [21]. However, the lumbar vertebral PDFF of the female volunteer in our study was comparable with the lumbar vertebral PDFF in 50-year old healthy females in other water-fat MR studies [22-24] and measurements of the adipose tissue served as an additional PDFF datapoint between 50% and 100%. Several factors could introduce inaccuracies in the measurement of BM PDFF using water-fat MR, including the effects of T1-related bias [25-27], T2* decay [27, 28], and spectral complexity of fat [29, 30]. Small flip angles ranging from 3 to 6° [25], the acquisition of six echoes [31], incorporation of a T2* map into the reconstruction algorithm, and the use of a multi-peak fat spectrum [32] accommodates for these factors. Another limitation of this study is that only two measurements were performed per scanner. Nevertheless, the validation outcomes are in line with earlier studies [21, 33]. Lastly, the most significant limitation is that only one female volunteer was scanned, which diminishes the robustness and generalizability of the findings. Caution should be exercised when extrapolating our results to different cohorts.

In conclusion, the water-fat MR scans used in this study have an excellent accuracy, repeatability, and reproducibility for PDFF phantom measurements, as well as an acceptable repeatability and reproducibility in vertebral BM fat quantification in a healthy 52-year old female volunteer. Since the detected biases in BM PDFF may have limited clinical relevance compared to the substantial increase in BM PDFF during chemoradiotherapy, our water-fat MR protocols show potential for quantification of BM

fat changes due to chemoradiotherapy in a prospective, longitudinal, multicenter study. Further studies with larger datasets are needed to validate and extend these findings for broader clinical applicability.

SUPPLEMENTARY MATERIAL

Supplementary material is available on

REFERENCES

1. Babyn PS, Ranson M, McCarville ME. Normal Bone Marrow: Signal Characteristics and Fatty Conversion. Magn Reson Imaging Clin N Am. 1998;6(3):473-95.
2. Green DE, Rubin CT. Consequences of irradiation on bone and marrow phenotypes, and its relation to disruption of hematopoietic precursors. Bone. 2014;63:87-94.
3. Naveiras O, Nardi V, Wenzel PL, Hauschka PV, Fahey F, Daley GQ. Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. Nature. 2009;460(7252):259-63.
4. Horeweg N, Creutzberg CL, Rijkmans EC, Laman MS, Velema LA, Coen VL, et al. Efficacy and toxicity of chemoradiation with image-guided adaptive brachytherapy for locally advanced cervical cancer. International Journal of Gynecologic Cancer. 2019;29(2).
5. Pötter R, Tanderup K, Schmid MP, Jürgenliemk-Schulz I, Haie-Meder C, Fokdal LU, et al. MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-I): a multicentre prospective cohort study. The lancet oncology. 2021;22(4):538-47.
6. Abu-Rustum NR, Lee S, Correa A, Massad LS. Compliance with and acute hematologic toxic effects of chemoradiation in indigent women with cervical cancer. Gynecol Oncol. 2001;81(1):88-91.
7. Damen PJ, Kroese TE, van Hillegersberg R, Schuit E, Peters M, Verhoeff JJ, et al. The influence of severe radiation-induced lymphopenia on overall survival in solid tumors: a systematic review and meta-analysis. International Journal of Radiation Oncology* Biology* Physics. 2021;111(4):936-48.
8. Wu ES, Oduyebo T, Cobb LP, Cholakian D, Kong X, Fader AN, et al. Lymphopenia and its association with survival in patients with locally advanced cervical cancer. Gynecol Oncol. 2016;140(1):76-82.
9. Bolan PJ, Arentsen L, Sueblinvong T, Zhang Y, Moeller S, Carter JS, et al. Water–fat MRI for assessing changes in bone marrow composition due to radiation and chemotherapy in gynecologic cancer patients. J Magn Reson Imaging. 2013;38(6):1578-84.
10. Carmona R, Pritz J, Bydder M, Gulaya S, Zhu H, Williamson CW, et al. Fat composition changes in bone marrow during chemotherapy and radiation therapy. International Journal of Radiation Oncology* Biology* Physics. 2014;90(1):155-63.
11. Wang C, Qin X, Gong G, Wang L, Su Y, Yin Y. Correlation between changes of pelvic bone marrow fat content and hematological toxicity in concurrent chemoradiotherapy for cervical cancer. Radiation Oncology. 2022;17(1):1-7.
12. Corbeau A, Nout RA, Mens JWM, Horeweg N, Godart J, Kerkhof EM, et al. PROTECT: prospective phase-II-trial evaluating adaptive proton therapy for cervical cancer to reduce the impact on morbidity and the immune system. Cancers. 2021;13(20):5179.
13. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27-67.
14. Hu HH, Yokoo T, Bashir MR, Sirlin CB, Hernando D, Malyarenko D, et al. Linearity and Bias of Proton Density Fat Fraction as a Quantitative Imaging Biomarker: A Multicenter, Multiplatform, Multivendor Phantom Study. Radiology. 2021;298(3):640-51.

15. Kukuk GM, Hittatiya K, Sprinkart AM, Eggers H, Gieseke J, Block W, et al. Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. Eur Radiol. 2015;25(10):2869-79.
16. Yokoo T, Serai SD, Pirasteh A, Bashir MR, Hamilton G, Hernando D, et al. Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. Radiology. 2018;286(2):486-98.
17. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.
18. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141-51.
19. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. J Magn Reson Imaging. 2019;49(7):e101-e21.
20. Carbonell G, Kennedy P, Bane O, Kirmani A, El Homsi M, Stocker D, et al. Precision of MRI radiomics features in the liver and hepatocellular carcinoma. Eur Radiol. 2021:1-11.
21. Schmeel FC, Vomweg T, Träber F, Gerhards A, Enkirch SJ, Faron A, et al. Proton density fat fraction MRI of vertebral bone marrow: Accuracy, repeatability, and reproducibility among readers, field strengths, and imaging platforms. J Magn Reson Imaging. 2019;50(6):1762-72.
22. Baum T, Rohrmeier A, Syväri J, Diefenbach MN, Franz D, Dieckmeyer M, et al. Anatomical Variation of Age-Related Changes in Vertebral Bone Marrow Composition Using Chemical Shift Encoding-Based Water–Fat Magnetic Resonance Imaging. Front Endocrinol (Lausanne). 2018;9.
23. Burian E, Syväri J, Dieckmeyer M, Holzapfel C, Drabsch T, Sollmann N, et al. Age- and BMI-related variations of fat distribution in sacral and lumbar bone marrow and their association with local muscle fat content. Sci Rep. 2020;10(1):9686.
24. Dieckmeyer M, Ruschke S, Cordes C, Yap SP, Kooijman H, Hauner H, et al. The need for T2 correction on MRS-based vertebral bone marrow fat quantification: implications for bone marrow fat fraction age dependence. NMR Biomed. 2015;28(4):432-9.
25. Liu C-Y, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: Correction of bias from T1 and noise. Magn Reson Med. 2007;58(2):354-64.
26. Yang IY, Cui Y, Wiens CN, Wade TP, Friesen-Waldner LJ, McKenzie CA. Fat fraction bias correction using T1 estimates and flip angle mapping. J Magn Reson Imaging. 2014;39(1):217-23.
27. Le Ster C, Gambarota G, Lasbleiz J, Guillin R, Decaux O, Saint-Jalmes H. Breath-hold MR measurements of fat fraction, T1, and T2* of water and fat in vertebral bone marrow. J Magn Reson Imaging. 2016;44(3):549-55.
28. Karampinos DC, Melkus G, Baum T, Bauer JS, Rummeny EJ, Krug R. Bone marrow fat quantification in the presence of trabecular bone: Initial comparison between water-fat imaging and single-voxel MRS. Magn Reson Med. 2014;71(3):1158-65.
29. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging. 2012;36(5):1011-4.

30. Karampinos DC, Ruschke S, Dieckmeyer M, Diefenbach M, Franz D, Gersing AS, et al. Quantitative MRI and spectroscopy of bone marrow. J Magn Reson Imaging. 2018;47(2):332-53.

31. Yu H, McKenzie CA, Shimakawa A, Vu AT, Brau ACS, Beatty PJ, et al. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. J Magn Reson Imaging. 2007;26(4):1153-61.

32. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R estimation with multifrequency fat spectrum modeling. Magn Reson Med. 2008;60(5):1122-34.

33. Jang JK, Lee SS, Kim B, Cho E-S, Kim YJ, Byun JH, et al. Agreement and reproducibility of proton density fat fraction measurements using commercial MR sequences across different platforms: a multivendor, multi-institutional phantom experiment. Invest Radiol. 2019;54(8):517-23.