NeuroImage 184 (2019) 555–565

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements

Hyunwoo Lee [a,*], Kunio Nakamura [b], Sridar Narayanan [a], Robert A. Brown [a], Douglas L. Arnold [a], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
[b] Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

A B S T R A C T

Objective: Longitudinal MRI studies are often subjected to mid-study scanner changes, which may alter image characteristics such as contrast, signal-to-noise ratio, contrast-to-noise ratio, intensity non-uniformity and geometric distortion. Measuring brain volume loss under these conditions can render the results potentially unreliable across the timepoint of the change. Estimating and accounting for this effect may improve the reliability of estimates of brain atrophy rates.

Methods: We analyzed 237 subjects who were scanned at 1.5 T for the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and were subject to intra-vendor or inter-vendor scanner changes during follow-up (up to 8 years). Sixty-three subjects scanned on GE Signa HDx and HDxt platforms were also subject to a T1-weighted sequence change from Magnetization Prepared Rapid Gradient Echo (MP-RAGE) to Fast Spoiled Gradient Echo with IR Preparation (IR-FSPGR), as part of the transition from ADNI-1 to ADNI-2/GO. Two-timepoint percentage brain volume changes (PBVCs) between the baseline "screening" and the follow-up scans were calculated using SIENA. A linear mixed-effects model with subject-specific random slopes and intercepts was applied to estimate the fixed effects of scanner hardware changes on the PBVC measures. The same model also included a term to estimate the fixed effects of the T1-weighted sequence change.

Results: Different hardware upgrade or change combinations led to different offsets in the PBVC (SE; p): Philips Intera to Siemens Avanto, −1.81% (0.30; $p < 0.0001$); GE Genesis Signa to Philips Intera, 0.99% (0.47, $p = 0.042$); GE Signa Excite to Signa HDx, 0.33% (0.095, $p = 0.0005$); GE Signa Excite to Signa HDxt, −0.023% (0.23, $p = 0.92$); GE Signa Excite to Signa HDx to Signa HDxt, 0.25% (0.095, $p = 0.010$) and 0.27% (0.16, $p = 0.098$), respectively; GE Signa HDx to Signa HDxt, −0.24% (0.25, $p = 0.34$); Siemens Symphony to Symphony TIM, −0.39% (0.16; $p = 0.019$). The sequence change from MP-RAGE to IR-SPGR was associated with an average −1.63% (0.12; $p < 0.0001$) change.

Conclusion: Inter-vendor scanner changes generally led to greater effects on PBVC measurements than did intra-vendor scanner upgrades. The effect of T1-weighted sequence change was comparable to that of the inter-vendor scanner changes. Inclusion of the corrective fixed-effects terms for the scanner hardware and T1-weighted sequence changes yielded better model goodness-of-fits, and thus, potentially more reliable estimates of whole-brain atrophy rates.

1. Introduction

Measuring brain atrophy using magnetic resonance imaging (MRI) is a topic of significant interest in the study of neurological disorders such as Alzheimer's disease (AD). These diseases result in a range of pathological processes that lead to progressive neuronal, axonal, and dendritic degeneration, and ultimately, central nervous system (CNS) tissue atrophy (Jack et al., 2013). For example, a longitudinal MRI study has reported that the annualized rates of volume loss in the whole-brain (WB) and hippocampus were on average more than two-fold and four-fold higher, respectively, in AD patients compared to age-matched healthy controls (Leung et al., 2013). Indeed, higher rates of brain volume loss in patients with AD or mild cognitive impairment (MCI) are associated with higher rates of decline in cognitive measures (Evans et al., 2010). For these reasons, MRI measures of brain volume change are widely recognized as markers of progression of neurodegeneration (Frisoni et al., 2010).

Various technical factors can influence MRI-based brain volume change measurements, especially when calculating longitudinal changes using serial images. For example, head motion (Preboske et al., 2006), inconsistent image contrast (Preboske et al., 2006), different levels of noise (Preboske et al., 2006), gradient non-linearity (Takao et al., 2010), intensity non-uniformity (Takao et al., 2010), inconsistent subject positioning (Caramanos et al., 2010), number of head coil channels (Krueger

* Corresponding author. 3801, Rue University, WB327, Montreal, Quebec, H3A 2B4, Canada.
E-mail address: hyunwoo.lee@mail.mcgill.ca (H. Lee).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.neuroimage.2018.09.062
Received 29 May 2018; Received in revised form 10 August 2018; Accepted 21 September 2018
Available online 22 September 2018

et al., 2012), and choice of image analysis methods (Nakamura et al., 2014; Popescu et al., 2016) all can affect measurement outcomes. Several single-site studies have investigated the effects of varying acquisition protocols on MRI outcomes. For example, Preboske and colleagues showed in a scan-rescan study that 1) implementing different flip angles and 2) switching from conventional to fast spoiled gradient echo (SPGR) sequence resulted in significant brain volume differences (Preboske et al., 2006). Han and colleagues showed that the average cortical thickness variability did not change significantly across an intra-vendor scanner upgrade from Siemens Sonata 1.5 T to Avanto 1.5 T (Han et al., 2006). However, different pulse sequences and image processing pipelines led to poorer scan-rescan reproducibility (Han et al., 2006).

Notably, Jovicich and colleagues assessed the effect of scanner vendor on brain volumes obtained with FreeSurfer. A scan-rescan analysis with Siemens Sonata 1.5 T and GE Signa 1.5 T showed that several structures had non-zero bias in the mean volume difference. Also, depending on the structure, the Sonata-Signa test set had bias in the mean volume difference compared to the Siemens Sonata scan-rescan reference. The authors noted that these effects need to be considered in longitudinal studies in which the effects of interest can be subtle (Jovicich et al., 2009).

The issue may become more significant when multiple MRI scanners are used, such as in a multi-center trial or when a significant mid-study change in the scanner occurs. For example, Kruggel and colleagues demonstrated that different 1.5 T and 3.0 T scanner platforms provide different levels of image quality, as measured by signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and mutual information of the joint histogram, and that these affected brain volume measurements (Kruggel et al., 2010). Therefore, interpretation of data from multiple scanning platforms must be done carefully since the effect of the platform change may confound the true effects of interest. Yet, despite the complexity, multi-center studies have important advantages of being able to recruit a larger number of participants across a wider range of population, and to increase the generalizability of the findings. Direct comparison of the MRI data acquired from multiple scanners, however, is likely more reliable when the data have been acquired in a similar manner using similar field strength and acquisition parameters. Recognizing this, a number of multi-center studies have implemented harmonized imaging protocols in order to reduce measurement variability due to protocol differences (Cannon et al., 2014; Mueller et al., 2005; Shinohara et al., 2017).

The Alzheimer's Disease Neuroimaging Initiative study (ADNI) is a longitudinal, multi-center study that acquired MRI data using a variety of 1.5 T and 3.0 T scanners from General Electric (GE), Siemens, and Philips (Jack et al., 2015). In particular, the study focused on designing and implementing standardized acquisition methods, as well as performing centralized image post-processing and quality control (Jack et al., 2015). These qualities make ADNI a suitable dataset to test the hypothesis that a change in the scanning platform can bias WB volume change measurements, and estimate the magnitude of this effect between unique pairs of MRI scanners. To do this, we identified all subjects from the ADNI-1 1.5 T study who had any MRI scanner change or upgrade during the follow-up. Also, we identified a subset of subjects who had a T1-weighted sequence change during the follow-up. Then, we used a linear mixed-effects (LME) model to estimate the rates of WB atrophy, as well as the effects of different MRI scanner changes or upgrade combinations and T1-weighted sequence change on percentage brain volume change (PBVC) measurements.

2. Methods

2.1. Data acquisition

The Alzheimer's Disease Neuroimaging Initiative (ADNI) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu) on 2015-01-21. Detailed MRI protocols are reported on the ADNI protocol website: (http://adni.loni.usc.edu/methods/documents/mri-protocols).

2.2. Subject selection

We started with N = 819 subjects (screening diagnosis, Normal = 229, MCI = 401, AD = 189; baseline diagnosis, Normal = 229, MCI = 398, AD = 192) officially enrolled in ADNI-1, who had baseline and follow-up visits conducted on 1.5 T scanners using the ADNI-specified 3-dimensional (3D) T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) sequence. Subject demographics are shown in Table 1. A subset of these subjects continued with 1.5 T MRI during the ADNI-Grand Opportunity (GO) and ADNI-2 phases. N = 818 of these subjects coincided with the 1.5 T ADNI-1 standard "screening-visits" dataset reported by Wyman and colleagues (Wyman et al., 2013). We excluded N = 46 subjects with only a single timepoint. For the remaining N = 773 subjects, the MRI scanner vendor (GE, Siemens, Philips) and the scanner model were identified for each timepoint. Scanner change was noted if the scanner models used during any of the follow-up timepoints did not match those of the baseline scan. Accordingly, N = 271 (Normal = 80, MCI = 141, AD = 50) subjects had scanner upgrades or changes (referred to as "Chg+" subgroup) versus N = 502 (Normal = 138, MCI = 241, AD = 123) who did not (referred to as "Chg-" subgroup) (Tables 2 and 3). Each timepoint had two back-to-back MP-RAGE scans, and we analyzed the first scan whenever possible.

2.3. MRI scanner information

Baseline scans were distributed between scanners as follows: GE (N = 381, 49.3%), Siemens (N = 294, 38.0%), and Philips (N = 98,

Table 1
Basic subject demographics – ADNI 1.5 T.

Characteristics	All subjects (N = 773)	Normal control subjects (N = 218)	Mild cognitive impairment subjects (N = 382)	Alzheimer's disease subjects (N = 173)
Mean age at baseline (SD) [range], yr	75.3 (6.8) [55.2–91.0]	76.0 (5.1) [60.0–89.7]	74.9 (7.3) [55.2–89.4]	75.2 (7.6) [55.2–91.0]
Sex, Female:Male	322:451	104:114	137:245	81:92
	Subjects with no scanner change or upgrade (N = 502)	Normal control subjects (N = 138)	Mild cognitive impairment subjects (N = 241)	Alzheimer's disease subjects (N = 123)
Mean age at baseline (SD) [range], yr	75.3 (7.0) [55.2–91.0]	76.0 (5.6) [60.0–89.7]	74.9 (7.3) [56.2–89.4]	75.2 (7.7) [55.2–91.0]
Sex, Female:Male	211:291	64:74	89:152	58:65
	Subjects with scanner change or upgrade (N = 271)	Normal control subjects (N = 80)	Mild cognitive impairment subjects (N = 141)	Alzheimer's disease subjects (N = 50)
Mean age at baseline (SD) [range], yr	75.3 (6.5) [55.2–87.8]	76.1 (4.0) [70.0–87.7]	75.0 (7.3) [55.2–87.8]	75.1 (7.5) [56.7–85.6]
Sex, Female:Male	111:160	40:40	48:93	23:27

Table 2
MRI scanner information for subjects without MRI scanner change or upgrade.

Subjects with no MRI scanner change or upgrade				
Baseline 1.5 T MRI scanner models	Total number of subjects (Total N = 502)	Number of normal control subjects (Total N = 138)	Number of MCI subjects (Total N = 241)	Number of Alzheimer's disease subjects (Total N = 123)
GE Genesis Signa	38	10	21	7
GE Signa Excite	117	33	47	37
GE Signa HDx	10		5	5
Siemens Avanto	56	18	25	13
Siemens Sonata	84	22	45	17
Siemens SonataVision	6	1	4	1
Siemens Symphony	108	31	55	22
Philips Achieva	19	5	10	4
Philips Intera	64	18	29	17

12.7%). A scanner change affected subjects who started scanning on one of seven scanner models distributed as follows: GE (N = 216, 79.7%), Siemens (N = 40, 14.8%), and Philips (N = 15, 5.5%). Thirteen combinations of inter- or intra-vendor scanner upgrade or change occurred. The majority of the cases involved GE scanners (Table 3). Fig. 1 shows an example pair of images acquired from a single subject using two different scanners from two different vendors (i.e. an inter-vendor scanner change). Similarly, Fig. 2 provides an example from two scanning platforms from a single vendor (i.e. an intra-vendor upgrade).

2.4. 3D T1-weighted sequence information

All Siemens and Philips scanners used the MP-RAGE sequence. It should be noted that GE scanners used a "work-in-progress" version of MP-RAGE during the ADNI-1 phase, and then switched to a GE product Fast Spoiled Gradient Echo with IR Preparation (IR-FSPGR) sequence for the ADNI-GO and ADNI-2 phases (Jack et al., 2010). This affected N = 63 Chg + subjects who had extended 1.5 T follow-up (e.g. ADNI-GO month 48 and beyond, or ADNI-2) on some of the GE Signa HDx or HDxt scanners. Note that the switch to the IR-FSPGR sequence occurred after the GE intra-vendor scanner upgrade. Fig. 3 shows an example pair of images acquired from a single subject using the two different sequences.

2.5. Image processing

We started with un-preprocessed 3D T1-weighted images (i.e. "Original") from the ADNI database. The images were preprocessed using the following steps: 1) nonparametric intensity non-uniformity normalization using N3 (Sled et al., 1998), 2) standard-space registration using the ICBM 2009c nonlinear symmetric template (Fonov et al., 2009), 3) brain extraction using BEaST (Eskildsen et al., 2012). Two-timepoint PBVCs were measured with SIENA (Smith et al., 2002), part of FSL (Smith et al., 2004). The baseline "screening" scans were designated as the reference (i.e. 100%), and all subsequent PBVCs were estimated with respect to the baseline. This produced a WB volume time course for each subject.

2.6. Statistical analysis

LME models with fixed (population-average) and random (subject-specific) effects have been frequently applied to model longitudinal brain atrophy outcomes (Chua et al., 2015; Jones et al., 2013; Leung et al.,

Table 3
MRI scanner information for subjects with MRI scanner change or upgrade.

Subjects with MRI scanner change or upgrade. The combinations included in the analysis are in boldface					
1.5 T MRI scanner model combination; "Original scanner" To "Changed scanner"	Total number of subjects (Total N = 271)	Number of normal control subjects (Total N = 105)	Number of MCI subjects (Total N = 176)	Number of Alzheimer's disease subjects (Total N = 50)	Type of scanner change
GE Genesis Signa To Siemens Avanto	26	7	12	7	Excluded due to nonconvergence; For all subjects, only one scan at baseline was acquired with GE Signa and the rest were acquired with Siemens Avanto
Philips Intera To Siemens Avanto	**15**	**6**	**7**	**2**	**Inter-vendor change; Included**
GE Genesis Signa To Philips Intera	**7**	**1**	**5**	**1**	**Inter-vendor change; Included**
GE Genesis Signa To GE Signa HDx	1		1		Excluded due to nonconvergence
GE Genesis Signa To Siemens Symphony	1		1		Excluded due to nonconvergence
GE Signa Excite To GE Signa HDx	**85**	**15**	**39**	**31**	**Intra-vendor upgrade; Included**
GE Signa Excite To GE Signa HDxt	**24**	**12**	**12**		**Intra-vendor upgrade; Included**
GE Signa Excite To GE Signa HDx To GE Signa HDxt	**60**	**25**	**35**		**Intra-vendor upgrade; two upgrades; Included**
GE Signa HDx To GE Signa HDxt	**12**	**2**	**9**	**1**	**Intra-vendor upgrade; Included**
Siemens Avanto To GE Signa HDxt	1	1			Excluded due to nonconvergence
Siemens Avanto To Siemens SonataVision	1	1			Excluded due to nonconvergence
Siemens Sonata To Siemens Espree	4	1	2	1	Excluded due to nonconvergence
Siemens Symphony To Siemens Symphony TIM	**34**	**9**	**18**	**7**	**Intra-vendor upgrade; Included**

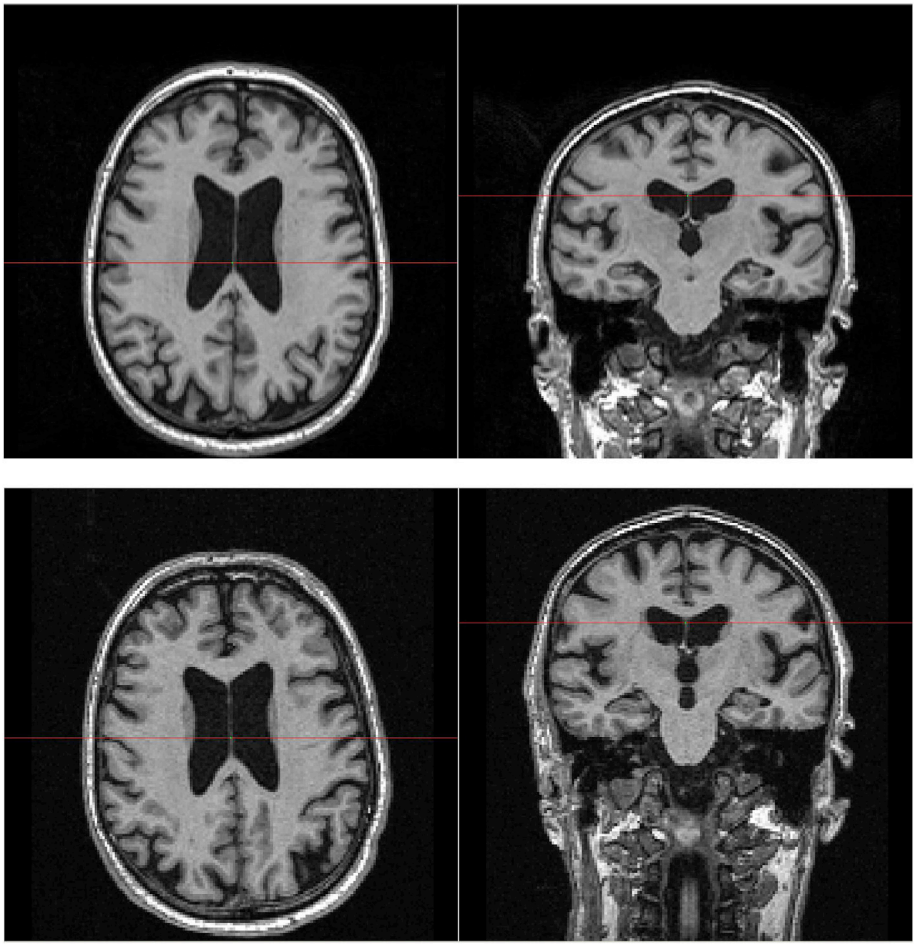

Fig. 1. Example pair of original input images from a single Normal subject (inter-vendor scanner change).
First row: 24 months follow-up, Philips Intera, MP-RAGE. PBVC: −2.4% from the baseline.
Second row: 36 months follow-up, Siemens Avanto, MP-RAGE. PBVC: −4.1% from the baseline.

2013; Nakamura et al., 2015). We considered several previous findings in building our LME model. Notably, in the case of analyzing data from multiple T1-weighted protocols, it was shown to be advantageous to include a categorical fixed-effect term for the different protocols (Jones et al., 2013). Another study tested various forms of LME models in terms of the Akaike Information Criterion (AIC) goodness-of-fit measure, and demonstrated the advantage of including subject-specific random intercepts and slopes with protocol-specific residual variance (Chua et al., 2015). Also, having protocol as a fixed effect led to a better model fit as opposed to having protocol-by-study time interactions (Chua et al., 2015). Finally, a study of N = 713 ADNI subjects found no evidence of acceleration in WB atrophy rates over three years of follow-up (Leung et al., 2013). Therefore, although many of the subjects in our study had more than three years of follow-up, we assumed linear time courses of atrophy.

The estimated PBVCs were modeled with an LME model that included subject-specific random slopes and intercepts, as well as fixed effects for the time from baseline and diagnosis groups, interaction between time and diagnosis group, MRI scanner models, and T1-weighted sequences. Scanner-specific residual variance was used. The model was as follows:

$$\Delta PBV = \beta_0 + \beta_1\,(\text{time}) + \beta_{DxGroup}\,(\text{DiagnosisGroup}) + \beta_{DxGroupRate}\,(\text{DiagnosisGroup} * \text{time}) + \beta_{ScannerChg}\,(\text{ScannerModel}) + \beta_{SequenceChg}\,(\text{T1Sequence}) + b_{oi} + b_{1i}\,(\text{time}) + \varepsilon_{ij}$$

where

“ΔPBV” was a continuous variable for the percentage WB volume change from the baseline “screening” reference point;
“time” was a continuous variable for years from the baseline scan date;
“DiagnosisGroup” was a categorical variable for the diagnosis group, i.e. normal control, MCI, or AD;
“ScannerModel” was a categorical variable for the MRI scanner model changes shown in Table 3;
“T1Sequence” was a categorical variable for the 3D T1-weighted sequence, which was MP-RAGE for all Siemens and Philips scanners and either MP-RAGE or IR-FSPGR for GE scanners;
“β_0”, “β_1”, “$\beta_{DxGroup}$”, “$\beta_{DxGroupRate}$”, “$\beta_{ScannerChg}$”, “$\beta_{SequenceChg}$” were the fixed-effects coefficients for the group intercept, group slope, additive effect of diagnosis group on PBVCs, atrophy rate associated with each diagnosis group, additive effect of MRI scanner change or upgrade on PBVCs, and additive effect of T1 sequence change on PBVCs, respectively;
“b_{oi}” and “b_{1i}” were the subject-specific random intercept and slope, respectively;
“ε_{ij}” was the error term.

This model was fit using the MIXED procedure, SAS v9.4.

To assess whether including the corrective terms for the MRI scanner or the T1-weighted sequence changes leads to better goodness-of-fit, an equivalent model without the corrective terms was also fit and AICs for each model were compared.

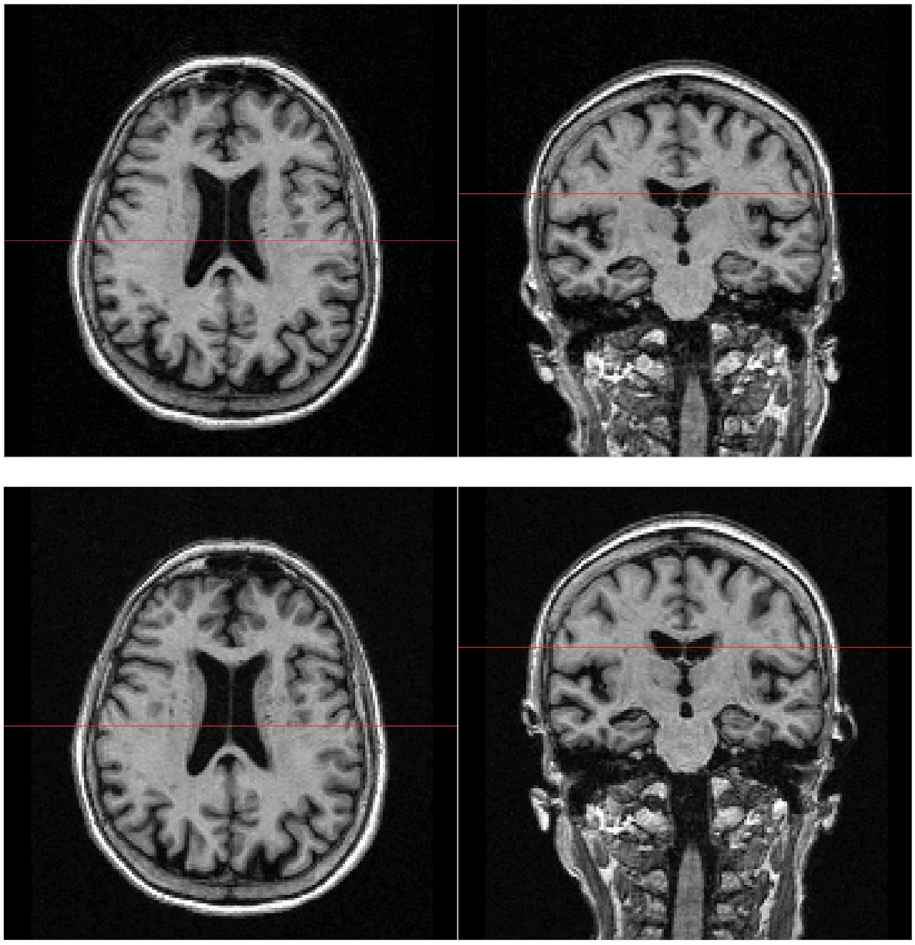

Fig. 2. Example pair of original input images from a single MCI subject (intra-vendor scanner upgrade).
Top row: 12 months follow-up, Siemens Symphony, MP-RAGE. PBVC: −1.4% from the baseline.
Bottom row: 18 months follow-up, Siemens Symphony TIM, MP-RAGE. PBVC: −2.7% from the baseline.

3. Results

The N = 773 subjects were divided into two subgroups: subjects who did not have a scanner upgrade or change during the follow-up ("Chg-", N = 502) versus those who did ("Chg+", N = 271). Overall, the modeled rates of WB PBVC between these two groups were not significantly different, Chg+, −1.15%/y (Standard Error: 0.05) vs. Chg-, −1.16%/y (0.04), p = 0.84, F-test. The rate for the Chg + group estimated without regard to scanner model or sequence was −1.24%/y (0.05).

3.1. Chg- subjects

Model-estimated WB PBVC rates by diagnosis group for the Chg- subjects are shown in Supplementary Table S1. There were no effects of scanner model change or T1 sequence change in this subgroup. The average rates were AD: −1.69%/y (SE: 0.097), MCI: −1.25%/y (0.077), and normal controls: −0.67%/y (0.060). Inline Supplementary Figure S1 plots the average rates for each diagnosis group overlaid on top of the actual PBVC measurement values for each subject.

3.2. Chg + subjects

Effects of the MRI scanner or T1-weighted sequence changes on PBVC measurements were estimated for seven combinations (subject N = 237). The remaining six combinations were excluded due to insufficient numbers of subjects or data points for model convergence; notably, the GE Genesis Signa to Siemens Avanto change occurred right after the baseline time point for everyone. For all the included cases, the model with the corrective terms attained a lower AIC compared to the one without the corrective terms. The same results were obtained with the −2 Res Log Likelihood, AICC, and BIC statistics. Figs. 4 and 5 illustrate the additive effects of scanner upgrade or change on group-average time courses (additional figures are available as Inline Supplementary Figures S2 to S6). Examples of the detailed model outcomes can be found in Tables 4 and 5 (additional tables are available as Supplementary Tables S2 to S6).

3.2.1. Effects of inter-vendor or intra-vendor scanner changes on PBVC

Philips Intera to Siemens Avanto (Fig. 4, Table 4): This inter-vendor scanner change led to an average decrease of −1.81% (0.30) in PBVC, p < 0.0001.

GE Genesis Signa to Philips Intera (Inline Supplementary Figure S2, Supplementary Table S2): This inter-vendor scanner change led to an average increase of 0.99% (SE: 0.47) in PBVC, p = 0.042.

GE Signa Excite to GE Signa HDx (Inline Supplementary Figure S3, Supplementary Table S3): This intra-vendor upgrade led to an average increase of 0.33% (0.095) in PBVC, p = 0.0005.

GE Signa Excite to GE Signa HDxt (Inline Supplementary Figure S4, Supplementary Table S4): This intra-vendor upgrade led to an insignificant decrease of −0.023% (0.23) in PBVC, p = 0.92.

GE Signa Excite to GE Signa HDx to GE Signa HDxt (Fig. 5, Table 5): The intra-vendor upgrade from Signa Excite to Signa HDx led to an average increase of 0.25% (0.095) in PBVC, p = 0.0103. There was a trend of PBVC increase (0.27%, (0.16), p = 0.0978) when calculating

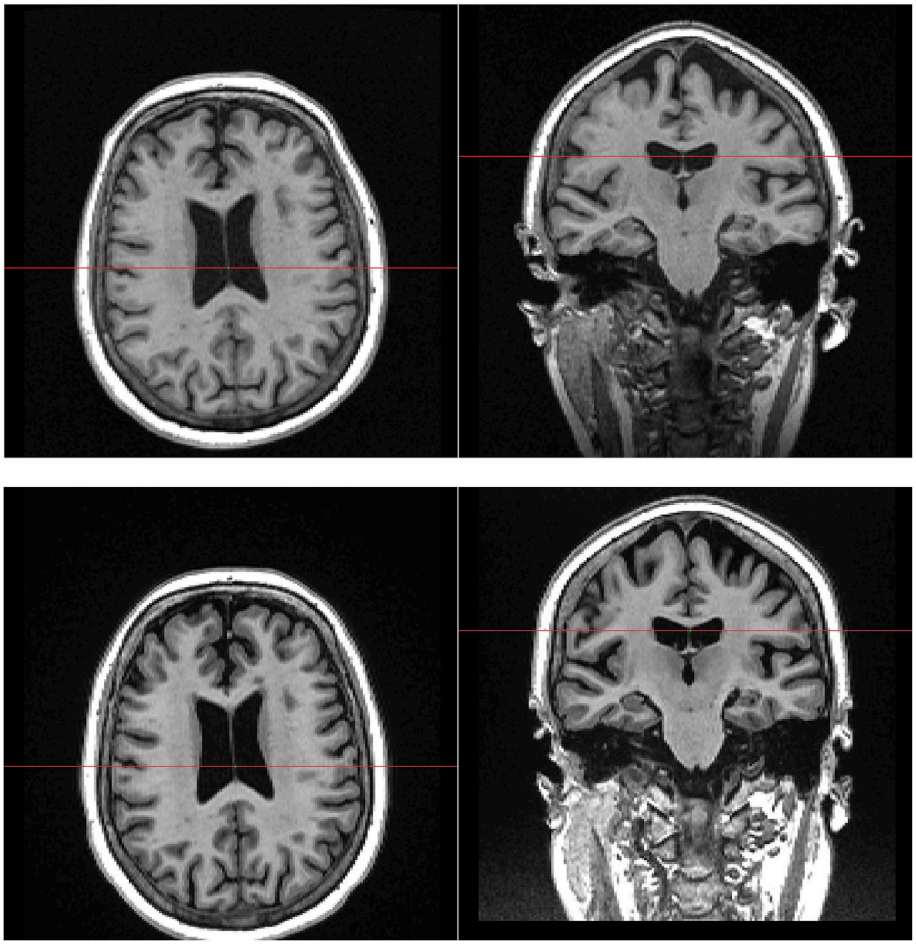

Fig. 3. Example pair of original input images from a single Normal subject (T1-weighted sequence change).
Top row: 48 months follow-up, GE Signa HDxt, MP-RAGE. PBVC: −1.2% from the baseline.
Bottom row: 60 months follow-up, GE Signa HDxt, IR-FSPGR. PBVC: −4.0% from the baseline.

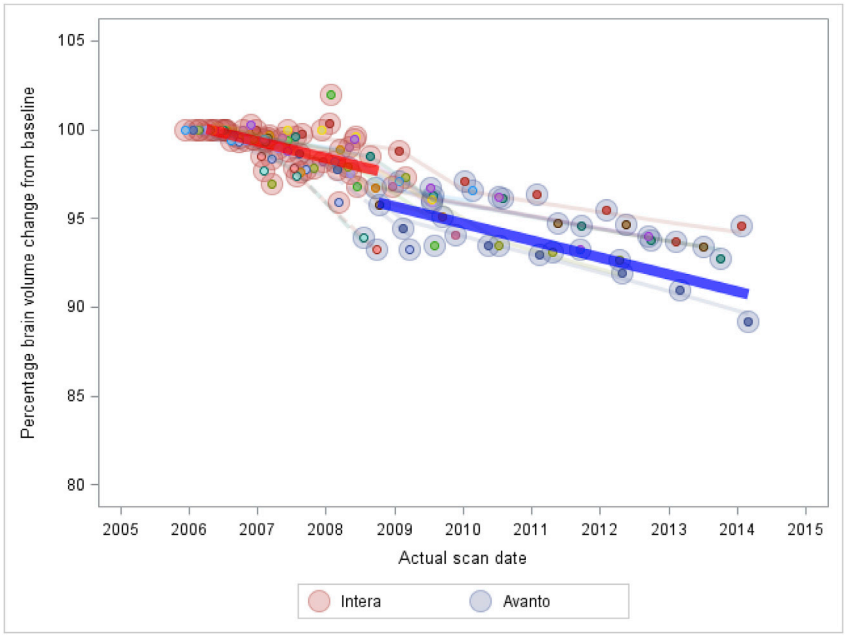

Fig. 4. Philips Intera to Siemens Avanto.
Model-fitted lines, plotted against actual scan dates for better visualization of the scanner change effects.
Colored large circles = Each color denotes the specific MRI scanner used at that timepoint.
Colored thick lines = Average model for all subjects, grouped by MRI scanner used during follow-up. Each color represents each MRI scanner. The line discontinuity represents the scanner upgrade/change effect.
Colored small dots and thin lines = Actual percentage brain volume change measurement values with respect to baseline for each subject, with the thin lines representing the fitted model for each subject. Each color represents each subject. The thin-line discontinuities ('jump downs') at later scan dates on Fig. 5 and Inline Supplementary Figures S3, S4, and S5 (involving Signa HDx or HDxt) represent the effects of T1-weighted sequence change, in addition to the scanner changes.

changes from Signa Excite to Signa HDxt.

GE Signa HDx to GE Signa HDxt (Inline Supplementary Figure S5, Supplementary Table S5): This intra-vendor upgrade led to an insignificant decrease of −0.24% (0.25) in PBVC, p = 0.34.

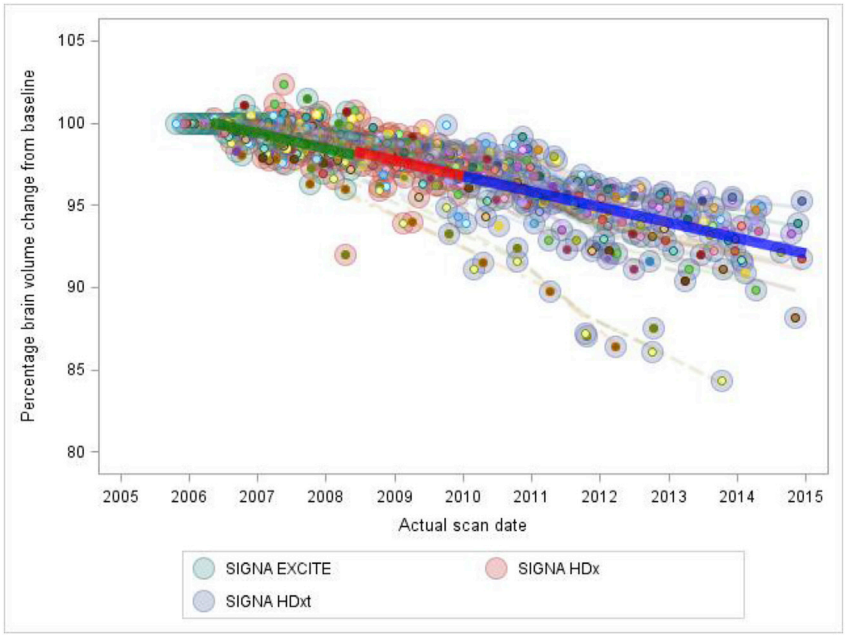

Fig. 5. GE Signa Excite to GE Signa HDx to GE Signa HDxt. Model-fitted lines, plotted against actual scan dates for better visualization of the scanner change effects.
Colored large circles = Each color denotes the specific MRI scanner used at that timepoint.
Colored thick lines = Average model for all subjects, grouped by MRI scanner used during follow-up. Each color represents each MRI scanner. The line discontinuity represents the scanner upgrade/change effect.
Colored small dots and thin lines = Actual percentage brain volume change measurement values with respect to baseline for each subject, with the thin lines representing the fitted model for each subject. Each color represents each subject. The thin-line discontinuities ('jump downs') at later scan dates on Fig. 5 and Inline Supplementary Figures S3, S4, and S5 (involving Signa HDx or HDxt) represent the effects of T1-weighted sequence change, in addition to the scanner changes.

Table 4
Effect of inter-vendor scanner change from Philips Intera to Siemens Avanto (Chg+).

Effect	MRI scanner model	T1-weighted sequence	Diagnosis group	Estimate [95% CI]	Standard error	p-value	Note
Intercept				100.09%	0.19	<.0001	Model AIC: 252.2
Reference T1-weighted sequence		MP-RAGE only		.	.	.	**Model without the corrective terms for scanner and sequence changes:** AIC: 281.3
MRI scanner change	**Philips Intera to Siemens Avanto**			**−1.81% [-2.41, −1.22]**	**0.30**	**<.0001**	Normal: −1.01%/y MCI: −1.41%/y
Reference MRI scanner	Philips Intera			.	.	.	AD: −2.91%/y
Diagnosis Group Effect			Normal	.	.	.	
			MCI	−0.04%	0.28	0.8974 from Normal	
			AD	−0.36%	0.47	0.4370 from Normal	
Estimated percentage brain volume change rates by diagnosis group			**Normal**	**−0.66%/y**	**0.098**	**0.0046 from zero**	
			MCI	**−1.06%/y**	**0.13**	**0.1222 from Normal**	
			AD	**−2.04%/y**	**0.35**	**0.0002 from Normal**	

Siemens Symphony to Siemens Symphony Total Imaging Matrix (TIM) (Inline Supplementary Figure S6, Supplementary Table S6**):** This major intra-vendor upgrade led to an average decrease of −0.39% (0.16) in PBVC, p = 0.0188.

3.2.2. Effects of T1-weighted sequence changes on PBVC

The four intra-vendor upgrade combinations, involving GE Signa HDx or HDxt platforms, were additionally affected by the T1-weighted sequence change that occurred after the corresponding upgrades.

GE Signa Excite to GE Signa HDx (Inline Supplementary Figure S3, Supplementary Table S3): The T1-weighted sequence change from MP-RAGE to IR-FSPGR (occurred on the Signa HDx platform) led to an average decrease of −1.56% (SE: 0.36) in PBVC, p < 0.0001.

GE Signa Excite to GE Signa HDxt (Inline Supplementary Figure S4, Supplementary Table S4**):** The T1-weighted sequence change from MP-RAGE to IR-FSPGR (occurred on the Signa HDxt platform) led to an average decrease of −2.47% (0.33) in PBVC, p < 0.0001.

GE Signa Excite to GE Signa HDx to GE Signa HDxt (Fig. 5, Table 5): The T1-weighted sequence change from MP-RAGE to IR-FSPGR (occurred on the Signa HDxt platform) led to an average decrease of −1.41% (0.15) in PBVC, p < 0.0001.

GE Signa HDx to GE Signa HDxt (Inline Supplementary Figure S5, Supplementary Table S5): The T1-weighted sequence change from MP-RAGE to IR-FSPGR (occurred on the Signa HDxt platform) led to an average decrease of −2.11% (0.32) in PBVC, p < 0.0001.

The average effect T1-weighted sequence change from MP-RAGE to IR-FSPGR was −1.63% (0.12) in PBVC, p < 0.0001 (Fig. 6). The model with the corrective terms for scanner and sequence changes provided a lower AIC (3940.0) compared to the model without the corrective terms (AIC: 4278.3). Also, the average PBVC rates in AD, MCI, and normal controls differed between these two models by 1%, 7%, and 17%, respectively.

The seven combinations analyzed above comprised N = 237 Chg + subjects. Model-based group-average PBVC rates by diagnosis group for these subjects, adjusted for scanner and sequence changes, were AD: −1.97%/y (0.13), MCI: −1.20%/y (0.090), and normal

Table 5
Effects of two intra-vendor scanner upgrades from GE Signa Excite to GE Signa HDx to GE Signa HDxt (Chg+).

Effect	MRI scanner model	T1-weighted sequence	Diagnosis group	Estimate [95% CI]	Standard error	p-value	Note
Intercept				99.95%	0.098	<.0001	Model AIC: 1228.7
T1-weighted sequence change		**IR-FSPGR**		**−1.41% [-1.69, −1.12]**	**0.15**	**<.0001**	**Model without the corrective terms for scanner and sequence changes:**
Reference T1-weighted sequence		MP-RAGE		.	.	.	AIC: 1354.9 Normal: −0.90%/y MCI: −1.20%/y
Second MRI scanner upgrade	**GE Signa Excite to GE Signa HDxt**			**0.27% [-0.05, 0.60]**	**0.16**	**0.0978**	
First MRI scanner upgrade	**GE Signa Excite to GE Signa HDx**			**0.25% [0.06, 0.43]**	**0.095**	**0.0103**	
Reference MRI scanner	GE Signa Excite			.	.	.	
Diagnosis Group			Normal	.	.		
Effect			MCI	0.13%	0.13	0.3165 from Normal	
Estimated percentage brain volume change rates by diagnosis group			**Normal**	**−0.76%/y**	**0.089**	**<.0001 from zero**	
			MCI	**−1.09%/y**	**0.11**	**0.0040 from Normal**	

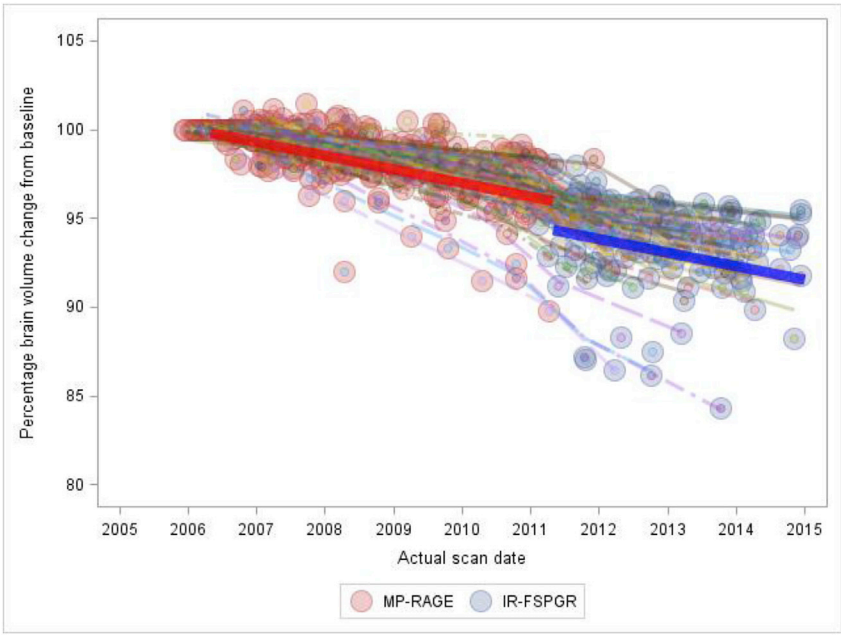

Fig. 6. Illustration of the effect of T1-weighted sequence change from MP-RAGE to IR-FSPGR.
Model-fitted lines, plotted against actual scan dates for better visualization of the sequence change effects.
Colored large circles = Each color denotes the specific T1-weighted sequence used at that timepoint.
Colored thick lines = Average model for all subjects, grouped by T1-weighted sequence. Each color represents each sequence. The line discontinuity represents the sequence change effect.
Colored small dots and thin lines = Actual percentage brain volume change measurement values with respect to baseline for each subject, with the thin lines representing the fitted model for each subject. Each color represents each subject.

controls: −0.68%/y (0.073). Detailed model outcomes are shown in Supplementary Table S7.

4. Discussion

We surveyed 819 normal control, MCI, and AD subjects enrolled in the ADNI-1 1.5 T study and identified those who had an MRI scanner upgrade or change during follow-up. Longitudinal PBVCs, with respect to the baseline, were measured from serial MRIs that reached up to 8 years of follow-up. An LME model was applied to model the time courses of WB PBVC while estimating the effects of inter-vendor scanner change, intra-vendor scanner upgrade, and T1-weighted sequence change from MP-RAGE to IR-FSPGR (subset of GE scanners only). The change of sequence from MP-RAGE to IR-FSPGR was associated with an average of −1.63% change in PBVC. Artifactual changes in PBVC were found across different scanner hardware upgrade or change combinations. Inclusion of the corrective terms for scanner and sequence changes always led to a better model fit (i.e. lower AIC).

We first modeled the time courses of WB PBVC in the Chg-subgroup to explore rates unaffected by changes in the scanning hardware or T1-weighted sequence. The average rate of PBVC in AD patients was 1.35x higher than those of MCI patients and 2.52x higher than those of normal controls (Supplementary Table S1). This is in line with a previous study by Leung and colleagues, in which they reported LME model-estimated WB volume change rates over 3-years follow-up in the ADNI subjects (Leung et al., 2013). Their model did not include terms for scanner or sequence, but these factors did not affect our Chg-group. Using a version of boundary-shift integral (KN-BSI), they found that the average rate in AD was 1.40x higher than those of MCI patients and 2.24x higher than those of normal controls. The actual rates differed due to the fact that SIENA tends to systematically give about 20% larger values compared to BSI (Smith et al., 2007).

The Chg + subgroup comprised N = 237 subjects in which each was followed-up with one of the seven 1.5 T MRI scanner model combinations. These seven combinations could be broadly classified into three categories: inter-vendor scanner change (GE to Philips; Philips to

Siemens), intra-vendor scanner upgrade (GE to GE; Siemens to Siemens), and T1-weighted sequence change (GE to GE). Presuming greater degrees of changes in hardware configuration lead to a larger effect on image characteristics, we hypothesized that inter-vendor scanner changes would induce greater bias on PBVC measurements compared to intra-vendor scanner upgrades. Furthermore, we hypothesized that the change in T1-weighted sequence from MP-RAGE to IR-FSPGR would have had a direct impact on image contrast and thus the PBVC measurements, even though the change occurred while on consistent MRI hardware platforms (GE Signa HDx or HDxt). These indeed was the case in our study.

We analyzed two cases of inter-vendor scanner change combinations (Table 3): GE Genesis Signa to Philips Intera, and Philips Intera to Siemens Avanto. Both cases represented a complete change in the hardware, including the main magnet and the coil. The average scanner change effects of +0.99% (Signa to Intera, $p = 0.042$) and −1.81% (Intera to Avanto, $p < 0.0001$) were significant and were roughly equivalent to a year's worth of volume loss in MCI and AD, respectively. As demonstrated in Fig. 1, contrast differences between images of the same subjects scanned on these different scanners from different vendors were subtle but present. When inter-vendor scanner changes occur, the bias due to the scanner change may exceed the magnitude of the main effect of interest, depending on the study.

There were five cases of intra-vendor scanner upgrade combinations (Table 3): GE Signa Excite to HDx, GE Signa Excite to HDxt, GE Signa HDx to HDxt, GE Signa Excite to HDx to HDxt, and Siemens Symphony to Symphony TIM. The upgrade from Excite to HDx included both hardware (e.g. receive chain architecture) and software components, whereas that from HDx to HDxt was mainly software-related. Upgrading from Excite to HDx exerted a significant effect on PBVC (+0.33%, $p = 0.0005$), whereas going from Excite to HDxt did not (−0.023%, $p = 0.92$). The average effect of a minor upgrade from HDx to HDxt was not significant (−0.24%, $p = 0.34$). A similar pattern was observed in the group of subjects who had two upgrades from Excite to HDx to HDxt; upgrading from Excite to HDx led to a significant effect (+0.25%, $p = 0.0103$) whereas going from Excite to HDxt did not (+0.27%, $p = 0.0978$). Overall, we found no evidence that the software-related upgrade from HDx to HDxt led to a significant systematic bias on PBVC.

The Siemens TIM upgrade was a major hardware change that affected the gradient system, radiofrequency coil, and software. This upgrade led to a significant effect (−0.39%, $p = 0.0188$) on PBVC, comparable to half a year's worth of normal aging in this group, and about 20% of the annual change in this population of AD subjects. Direct hardware changes led to effects with both positive and negative directions.

Intra-vendor upgrades are the most common scenarios that occur in longitudinal studies, and our results suggest that, although the effects may differ from one upgrade to another, they may be ignorable if the rates of brain volume loss of interest are relatively large, such as in AD. Understanding the exact circumstances in which these effects are best ignored is not trivial, as it depends not only on the magnitude of the effect of interest, but also on study design factors such as the number of subjects affected by the change in comparison to the total number of subjects in the study.

The transition from ADNI-1 to ADNI-2/GO protocols included a change in the 3D T1-weighted sequence from MP-RAGE to IR-FSPGR in $N = 63$ subjects scanned on select 1.5 T GE HDx and HDxt platforms ($N = 4$ HDx and $N = 59$ HDxt). Briefly, the GE T1-weighted sequence that most closely resembled the ADNI MP-RAGE sequence was the GE product IR-FSPGR sequence, which was not available until July 2006 (https://adni.loni.usc.edu/wp-content/themes/freshnews-dev-v2/documents/mri/ADNI_MRI_Methods_Non-ADNI_Studies.pdf). Until the implementation of this sequence, the GE sites needed to use the earlier "works-in-progress" version of the ADNI MP-RAGE that focused on maximizing inter-vendor protocol standardization, but at the expense of replicability of the exact ADNI methods on other GE scanners. For the ADNI-2/GO phases, a complete switch was made to the GE product IR-FSPGR sequence (Jack et al., 2010). This would have resulted in an alteration of the SNR and CNR (Lin et al., 2006). Fig. 3 shows example images from a single subject who had scans available from both sequences. Indeed, the T1-sequence change effect was significant in all cases with the average effect of −1.63% on PBVC, $p < 0.0001$, estimated from the model that included all Chg + subjects (Fig. 6). This type of sequence change was specific to the ADNI study planning process and is unlikely to occur in a typical prospective longitudinal study. However, such changes may affect retrospective studies or studies in which standardized acquisitions are not performed. Our results suggest that the effect of sequence changes, even within the same hardware platform, may even exceed that observed in inter-vendor scanner changes and produce significant errors in brain volume measurements. Yet, it should be noted that our findings are specific to the MP-RAGE to IR-FSPGR sequence change on GE platforms, for which data were available. In general, the impact of a sequence change on PBVC would depend on the specific differences in the sequences and would need to be evaluated for each combination observed in practice.

It is possible that the step changes in PBVC resulted from a change in the brain-CSF boundary delineation after the scanner or sequence change, since the measurement of brain volume changes over time generally depends on the detection of edge motion between registered scans. For example, SIENA uses the derivative of the gradient across the brain-CSF boundary to estimate the brain/non-brain edge motion between two timepoints and converts the mean edge displacement into the PBVC value (Smith et al., 2002). An example scenario would be an improved boundary delineation (e.g. due to increased contrast or a reduction in the brain-CSF partial volume effects) which may result in an apparently reduced WB volume; in this case, the pre-upgrade volume would have been an overestimation due to the partial CSF volume being included in the brain volume.

Our observation has important implications for longitudinal studies of brain volume change. In all cases, models with corrective terms for scanner and sequence changes yielded a lower AIC compared to those without, despite the two additional parameters being included in the model. This suggests that the atrophy rates estimated using the model with the corrective terms better represent the data, and that the rates from models without the corrective terms may be over- or under-estimated, depending on the direction of the effect of the change. For example, the average model-estimated rates in the subjects who had the intra-vendor upgrade from Symphony to Symphony TIM were 10%, 6%, and 11% different for AD, MCI and normal respectively, compared to the rates estimated ignoring the changes. In the subjects who switched from Philips Intera to Siemens Avanto, these differences were 35%, 28%, and 42% for AD, MCI and normal respectively. It should be noted that the effect of scanning platform change is an aggregation of effects from various sources that contribute to image inconsistency, such as inconsistent image contrast, gradient nonlinearity, geometric distortion, intensity nonuniformity, and subject positioning. In principle, they may be corrected using site-specific living phantoms to estimate the scanner change effect and calibrate the measurements. However, this is generally not feasible and is infrequently done in practice. Whenever any scanner upgrade or changes occur, the researcher needs to identify the potential source(s) of image inconsistency and incorporate image processing schemes that may help alleviate the inconsistency. Yet, it should be recognized that the harmonized acquisition protocol and image processing steps alone may not be sufficient to remove the image variability originating from changes in scanning platforms. For example, a recent work from the North American Imaging in Multiple Sclerosis (NAIMS) Cooperative, which implemented a standardized protocol on 3T Siemens scanners, has reported significant scanner model-related variations in T1-hypointense and T2-hyperintense lesions as well as brain volume measures (Shinohara et al., 2017). Other works from the Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) dataset has reported significant site/scanner effects on cortical thickness as well as functional connectivity measures (Fortin

et al., 2018; Yu et al., 2018). These results suggest that unwanted scanner-related non-biological variability is pervasive in many areas of brain imaging research. Additional steps, such as incorporating corrective terms into the statistical analysis model, may be necessary to attain a more reliable outcome. However, the modeling requires that there be sufficient numbers of subjects affected by any modeled change to be able to estimate the effect. This can be a challenge in some multi-center drug trials with many sites and not many subjects per site.

Study limitations and future directions: There were several limitations to our study. First, the ADNI study provided a valuable large-scale, multi-site dataset acquired using a standardized protocol and quality control. We surveyed the data 'as-is' and could analyze seven different scanner upgrade/change combinations. This approach to subject selection inherently resulted in an unbalanced design. This issue was partially alleviated by the use of the LME model, which can accommodate unbalanced data (Fitzmaurice and Ravichandran, 2008). Second, the average age at baseline was around 75 for our subjects, and this is when the rate of normal aging-related WB atrophy begins to accelerate (Scahill et al., 2003). We kept the model as parsimonious as possible and did not take the potentially non-linear pattern of WB atrophy into account. Although this effect may not be apparent over the short-term (Leung et al., 2013), it could have affected subjects who had 8 years of follow-up. Third, it is unknown whether the specific estimates of the effects for different scanner changes obtained from this study are generalizable, as various designs exist with regards to image acquisition, processing, and analysis pipeline. Yet, our study pipeline can be fairly easily replicated. The ADNI acquisition protocol is widely available and is increasingly being used in clinical trials, and our pre-processing steps and the SIENA method also have been commonly used. Whether our specific estimates apply to different image processing pipelines or atrophy measurement techniques (e.g. BSI, Jacobian Integration) needs to be further explored. Fourth, our study analyzed only 1.5 T scans, but 3.0 T systems are rapidly being adopted. In fact, newly enrolled subjects in the ADNI-2/GO have been entirely scanned at 3.0 T (Jack et al., 2010). Potential effects of 3.0 T scanner change/upgrade need to be investigated. Finally, there is a growing interest in measuring grey- and white-matter volumes separately, as grey-matter atrophy may be better correlated with disability progression and cognitive impairment than WB atrophy (Fisher et al., 2008). The measurement of grey-matter change itself is technically challenging, and any scanner upgrade or change can add further complexity to the analysis. Kruggel and colleagues demonstrated significant within-subject variability of grey- and white-matter compartmental volumes (and thus, WB volume also) on different 1.5 T and 3.0 T scanners used in the ADNI (Kruggel et al., 2010). Moreover, Nakamura and colleagues revealed that the presence of white-matter lesions can significantly bias grey- and white-matter segmentation (Nakamura and Fisher, 2009). Further research on this important topic is warranted.

In conclusion, we demonstrated that different scanner hardware upgrades can exert different bias effects on WB PBVC. Inter-vendor scanner changes generally led to greater effects compared to intra-vendor scanner upgrades. Change in the 3D T1-weighted sequence from MP-RAGE to IR-FSPGR, within the same scanning platform, also led to a significant effect, comparable to that from inter-vendor scanner changes. Modeling brain volume loss with an LME model that includes corrective terms for scanner and sequence changes yields better model fits and more reliable estimates of WB atrophy rates.

Declaration of interest

Dr. Hyunwoo Lee reports no disclosures. Dr. Kunio Nakamura reports research grants from NIH, DOD, NMSS, Biogen, Sanofi Genzyme, and Novartis, and personal fees from Sanofi Genzyme (speaking), NeuroRx (consulting), and Biogen (license royalties). Dr. Sridar Narayanan reports research grants from the Canadian Institutes of Health Research and the Myelin Repair Foundation, and personal fees from NeuroRx Research and a speaker's honorarium from Novartis Canada, unrelated to the submitted work. Dr. Robert Brown reports consultation fees from NeuroRx Research and Biogen. Dr. Douglas Arnold reports consultant fees and/or grants from Acorda, Adelphi, Alkermes, Biogen, Celgene, Frequency Therapeutics, Genentech, Genzyme, Hoffman LaRoche, Immune Tolerance Network, Immunotec, MedDay Merck-Serono, Novartis, Pfizer, Receptos, Roche, Sanofi-Aventis, Canadian Institutes of Health Research, MS Society of Canada, International Progressive MS Alliance, and an equity interest in NeuroRx Research.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Hyunwoo Lee was supported by the Doctoral Training Award from Fonds de Recherché du Quebec - Santé.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2018.09.062.

References

Cannon, T.D., Sun, F., McEwen, S.J., Papademetris, X., He, G., van Erp, T.G.M., Jacobson, A., Bearden, C.E., Walker, E., Hu, X., Zhou, L., Seidman, L.J., Thermenos, H.W., Cornblatt, B., Olvet, D.M., Perkins, D., Belger, A., Cadenhead, K., Tsuang, M., Mirzakhanian, H., Addington, J., Frayne, R., Woods, S.W., McGlashan, T.H., Constable, R.T., Qiu, M., Mathalon, D.H., Thompson, P., Toga, A.W., 2014. Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis. Hum. Brain Mapp. 35, 2424–2434. https://doi.org/10.1002/hbm.22338.

Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, G.B., Collins, D.L., Arnold, D.L., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. Neuroimage 49, 1601–1611. https://doi.org/10.1016/j.neuroimage.2009.08.008.

Chua, A.S., Egorova, S., Anderson, M.C., Polgar-Turcsanyi, M., Chitnis, T., Weiner, H.L., Guttmann, C.R.G., Bakshi, R., Healy, B.C., 2015. Handling changes in MRI acquisition parameters in modeling whole brain lesion volume and atrophy data in multiple sclerosis subjects: comparison of linear mixed-effect models. NeuroImage Clin 8, 606–610. https://doi.org/10.1016/j.nicl.2015.06.009.

Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Østergaard, L.R., Collins, D.L., Alzheimer's Disease Neuroimaging Initiative, 2012. BEaST: brain extraction based on nonlocal segmentation technique. Neuroimage 59, 2362–2373. https://doi.org/10.1016/j.neuroimage.2011.09.012.

Evans, M.C., Barnes, J., Nielsen, C., Kim, L.G., Clegg, S.L., Blair, M., Leung, K.K., Douiri, A., Boyes, R.G., Ourselin, S., Fox, N.C., Alzheimer's Disease Neuroimaging Initiative, 2010. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. Eur. Radiol. 20, 674–682. https://doi.org/10.1007/s00330-009-1581-5.

Fisher, E., Lee, J.-C., Nakamura, K., Rudick, R.A., 2008. Gray matter atrophy in multiple sclerosis: a longitudinal study. Ann. Neurol. 64, 255–265. https://doi.org/10.1002/ana.21436.

Fitzmaurice, G.M., Ravichandran, C., 2008. A primer in longitudinal data analysis. Circulation 118, 2005–10. https://doi.org/10.1161/CIRCULATIONAHA.107.714618.

Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 47, S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 167, 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024.

Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. 6, 67–77. https://doi.org/10.1038/nrneurol.2009.215.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32, 180–194. https://doi.org/10.1016/j.neuroimage.2006.02.051.

Jack, C.R., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., Desikan, R.S., Fennema-Notestine, C., Fjell, A.M., Fletcher, E., Fox, N.C., Gunter, J., Gutman, B.A., Holland, D., Hua, X., Insel, P., Kantarci, K., Killiany, R.J., Krueger, G., Leung, K.K., Mackin, S., Maillard, P., Malone, I.B., Mattsson, N., McEvoy, L., Modat, M., Mueller, S., Nosheny, R., Ourselin, S., Schuff, N., Senjem, M.L., Simonson, A., Thompson, P.M., Rettmann, D., Vemuri, P., Walhovd, K., Zhao, Y., Zuk, S., Weiner, M., 2015. Magnetic resonance imaging in Alzheimer's disease neuroimaging initiative 2. Alzheimers. Dementia 11, 740–756. https://doi.org/10.1016/j.jalz.2015.05.002.

Jack, C.R., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., Killiany, R.J., Decarli, C.S., Dale, A.M., Carmichael, O.W., Tosun, D., Weiner, M.W., Alzheimer's Disease Neuroimaging Initiative, 2010. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimers. Dement. 6, 212–220. https://doi.org/10.1016/j.jalz.2010.03.004.

Jack, C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L.M., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Pankratz, V.S., Donohue, M.C., Trojanowski, J.Q., 2013. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 12, 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0.

Jones, B.C., Nair, G., Shea, C.D., Crainiceanu, C.M., Cortese, I.C.M., Reich, D.S., 2013. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. NeuroImage Clin 3, 171–179. https://doi.org/10.1016/j.nicl.2013.08.001.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46, 177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010.

Krueger, G., Granziera, C., Jack, C.R., Gunter, J.L., Littmann, A., Mortamet, B., Kannengiesser, S., Sorensen, A.G., Ward, C.P., Reyes, D.A., Britson, P.J., Fischer, H., Bernstein, M.A., 2012. Effects of MRI scan acceleration on brain volume measurement consistency. J. Magn. Reson. Imag. 36, 1234–1240. https://doi.org/10.1002/jmri.23694.

Kruggel, F., Turner, J., Muftuler, L.T., Initiative, A.D.N., 2010. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49, 2123–2133. https://doi.org/10.1016/j.neuroimage.2009.11.006.

Leung, K.K., Bartlett, J.W., Barnes, J., Manning, E.N., Ourselin, S., Fox, N.C., 2013. Cerebral atrophy in mild cognitive impairment and Alzheimer disease: rates and acceleration. Neurology 80, 648–654. https://doi.org/10.1212/WNL.0b013e318281ccd3.

Lin, C., Watson, R.E., Ward, H.A., Rydberg, C.H., Witte, R.J., 2006. MP-RAGE compared to 3D IR SPGR for optimal T1 contrast and image quality in the brain at 3T. Proc. Intl. Sco. Mag. Reson. Med. 981.

Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. 15, 869–877 xi–xii. https://doi.org/10.1016/j.nic.2005.09.008.

Nakamura, K., Brown, R.A., Narayanan, S., Collins, D.L., Arnold, D.L., Alzheimer's Disease Neuroimaging Initiative, 2015. Diurnal fluctuations in brain volume: statistical analyses of MRI from large populations. Neuroimage 118, 126–132. https://doi.org/10.1016/j.neuroimage.2015.05.077.

Nakamura, K., Fisher, E., 2009. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. Neuroimage 44, 769–776. https://doi.org/10.1016/j.neuroimage.2008.09.059.

Nakamura, K., Guizard, N., Fonov, V.S., Narayanan, S., Collins, D.L., Arnold, D.L., 2014. MRI-based simulation of central brain atrophy for evaluation of brain atrophy measurement methods. In: Proceedings of the International Society for Magnetic Resonance in Medicine, p. 4314.

Popescu, V., Schoonheim, M.M., Versteeg, A., Chaturvedi, N., Jonker, M., Xavier de Menezes, R., Gallindo Garre, F., Uitdehaag, B.M.J., Barkhof, F., Vrenken, H., 2016. Grey matter atrophy in multiple sclerosis: clinical interpretation depends on choice of analysis method. PLoS One 11 e0143942. https://doi.org/10.1371/journal.pone.0143942.

Preboske, G.M., Gunter, J.L., Ward, C.P., Jack, C.R., 2006. Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. Neuroimage 30, 1196–1202. https://doi.org/10.1016/j.neuroimage.2005.10.049.

Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C., 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch. Neurol. 60, 989–994. https://doi.org/10.1001/archneur.60.7.989.

Shinohara, R.T., Oh, J., Nair, G., Calabresi, P.A., Davatzikos, C., Doshi, J., Henry, R.G., Kim, G., Linn, K.A., Papinutto, N., Pelletier, D., Pham, D.L., Reich, D.S., Rooney, W., Roy, S., Stern, W., Tummala, S., Yousuf, F., Zhu, A., Sicotte, N.L., Bakshi, R., Cooperative, N.A.I.M.S., 2017. Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. AJNR. Am. J. Neuroradiol. 38, 1501–1509. https://doi.org/10.3174/ajnr.A5254.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imag. 17, 87–97. https://doi.org/10.1109/42.668698.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl. 1), S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Smith, S.M., Rao, A., De Stefano, N., Jenkinson, M., Schott, J.M., Matthews, P.M., Fox, N.C., 2007. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. Neuroimage 36, 1200–1206. https://doi.org/10.1016/j.neuroimage.2007.04.035.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17, 479–489. https://doi.org/10.1006/nimg.2002.1040.

Takao, H., Abe, O., Hayashi, N., Kabasawa, H., Ohtomo, K., 2010. Effects of gradient non-linearity correction and intensity non-uniformity correction in longitudinal studies using structural image evaluation using normalization of atrophy (SIENA). J. Magn. Reson. Imag. 32, 489–492. https://doi.org/10.1002/jmri.22237.

Wyman, B.T., Harvey, D.J., Crawford, K., Bernstein, M.A., Carmichael, O., Cole, P.E., Crane, P.K., DeCarli, C., Fox, N.C., Gunter, J.L., Hill, D., Killiany, R.J., Pachai, C., Schwarz, A.J., Schuff, N., Senjem, M.L., Suhy, J., Thompson, P.M., Weiner, M., Jack, C.R., Alzheimer's Disease Neuroimaging Initiative, 2013. Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimers. Dement. 9, 332–337. https://doi.org/10.1016/j.jalz.2012.06.004.

Yu, M., Linn, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., Sheline, Y.I., 2018. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Hum. Brain Mapp. https://doi.org/10.1002/hbm.24241.