ORIGINAL RESEARCH

Accuracy and longitudinal consistency of PET/MR attenuation correction in amyloid PET imaging amid software and hardware upgrades

Chunwei Ying, Yasheng Chen, Yan Yan, Shaney Flores, Richard Laforest, Tammie L. S. Benzinger, and Hongyu An

ABSTRACT

BACKGROUND AND PURPOSE: Integrated PET/MR allows the simultaneous acquisition of PET biomarkers and structural and functional MRI to study Alzheimer disease (AD). Attenuation correction (AC), crucial for PET quantification, can be performed using a deep learning approach, DL-Dixon, based on standard Dixon images. Longitudinal amyloid PET imaging, which provides important information about disease progression or treatment responses in AD, is usually acquired over several years. Hardware and software upgrades often occur during a multiple-year study period, resulting in data variability. This study aims to harmonize PET/MR DL-Dixon AC amid software and head coil updates and evaluate its accuracy and longitudinal consistency.

MATERIALS AND METHODS: Tri-modality PET/MR and CT images were obtained from 329 participants, with a subset of 38 undergoing tri-modality scans twice within approximately three years. Transfer learning was employed to fine-tune DL-Dixon models on images from two scanner software versions (VB20P and VE11P) and two head coils (16-channel and 32-channel coils). The accuracy and longitudinal consistency of the DL-Dixon AC were evaluated. Power analyses were performed to estimate the sample size needed to detect various levels of longitudinal changes in the PET standardized uptake value ratio (SUVR).

RESULTS: The DL-Dixon method demonstrated high accuracy across all data, irrespective of scanner software versions and head coils. More than 95.6% of brain voxels showed less than 10% PET relative absolute error in all participants. The median [interquartile range] PET mean relative absolute error was 1.10% [0.93%, 1.26%], 1.24% [1.03%, 1.54%], 0.99% [0.86%, 1.13%] in the cortical summary region, and 1.04% [0.83%, 1.36%], 1.08% [0.84%, 1.34%], 1.05% [0.72%, 1.32%] in cerebellum using the DL-Dixon models for the VB20P-16-channel-coil, VE11P-16-channel-coil and VE11P-32-channel-coil data, respectively. The within-subject coefficient of variation and intra-class correlation coefficient of PET SUVR in the cortical regions were comparable between the DL-Dixon and CT AC. Power analysis indicated that similar numbers of participants would be needed to detect the same level of PET changes using DL-Dixon and CT AC.

CONCLUSIONS: DL-Dixon exhibited excellent accuracy and longitudinal consistency across the two software versions and head coils, demonstrating its robustness for longitudinal PET/MR neuroimaging studies in AD.

ABBREVIATIONS: AC = attenuation correction; AD = Alzheimer disease; HU = Hounsfield unit; ICC = intraclass correlation coefficient; MAE = mean absolute error; MRAE = mean relative absolute error; pCT = pseudo-CT; PiB = Pittsburgh Compound B; SD = standard deviation; SUVR = standardized uptake value ratio; wCV = within-subject coefficient of variation.

Received month day, year; accepted after revision month day, year.
From Mallinckrodt Institute of Radiology (Chunwei Ying, Shaney Flores, Tammie L. S. Benzinger, Hongyu An), Department of Neurology (Yasheng Chen, Hongyu An), Department of Surgery (Yan Yan), Knight Alzheimer Disease Research Center (Tammie L. S. Benzinger), and Department of Neurosurgery (Tammie L. S. Benzinger), Washington University School of Medicine, St. Louis, Missouri, USA.

Author Hongyu An, D.Sc. has received research funding from the NIH and Siemens Healthineers. Author Tammie Benzinger, MD, PhD, has received investigator-initiated research funding from the NIH, the Alzheimer's Association, the Foundation at Barnes-Jewish Hospital, Siemens Healthineers and Avid Radiopharmaceuticals (a wholly-owned subsidiary of Eli Lilly and Company). She participates as a site investigator in clinical trials sponsored by Eli Lilly and Company, Biogen, Eisai, Jaansen, and Roche. She has served as a paid and unpaid consultant to Eisai, Siemens, Biogen, Janssen, and Bristol-Myers Squibb.

Please address correspondence to Hongyu An, DSc, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S Kingshighway, WPAV CCIR, CB 8131, St. Louis, MO, 63110, USA; E-mail: hongyuan@wustl.edu.

SUMMARY SECTION

PREVIOUS LITERATURE: Several studies have proposed deep learn-based MR attenuation correction methods using MR images acquired with the same scanner software version and head coil. Thus far, no study has evaluated the accuracy and longitudinal consistency across both software and hardware upgrades, which often occur in longitudinal studies over several years in patients with Alzheimer disease.

KEY FINDINGS: DL-Dixon demonstrated high accuracy across data acquired using two scanner software versions and head coils, passing all four qualification criteria proposed by a recent consensus paper. Moreover, the longitudinal consistency of DL-Dixon attenuation correction is similar to that of CT attenuation correction over three years.

KNOWLEDGE ADVANCEMENT: DL-Dixon exhibited excellent accuracy and longitudinal consistency across two software versions and

two coils, demonstrating its efficacy as a robust MR-based PET attenuation correction method for longitudinal Alzheimer disease research and clinical trials using PET/MR.

INTRODUCTION

In-vivo amyloid PET imaging plays a crucial role in Alzheimer disease (AD) diagnosis and treatment[1-5]. A recently FDA-approved Lecanemab amyloid reduction therapy uses PET or cerebrospinal fluid tests to determine patients' eligibility for treatment[5]. Integrated PET/MR imaging is beneficial as it allows the simultaneous acquisition of PET biomarkers and structural and functional MRI data in a single examination. Quantitative PET imaging requires accurate attenuation correction (AC), one of the most critical factors. MR-based methods have been explored to synthesize pseudo-CT (pCT) for PET AC using a variety of deep learning architectures, including convolutional encoder-decoder[6-8], UNet[9-12], generative adversarial networks[13], CycleGAN[14], and Bayesian deep learning[15]. Recently, a three-dimensional patch-based residual UNet method demonstrated highly accurate PET AC using ultra-short echo MRI, T1 MPRAGE, or Dixon images[9]. Among these methods, the deep learning-based T1-enhanced selection of linear attenuation coefficients (DL-TESLA) network achieved the highest AC accuracy by including quantitative R1 maps derived from a dual-flip-angle and dual-echo ultrashort echo time MRI sequence. Chen et al. also demonstrated high PET/MR AC accuracy from a network using vendor-provided Dixon images as inputs (DL-Dixon). The ultrashort echo sequence, with an acquisition time of 3 minutes and 50 seconds, is a custom sequence not widely available. In contrast, the Dixon sequence, a standard PET/MR AC sequence employed by the vendor, has an acquisition time of only 19 seconds. An accurate DL-Dixon method offers a practical solution for many existing PET/MR brain images using only a standard Dixon scan, making it a promising candidate for adoption in PET/MR clinical applications.

Longitudinal amyloid PET scans using a variety of PET tracers, including ^{18}F-labeled Florbetapir, Florbetaben, and Flutemetamol, and ^{11}C-labeled Pittsburgh Compound B (PiB), are usually acquired over several years to monitor disease progression or treatment response. The annual mean changes of PET-measured amyloid deposition for patients with AD have been reported to be 1%-4%[4, 5, 16, 17]. Knowledge of the test-retest repeatability of amyloid PET is crucial to distinguishing methodology variability from true pathophysiological longitudinal changes[18-21]. Recently, a Radiological Society of North America Quantitative Imaging Biomarkers Alliance profile was proposed to improve the test-retest repeatability of amyloid PET imaging by standardizing the imaging acquisition approach[22].

MR scanner upgrades, including software and hardware upgrades, are often introduced by vendors. For example, the scanner software version of the Siemens Biograph mMR scanner at our institution was upgraded, and a new 32-channel head coil was introduced. Images acquired using different software versions or head coils have different spatial distributions of signal and noise, leading to increased PET/MR AC variabilities in longitudinal studies. Thus far, the longitudinal consistency of PET/MR AC has not yet been evaluated across both software and hardware upgrades.

This study aims to evaluate the accuracy and longitudinal consistency of DL-Dixon AC in amyloid PET with software and hardware upgrades between visits over approximately three years. We also performed power analyses to estimate the sample size required to detect longitudinal PET standardized uptake value ratio (SUVR) changes.

MATERIALS AND METHODS

The methodology proposed in the TRIPD checklist was followed in this study.

Participants and image acquisition

Tri-modality PET/MR and CT images were acquired from 329 participants at Washington University School of Medicine with institutional review board approval and participants' written consent.

PET and MR images were acquired simultaneously using a Siemens Biograph mMR PET/MR scanner (Siemens Healthineers, Erlangen, Germany) between July 2014 and September 2022. Over this period, the PET/MR scanner had a software upgrade from Syngo VB20P to Syngo VE11P and a coil upgrade from a 16-channel head/neck coil to a 32-channel head coil. PET listmode data were acquired with ^{18}F-Florbetapir (Amyvid [Avid], Eli Lilly, Indianapolis, IN) or ^{11}C-PiB tracer. T1 MPRAGE images were acquired (TE/TR = 2.95/2300 ms, TI = 900 ms, flip angle = 9°, matrix size = 240 × 256 × 176, voxel size = 1.05 × 1.05 × 1.2 mm^3). In- and opposed-phase Dixon MR images were acquired using the vendor-provided standard Dixon AC scan (TE1/TE2/TR = 1.23/2.46/3.6 ms, flip angle = 10°, matrix size = 192 × 126 × 128, voxel size = 2.6 × 2.6 × 3.1 mm^3, acquisition time = 19 seconds). Low-dose CT images were acquired using a Siemens Biograph 40 or Biograph Vision PET/CT scanner (Siemens Healthineers, Erlangen, Germany) at 120 kVp, with voxel size = 0.59 × 0.59 × 3.0 mm^3 or 0.59 × 0.59 × 2.0 mm^3.

Image processing

T1 MPRAGE images were segmented and parcellated using FreeSurfer 5.3 for regional analysis. Dixon head masks were determined using the in-phase Dixon images with an empirically determined threshold to remove the background. Bias filed correction was performed using the FMRIB's Automated Segmentation Tool[23] in the FSL toolbox (FMRIB, Oxford, UK). The level-set segmentation tool[24] in the Computational Morphometry Toolkit was used to segment the head region from the background in CT images. CT images were aligned to the Dixon images using a rigid registration with the FSL's FLIRT[25].

Deep learning network and models for pCT estimation

A three-dimensional residual UNet with Dixon in- and opposed-phase images as inputs was developed to estimate pCT. The network structure and training strategy, including hyperparameter initialization, objective function, optimizer, learning rate, patch size, and patch combination approach, were described previously[9]. Means and standard deviations (SD) of CT HU were calculated from all participants. Means and SD of Dixon images were obtained from Dixon in- and opposed-phase image pairs to preserve the relative contrast for each

participant. The normalized image was calculated as (image – mean) / (2 × SD) and then used in the deep learning network training.

As summarized in Table 1, three DL-Dixon models were trained using Dixon images acquired using different software versions and head coils. PET/MR data were acquired with VB20P using a 16-channel head/neck coil from 176 participants (median [interquartile range] age: 70 [65, 75], 101 females) -6 [-30.2, 0.2] days from CT (negative numbers indicate that the PET/MR scan was performed earlier than the CT scan). A VB20P-16Ch model was trained with 69 participants for training and 18 participants for validation. The model was applied to the remaining 89 participants for testing. PET/MR data were acquired with VE11P using a 16-channel head/neck coil from 105 participants (median [interquartile range] age: 71 [65, 76], 58 females) 1 [-6, 20] days from CT. A VE11P-16Ch model was obtained using transfer learning from the VB20P-16Ch model, with 42 participants for training and 11 participants for validation. The model was applied to the remaining 52 participants for testing. PET/MR data were acquired with VE11P using a 32-channel head coil from 48 participants (median [interquartile range] age: 72.5 [68, 78], 27 females) 4 [-7.8, 38.2] days from CT. A VE11P-32Ch model was obtained using transfer learning from the VE11P-16Ch model, with 19 participants for training and 5 participants for validation. The model was applied to the remaining 24 participants for testing.

Table 1. Three DL-Dixon models were trained based on the software and coil used. The total number of participants for each model is summarized.

N = 329	16-channel coil	32-channel coil
VE20P	VB20P-16Ch (N = 176)	-
VE11P	VE11P-16Ch (N = 105)	VE11P-32Ch (N = 48)

A subset of participants (N = 38; median [interquartile range] age: 71 [68, 75] years, 22 females) underwent triple-modality images at two time points (PET1/MR1/CT1 and PET1/MR2/CT2). These data were used to evaluate the longitudinal consistency of DL-Dixon as network testing data. The median [interquartile range] time between the same participant's first and second PET1/MR (PET/MR1 vs. PET2/MR2) and first and second CT scans (CT1 vs. CT2) were 39 [36, 47] and 39 [36, 47] months, respectively. The details of the software version and coil used in the data acquisition at PET/MR1 and PET/MR2 are summarized in Table 2. The VB20P-16Ch, VE11P-16Ch, or VE11P-32Ch model was applied to Dixon images acquired with the corresponding software version and head coil.

Table 2. The longitudinal consistency of DL-Dixon was evaluated in 38 participants with repeated scans over approximately three years. The software version and head coil used during the two visits are summarized.

N = 38		**Visit 2**		
		VB20P	VE11P	VE11P
		16-channel coil	16-channel coil	32-channel coil
Visit 1	VB20P	10	15	4
	16-channel coil			
	VE11P	-	-	9
	16-channel coil			

μ-map generation and PET reconstruction

A piecewise linear conversion was used to convert CT and DL-Dixon pCT images to μ-maps for AC[26]. Using the vendor-provided e7tools software (Siemens Medical Solutions, Knoxville, TN), PET listmode data acquired from 50 to 70 minutes or 30 to 60 minutes after tracer injection for ^{18}F-Florbetapir and ^{11}C-PiB PET, respectively, were reconstructed with Poisson ordered subset expectations maximization algorithm (3 iterations, 21 subsets) with a 5 mm post-reconstruction Gaussian filter.

To evaluate the longitudinal consistency of CT AC and MR AC, the CT μ-maps and DL-Dixon μ-maps at two time points of each participant were first aligned using FSL's FLIRT[25]. The same PET listmode data (^{18}F-Florbetapir: N = 25, ^{11}C-PiB: N = 13) was then reconstructed with the CT μ-maps and DL-Dixon μ-maps from two scan visits.

Accuracy analysis

The pCT images were visually inspected for artifacts. The accuracy of pCT was evaluated using the acquired CT images as the gold standard. The whole head pCT mean absolute error (MAE) was calculated as,

$$CT\ MAE = \frac{\sum_{i=1}^{N}|pCT_i - CT_i|}{N} \quad (1)$$

The accuracy of DL-Dixon PET AC was evaluated using CT PET AC as the gold standard. The PET images of individual participants were first aligned to their T1 MPRAGE images using FSL's FLIRT and then aligned to the International Consortium for Brain Mapping Atlas using Advanced Normalization Tools[27, 28].

The voxel-wise PET relative error was calculated as,

$$PET\ relative\ error\ (\%) = \frac{PET_{pCT} - PET_{CT}}{PET_{CT}} \times 100\%, \quad (2)$$

and the voxel-wise PET relative absolute error was calculated as,

$$PET\ relative\ absolute\ error\ (\%) = \frac{\left|PET_{pCT} - PET_{CT}\right|}{PET_{CT}} \times 100\%. \quad (3)$$

The regional PET mean relative absolute error (MRAE) was calculated in six FreeSurfer-defined ROIs used by the Alzheimer's Disease Neuroimaging Initiative (ADNI) pipeline[29] and the medial temporal lobe[30]. Among these ROIs, the cortical summary region is often used to examine global amyloid deposition, while the cerebellum is a reference region[29, 31].

The accuracy of DL-Dixon PET/MR AC was evaluated following the four qualification criteria recommended by a consensus paper[32]. These criteria include 1) the MRI-based AC maps and corresponding PET should be free of artifacts and without misregistration; 2) PET relative absolute error should be less than 10% in over 90% brain voxels; 3) PET MRAE should be below 10% in all study-specific ROIs and 4) PET MRAE should be below 5% in the reference ROI if reference tissue analysis is involved.

The accuracy of DL-Dixon models was compared using the two-sample t-test with the Benjamini–Hochberg procedure to control for false discovery rate in multiple comparisons using R 4.3.2 (Foundation for Statistical Computing, Vienna, Austria).

Longitudinal consistency analysis

PET SUVR in the cortical summary region was calculated using the cerebellum as the reference region[29, 31]. The longitudinal consistency of the CT-based and DL-Dixon methods was assessed using the Bland and Altman method[33] and the intra-class correlation coefficient (ICC; single rater, absolute-agreement, 2-way mixed-effects model[34]) using MATLAB 2021a (The MathWorks, Natick, MA) and R 4.3.2. The mean and SD of the PET SUVR relative differences between the two scans were calculated for CT or DL-Dixon AC. In addition, the within-subject coefficient of variation (wCV) was obtained. Furthermore, a power calculation by accounting for longitudinal consistency was performed to estimate the number of participants needed to detect certain levels of PET SUVR changes with 80% power.

Data Availability

Investigators can access the data by following the steps outlined on the Knight ADRC website at our institution (https://knightadrc.wustl.edu/professionals-clinicians/request-center-resources/). Data access will be available upon the request's approval by the Knight ADRC. The authors will share the code used in this study upon the publication of this manuscript.

RESULTS

Figure 1 shows the Dixon in-phase MR images, DL-Dixon pCT images, acquired CT images, and the difference map between pCT and CT images from three representative participants. MR images were acquired using different software versions and head coils. All three models generated pCT maps similar to the acquired CT maps without artifacts. The whole head pCT MAE (mean ± SD) was 64.7 ± 9.2 HU, 61.6 ± 7.2 HU, and 62.3 ± 10.0 HU for VB20P-16Ch, VE11P-16Ch, and VE11P-32Ch models, respectively. There was no significant difference in pCT MAE among different DL-Dixon models ($p > 0.1$).

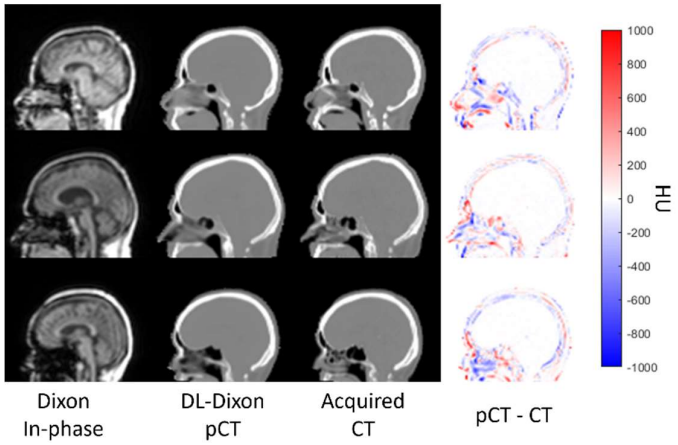

FIG 1. Dixon in-phase MR images (first column), DL-Dixon pCT images (second column), CT images (third column), and HU difference map between pCT and CT (fourth column) from 3 representative participants. The PET/MR scans were acquired using the VB20P software version and a 16-channel head/neck coil (first row), the VE11P software version and a 16-channel head/neck coil (second row), and the VE11P software version and a 32-channel head (third row).

Accuracy of DL-Dixon PET AC

Figure 2 shows the PET images reconstructed using the DL-Dixon AC and CT AC. PET/MR images were acquired using different software versions and head coils. As demonstrated in Figure 3, the mean PET relative error was between -1% and 1% in most brain regions for all three models. Figure 4 demonstrates the cumulative voxelwise relative absolute error of PET reconstructed using DL-Dixon AC. 99.81% ± 0.42%, 99.64% ± 0.67%, and 99.91% ± 0.14% of brain voxels had PET relative absolute error less than 10% for VB20P-16Ch, VE11P-16Ch, and VE11P-32Ch models, respectively. All participants had over 95.6% brain voxels with PET relative absolute error of less than 10%.

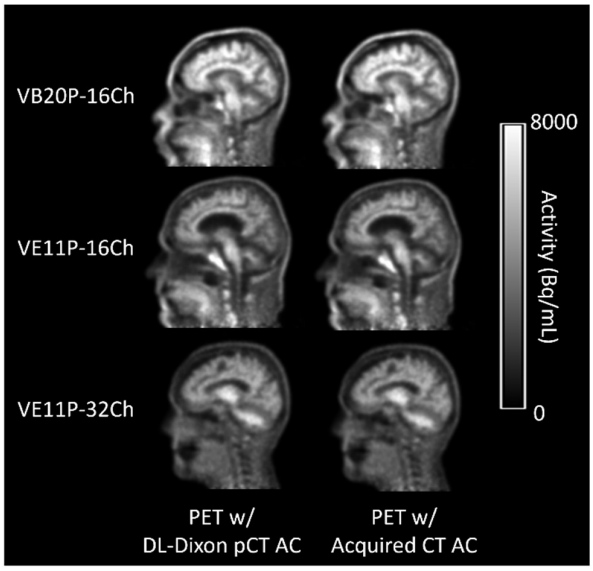

FIG 2. PET reconstructed with the DL-Dixon AC (first column) or the CT AC (second column) from 3 representative participants. The PET/MR scans were acquired using the VB20P software version and a 16-channel head/neck coil (first row), the VE11P software version and a 16-channel head/neck coil (second row), and the VE11P software version and a 32-channel head (third row).

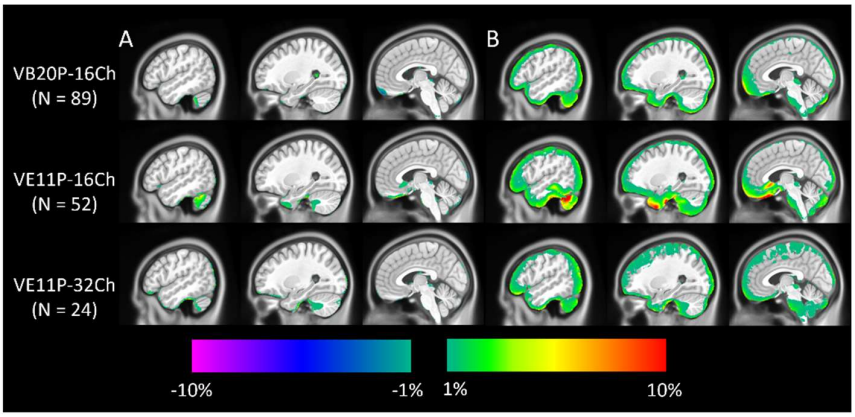

FIG 3. Mean (A) and standard deviation (B) of PET relative error on the voxel basis across testing participants of the VB20P-16Ch model (N = 89), the VE11P-16Ch model (N = 52) and the VE11P-32Ch model (N = 24). The CT AC method is used as the gold standard reference.

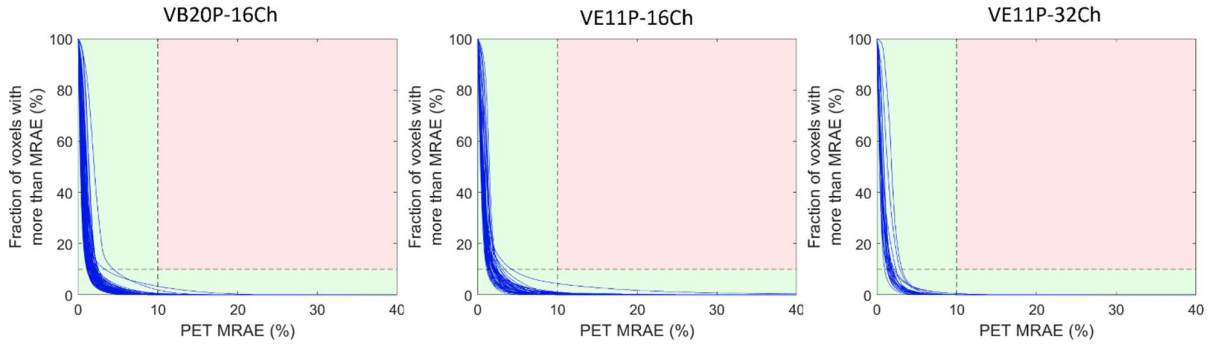

FIG 4. Cumulative voxel-wise PET MRAE using DL-Dixon AC. The acquired CT AC method is used as the gold standard reference. Each blue curve represents one participant. If the line stayed within the green region, the participant passed the qualification criteria.

Figure 5 demonstrates PET MRAE in seven amyloid PET-related ROIs. The median [interquartile range] PET MRAE was 1.10% [0.93%, 1.26%], 1.24% [1.03%, 1.54%], 0.99% [0.86%, 1.13%] in the cortical summary region, and 1.04% [0.83%, 1.36%], 1.08% [0.84%, 1.34%], 1.05% [0.72%, 1.32%] in cerebellum using the VB20P-16Ch, VE11P-16Ch and VE11P-32Ch models, respectively. Except VE11P-16Ch had significantly higher PET MRAE than VE11P-32Ch in the frontal cortex region ($p = 0.02$) and the medial temporal lobe

(p = 0.05), and VB20P-16Ch in the lateral temporal cortical region (p = 0.04) and the medial temporal lobe (p = 0.03), three models had comparable PET MRAE in the remaining ROIs.

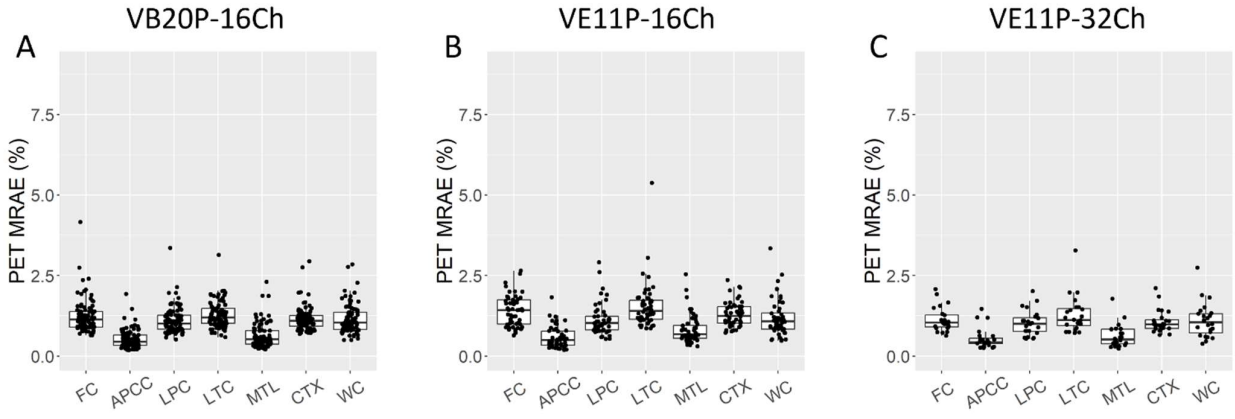

FIG 5. PET MRAE in seven cerebral cortical and cerebellum ROIs. The boxplots show the 25th, 50th (median), and 75th percentiles. FC: frontal cortical region, APCC: anterior and posterior cingulate cortical region, LPC: lateral parietal cortical region, LTC: lateral temporal cortical region, MTL: medial temporal lobe, CTX: cortical summary region, WC: whole cerebellum region.

Longitudinal consistency of DL-Dixon

The Bland-Altman plots and ICC plots in Figure 6 show the longitudinal consistency of regional PET SUVR using CT AC and DL-Dixon AC in the cortical summary region. Table 3 summarizes the mean SUVR relative differences, wCV, and ICC. PET SUVR using CT AC and DL-Dixon AC had similar wCV and ICC. Moreover, ^{18}F-Florbetapir PET (Blue symbols) and ^{11}C-PiB PET (Red symbols) had comparable longitudinal consistency with either CT AC or DL-Dixon AC.

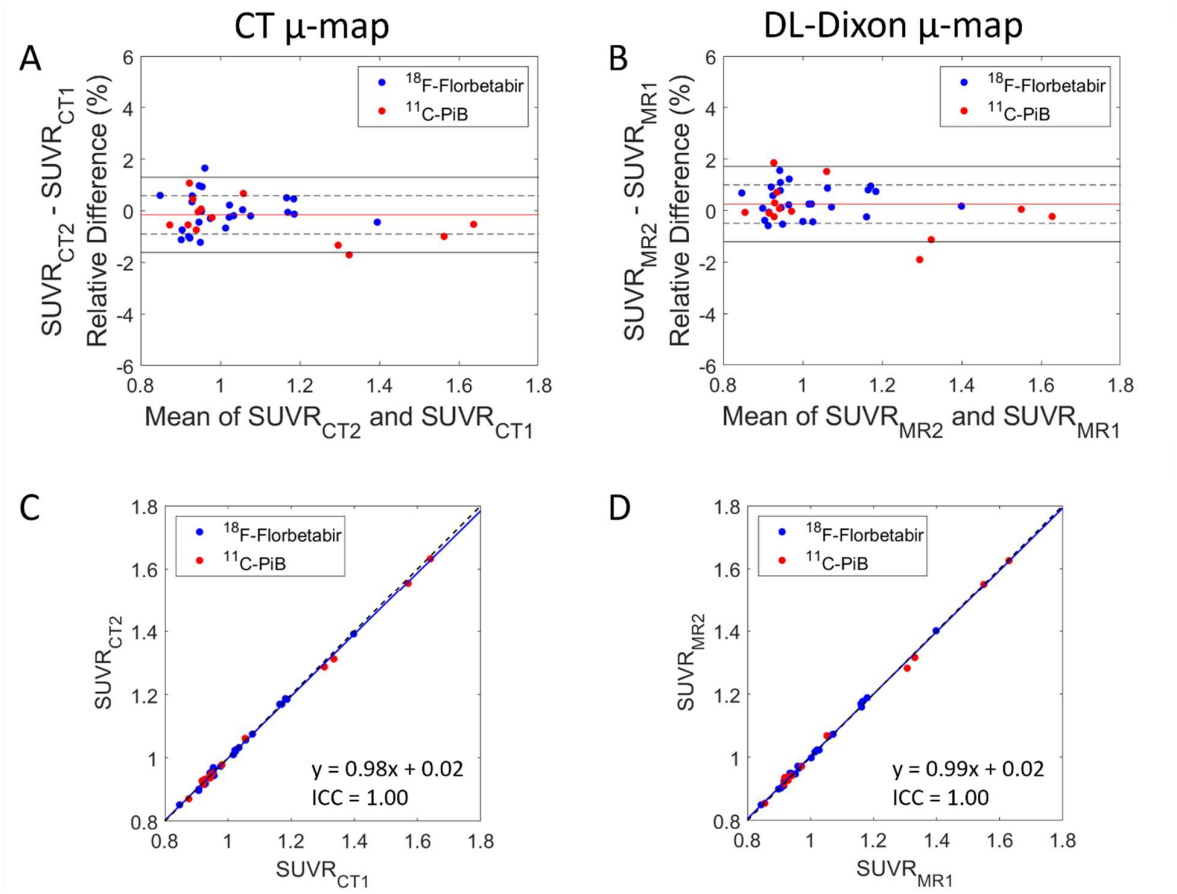

FIG 6. Longitudinal consistency of PET using CT and DL-Dixon AC. The Bland-Altman plots of the PET SUVR difference between two CT (A) and two DL-Dixon (B) ACs in the cortical summary region are shown. The red horizontal line, dotted black horizontal lines, and solid black horizontal lines represent the mean, ± SD, and ± 1.96SD of the PET SUVR differences, respectively. Scatter plots

of the PET SUVR between two CT (C) and two DL-Dixon (D) ACs in the cortical summary region are shown. The solid blue line and dotted black line represent the linear fitting line and line of identity, respectively. Symbol colors indicate different tracers (Blue symbols: ^{18}F-Florbetapir PET, Red symbols: ^{11}C-PiB PET).

Table 3. PET SUVR longitudinal consistency of CT AC and DL-Dixon AC in the cortical summary region. PET SUVR difference, wCV, and ICC are included.

	CT	DL-Dixon
SUVR Difference (Mean ± SD)	-0.16% ± 0.74%	0.25% ± 0.75%
wCV	0.53%	0.55%
ICC	1.00	1.00

The number of participants required to detect real longitudinal PET SUVR changes in the cortical summary region with 80% power is shown in Figure 7. Assuming the correlation between the paired measurements from a participant of 0.3, 0.5, 0.7, and 0.9, the required numbers of participants needed to detect a 3% change in SUVR in the cortical summary region are 388, 278, 168, and 58 using CT AC and 392, 280, 169 and 58 using DL-Dixon AC.

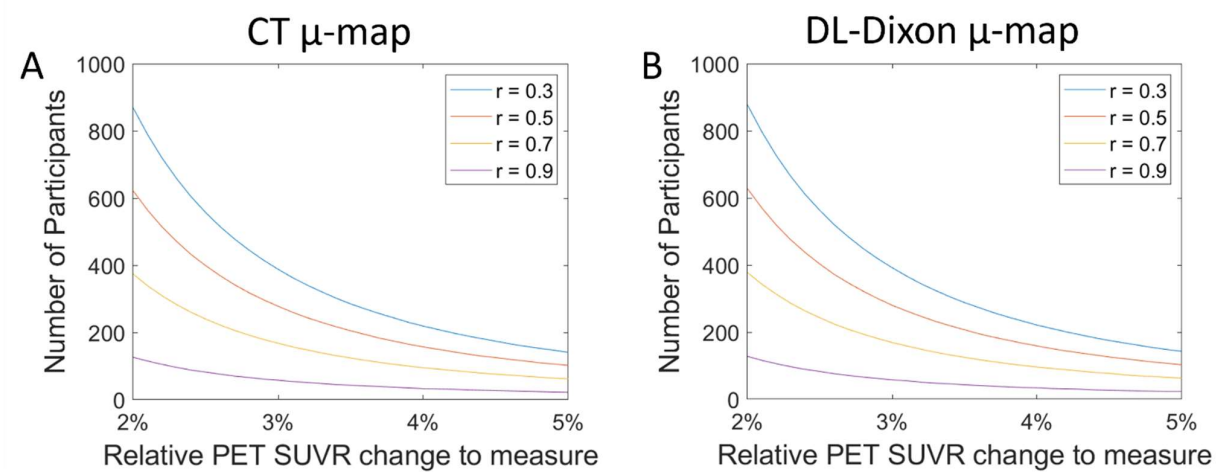

FIG 7. Number of participants required to detect longitudinal change in the cortical summary region with 80% power. The "r" is the assumed correlation between the paired measures from a participant.

DISCUSSION

Deep learning-based image synthesis has been widely implemented for transforming imaging between MR and CT for PET/MR AC and radiation therapy planning[9, 13, 14, 35]. However, there is no consensus on the extent to which deep neural network-synthesized pseudo-images should be accepted. Addressing this question requires rigorous evaluation. In this study, we used the acquired CT images as the ground truth for such evaluation. We demonstrated that MR-synthesized pCT closely resembles the acquired CT, and the proposed method meets the qualification criteria outlined in a consensus paper[32]. Furthermore, excellent longitudinal consistency of MR-based PET AC over several years was achieved across software and hardware upgrades, which is crucial for the use of PET/MR in AD longitudinal trials. To the best of our knowledge, our study is the first to evaluate the accuracy and longitudinal consistency of a PET/MR AC approach across both scanner software and head coil updates.

Several existing deep learning methods achieved PET MRAEs of 1% to 3% in cortical and cerebellum ROIs using the same software version and head coil [9, 11, 12, 14]. It is unclear whether these methods may be generalized to MR images acquired after MR scanner software and hardware upgrades. One study used the MR images acquired using the same head coil but two software versions[10]. This method showed PET MRAE of 1.5% to 2% in the cortical ROIs and over 2% in the cerebellum ROI. The proposed DL-Dixon models have a median PET MRAE from 0.99% to 1.24% and 1.04% to 1.08% in the cortical summary region and the cerebellum, respectively, across both software and hardware upgrades (Figure 5). Moreover, all three DL-Dixon models passed the recommended qualification criteria for all participants[32].

Test-retest repeatability is crucial for including a quantitative biomarker in longitudinal research and clinical trials. The test-retest repeatability of amyloid PET using PET/CT or standalone PET scanners has been investigated in previous studies. Joshi et al. found an ^{18}F-Florbetapir cortical SUVR wCV of 1.94% and 1.20% for patients with AD and healthy controls over four weeks[18]. Vandenberghe et al. found an ^{18}F-Flutemetamol SUVR wCV of 1.15% in the composite cortical ROI for patients with AD over 7-13 days[19]. The long-term

cortical SUVR wCV was reported to be 1.25%-3.38% for cognitively normal subjects over two years using the ^{18}F-Florbetapir tracer [20-22]. Based on these studies, the Quantitative Imaging Biomarkers Alliance profile suggests that ^{18}F-labeled amyloid PET SUVR should have a wCV of less than 1.94%[22].

Several studies evaluated the repeatability or longitudinal consistency of MR-based PET AC methods. One study found a whole brain SUVR difference of 0.65% ± 0.95% over ten days[36], while another study found an SUVR difference of -0.65% ± 1.62% and wCV of 1.15% in the mean cortical region over three years [9]. In this study, longitudinal SUVR difference and wCV of DL-Dixon AC (0.25% ± 0.75% and 0.55%) are similar to those of CT AC (-0.16% ± 0.74% and 0.53%) despite software and hardware updates over three years (Figure 6, Table 3). To detect a specific level of longitudinal change in amyloid PET SUVR, a similar number of participants would be needed using DL-Dixon compared to CT AC (Figure 7). The longitudinal consistency of DL-Dixon meets the Quantitative Imaging Biomarkers Alliance recommendation.

Recently, the Centiloid approach was proposed to normalize the amyloid burden measured using various tracers on different scanners to a standard scale[37]. The annualized absolute Centiloid change was reported to be 2.2 to 3 Centiloid in dominantly inherited AD mutation-positive participants[38] and 2.43 Centiloid in patients with mild dementia or mild cognitive impairment due to AD[5]. Using the SUVR-to-Centiloid transformations for the ADNI FreeSurfer 5.3 pipeline[31, 37], the longitudinal Centiloid difference was -0.39 ± 1.58 for CT AC and 0.45 ± 1.59 for MR AC. It is worth noting that the studies mentioned above used separately acquired PET data at two time points, while this study used the same PET data but separately acquired CT or MR to derive μ-maps. The longitudinal differences might be higher if two separately acquired PET data were used. This study used the MR and CT images acquired twice over three years. Possible structural changes over this period may partially affect the longitudinal consistency.

Numerous deep learning-based methods have been developed to synthesize pseudo-CT images using MRI[6-11, 13-15]. CT images measure tissue electron density, while MR signal depends on magnetic properties, such as proton density and tissue relaxation rates. There is no direct relationship between the signal intensity of Dixon MR and CT HU, which results in challenges in intensity-based methods. Despite differences in imaging physics, the paired MR and CT images are obtained from the same patients. Therefore, MR and CT images share the same anatomical structures. In this study, a three-dimensional residual UNet was trained to learn complex nonlinear relationships between MR and the corresponding CT image by minimizing differences between the predicted pCT and CT. The transformation of MR to CT involves anatomical, geometrical, image contrast, and texture information derived from millions of paired MR and CT patches.

This study has several limitations. First, all data in this study were obtained from elderly participants using amyloid tracers at a single research site. The accuracy and longitudinal consistency of DL-Dixon should be further evaluated using different scanners in multicenter studies with patients of a broader age range. Second, participants in this study do not have bone abnormalities. The performance of DL-Dixon in such cases is unclear.

CONCLUSIONS

In conclusion, DL-Dixon exhibited excellent accuracy and longitudinal consistency across two software versions and two coils, demonstrating its efficacy as a robust MR-based AC method for longitudinal research and clinical trials using PET/MR.

ACKNOWLEDGMENTS

We acknowledge the following funding sources: NIH grants P30AG066444, P01AG003991, P01AG026276, RF1NS116565, R01EB032713, and R21NS127425. Image acquisition received additional support from UL1TR00044, P30NS098577, and Siemens Healthineers. Avid Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly) provided the doses of 18F-Florbetapir and partially funded the cost of the PET scans.

REFERENCES

1. Jagust WJ, Bandy D, Chen K, et al. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimers Dement* 2010;6:221-229

2. McKay NS, Gordon BA, Hornbeck RC, et al. Neuroimaging within the dominantly inherited Alzheimer's network (DIAN): PET and MRI. *BioRxiv* 2022:2022.2003. 2025.485799

3. Sims JR, Zimmer JA, Evans CD, et al. Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA* 2023;330:512-527

4. Budd Haeberlein S, Aisen PS, Barkhof F, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *J Prev Alzheimers Dis* 2022;9:197-210

5. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in Early Alzheimer's Disease. *N Engl J Med* 2023;388:9-21

6. Jang H, Liu F, Zhao G, et al. Technical Note: Deep learning based MRAC using rapid ultrashort echo time imaging. *Med Phys* 2018;45:3697-3704

7. Shiri I, Ghafarian P, Geramifar P, et al. Direct attenuation correction of brain PET images using only emission data via a deep convolutional encoder-decoder (Deep-DAC). *Eur Radiol* 2019;29:6867-6879

8. Liu F, Jang H, Kijowski R, et al. Deep Learning MR Imaging-based Attenuation Correction for PET/MR Imaging. *Radiology* 2018;286:676-684

9. Chen Y, Ying C, Binkley MM, et al. Deep learning-based T1-enhanced selection of linear attenuation coefficients (DL-TESLA) for PET/MR attenuation correction in dementia neuroimaging. *Magn Reson Med* 2021;86:499-513

10. Ladefoged CN, Hansen AE, Henriksen OM, et al. AI-driven attenuation correction for brain PET/MRI: Clinical evaluation of a dementia cohort and importance of the training group size. *Neuroimage* 2020;222:117221

11. Gong K, Han PK, Johnson KA, et al. Attenuation correction using deep Learning and integrated UTE/multi-echo

Dixon sequence: evaluation in amyloid and tau PET imaging. *Eur J Nucl Med Mol Imaging* 2021;48:1351-1361

12. Gong K, Yang J, Kim K, et al. Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images. *Phys Med Biol* 2018;63:125011

13. Arabi H, Zeng G, Zheng G, et al. Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI. *Eur J Nucl Med Mol Imaging* 2019;46:2746-2759

14. Gong K, Yang J, Larson PEZ, et al. MR-based Attenuation Correction for Brain PET Using 3D Cycle-Consistent Adversarial Network. *IEEE Trans Radiat Plasma Med Sci* 2021;5:185-192

15. Leynes AP, Ahn S, Wangerin KA, et al. Attenuation Coefficient Estimation for PET/MRI With Bayesian Deep Learning Pseudo-CT and Maximum-Likelihood Estimation of Activity and Attenuation. *Ieee T Radiat Plasma* 2022;6:678-689

16. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal beta-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *J Nucl Med* 2015;56:567-574

17. Jack CR, Jr., Wiste HJ, Lesnick TG, et al. Brain beta-amyloid load approaches a plateau. *Neurology* 2013;80:890-896

18. Joshi AD, Pontecorvo MJ, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with alzheimer's disease and cognitively normal subjects. *J Nucl Med* 2012;53:378-384

19. Vandenberghe R, Van Laere K, Ivanoiu A, et al. 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann Neurol* 2010;68:319-329

20. Brendel M, Hogenauer M, Delker A, et al. Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. *Neuroimage* 2015;108:450-459

21. Chen K, Roontiva A, Thiyyagura P, et al. Improved power for characterizing longitudinal amyloid-beta PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J Nucl Med* 2015;56:560-566

22. Smith AM, Obuchowski NA, Foster NL, et al. The RSNA QIBA Profile for Amyloid PET as an Imaging Biomarker for Cerebral Amyloid Quantification. *J Nucl Med* 2023;64:294-303

23. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *Ieee T Med Imaging* 2001;20:45-57

24. Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 1979;9:62-66

25. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Medical image analysis* 2001;5:143-156

26. Carney JP, Townsend DW, Rappoport V, et al. Method for transforming CT images for attenuation correction in PET/CT imaging. *Med Phys* 2006;33:976-983

27. Avants BB, Tustison NJ, Song G, et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 2011;54:2033-2044

28. Fonov VS, Evans AC, McKinstry RC, et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* 2009;47:Page S102

29. Landau S, Jagust W. Florbetapir processing methods. *Los Angeles: Alzheimer's Disease Neuroimaging Institute* 2015

30. Albrecht DS, Sagare A, Pachicano M, et al. Early neuroinflammation is associated with lower amyloid and tau levels in cognitively normal older adults. *Brain, behavior, and immunity* 2021;94:299-307

31. Kolibash S, Minhas D, Lopresti B. Centiloid Level-2 Analysis of [18F] Florbetaben (FBB) and [18F] Florbetapir (FBP) PET Image Data using the ADNI Pipeline. pp. *Centiloid Level-2 Analysis of [18F] Florbetaben (FBB) and [18F] Florbetapir (FBP) PET Image Data using the ADNI Pipeline*

32. Catana C, Laforest R, An H, et al. A Path to Qualification of PET/MRI Scanners for Multicenter Brain Imaging Studies: Evaluation of MRI-Based Attenuation Correction Methods Using a Patient Phantom. *J Nucl Med* 2022;63:615-621

33. Martin Bland J, Altman D. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet* 1986;327:307-310

34. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-163

35. Liu Y, Chen A, Shi H, et al. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Computerized medical imaging and graphics* 2021;91:101953

36. Chen KT, Izquierdo-Garcia D, Poynton CB, et al. On the accuracy and reproducibility of a novel probabilistic atlas-based generation for calculation of head attenuation maps on integrated PET/MR scanners. *Eur J Nucl Med Mol Imaging* 2017;44:398-407

37. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement* 2015;11:1-15 e11-14

38. Chen CD, McCullough A, Gordon B, et al. Longitudinal head-to-head comparison of (11)C-PiB and (18)F-florbetapir PET in a Phase 2/3 clinical trial of anti-amyloid-beta monoclonal antibodies in dominantly inherited Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 2023;50:2669-2682

SUPPLEMENTAL FILES