



Published in final edited form as:

Neuroimage. 2008 February 15; 39(4): 1752–1762. doi:10.1016/j.neuroimage.2007.10.026.

Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils

Richard G. Boyes^{a,*}, Jeff L. Gunter^b, Chris Frost^{a,c}, Andrew L. Janke^d, Thomas Yeatman^a, Derek L.G. Hill^e, Matt A. Bernstein^b, Paul M. Thompson^f, Michael W. Weiner^{g,h}, Norbert Schuff^g, Gene E. Alexanderⁱ, Ronald J. Killiany^j, Charles DeCarli^k, Clifford R. Jack^b, and Nick C. Fox [on behalf of for the ADNI Study]^{a,1}

^aDementia Research Centre, Institute of Neurology, Box 16, University College London, Queen Square, London, WC1N 3BG, UK

^bMayo Clinic College of Medicine, Rochester, MN 55905, USA

^cLondon School of Hygiene and Tropical Medicine, University College London, UK

^dMontreal Neurological Institute, McGill University, Quebec, Canada

^eCentre for Medical Image Computing, University College London, UK

^fLaboratory of Neuro Imaging, Brain Mapping Division, Department of Neurology, UCLA School of Medicine, Los Angeles, CA 90095-7332, USA

^gDepartment of Radiology, UC San Francisco, San Francisco, CA 94143, USA

^hDepartment of Medicine and Psychiatry, UC San Francisco, San Francisco, CA 94143, USA

ⁱDepartment of Psychology and Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ 85721, USA

^jBoston University School of Medicine, Boston, MA 02118, USA

^kDepartment of Neurology and Center for Neuroscience, UC Davis, Sacramento, CA 95817, USA

Abstract

Measures of structural brain change based on longitudinal MR imaging are increasingly important but can be degraded by intensity non-uniformity. This non-uniformity can be more pronounced at higher field strengths, or when using multichannel receiver coils. We assessed the ability of the non-parametric non-uniform intensity normalization (N3) technique to correct non-uniformity in 72 volumetric brain MR scans from the preparatory phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI). Normal elderly subjects ($n = 18$) were scanned on different 3-T scanners with a multichannel phased array receiver coil at baseline, using magnetization prepared rapid gradient echo (MP-RAGE) and spoiled gradient echo (SPGR) pulse sequences, and again 2 weeks later.

When applying N3, we used five brain masks of varying accuracy and four spline smoothing distances ($d = 50, 100, 150$ and 200 mm) to ascertain which combination of parameters optimally reduces the non-uniformity. We used the normalized white matter intensity variance (standard deviation/mean) to ascertain quantitatively the correction for a single scan; we used the variance of the normalized difference image to assess quantitatively the consistency of the correction over time from registered scan pairs.

*Corresponding author. Fax: +44 0 207 676 2066. E-mail address: r.boyes@ucl.ac.uk (R.G. Boyes).

¹Additional contributors to the ADNI Study are listed at http://www.loni.ucla.edu/ADNI/Data/ADNI_Authorship_List.pdf.

Our results showed statistically significant ($p < 0.01$) improvement in uniformity for individual scans and reduction in the normalized difference image variance when using masks that identified distinct brain tissue classes, and when using smaller spline smoothing distances (e.g., 50-100 mm) for both MP-RAGE and SPGR pulse sequences. These optimized settings may assist future large-scale studies where 3-T scanners and phased array receiver coils are used, such as ADNI, so that intensity non-uniformity does not influence the power of MR imaging to detect disease progression and the factors that influence it.

Introduction

MRI is central in the diagnosis of neurodegenerative diseases as the high spatial resolution and excellent tissue contrast of structural MRI allows detection of tissue change caused by these diseases. Serial scanning aids this process, allowing detection of change over time, and is increasingly used to monitor disease progression and effects of therapeutic intervention. Many techniques have been developed to quantify disease progression from serial MRI in neurological disorders (Freeborough and Fox, 1997; Smith et al., 2001; Ashburner et al., 2003; Wang et al., 2002; Jack et al., 2004; Janke et al., 2001; Thompson et al., 2004). These techniques, however, rely on the consistency of signal intensity from one scan to the next.

Consistency of signal intensity can be affected to varying degrees by a common artefact known as MR intensity non-uniformity (or simply non-uniformity), also alternately referred to as bias, inhomogeneity, or shading artefact. It is manifested as a smooth spatially varying signal intensity across an MR image. Many factors contribute to it, including inhomogeneous radio-frequency (RF) fields (caused by distortion of the RF field by the object being scanned, or non-uniformity of the transmission field), inhomogeneous reception sensitivity and electromagnetic interaction with the object being scanned. A review of the causes of non-uniformity is discussed in Belaroussi et al. (2006). Non-uniformity does not necessarily impact on a radiologist's diagnosis, but it can affect computational analysis of the image due to the variance in signal intensity. Many segmentation, registration and analysis algorithms make assumptions about voxel intensities, and their results are frequently sensitive to non-uniformity. In cases where non-uniformity is severe the results from analysis may become meaningless.

Many postprocessing correction techniques have been proposed to reduce the impact of non-uniformity on brain MR scans, usually based on the assumption that the non-uniformity is a spatially varying multiplicative field and/or that the intensity within the tissue classes of cerebrospinal fluid (CSF), grey matter (GM) and white matter (WM) should be uniform. Some approaches define points or regions of known tissue type and then fit a field to those points and extrapolate it to the rest of the scan (Tincher et al., 1993). Other approaches derive the non-uniformity by modeling tissue to calculate the expected intensities and then estimating the non-uniformity by calculating the difference between the actual and expected voxel values. This is usually achieved using an expectation maximization technique (Guillemaud and Brady, 1997; Wells et al., 1996; Zhang et al., 2001). Alternatively, or additionally, serially acquired scans can be registered together and subtracted, giving a differential bias field that can be used to approximate the non-uniformity (or, in the case of additional correction, the residual non-uniformity) of each scan (Lewis and Fox, 2004). Recent reviews of non-uniformity correction techniques can be found in Belaroussi et al. (2006) and Vovk et al. (2007).

Non-parametric non-uniform intensity normalization (N3) (Sled et al., 1998) is a non-uniformity correction technique that finds a multiplicative field that maximizes the frequency content of the intensity distribution of the corrected scan. N3 proceeds by estimating a Gaussian distribution of the true scan intensities (uncorrupted by non-uniformity) by deconvolution and then uses this distribution and the original scan to obtain an estimate of the non-uniformity

field. This field is then smoothed by fitting a cubic B-spline intensity field to the estimate using a selected basis point distance. This smoothed estimate is then removed from the original scan, and the process begins again, iterating until the non-uniformity estimate has converged.

In a comparison of several correction techniques N3 performed well (Arnold et al., 2001). Also, the algorithm and software are in the public domain (<http://www.bic.mni.mcgill.ca/software/N3/>) and is probably the most widely used non-uniformity correction technique in neurological imaging.

In this study we corrected non-uniformity in images obtained from 3-T scanners using N3. N3 was initially developed for correcting 1.5-T data as 3-T scanners were not commonplace at the time, and thus it is not widely known how well it corrects non-uniformity in higher field scanners and a new generation of receiver coils. 3-T and higher field scanners offer an increased signal to noise ratio resulting in greater anatomical detail and are becoming increasingly available in clinical settings; however artefacts such as non-uniformity may be more pronounced due to the shorter RF wavelength compared to scanning at 1.5 T. This is commonly known as a “dielectric resonance” artefact, but recently that term has been discouraged in favor of “central brightening artefact” (Collins et al., 2005). A review of this and other artefacts commonly observed at 3 T is provided in Bernstein et al. (2006). Likewise, newer generation phased array receiver coils can provide increased signal to noise but may have problems with non-uniformity due the smaller size of their elements and proximity to the scalp, as well as interaction in reception sensitivity between the individual elements. Therefore it is paramount that problems with non-uniformity in newer generation scanners be addressed, so that longitudinal imaging can benefit from the advantages of higher field scanners and more modern coils.

The accuracy and performance of N3 are affected by two key factors: firstly the choice of mask used to identify the region of the scan over which N3 works and secondly the estimate of the distance over which the non-uniformity field varies, which will in general be smaller at 3 T due to the shorter RF wavelength. We wished to ascertain the importance of these two factors using scans obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI; see <http://www.loni.ucla.edu/ADNI>) preparatory phase.

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. 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). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principle Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research — approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

The purpose of the MRI portion of the preparatory phase of ADNI was to evaluate various technical options for scanning in order to select optimum protocol parameters prior to initiating patient enrolment. Further reading in the preparatory phase of ADNI can be found in Leow et al., 2006. One of the features of the preparatory phase was an evaluation of test-retest stability and this was accomplished by scanning healthy elderly volunteers twice at a short interval. This generated serial MRI data in which little biological change was expected, but which would contain change that might occur due to short-term drifts in scanner calibration. In this study, we used these serial images obtained from 3-T scanners equipped with phased array receive coils to assess the ability of different invocations of N3 to reduce signal non-uniformity.

Materials and methods

Subjects and Scans

Four sites with 3-T scanners (vendors included GE, Siemens and Philips) scanned 18 normal elderly subjects twice at baseline using different pulse sequences and again 2 weeks later, using multichannel phased array coils (Table 1). Inclusion criteria required that all subjects were between 55 and 90 years of age, with an informant/caregiver able to provide an independent evaluation of functioning. All enrolled subjects had mini-mental state examination (MMSE) (Folstein et al., 1975) scores of between 28 and 30 and a clinical dementia rating (CDR) of 0, without symptoms of depression, mild cognitive impairment (MCI) or other dementia and no current use of psychoactive medications.

T1-weighted sagittal volumes were obtained using magnetization prepared rapid gradient echo (MP-RAGE) and spoiled gradient echo (SPGR) pulse sequences.

Representative imaging parameters for the MP-RAGE sequence were TR = 2300 ms, TI = 900 ms, minimum full TE, flip angle = 8°, 170 sagittal 1.2-mm-thick slices, 256 × 256 matrix with a field of view of 256-260 mm, yielding a spatial resolution of 1.0 × 1.0 × 1.2 mm³. Detailed lists of imaging protocol parameters for the 3-T MP-RAGE on the three vendor's systems are available at <http://www.loni.ucla.edu/ADNI/Research/Cores/>. On the GE scanners, a pulse sequence that operates similarly (IR-SPGR) was modified to operate equivalently to the MP-RAGE sequence that is available from the other two vendors. In particular, a linear view ordering was used, and a recovery time was made available after the gradient echo readout train. We have called this sequence MP-RAGE for ease of discussion.

Representative imaging parameters for the SPGR sequence were identical to the MP-RAGE except that TR=16 ms, flip angle = 25° and 164 slices were acquired, and TI is not applicable because no inversion pulse is used. For this protocol, the SPGR pulse sequence is somewhat less efficient than MP-RAGE, so its typical acquisition time was 11 min, as opposed to 9.5 min.

As described in Bernstein et al. (2006), all of the image data here were pre-processed (prior to application of N3) with a B1-receive field correction (Narayana et al., 1988), that operates in a similar way to widely available commercial packages known as CLEAR (Philips), PURE (GE Healthcare), and Prescan Normalize (Siemens). Furthermore, the low flip angle (e.g., < 10°) of the MP-RAGE readout train minimized contrast variation from the central brightening artefact, which in general, is more difficult to correct than spatial modulation of the intensity alone. For the SPGR, which had a flip angle of 25°, a small but noticeable amount of contrast variation was expected.

Methods of correction

The N3 algorithm described in the introduction has several parameters that might require tuning: convergence criteria; maximal number of iterations; the shape of the Gaussian and

distance between the knot locations of the spline basis functions. In addition a mask can be provided to identify the region that N3 works over (e.g., brain as opposed to whole head or entire image). As N3 uses a histogram sharpening method, greater clarity in the histogram with respect to the form of the non-uniformity field should improve the method. Restricting the algorithm to the intracranial cavity is ideal, as it should consist of three distinct tissue classes: CSF, GM and WM. Thus a mask that correctly identifies the brain and CSF spaces should in theory help N3 to improve the correction. For all scans obtained, we used the following techniques to obtain five different masks with varying ability in identifying the intracranial volume (we also include a dummy control mask that indicates the original, uncorrected scan):

1. *Uncorrected*: No mask/no scan correction.

2. *Thresholded*: A lower voxel intensity threshold was calculated (Otsu, 1979), and any voxels in the scan that had an intensity less than this threshold were masked out. A largest connected component algorithm was then applied so any small islands of disconnected voxels were removed. This mask should exclude background voxels only, therefore regions of the image that are not needed for subsequent analysis (including neck, skull and exterior facial features) normally remain and were therefore included in the subsequent non-uniformity correction.

3. *Template 6 dof*: The ICBM 152 scan (Mazziotta et al., 2001) (also see http://www.bic.mni.mcgill.ca/cgi/icbm_view/) was registered to the incoming scan using 6 degrees of freedom (dof) (translations and rotations in the three coordinate axes of the scan), maximizing the normalized mutual information (NMI) (Studholme et al., 1999). The registration software used was FLIRT (Jenkinson and Smith, 2001) from the FSL toolkit, which has been shown to be highly robust (Smith et al., 2002). The registration parameters were then applied to the brain mask of the ICBM 152 template, using trilinear interpolation and thresholding at the 0.5 level, so it approximately overlaid the brain on the incoming scan. The mask included most CSF and intra-sulcal and ventricular spaces. The mask and the ICBM 152 template are shown in Fig. 1. This masking technique aims to include a volume of brain, CSF and (possibly) other parts of the head on the incoming scan. The volume of the masks should be the same for all scans, and thus will include different amounts external to the intracranial cavity for each person.

4. *Template 12 dof*: The ICBM 152 scan was registered to the incoming scan using 12 dof instead of 6 (i.e., including scalings and shears in the 3 coordinate axes of the scan as well as rotations/translations), and the mask transformed. This masking technique aims to improve on the template 6-dof mask by including scaling and shear factors to account for intersubject variability. Thus it should include brain and CSF spaces while parts external to the intracranial cavity are kept to a minimum.

5. *Manual-CSF*: A brain segmentation was performed using a semi-automated iterative morphological technique (Freeborough et al., 1997). This technique aims to include only grey and white matter, removing all ventricular and intra-sulcal CSF.

6. *Manual + CSF*: The above segmentation was filled, adding in ventricular and intra sulcal CSF spaces. This mask includes some CSF, adding another class to the histogram.

Note that the Manual-CSF and Manual+CSF masks are semi-automated, but for simplicity they are referred to as 'Manual.' Examples of each mask can be seen in Fig. 2. To assess the quality of the masks (and thus to check if the respective masking techniques have worked), we visually inspected all of the masks for each scan. We gave each mask a pass/fail rating based on whether or not it had fulfilled its criteria to an acceptable level. For the five masks, we used the following minimum criteria in order for each mask to be considered acceptable:

- *Thresholded*: All of the brain must be included.
- *Template 6 dof*: All brain must be included, inclusion of skull and small amounts of anatomy inferior to brain is acceptable.
- *Template 12 dof*: All brain must be included and significant amounts of CSF. Inclusion of skull and parts of anatomy inferior to brain should be minimal.
- *Manual-CSF*: Entire brain must be included, CSF should be excluded.
- *Manual+CSF*: Entire brain, ventricular and intra-sulcal CSF must be included, no other part of the anatomy should be present.

We chose four different spline distances for use in N3: 50 mm, 100 mm, 150 mm and 200 mm, which we chose to cover an appropriate range to highlight method variance, stepping down from the default of 200 mm (for 1.5-T data), reasoning that 3-T non-uniformity on phased array coils would have a shorter wavelength.

Thus there were 20 different ways N3 was applied to the 72 scans, resulting in a total of 1440 corrected images. N3 has other parameters that can be chosen, most of which are associated with a trade off between speed and accuracy. We chose the remaining parameters to favor accuracy over speed, but allowing the correction time not to be prohibitively expensive. The selection of the other parameters was:

1. FWHM — the full width at half maximum of the assumed distribution, chosen to be 0.05 (normalized voxel intensity).
2. Stopping threshold — 0.0001 (the coefficient of variation in the ratio between subsequent field estimates at each iteration, $e=sd(r_n)/(\text{mean}(r_n))$, $n = 1 \dots N$, where r_n is the ratio between subsequent non-uniformity field estimates at the n th location).
3. Resampling — 2 (scans are resampled to have a lower resolution, so we reduced resolution by a factor of 2 in x , y and z).
4. Maximum number of iterations — 1000.

Measures of comparison

Variation of white matter—To compare how well the different combination of masks and spline smoothing distances reduced the non-uniformity in the scans, we evaluated the intensity variation over the white matter. This tissue class is an ideal region over which to evaluate non-uniformity as it is relatively pure and reasonably homogeneous within a spatially extensive and contiguous region, unlike GM or CSF.

While this measure cannot distinguish between intensity non-uniformity, anatomical variation (such as regional differences in myelination density, which should be relatively small compared to the non-uniformity) and noise, we can assume that any reduction is at least partly due to the associated reduction in non-uniformity, and therefore that a larger reduction in the coefficient of variation is due to N3 working more effectively.

For each subject, a white matter segmentation was obtained for the different pulse sequences by an experienced image analyst using the original, uncorrected scan and the Manual-CSF mask as a starting point. The same iterative morphological technique used in obtaining the Manual-CSF mask (Freeborough et al., 1997) was then applied to obtain a white matter segmentation. This segmentation was then eroded once to remove any voxels on the GM/WM

interface that would have partial volume effects. Each baseline scan was then affinely registered (12 dof) to the repeat for both pulse sequences using the Manual-CSF masks, maximizing the NMI between the two. This registration was then used to transform the white matter mask (using trilinear resampling and thresholding at the 0.5 level) to each pulse sequence's respective repeat scan, so that the white matter masks on the baseline and repeat scans were the same.

The coefficient of variation $\lambda = \frac{\text{sd}(\text{WM}(x))}{\text{mean}(\text{WM}(x))}$ was then calculated for the original and all corrected scans over the white matter mask, to give a result for each scan and each type of correction. The geometric mean of λ for the MP-RAGE and SPGR pulse sequences was then plotted for each mask type and distance on a log scale.

Variance of normalized registered difference image—Given the importance of serial imaging to longitudinal studies such as ADNI, we also performed a similar test on registered scans from the two time points.

As each subject was scanned on two occasions 2 weeks apart, there should be negligible anatomical difference between the respective pulse sequence's scans at the two time points. If a normalized, registered, difference image were to be created for the MP-RAGE and SPGR sequences, it should be composed of registration error, physiological and instrumentation noise, and the difference between the two non-uniformity fields. If the non-uniformity had been removed or significantly reduced using any of the 20 different techniques, there should be a reduction in the variance of this normalized difference image that should relate to how well the non-uniformity had been removed from each pulse sequence's scan pairs. We assume that variability in subject positioning between two time points will lead to differences in the location and orientation of the non-uniformity with respect to the subject.

We therefore registered each MP-RAGE and SPGR repeat scan to their baselines, masking out all non-brain voxels before registration (using the Manual-CSF brain masks), using 12 dof and maximizing NMI. Using the transformation the repeat brain mask was transformed to the baseline using trilinear interpolation and thresholded at the 0.5 level to re-binarize the mask. This transformed repeat mask was then intersected with the baseline mask to create a joint mask. The repeat scan was then transformed to the baseline using renormalized windowed sinc interpolation (Thacker et al., 1999). The intensities of both the baseline and registered repeat scans were then normalized by dividing each scan's voxels by its respective mean in the joint mask following a single erosion. The difference image for the MP-RAGE and SPGR sequences was then calculated by subtracting the registered repeat normalized scan from the baseline normalized scan. This difference image was then calculated for both pulse sequences' scan pairs, and the variance $\omega = \text{var}(\text{diff}(x))$ was calculated over the joint mask to obtain a value for each scan that could be used in statistical analysis. The geometric mean of ω across all subjects was also plotted for each mask type and distance on a log scale on different graphs for the MP-RAGE and SPGR scans.

Statistical analysis

To take into account the repeated measures nature of the study design linear mixed models were used to analyze the white matter and difference image variance of the MP-RAGE and SPGR pulse sequences separately. Both the WM and paired difference variance measures were log transformed to facilitate comparisons on a percentage scale. The masking type and spline smoothing distance were considered as six and four category fixed effects, respectively. Interaction terms were included to investigate whether the differences between masks depended on the smoothing distance. The model for difference image variance included four types of random effects. These were (i) subject-specific effects (impacting on all measures on a particular subject), (ii) subject-specific "mask/control" effects (one impacting on the single

measure made from the uncorrected scan and another one all other measures made on that subject), (iii) effects for each mask and subject combination (impacting on all measures for a particular mask on a given subject) and (iv) effects for each smoothing distance and subject combination. A similar model was used for the white matter variance measures with the addition that each of the random effects was partitioned into subject and scan effects (because there were two scans for each subject-mask-smoothing distance combination). For both models a global test for interaction between mask type and smoothing distance was performed for the MP-RAGE and SPGR pulse sequences. Provided that this achieved statistical significance ($p < 0.05$) pair-wise comparisons (without adjustment for multiple comparisons) were made between the different masks at each smoothing distance. Differences were expressed as percentage differences in geometric mean levels with 95% confidence intervals.

Results

Fig. 3 shows the MP-RAGE-uncorrected and N3-corrected scan and the associated non-uniformity field when using the template 12-dof mask and a 50-mm smoothing distance. Fig. 4 shows the equivalent SPGR scan (of the same person — note that the SPGR scan is registered to the MP-RAGE for visualization, thus the slices of Figs. 3 and 4 are equivalent).

Note in the uncorrected MP-RAGE scan the bright, deep central region typical of 3-T non-uniformity, affecting image intensity in the thalamus and cerebellar white matter, and how in the corrected scans the spatial variation in intensity has been removed. The SPGR scan is not affected to the same degree; however, non-uniformity is still apparent.

Table 2 shows the number of masking failures using each masking technique for the two pulse sequences. We found when using the thresholding technique on both MP-RAGE and SPGR scans that the value calculated usually led to the neck, skull and dura being included in the mask, but sometimes large parts of CSF were classed as background and excluded. However, in the MP-RAGE scans the severity of the non-uniformity field (for an example, see Fig. 3) led to areas of grey matter approaching the intensity of CSF, which led to parts of brain being excluded and thus the masks were rated as a fail. SPGR scans were not as severely affected by the non-uniformity field and thus all regions created using the thresholding technique included all of the brain.

The two failures in the MP-RAGE and SPGR template 12-dof masks were from the same subject. The mask included the eyes and base of skull anterior to brainstem for both baseline and repeat scans (the aim of the masking technique is to include the brain and CSF spaces only). This was probably related to the subject's unusually large frontal sinus and enlarged superior CSF spaces, which confounded the 12-dof registration.

MP-RAGE white matter variation

Fig. 5 shows mean levels of normalized white matter variance according to mask and distance for the MP-RAGE sequence. Table 3 shows the percentage difference between geometric mean levels for each mask relative to the template 12-dof mask at the same smoothing distance. A global test of interaction provided evidence ($p < 0.0001$) that the way in which white matter variance varies with smoothing distance does differ between masks. Both Thresholded and Uncorrected scans give rise to variances that are on average more than 200% greater than those for the template 12-dof mask at all smoothing distances. The template 6-dof mask is also associated with statistically significantly higher levels at each distance but the magnitude of the difference is reduced. The Manual-CSF mask resulted in levels that were broadly similar to those with the template 12-dof mask, although results differed slightly by smoothing distance. At a distance of 100 mm Manual-CSF gave results that were on average 23% higher than those with the template 12-dof mask (a difference that just achieved statistical significance,

$p < 0.05$). At smoothing distances of 50 and 150 mm the template 12-dof mask still performed best, but differences were small and not statistically significant. At a smoothing distance of 200 mm the Manual-CSF mask was associated with levels that were 14% lower, but this difference was also not statistically significant. Results for Manual+CSF were similar to those for Manual-CSF.

MP-RAGE paired difference variance

Fig. 6 shows mean levels of the variance of the difference image according to mask and smoothing distance for the MP-RAGE sequence. Table 4 shows the percentage difference between geometric mean levels for each mask relative to the template 12-dof mask at the same smoothing distance. A global test of interaction provided evidence ($p=0.013$) that the way in which difference image variance varies with smoothing distance does differ between masks. The Thresholded, template 6 dof and Uncorrected scans have greater variance at all smoothing distances when compared to the template 12-dof scans. In contrast, variance with the Manual +/-CSF masks is less than that of the template 12 dof at all smoothing distances. At larger smoothing distances the reductions are of the order of 10% and statistically significant ($p < 0.05$). However, for the smaller distances, the differences are small and not statistically significant.

SPGR white matter variance

Fig. 7 shows mean levels of normalized white matter variance according to mask and distance for the SPGR sequence. Table 5 shows the percentage difference between geometric mean levels for each mask relative to the template 12-dof mask at the same smoothing distance. A global test of interaction provided evidence ($p = 0.0007$) that the way in which white matter variance varies with smoothing distance does differ between masks. Uncorrected scans give rise to variances that are on average more than 200% greater than those for the template 12-dof mask at all smoothing distances, with the largest percentage difference at a distance of 50 mm. For thresholded scans variances are about double those for the template 12-dof mask at all smoothing differences, with the percentage difference increasing with smoothing difference. Template 6-dof masks are also associated with statistically significantly higher variances at each distance but the magnitude of the differences is reduced. The Manual-CSF mask and Manual+CSF masks result in levels that are similar to those with the template 12-dof mask, with no pair-wise differences achieving statistical significance.

SPGR paired difference variance

Fig. 8 shows mean levels of the variance of the difference image according to mask and smoothing distance for the SPGR sequence. Table 6 shows the percentage difference between geometric mean levels for each mask relative to the template 12-dof mask at the same smoothing distance. A global test of interaction provided evidence ($p=0.012$) that the way in which difference image variance varies with smoothing distance does differ between masks. The Thresholded, template 6 dof and Uncorrected scans have greater variance at all smoothing distances when compared to the template 12-dof scans, although only the differences with the Uncorrected scans (at all distances) and the Thresholded scan at a distance of 50 mm achieved statistical significance ($p < 0.05$). In contrast, variance with the Manual +/-CSF masks is less than that of the template 12 dof at all smoothing distances. These differences are generally small with only that with the Manual+CSF mask at a distance of 150 mm achieving statistical significance ($p=0.044$).

Discussion

Upon initial inspection of the scans, we noted that after the B1-receive field correction a high degree of residual non-uniformity centred in the brain persisted. This was more pronounced in

the MP-RAGE compared to the SPGR sequence and is quantitatively shown by Figs. 5 and 7. Thus it is likely that different pulse sequences are affected to varying degrees by non-uniformity in 3-T scanning.

Visual inspection of the N3-corrected scans indicated the residual non-uniformity was considerably reduced when using accurate masks and smaller smoothing distances, as indicated in Figs. 5-8. The quantitative measures support our assumption that masks with greater clarity in the histogram improves the correction, and also that reducing the smoothing distance encapsulates the form of the non-uniformity field more accurately. For both the white matter and normalized difference images, the variance is significantly reduced as the accuracy of the masks increase and the smoothing distance decreases. These quantitative measurements are consistent with the findings of our visual assessments.

Upon inspection of the masks that were created (Table 2), the most notable problem is the number of failures in the thresholded masking technique for the MP-RAGE scans. Overall the mask included a significant amount of non-brain (also the case for the SPGR thresholded masks), with CSF and sometimes brain being excluded. This may explain the poor performance of the non-uniformity correction associated with this mask for both pulse sequences, as not enough distinct tissue classes are made available to N3. The sensitivity of the threshold on the MP-RAGE scans to the non-uniformity is an important problem; clearly a masking technique should be independent of any non-uniformity to be of any practical use. For both masks derived from the 6-dof and 12-dof template registration techniques, we used an entropy-based measure (NMI) as the cost function for the registration, which should be largely independent of any non-uniformity present in the scan (Maes et al., 1997). Despite this the four masking failures on the single subject when using the 12-dof registration show that the technique is not completely robust, and suggest that the use of a different template (e.g., one created from a group of aged normal controls) may be more appropriate. Although the registration technique has been shown to be highly robust (Smith et al., 2002), it would be advisable to check the generated masks prior to N3 correction (e.g., range checking of the volume or visual inspection), and any of poor quality interactively edited.

When comparing the correction using the template 12-dof mask to the semi-automated (Manual) correction, it is apparent on both MP-RAGE and SPGR scans that the variance results are more tightly clustered over the different smoothing distances. For the MP-RAGE corrections at the 100- and 50-mm distances there is little to choose between the template 12 dof and Manual masks, while at the 150- and 200-mm distances the template 12-dof mask is inferior. For the SPGR sequence the 50-mm smoothing distance performed better than the 100 mm in both WM and paired difference variance assessments.

We initially expected that including CSF in any mask would improve the performance of N3 as another, distinct tissue class would be included, and the greater volume would encapsulate more of the non-uniformity, leading to higher accuracy. However, this was not the case, as there was no significant difference between the Manual-CSF and Manual+CSF masks at any smoothing distance, indicating that two tissue classes suffice. This finding suggests that the template MNI 152 template mask we created could be altered to include less extra-sulcal CSF, which may reduce the amount of tissue external to the brain cavity included in the incoming scan, further improving the template-based correction.

In determining an ideal smoothing distance, the quantitative results of the MP-RAGE indicate that 50 mm and 100 mm are significantly better than 150 or 200 mm, with a minimal improvement between 50 and 100 mm in our assessment. For the SPGR scans, there is greater benefit in using a smaller smoothing distance of 50 mm, as shown in Figs. 7 and 8. However, smaller distances will naturally smooth out any variance; the key is that it must smooth only

non-uniformity, not noise or anatomical variation. We believe that no slowly varying anatomical variance should occur over a spatial scale of 50 mm, and thus N3 should not be capable of removing any anatomical features in the scans. We checked this by visually inspecting some of the 50 mm corrected scans and found that no naturally occurring variance was eliminated.

One of the parameters of N3 that was not considered in this paper is the FWHM, which is used to describe the assumed distribution of the non-uniformity field. In this paper it was chosen to be 0.05, in order to favor accuracy over speed. Sled noted this and also suggested a varied approach to the FWHM parameter during the correction procedure, leading to faster convergence but reduced accuracy of the estimated non-uniformity field. This idea was not explored and could be in future to obtain more accuracy with less computational cost.

The main finding of this work is that the reliability and robustness of the N3 algorithm for correcting non-uniformity in 3-T head scans for different pulse sequences can be improved by considered selection of brain masks and the smoothing parameter. N3 cannot be assumed to perform well for a wide range of scanners and coils using just the default parameters. We conclude that for 3-T brain imaging using phased array receiver coils a spline distance of 50 mm should be chosen and a brain mask provided which excludes non-brain tissues, as inclusion of CSF spaces does not appear essential. Automated brain masks created using 12-dof registration of a brain template performed well although checking of the masks or anomalous results is recommended. An informed selection of the smoothing distance can be made using visual inspection and knowledge of the sequence, coil and field strength. These factors are likely to be particularly important for other higher field scanners and phased array coils.

Acknowledgments

This project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; Principal Investigator: Michael Weiner; NIH grant number U01 AG024904). ADNI is funded by the National Institute of Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the Foundation for the National Institutes of Health, through generous contributions from the following companies and organizations: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck and Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, the Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories and the Institute for the Study of Aging (ISOA), with participation from the U.S. Food and Drug Administration. We would also like to thank Dr. John Stevens, consultant neuroradiologist at the Institute of Neurology, UCL, UK, for assessment of scans.

References

- Arnold JB, Liow JS, Schaper KA, Stern JJ, Sled JG, Shattuck DW, Worth AJ, Cohen MS, Leahy RM, Mazziotta JC, Rottenberg DA. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage* 2001;13:931–943. [PubMed: 11304088]
- Ashburner J, Csernansky JG, Davatzikos C, Fox NC, Frisoni GB, Thompson PM. Computer-assisted imaging to assess brain structure in healthy and diseased brains. *Lancet Neurol* 2003;2:79–88. [PubMed: 12849264]
- Belaroussi B, Milles J, Carme S, Zhu YM, Benoit-Cattin H. Intensity non-uniformity correction in MRI: existing methods and their validation. *Med. Image Anal* 2006;10:234–246. [PubMed: 16307900]
- Bernstein MA, Huston J, Ward HA. Imaging artifacts at 3.0T. *J. Magn. Reson. Imaging* 2006;24:735–746. [PubMed: 16958057]
- Collins CM, Liu WZ, Schreiber W, Yang QX, Smith MB. Central brightening due to constructive interference with, without, and despite dielectric resonance. *J. Magn. Reson. Imaging* 2005;21:192–196. [PubMed: 15666397]
- Folstein M, Folstein S, McHughes P. The “Mini mental state”: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res* 1975;12:189–198. [PubMed: 1202204]

- Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imag* 1997;16:623–629.
- Freeborough PA, Fox NC, Kitney RI. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput. Methods Programs Biomed* 1997;53:15–25. [PubMed: 9113464]
- Guillemaud R, Brady M. Estimating the bias field of MR images. *IEEE Trans. Med. Imag* 1997;16:238–251.
- Jack CRJ, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, Knopman DS, Boeve BF, Ivnik RJ, Smith GE, Cha RH, Tangalos EG, Petersen RC. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* 2004;62:591–600. [PubMed: 14981176]
- Janke AL, de Zubicaray G, Rose SE, Griffin M, Chalk JB, Galloway GJ. 4D deformation modeling of cortical disease progression in Alzheimer's dementia. *Magn. Reson. Med* 2001;46:661–666. [PubMed: 11590641]
- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal* 2001;5:143–156. [PubMed: 11516708]
- Leow AD, Klunder AD, Jack CR, Toga AW, Dale AM, Bernstein MA, Britson PJ, Gunter JLG, Ward CP, Whitwell JL, Borowski BJ, Fleisher AS, Fox NC, Harvey D, Kornak J, Schuff N, Studholme C, Alexander GE, Weiner MW, Thompson PM. Longitudinal stability of MRI for mapping brain change using tensor based morphometry. *NeuroImage* 2006;31:627–640. [PubMed: 16480900]
- Lewis EB, Fox NC. Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage* 2004;23:75–83. [PubMed: 15325354]
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag* 1997;16:187–198.
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond., B Biol. Sci* 2001;356:1293–1322. [PubMed: 11545704]
- Narayana PA, Brey WW, Kulkarni MV, Sievenpiper CL. Compensation for surface coil sensitivity variation in magnetic-resonance imaging. *Magn. Reson. Imaging* 1988;6:271–274. [PubMed: 3398733]
- Otsu N. Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern* 1979;9:62–66.
- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imag* 1998;17:87–97.
- Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. *J. Comput. Assist. Tomogr* 2001;25:466–475. [PubMed: 11351200]
- Smith SM, Zhang YY, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 2002;17:479–489. [PubMed: 12482100]
- Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recogn* 1999;32:71–86.
- Thacker NA, Jackson A, Moriarty D, Vokurka EA. Improved quality of re-sliced MR images using re-normalized sinc interpolation. *J. Magn. Reson. Imaging* 1999;10:582–588. [PubMed: 10508326]
- Thompson PM, Hayashi KM, Sowell ER, Gogtay N, Giedd JN, Rapoport JL, de Zubicaray GI, Janke AL, Rose SE, Semple J, Doddrell DM, Wang Y, van Erp TG, Cannon TD, Toga AW. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage* 2004;23 (Suppl 1):S2–S18. [PubMed: 15501091]
- Tincher M, Meyer CR, Gupta R, Williams DM. Polynomial modeling and reduction of RF body coil spatial inhomogeneity in MRI. *IEEE Trans. Med. Imag* 1993;12:361–365.
- Vovk U, Pernus F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imag* 2007;26:405–421.
- Wang D, Chalk JB, Rose SE, de Zubicaray G, Cowin G, Galloway GJ, Barnes D, Spooner D, Doddrell DM, Semple J. MR image-based measurement of rates of change in volumes of brain structures: Part

II. Application to a study of Alzheimer's disease and normal aging. *Magn. Reson. Imaging* 2002;20:41–48. [PubMed: 11973028]

Wells WM III, Grimson WEL, Kikinis R, Jolesz FA. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imag* 1996;15:429–442.

Zhang YY, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag* 2001;20:45–57.

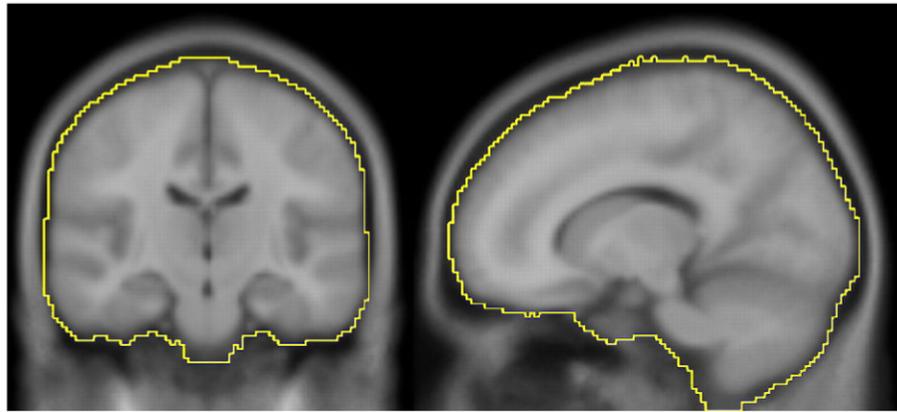


Fig. 1. ICBM152 scan and mask used to obtain the template 6-dof and 12-dof masks.

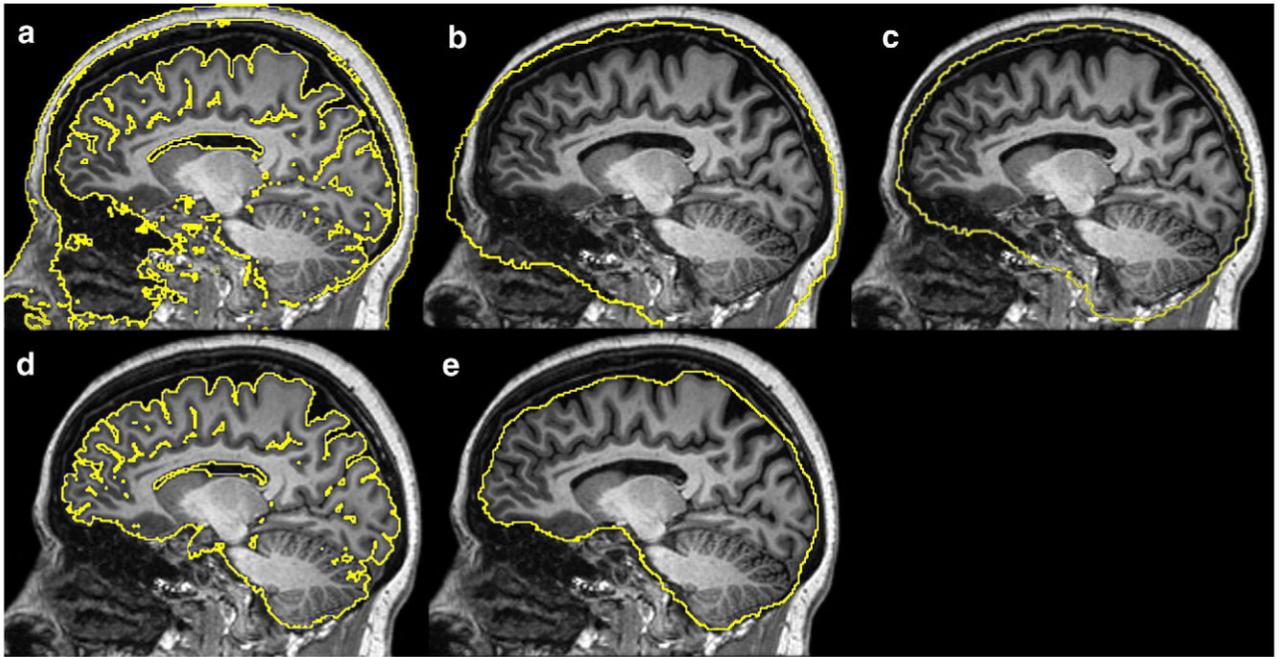


Fig. 2. A representative example of the masks obtained for one particular MP-RAGE scan using the different techniques: (a) thresholded; (b) template 6 dof; (c) template 12 dof; (d) Manual-CSF; (e) Manual+CSF. Note the scale difference between panels b and c, with b being much larger, reflecting the scaling change in the 12-dof registration.

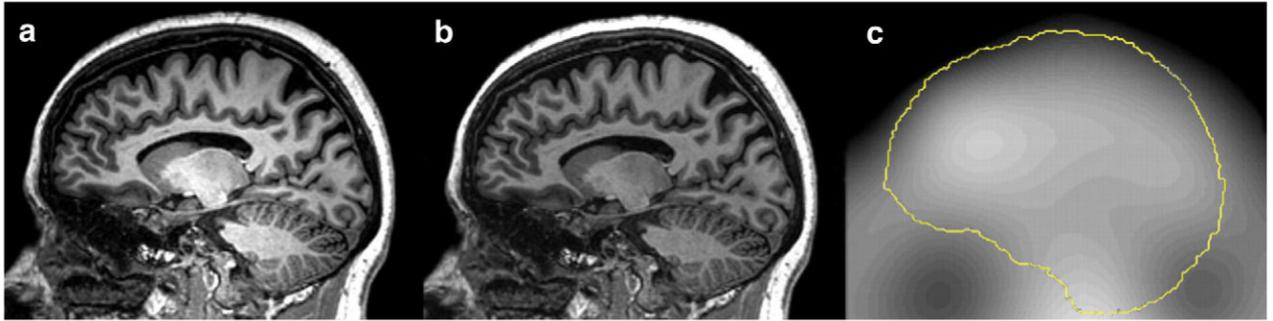


Fig. 3. MP-RAGE (a) initial scan, (b) corrected scan (using the template-12-dof mask and 50-mm smoothing distance) and (c) the estimated non-uniformity field obtained with the template 12-dof mask overlaid.

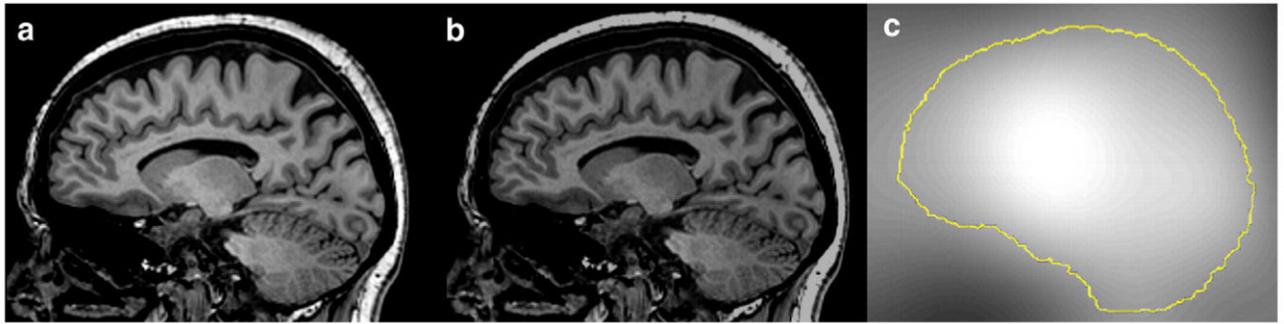


Fig. 4. SPGR (a) initial scan, (b) corrected scan (using the template 12-dof mask and 50-mm smoothing distance) and (c) the estimated non-uniformity field obtained with the template 12-dof mask overlaid.

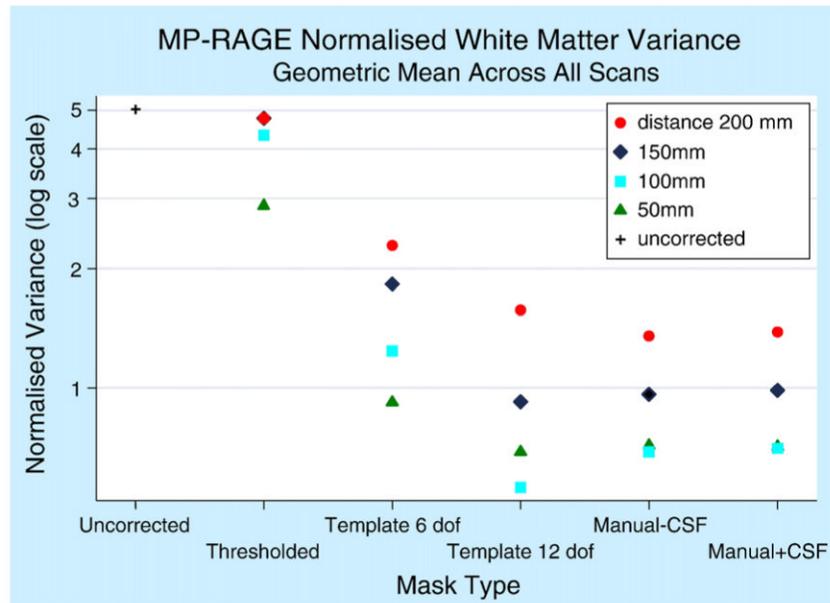


Fig. 5. MP-RAGE mean normalized white matter intensity variances when using N3 for different masks and smoothing distances.

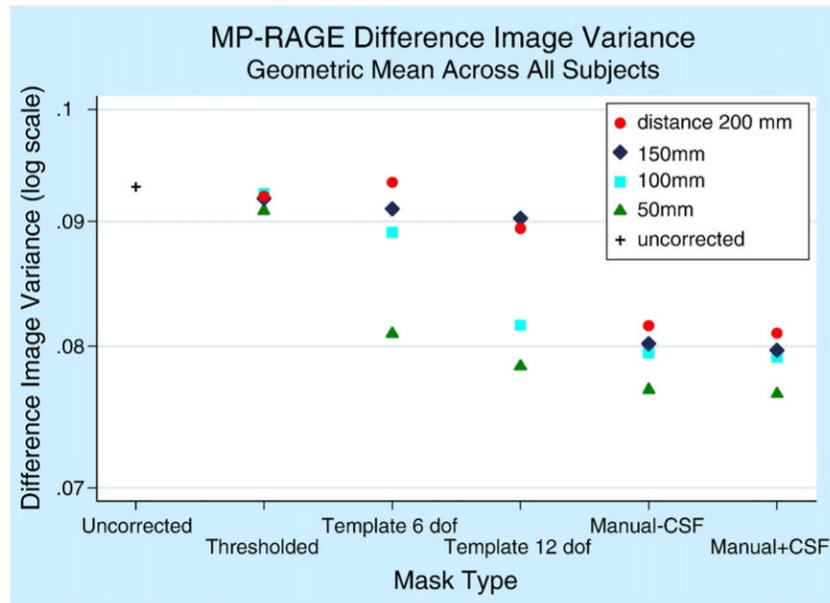


Fig. 6. MP-RAGE mean difference image variances when using N3 for different masks and smoothing distances.

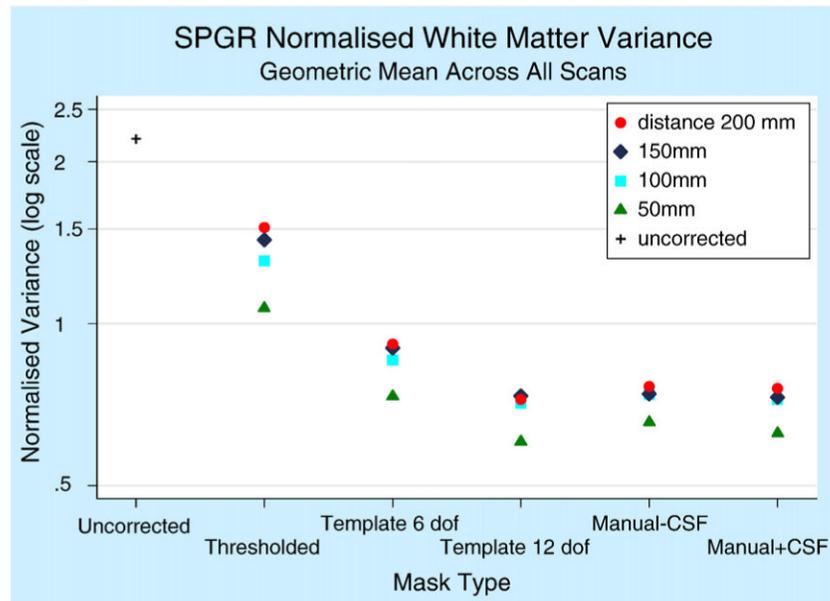


Fig. 7. SPGR mean normalized white matter intensity variances when using N3 for different masks and smoothing distances.

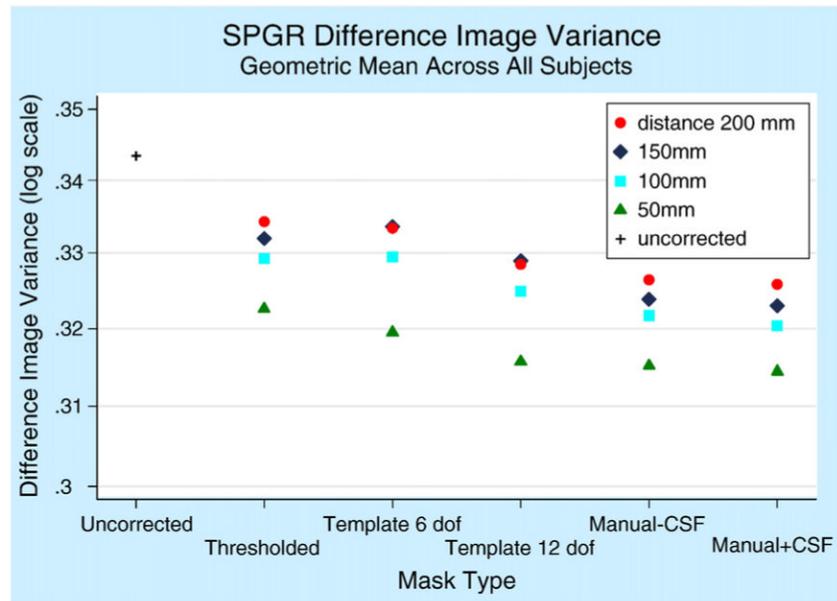


Fig. 8. SPGR mean difference image variances when using N3 for different masks and smoothing distances.

Table 1

Sites involved in 3-T scanning in the ADNI preparatory phase

Site 1 (Mayo Clinic, Rochester, Minnesota, USA)	Site 2 (Johns Hopkins University, Baltimore, Maryland, USA)	Site 3 (Boston University, Boston, Massachusetts, USA)	Site 4 (University of California, Davis, USA)
GE Signa 3 T Subjects: 9	Philips Intera 3 T Subjects: 3	Philips Intera 3 T Subjects: 3	Siemens Trio 3 T Subjects: 3

The scanner type and the number of subjects (who were scanned twice 2 weeks apart, using MP-RAGE and SPGR pulse sequences, resulting in four scans per subject) at each site are indicated.

Table 2

Pass/Fail of each masking technique for the two different pulse sequences (baseline and repeat timepoints combined)

Masking technique	Total passed MP-RAGE	Total failed MP-RAGE	Total passed SPGR	Total failed SPGR
Thresholded	29	7	36	0
Template 6 dof	36	0	36	0
Template 12 dof	34	2	34	2
Manual-CSF	36	0	36	0
Manual+CSF	36	0	36	0

MP-RAGE percentage increase in (geometric) mean white matter variance (95% confidence interval) score for each correction method relative to that of the Template 12 dof mask, comparing like for like smoothing distances

Table 3

Distance/Mask (mm)	Uncorrected	Thresholded	Template 6 dof	Manual-CSF	Manual+CSF
50	628 (473, 826)	317 (243, 407)	33 (10, 162)	4 (-15, 26)	3 (-15, 25)
100	795 (605, 1038)	673 (535, 937)	121 (82, 168)	23 (1, 49)	25 (3, 52)
150	444 (328, 792)	417 (326, 528)	98 (63, 140)	4 (-14, 27)	7 (-13, 30)
200	220 (152, 307)	205 (151, 270)	46 (20, 77)	-14 (-29, 4)	-12 (-27, 7)

Table 4

MP-RAGE percentage increase in (geometric) mean normalized difference variance (95% confidence interval) score for each correction method relative to that of the Template 12 dof mask, comparing like for like smoothing distances

Distance/Mask (mm)	Uncorrected	Thresholded	Template 6 dof	Manual-CSF	Manual+CSF
50	18 (9, 28)	15 (7, 25)	3 (-5, 12)	-2 (-10, 6)	-3 (-10, 5)
100	14 (5, 23)	13 (5, 22)	9 (1, 18)	-3 (-10, 5)	-3 (-10, 5)
150	3 (-5, 12)	2 (-6, 10)	1 (-7, 9)	-11 (-18, -4)	-12 (-18, -4)
200	4 (-4, 13)	3 (-5, 11)	4 (-3, 13)	-9 (-16, -1)	-9 (-16, -2)

Table 5

SPGR percentage increase in (geometric) mean white matter variance (95% confidence interval) score for each correction method relative to that of the Template 12 dof mask, comparing like for like smoothing distances

Distance/Mask (mm)	Uncorrected	Thresholded	Template 6dof	Manual-CSF	Manual+CSF
50	266 (203, 342)	77 (57, 100)	21 (7, 37)	9 (-4, 23)	4 (-8, 17)
100	210 (157, 275)	84 (63, 108)	20 (7, 36)	4 (-8, 18)	2 (-10, 15)
150	201 (149, 263)	95 (73, 121)	23 (9, 39)	1 (-11, 14)	-1 (-12, 12)
200	205 (152, 268)	109 (85, 136)	27 (12, 43)	6 (-7, 19)	5 (-7, 18)

Table 6

SPGR percentage increase in (geometric) mean normalized difference variance (95% confidence interval) score for each correction method relative to that of the Template 12 dof mask, comparing like for like smoothing distances

Distance/Mask (mm)	Uncorrected	Thresholded	Template 6dof	Manual-CSF	Manual+CSF
50	8.8 (5.3, 12.4)	2.2 (0.3, 4.0)	1.2 (-0.6, 3.0)	-0.2 (-1.9, 1.6)	-0.4 (-2.2, 1.4)
100	5.7 (2.7, 9.3)	1.4 (-0.4, 3.2)	1.4 (-0.4, 3.3)	-1.0 (-2.7, 0.8)	-1.4 (-3.1, 0.4)
150	4.4 (1.0, 7.9)	0.9 (-0.9, 2.7)	1.4 (-0.4, 3.3)	-1.6 (-3.3, 0.2)	-1.8 (-3.6, -0.1)
200	4.6 (1.2, 8.1)	1.8 (-0.1, 3.6)	1.5 (-0.3, 3.3)	-0.6 (-2.4, 1.2)	-0.8 (-2.6, 1.0)