schizophrenia (SZ) based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria alone (Zimmermann et al., 2009). Although many clinical symptoms are common to various psychiatric disorders, depressive symptoms are particularly ubiquitous in the disease process or clinical staging of various psychiatric disorders (Hafner et al., 2005). For instance, differentiation between BP presenting with depressive symptoms and unipolar MDD is a topical issue (Akiskal et al., 1995). Indeed, most patients with BP with depressive symptoms are initially diagnosed with and treated for MDD (Akiskal et al., 1995; Goldberg et al., 2001). Therefore, biomarkers that can facilitate early and accurate differentiation of BP with depressive symptoms from MDD are necessary.

In addition, depressive symptoms that fulfil the operational diagnostic criteria for a depressive episode/major depression can also occur at any stage of SZ and can contribute substantially to its associated morbidity and even mortality (an der Heiden et al., 2005). The differentiation of SZ from MDD, especially in the early stages, is also important because patients with SZ also exhibit non-psychotic and non-specific prodromal symptoms (e.g., depressive or negative symptoms and cognitive deficits) for several years before the onset of full-blown psychosis (McGorry et al., 2008). Therefore, the availability of clinically useful and cost-effective biomarkers for the differential diagnosis of major psychiatric disorders would likely enhance patient management, improve treatment/therapeutic response and lead to targeted therapies tailored to the individual (Holsboer, 2008). Despite their potential, to date, no such biomarkers have been established.

Functional imaging studies are one source of potential biomarkers (Gur et al., 2007; Phillips and Vieta, 2007); these studies have previously elucidated subtle brain abnormalities in patients with major psychiatric disorders relative to healthy control (HC) individuals and have been applied to the differential diagnosis of psychiatric disorders (e.g., to differentiate MDD from SZ, Barch et al., 2003 or BP, Almeida et al., 2009). However, some functional neuroimaging techniques are limited by the fact that, during the procedure, the individuals need to be placed in an uncomfortable or unnatural setting (e.g., lying in a supine position in a narrow gantry with the head fixed during the entire examination), for accurate measurement.

In contrast, multi-channel near-infrared spectroscopy (NIRS) using near-infrared light provides a completely non-invasive measurement of the spatiotemporal characteristics of brain function in ordinary clinical settings and allows patients to be comfortably seated in a well-lit room; therefore, it is considered a method for 'real-world neuroimaging'. Additionally, NIRS has relatively low maintenance costs and does not involve ionising radiation or objectionable noise; thus, it can be repeated as needed even for patients with psychiatric disorders. The utility and limitations of NIRS have been discussed extensively in previous reports (Ferrari and Quaresima, 2012; Obrig and Villringer, 2003; Strangman et al., 2002a). NIRS allows the measurement of haemoglobin concentration changes (1) only in the cortical surface area located immediately beneath the probes, but not in deeper brain structures, and (2) with limited spatial resolution, although it has a high temporal resolution. In NIRS, typical cortical activation represents not only decreased concentration of deoxy-haemoglobin ([deoxy-Hb]), which is considered the main source of blood oxygenation level-dependent (BOLD) contrast increase in functional magnetic resonance imaging (fMRI), but also a relatively larger increase in oxy-haemoglobin concentration ([oxy-Hb]) (Fig. 1).

The verbal fluency task (VFT) is a cognitive task that is used as a neuropsychological test or a neuroimaging task. The VFT elicits different abnormalities relevant to each diagnostic group of major psychiatric disorders (Curtis et al., 2001; Zanelli et al., 2010). We previously developed a very brief (<3 min) VFT and used it to investigate the differential fronto-temporal haemodynamic pattern between MDD and SZ (Suto et al., 2004) or MDD and BP (Kameyama et al., 2006), as well as the relationship between NIRS signals and functional impairment in SZ (Takizawa et al., 2008). We also found functional NIRS abnormalities in individuals at ultra-high risk for SZ and patients with first-episode psychosis (Koike et al., 2011). However, the clinical applicability of NIRS to the differential diagnosis of individuals remains uncertain. In this study, we extended our translational approach to replicate our previous findings (Kameyama et al., 2006; Suto et al., 2004) in a seven-site collaborative study using a large, fully independent sample set, and to evaluate the application of NIRS to psychiatric differential diagnosis in natural clinical settings.

Specifically, we used NIRS with wide coverage of the prefrontal and temporal cortices to investigate whether the frontal and temporal brain haemodynamic responses induced by cognitive activation could serve as biomarkers of underlying major psychiatric disorders with depression. To validate the reproducibility and generalisability of the results, we applied an algorithm developed using the data generated at the initial site to the test data derived from the remaining 6 sites. We hypothesised that the spatiotemporal characteristics of the haemodynamic responses detected by NIRS would not only differentiate patients with psychiatric disorders from HCs with acceptable sensitivity and specificity, but would also differentiate correctly and with a high concordance rate patients with MDD from patients with bipolar disorder and schizophrenia who present with depressive symptoms.

Material and methods

Participants

This multi-site study was performed in 7 hospitals: 6 were affiliated with universities (Fukushima, Gunma, Mie, Tokyo, Showa, and Tottori) and one was affiliated with the National Centre of Neurology and Psychiatry of Japan. The sites were situated in the Tokyo metropolitan area and in moderate-scale prefectural capital cities (Fukushima, Maebashi, Tsu and Yonago). The participants were recruited from June 2004 to June 2009, with the exception of recruitment at the initial site (Gunma University Hospital in Maebashi City), which was conducted over 6 years (March 2003 to March 2009). The ethics committees of the participating hospitals approved this collaborative study. In accordance with the Declaration of Helsinki, all participants gave written informed consent after receiving a complete explanation of the study.

Six hundred and seventy-three in-patients and out-patients with psychiatric disorders (MDD, BP and SZ), in addition to 1007 HC volunteers (Flow diagram (1)), were initially enrolled. Of note, these individuals were not the same as those included in our previous studies (Kameyama et al., 2006; Suto et al., 2004). The patients were diagnosed by experienced psychiatrists based on the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (First et al., 1997). The HC volunteers were hospital staff members, university students and members of the general population who responded to website or newspaper advertisements in each city. The SCID non-patient edition was used to screen HC individuals. The exclusion criteria of the initial enrolment were neurological illness, traumatic brain injury with any known cognitive consequences and alcohol/substance abuse or addiction. All participants were native Japanese speakers who were capable of performing a Japanese version of the VFT easily.

On the day of NIRS measurement, the depressive symptoms of participants were evaluated using the 17-item Hamilton Rating Scale for Depression (HAMD) (Hamilton, 1960) and their psychotic and manic symptoms were evaluated using both the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1991) and the Young Mania Rating Scale (YMRS) (Young et al., 1978), respectively, by well-trained psychiatrists with no knowledge of the NIRS data. During the study, all patients with psychiatric disorders were medicated with one or more agents (anti-psychotics, anti-depressants, anxiolytics and/or anti-parkinsonian agents), with the exception of 10 drug-free patients with MDD and 5 drug-free patients with SZ.

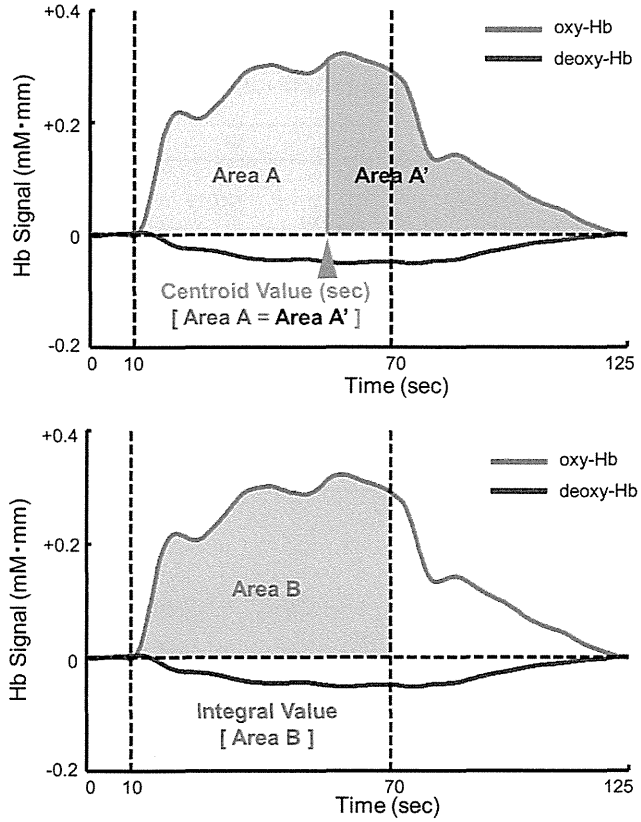


Fig. 1. Typical time-course pattern of near-infrared spectroscopy (NIRS) signals coupled with the verbal fluency task. The 'centroid value' is indicated by the time [s], which is indicated by a perpendicular line from the centroid of an NIRS signal-change area (calculated with positive change) throughout all the task periods. Oxy-Hb: oxygenated haemoglobin signal; deoxy-Hb: deoxygenated haemoglobin signal.

To minimise the influence of confounding factors, we performed group matching for age and gender among the 4 diagnostic groups using one-way analysis of variance (ANOVA) and a chi-squared test, which excluded randomly selected individuals and brought the mean age of the HC individuals and patients with MDD or SZ in closer alignment with that of the patients with BP (44.0 ± 14.9 years old (y.o.)), which was the group with the fewest individuals (Table 1 and Flow diagram (2)). For confirmation, we also analysed demographically non-matched samples that are identified in the Supplementary Material (I). The overall results were the same as those described in the main manuscript and the reduction in the total number of study participants after demographical matching did not appear to have an influence on the development of the algorithm (see Supplementary Material (I)).

Because our clinically valuable target were help-seeking unremitted patients, subsequently we excluded study participants with extremely mild symptoms (HAMD score ≤ 5, PANSS depression item score ≤ 1, PANSS negative symptom score ≤ 11, PANSS general psychopathology score ≤ 21, or PANSS positive symptom score − negative symptom score ≤ 11; the latter 3 criteria were based on the criteria from the PANSS manual for the 5th percentile of patients with mild SZ, Kay et al., 1991). We also excluded patients in a manic phase (YMRS score > 10) from the NIRS measurement; rather, we focused on patients with BP who were in the depressive phase because the different phases may produce different brain dysfunctions in patients with BP (Phillips and Vieta, 2007), and manic patients with BP were diagnosed without apparent difficulty (Flow diagram (3)).

Activation task

The activation task used in this study was similar to that used in our previous studies (Kameyama et al., 2006; Suto et al., 2004; Takizawa et al., 2008). Briefly, a VFT (letter version) was administered and NIRS signal ([oxy-Hb] and [deoxy-Hb]) changes were measured during a 10 s pre-task baseline period, a 60 s activation period and a 55 s post-task baseline period. During the activation period, the participants were instructed to utter as many Japanese words beginning with a designated syllable as possible. For the pre- and post-task baseline periods, the individuals were instructed to simply repeat Japanese vowels out loud. The total number of correct words generated during the 60 s activation period was used as the measure of task performance (Table 1).

Among the many neuropsychological tasks used for detecting neurocognitive deficits in patients with major psychiatric disorders (Barrett et al., 2009; Zanelli et al., 2010), we selected the VFT because it is an executive task that exhibits distinct differences in performance and neuroimaging data among each diagnostic group of major psychiatric disorders (Costafreda et al., 2006; Curtis et al., 2001; Zanelli et al., 2010). In addition, the VFT is easy to understand and execute; in fact, all participants generated more than one word during the VFT. Therefore, this task is suitable for translational research aimed at identifying practical biomarkers.

NIRS measurement

The NIRS apparatus and measurement procedure were described in full previously (Takizawa et al., 2008). Briefly, we used a 52-channel NIRS system (ETG-4000; Hitachi Medical Co., Tokyo, Japan). The preparation of the apparatus, including the audiovisual on-screen instructions, usually took less than 7 min and our brief version of the VFT took less than 3 min, which is less demanding for participants (10–15 min is necessary for the entire procedure).

NIRS is based on the principle that near-infrared light is preferentially absorbed by haemoglobin and less so by other tissues. Near-infrared light emitted from the skin travels into the body, is reflected and absorbed by the internal tissues and reappears on the skin. Thus, the absorption of near-infrared light reflects haemoglobin concentration ([Hb]) in the tissue placed beneath emission and detection probe pairs. Measurements taken using 2 or more wavelengths of near-infrared light enable the determination of [oxy-Hb] and [deoxy-Hb] changes because their absorptions are different at different wavelengths. The ETG-4000 system measures relative changes in [oxy-Hb] and [deoxy-Hb] using 2 wavelengths (695 and 830 nm) of infrared light, based on the modified Beer–Lambert law.(Yamashita et al., 1996) In this continuous-wave NIRS system, these [Hb] values include a differential pathlength factor (DPF); therefore, the unit of this form of NIRS measurement is mM·mm. The distance between pairs of source-detector probes was set to 3.0 cm and each measurement area located between pairs of source-detector probes was defined as one 'channel'. It is assumed that a machine in which the source-detector spacing is 3.0 cm measures points at a depth of 2–3 cm from the scalp (i.e., the surface of the cerebral cortex) (Okada and Delpy, 2003). The temporal resolution of NIRS was set to 0.1 s.

The arrangement of the probes measured relative [oxy-Hb] and [deoxy-Hb] signal changes in the bilateral prefrontal cortical area (i.e., dorso-lateral [Brodmann areas (BAs) 9 and 46], ventro-lateral [BAs 44, 45, and 47] and fronto-polar [BA 10] regions) and in the superior and middle temporal cortical surface regions, which was corroborated by a multi-individual study of anatomical cranio-cerebral correction via the international 10–20 system (Fig. 2 and Table S1) (Tsuzuki et al., 2007). However, in the 10–20 system, the anterior parts of the probes (e.g., Fpz) can be positioned precisely, whereas the position errors of more lateral probes might be increased due to inter-individual head size variability. In addition, although we initially aimed to analyse single-individual and single-channel levels in this study, studies of repeated NIRS measures have demonstrated acceptable reliability of the NIRS signal at the group and cluster levels, whereas retest reliability was unsatisfactory at the single-individual and single-channel levels (Schecklmann et al., 2008).

Table 1
Demographic and clinical characteristics of the 4 age- and gender-matched diagnostic groups at all 7 study sites.

	Major depressive disorder		Schizophrenia		Bipolar disorder		Healthy control		Group difference
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*-Value
n	153		136		134		590		
Age years	43.8	12.7	43.7	12.1	44.0	14.9	43.9	15.7	0.99
Gender, women/men	77/76		67/69		69/65		314/276		0.81[a]
Education, years	14.0	1.9	15.2	2.0	15.6	2.0	16.1	2.4	<0.01
Estimated premorbid IQ	106.0	10.1	103.7	11.2	106.9	8.6	107.2	10.1	0.23
Task performance	13.0	3.8	13.6	4.4	12.0	3.6	15.3	4.8	<0.01
Age at onset, years	39.2	11.3	23.4	7.4	32.9	12.4	–		
PANSS									
Positive	–		16.3	5.0	–		–		
Negative	–		21.0	6.0	–		–		
General psychopathology	–		37.0	7.6	–		–		
HAM-D	14.1	6.7	–		8.4	7.0	–		
YMRS	–		–		4.7	5.9	–		
GAF	53.9	9.7	47.3	11.4	55.5	13.3	–		

Abbreviations: IQ, intelligence quotient; PANSS, Positive and Negative Syndrome Scale; GAF, global assessment of functioning.
[a] Chi-square test was used for testing group difference. Otherwise, one-way ANOVA was used.

Therefore, instead of undertaking a full analysis at the single-individual and single-channel levels, here we performed an analysis of NIRS signals at the single-individual and cluster levels. A principal component analysis (PCA) of NIRS [oxy-Hb] signal changes in targeted fronto-temporal channels was performed at the initial study site as a preliminary analysis to capture a channel cluster of the analogous time-course pattern in HC individuals. Subsequently, the weight maps of the first and second principal component graphs were used to identify 2 cluster components.

These analyses suggested that 2 cluster components were identified and that the 2 clusters included the frontal region (11 channels) and the bilateral temporal region (10 channels each) (see Supplementary Material (II) and eFig. S1). The channels in these 2 respective regions of interest were averaged and transformed into representative 'Region 1 (R1)' and 'Region 2 (R2)' NIRS signals for each individual (Fig. 3). According to registration into the LONI Probabilistic Brain Atlas 40 (LBPA40) (Fig. 2) (see Supplementary Table S1 for LBPA40 anatomical labels) (Shattuck et al., 2008), the R1 NIRS signal consisted of signals from channels located approximately in the fronto-polar and dorsolateral prefrontal cortical regions (i.e., superior and middle frontal gyri), whereas the R2 NIRS signal consisted of signals from channels located approximately in the ventro-lateral prefrontal cortex and the superior and middle temporal cortical regions (i.e., inferior frontal gyrus and superior and middle temporal gyri).

An automatic artefact-rejection procedure (see Supplementary Material (III)) was followed and individual data were excluded when there were fewer than 6 remaining channels from each of the 2 cluster regions (Flow diagram (4)).

Statistical analyses

Taking into consideration the potential application of the technique in ordinary clinical settings and personalised care, a conservative receiver operating characteristic (ROC) analysis was performed and used to generate simple indices of NIRS signal patterns, to aid individual diagnoses.

The spatiotemporal characteristics of the frontal and temporal haemodynamic responses induced by VFT were assessed and subsequently applied to an algorithm using the simplest and fewest variables for differential diagnosis. Because previous studies have shown that the best way to differentiate patients with MDD from those with BP or SZ is to describe the time-course of changes in the NIRS signal associated with the VFT (Kameyama et al., 2006; Suto et al., 2004), we chose to create 2 simple visual indices, referred to here as 'integral value' and 'centroid value', to capture these time-course changes.

The integral value describes the size of the haemodynamic response during the 60 s activation task period, whereas the centroid value serves as an index of time-course changes throughout the task, with periods representing the timing of the haemodynamic response. The centroid value is indicated by a time shown with a perpendicular line from the centroid of the NIRS signal change area (calculated as a positive change) throughout the task periods (from 0 [s] to 125 [s]

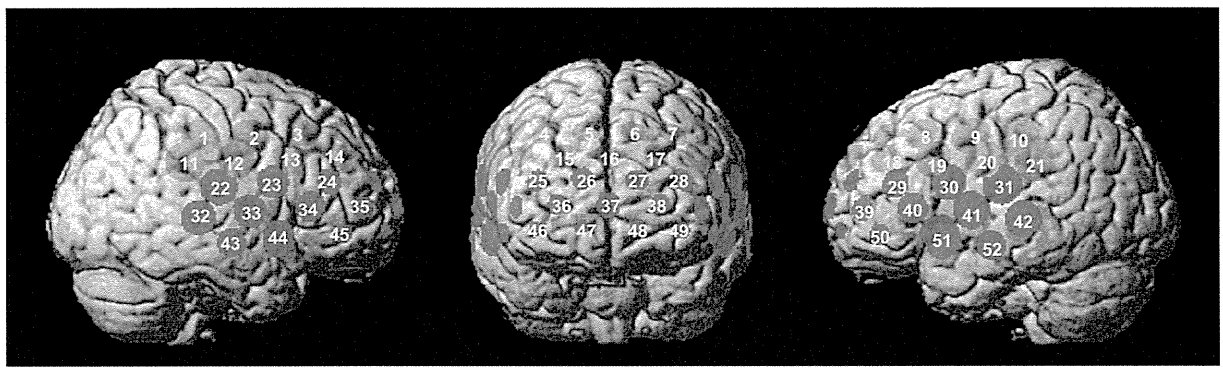


Fig. 2. Regions of interest (Regions 1 and 2) of the near-infrared spectroscopy (NIRS) signals. The locations of near-infrared spectroscopy (NIRS) measurements were probabilistically estimated and anatomically labelled in the standard brain space (LBPA40) according to Tsuzuki et al. (2007). Region 1: (ch 25–28, ch 36–38 and ch 46–49); Region 2, Right: (ch 22–24, ch 32–35 and ch 43–45); Left: (ch 29–31, ch 39–42 and ch 50–52).

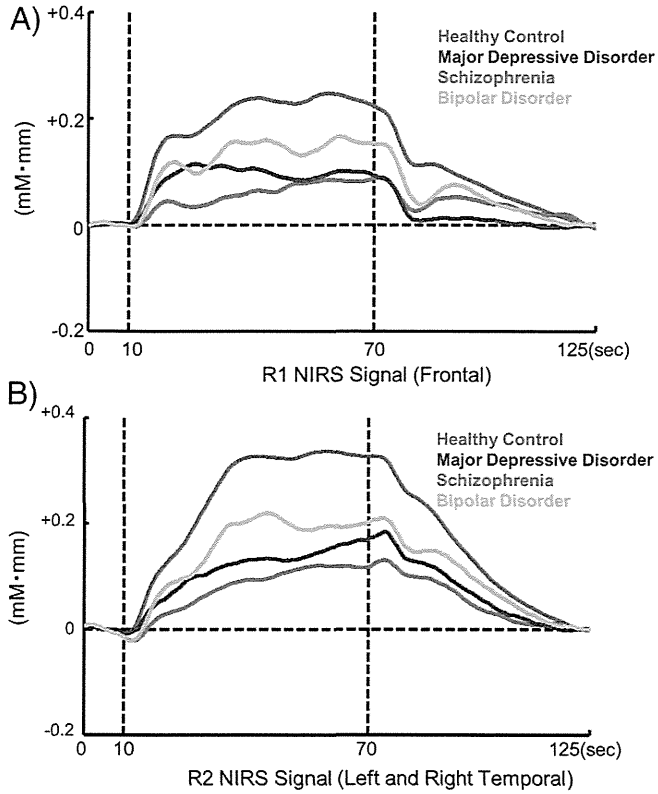


Fig. 3. Time courses of the haemodynamic responses in Region 1 (R1) and Region 2 (R2) in the 4 diagnostic groups. Panels A and B show the time courses of the haemodynamic responses in R1 and R2, respectively.

(= 10 [s] + 60 [s] + 55 [s])) (see Fig. 1). To confirm the reproducibility of each single index between the 2 measurements, a test–retest analysis (single-measure intra-class correlation (ICC) analysis using a one-way random effect model) revealed the presence of significant intra-class correlation coefficients for both the R1 and R2 integral values [r = 0.47, $p = 0.01$; r = 0.59, $p < 0.01$, respectively] and the R1 centroid value [r = 0.65, $p < 0.01$], but not for the R2 centroid value [r = 0.20, $p = 0.19$] (see Supplementary material (IV)). PCA and ICC analyses revealed that the 2 indices of NIRS analysis during VFT were acceptable at the single-individual and cluster levels. Thus, the 3 significant variables were used for further analysis.

The 2 representative R1 and R2 NIRS signals obtained from each individual were averaged separately for each type of [Hb] and the integral and centroid values were calculated using parametric statistical tests. Further analyses focused on the increases in [oxy-Hb], because these appear to reflect task-related cortical activation more directly than do decreases in [deoxy-Hb], as evidenced by the stronger correlation between the former and the blood oxygenation level-dependent signal measured by fMRI (Strangman et al., 2002b) and by the results of animal studies (Hoshi et al., 2001). As the typical [oxy-Hb] activation pattern had a positive direction (Fig. 1), data with positive [oxy-Hb] changes (i.e., data with an integral value ≥ 0) in R1 and R2 were used to create an algorithm (Flow diagram (4)). Data exhibiting negative [oxy-Hb] changes were added to the analysis and were described in the results as being appropriate. The analysis of [deoxy-Hb] changes was reported in Supplementary Material (V); however, no significant variable was found regarding [deoxy-Hb] changes.

First, as a preliminary analysis to identify the variable that differentiates patients with psychiatric diseases from HCs most robustly, the 3 variables, including both integral and centroid values of the R1 NIRS signal and the integral value of the R2 NIRS signal, were compared among all of the patients and the age- and gender-matched controls at the initial study site using ANOVA. The resulting significant variables were applied to ROC analyses at the remaining 6 sites.

Because mental health professionals in real clinical settings must differentiate patients with MDD from those with BP or SZ presenting with depression as accurately as possible, the second main analysis performed here aimed to determine the most informative variable and the optimal threshold to discriminate patients with MDD from those with non-MDD disorders. In the present study, the 3 variables, including both integral and centroid indices of R1 and the integral R2 index of the NIRS signal, were compared among patients with MDD and those with either of the other 2 disorders using ANOVA; the variables that were deemed to be significant were applied to the ROC analysis. The preliminary data from the initial site were used to determine an optimal threshold, which was then validated using the test data from the remaining 6 sites.

Third, Pearson's correlation analysis was performed between the significant variables and demographic confounding factors. Data were tested for a normal distribution using the Kolmogorov–Smirnov test. Data that were not normally distributed were analysed using Spearman's correlation analysis.

In particular, regarding clinical confounding factors, such as symptoms (HAMD, YMRS and PANSS scores) and medication doses (antidepressants: imipramine (IMP) equivalent dose; antipsychotics: chlorpromazine (CPZ) equivalent dose; anxiolytics: diazepam equivalent dose; and anti-parkinsonian drugs: biperiden equivalent dose, lithium dose, sodium valproate dose and carbamazepine dose), a stepwise multiple linear regression analysis was performed with a probability of F for conservative entry and removal criteria of 0.01 and 0.05, respectively, to elucidate the complicated relationships among these clinical confounding factors in each diagnostic group.

All data are expressed as mean and standard deviation (SD). The significance level was set to alpha = 0.05. When a difference was considered significant, we presented both the effect size (Cohen's d) and the 95% confidence interval (CI). Statistical analyses were performed using the SPSS 16.0.1J software (SPSS Inc., Tokyo, Japan).

Results

Demographic characteristics

Table 1 shows the demographic and clinical characteristics of the 4 age- and gender-matched diagnostic groups used in this study. One-way ANOVA revealed an absence of significant age differences among the groups ($p = 0.99$) and a chi-squared test showed an absence of gender differences among the groups ($p = 0.81$). In addition, the age and gender distributions among the 4 diagnostic groups were not significantly different at the initial site (Gunma University, MDD: 39.9 (11.7) y.o., 12/15; BP: 41.1 (13.2) y.o., 22/15; SZ: 40.1 (14.9) y.o., 11/20; and HC: 40.0 (4.2) y.o., 7/10) (age, $p = 0.98$; gender, $p = 0.24$) and at the other 6 sites (MDD: 44.6 (12.7) y.o., 65/61; BP: 45.1 (15.4) y.o., 47/50; SZ: 44.8 ± 11.0 y.o., 56/49; and HC: 44.0 ± 15.9 y.o., 307/266) (age, $p = 0.89$; gender, $p = 0.81$).

Preliminary test of the difference between HCs and patients

Although it was not the main theme of this study, to compare our results with those of studies of biomarkers performed only to detect functional abnormalities in patients against a control group, we also analysed the differences between HC individuals and patients to confirm the significance of the 3 variables chosen for analysis. Full analyses are described in Supplementary Material (VI).

From the analyses performed using data from the initial site, we adopted both R1 and R2 integral values as statistically significant variables for the algorithm. Thresholds were dependent on the

purpose for which the variables were used (Table 2). For example, if the optimal thresholds of the integral values of R1 and R2 derived from the initial site (73 and 104) were applied to the independent test data from the remaining 6 sites, the sensitivities were 0.73 (proportion of patients/measurement: 96/131) and 0.79 (104/131) and the specificities were 0.63 (proportion of HCs/measurement: 326/514) and 0.63 (324/514) for R1 (positive predictive value (PPV) = 0.37, negative predictive value (NPV) = 0.90) and R2 (PPV = 0.40, NPV = 0.92), respectively.

Test for differentiation of patients with unipolar MDD from those with BP and SZ

Using the preliminary data from the initial site, one-way ANOVA performed between the patients with MDD and those with one of the other 2 disorders of interest (BP or SZ) revealed a significant difference in the R1 centroid values [$F(1,53) = 9.54, p < 0.01$; $d = 0.96$, 95% CI, (0.25 to 1.62)], but not in the R1 [$F(1,53) = 0.14, p = 0.71$] or the R2 [$F(1,53) = 0.05, p = 0.83$] integral values.

As the significant R1 centroid value proved to be the most useful variable, we applied it to ROC analysis for the differentiation of patients with unipolar MDD from those with non-MDD disorders. The resulting area under the ROC curve (Az) value was 0.74 [95% CI, (0.61 to 0.87)] and the optimal threshold was 54 [s] from the extreme top left point of the ROC curve (eFig. S4).

To validate the optimal threshold calculated, we applied it to the independent test data of the remaining 6 sites, to differentiate the patients with MDD from those with SZ and BP [Az = 0.81, 95% CI, (0.74 to 0.89); $d = 1.17$, 95% CI, (0.79 to 1.54); optimal threshold = 54 [s], PPV = 0.79, NPV = 0.82; Fig. 4]. Using this threshold (54 [s]), 74.6% of the patients with MDD (proportion of patients/measurement: 41/55) and 85.5% of those with SZ or BP (65/76) were classified correctly [76.9% of BP patients (20/26) and 90.0% of SZ patients (45/50)] (Fig. 5). The ROC curves of MDD v. BP [Az = 0.74, 95% CI, (0.62 to 0.85); $d = 0.81$, 95% CI, (0.32 to 1.29); optimal threshold = 54 [s], PPV = 0.87, NPV = 0.59] and MDD v. SZ [Az = 0.86, 95% CI, (0.78 to 0.93); $d = 1.40$, 95% CI, (0.96 to 1.82); optimal threshold = 54 [s], PPV = 0.89, NPV = 0.78] are shown separately in eFig. S5.

For reference, the test performed for the differentiation between patients with BP and those SZ is shown in Supplementary Material (VII).

Correlational analysis of demographic and clinical confounding factors

Correlational analysis showed no significant correlations between any of the significant dependent variables (among the R1 and R2 integral values and the R1 centroid value of NIRS signals) and any of the demographic confounding factors [performance (number of correct words), education years and pre-morbid IQ; $p > 0.05$] for all patients with psychiatric disorders (MDD, BP and SZ).

Regarding clinical confounding factors, a stepwise regression analysis of each significant dependent variable for each disorder revealed that there was no entry clinical variable in the linear regression models, with the exception of the global assessment of functioning (GAF) score (beta = 0.50, $p < 0.01$) for the R2 integral value ($F = 10.73$, $p < 0.01$; $R = 0.50$, $R^2 = 0.25$, adjusted $R^2 = 0.23$) in patients with MDD, and the GAF score (beta = 0.58, $p = 0.01$) for the R2 integral value ($F = 8.43$, $p = 0.01$; $R = 0.59$, $R^2 = 0.35$, adjusted $R^2 = 0.30$) in patients with BP who exhibited depressive symptoms. Thus, only one clinical variable (i.e., GAF score) among all of the medication and clinical variables examined had a significant impact on the R2 integral values for patients with MDD or BP who exhibited depressive symptoms.

Table 2
Sensitivities and specificities of the integral values of Region 1 (R1) and Region 2 (R2) signals between healthy controls and all patients with psychiatric disorders, based on the independent data collected from the 6 additional sites.

Integral value	R1		R2	
	Sensitivity	Specificity	Sensitivity	Specificity
160	0.95	0.27	0.90	0.39
150	0.93	0.30	0.89	0.43
140	0.92	0.34	0.88	0.47
130	0.90	0.37	0.87	0.52
120	0.88	0.42	0.85	0.57
110	0.85	0.46	0.82	0.61
100	0.82	0.50	0.78	0.64
90	0.78	0.54	0.76	0.68
80	0.74	0.61	0.73	0.73
70	0.72	0.65	0.64	0.76
60	0.66	0.71	0.57	0.78
50	0.57	0.75	0.52	0.81
40	0.47	0.80	0.43	0.86
30	0.38	0.84	0.33	0.89

Discussion

The present multi-site study is the first large-scale, case–control study that demonstrates the utility of NIRS for the differential diagnosis of major psychiatric disorders. The main strengths of this study include the application of a neuroimaging biomarker in clinical practice that allows the clinically useful differential diagnosis of depressive states. The frontal centroid value, which represents the timing of frontal NIRS signal patterns, was a significant variable for differential diagnosis and the optimal threshold derived from the ROC analysis correctly discriminated patients with unipolar MDD (74.6%) from those with non-MDD disorders (85.5%; BP, 76.9% and SZ, 90.0%).

Single-individual diagnostic classification analyses among various psychiatric disorders

The present study was not only a case–control study of group comparisons, but also a study specifically designed for examining the practical utility of single-individual diagnostic classification in various psychiatric disorders. Several studies have reported the single-individual diagnostic classification of one psychiatric disorder compared with HCs by applying multivariate statistical methods (e.g.,

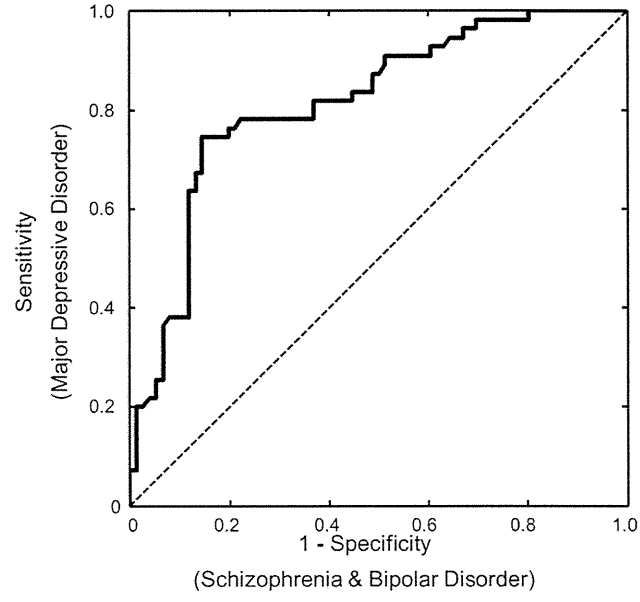


Fig. 4. Receiver operating characteristic analysis of the centroid value of Region 1 (R1) near-infrared spectroscopy signal between patients with major depressive disorder and those with either of the other 2 disorders of interest (bipolar disorder and schizophrenia) based on the independent data collected from the 6 additional sites.

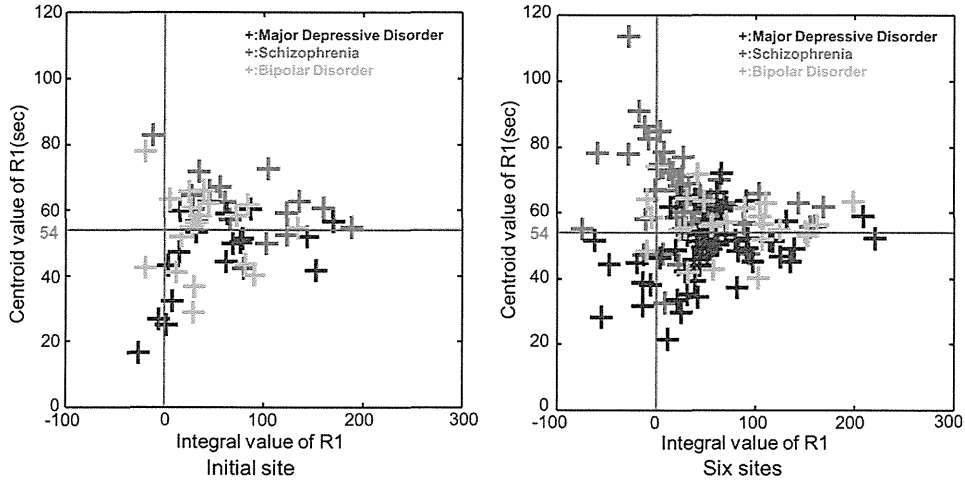


Fig. 5. Scatter plots of the centroid and integral values of Region 1 (R1) signal in the patients, both at the initial site (Gunma University) and at the 6 additional sites.

neuroanatomical pattern classification) to structural MRI data (Davatzikos et al., 2005) and NIRS data (Hahn et al., 2013) from SZ and high-risk psychosis samples (Koutsouleris et al., 2009), as well as to functional MRI data from patients with depression (Hahn et al., 2011). These studies were technically sophisticated; however, more research must be performed to test their reproducibility and generalisability in the advanced stage of clinical application, because (1) they were designed for the analysis of one diagnostic classification based on comparison to HCs, and not for differential diagnosis among multiple psychiatric disorders; and (2) they were performed using one relatively small cohort; thus, they must be replicated in another cohort including larger sample groups.

Furthermore, we will discuss briefly our results in comparison with those of other single-individual diagnostic classification studies (Davatzikos et al., 2005; Fu et al., 2008; Hahn et al., 2011; Koutsouleris et al., 2009). We used only a single variable (simple 'centroid value' of NIRS signals) and found that the classification rates (unipolar MDD: 74.6% correct classification; the 2 other disorders: 85.5% correct classification (BP, 76.9%; SZ, 90.0%)) were almost equivalent to the rates reported in the previous MRI studies using multivariate statistical methods (which had 80–90% classification rates in the patient group compared with the HC group).

To determine whether a higher disease classification rate could be achieved by using a multivariate pattern analysis (compared with that obtained using one simple variable), which was used in previous MRI studies, we confirmed the results using the multivariate pattern classification analysis described in Supplementary Material (VIII). The leave-one-out cross-validation method revealed that 4 significant variables, or even one variable (the R1 centroid value), could differentiate patients with unipolar MDD from those with either of the 2 other disorders (non-MDD) with a similar degree of mean accuracy (76.8% (unipolar MDD: 73.0% (54/74), non-MDD: 74.8% (83/111))).

Clinical importance and implications

Another clinically valuable feature of our work is that it aimed to facilitate diagnosis among patients with similar depressive symptoms, which psychiatrists often find to be a difficult task. Most BP patients with depressive symptoms are initially diagnosed with and treated for MDD (Akiskal et al., 1995; Goldberg et al., 2001). Therefore, our findings may help differentiate BP with depressive symptoms from MDD. Depressive symptoms and cognitive deficits are also common early signs of SZ (Hafner et al., 2005). Of particular clinical relevance is the observation that SZ patients with concomitant depression have a greater risk of suicide or an unfavourable disease course (an der Heiden et al., 2005). Therefore, sufficient attention must be given to the diagnosis and treatment of depression in SZ patients.

The results of the present study may draw attention to the heterogeneity observed among MDD patients. Rather than simply being misclassified, approximately 25% of patients with unipolar MDD who were classified by the system as having a non-MDD disorder may have a brain pathophysiology that is biologically different from that of the majority of MDD patients. Evidence suggests that 25–50% of individuals with recurrent major depression (particularly those in atypical early-onset or treatment-refractory subgroups) may in fact have broadly defined BP (Angst, 2007). In this study, 74.6% of the patients with MDD were classified correctly; the remaining 25.4% might include either patients who would progress to a diagnosis of one of the 2 other disorders or patients with a broadly defined BP who were diagnosed with MDD according to the DSM criteria. This explanation might be justified by the finding of a correct classification rate of 75% for patients with MDD. For practical purposes, among patients diagnosed clinically with MDD, the early suspicion of the possibility of a diagnosis of a non-MDD disorder with depression would also provide an opportunity to reduce the hazardous effects of the illness on personal, social and occupational aspects; therefore, our results should be of great clinical importance in practical applications. Thus, a prospective study aimed at elucidating the heterogeneity of unipolar MDD is required.

Advantages of the NIRS method

We used the same NIRS system (a non-invasive, portable and user-friendly device) and the same concise measurement procedure at every site; therefore, inter-site compatibility was not an issue here; however, it may be an obstacle in other neuroimaging multi-site studies. Furthermore, we used a high temporal resolution (0.1 s) in the NIRS system for measuring time-specific characteristics of dynamic prefrontal cortical functions; this enabled analyses that included more detailed time-course comparisons of NIRS signal changes. We created and adopted new variables, such as the 'centroid value', to determine the timing of the haemodynamic response (Fig. 1). The high temporal resolution of NIRS might allow not only the detection of functional abnormalities (e.g., hypofrontality), but also the capture of the specific haemodynamic activation time courses of each psychiatric disorder and aid differential diagnosis.

The practical application of biomarkers requires that they be relatively simple. The simplicity of both the test procedure and the associated data analysis is important not only for the participants, but also for their caretakers and clinicians. Therefore, rather than using complicated multivariate statistical methods, we developed a robust classification algorithm for real-time visual evaluation of patients using the simplest, and lowest number of variables on the basis of a ROC analysis. This was important because we sought to develop a psychiatric practice empowered by the initiative of patients by sharing the 'comprehensively visualised' results that can be easily recognisable by patients and caretakers, rather than results from complicated 'black-box' analyses. In addition, using the condensed VFT (<3 min) developed previously by us, we designed a diagnostic support system in a way that the results are available to clinicians in less than 15 min. The availability of such a 'comprehensively visualised' report to clinicians, patients and their caretakers at a first visit, while laying out a future treatment plan, would likely lead to a paradigm shift to a patient-centred approach in clinical psychiatry.

Limitations

The methodological aspects of the present study warrant commentary. First, most of the patients included in the study were taking medications at the time of measurement. To our knowledge, no clear evidence of the effects of medication on NIRS signals has been demonstrated. We found that none of the medications at any dose was significantly correlated with NIRS signals in this study; however, we cannot fully exclude the effects of medication on haemodynamic signals. For confirmation, the application of the algorithm described above (optimal cut-off of the R1 centroid value) to the drug-free patients exclusively, 6 out of 10 patients with MDD patients (60%) and 4 out of 5 patients with SZ (80%) were classified correctly. Second, the size of the sample included in our final analysis was substantially reduced from that initially recruited, because we tried to minimise the confounding factors of age and gender by matching the groups and excluded patients in remission, as well as patients in the manic phase (see Flow diagram). In our confirmatory analysis, we included all non-matched and in-remission patients and found that the results were quite similar, although this analysis had a lower detection power. The optimal threshold of the sample sets before demographical matching was also the same as that calculated originally. These results suggest that the reduction in the total number of study participants after demographical matching did not affect the development of the algorithm [see Supplementary Material (I)]. However, we must consider the possibility that this diagnostic support system is best suited for young and middle-aged patients with moderate or severe symptoms (e.g., aged between 23 and 65 years (mean ± 1.5 SD)). Third, a PCA of haemodynamic response performed to capture a channel cluster led to the identification of 2 cluster regions. Nonetheless, as we thought that pooling many NIRS signals together into only 2 representative regions of interest (R1 (frontopolar and dorsolateral prefrontal regions) and R2 (ventrolateral prefrontal and temporal regions)) might oversimplify the results (see the Discussion of Supplementary Material (II)), we sought to confirm the reliability of the 2 clusters by performing a test–retest analysis in a portion of the samples. We found significant ICCs for both the R1 and R2 integral values and for the R1 centroid value between 2 measurements (see Supplementary Material (IV)). Therefore, we used the two data-derived clusters that reflected a fronto-temporal haemodynamic response during VFT. Fourth, we have controlled some well-known confounders in the analyses. However, the NIRS signal might be affected by the other systemic confounders, such as autonomic function, neuroendocrine function, diet and physical activity. In addition, brain anatomical factors, such as scalp–cortex distance and frontal sinus volume, as well as genetic variants might also be potential confounders. Further studies are required to address the relationship between the NIRS signal and these confounders. If these findings are fully replicated, the development of methods of integrating confounding factors into NIRS signal in the future will be ideal. Fifth, we did not use the exclusion criterion of first-degree relatives with axis I psychiatric disorders for healthy controls. This could give a bias to the data in healthy controls, which means that some of the first-degree relatives of persons with axis I psychiatric disorders might have been included as a healthy control in the present study. However, as the same situations are assumed in real clinical settings, we daringly recruited healthy controls without applying that strict exclusion criterion.

Conclusions and future implications

In conclusion, this multi-site study provided evidence that the fronto-temporal NIRS signal may be used as a tool in assisting the diagnosis of major psychiatric disorders with depressive symptoms. Future NIRS research should be performed to study the applicability of this method to (1) the identification of a need for therapy, (2) the assessment of the efficacy of various treatments, (3) the establishment of prognostic predictions that may be clarified by longitudinal follow-up assessments of patients in various clinical stages and (4) the examination of the use of NIRS as a screening tool.

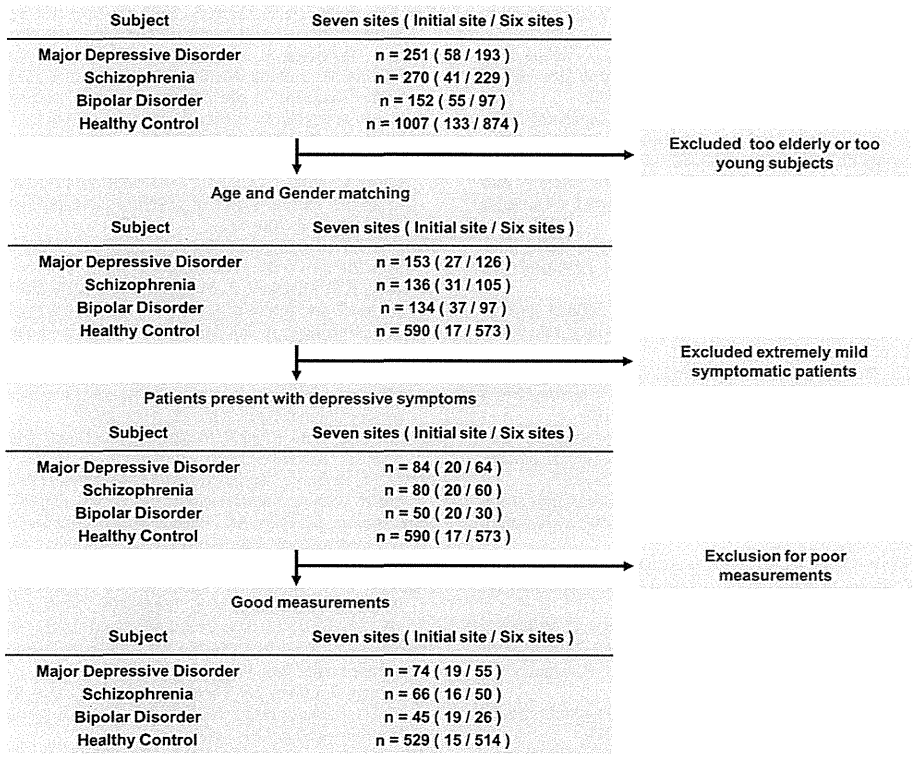
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.05.126.

Acknowledgments

The authors would like to thank all the participants in this study. The authors also thank Makoto Ito, M.D., Ph.D., Tomohiro Suto, M.D., Ph.D., Yutaka Yamagishi, M.D., Naoki Hanaoka, M.D., Ph.D., Toshimasa Sato, Ed.M. and Noriko Sakurai of Gunma University; Kohei Marumo, M.D., Ph.D. and Yuki Kawakubo, Ph.D., of The University of Tokyo; Hitomi Kobayashi, Ph.D. and Junko Motoki, Ph.D., of Showa University; Osamu Saitoh, M.D., Ph.D., Kimitaka Anami, M.D., Ph.D., Yohtaro Numachi, M.D., Ph.D., Yuji Sugimura and Masaru Ogawa of the National Centre of Neurology & Psychiatry Hospital; Daisuke Gotoh, M.D., Yojiro Sakai, M.D., Ph.D. and Emi Yoshida, M.D., of Fukushima Medical University, for technical assistance during data collection.

Role of the funding source

This study was supported in part by grants-in-aid for scientific research from the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (Nos. 18390318, 19659289, 20390310 and 22659209 to MF; Nos. 21249064 and 221S003 [Innovative Areas (Comprehensive Brain Science Network)] to KK; No. 17019029 [Priority Areas (Applied Genomics)] to YO; and No. 23791309 to RT), and by grants-in-aid from the Ministry of Health, Labour, and Welfare (MHLW) of Japan (H20-kokoro-ippan-001, H20-3, and H22-seishin-ippan-015 to KK). This study was also supported in part by Health and Labour Sciences Research Grants for Comprehensive Research on Disability Health and Welfare (previously, Health and Labour Science Research Grant for Research on Psychiatric and Neurological Disease and Mental Health; H20-001 to MF and H23-seishin-ippan-002 to RT) and by the Intramural Research Grants for Neurological and Psychiatric Disorders of The National Centre of Neurology and Psychiatry (previously, The Research Grant for Nervous and Mental Disorders from the MHLW; 21B-1 to MF and 20B-3 to TN). In addition, this study was supported in part by an Intramural Research Grant for Neurological and Psychiatric Disorders of NCNP (No. 23-10 to RT). A part of this study was also the result of a project entitled 'Development of biomarker candidates for social behavior', which was carried out under the Strategic Research Program for Brain Sciences by MEXT, Japan. This study was also supported in part



Flow diagram. Main analyses were based on matched samples according to the flow diagram.

by grants from the Japan Research Foundation for Clinical Pharmacology (to RT).

The funders had no role in the design of the study, data collection, data analysis, decision to publish, or preparation of the manuscript. The corresponding author had full access to all of the data in this study and takes responsibility for the integrity of the data, the accuracy of the data analysis and the decision to submit the manuscript for publication.

Author contributions

Masato Fukuda designed the experiments and organized the multi-site collaborative study. Ryu Takizawa, Masato Fukuda, Shingo Kawasaki, and Kiyoto Kasai analysed the data and wrote the first draft of the paper. The other contributors performed data acquisition and revised the first draft critically for important intellectual content. All contributors have approved the final version of the manuscript.

Contributors from the Joint Project for Psychiatric Application of Near-Infrared Spectroscopy (JPSY-NIRS) Group

Gunma University: Masato Fukuda, M.D., Ph.D., Masashi Suda, M.D., Ph.D., Yuichi Takei, M.D., Ph.D., Yoshiyuki Aoyama, M.D., Ph.D., Kosuke Narita, M.D., Ph.D., Masahiko Mikuni, M.D., Ph.D., Masaki Kameyama, M.D., Ph.D., and Toru Uehara, M.D., Ph.D.

The University of Tokyo: Ryu Takizawa, M.D., Ph.D., Kiyoto Kasai, M.D., Ph.D., Masaru Kinou, M.D., Ph.D., Shinsuke Koike, M.D. and Ayaka Ishii-Takahashi, M.D.

Hitachi Medical Corporation: Shingo Kawasaki, MS, Noriyoshi Ichikawa, BS, and Michiyuki Fujiwara, AS.

Showa University: Haruhisa Ohta, M.D., Ph.D., Hiroi Tomioka, M.D., Ph.D., Bun Yamagata, M.D., Ph.D., and Kaori Yamanaka, M.D.

Keio University: Masaru Mimura, M.D., Ph.D.

Tottori University: Shenghong Pu, Ph.D.

The National Centre of Neurology and Psychiatry: Kazuyuki Nakagome, M.D., Ph.D., Takamasa Noda, M.D., Taro Matsuda, M.D., and Sumiko Yoshida, M.D., Ph.D.

Fukushima Medical University: Shin-Ichi Niwa, M.D., Ph.D., Soichi Kono, M.D., Hirooki Yabe, M.D., Ph.D., and Sachie Miura, M.D.

Mie University and Tokyo Metropolitan Matsuzawa Hospital: Yuji Okazaki, M.D., Yukika Nishimura, Ph.D., Hisashi Tanii, M.D., Ph.D., Ken Inoue, M.D., Ph.D., Chika Yokoyama, MA, Yoichiro Takayanagi, M.D., Ph.D., Katsuyoshi Takahashi, M.D., and Mayumi Nakakita, M.D.

Conflict of interest

We would like to disclose potential conflicts of interest regarding all financial and material support for the present study. The principal investigators of each site (Masato Fukuda of Gunma University, Kiyoto Kasai of The University of Tokyo, Masaru Mimura of Keio university and Showa University, Kazuyuki Nakagome of The National Center of Neurology and Psychiatry and Tottori University, Shin-ichi Niwa of Fukushima Medical University, Yuji Okazaki of both Mie University and Tokyo Metropolitan Matsuzawa Hospital and Takamasa Noda of The National Center of Neurology and Psychiatry) have potential conflicts of interest in the submitted work. Each site has had an official contract with the Hitachi Group (Advanced Research Laboratory, Hitachi, Ltd., and The Research and Developmental Center, Hitachi Medical Corporation) for a collaborative study of the clinical application of NIRS in psychiatric disorders. For this study, the Hitachi Group provided a project grant (JPY 300,000–2,000,000 per year) and material support (temporary rental of a NIRS (Optical Topography) ETG-4000 system) for each site. Shingo Kawasaki, Noriyoshi Ichikawa and Michiyuki Fujiwara are employees of Hitachi Medical Corporation.

The other authors have no financial relationships with any organisations that might have an interest in the submitted work in previous years and no other relationships or activities that could appear to have influenced the submitted work.

References

Akiskal, H.S., Maser, J.D., Zeller, P.J., Endicott, J., Coryell, W., Keller, M., Warshaw, M., Clayton, P., Goodwin, F., 1995. Switching from 'unipolar' to bipolar II. An 11-year

prospective study of clinical and temperamental predictors in 559 patients. Arch. Gen. Psychiatry 52, 114–123.
Almeida, J.R., Versace, A., Mechelli, A., Hassel, S., Quevedo, K., Kupfer, D.J., Phillips, M.L., 2009. Abnormal amygdala–prefrontal effective connectivity to happy faces differentiates bipolar from major depression. Biol. Psychiatry 66, 451–459.
an der Heiden, W., Konnecke, R., Maurer, K., Ropeter, D., Hafner, H., 2005. Depression in the long-term course of schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. 255, 174–184.
Angst, J., 2007. The bipolar spectrum. Br. J. Psychiatry 190, 189–191.
Barch, D.M., Sheline, Y.I., Csernansky, J.G., Snyder, A.Z., 2003. Working memory and prefrontal cortex dysfunction: specificity to schizophrenia compared with major depression. Biol. Psychiatry 53, 376–384.
Barrett, S.L., Mulholland, C.C., Cooper, S.J., Rushe, T.M., 2009. Patterns of neurocognitive impairment in first-episode bipolar disorder and schizophrenia. Br. J. Psychiatry 195, 67–72.
Costafreda, S.G., Fu, C.H., Lee, L., Everitt, B., Brammer, M.J., David, A.S., 2006. A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. Hum. Brain Mapp. 27, 799–810.
Curtis, V.A., Dixon, T.A., Morris, R.G., Bullmore, E.T., Brammer, M.J., Williams, S.C., Sharma, T., Murray, R.M., McGuire, P.K., 2001. Differential frontal activation in schizophrenia and bipolar illness during verbal fluency. J. Affect. Disord. 66, 111–121.
Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E., 2005. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry 62, 1218–1227.
Ferrari, M., Quaresima, V., 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. Neuroimage 63 (2), 921–935.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1997. Structured Clinical Interview for DSM-IV Axis I Disorders. Biometric Research Department, New York State Psychiatric Institute, New York, U.S.A.
Fu, C.H., Mourao-Miranda, J., Costafreda, S.G., Khanna, A., Marquand, A.F., Williams, S.C., Brammer, M.J., 2008. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol. Psychiatry 63, 656–662.
Goldberg, J.F., Harrow, M., Whiteside, J.E., 2001. Risk for bipolar illness in patients initially hospitalized for unipolar depression. Am. J. Psychiatry 158, 1265–1270.
Gur, R.E., Keshavan, M.S., Lawrie, S.M., 2007. Deconstructing psychosis with human brain imaging. Schizophr. Bull. 33, 921–931.
Hafner, H., Maurer, K., Trendler, G., an der Heiden, W., Schmidt, M., 2005. The early course of schizophrenia and depression. Eur. Arch. Psychiatry Clin. Neurosci. 255, 167–173.
Hahn, T., Marquand, A.F., Ehlis, A.C., Dresler, T., Kittel-Schneider, S., Jarczok, T.A., Lesch, K.P., Jakob, P.M., Mourao-Miranda, J., Brammer, M.J., Fallgatter, A.J., 2011. Integrating neurobiological markers of depression. Arch. Gen. Psychiatry 68, 361–368.
Hahn, T., Marquand, A.F., Plichta, M.M., Ehlis, A.C., Schecklmann, M.W., Dresler, T., Jarczok, T.A., Eirich, E., Leonhard, C., Reif, A., Lesch, K.P., Brammer, M.J., Mourao-Miranda, J., Fallgatter, A.J., 2013. A novel approach to probabilistic biomarker-based classification using functional near-infrared spectroscopy. Hum. Brain Mapp. 34, 1102–1114.
Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.
Holsboer, F., 2008. How can we realize the promise of personalized antidepressant medicines? Nat. Rev. Neurosci. 9, 638–646.
Hoshi, Y., Kobayashi, N., Tamura, M., 2001. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. J. Appl. Physiol. 90, 1657–1662.
Kameyama, M., Fukuda, M., Yamagishi, Y., Sato, T., Uehara, T., Ito, M., Suto, T., Mikuni, M., 2006. Frontal lobe function in bipolar disorder: a multichannel near-infrared spectroscopy study. NeuroImage 29, 172–184.
Kay, S.R., Opler, L.A., Fiszbein, A., 1991. Positive and Negative Syndrome Scale (PANSS) Rating Manual. Multi-health Systems Inc., Toronto.
Koike, S., Takizawa, R., Nishimura, Y., Takano, Y., Takayanagi, Y., Kinou, M., Araki, T., Harima, H., Fukuda, M., Okazaki, Y., Kasai, K., 2011. Different hemodynamic response patterns in the prefrontal cortical sub-regions according to the clinical stages of psychosis. Schizophr. Res. 132, 54–61.
Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Moller, H.J., Gaser, C., 2009. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch. Gen. Psychiatry 66, 700–712.
Mathers, C.D., Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, e442.
McGorry, P.D., Killackey, E., Yung, A., 2008. Early intervention in psychosis: concepts, evidence and future directions. World Psychiatry 7, 148–156.
Obrig, H., Villringer, A., 2003. Beyond the visible—imaging the human brain with light. J. Cereb. Blood Flow Metab. 23, 1–18.
Okada, E., Delpy, D.T., 2003. Near-infrared light propagation in an adult head model. II. Effect of superficial tissue thickness on the sensitivity of the near-infrared spectroscopy signal. Appl. Opt. 42, 2915–2922.
Phillips, M.L., Vieta, E., 2007. Identifying functional neuroimaging biomarkers of bipolar disorder: toward DSM-V. Schizophr. Bull. 33, 893–904.
Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M.R., Rahman, A., 2007. No health without mental health. Lancet 370, 859–877.
Schecklmann, M., Ehlis, A.C., Plichta, M.M., Fallgatter, A.J., 2008. Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. Neuroimage 43 (1), 147–155.
Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080.
Strangman, G., Boas, D.A., Sutton, J.P., 2002a. Non-invasive neuroimaging using near-infrared light. Biol. Psychiatry 52, 679–693.
Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A., 2002b. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage 17, 719–731.
Suto, T., Fukuda, M., Ito, M., Uehara, T., Mikuni, M., 2004. Multichannel near-infrared spectroscopy in depression and schizophrenia: cognitive brain activation study. Biol. Psychiatry 55, 501–511.
Takizawa, R., Kasai, K., Kawakubo, Y., Marumo, K., Kawasaki, S., Yamasue, H., Fukuda, M., 2008. Reduced frontopolar activation during verbal fluency task in schizophrenia: a multi-channel near-infrared spectroscopy study. Schizophr. Res. 99, 250–262.
Tsuzuki, D., Jurcak, V., Singh, A.K., Okamoto, M., Watanabe, E., Dan, I., 2007. Virtual spatial registration of stand-alone fNIRS data to MNI space. NeuroImage 34, 1506–1518.
Yamashita, Y., Maki, A., Ito, Y., Watanabe, E., Koizumi, H., 1996. Noninvasive near-infrared topography of human brain activity using intensity modulation spectroscopy. Opt. Eng. 35, 1046–1049.
Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, 429–435.
Zanelli, J., Reichenberg, A., Morgan, K., Fearon, P., Kravariti, E., Dazzan, P., Morgan, C., Zanelli, C., Demjaha, A., Jones, P.B., Doody, G.A., Kapur, S., Murray, R.M., 2010. Specific and generalized neuropsychological deficits: a comparison of patients with various first-episode psychosis presentations. Am. J. Psychiatry 167, 78–85.
Zimmermann, P., Bruckl, T., Nocon, A., Pfister, H., Lieb, R., Wittchen, H.U., Holsboer, F., Angst, J., 2009. Heterogeneity of DSM-IV major depressive disorder as a consequence of subthreshold bipolarity. Arch. Gen. Psychiatry 66, 1341–1352.

Medical Physics
The International Journal of Medical Physics Research and Practice

Improved volumetric measurement of brain structure with a distortion correction procedure using an ADNI phantom

Norihide Maikusa, Fumio Yamashita, Kenichiro Tanaka, Osamu Abe, Atsushi Kawaguchi, Hiroyuki Kabasawa, Shoma Chiba, Akihiro Kasahara, Nobuhisa Kobayashi, Tetsuya Yuasa, Noriko Sato, Hiroshi Matsuda, Takeshi Iwatsubo, and The Japanese Alzheimer's Disease Neuroimaging Initiative

Citation: Medical Physics **40**, 062303 (2013); doi: 10.1118/1.4801913
View online: http://dx.doi.org/10.1118/1.4801913
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/40/6?ver=pdfcov
Published by the American Association of Physicists in Medicine

Improved volumetric measurement of brain structure with a distortion correction procedure using an ADNI phantom

Norihide Maikusa[a)]
Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo 187-855, Japan

Fumio Yamashita
Iwate Medical University, Morioka 028-3694, Japan

Kenichiro Tanaka
Research Association for Biotechnology, Tokyo 105-0003, Japan

Osamu Abe
Department of Radiology, Nihon University School of Medicine, Tokyo 173-8610, Japan

Atsushi Kawaguchi
Biostatistics Center, Kurume University, Kurume 830-0011, Japan

Hiroyuki Kabasawa
Applied Science Laboratory, GE Healthcare, Tokyo 191-8503, Japan

Shoma Chiba
Research Association for Biotechnology, Tokyo 105-0003, Japan

Akihiro Kasahara
Imaging Center, The University of Tokyo Hospital, Tokyo 113-8655, Japan

Nobuhisa Kobayashi
Oizumi Hospital, Tokyo 178-0061, Japan

Tetsuya Yuasa
Department of Bio-systems Engineering, Graduate School of Science and Engineering, Yamagata University, Yonezawa 992-8510, Japan

Noriko Sato
Department of Radiology, National Center of Neurology and Psychiatry, Tokyo 187-855, Japan

Hiroshi Matsuda
Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo 187-855, Japan

Takeshi Iwatsubo
Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

The Japanese Alzheimer's Disease Neuroimaging Initiative

(Received 23 August 2012; revised 1 April 2013; accepted for publication 2 April 2013; published 28 May 2013)

Purpose: Serial magnetic resonance imaging (MRI) images acquired from multisite and multivendor MRI scanners are widely used in measuring longitudinal structural changes in the brain. Precise and accurate measurements are important in understanding the natural progression of neurodegenerative disorders such as Alzheimer's disease. However, geometric distortions in MRI images decrease the accuracy and precision of volumetric or morphometric measurements. To solve this problem, the authors suggest a commercially available phantom-based distortion correction method that accommodates the variation in geometric distortion within MRI images obtained with multivendor MRI scanners.

Methods: The authors' method is based on image warping using a polynomial function. The method detects fiducial points within a phantom image using phantom analysis software developed by the Mayo Clinic and calculates warping functions for distortion correction. To quantify the effectiveness of the authors' method, the authors corrected phantom images obtained from multivendor MRI scanners and calculated the root-mean-square (RMS) of fiducial errors and the circularity ratio as evaluation values. The authors also compared the performance of the authors' method with that of a distortion correction method based on a spherical harmonics description of the generic gradient design parameters. Moreover, the authors evaluated whether this correction improves the test–retest reproducibility of voxel-based morphometry in human studies.

Results: A Wilcoxon signed-rank test with uncorrected and corrected images was performed. The root-mean-square errors and circularity ratios for all slices significantly improved ($p < 0.0001$)

after the authors' distortion correction. Additionally, the authors' method was significantly better than a distortion correction method based on a description of spherical harmonics in improving the distortion of root-mean-square errors ($p < 0.001$ and 0.0337, respectively). Moreover, the authors' method reduced the RMS error arising from gradient nonlinearity more than gradwarp methods. In human studies, the coefficient of variation of voxel-based morphometry analysis of the whole brain improved significantly from 3.46% to 2.70% after distortion correction of the whole gray matter using the authors' method (Wilcoxon signed-rank test, $p < 0.05$).

Conclusions: The authors proposed a phantom-based distortion correction method to improve reproducibility in longitudinal structural brain analysis using multivendor MRI. The authors evaluated the authors' method for phantom images in terms of two geometrical values and for human images in terms of test–retest reproducibility. The results showed that distortion was corrected significantly using the authors' method. In human studies, the reproducibility of voxel-based morphometry analysis for the whole gray matter significantly improved after distortion correction using the authors' method.
 [http://dx.doi.org/10.1118/1.4801913]

Key words: Alzheimer's disease, magnetic resonance imaging, geometric distortion, image normalization

I. INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia, and typically shows memory impairment at the earliest clinical stage. Magnetic resonance imaging (MRI) has shown much promise as a biomarker method of quantifying AD progression. Use of MRI in morphometric or volumetric measurement of brain atrophy, such as changes in cortical thickness, hippocampus volume, whole brain volume, voxel-based morphometry, or tensor-based morphometry, has resulted in improved diagnosis.[1–9] These measurements can also be used to assess the effectiveness of applied therapies.[10–12] A large multisite longitudinal study named the Alzheimer's Disease Neuroimaging Initiative (ADNI) was launched in the United States (US) in 2005.[13] A goal of the US-ADNI is to validate the MRI scan as a surrogate marker of AD. The Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) is another large multisite imaging study with the same goals as the US-ADNI.[14] Following the ADNI's study protocols, the J-ADNI recruited 150 elderly controls, 300 subjects with mild cognitive impairment, and 150 AD patients. Three-dimensional T1-weighted images were acquired from subjects and a specifically designed ADNI phantom (Phantom Laboratory, Salem, New York) was used to inspect image quality and artifacts as well as the stability of acquisitions.

MRI scans often contain geometrical distortions. The most prominent factors are image gradient nonlinearity, static magnetic field inhomogeneity, and magnetic susceptibility as typical image reconstruction relies on linear approximation of a magnetic field gradient.[15] A distortion can cause a superficial local volume change, which affects the precision and accuracy of volumetric and morphometric analysis.[13,15–19] The US-ADNI corrects image geometry for gradient nonlinearity using the gradwarp (GW) correction, which involves a gradient coil design and a phantom-based scaling correction method.[13,15] The former relies on the geometry of gradient coil construction and can only correct the nonlinear component of geometrical distortion. The latter corrects linear scaling changes of images through an affine transformation of nine degrees of freedom. However, GW correction requires information of the coil construction; the scanner is limited if this information is not available. Additionally, GW correction does not correct distortion variability between scanners or changes because of aging within the same scanner, and requires prior information of the gradient coil construction from the manufacture and is not available for all scanners.

On the other hand, several phantom-based distortion-correction methods have been reported in the literature.[16,17,20–22] Baldwin *et al.* characterized and corrected distortion using a three-dimensional (3D) grid phantom and elastic-body spline-kernel transformation function.[17] Carmanos *et al.* constructed a DUPLO-based phantom and proposed distortion correction using information characterized by that phantom and spherical harmonic expansion.[16] Schad *et al.*[20] carried out an early investigation on two-dimensional (2D) MRI distortion correction using a 2D polynomial equation. Menuel *et al.*[22] used phantom including cylindrical rods as fiducial points. Langlois *et al.* proposed a correction method based on the Fourier transform and a simple cubic phantom.[21] This method obtained information about distortion to allow correction of both gradient nonlinearity and background field inhomogeneity for many subsequent patients via scans of a phantom of well-known geometry. However, Gunter *et al.* reported that the distortion factor drifts.[19] Therefore, in a phantom-based method, it is desirable to periodically scan the phantom to correct the drift. The above studies used the phantom only to assess characteristics of the geometrical distortion.

In J-ADNI MRI scanning protocol, ADNI phantom images are acquired consecutively after patients are scanned in accordance with the US-ADNI protocol.[13] The ADNI phantom image was used to check the signal-to-noise ratio (SNR), contrast, and geometric distortion for all subjects at each scan time as done by the US-ADNI. Gunter reported that an ADNI phantom can be measured in a multisite study to identify scanner errors through central monitoring, and in the latest result, these errors would have contributed to imprecision in quantitative metrics of more than 25%.[19] The present report states

TABLE I. Range of parameters for MRI acquisition of human subjects and phantoms.

Manufacturer	Model name	Slice num	TR[a] (ms)	TI (ms)	Flip (deg)	Matrix [in-plane resolution (mm^2)]	Fov (mm^2)	Slice thickness (mm)	Plane
GE	GENESIS SIGNA SIGNA EXCITE SIGNA HDx/HDxt	166–180							
Siemens	Avanto MAGNETOM VISION Sonata Symphony Symphony Vision Symphony Tim	160	MA[b]:2400 BC[c]:3000	1000	8	192×192 [1.25×1.25] or 256×256 [0.9375×0.9375]	240×240	Human: 1.2 Phantom 1.3	Sagittal PE[d] = A/P
Toshiba	Excelart Vantage	180							
Philips	Achieva Intera	170	MA:2300						
Hitachi	ECHELON Vega	170	MA:2600	1100					

[a]TR is defined here as the receptition time for the inversion pulses.
[b]MA = multicoil phased-array head coil.
[c]BC = birdcage or volume head coil.
[d]PE is the phase encoding direction.

that ADNI phantom measurement is necessary to ensure consistency of MRI data acquired in a multisite longitudinal study like the ADNI project. If the distortion can be corrected using an ADNI phantom image, the correction can be applied to the ADNI protocol using the ADNI phantom that was already being employed for other purposes. In other words, the ADNI phantom-based distortion correction method can serve dual purposes: correction of geometrical distortion and monitoring of the scanner condition.

For the reason above, we suggest an ADNI phantom-based distortion correction using a polynomial equation to enhance the reliability and accuracy of structural analysis of brain images. Our method can correct nonlinear and linear components of geometrical distortion using the polynomial function. ADNI phantom scanning is available commercially and is recommended to monitor several MRI scanner conditions in multisite and longitudinal studies compliant with the ADNI study protocol. In this paper, we refer to this correction as the J-ADNI method. We evaluate the effectiveness of the method in terms of fiducial errors and circularity in phantom studies. For human studies, we use voxel-based morphometry (VBM). Finally, we compare improvements in distortion correction. Our method uses a commercially available phantom, and therefore attains distortion correction for all scanners without requiring information of the coil design from the manufacturer.

II. METHODS

II.A. MRI data and protocols

The pulse sequence used to acquire 3D, 1.5-T T1-weighted images was the magnetization-prepared rapid gradient echo (MPRAGE) sequence. MRI scanners were accepted via checking conformance with J-ADNI protocols, and we checked whether any image suffered serious degradation due to a motion artifact, warping around into the side of the skull, low SNR, signal loss, or metal artifact. This strengthened the longitudinal and cross-sectional analysis. The MPRAGE sequence is used to enhance gray/white contrast to noise for superior performance in applications requiring cortical segmentation.[13] We thus introduced the MPRAGE sequence to acquire magnetic resonance images in the J-ADNI. The parameters used to obtain magnetic resonance images with the MPRAGE sequence are listed in Table I for each manufacturer and model. For Siemens Magnetom Vision, Toshiba Excelart Vantage, and Hitachi Echelon Vega systems, a customized sequence designed to retain compatibility with the ADNI MPRAGE sequence was used. The parameters were chosen to be as close as possible to the parameters of the US-ADNI's MRI sequence to merge MRI data for global analysis and/or comparison of the population at a future date.

II.B. ADNI phantom design

The ADNI phantom contains 165 polycarbonate spheres filled with copper sulfate inside a water-filled clear urethane shell (20 cm in diameter). Four contrast spheres (3.0 cm in diameter) with copper sulfate concentrations of 0.9, 1.2, 1.7, and 2.4 mM provided scanner contrast. A large sphere (6.0 cm in diameter) containing 3.3 mM copper sulfate solution at the center of the phantom was used to calculate the SNR of a scanner and define the origin of the coordinate system. Inclusions (158 with diameters of 1.0 cm, and two with diameters of 1.5 cm) in 3.3 mM copper sulfate solutions, known as fiducial or small spheres, defined the geometrical coordinates of the phantom image in addition to

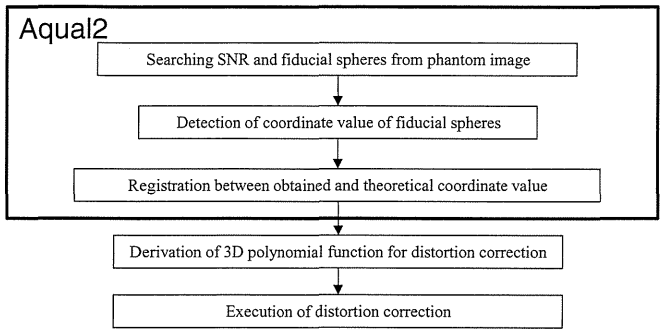


FIG. 1. Flowchart of distortion correction including mass-preservation resampling.

the SNR sphere. The phantom (Phantom Laboratory) was manufactured via numerous inspections of batches of both machined and molded components along with assembly inspections. The tolerance for the sphere locations was ±0.125 mm in the phantom's x–y directions (parallel to the equatorial plane) and ±0.3 mm in the z direction.

II.C. Method for correcting phantom-based geometrical distortion

A flow chart of the proposed distortion correction method is shown in Fig. 1. First, we performed phantom image analysis using Aqual2 (Acceptance Qualification tool), which was developed at the Mayo Clinic.[19] Aqual2 was written in MATLAB (MathWorks, Natick, Massachusetts) and is available via the Internet (http://adni.loni.ucla.edu/research/research-tools/aqual2/). This program finds 160 fiducial spheres and a SNR sphere to obtain coordinate values relative to the center of phantom images. The program is then able to register the obtained and designed coordinate values. By comparing the actual measured coordinate value of the fiducial and SNR spheres with the designed coordinate value, we can calculate the distortion field of the scanner. The warping function from a distorted coordinate to a design (i.e., nondistorted) coordinate value is given by

$$\begin{cases} X = \sum \sum_{p+q+r\le n} \sum a_{pqr} x^p y^q z^r + \varepsilon \\ Y = \sum \sum_{p+q+r\le n} \sum b_{pqr} x^p y^q z^r + \varepsilon \\ Z = \sum \sum_{p+q+r\le n} \sum c_{pqr} x^p y^q z^r + \varepsilon \end{cases}, \quad (1)$$

where x, y, z and X, Y, Z are the obtained and designed coordinate values, respectively, along three orthogonal dimensions; p, q, and r are orders of the polynomial equation and n is the polynomial degree of a function; a_{pqr}, b_{pqr}, and c_{pqr} are coefficients of the polynomial equation; and ε is a Gaussian random variable. Because designed coordinate values of fiducial and SNR spheres are known, fiducial errors between design and obtained coordinate values reflect the geometric distortion of the scanner due to gradient nonlinearity, background field inhomogeneity, and magnetic susceptibility. We assume that the contribution of susceptibility in the ADNI phantom image is negligible. Thus, we can correct distortion in a MRI brain image for the same scanner using the equation derived from the phantom image. Note that our correction model has Gaussian distribution errors as described by ε in Eq. (1). Polynomial coefficients, a_{pqr}, b_{pqr}, and c_{pqr}, are calculated from the correspondence between x, y, z and X, Y, Z using Gaussian elimination to minimize Gaussian random variables in each of the three directions. Moreover, to use this polynomial equation, we must decide on the degrees of the polynomial (i.e., p, q, and r). To solve this problem, we use the Bayesian information criterion (BIC). The BIC is a criterion for model selection among a class of parametric models with different numbers of parameters.[23,24] In the present model, we assumed error variance between obtained and design sphere coordinate values given by a Gaussian distribution. The BIC is defined as

$$\mathrm{BIC}(n) = n \cdot \ln(\hat{\sigma}_e^2) + k \cdot \ln(n), \quad (2)$$

where σ is the error variance, n is the number of data points (i.e., the number of fiducial spheres), and k is the number of free parameters to be estimated. The best parameter of the polynomial equation is the one that minimizes Eq. (2). Hence, we selected an optimized set of polynomial functions for the three directions individually.

The polynomial function corrects geometrical distortion in a brain image, which is acquired immediately before phantom scanning with the same scanner. The resampling method used when an image is deformed to correct geometric distortion is trilinear interpolation. Additionally, it is necessary to correct the intensity to compensate for the change in effective voxel size with distortion. This intensity is corrected by multiplying the image intensity and the Jacobian determinant, $|J|$, calculated from the transformation equation at each voxel.[15,25] This equation is the polynomial function. We can calculate the Jacobian determinant, $|J|$, of the polynomial function associated with Eq. (1) as

$$|J| = \begin{vmatrix} \frac{\partial X'}{\partial x} & \frac{\partial X'}{\partial y} & \frac{\partial X'}{\partial z} \\ \frac{\partial Y'}{\partial x} & \frac{\partial Y'}{\partial y} & \frac{\partial Y'}{\partial z} \\ \frac{\partial Z'}{\partial x} & \frac{\partial Z'}{\partial y} & \frac{\partial Z'}{\partial z} \end{vmatrix}, \quad (3)$$

where $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$ are the partial derivative operators with respect to the orthogonal components, and X', Y', and Z' are the coordinates corrected by polynomial functions.

This correction software, written in MATLAB, executes automatically in about 15 min on a Xeon Processor E5540 (2.53 GHz) with RedHat Enterprise Linux 5.3 as the operating system using only input from phantom and brain images.

II.D. Evaluation metrics

For quantitative assessment of our correction method for phantom images acquired by all scanners ($n = 41$) at 38 J-ADNI clinical sites, we used two functions. The first function

was the root mean square (RMS) error in the determined location of fiducial spheres. The second function was the circularity ratio (CR) of an outer shell in a phantom image for three slices: axial, coronal, and sagittal. Polynomial functions for the correction procedure obtained from one phantom image were used to deform other phantom images acquired 1 week later by the same scanner.

To obtain RMS errors, coordinate values of fiducial spheres were acquired from uncorrected and corrected phantom images using Aqual2. We compared obtained and designed coordinate values, and calculated RMS errors of fiducial spheres.

Our method used geometrical information of the positions of fiducial spheres in phantom images. Therefore, it was expected that our method corrects distortion around the fiducial spheres. The overfitting of the distortion correction is not as great as predicted; accordingly, RMS errors are less than expected but the whole-image structures have unforeseen shapes. To assess the adequacy of the image obtained after distortion correction around fiducial points, we used the CR metric. The CR is defined as the amount of geometrical circle distortion according to Japanese industrial standards. The outer shell of a phantom is spherical and its orthogonal plane images are thus truly circular. Therefore, CRs of the outer shell in axial, coronal, and sagittal slices were used as other metrics to assess distortion. The CR is useful in assessing the validity of our method outside the fiducial sphere. The CR can be calculated as

$$\mathrm{CR} = \frac{r_i}{r_c}, \tag{4}$$

where r_i and r_c are radii of the incircle and circumcircle, respectively. If the ratio of the radii of the two concentric circles (incircle and circumcircle) is 1.0, the boundary is a true circle. To calculate the CR, we draw an incircle and circumcircle manually in three orthogonal images (i.e., the axial, sagittal, and coronal images) with origins at the center of the 3D phantom image. An operator carried out this task after only being instructed to draw two circles inscribed and circumscribed on the boundary between the outer shell of the phantom and background. To demonstrate the performance of our method, the Wilcoxon signed-rank test was performed for uncorrected and corrected images because the data may not be normally distributed because of the small sample size. A two-sided p-value less than 0.05 was considered statistically significant.

II.E. Comparison of geometrical distortion correction methods of the J-ADNI and US-ADNI

The US-ADNI employs the GW correction method[13] and ADNI phantom-based scaling correction. In GW correction, a set of spherical harmonic coefficients is computed for a particular gradient coil design, and can be used to correct the distortion arising from gradient nonlinearity embedded in acquired images.[18,25] Moreover, ADNI phantom-based linear scaling correction involves reducing the observed geometric drift or voxel size adjustment employing an affine transformation of nine degrees of freedom.[19] The parameters of affine transformation were acquired from design and obtained fiducial points of the ADNI phantom. Because the scaling correction program is not publicly available, we wrote MATLAB-based software for affine transformation of the magnetic resonance image using voxel size information obtained by Aqual2. We checked the adequacy of this program by comparing scaling corrected images available from the US-ADNI (www.loni.ucla.edu/ADNI/).

GW correction is only supported for scanners manufactured by GE and Siemens. GW correction uses the same spherical harmonic coefficients for the same gradient design to correct distortion only arising from gradient nonlinearity. Therefore, it cannot correct distortion variations between scanners of the same model or changes with time for the same scanner. These variations are generated by machine maintenance, such as magnetic field adjustment, repair of the magnetic coil, and quenching. We compared our proposed correction method with the GW method and the GW-plus-scaling method for corrected phantom images. Here, we need to separate out the contribution of gradient nonlinearity from other distortion components because our method corrects inhomogeneity in the magnetic background field as well as gradient nonlinearity. Bakker *et al.* reported that distortion only arose from gradient nonlinearity in the phase-encoding direction in a two-dimensional magnetic resonance image.[26] Subsequently, Baldwin *et al.* reported that the slice-encoding direction as well as the phase-encoding direction provides distortion when this principle is extended to a 3D magnetic resonance image.[17] Here, we assess RMS errors along the phase- and slice-encoding directions (i.e., the A/P and R/L directions, respectively), and also the CR in the axial image to compare the efficacy of our method with that of the GW and GW-plus-scaling correction procedures.

GW software is publicly available to researchers and we were able to obtain gradient information from GE Healthcare Japan, but we could only correct images acquired by GE MRI scanners using GW. For a statistical comparison of the J-ADNI, GW, and GW-plus-scaling correction methods, a Steel–Dwass signed-rank test was performed for the RMS error and CR.

II.F. Application of the distortion correction method to human studies

Geometrical distortion is induced by background field inhomogeneity and magnetic susceptibility as well as gradient nonlinearity. Magnetic susceptibility is dependent on the imaged object. Our method assumes that the contribution of susceptibility to geometrical distortion of the ADNI phantom image is negligible. Therefore, to evaluate our simple and practical method of correcting images of human subjects, we performed VBM analysis using software developed by Matsuda *et al.*[8] This approach allows us to automatically detect early specific atrophy in AD using three-dimensional T1-weighted MRI data as a series of segmentations using statistical parametric mapping 8 (SPM8) with the toolbox DARTEL. The following is a brief explanation of the procedure. Gray/white matter can be anatomically segmented employing

a unified tissue-segmentation procedure after correction of image-intensity nonuniformity. Segmented images are then normalized to the MNI (Montreal Neurological Institute) space via a nonlinear transformation to a customized template created by DARTEL, modulation by the Jacobian determinant, and linear transformation from the customized template to MNI space. Finally, the images are smoothed using an 8-mm full-width-at-half-maximum Gaussian kernel. To perform Z-score analysis for gray matter, brain images of 80 healthy control volunteers without memory impairment or cognitive disorders (37 men and 43 women, 70.4 ± 7.8 years of age) were processed using the above procedures, and the mean and standard deviation of these smoothed modulated intensity values were calculated voxel by voxel in MNI space, preliminarily. Finally, a Z-score map was generated by comparing an individual image with a control mean and standard deviation at each voxel in a gray/white matter image. The Z-score was calculated as [control mean] − [individual value]/(control SD) and it shows atrophy relative to the normal control.

Directly assessing the usefulness of VBM analysis is difficult because the true brain atrophy of the subject is unknown. Therefore, we assumed that geometric distortion reduces the reproducibility of the Z-score. Segmentation and anatomical standardization errors may be increased by an anatomical volume change and partial volume effects resulting from distortion. We assessed the reproducibility of VBM analysis for brain images corrected using the J-ADNI method for 38 healthy volunteers (32 men and six women, 33.6 ± 10.7 years of age) scanned twice at an interval of 1 week. The phantom image was obtained immediately after each volunteer was scanned and the imaging sequence is identical with phantom scanning as described in Table I. Using the automated VBM software, we obtained two indicators for characterizing atrophy in medial temporal structures and in the whole brain.[8] The first indicator was the averaged Z-score in the medial temporal structures, which indicates the severity of atrophy in the entorhinal cortex and hippocampus. The second indicator was the extent of significant atrophy in the region of the whole brain, essentially the percentage of coordinates with a Z-score exceeding a threshold value (2.0) in the entire brain. We defined reproducibility as a coefficient of variation (CV) for these two indicators between the first and second scans.

III. RESULTS

III.A. Phantom studies

All phantom images were acquired with MRI scanners approved by the J-ADNI at 38 clinical sites. Polynomial functions were determined from phantom images acquired at week 0 and applied to other phantom images acquired by the same scanner 1 week later. The polynomial function was of third order along the A/P and R/L directions, and second order along the S/I direction. RMS errors of corrected phantom images obtained using a fourth- or fifth-order polynomial function vary less than RMS errors for the result obtained using a polynomial function of third order.

The resulting uncorrected images and images corrected using the J-ADNI, GW, and GW-plus-scaling correction methods are shown in Fig. 2. A red circle and yellow circle are superimposed on the outer and inner edges of the phantom shell, respectively. In the figure, although the outline of a phantom should be spherical, uncorrected and corrected GW phantom images are not circular. However, the outline of a corrected phantom image subject to the J-ADNI method appears to be circular. An example of the spatial distribution of fiducial sphere errors in a corresponding phantom image is shown in Fig. 3. Error levels corresponding to fiducial sphere positions are indicated by arrows of different color. A red arrow indicates errors over 3 mm, yellow indicates errors of 2–3 mm, blue indicates errors of 1–2 mm, and gray indicates errors of 0–1 mm. The direction of the arrow indicates the direction of the highest error at the orthogonal coordinate. Figures 3(b) and 3(c) show remaining errors over 1.0 mm at positions of designed fiducial spheres. However, the result of the J-ADNI method shows the errors of fiducial spheres under a voxel resolution of 1.0 mm. Therefore, the results show that our distortion correction method is qualitatively more efficient than other correction methods.

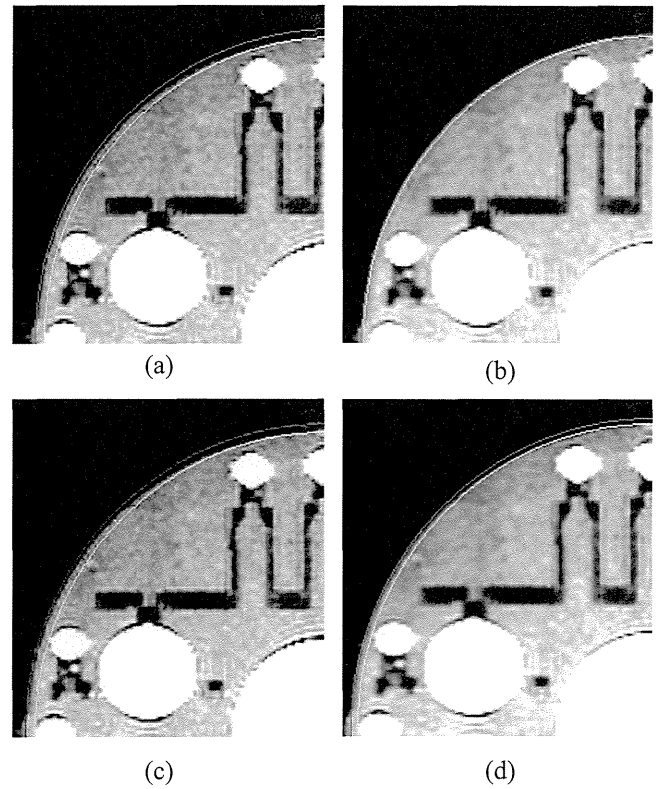


FIG. 2. Phantom across the axial plane. (a) Uncorrected; (b) corrected with the J-ADNI method; (c) GW correction; and (d) GW (nonlinear) and scaling correction (linear). The distance between the yellow and red circles shows the distortion level at the outer shell of the phantom.

For quantitative assessment of our method of correcting phantom images, we calculated the RMS error distribution. The distribution was obtained from uncorrected phantom images and phantom images corrected using Aqual2 and the CR. An incircle and circumcircle were manually drawn on the orthogonal plane of a phantom image (Fig. 4). The median RMS error was 0.615 mm for uncorrected phantom images and 0.257 mm for corrected images. The median CR was 0.985

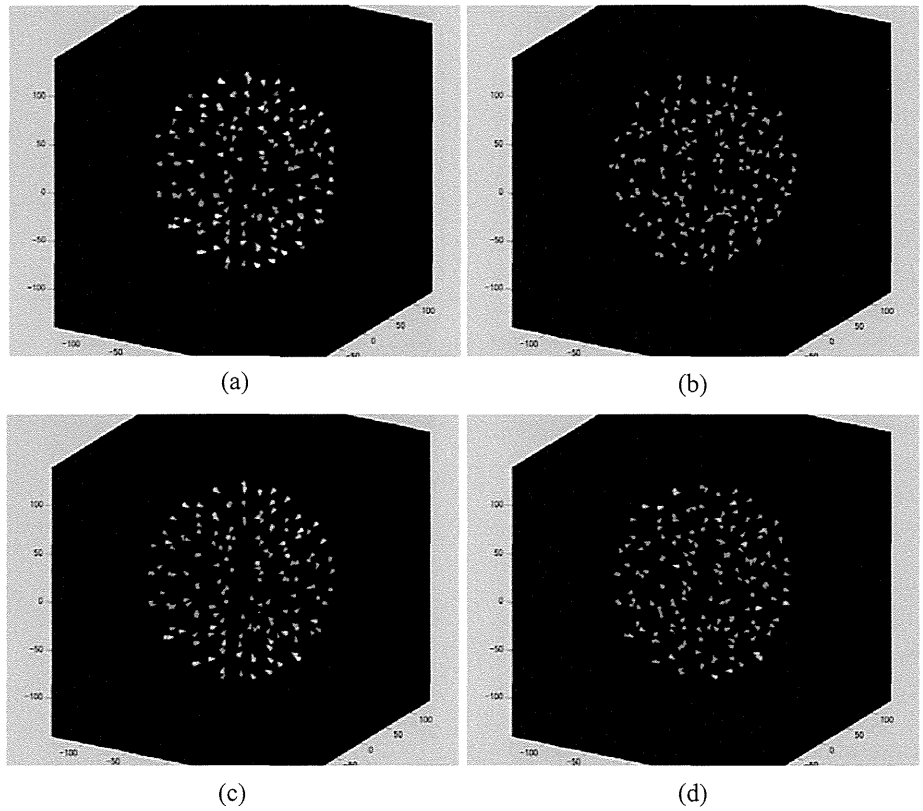


FIG. 3. Spatial distributions of fiducial sphere errors (differences) between obtained and designed coordinate values in phantom images. The arrows indicate the direction and magnitude of error. Colors indicate error levels at corresponding positions of fiducial spheres. The direction of an arrow corresponds to the basis vector with the highest magnitude of error. Red indicates error exceeding 3 mm; yellow indicates error of 2–3 mm; blue indicates error of 1–2 mm; and gray indicates error of 0–1 mm. (a) Uncorrected; (b) corrected with the J-ADNI method; (c) GW correction; and (d) GW and scaling correction.

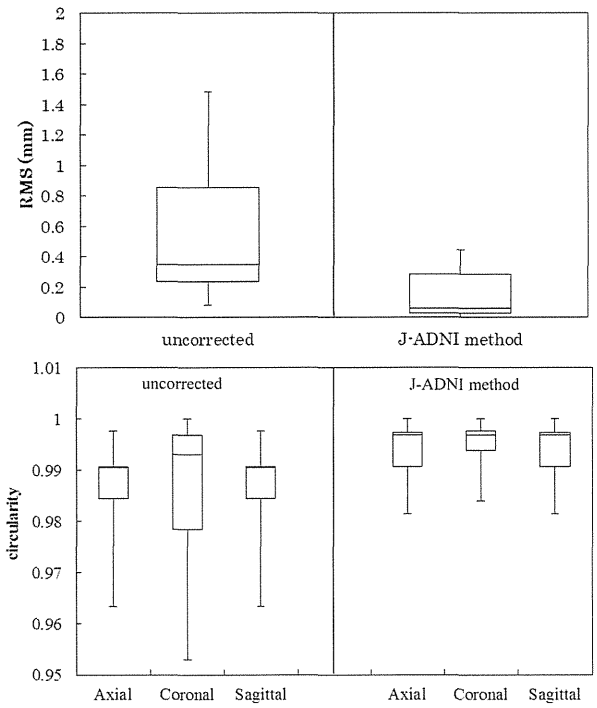


FIG. 4. RMS error and CR plots of uncorrected and corrected (J-ADNI method) phantom images ($N = 42$) obtained from all scanners for evaluation of effectiveness. (a) and (b) RMS errors obtained from analysis of phantom images by Aqual2. (c) and (d) CR plots for axial, coronal, and sagittal sections that were uncorrected (c) and corrected with the J-ADNI method (d).

for uncorrected axial images, 0.993 for uncorrected coronal images, and 0.991 for uncorrected sagittal images. After distortion correction, the median CR was 0.997 for axial images, 0.997 for coronal images, and 0.997 for sagittal images. RMS errors and the CR for all slices significantly improved (Wilcoxon signed-rank test, $p < 0.001$) after J-ADNI distortion correction.

These results show that the CR of a corrected image following application of the J-ADNI method was better than that of an uncorrected image for all slices. It is clear from Fig. 4 that the J-ADNI method improved RMS errors and the CR, and effectively corrected geometrical distortion in all phantom images.

To compare the performance of our method with that of GW correction and that of phantom-based scaling correction in addition to GW correction, we calculated RMS errors and the CR for phantom images as described previously. We obtained this information from GE Healthcare Japan and were therefore only able to apply GW correction to 11 images acquired by a group of GE scanners. The RMS errors in phase-encoding and slice-encoding directions, and CR results in the axial plane, both uncorrected and corrected with each method, are shown in Fig. 5. The top panel indicates that median RMS errors were 0.259 mm for uncorrected phantom images, 0.0976 mm for J-ADNI-corrected images, 0.268 mm for GW-corrected images, and 0.187 mm for GW-plus-scaling-corrected images along the phase-encoding direction.

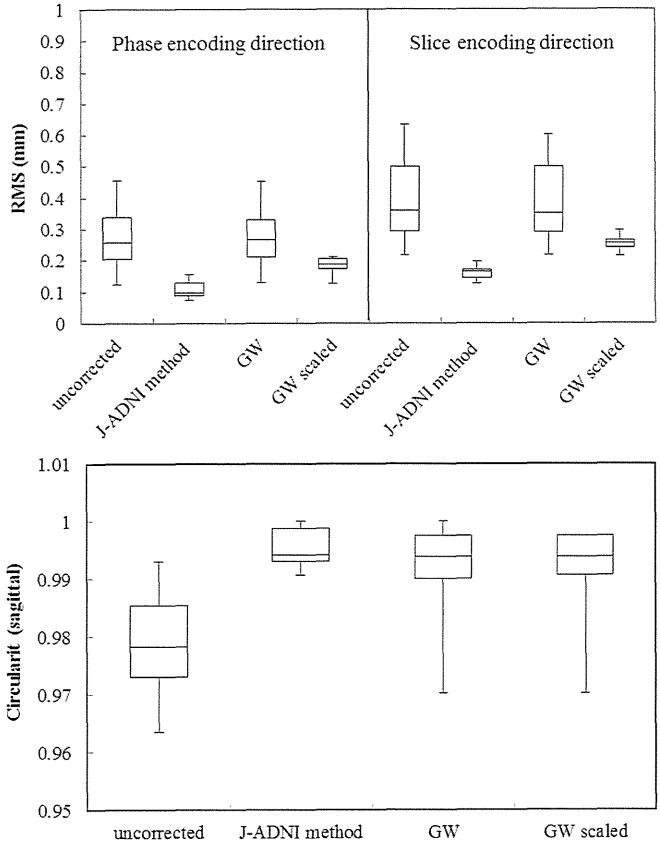


FIG. 5. RMS errors and CR plots ($N = 12$) for comparison of effectiveness of the J-ADNI, GW, and GW-plus-scaling correction methods for a confined group of GE scanners. The top panel shows RMS errors calculated by Aqual2 in the phase-encoding, A/P, direction (a) and slice-encoding, R/L, direction (b). The bottom panel shows CR plots for axial phantom images (c). These directions and (axial) plane suffered only the contribution of the gradient nonlinearity of distortion.

The corresponding values were 0.359 mm for uncorrected phantom images, 0.164 mm for J-ADNI-corrected images, 0.350 mm for GW-corrected images, and 0.255 mm for GW-plus-scaling-corrected images along the slice-encoding direction. The bottom panel shows the median CRs for the axial images: 0.978 for uncorrected images, 0.994 for J-ADNI-corrected images, 0.994 for GW-corrected images, and 0.994 for GW-plus-scaling-corrected images. The J-ADNI method improved RMS errors significantly compared with using uncorrected, GW, and GW-scaling correction methods along the phase-encoding direction and slice-encoding direction ($p < 0.001$).

CRs of the phantom image corrected with the J-ADNI method, GW, and GW-plus-scaling correction showed significant error reduction compared with the uncorrected image (Steel–Dwass test; $p < 0.001$, $p = 0.028$, and $p = 0.014$, respectively). However, there was almost no difference in the CR when comparing the J-ADNI method with the GW and GW-plus-scaling corrections despite there being CR differences for a specific image as shown in Fig. 2.

III.B. Reproducibility of VBM analysis

We performed Z-score analysis using automated VBM software. Ideally, CV for indicators in test–retest studies should be zero for a subject where there is no expectation of morphological changes associated with brain atrophy. CV maps drawn by comparing the initial and second Z-score analyses of uncorrected and corrected brain images of 38 healthy volunteers are shown in Fig. 6. The top and bottom rows show results for uncorrected and corrected brain images, respectively. In these maps, color voxels have a median value of 38 for the volunteer's CV of the Z-score between the first and second scans and the voxels are overlaid with orthogonal sections of the anatomically standardized MRI. The color scale range was −25.0% to −7.5% and 7.5% to 25.0%. Areas enclosed by green lines indicate medial temporal structures having the most significant decline of gray matter concentration at the very early stages of AD (Matsuda *et al.*[8]).

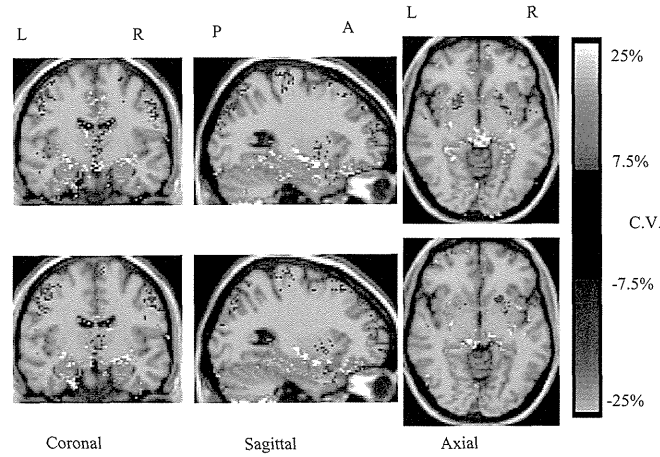


FIG. 6. Map of median CVs obtained by VBM Z-score analysis. Top and bottom rows show CV maps for an uncorrected image and a distortion-corrected image (J-ADNI method), respectively. The color-scaled CV map ranges from −25.0% to −7.5% and from 7.5% to 25.0% are displayed by overlaying orthogonal sections of the anatomically standardized MRI. Areas enclosed by green lines show medial temporal structures. Note the decrease in CV following distortion correction for the whole brain.

We obtained CVs of the averaged Z-score in the medial temporal structures. The extent of a region showing significant atrophy between first and second scans obtained using automated VBM software is shown in Fig. 7. The median CV of the averaged Z-score in the medial temporal structures after distortion correction using our method was 2.68%, which is not a significant improvement when comparing with uncorrected images. The median CV of the percentage of coordinates with a Z-score exceeding the threshold in the entire brain improved from 3.46% to 2.70% after distortion correction using the J-ADNI method (Wilcoxon signed-rank test, $p < 0.05$).

IV. DISCUSSION

To accurately measure the atrophy of brain structures such as the hippocampus, and to explore surrogate biomarkers obtained from MRI, geometrical distortion needs to be corrected. The US-ADNI uses two types of distortion correction methods: GW and phantom-based scaling correction. GW is only used to correct nonlinearities of geometrical distortion,

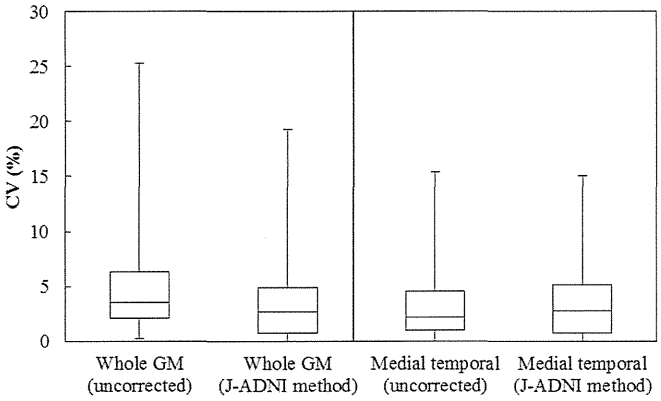


FIG. 7. CV between the baseline and second scan for uncorrected and corrected brain images ($N = 42$). The extent of atrophy variance in the whole-brain gray matter (GM) is shown (a), alongside the average Z-score for the medial temporal structures of the gray matter, which was determined using automated VBM software (b).

whereas scaling correction is used to correct linear scaling distortion.[19] According to Eq. (1), the first order of the polynomial function corresponds to the scaling factor of transformation and can correct linear scaling correction. Higher orders correct residual distortion; i.e., nonlinearities of distortion. Therefore, the J-ADNI method, which is based on a polynomial equation, can correct both nonlinear and linear geometrical distortions concurrently, and the method corresponds to phantom-based scaling correction and GW correction used by the US-ADNI.

Figure 4 shows that RMS errors and CRs in phantom studies significantly improved after J-ADNI distortion correction. In particular, RMS errors reduced by several millimeters to below the voxel resolution of 1.0 mm following application of the J-ADNI method. The structural analysis of brain images, such as voxel-based volume measurements within a region of interest, requires accuracy below the voxel resolution. Figure 4 shows that our method has enhanced reliability for volumetric brain image analysis.

We examined the ability of the J-ADNI, GW, and GW-plus-scaling correction methods to correct distortion arising from gradient nonlinearity. Figure 5 shows that RMS errors after GW correction did not change. This may be because of the efficiency limit of GW, as subsequent scaling correction reduced RMS errors significantly. These results show that the distortion arising from gradient nonlinearity includes a component that cannot be described by generic coil information and demonstrate the necessity of ADNI phantom-based scaling correction in addition to GW correction. In fact, the necessity of scaling correction through the registration of nine degrees of freedom has been reported.[19,27] Furthermore, GW corrects geometrical distortion according to the geometry of the generic gradient coil construction. If the gradient nonlinearity changes for the same scanner model, GW cannot correct this variability. Figure 5 shows that our method can reduce RMS errors more than GW and GW-plus-scaling correction. Hence, our method can describe distortion arising from gradient nonlinearity and correct it.

In contrast, the CR following implementation of the J-ADNI method was not statistically significant compared with the CRs following GW correction and GW-plus-scaling correction. The CR was obtained by drawing two circles on the border of the phantom outer shell and background, and its sensitivity was the pixel unit. Moreover, these circles were drawn manually and there were user-introduced errors at the level of the pixel unit. For the above reasons, the CR of the phantom image can be used to assess the adequacy of correction in the overall region, but may not be a good metric with which to correctly quantify and compare distortion correction methods.

However, the J-ADNI method detected fiducial sphere position errors and estimated geometrical distortion fields in phantom images for each scan, and corrected the nonlinearity and linearity of distortion. It was therefore used successfully to correct gradient variability.

The J-ADNI method demonstrates improved RMS error assessment and can thus correct geometrical distortion, thereby improving the reliability and accuracy of volumetric and/or morphometric analysis of structural MRI brain images.

In human studies, the reproducibility of VBM analysis for whole-brain gray matter was significantly improved after distortion correction using the J-ADNI method. We assumed that the CR of the outer shell of the phantom affects the measurement accuracy of the broad structure volume, which is necessary for whole-brain imaging. The CR was improved using either the J-ADNI or GW method, corresponding to improved accuracy in describing the broad structure. Geometrical distortion resulted in apparent volume changes and partial volume effects at each voxel. Our method corrected these erroneous volume changes and partial volume effects through image warping and mass-preservation resampling, respectively. Repeatability of subsequent analyses was improved after distortion correction. However, the reproducibility of VBM analysis for medial temporal structures was not significantly improved using our method because fiducial position error was more common in the outer region of the phantom image, as shown in Fig. 2. In fact, medial temporal structures originally showed low variation in VBM analysis (Fig. 6). Therefore, effects of distortion correction may be smaller than those observed for the whole brain. Alternatively, the efficacy of distortion correction is seen for the whole-brain gray matter including the region of large distortion. Our future task is to investigate the performance of our method for anatomical structures around the surface of the brain that are expected to suffer from distortion in human studies; e.g., the cortical thickness, lateral temporal lobe, and inferior parietal lobe.

The GW method requires information about the gradient coil, which can only be provided by the manufacturer, and the method can only be applied to MRI scanners produced by GE and Siemens. Our method requires scanning of a phantom image to correct geometric distortion of a subject's brain image and can be adapted to all scanners. The ADNI phantom is commercially available and can be easily purchased. However, the phantom-based correction method requires additional scanning to obtain a phantom image for correction.

A multisite study such as the ADNI can monitor scanner performance by scanning the ADNI phantom. Our method uses this scan, and we can thus correct distortion by making one phantom scan serve a dual purpose.

In future studies, we will investigate the period of validity for the distortion field obtained from a phantom image. This would assist in reducing the cost of conducting a phantom scan. We also intend to propose an optimal phantom scanning design for distortion correction using our developed method associated with scanner performance monitoring.

ACKNOWLEDGMENTS

This study is part of the "Translational Research Promotion Project/Research project for the Development of a Systematic Method for the Assessment of Alzheimer's Disease," sponsored by the New Energy and Industrial Technology Development Organization (NEDO) of Japan. The J-ADNI is also supported by a Grant-in-Aid for Comprehensive Research on Dementia from the Japanese Ministry of Health, Labour, and Welfare, as well as by grants from J-ADNI Pharmaceutical Industry Scientific Advisory Board (ISAB) companies. The authors would like to thank the J-ADNI Imaging ISAB and other organizations for their support of this work.

[a]Author to whom correspondence should be addressed. Electronic mail: maikusa@ncnp.go.jp

[1]C. Fennema-Notestine, D. J. Hagler, Jr., L. K. McEvoy, A. S. Fleisher, E. H. Wu, D. S. Karow, A. M. Dale, and Alzheimer's Disease Neuroimaging Initiative, "Structural MRI biomarkers for preclinical and mild Alzheimer's disease," Hum. Brain. Mapp. **30**(10), 3238–3253 (2009).

[2]B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, D. A. Bennett, L. A. Beckett, and L. deToledo-Morrell, "MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease," Neurobiol. Aging **22**, 747–754 (2001).

[3]D. Holland, J. B. Brewer, D. J. Hagler, C. Fennema-Notestine, A. M. Dale, and Alzheimer's Disease Neuroimaging Initiative, "Subregional neuroanatomical change as a biomarker for Alzheimer's disease," Proc. Natl. Acad. Sci. U.S.A. **106**(49), 20954–20959 (2009).

[4]M. Grundman, D. Sencakova, C. R. Jack, Jr., R. C. Petersen, H. T. Kim, A. Schultz, M. F. Weiner, C. DeCarli, S. T. DeKosky, C. Dyck, R. G. Thomas, and L. J. Thal, "Alzheimer's disease cooperative study brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial," J. Mol. Neurosci. **19**, 23–27 (2002).

[5]C. R. Jack, Jr., R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen, "Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease," Neurology. **49**, 786–794 (1997).

[6]S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, Jr., J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," Brain **131**, 681–689 (2008).

[7]B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," Neuroradiology **51**, 73–83 (2009).

[8]H. Matsuda, S. Mizumura, K. Nemoto, F. Yamashita, E. Imabayashi, N. Sato, and T. Asada, "Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated Lie algebra improves the diagnosis of probable Alzheimer disease," AJNR Am. J. Neuroradiol. **33**, 1109–1114 (2012).

[9]P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, Jr., "Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies," Neuroimage **39**, 1186–1197 (2008).

[10]B. Dubois, H. H. Feldman, C. Jacova, S. T. Dekosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P. Scheltens, "Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria," Lancet Neurol. **6**(8), 734–746 (2007).

[11]R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M. O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," Neuroimage **56**, 766–781 (2011).

[12]S. K. Madsen, A. J. Ho, X. Hua, P. S. Saharan, A. W. Toga, C. R. Jack, Jr., M. W. Weiner, and P. M. Thompson, "3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects," Neurobiol. Aging **31**, 1312–1325 (2010).

[13]C. R. Jack, Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," J. Magn. Reson. Imaging **27**(4), 685–691 (2008).

[14]T. Iwatsubo, "Japanese Alzheimer's Disease Neuroimaging Initiative: Present status and future," Alzheimers Dement. **6**, 297–299 (2010).

[15]A. Janke, H. Zhao, G. J. Gowin, G. J. Galloway, and D. M. Doddrell, "Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images," Magnet. Reson. Med. **52**, 115–122 (2004).

[16]Z. Caramanos, V. S. Fonov, S. J. Francis, S. Narayanan, G. B. Pike, D. L. Collins, and D. L. Arnold, "Gradient distortions in MRI: Characterizing and correcting for their effects on SENA-generated measures of brain volume change," Neuroimage **49**, 1601–1611 (2010).

[17]L. N. Baldwin, K. Wachowicz, and S. D. Thomas, "Characterization, prediction and correction of geometric distortion in 3 T MRI images," Med. Phys. **34**, 388–399 (2007).

[18]A. D. Leow, A. D. Klunder, C. R. Jack, Jr., A. W. Toga, A. M. Dale, M. A. Bernstein, P. J. Britson, J. L. Gunter, C. P. Ward, J. L. Whitwell, B. J. Borowski, A. S. Fleisher, N. C. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G. E. Alexander, M. W. Weiner, P. M. Thompson, and ADNI Preparatory Phase Study, "Longitudinal stability of MRI for mapping brain change using tensor-based morphometry," Neuroimage **31**, 627–640 (2006).

[19]J. L. Gunter, M. A. Bernstein, B. J. Borowski, C. P. Ward, P. J. Britson, and J. P. Felmlee, "Measurement of MRI scanner performance with the ADNI phantom," Med. Phys. **36**, 2193–2205 (2009).

[20]L. Schad, S. Lott, F. Schmitt, V. Sturm, and W. J. Lorenz, "Correction of spatial distortion in MRI image: A prerequisite for accurate stereotaxy," J. Comput. Assist. Tomogr. **11**, 499–505 (1987).

[21]S. Langlois, M. Desvigness, J. M. Constans, and M. Revenu, "MRI geometric distortion: A simple approach to correcting the effects of non-linear gradient fields," J. Comput. Assist. Tomogr. **9**, 821–831 (1999).

[22]C. Menuel, L. Garnero, E. Bardinet, F. Poupon, D. Phalippou, and D. Dormont, "Characterization and correction of distortions in stereotactic magnetic resonance imaging for bilateral subthalamic stimulation in Parkinson disease," J. Neurosurg. **103**, 256–266 (2005).

[23]A. R. Liddle, "Information criteria for astrophysical model selection," Mon. Not. Roy. Astron. Soc. Lett. **377**, L74-L78 (2007).

[24]G. Schwarz, "Estimating the dimension of a model," Ann. Stat. **6**, 461–464 (1978).

[25]J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. MacFall, B. Fischl, and A. Dale, "Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data," Neuroimage **30**, 436–443 (2006).

[26]C. J. C. Bakker, M. A. Moerland, R. Bhagwandien, and R. Beersma, "Analysis of machine-dependent and object-induced geometric distortion if 2DFT MR imaging," Magn. Reson. Imaging **10**, 597–608 (1992).

[27]M. J. Clarkson, S. Ourselin, C. Nielsen, K. K. Leung, J. Barnes, J. L. Whitwell, J. L. Gunter, D. L. Hill, M. W. Weiner, C. R. Jack, Jr., N. C. Fox, and Alzheimer's Disease Neuroimaging Initiative, "Comparison of phantom and registration scaling corrections using the ADNI cohort," Neuroimage **47**, 1506–1513 (2009).