10.1148/radiol.2018180940

Appendix E1

Materials and Methods–MR Data Acquisition

The T1-weighted sequence was a sagittal spoiled gradient recalled acquisition with the following parameters: echo time (TE) = 3.092 ms, repetition time (TR) = 7.648 ms, inversion time (TI) = 400 ms, number of averages = 1, flip angle = 11 degrees, acquisition time = 5 min 47 sec; the T2-weighted sequence is an axial fast spin echo acquisition with the following parameters: TE = 100.896 ms, TR = 3484 ms, number of averages = 1, flip angle = 111 degrees, acquisition time = 4 min 40 sec; the T2 FLAIR-weighted sequence is acquired with the following parameters: TE = 161.488 ms, TR = 6000 ms, TI = 1773 ms, number of averages = 1, flip angle = 90 degrees, acquisition time = 7 min 40 sec.

Materials and Methods–Software

Image preprocessing and analyses were done using the open-source software FSL 5.0 (FMRIB, Oxford, UK, 2012) and FreeSurfer 5.3.0 (Martinos Center, MGH, Charlestown, MA, 2013).

Statistical analyses were done using Stata 15.1 (StataCorp LP, College Station, TX) and R version 3.3.1 (r-project.org) with version 1.3 of the "ExactCIdiff" package.

Results–Assessment of Image Quality

In certain images, there appears to be "super-resolution" effects in the cortical regions while having blurrier uptake patterns in the white matter regions. While U-Nets nominally preserve the resolution of the learned gold standard image, certain areas might appear to have lower or higher resolution due to the multiple-channel inputs from new data. In this case, the higher resolution in the cortical areas might be due to the high-resolution MR images and the blurriness in the white matter regions might be due to the network outputting smooth patches in response to the noisy nature of the input. In the pixelwise difference maps (Fig E1) between the synthesized PET+MR images and the full-dose images, though, we have not seen obvious edge enhancement. In the future, further analysis could be done using the frequency domain blur measure (FBM) or edge blur measure (EBM) to quantitatively assess the edge effects, possibly in tandem with including or excluding contrasts for training.

An amyloid-negative participant is shown (Figure E2) in addition to the amyloid-positive participant in Figure 3.

An image of the "blotches" observed (Fig E3, indicated by the arrows) in a representative study participant is shown below. The physicians observed blotches in 12/40 datasets in the PET+MR images and 10/40 in the PET-only images. For each volume with blotches, relatively few such blotches were observed (5.25 ± 3.49 and 5.80 ± 4.73 blotches in the brain for the PET+MR and PET-only images, respectively); presence of the blotches, if any, was spread throughout the whole volume. The location of the blotches corresponded with high-activity speckles in the low-dose PET images. The presence of these blotches did not interfere with the clinical interpretation of the image.

We also assessed whether the blotches would affect voxel-wise SUV variability in the cerebellar cortex, as increased variability in this region would affect SUVR normalization. For this we calculated the coefficient of variation (CV, voxel-wise mean/voxel-wise SD) in the FreeSurfer-defined left and

right cerebellar cortices. The mean left cerebellar cortex CV across subjects is 0.35, 0.57, 0.36, 0.33 for the full-dose, low-dose, PET-only, and PET+MR images respectively. For the right cerebellar cortex the values are 0.33, 0.46, 0.32, 0.31. The mean of the CV across subjects shows that the variability in cerebellar SUV is comparable for the full-dose and PET-only images, while the PET+MR images have the smallest CV (tested with the two-tailed paired *t* test at the $P = .05$ level). In addition, this shows that the blotches did not affect variability in the synthesized images as we would have expected increased CV otherwise.

Results–Clinical Readings

Kendall's tau-b, Krippendorff's alpha, and symmetry tests were used to evaluate between/within raters' agreements. The intrareader agreement on amyloid uptake status was first shown to be high (Table E1). Next, the interreader agreement on amyloid uptake status and image quality scores were high (Tables E2, E3). These results were used as the basis of pooling the two readers' readings (for the first time point only) for subsequent analyses. Confusion matrices for the image quality proportions are provided in Table E4. The predicted percentage of high image quality score (4-5) on each method and the difference between methods is shown in Table E5.

Results–Region-based Analyses

The concordance analyses and Bland-Altman plots of the regional SUVR comparisons between image types are provided in Figures E4–E6 and Tables E6, E7.

Table E1: Confusion Matrices for Intrareader Agreement of the Amyloid Status (Positive/Negative) for the Readers

Reader 1		Time 1		
		Negative	Positive	Total
Time 2	Negative	28	3	31
	Positive	0	9	9
	Total	28	12	40
Reader 2		Time 1		
		Negative	Positive	Total
Time 2	Negative	28	3	31
	Positive	1	8	9
	Total	29	11	40

Only the full-dose images are compared. For the first reader, the tau-b statistic was 0.823 ($P < .001$), Krippendorff's alpha was 0.809 (95% CI 0.552 ∼ 1), and the $P = .25$ for the symmetry test. For the second reader, the tau-b statistic was 0.741 ($P < .001$), Krippendorff's alpha was 0.737 (95% CI 0.433 ∼ 0.94), and the $P = .625$ for the symmetry test.

Table E2: Confusion Matrices for Interreader Agreement of the Amyloid Status (Positive/Negative), Showing That the Readers Agreed on the Amyloid Status for Readings at Both Time Points

Time 1		Reader 2		
		Negative	Positive	Total
Reader 1	Negative	74	5	79
	Positive	7	40	47
	Total	81	45	126
Time 2		Reader 2		
		Negative	Positive	Total
Reader 1	Negative	29	2	31
	Positive	2	7	9
	Total	31	9	40

All interpretable images were compared for the first time point and only the full-dose images were read a second time. For the first time point the tau-b statistic was 0.795 ($P < .001$), Krippendorff's alpha was 0.795 (95% CI 0.673 ∼ 0.902), and the $P = .774$ for the symmetry test. For the second time point the tau-b statistic was 0.713 ($P < .001$), Krippendorff's alpha was 0.717 (95% CI 0.419 ∼ 0.934), and the $P = 1$ for the symmetry test.

Table E3: Confusion Matrix for Interreader Agreement of the Image Quality Score (5 = Excellent)

		Reader 2					
		1	2	3	4	5	Total
Reader 1	1	22	6	0	0	0	28
	2	6	7	4	0	0	17
	3	0	1	20	10	0	31
	4	0	0	10	38	13	61
	5	0	0	0	7	16	23
	Total	28	14	34	55	29	160

The tau-b statistic was 0.798 ($P < .001$), Krippendorff's alpha was 0.867 (95% CI 0.814 ∼ 0.904), and the $P = .494$ for the symmetry test. This shows that the readers agreed strongly on scoring and did not systemically over-or under-call scores with respect to each other. No scores were over 1 category apart for any of the studies. Only the first time point was scored.

Table E4: Confusion Matrices for Image Quality Score (5 = Excellent) Dichotomization between the Full-Dose Images and the PET+MR Images and between the Full-Dose Images and the PET-Only Images

		Images from the PET+MR model		
		1–3	4–5	Total
Full-dose images	1–3	2	2	4
	4–5	10	66	76
	Total	12	68	80
		Images from the PET-only model		
		1–3	4–5	Total
Full-dose images	1–3	4	0	4
	4–5	52	24	76
	Total	56	24	80

Only the first time point was scored. 95% of Full-dose images are scored 4–5 while it is 85% for the PET+MR images and 30% for the PET-only images. The 95% CI of the difference of proportions (-10% and-65%) is-20%∼-1% and-75%∼-53% for the PET+MR images and the PET-only images respectively.

Table E5: Predicted Percentage of High Image Quality Score (4–5) on Each Method and the Difference between Methods

Method	Predicted percentage (proportions)	*P* value	95% CI (LL, UL)
Full-dose images	0.95 (76/80)		(0.90,1.00)
Images from the PET+MR model	0.85 (68/80)		(0.75,0.95)
(PET+MR)-(Full-dose)	–0.1 (8/80)	<0.001	(-0.15,-0.05)
Method	Predicted percentage (proportions)	*P* value	95% CI (LL, UL)
Full-dose images	0.95 (76/80)		(0.90,1.00)
Images from the PET-only model	0.3 (24/80)		(0.15,0.45)
(PET-only)-(Full-dose)	–0.65 (52/80)	<0.001	(-0.85,-0.45)

The criterion for *P* value here is 0.025 (0.05/2), corrected for two comparisons (Full-dose images versus Images from the PET+MR model, Full-dose images versus Images from the PET-only model). Significant *P* value means the difference of proportion between two methods is significantly different from zero. Abbreviations used: CI = confidence interval, LL = lower limit, UL = upper limit.

Table E6: Region-Based Intraclass Correlation of the SUVR Values, Subdivided by Patient Diagnosis

Full-dose images versus	Group	Intraclass correlation	95% CI (LL, UL)
Low-dose images	AD	0.82	0.69, 0.91
	HC1	0.76	0.59, 0.87
	HC2	0.8	0.69, 0.88
	Lewy	0.85	0.77, 0.90
	MCI	0.78	0.67, 0.85
	PD	0.79	0.67 0.88
	P_L_D	0.93	0.93, 0.97
Images from the PET+MR model	AD	0.97	0.96, 0.98
	HC1	0.95	0.91, 0.98
	HC2	0.9	0.84, 0.94
	Lewy	0.92	0.89, 0.94
	MCI	0.94	0.89, 0.97
	PD	0.96	0.95, 0.97
	P_L_D	0.94	0.92, 0.96
Images from the PET-only model	AD	0.96	0.94, 0.97
	HC1	0.94	0.92, 0.95
	HC2	0.9	0.82, 0.95
	Lewy	0.94	0.92, 0.96
	MCI	0.94	0.90, 0.97
	PD	0.92	0.89, 0.95
	P_L_D	0.94	0.92, 0.96

Abbreviations used: AD = Alzheimer's disease, HC1 = healthy control from cohort 1, HC2 = healthy control from cohort 2, Lewy = Dementia with Lewy bodies, MCI = mild cognitive impairment, PD = Parkinson's disease, P_L_D = Parkinsonism with language decline, CI = confidence interval, LL = lower limit, UL = upper limit.

Table E7: Region-Based SUVR Value Differences between Image Types, Subdivided by Patient Diagnosis

Full-dose images versus	Group	No. of regions	Mean (SD)	95% CI (LL, UL)
Low-dose images	AD	879	0.18 (0.29)	−0.39, 0.75
	HC1	438	0.29 (0.27)	−0.24, 0.81
	HC2	880	0.17 (0.26)	−0.33, 0.67
	Lewy	111	0.15 (0.13)	−0.12, 0.41
	MCI	330	0.23 (0.22)	−0.21, 0.67
	PD	1649	0.16 (0.23)	−0.28, 0.6
	P_L_D	109	0.06 (0.08)	−0.1, 0.21
	Total	4396	0.18 (0.25)	−0.31, 0.67
Images from the PET+MR model	AD	879	0.03 (0.14)	−0.25, 0.31
	HC1	438	0.02 (0.18)	−0.33, 0.38
	HC2	880	0.10 (0.18)	−0.26, 0.45
	Lewy	111	−0.13 (0.09)	−0.3, 0.05
	MCI	330	0.03 (0.16)	−0.3, 0.35
	PD	1649	0.02 (0.12)	−0.21, 0.25
	P_L_D	109	0.07 (0.07)	−0.06, 0.21
	Total	4396	0.04 (0.15)	−0.26, 0.33
Images from the PET-only model	AD	879	0.09 (0.14)	−0.19, 0.37
	HC1	438	0.11 (0.17)	−0.23, 0.44
	HC2	880	0.10 (0.18)	−0.26, 0.46
	Lewy	111	0 (0.12)	−0.22, 0.23
	MCI	330	0.05 (0.15)	−0.26, 0.35
	PD	1649	0.07 (0.15)	−0.23, 0.37
	P_L_D	109	0.07 (0.07)	−0.07, 0.22
	Total	4396	0.08 (0.16)	−0.23, 0.39

Abbreviations used are same as those in Table E6.