Received: 5 October 2023 | Revised: 15 January 2024 | Accepted: 19 January 2024

DOI: 10.1002/dad2.12561

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Inter-scanner Aβ-PET harmonization using barrel phantom spatial resolution matching

Gihan P. Ruwanpathirana[1,2] | **Robert C. Williams**[2] | **Colin L. Masters**[3,4] | **Christopher C. Rowe**[3,4,5] | **Leigh A. Johnston**[1,2] | **Catherine E. Davey**[1,2]

[1]Department of Biomedical Engineering, The University of Melbourne, Melbourne, VIC, Australia

[2]Melbourne Brain Centre Imaging Unit, The University of Melbourne, Melbourne, VIC, Australia

[3]Florey Institute of Neurosciences and Mental Health, The University of Melbourne, Melbourne, VIC, Australia

[4]The Australian Dementia Network (ADNET), Melbourne, Australia

[5]Department of Molecular Imaging & Therapy, Austin Health, Melbourne, VIC, Australia

Correspondence
Catherine Davey, and Gihan Ruwanpathirana, Department of Biomedical Engineering, The University of Melbourne, Melbourne, VIC, Australia.
Email: cedavey@unimelb.edu.au and gruwanpathir@student.unimelb.edu.au

Funding information
National Health and Medical Research Council (NHMRC), Grant/Award Numbers: 1152623, 1140853, 1132604

Abstract

INTRODUCTION: The standardized uptake value ratio (SUVR) is used to measure amyloid beta–positron emission tomography (Aβ-PET) uptake in the brainDifferences in PET scanner technologies and image reconstruction techniques can lead to variability in PET images across scanners. This poses a challenge for Aβ-PET studies conducted in multiple centers. The aim of harmonization is to achieve consistent Aβ-PET measurements across different scanners. In this study, we propose an Aβ-PET harmonization method of matching spatial resolution, as measured via a barrel phantom, across PET scanners. Our approach was validated using paired subject data, for which patients were imaged on multiple scanners.

METHODS: In this study, three different PET scanners were evaluated: the Siemens Biograph Vision 600, Siemens Biograph molecular computed tomography (mCT), and Philips Gemini TF64. A total of five, eight, and five subjects were each scanned twice with [^{18}F]-NAV4694 across Vision-mCT, mCT-Philips, and Vision-Philips scanner pairs. The Vision and mCT scans were reconstructed using various iterations, subsets, and post-reconstruction Gaussian smoothing, whereas only one reconstruction configuration was used for the Philips scans. The full-width at half-maximum (FWHM) of each reconstruction configuration was calculated using [^{18}F]-filled barrel phantom scans with the Society of Nuclear Medicine and Molecular Imaging (SNMMI) phantom analysis toolkit. Regional SUVRs were calculated from 72 brain regions using the automated anatomical labelling atlas 3 (AAL3) atlas for each subject and reconstruction configuration. Statistical similarity between SUVRs was assessed using paired (within subject) *t*-tests for each pair of reconstructions across scanners; the higher the *p*-value, the greater the similarity between the SUVRs.

RESULTS: Vision-mCT harmonization: Vision reconstruction with FWHM = 4.10 mm and mCT reconstruction with FWHM = 4.30 mm gave the maximal statistical similarity (maximum *p*-value) between regional SUVRs. **Philips-mCT harmonization**: The FWHM

of the Philips reconstruction was 8.2 mm and the mCT reconstruction with the FWHM of 9.35 mm, which gave the maximal statistical similarity between regional SUVRs. **Philips–Vision harmonization**: The Vision reconstruction with an FWHM of 9.1 mm gave the maximal statistical similarity between regional SUVRs when compared with the Philips reconstruction of 8.2 mm and were selected as the harmonized for each scanner pair.

CONCLUSION: Based on data obtained from three sets of participants, each scanned on a pair of PET scanners, it has been verified that using reconstruction configurations that produce matched-barrel, phantom spatial resolutions results in maximally harmonized Aβ-PET quantitation between scanner pairs. This finding is encouraging for the use of PET scanners in multi-center trials or updates during longitudinal studies.

KEYWORDS
amyloid beta, PET scanner harmonization, Philips TF64, SIEMENS mCT, SIEMENS Vision

Highlights

- **Question**: Does the process of matching the barrel phantom-derived spatial resolution between scanners harmonize amyloid beta–standardized uptake value ratio (Aβ-SUVR) quantitation?
- **Pertinent findings**: It has been validated that reconstruction pairs with matched barrel phantom-derived spatial resolution maximize the similarity between subjects paired Aβ-PET (positron emission tomography) SUVR values recorded on two scanners.
- **Implications for patient care**: Harmonization between scanners in multi-center trials and PET camera updates in longitudinal studies can be achieved using a simple and efficient phantom measurement procedure, beneficial for the validity of Aβ-PET quantitation measurements.

1 | INTRODUCTION

The imaging of amyloid beta (Aβ) protein plaques in the brain using positron emission tomography (PET) is of vast importance in the study of Alzheimer's disease (AD), with the development of Aβ-specific radiotracers including [^{18}F]-florbetaben, [^{18}F]-florbetapir, [^{18}F]-NAV4694, and [^{18}F]-flutemetamol enabling better monitoring in patients with AD.[1–4] Although the extent of Aβ-PET accumulation in the brain is commonly measured using a standardized uptake value ratio (SUVR),[2] it is known to be dependent on the choice of Aβ-PET radiotracer and scanner hardware.[5,6] The spatial resolution of PET data is governed by physical factors including positron range of the radiotracer, photon scattering, and hardware-specific limitations, as well as the choice of reconstruction algorithm and associated parameter settings.[7,8] It has been well established that PET quantitation is impacted by the partial volume effect (PVE), in which voxel intensity is determined by both the dominant tissue within the voxel, and any surrounding tissue that falls within the voxel boundaries. Therefore, choice of reconstruction parameters governing spatial resolution—and hence determining the PVE extent—influences A β quantitation.[6,9,10] Furthermore, we have established that spatial resolution differentially impacts the Aβ– and Aβ+ cohorts.[6,11] Together, these scanner-dependent effects, and the rapid advancement of PET technologies that is leading to significant performance variation across scanners, represent a problem for multi-site Aβ-PET studies; the same patient imaged on different PET scanners can exhibit significant image and Aβ-PET metric variability, even when attempts are made to match reconstruction parameters across scanning sites.

Problems associated with inter-scanner variability have long been recognized by the oncology community. The European Association of Nuclear Medicine (EANM) launched the EANM Research Ltd. (EARL) initiative to standardize images and promote multicenter studies.[12,13] The EARL initiative employed the recovery coefficient, defined as the ratio of observed to true activity in PET, to compare PET reconstructions. They defined upper and lower limits for recovery coefficients using the National Electrical Manufacturers Association's (NEMA) NU

2 body phantom imaging, and participating scanning sites choose a PET reconstruction configuration such that the recovery coefficients of their phantom images fall within the limits. EARL phantom criteria are not necessarily suitable for brain PET, as the NEMA NU 2 phantom contains multiple spheres with high radioactivity values within uniform background radioactivity. Although the NEMA NU 2 setup is suitable to simulate tumor uptake, it does not simulate brain uptake, which is characteristically different, with a more dispersed uptake across larger gray matter and white matter compartments.[14]

In the neuroimaging domain, Alzheimer's Disease Neuroimaging Initiative (ADNI) developed one of the first frameworks to reduce inter-scanner differences for [^{18}F]-fluorodeoxyglucose (FDG) imaging by post-reconstructing smoothing data to a pre-defined spatial resolution, as determined using a Hoffman phantom.[5] There has been no definitive validation of this procedure for Aβ-PET, however. Furthermore, we have demonstrated that post-reconstruction smoothing differentially affects Aβ– and Aβ+ groups.[6] Since the ADNI initiative, several initiatives have recommended acquisition protocols and image quality procedures to standardize both Aβ-PET and FDG imaging for multicenter studies.[15,16] Akamatsu et al. proposed a phantom procedure to optimize image quality of Aβ-PET and FDG imaging across different scanners, where participating sites choose reconstruction settings to meet predefined criteria for image noise, uniformity, contrast, and spatial resolution using phantom scans.[17] A recent study proposed the use of an FDG-filled Hoffman phantom to define the upper and lower limits for recovery coefficients in gray matter and a gray-to-white matter ratio to standardize quantification across scanners, similar to the EARL initiative.[14] The overhead associated with acquiring Hoffman phantom data, in conjunction with the likelihood of experimental errors, led Lodge et al. to propose a simple method for calculating PET spatial resolution using an ^{18}F-filled barrel phantom readily available as part of routine quality assurance.[11] These studies were based on phantom data; the impact of standardizing quantification metrics on humans has not yet been evaluated.

In the current study, we hypothesized that multi-scanner Aβ-PET harmonization can be achieved by matching spatial resolution as determined by the Lodge barrel phantom measurement. We interrogated this hypothesis using a three-camera comparison data set of [^{18}F]-NAV4694 radiotracer, where each participant was scanned on two out of a set of three scanners. Raw PET data and vendor-supplied PET image reconstruction toolboxes enabled comparison of a broad range of reconstruction settings and spatial resolutions. The degree of harmonization was indicated by the similarity of participants Aβ-SUVR measures from each pair of scanners.

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the existing literature using conventional sources, such as PubMed and Google scholar. Alzheimer's disease (AD) is a debilitating disease that is increasing in prevalence, particularly given that many countries have an aging population. New drugs are being released to combat the cognitive decline associated with AD. It is crucial that both the research and clinical communities can diagnose and monitor AD. With the improvement of positron emission tomography (PET), AD is being diagnosed earlier and with increasing accuracy. However, a persistent issue is the difficulty in comparing imaging results obtained from different scanners. We attempt to address this issue by proposing a simple technique that harmonizes imaging quantification for AD across scanners, without compromising spatial resolution.
2. **Interpretation**: Our data validates that the reconstruction configurations with matched barrel phantom derived spatrial resolutions maximise the similarity between subjects' Aβ-PET SUVR values recorded on two scanners.
3. **Future directions**: Proposed harmonization method needs to be validated on other Aβ-PET tracers. Further on, several Aβ-PET processing pipelines could be employed to validate the harmonization method.

2 | METHODS

Three scanners were used in this study: (1) Siemens Biograph Vision 600, (2) Siemens Biograph mCT, and (3) Philips Gemini TF64. Hereinafter, these scanners are referred to as the Vision, mCT, and Philips, respectively.

2.1 | Phantom spatial resolution calculation

A barrel phantom was used to measure the radial and axial PET spatial resolution of a given reconstruction configuration by calculating the full width at half maximum (FWHM) along the two directions, as proposed by Lodge et al.[11]: A barrel phantom filled with ^{18}F was placed at the center of the scanner field-of-view, with the long axis parallel to the axis of the scanner. One end of the phantom was lifted to create a small angle with the scanner axis. Phantom data were reconstructed with a set of reconstructions and uploaded to the Society of Nuclear Medicine and Molecular Imaging's (SNMMI) phantom analysis toolkit (http://www.snmmi.org/PAT) to calculate the FWHM (average of axial and radial FWHMs) of each configuration. Hereinafter, 'spatial resolution' refers to the Lodge barrel phantom-computed spatial resolution.

2.2 | Data collection and image reconstructions

The use of three scanners allowed three pairwise comparisons: (1) Vision-mCT; (2) mCT-Philips, and (3) Vision–Philips. The same subjects were scanned for Aβ twice on the selected pair of scanners. All subjects were injected with [^{18}F]-NAV4694 radiotracer 50 min prior to 20 min of continuous scanning. Ethics approval and consent to participate in

TABLE 1 Demographics of the Vision–mCT, mCT–Philips, and Vision–Philips data sets.

Study type	Vision–mCT comparison	mCT–Philips comparison	Vision–Philips comparison
Sample size	5	8	5
Aβ+ (%)	2 (40)	0(0)	2 (40)
Age ± Standard Deviation (SD)	77.20 ± 3.49	75.38 ± 6.09	75.80 ± 2.59
Sex, F (%)	1 (20)	4 (50)	3 (60)

this study were approved by the Austin Health Human Research Ethics Committee (HREC/18/Austin/201).

Vision–mCT: Five subjects (Table 1) were scanned on both the Vision and mCT scanners with an average inter-scan interval of 45.6 ± 4.27 weeks. [^{18}F]-NAV4694 radiotracer doses of 189.5 ± 4.21MBq and 104.02 ± 3.35MBq were used in the mCT and Vision scans, respectively. Data were reconstructed with ordered subset expectation maximization with time-of-flight information (OPTOF). For the Vision, 12 iteration (i) values were used, {1, 2, 3, 4, 5, 6, 12, 18, 24, 30, 36, 42}, with the number of subsets, $s = 5$. For the mCT, $s = 21$ and $i \in \{2, 4, 6, 8, 10, 12\}$. No post-reconstruction Gaussian smoothing was used. The constant number of subsets reflected scanner software constraints. All scans were reconstructed at a voxel size of 1.65 mm × 1.65 mm × 2 mm.

mCT–Philips: Eight subjects (Table 1) were scanned on both the mCT and Philips scanners with an inter-scan interval of 41.6 ± 8.24 weeks. [^{18}F]-NAV4694 radiotracer doses of 188.79 ± 7.96MBq and 206.63 ± 4.27MBq were injected in the mCT and Philips scans, respectively. Philips data were reconstructed using a single reconstruction configuration, the line-of-response row-action maximum likelihood (LOR RAMLA) with smoothing parameter set to 'SHARP'. This is the default brain imaging protocol of the Philips T64, which cannot be altered. The spatial resolution of the Philips reconstruction configuration was 8.2 mm. The mCT data were reconstructed using OPTOF with $i \in \{2, 4, 6, 8, 10, 12\}$ and $s = 21$. To further reduce the resolution of the mCT reconstructions, post-reconstruction Gaussian smoothing was applied to the lowest resolution mCT reconstruction ($i = 2$, $s = 21$) from 1to13 mm in steps of 1 mm. Gaussian smoothing was applied prior to the processing steps detailed in the Data Analysis section. All scans were reconstructed at a voxel size of 2 mm × 2 mm × 2 mm.

Vision–Philips: Five subjects (Table 1) were scanned across both Vision and Philips scanners with an inter-scan interval of 31.66 ± 6.98 weeks. [^{18}F]-NAV4694 radiotracer doses of 137.08 ± 51.84MBq and 194.00 ± 11.70MBq were used when acquiring the Vision and Philips scans, respectively. Philips data were reconstructed using the only reconstruction configuration available, the LOR RAMLA with smoothing parameter set to 'SHARP.'. This is the default brain imaging protocol of the Philips T64, which cannot be altered. Vision data were reconstructed using OPTOF with $i \in \{1, 2, 3, 4, 5, 6, 12, 18, 24, 30, 36, 42\}$ and $s = 5$. The lowest resolution achieved when only changing the number of iterations, i, was 9.1 mm ($i = 1$, $s = 5$). To further reduce the resolution of the Vision reconstructions, post-reconstruction Gaussian smoothing was applied from 2to8 mm in steps of 2 mm. Compared to 1 mm step used in post-reconstruction smoothing of mCT data, Vision data were smoothed with 2 mm increments; the aim was to encompass a broader spatial resolution range efficiently, given that the vision reconstruction (9.1 mm) had already achieved the Philips resolution (8.2 mm). Gaussian smoothing was applied prior to the processing steps detailed in the Data Analysis section. All scans were reconstructed at a voxel size of 2 mm × 2 mm × 2 mm.

ADNI reconstruction comparison: All data from the three scanners were reconstructed using ADNI-proposed reconstruction configurations: the mCT scanner data were reconstructed using 4*i*, 24*s* without enabling TOF information and post-reconstruction smoothing; the Vision data were reconstructed without post-reconstruction smoothing using 8*i*,5*s* enabling time of flight (TOF); Philips data were reconstructed using the Brain protocol with LOR-RAMLA, setting the smoothing parameter to 'SHARP.' [18]

2.3 | Data analysis

All scans were registered to a computerized tomography (CT)–based normalization template: The skull was stripped to remove off-target binding in non-brain regions. CT scans, acquired for attenuation correction, were used to generate a skull-stripped mask using the FSL Brain Extraction Tool, and the mask was transformed to the PET domain. Results were manually checked for both registration and stripping faults. Each scan was non-linearly registered to the FSL MNI152 1 mm brain template using Advanced Normalization Tools with the CT image used as an intermediate step between the PET image and the template.

Aβ-PET SUVR values (hereinafter termed 'SUVR values') were generated using the whole cerebellum as the reference region, in accordance with the standard Centiloid (CL) method.[19] For each scan, SUVR values were computed across a total of 72 brain regions spread across the temporal lobe, frontal lobe, parietal lobe, cingulate, and occipital lobe (Automated Anatomical Labelling Atlas 3 [AAL3]).[20] For the Vision–mCT data set, a single data set defined by its reconstruction configuration, was comprised of 360 regional SUVR values (5 subjects × 72 brain regions). Each data set from the Vision scanner was compared with every data set from the mCT scanner, giving a total of 72 comparisons between the two scanners (6 mCT × 12 Vision reconstructions). For the mCT–Philips data set, a data set from a single reconstruction configuration comprised a total of 576 regional SUVR values (8 subjects × 72 brain regions). Each mCT data set was compared with the Phillips data set, giving a total of 20 comparisons (1 Philips × 20 mCT reconstructions) between the scanners. For the Vision–Philips data set, each data set comprised 360 SUVR values (5 subjects × 72 brain regions). Each Vision data set was compared with the Phillips data set, giving a total of 16 comparisons (1 Philips × 16 Vision reconstructions) between the scanners. In addition, global SUVR values were computed for each data set by calculating the mean SUVR inside the neocortical mask defined by the CL project.[19]

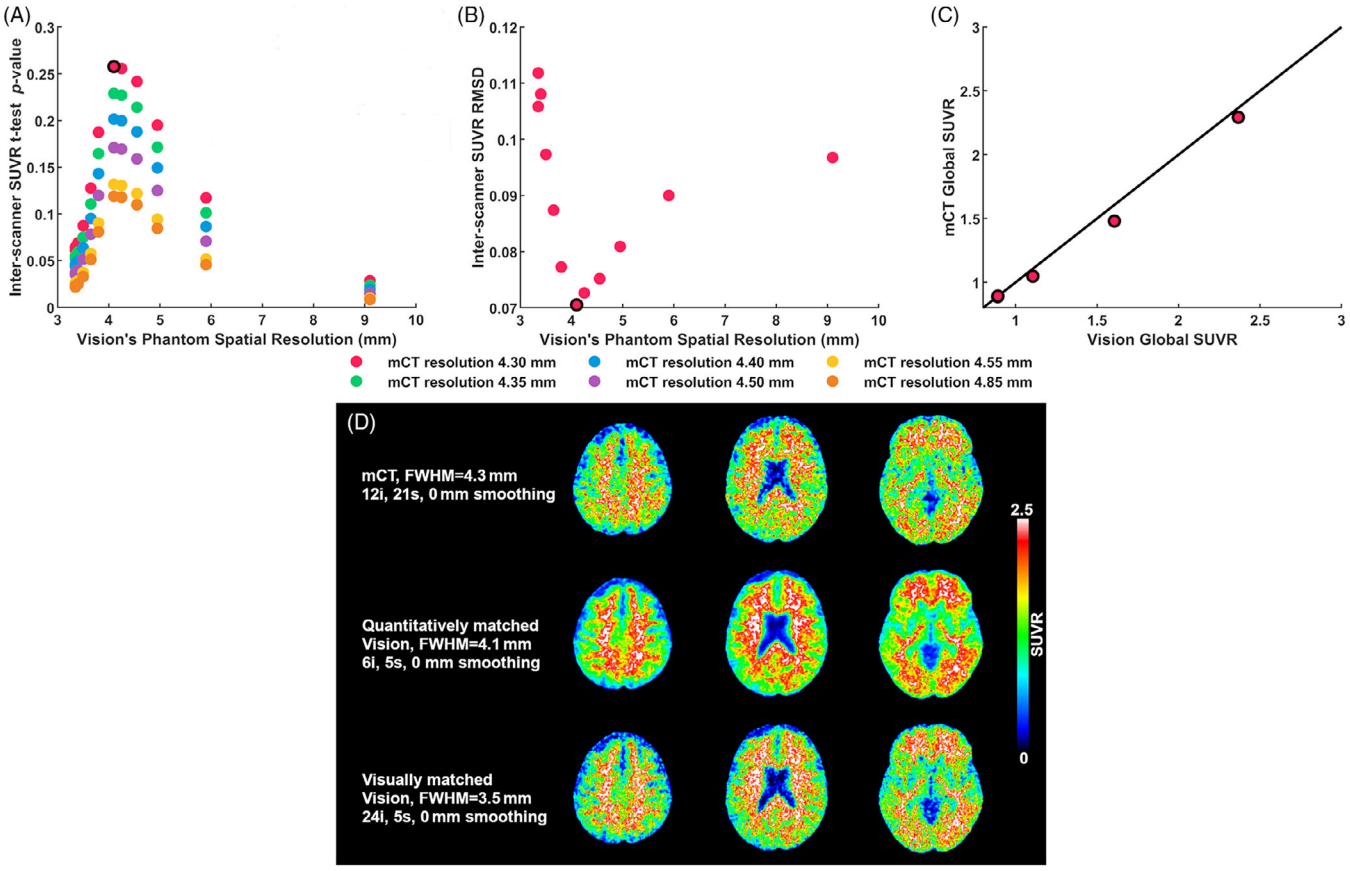

FIGURE 1 Vision–mCT harmonization. (A) Statistical similarity between paired Vision–mCT regional SUVRs, evaluated using a within-subject *t*-test (higher *p*-value indicates greater similarity). (B) Root mean square difference between global mCT and Vision SUVRs (mCT: FWHM = 4.3 mm across Vision FWHM resolutions). (C) Global SUVR scatterplot for five subjects using the harmonized reconstruction configuration pair, generated by calculating mean SUVR inside the neocortical mask defined by the CL project.[19] Four data points visible as two subjects' data overlap at (0.89, 0.89). (D) Visual comparison of the harmonization process between mCT and Vision using brain slices from a representative subject. First row: mCT reconstruction (FWHM = 4.3 mm); second row: quantitatively matched Vision reconstruction (FWHM = 4.1 mm); and third row: visually matched Vision reconstruction (FWHM = 3.5 mm). In A–C, black circles denote harmonized (maximally similar) pair of reconstruction configurations. The black line denotes the line of equal SUVRs between the scanners. FWHM, full width at half maximum; SUVR, standardized uptake value ratio.

Assessment of harmonization: Statistical similarity between SUVRs was assessed using paired (within subject) *t*-tests for each pair of reconstructions across scanners; the higher the *p*-value, the greater the similarity between the SUVRs. A second measure of harmonization, that of root mean-squared-differences (RMSD) between paired global SUVR values, was employed.

For completeness, Computational Analysis of PET by the Australian Imaging, Biomarker and Lifestyle (CapAIBL) was used to generate CL values for all reconstructions on each set of data (http://milxcloud.csiro.au).[21]

3 | RESULTS

In order to test our hypothesis that harmonization of Aβ-SUVR values between two scanners could be achieved by matching spatial resolutions, in the following we present SUVR similarity as a function of phantom-derived FWHM (resolution) and demonstrate that maximal SUVR harmonization (statistical similarity) is achieved at matched spatial resolution. We further verify that matched spatial resolution provides SUVRs with minimal RMSD.

3.1 | Vision-mCT harmonization

The harmonization results between the mCT and Vision scanners are presented in Figure 1. The barrel phantom-derived spatial resolutions of the mCT reconstructions were between 4.30 mm and 4.85 mm. The Vision reconstruction with spatial resolution of 4.1 mm (6*i*, 5*s*, 0 mm smoothing) gave rise to a maximal statistical similarity (highest *p*-value, $p = 0.26$, $t = -1.13$) between regional SUVRs (Figure 1A). The optimally harmonized reconstruction pair is marked in a black circle (Figure 1A,B), and resulted in the minimal RMSD across all reconstruction configuration pairs (Figure 1B). Paired global SUVRs calculated using the optimally harmonized reconstruction configuration pair were closely matched (Figure 1C).

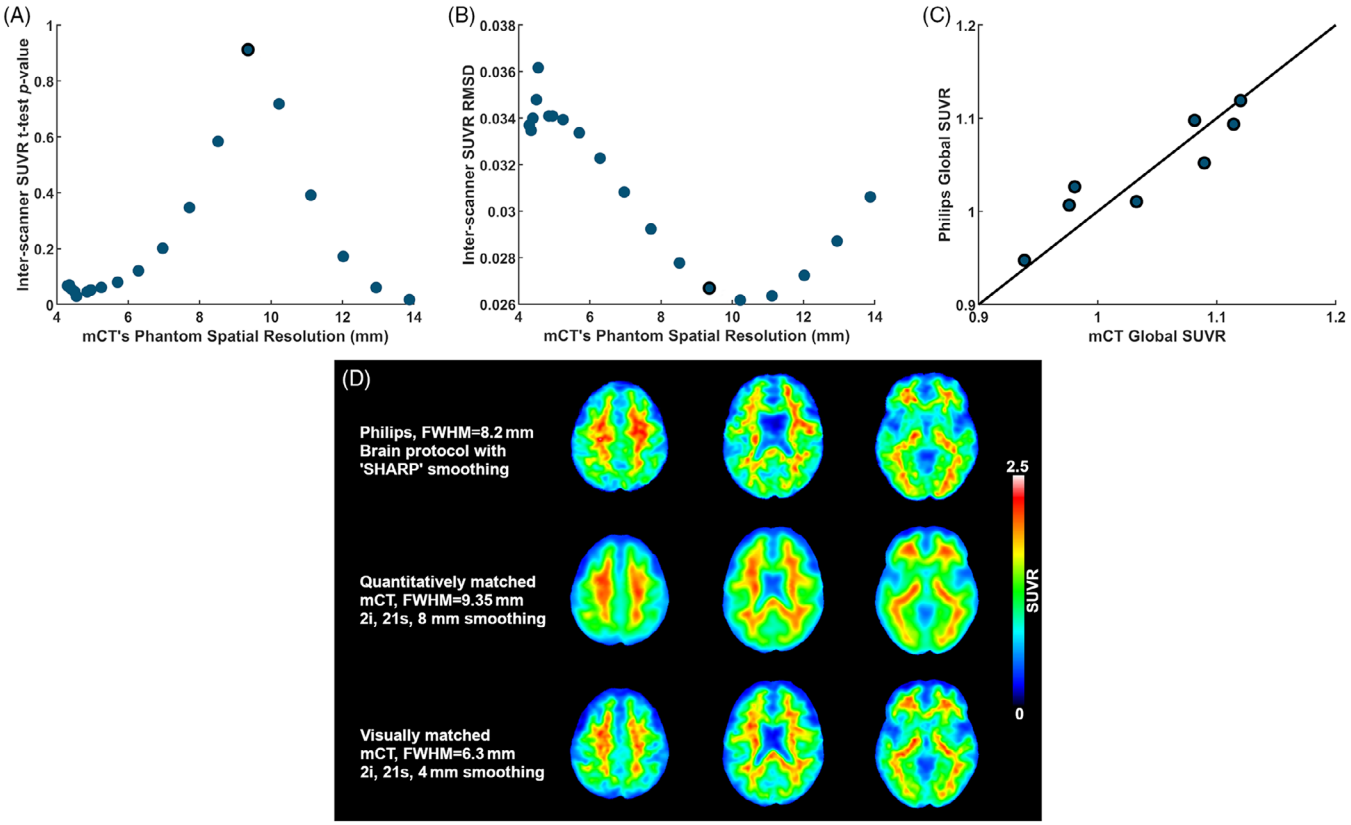

FIGURE 2 mCT-Philips harmonization. (A) Statistical similarity between paired mCT–Philips regional SUVRs, evaluated using a within subject *t*-test (higher *p*-value indicating greater similarity). (B) Root mean square difference between paired mCT and Philips Global SUVRs (Philips: FWHM = 8.2 mm across mCT FWHM resolutions). (C) Global SUVR scatterplot for eight subjects using the harmonized reconstruction configuration pair, generated by calculating mean SUVR inside the neocortical mask defined by the CL project.[19] (D) Visual comparison of the harmonization process between mCT and Philips using brain slices from a representative subject; First row: Philips reconstruction (FWHM = 8.2 mm), second row: quantitatively matched mCT reconstruction (FWHM = 9.35 mm), third row: visually matched mCT reconstruction (FWHM = 6.3 mm). In A–C, black circles denote harmonised (maximally similar) pair of reconstruction configurations. The black line denotes the line of equal SUVRs between the scanners. CL, Centiloid; FWHM, full width at half maximum; SUVR, standardized uptake value ratio.

Representative brain slices from a subject scanned by both mCT and Vision are shown in Figure 1D. The statistically matched mCT (FWHM = 4.3 mm, 12*i*, 21*s*, 0 mm smoothing, first row of Figure 1D) and Vision (FWHM = 4.1 mm, 24*i*, 5*s*, 0 mm smoothing, second row of Figure 1D) reconstruction configurations are less visually similar than the Vision reconstruction configuration with FWHM = 3.5 mm (24i, 5s, 0 mm smoothing, third row of Figure 1D), demonstrating that visual matching between mCT and Vision does not quantitatively harmonize Aβ-SUVR values.

3.2 | mCT-Philips harmonization

Figure 2 shows the harmonization results for the mCT-Philips scanner pair. There was only one standard brain mode clinical reconstruction algorithm for the Philips, with spatial resolution 8.2 mm. The mCT reconstruction configuration with FWHM = 9.35 mm (2*i*, 21*s*, 8 mm smoothing) resulted in maximal similarity (highest *p*-value, $p = 0.91$, $t = 0.11$) between regional SUVRs (Figure 2A). The RMSD between mCT and Philips global SUVRs across mCT resolutions demonstrates a broad trough with a minimum greater than 9.35 mm (Figure 2B). The eight subjects global SUVRs for the pair of harmonized reconstruction configurations were well matched (Figure 2C).

Representative brain slices from a subject scanned by both mCT and Philips are depicted in Figure 2D. The quantitatively harmonized Philips (FWHM = 8.2 mm, first row of Figure 2D) and mCT (FWHM = 9.35 mm, 2*i*, 21*s*, 8 mm smoothing, second row of Figure 2D) reconstruction configurations are less visually similar than the mCT reconstruction configuration with FWHM = 6.30 mm (2*i*, 21*s*, 4 mm smoothing, third row of Figure 2D), again demonstrating the necessity of quantifiable metrics for harmonization.

3.3 | Vision-Philips harmonization

Figure 3 shows the harmonization results for the Vision–Philips scanner pair. There was only one standard brain mode clinical reconstruction algorithm for the Philips, with spatial resolution 8.2 mm. The Vision reconstruction configuration with FWHM = 9.1 mm (1*i*, 5*s*, 0 mm smoothing) resulted in maximal similarity (highest *p*-value,

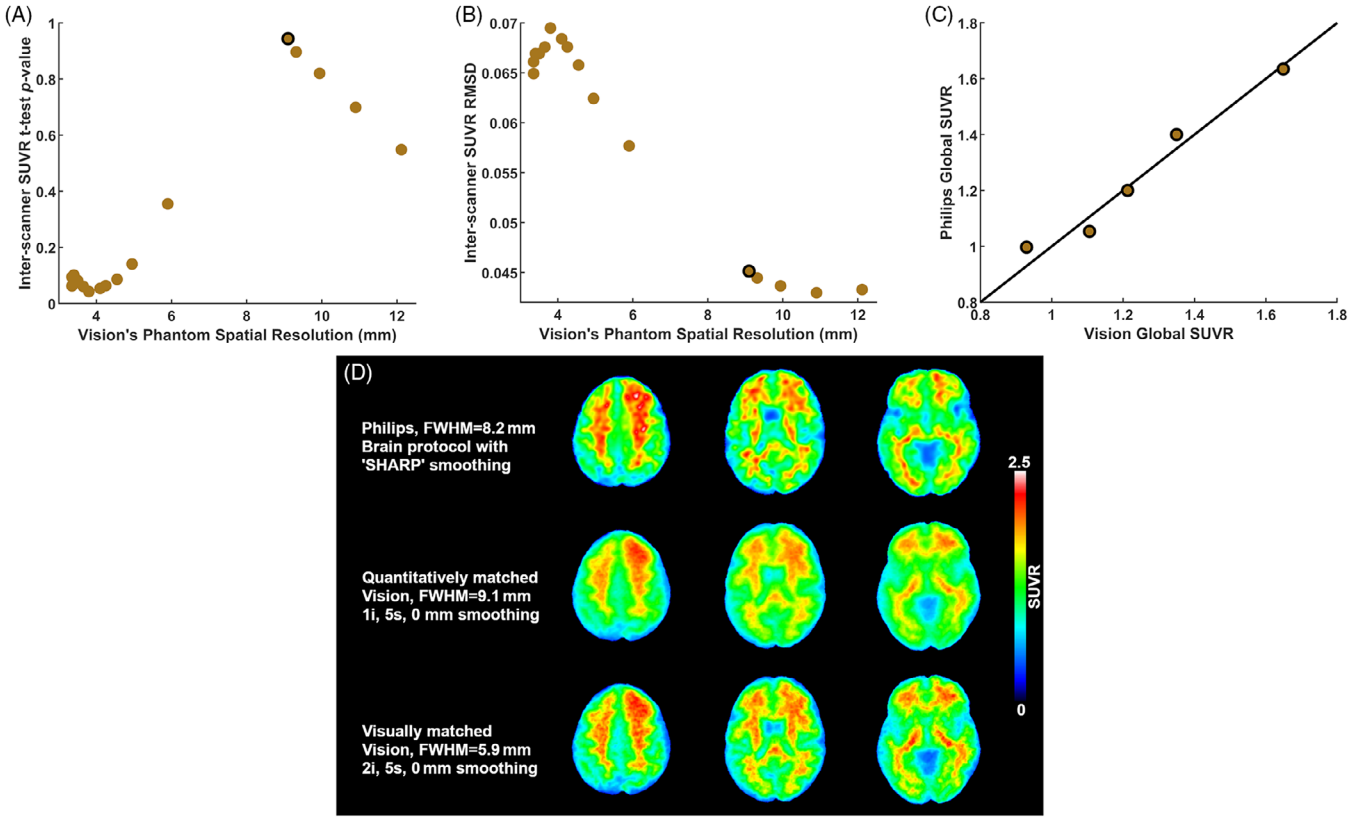

FIGURE 3 Vision–Philips harmonization. (A) Statistical similarity between paired Vision–Philips regional SUVRs, evaluated using a within-subject *t*-test (higher *p*-value indicating greater similarity). (B) Root mean square difference between paired Vision and Philips Global SUVRs (Philips: FWHM = 8.2 mm across Vision FWHM resolutions). (C) Global SUVR scatterplot for five subjects using the harmonized reconstruction configuration pair, generated by calculating mean SUVR inside the neocortical mask defined by the CL project.[19] (D) Visual comparison of the harmonization process between Vision and Philips using brain slices from a representative subject. First row: Philips reconstruction (FWHM = 8.2 mm), second row: quantitatively matched Vision reconstruction (FWHM = 9.1 mm), and third row: visually matched Vision reconstruction (FWHM = 5.9 mm). In A–C, black circles denote a harmonized (maximally similar) pair of reconstruction configurations. The black line denotes the line of equal SUVRs between the scanners. CL, Centiloid; FWHM, full width at half maximum; SUVR, standardized uptake value ratio.

$p = 0.94$, $t = 0.08$) between regional SUVRs (Figure 3A). The RMSD between Vision and Philips global SUVRs across Vision resolutions demonstrates a broad trough with minimum greater than 9.1 mm (Figure 3B). The five subjects' global SUVRs for the pair of harmonized reconstruction configurations were well matched (Figure 3C).

Representative brain slices from a subject scanned by both Vision and Philips are depicted in Figure 3D. The quantitatively harmonized Philips (FWHM = 8.2 mm, first row of Figure 3D) and Vision (FWHM = 9.1 mm, 1*i*, 5*s*, 0 mm smoothing, second row of Figure 3D) reconstruction configurations are less visually similar than the Vision reconstruction configuration with FWHM = 5.9 mm (2*i*, 5*s*, 0 mm smoothing, third row of Figure 3D), again demonstrating the necessity of quantifiable metrics for harmonization.

3.4 | Comparison with ADNI reconstructions

Regional SUVR-based statistical similarity and RMSD values of global SUVRs for ADNI reconstruction configurations are provided in Table S1. The barrel phantom-derived FWHM values for ADNI reconstructions are 4.60 mm, 3.95 mm, and 8.2 mm for mCT, Vision, and Philips, respectively. For the Vision-mCT scanner pair, regional SUVRs between the ADNI reconstructions were not significantly different ($p = 0.255$, $t = -1.139$) and RMSD between global SUVRs was 0.071. In contrast, regional SUVRs between the Philips and mCT reconstructions trended toward statistical significance (non-harmonized, $p = 0.058$, $t = 1.898$) and RMSD between global SUVRs was 0.031. Regional SUVR between the Vision and Philips were significantly different (non-harmonized, $p = 0.048$, $t = 1.981$) with RMSD between global SUVRs of 0.07.

3.5 | Centiloid harmonization comparison

The CL scale, as a linear transformation of global SUVR values, should be harmonized for harmonized SUVRs. Comparisons of CL values calculated using the three sets of quantitatively harmonized scanner reconstruction pairs are provided in Figure S1. CL values between

the mCT–Philips pair and between the Vision–Philips pair are well-matched. Comparison of Vision–mCT pair shows CL values that are larger for the Vision than mCT.

4 | DISCUSSION

In this study, we hypothesized that matching the spatial resolution of scanners, as measured by the Lodge barrel phantom method,[11] harmonizes Aβ-PET quantitation across scanners. This hypothesis was motivated by the knowledge that spatial resolution is affected by photon scattering, positron range of the radioisotope, hardware-specific limitations, and choice of reconstruction algorithm and parameter configuration[7,8]; consequently, partial volume effects affect PET quantitation.[6,9,10] ADNI introduced the process of post-reconstruction smoothing to achieve the same spatial resolution across scanners to harmonize FDG brain scans.[5] This procedure has not been validated for Aβ-PET imaging, and our recent study has demonstrated that smoothing may not be optimal for matching the spatial resolution, as it changed the Aβ-PET SUVR in the Aβ+ group and not in the Aβ– group.[6]

Our results have provided significant evidence in support of the hypothesis of harmonization via barrel phantom-derived spatial resolution, using two metrics of harmonization: statistical similarity (p-value) between paired regional SUVRs and RMSD between paired global SUVRs. This is an important and promising outcome, as our proposed harmonization procedure depends solely on phantom scans and no requirement for scanning subjects in multiple scanners. Furthermore, the overhead associated with acquiring Hoffman phantom data is greatly reduced in the Lodge method based on an ^{18}F-filled barrel phantom readily available in every center as a part of routine quality assurance.[11]

The two metrics of harmonization, statistical similarity and RMSD, provided minor differences in harmonized reconstruction configurations for the Vision–Philips and mCT–Philips comparisons. This may be due to differences in the post-reconstruction smoothing processes employed by Philips and Siemens software; Siemens reconstructions uses uniform smoothing throughout the brain, whereas Philips uses non-uniform, region-based smoothing throughout the brain. Although our recent study suggests that post-reconstruction smoothing may not be optimal to harmonize Aβ-SUVR between scanners,[6] in this research we applied smoothing in the harmonization process between mCT and Philips scanners, out of necessity as the lowest resolution that could be achieved by the mCT by changing the number of iterations and subsets with TOF enabled was much higher than the Philips resolution. In the future it can be explored whether harmonization can be achieved without enabling the TOF in the reconstruction process, as it may provide a wider range of spatial resolution values. In the current study, TOF was enabled for every scanner to make reconstructions consistent.

For comparison, we implemented the ADNI-proposed reconstruction configurations for the mCT, Vision, and Philips. They were shown to harmonize only the Aβ-PET SUVR quantitation between the Vision–mCT pair, and not the mCT–Philips and Vision–Philips pairs. This may be because ADNI reconstructions were proposed to standardize the PET image reconstruction across multiple centers; adhering to the proposed ADNI reconstruction configurations may effectively fulfill the intended purpose of standardizing image reconstruction protocol; however it may not achieve harmonization of Aβ-PET quantitation between the scanners.

We have demonstrated that visual matching of Aβ-PET images across scanners does not robustly harmonize Aβ-PET quantitation. Our proposed process of matching spatial resolution has been shown to be more effective in harmonizing Aβ-PET quantitation. One possible explanation for this discrepancy is that visual matching focuses on the similarity of noise characteristics in the reconstructed images, which may not impact PVE across the white matter and gray matter regions of the brain. In contrast, spatial image resolution primarily governs PVE.

It is important to note that there is significant variability in the performance of PET scanners used in clinical settings, and the proposed harmonization scheme may compromise the high spatial resolution capabilities of newer systems. In our previous research, we observed that high-resolution imaging is beneficial in both cross-sectional and longitudinal Aβ-PET studies.[6] Because storing raw PET data for retrospective reconstruction can be a significant overhead in clinical settings, a potential solution is to reconstruct PET data in three different settings: high resolution for center-specific studies, a harmonized reconstruction configuration for multicenter-specific studies, and a separate reconstruction for visual interpretation. This approach allows for utilization of the most appropriate reconstruction method based on the context of the study.

The choice of reconstruction configuration influences the spatial resolution of the PET image, making the Aβ-PET quantitation different across the reconstruction. [6] Therefore, it is important to note that, even though the two scanners are identical in brand and model, the Aβ-PET quantitation between the scanners may differ if the choice of the reconstruction configurations are different between the scanners. Consequently, harmonization between the same scanners would be necessary.

Limitations of the current study include: (1) the mCT–Philips scanner comparison lacks Aβ+ subjects due to unavailability of participants; (2) a limited number of subjects were used in each scanner pair comparison; (3) only one Aβ-PET tracer was used, and the results may change with the level of off-target binding (further studies can be explored in future to validate the proposed harmonization method on the other tracers);, and (4) analysis was done using one Aβ-PET processing pipeline; validation of several Aβ-PET pipelines could be conducted in future studies. Irrespective of these limitations, the results provide evidence in support of the hypothesis that barrel phantom-derived spatial resolution can be used for scanner harmonization in Aβ-PET imaging to reduce quantitation differences.

5 | CONCLUSION

Using Aβ-PET data from three sets of participants, each scanned on a pair of PET scanners, we conclude that the process of matching spatial resolution measured by a barrel phantom can be used to minimize the

Aβ-PET quantitation differences between the scanners. These promising results encourage application in Aβ-PET multi-center trials and for PET camera updates during longitudinal studies.

ACKNOWLEDGMENTS

The authors acknowledge the facilities and scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Melbourne Brain Center Imaging Unit, the University of Melbourne. The first author would also like to acknowledge the Rowden White scholarship for its assistance in his research. The authors acknowledge the contributions of the the research teams at the Australian Imaging, Biomarker and Lifestyle (AIBL). The research was supported by National Health and Medical Research Council grants 1152623, 1140853 and 1132604.

Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

ORCID

Catherine E. Davey https://orcid.org/0000-0002-5672-7941

REFERENCES

1. Villemagne VL, Ong K, Mulligan RS, et al. Amyloid imaging with 18F-florbetaben in alzheimer disease and other dementias. *J Nucl Med*. 2011;52(8):1210-1217.
2. Rowe CC, Villemagne VL. Brain amyloid imaging. *J Nucl Med Technol*. 2013;41(1):11-18.
3. Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). *J Nucl Med*. 2010;51(6):913-920.
4. 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(3):319-329.
5. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. *Neuroimage*. 2009;46(1):154-159.
6. Ruwanpathirana G, Williams R, Masters C, Rowe C, Davey C, Johnston L. Impact of PET reconstruction parameters on the quantitation of PET Aβ-amyloid. *J Nucl Med*. 2022;63(2):3281.
7. Andersen FL, Klausen TL, Loft A, Beyer T, Holm S. Clinical evaluation of PET image reconstruction using a spatial resolution model. *Eur J Radiol*. 2013;82(5):862-869.
8. Thomas BA, Erlandsson K, Modat M, et al. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2011;38(6):1104-1119.
9. Akamatsu G, Ikari Y, Nishio T, et al. Optimization of image reconstruction conditions with phantoms for brain FDG and amyloid PET imaging. *Ann Nucl Med*. 2016;30(1):18-28.
10. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med*. 2007;48(6):932-945.
11. Lodge MA, Leal JP, Rahmim A, Sunderland JJ, Frey EC. Measuring PET spatial resolution using a cylinder phantom positioned at an oblique angle. *J Nucl Med*. 2018;59(11):1768-1775.
12. Kaalep A, Sera T, Oyen W, et al. EANM/EARL FDG-PET/CT accreditation—summary results from the first 200 accredited imaging systems. *Eur J Nucl Med Mol Imaging*. 2018;45(3):412-422.
13. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: eANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42(2):328-354.
14. Verwer EE, Golla SSV, Kaalep A, et al. Harmonisation of PET/CT contrast recovery performance for brain studies. *Eur J Nucl Med Mol Imaging*. 2021;48:2856-2870.
15. Jagust WJ, Bandy D, Chen K, et al. The Alzheimer's disease neuroimaging initiative positron emission tomography core. *Alzheimer's Dement*. 2010;6(3):221-229.
16. Minoshima S, Drzezga AE, Barthel H, et al. SNMMI procedure standard/EANM practice guideline for amyloid PET imaging of the brain 1.0. *J Nucl Med*. 2016;57(8):1316-1322.
17. Ikari Y, Akamatsu G, Nishio T, et al. Phantom criteria for qualification of brain FDG and amyloid PET across different cameras. *EJNMMI Phys*. 2016;3(1):3-23.
18. ADNI. Alzheimer ' s Disease Neuro-Imaging III (ADNI3) Study. Published online 2016:1-33.
19. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's Dement*. 2015;11(1):1-15.e4.
20. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. *Neuroimage*. 2020;206:116189. August 2019.
21. Bourgeat P, Doré V, Fripp J, et al. Implementing the centiloid transformation for 11C-PiB and β-amyloid 18F-PET tracers using CapAIBL. *Neuroimage*. 2018;183(March):387-393.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ruwanpathirana GP, Williams RC, Masters CL, Rowe CC, Johnston LA, Davey CE. Inter-scanner Aβ-PET harmonization using barrel phantom spatial resolution matching. *Alzheimer's Dement*. 2024;16:e12561. https://doi.org/10.1002/dad2.12561