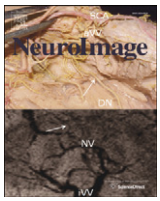
NeuroImage 60 (2012) 1296–1306

Contents lists available at SciVerse ScienceDirect

NeuroImage

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



Full Length Article

Analysis of automated methods for spatial normalization of lesioned brains

P. Ripollés [a,b,*], J. Marco-Pallarés [a,b], R. de Diego-Balaguer [a,b,d,e], J. Miró [a,b,c], M. Falip [c], M. Juncadella [c], F. Rubio [c], A. Rodriguez-Fornells [a,b,d]

[a] Cognition and Brain Plasticity Group [Bellvitge Biomedical Research Institute-] IDIBELL, L'Hospitalet de Llobregat, Barcelona, 08097, Spain
[b] Dept. of Basic Psychology, Campus Bellvitge, University of Barcelona, L'Hospitalet de Llobregat, Barcelona 08097, Spain
[c] Neurology Section, Hospital Universitari de Bellvitge (HUB), 08097, L'Hospitalet (Barcelona), Spain
[d] Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain
[e] Departement d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France and INSERM U955, UPEC, France

ARTICLE INFO

Article history:
Received 4 February 2011
Revised 16 January 2012
Accepted 18 January 2012
Available online 28 January 2012

Keywords:
Normalization
Epilepsy
Diffeomorphic
Cost function masking
Stroke
Unified Segmentation

ABSTRACT

Normalization of brain images is a crucial step in MRI data analysis, especially when dealing with abnormal brains. Although cost function masking (CFM) appears to successfully solve this problem and seems to be necessary for patients with chronic stroke lesions, this procedure is very time consuming. The present study sought to find viable, fully automated alternatives to cost function masking, such as Automatic Lesion Identification (ALI) and Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL). It also sought to quantitatively assess, for the first time, Symmetrical Normalization (SyN) with constrained cost function masking. The second aim of this study was to investigate the normalization process in a group of drug-resistant epileptic patients with large resected regions (temporal lobe and amygdala) and in a group of stroke patients. A dataset of 500 artificially generated lesions was created using ten patients with brain-resected regions (temporal lobectomy), ten stroke patients and twenty five-healthy subjects. The results indicated that although a fully automated method such as DARTEL using New Segment with an extra prior (the mean of the white matter and cerebro-spinal fluid) obtained the most accurate normalization in both patient groups, it produced a shrinkage in lesion volume when compared to Unified Segmentation with CFM. Taken together, these findings suggest that further research is needed in order to improve automatic normalization processes in brains with large lesions and to completely abandon manual, time consuming normalization methods.

Introduction

Spatial normalization is one of the most important steps in second-level group magnetic resonance imaging (MRI) analyses. Structural images of participants are normalized to a template (standard or group), ensuring that a one-to-one correspondence among the brains of each individual in the group is created. Normalization becomes more complex when it has to deal with patients with brain lesions. These brains have often greater differences than those individual variations characterizing healthy brains due to important lesions or pathologies (Brett et al., 2001). Correct normalization of individual brains is essential to ensure that brain areas are properly aligned, maximizing sensitivity and minimizing false-negative results. To this end, multiple normalization algorithms have been implemented in fully automated software programs.

Two of the most used normalization algorithms are the Diffeomorphic Anatomical Registration using Lie Algebra (DARTEL) (Ashburner, 2007) and its predecessor, Unified Segmentation (Ashburner and Friston, 2005), implemented in the Statistical Parametric Mapping software (SPM, Wellcome Department of Imaging Neuroscience, University College, London, UK, www.fil.ion.ucl.ac.uk/spm/). Unified Segmentation combines segmentation, bias correction and spatial normalization under the same iterative model using white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) tissue maps as priors (TPMs). These TPMs are deformed by a linear combination of a thousand cosine transform bases, and several Gaussian distributions are used to model the intensity of each tissue class. Unlike Unified Segmentation, DARTEL utilizes a large deformation framework to preserve topology, assuring that the deformations are invertible, diffeomorphic and parameterised by a flow field. Rather than using a thousand parameters for the registration process as Unified Segmentation, DARTEL uses about six million and the registration itself involves alternating between computing an average template of the GM and WM TPMs from all subjects and warping all subjects' TPMs into a better alignment with the template created (Ashburner, 2009). Both of these algorithms are segmentation-dependant, as

* Corresponding author at: Cognition and Brain Plasticity Group (IDIBELL), Campus Bellvitge, Feixa Llarga, s/n (08907), L'Hospitalet de Llobregat, Spain. Fax: +34 93 4024268.
E-mail address: pablo.ripolles.vidal@gmail.com (P. Ripollés).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.01.094

DARTEL needs the segmentations of all subjects in the group to create the average template and Unified Segmentation combines segmentation with normalization. Besides Unified Segmentation, another way to provide DARTEL with the GM and WM segmentations it needs is the New Segment toolbox under the SPM8 distribution. This algorithm is essentially the same as that described in the Unified Segmentation model, except for a different treatment of the mixing proportions, the use of an improved registration model, the ability to use multi-spectral data and an extended set of tissue probability maps (The FIL Methods Group, 2010). The default set includes TPMs for gray matter, white matter, CSF, bone, soft tissue and air/background, but allows the user to define as many extra TPMs as desired (The FIL Methods Group, 2010). New Segment can also provide deformation fields which can be later used to spatially normalize images, but in this manuscript it has been only used to provide DARTEL with the segmentations it needs.

In addition, as part of the Advanced Normalization Tools (ANTS) (Avants et al., 2011a), the Symmetric normalization algorithm (SyN) (Avants et al., 2008), also based on large deformations, has shown to perform at least as good as DARTEL when dealing with healthy subjects (Klein et al., 2009). SyN keeps symmetry when connecting two images in a geodesic (shortest distance in space) link, meaning that the path from A to B is the same than the one from B to A, irrespective of the optimisation or the similarity metric used (Avants et al., 2008). SyN has about 28 million degrees of freedom (Klein et al., 2009) and uses a gradient optimization scheme which is basically an iteration over time of three steps: computing the similarity gradient, updating the deformation field and regularizing the deformation field. Within ANTS, SyN can work with different similarity metrics as cross-correlation, mutual information or mean square difference (Avants et al., 2011b) and can use different types of regularization based on Gaussian or Bsplines. ANTS also provides optimal template construction and image segmentation, among other things.

In healthy subjects these normalization methods function optimally but, in contrast, spatial normalization suffers from some limitations when normalizing images of patients with large lesions, such as those found in stroke patients (Andersen et al., 2010) or in patients with tumors, cortical dysplasia or atrophy (Crinion et al., 2007). Different procedures have been used when trying to normalize abnormal brains. Initially, cost function masking (CFM) with standard SPM normalization was proposed for routine use when normalizing brains with regions containing abnormal signal intensities (Brett et al., 2001). Most normalization methods calculate a cost function, a measure of the signal intensity difference between a source image and a template, which has to be minimized (Brett et al., 2001). CFM is based on creating a binary mask of the lesioned area and taking the signal under the masked area out of the calculation of the transformations needed to normalize the image. Later, Crinion et al. (2007) proposed that, although for low regularization the use of the Unified Segmentation model with CFM provided a better registration than Unified Segmentation without CFM, when using medium regularization the use of CFM did not improve normalization. These results were assessed in a set of ten patients with different types of brain injuries (including tumor, stroke, cortical atrophy and dysplasia) and the regularization process was formulated as the precision of the bending energy priors on the deformation relative to the squared difference between the observed and normalized images under the Gaussian definition of noise. Nevertheless, a very recent study by Andersen et al. (2010) using a database of 49 chronic stroke patients showed that Unified Segmentation and medium regularization with CFM yielded better normalization results than those of Unified Segmentation with medium regularization alone, demonstrating the need for CFM when dealing with large lesions.

Interestingly, a different approach was recently presented by Seghier et al. (2008) in which adding an extra fourth tissue prior, which was defined as the mean of the cerebrospinal fluid (CSF) and white matter (WM) tissue probability maps (provided by SPM), improved the segmentation using the unified model. Unified Segmentation is computed twice: first, an iteration is computed with the predefined extra class, and then a second iteration is calculated with an updated definition (subject-specific) of the extra class (Seghier et al., 2008). Following this work, the Automated Lesion Identification (ALI) toolbox was developed, which allows the user to implement this type of normalization while also being capable of automatically identifying and delineating lesions.

Although spatial normalization of abnormal brains has been assessed with small-deformation methods such as Unified Segmentation, it has not yet been studied with large deformation techniques such as DARTEL, which have proven to achieve a better normalization in healthy volunteers (Klein et al., 2009; Tahmasebi et al., 2009; Yassa and Stark, 2009). Alternatively, the SyN method allows the possibility to use a technique called constrained cost function masking (CCFM) in order to normalize lesioned brains (Kim et al. 2007), which has been reported to give optimal results when normalizing a group of stroke patients (Schwartz et al., 2009). As in CFM, a lesion mask must be depicted and applied to the registration. In CCFM, the velocity field parameters calculated by SyN under the mask are treated as unknown values and estimated using the information given by the velocity fields of the healthy tissue near the masked area.

In addition to the methodological issues previously discussed, there is a limitation in the generalization of the results obtained by previous studies to all brain lesions. It is important to note that most of the previous analysis dealt with patients with a variety of brain injuries, especially with lesions derived from vascular events, but none of them directly investigated these issues with a set of patients with large brain resections. After surgery, CSF mostly invades the empty space left by the resected tissue; however, the remaining brain tissue also tries to occupy this new empty volume. These variations produce a great variety of morphological brain changes, which might create serious difficulties when trying to normalize images of these injured brains. Because many studies involve tumor-resected patients or epileptic patients with removal of the epileptic focus, these problems are not infrequent (Cheung et al., 2009; Immonen et al., 2010; Yogarajah et al., 2010). Achieving an optimal normalization for these patients is crucial, especially for studies aiming to compare pre- and post-surgery brains.

The first aim of the present study was to evaluate which of the previously described approaches is more effective for abnormal brain normalization. We are particularly interested in automated methods that could lead to the abandonment of manual lesion depicting methods such as CFM or CCFM, which are extremely time consuming as the lesions must be defined manually by an expert neurologist or neuroradiologist and might also be subject to expert biases. We expected that the more sophisticated approach given by DARTEL (Klein et al., 2009; Tahmasebi et al., 2009; Yassa and Stark, 2009) and SyN (Klein et al., 2009), which has proven to provide a better normalization than other traditional SPM based methods in healthy subjects, would achieve a more accurate normalization. We also expected that using an extra class as a prior would cause the ALI toolbox to improve normalization results compared to regular Unified Segmentation. The second aim of this study was to determine the impact of the type of lesion in the performance of the normalization methods, especially when dealing with brain-resected patients. With this purpose in mind, normalization methods were tested on three different groups of participants (healthy patients, stroke patients and epileptic patients with brain-resected regions) to study whether there are differences in normalization performance of the different methods when dealing with different types of lesions.

Materials and methods

Subjects

Three different sets of patients participated in this study: (1) ten stroke patients (8 chronic, 2 acute) who suffered focal damage due

to a single vascular event (2 women; mean age 66.3; age range 46–75; mean time since ictus 45.3 months, see Table 1); (2) ten mesial temporal lobe epileptic patients who were therapeutic drug resistant and had undergone surgery (temporal lobectomy or amygdalectomy) to have the epileptic focus regions resected (6 women; mean age 44.7; age range 29–66, see Table 1); (3) twenty-five healthy participants (17 women, mean age 34.9, age range 21–71).

MRI data acquisition

Structural MRI data were obtained using a 3.0 Tesla Siemens Trio MRI system. A whole brain 3D magnetisation-prepared rapid gradient echo (MPRAGE) scan was acquired for each individual of each group (flip angle = 9°, TR = 2300 ms, TE = 3 ms, voxel resolution = 1.00 × 1.00 × 1.00 mm, field of view = 244 mm).

Dataset creation and patient normalization

One way to assess the goodness of normalization when dealing with lesioned brains is to apply a lesion to a healthy brain and then compare the normalization parameters obtained for the healthy and lesioned versions of the image (Andersen et al., 2010; Brett et al., 2001; Crinion et al., 2007; Nachev et al. 2008). With this aim, a dataset of 500 artificially lesioned images was created using the ten stroke lesions, the ten resections and the twenty-five brain images from the healthy subjects.

As in Nachev et al. (2008) we inserted, in each healthy brain image, 20 different lesions (10 from the stroke patients and 10 from the resections), thus yielding 500 images. We first normalized the twenty images from both the resected and stroke groups using Unified Segmentation with CFM. Because the use of cost function masking has proven to be effective in the normalization of abnormal brains, a trained neurologist created binary masks of the lesioned areas by manually depicting the precise boundaries of the lesion directly into the T1 image using the MRIcron software (Rorden and Brett, 2000; http://www.cabiatl.com/mricro/mricron/index.html). It has been shown that using CFM with roughly or precisely defined masks as well as smoothed or unsmoothed masks has no effects on normalization (Andersen et al., 2010). In the present study, smoothing was not applied to the masks, and they were drawn as accurately as possible. All structural images of both stroke and resected epileptic patients groups, with their respective masks, are displayed in Figs. 1a and b, respectively.

Table 1
Demographic information and lesion description for the patient dataset.

Group	Patient	Sex	Age	Lesion location	Time since ictus (months)
Stroke	S1	Male		Left middle cerebral artery	2
Stroke	S2	Female	68	Right middle cerebral artery	10
Stroke	S3	Male	75	Right middle cerebral artery	92
Stroke	S4	Male		Left Frontal Haematoma	4
Stroke	S5	Male	61	Right middle cerebral artery	87
Stroke	S6	Male	68	Left thalamus	12
Stroke	S7	Male	69	Left capsule	172
Stroke	S8	Male	58	Left middle cerebral artery	60
Stroke	S9	Male	70	Right thalamus	6
Stroke	S10	Female	51	Left basal ganglia	8

Group	Patient	Sex	Age	Lesion location	Time since surgery (months)
Epileptic	R1	Male	45	Right temporal lobectomy	3.5
Epileptic	R2	Female	37	Left temporal lobectomy	4
Epileptic	R3	Male	52	Right temporal lobectomy	3.5
Epileptic	R4	Male	37	Right temporal lobectomy	3
Epileptic	R5	Female	29	Left amygdalectomy	3.5
Epileptic	R6	Female	35	Left amygdalectomy	4
Epileptic	R7	Female	51	Right temporal lobectomy	3.5
Epileptic	R8	Female	62	Left temporal lobectomy	4
Epileptic	R9	Female	33	Left temporal lobectomy	3
Epileptic	R10	Male	66	Right temporal lobectomy	3

All the twenty-five images from the healthy group were also brought to MNI space using Unified Segmentation. Once in the same space, lesions were inserted into healthy brains and, using the inverse transformation obtained for the healthy brain images in the previous step, they were brought back to native space. To take into account the difference in intensities between lesioned and healthy brain images, we applied a correction factor to the lesion to be inserted prior to reslicing the artificially simulated image to native space. This factor was calculated, as in Brett et al. (2001), as the division of the mean of all brain tissue voxels outside the place where the lesioned area should be inserted in the healthy brain divided by the mean of all brain tissue voxels outside the lesioned area in the lesioned brain. After repeating this process for all healthy brain images we obtained two large datasets of artificially generated brain images: 250 images from stroke lesions and 250 images from resected lesions.

SPM8 was used to implement all procedures except for SyN which works under ANTS. In every case, each structural image was normalized to the MNI stereotaxic space using the ICBM452 template, except for the registrations carried out by SyN which were normalized to a subject-specific template (see below). All the parameters in the algorithms described in this section were used when normalizing both the resected and stroke groups, as well as the healthy dataset. For more details on the code used to implement the algorithms please see Supplementary Material Section 3.

Ten different types of segmentation–normalization combinations were then applied to the resected and stroke groups of artificially generated lesions (see Table 2). In total, almost 5000 registrations were computed. First, combinations involving Unified Segmentation were used. The first approach implemented was Unified Segmentation, as it has been proven to be a good option when normalizing abnormal brains (Crinion et al., 2007). Using the lesion masks, Unified Segmentation with CFM was performed (Crinion et al., 2007). Default parameters were used for segmentation (number of Gaussians per class, Bias FWHM or warp frequency cut-off), except for regularization which was set to medium for both normal and CFM approaches.

An extra class, the result of the mean of the WM and CSF tissue probability maps provided by SPM, has been demonstrated to give the segmentation procedure more flexibility when dealing with damaged tissue (Seghier et al., 2008). However, this extra class was an empirical definition that worked well for a specific dataset. Different datasets with different contrasts and different types of lesions or locations can respond better to a different extra class, and priors can be adapted to increase accuracy of the normalization (Mohamed Seghier, personal communication, October 13 2010). To see the effect of using different extra classes as priors, the ALI toolbox was used with two different extra classes, one of which was the mean of the WM and CSF TPMs and the other was the mean of the GM and CSF TPMs provided by SPM8. Adding the GM based extra prior did not improve ALI toolbox performance (see Supplementary Tables 1–3). Normalization within the unified model was performed, as ALI normalization is basically a Unified Segmentation run twice with an extra prior. Threshold probability and threshold size were kept to default values and as a small number of iterations seems to be a reasonable trade-off between definition of the extra class and proper classification of normal tissue (Seghier et al., 2008) it was also set to the default which is 2 iterations.

The GM and WM tissue maps provided by these three previous combinations under the Unified Segmentation model (medium regularization with and without cost function masking and ALI toolbox) were used to provide DARTEL with the needed initial segmentations. In addition, New Segment, was also used to provide for GM and WM images. All the segmentation parameters used in New Segment were

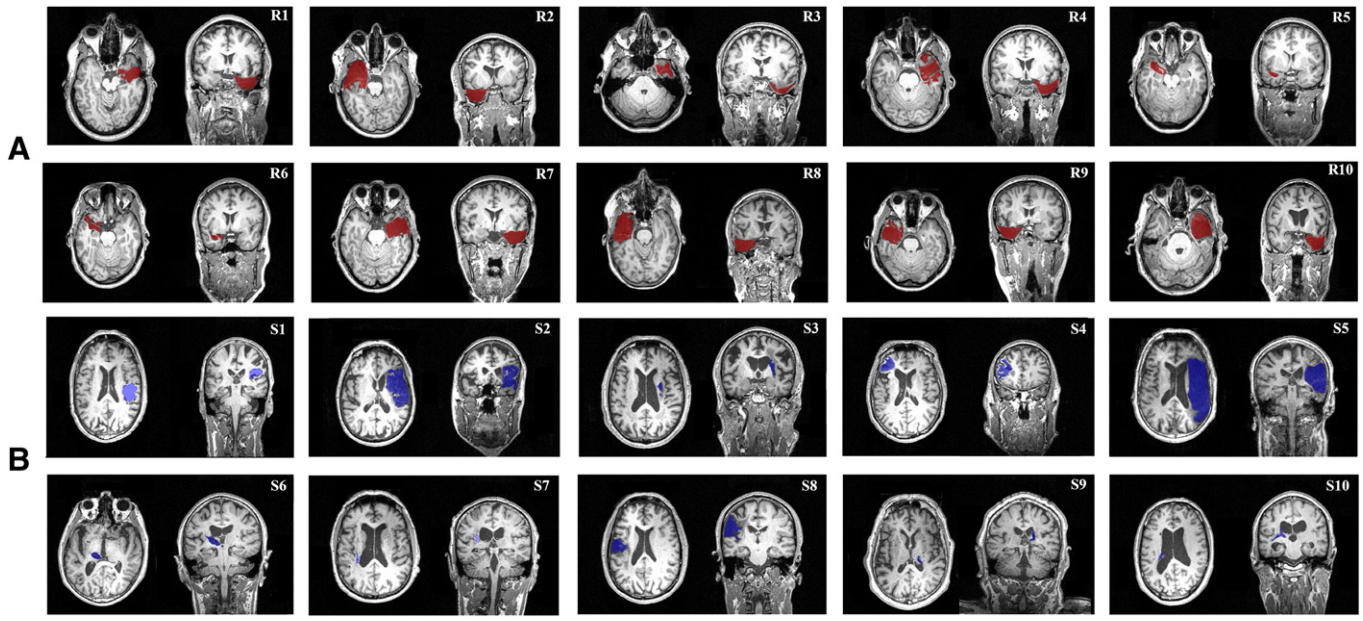


Fig. 1. A. T1 axial and coronal slices of the group of drug-resistant epileptic patients illustrating the temporal lobe and amygdala regions resected (red masks). B.T1 axial and coronal slices of the group of chronic stroke patients with their respective lesion masks covering all of the injured areas (blue masks). Neurological convention is used in both datasets.

the default except again for regularization which was set to medium as in Unified Segmentation (Crinion et al., 2007).

An extra segmentation was performed under New Segment, following the rationale of the ALI toolbox (Seghier et al., 2008). An extra tissue class defined as the mean of the WM and CSF TPMs was added to the six default TPMs (results of New Segment with an extra prior defined as the mean of the GM and CSF TPMs are provided in the supplementary material). These two sets of GM and WM tissue maps provided by New Segment were also used as an input to implement DARTEL normalization. All parameters were kept to default except for the regularization term. DARTEL penalizes the registration with an energy form which can be linear elastic, membrane or bending. The default option, linear elastic, was chosen as it was also the one used by Klein et al. (2009) in their evaluation of the performance of DARTEL when dealing with healthy subjects. DARTEL linear elastic regularization allows the user to increase a parameter μ which will penalize shearing and scaling (Ashburner, 2007). Four times the default μ value was used after carrying out a small parameter search, as increasing regularization has improved other algorithms performance in past studies (Crinion et al., 2007). For this parameter search, ten different stroke artificially generated images were first normalized using Unified Segmentation with CFM. The gray and white matter images generated were provided to DARTEL and the artificially generated lesioned images and their healthy counterparts were normalized. Three different regularizations were applied using twice, four times and default DARTEL regularization. Normalizations parameters between healthy and lesioned images were compared in order to assess the effect of augmenting the regularization and four times default was chosen (see Results section).

SyN with and without CCFM was also used to normalize both lesioned groups. Cross-correlation as a similarity metric, Gaussian regularization and 30 iterations for the first coarsest, 99 for the next and 11 at full resolution were chosen, as these were the ones used by Klein et al. (2009) in their normalization study. Gradient-descent step was kept to default value 0.25 and, as recommended in ANTS manual (Avants et al., 2011b), the radius of the correlation window in the similarity metric was increased to eight which has been shown to improve registration performance when dealing with brains with severe damage.

As in all Unified Segmentation and DARTEL based normalizations, all 500 simulated lesions using SyN CCFM were initially registered to the ICBM452 template, but more than a hundred normalizations ended with awkward distortions. This result was probably obtained because the ICBM452 template is too smooth and lacks the anatomical detail needed to be used as target template for SyN. Then, ANTS utility for template creation (Avants et al., 2009) was used to create a template with the 25 healthy subjects' images, which was later used as the target template for SyN with and without CCFM.

Table 2
Segmentation–normalization procedures used in the study. With DARTEL, the GM and WM images generated by the different methods in the table are entered as input. Methods marked with an asterisk are also used to normalize the healthy group.

Unified segmentation	Medium regularization*
	Medium regularization CFM
	ALI toolbox + (WM + CSF) /2*
DARTEL	Medium regularization*
	Medium regularization CFM
	ALI toolbox + (WM + CSF) /2*
	New Segment*
	New Segment + (WM + CSF) /2*
ANTS	SyN*
	SyN CCFM

Healthy participants' normalization

For the set of 25 healthy volunteers, only seven different types of segmentation–normalization combinations were applied, as the methods involving CFM or CCFM were not performed. In Table 2 methods marked with an asterisk are the ones also assessed for healthy participants.

Performance evaluation

Visual inspection of the normalized images was performed on every single image to identify poor normalization and segmentation. In addition, root mean square displacement (RMSD) was used to quantify the performance of the different normalization methods in the 250 artificially generated images of the resected and stroke groups. To compute it, a deformation field, which is an image with a

mapping from a template to a source, was generated from the normalization parameters for every single registration calculated. Then, the deformation field for healthy subject *H* normalized with algorithm *A* and the deformation field of image *L*, artificially generated from image *H*, and also normalized with algorithm *A*, were compared. This was done by calculating the summed squared differences (distance) in each direction x, y and z between voxel *n* of image *H* and its counterpart on image *L*. Finally the RMSD (Eqs. 1 and 2) of all the differences in the voxels was taken (Brett et al., 2001):

$$D_n = \sqrt{\left(x_n^{LA} - x_n^{HA}\right)^2 + \left(y_n^{LA} - y_n^{HA}\right)^2 + \left(z_n^{LA} - z_n^{HA}\right)^2} \quad (1)$$

$$RMSD = \sqrt{\sum_{n=1}^{T} {(D_n)^2}\Big/{T}} \quad (2)$$

where D_n is the distance and T is the total number of voxels in the image. This measure was obtained for each of the 500 artificially generated images and each of the 10 segmentation–normalization algorithms (see Table 2). In some cases, normalization of lesion volume is especially important, as for example, when trying to make inferences about the behavioral effect of a lesion. For example, lesion behavioral mapping (Bates et al., 2003; Rorden et al., 2007), has been widely used to report relationships between brain injury location and behavioral measures. Usually, in this type of analysis, all patients' lesion definitions (LD) must necessarily be in the same normalized space. Thus, normalization of lesion volume becomes critical. Therefore, RMS displacement measure was calculated inside and outside the lesion as well as for the whole brain tissue. To this purpose, each of the LDs, which was manually depicted in native space, was resliced into normalized space obtaining a normalized lesion definition (NLD). To perform these transformations, the normalization parameters calculated previously for the ten normalization combinations analyzed (see Table 2) for every simulated lesion in both patient datasets were used, and one NLD per simulated lesion and normalization algorithm was obtained. RMSD between healthy and lesioned brains was calculated inside and outside these NLDs for all subjects in both stroke and resected groups. Repeated measures ANOVA was performed to check for a significant effect of the method used and Greenhouse–Geisser correction was applied. Paired t-tests were subsequently calculated between pairs of methods to evaluate significant differences between them.

RMSD was also calculated for the transformations done within the regularization parameter search for DARTEL and repeated measures ANOVA with Greenhouse–Geisser correction was performed for DARTEL with default, double and four times regularization. After this, paired *t*-tests were also calculated between the different regularized DARTEL data.

Post-normalized lesion volume has also proven to be a good way to quantitatively assess the goodness of normalization (Andersen et al., 2010). Thus, for each of the ten algorithms, to account for differences in lesion size post-normalization, the volume of 200 lesions after normalization (10 strokes and 10 resections per 10 algorithms) were manually delineated. While a direct comparison can be done for all SPM based methods because they are normalized to the same MNI template, a correction factor must be applied to the lesion volumes calculated on images using ANTS based methods, as they are normalized to a subject-specific template. This factor was calculated as the division of the number of brain tissue voxels in each healthy subject image normalized with an SPM based method and the same parameter for ANTS based normalized images. Percent of change of each volume respect to the volume of Unified Segmentation CFM was computed, as it is considered the gold-standard (Andersen et al., 2010). Repeated measures ANOVA with Greenhouse–Geisser correction was performed to check for a significant effect of the nine remaining methods. Paired *t*-tests were also calculated between pairs of methods to evaluate significant differences. As manual definition of lesioned areas can introduce a bias, lesion volume was also calculated using the previously defined NLDs, only for the same lesions that were manually depicted in normalized space. Paired t-tests were performed, for each algorithm, between manually defined lesions in normalized space and NLDs.

The non-use of a lesion mask has proven to distort the area around the lesion (Brett et al., 2001; Kim et al., 2007). To control for this effect, Dice similarity index (DSI; Dice, 1945) was used to calculate the area of overlap among the normalized lesion definitions for methods using and not using a lesion mask. DSI ranges between 0 and 1 (being 0 no overlap and 1 a perfect similarity) and takes into account both false negatives and false positives. DSI between masked (M) and non-masked (NM) algorithms was calculated as twice the overlap between both NLDs divided by their sum (Eq. 3):

$$DICE = \frac{2(M \cap NM)}{(M + NM)} \quad (3)$$

To calculate this metric, all NLDs were used. For the SPM based algorithms DSI was computed taking Unified Segmentation with CFM as the ground truth as it is the only algorithm using a lesion mask. For the same reason, for ANTS based algorithms, SyN with CCFM was taken as the reference. Results were compared using the non-parametrical Mann–Whitney U test (Wilke et al., 2011).

For the healthy subjects group, normalized cross-correlation (NCC) was used, as it has proven to be a good measure of similarity between warped and template images (Tahmasebi et al., 2009). In this case, it was used to assess the similarity and thus the accuracy of the normalization process among all warped images within the healthy group. First of all, a pair of images was normalized to have a mean of zero and unit variance, and then a cross correlation was computed following (Eq. 4):

$$NCC(X,Y) = \frac{\sum_{i=1}^{N} (x_n - \mu_x)\left(y_n - \mu_y\right)}{\sigma_x \sigma_y} \quad (4)$$

where *N* is the number of voxels, x_n and y_n are the values of voxel n in the images X and Y, μ_y and μ_x are their respective means and σ_x and σ_y are their respective standard deviations. For each computation between two different warped images, a whole brain NCC score was obtained. Within the healthy group, NCC scores were computed by comparing each warped image against every single other image in the group, therefore obtaining *n!/2(n-2)!* NCC scores, where *n* stands for the number of subjects in the dataset (i.e. 25). Because each segmentation–normalization combination uses the same pairs of images to compute each NCC score, once again repeated measures ANOVA with Greenhouse–Geisser correction was performed before paired t-tests were calculated.

Results

Visual inspection

Visual inspection of the GM and WM images showed poor segmentation and normalization for very few images in three different algorithms: Unified Segmentation without CFM (five resections and four strokes), Unified Segmentation with CFM (five resections and four strokes) and ALI toolbox with (WM + CSF)/2 (five resections and four strokes). All these images were artificially generated from the same healthy subject. The transformations calculated for these images were not included in the analysis and their respective GM and WM segmentations were not provided to DARTEL as inputs either. A visual example of the ten normalization methods on a resection can be seen in Fig. 2.

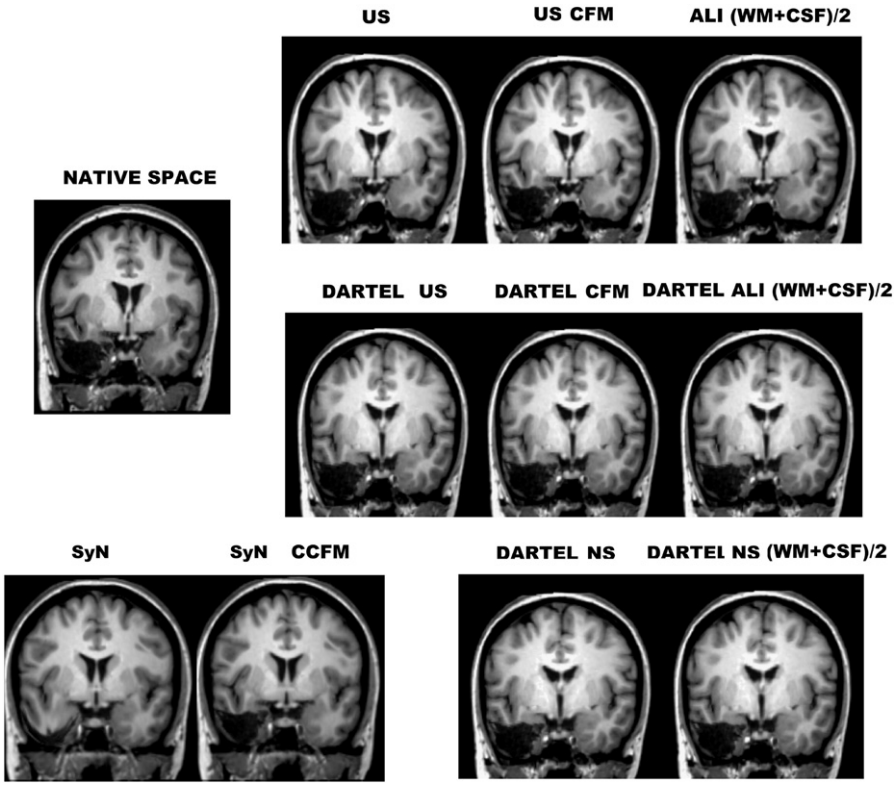


Fig. 2. Ten different normalizations applied to the same artificially generated image from the resected dataset. US: Unified Segmentation. CFM: Cost Function Masking. ALI: Automatic Lesion Identification. WM: White matter. CSF: Cerebrospinal fluid. DARTEL: Diffeomorphic Anatomical Registration using Exponentiated Lie algebra. SyN: Symmetrical Normalization. CCFM: Constrained Cost Function Masking. NS: New Segment.

From visual assessment it is hard to evaluate which normalization procedure performed better as they all seem to do a proper registration. Nonetheless, a closer look shows that some of the algorithms warp some lesioned areas out of what brain tissue should be. This specially happens when the lesion is near the skull, as in the resected dataset. To get a better view on this effect, brain tissue masks were created from the ICBM452 template used for SPM based methods and from the subject-specific template used for ANTS based methods (see below). In Fig. 3, the lesion volume depictions of the same resected image from Fig. 2 are shown over their respective brain tissue maps. It can be seen that Unified Segmentation without CFM, ALI + (WM + CSF) / 2, DARTEL with Unified Segmentation with and without CFM and DARTEL with ALI + (WM + CSF) / 2 all warp the lesioned area away from what the normalized brain tissue should be causing an out-of-brain distortion (OoBD) effect. Because of this OoBD effect, only voxels within a defined brain tissue mask (BTM) were later computed when calculating the lesion volumes post-normalization. For SPM based methods a resliced version, to match voxel size of normalized T1s, of the standard SPM8 brain tissue mask was used as TBM. On the other hand, for SyN based methods, a specific TBM was created after segmenting, binarizing and smoothing the subject specific template created using ANTS. Importantly, the volumes reported here are the volumes of the lesion, masked to include only voxels inside the brain. If an algorithm has a tendency to create an OoBD effect, the volume measure presented here will not penalize this algorithm as long as the remaining volume inside the brain is similar to the standard.

DARTEL regularization parameter search

Repeated measures ANOVA analysis with Greenhouse–Geisser correction on the RMSD measures for the ten stroke images used in the parameter search normalized with DARTEL with default, double and four times regularization showed a significant effect of the regularization ($F(2,18) = 11.284$, $p<0.01$, see Table 3). DARTEL with 4 times regularization had the lowest RMSD score, which was lower than DARTEL with default and double regularization ($t(9) = -3.6$, $p<0.01$ and $t(9) = -3.2$, $p<0.05$, respectively). When DARTEL with segmentations provided by Unified Segmentation with CFM and four times regularization is compared to the RMSD scores for Unified Segmentation with CFM, which is supposed to be the gold standard, a paired *t*-test failed to find significant differences between them.

Epilepsy group

Repeated measures ANOVA analysis with Greenhouse–Geisser correction on root mean squared displacement scores for the resected epileptic group showed a significant effect of the normalization method ($F(9,2196) = 58.322$ $p<0.001$, see Fig. 4). DARTEL with New Segment + (WM + CSF) / 2 as input was found to have the lowest RMSD score, which was lower than Unified Segmentation with CFM and SyN with CCFM ($t(244) = -6.99$, $p<0.001$ and $t(249) = -7.69$, $p<0.001$, respectively). Unified Segmentation with CFM obtained a lower RMSD score than SyN with CCFM ($t(244) = -2.29$, $p<0.03$). RMSD scores outside the lesion were in agreement with whole brain scores, being DARTEL with New Segment + (WM + CSF) / 2 the normalization method that achieved the smallest RMSD. However, inside the NLDs, Unified Segmentation with CFM was found to achieve the lowest RMSD score (1.35 mm error), compared to SyN with CCFM and DARTEL with New Segment + (WM + CSF) / 2 as input ($t(244) = -5.24$, $p<0.001$ and $t(244) = -4.45$ $p<0.001$, respectively), which achieved the second lowest errors. There were not significant differences between DARTEL with New Segment + (WM + CSF) / 2 (1.63 mm error) and SyN with CCFM (1.57 mm error) scores ($t(249) = 0.60$, $p<0.54$). A detailed report on this analysis can be found in Section 1 and Table S4 of the Supplementary Material.

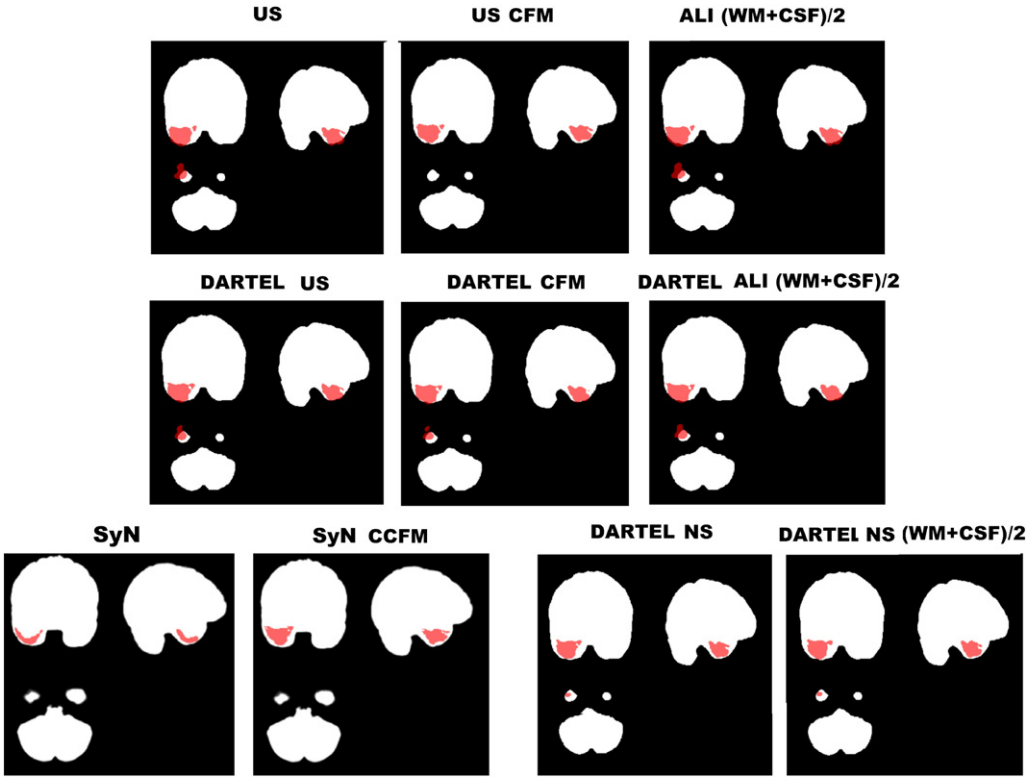


Fig. 3. Lesion volumes over brain tissue maps for the different normalizations of the same artificially generated image shown in Fig. 2 from the resected dataset.

All *t*-test comparisons failed to show differences in volume between the manual depicted lesions in normalized space and their respective NLDs. Mean volume for the NLD analysis are included in Table S6 of the Supplementary Material. The analysis of the volume of the lesions manually depicted in normalized space revealed that there were significant differences between normalization methods (see Fig. 7A). Repeated measures ANOVA showed a significant effect of method ($F(8,72)=96.6$, $p<0.001$). Fig. 7A revealed that clearly, SyN presented smaller volume differences compared to the other methods ($t(9)>12$, $p<0.001$). If this method was removed from the comparison, differences in volume change among methods still remained significant ($F(7,63)=6.6$, $p<0.01$). The method that presented the smallest change in volume was Unified Segmentation with medium regularization compared to the other methods ($t(9)>2.3$, $p<0.05$) except to ALI toolbox ($t(9)=0.96$, *n.s.*) which the differences were not significant.

All algorithms achieved high DSI values (see Table 4) with Unified Segmentation without CFM and ALI + (WM + CSF)/2 achieving the highest scores. No significant DSI differences ($p<0.7$) were found between them, but their respective DSIs were significantly higher than the ones obtained by the other algorithms (all $p<0.001$). SyN without CCFM obtained the lowest (all $p<0.001$) DSI.

Table 3
Root mean square displacement (RMSD) for the ten stroke artificially generated images within the parameter search data.

Method	Regularization	Mean
Unified Segmentation with CFM	Medium	0.56
DARTEL CFM	Default	0.89
	Double default	0.79
	Four times default	0.52

Stroke patients

Repeated measures ANOVA analysis with Greenhouse–Geisser correction on root mean squared displacement scores for the stroke group showed a significant effect of the normalization method ($F(9,2205)=259.963$, $p<0.001$, see Fig. 5). DARTEL with New Segment + (WM + CSF)/2 as input was again the method with the lowest RMSD score, which was lower than Unified Segmentation with CFM and SyN with CCFM ($t(245)=-15.84$, $p<0.001$ and $t(249)=-14.69$, $p<0.001$, respectively). No differences were found between the RMSD scores of Unified Segmentation with CFM and SyN with CCFM. In this case, the pattern of results inside and outside NLDs was exact to the one obtained from the whole brain analysis, being DARTEL with New Segment + (WM + CSF)/2 the normalization algorithm which achieved the lowest RMSD score. However, inside NLDs, no differences were found between

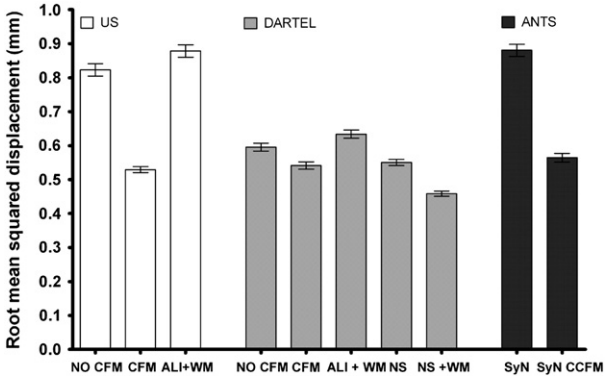


Fig. 4. Mean (and standard error of the mean, SEM) for RMSD scores for all algorithms and the epileptic resected dataset. Lower values indicate better performance of the normalization procedure.

Table 4
DSI mean and SD for epileptic and stroke datasets. SPM based methods are compared against Unified Segmentation with CFM, while ANTS based results are compared against SyN with CCFM.

Method		Resected epileptic	Stroke
Unified Segmentation	Med. regularization	0.90 ± 0.03	0.97 ± 0.03
	ALI + (WM + CSF)/2	0.89 ± 0.03	0.93 ± 0.03
DARTEL	Med. regularization	0.88 ± 0.04	0.87 ± 0.03
	Med. regularization CFM	0.90 ± 0.02	0.88 ± 0.03
	ALI + (WM + CSF)/2	0.87 ± 0.02	0.86 ± 0.06
	New Segment	0.88 ± 0.02	0.86 ± 0.06
	New Segment + (WM + CSF)/2	0.85 ± 0.02	0.84 ± 0.07
ANTS	SyN	0.80 ± 0.05	0.81 ± 0.07

the RMSD scores ($p<0.5$) achieved by Unified Segmentation with CFM (1.07 mm error) and the ones obtained by DARTEL with New Segment + (WM + CSF)/2 (1.03 mm error). A complete report on this analysis can be encountered in Section 1 and Table S5 of the Supplementary Material.

Once again, *t*-test comparisons failed to find differences in lesion volume between NLDs and manually depicted lesions in normalized space in the stroke group. Mean volume for the NLDs can be found in Table S6 of the Supplementary Material. Repeated measures ANOVA on the lesion volume change for manually depicted lesions in normalized space revealed that there were significant differences between normalization methods ($F(8,72)=31.1$, $p<0.001$; see Fig. 7B). Fig. 7B shows that again, SyN presented greater decrease of volume compared to all the other methods ($t(9)>10$, $p<0.001$). The repeated measures ANOVA without the SyN method also showed significant differences between methods ($F(7,63)=5.5$, $p<0.01$). In the stroke group, the method that presented smaller change in volume was the ALI toolbox. It showed smaller changes than New Segment ($t(9)=5.6$, $p<0.001$) and DARTEL Medium Regularization ($t(9)>5.0$, $p<0.001$), and marginal significant differences compared to New Segment WM ($t(9)=1.96$, $p<0.1$) and SyN CCFM ($t(9)=1.90$, $p<0.1$).

Once again, all algorithms obtained high DSI scores and Unified Segmentation without CFM showed the highest one. This score was significantly higher than the ones obtained by the rest of the other normalization algorithms (all $p<0.001$). As in the resected dataset, SyN without CCFM obtained the lowest (all $p<0.001$) DSI.

Healthy participants

Repeated measures ANOVA analysis on NCC scores for the healthy group showed a significant effect of the normalization method ($F(6,1794)=220.760$, $p<0.001$, see Fig. 6). SyN was the method with the highest NCC score, which was higher than DARTEL with New Segment and Unified Segmentation ($t(299)=3.48$, $p<0.001$ and $t(299)=14.69$, $p<0.001$, respectively).

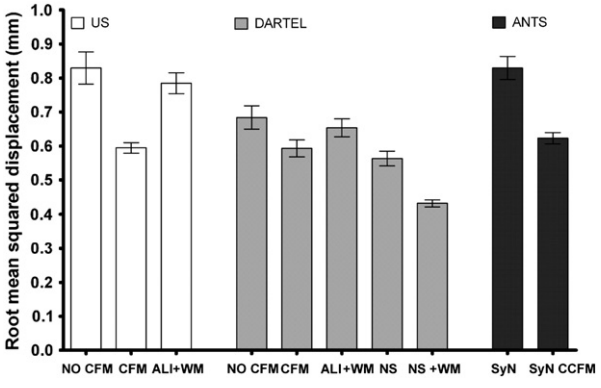


Fig. 5. Mean (and SEM values) RMSD scores for all algorithms and the stroke dataset. Lower values indicate better performance of the normalization procedure.

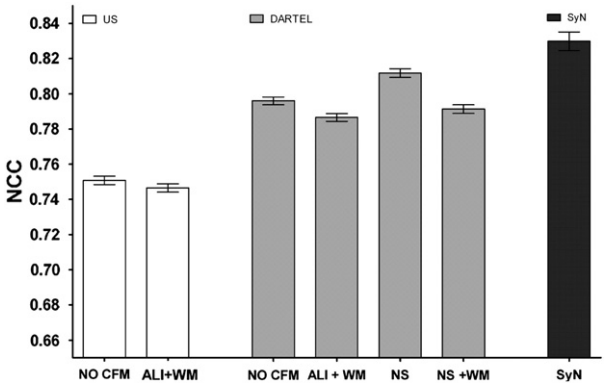


Fig. 6. Mean (and SEM values) NCC scores for all algorithms, except for those using CFM or CCFM, for the healthy group dataset. Higher values indicate better performance of the normalization procedure.

Discussion

In the present study, ten different methods for normalizing brains with lesions were compared in three different sets of participants, including brain-resected drug-resistant epileptic patients, stroke patients and healthy individuals. With this aim, two datasets of 250 artificially generated lesioned brains were created from the lesions of the two groups of patients (ten stroke subjects and ten brain resections) and the twenty-five brains of the healthy subjects, yielding a total amount of 5000 normalized brains. When assessing normalization performance in lesioned brains, it is not possible to know which algorithm has the lowest error compared to some unknown gold standard. Nevertheless, one can compare variability of the algorithms across brains and between lesioned-normal images. For this purpose, root mean square displacement and post normalization lesion volume changes have been widely used (Andersen et al., 2010; Brett et al., 2001; Crinion et al., 2007; Nachev et al. 2008). Using these measures of displacement and volume changes after normalization, the overall results indicated that there is not a single algorithm which outperforms the others across healthy and patient groups, and across measures.

When dealing with *healthy subjects*, SyN obtained the highest NCC mean score, followed by DARTEL with New Segment. For both *resected and stroke datasets*, the lowest RMSD value (whole brain, inside and outside lesion) was obtained by DARTEL using New Segment with (WM + CSF)/2 as a prior followed by Unified Segmentation with CFM and SyN with CCFM, except for the RMSD scores inside lesion volume for the resected dataset, in which Unified Segmentation with CFM achieved the lowest error. However, for the post-normalization lesion volume measure, Unified Segmentation with CFM showed the highest mean in both patient groups. In addition, only DARTEL with its two New Segment-based implementations, SyN with and without CCFM and Unified Segmentation with CFM do not suffer from an out-of-brain distortion effect which displaces lesioned voxels out of the brain tissue normalized area (see Fig. 3). All SPM non-masked algorithms obtained high DSI values for both resected and stroke datasets, with respect to Unified Segmentation with CFM (see Table 4). SyN without CCFM obtained the lowest DSI score, which is in agreement with previous results that showed that SyN without CCFM produced distortions around lesioned areas (Kim et al., 2007).

Recent studies have demonstrated that DARTEL (Klein et al., 2009; Tahmasebi et al., 2009; Yassa and Stark, 2009) and SyN (Klein et al., 2009) perform better than Unified Segmentation when dealing with healthy subjects. More precisely, Tahmasebi et al. (2009) showed that the NCC score of DARTEL was significantly increased by 3% over the NCC score of Unified Segmentation in the registration of 17 subject images against a given template. Our NCC scores for the healthy dataset are consistent with these findings, as DARTEL[1] with New Segment and SyN had a 6% and a 8% greater NCC score than Unified Segmentation, respectively (see Fig. 6). Klein et al. (2009) showed how SyN was as good as DARTEL or even better when dealing with healthy subjects. This is also consistent with our results as SyN yields the highest NCC mean score, almost a 2% better than DARTEL with New Segment.

Although there are studies of normalization dealing with chronic stroke patients (Andersen et al., 2010) or with patients with a wide range of lesions types (Brett et al., 2001; Crinion et al., 2007), to our knowledge, this is the first time normalization methods have been tested in a group of individuals with brain-resected regions (a group of therapeutic drug resistant mesial temporal lobe epileptic patients). Different results have been obtained using Unified Segmentation with CFM on different groups of patients with different types of lesions. Although Crinion et al. (2007) showed that there was not an improvement in registration between Unified Segmentation with medium regularization with and without CFM for patients with a range of lesions types, Andersen et al. (2010) showed that CFM obtained better results over medium regularization when normalizing a large group of chronic stroke patients. These results seem to show that when using Unified Segmentation, CFM becomes the gold standard when treating with patients that have large lesions and significant enlargement of CSF filled spaces, such as those found in individuals suffering from chronic stroke (Andersen et al., 2010). The results of the present study are consistent with the findings reported in Andersen et al. (2010) as Unified Segmentation with CFM showed a RSMD mean score almost 0.29 mm lower than Unified Segmentation without CFM for the post-surgery epileptic group (see Fig. 4) and 0.24 mm lower for the stroke dataset (see Fig. 5). However, none of these Unified Segmentation-based methods improved the performance of DARTEL with New Segment with (WM + CSF)/2 as a prior which yielded the lowest RMSD scores, being 0.07 mm and 0.16 mm lower than Unified Segmentation with CFM for resected and stroke datasets, respectively. Using the ALI toolbox did not differ from using Unified Segmentation alone, although for the stroke dataset it was the algorithm obtaining the smallest mean volume reduction post-normalization.

One might think, looking as the RMSD results (see Figs. 4 and 5), that DARTEL with New Segment using (WM + CSF)/2 as a prior, which is fully automated, is the best choice to normalization of lesioned brain images, followed by two non-automatic methods as Unified Segmentation with CFM and SyN with CCFM which need lesion mask depiction. However, the post-normalization lesion volume measure (Fig. 7) casts doubt on the former statement as Unified segmentation with CFM gets the highest mean volume for both resection and stroke groups. In fact, when compared with Unified Segmentation with CFM, DARTEL with New Segment using (WM + CSF)/2 as a prior produces a shrinkage in lesion volume of about a 10% and a 14% for resected and stroke datasets respectively (see Fig. 7). Therefore the present results seem to suggest that there is a complex trade-off among methods: an automatic method such as DARTEL, provides lower RMSD values but higher reduction of lesions, while those methods providing better conservation of lesion volumes (such as

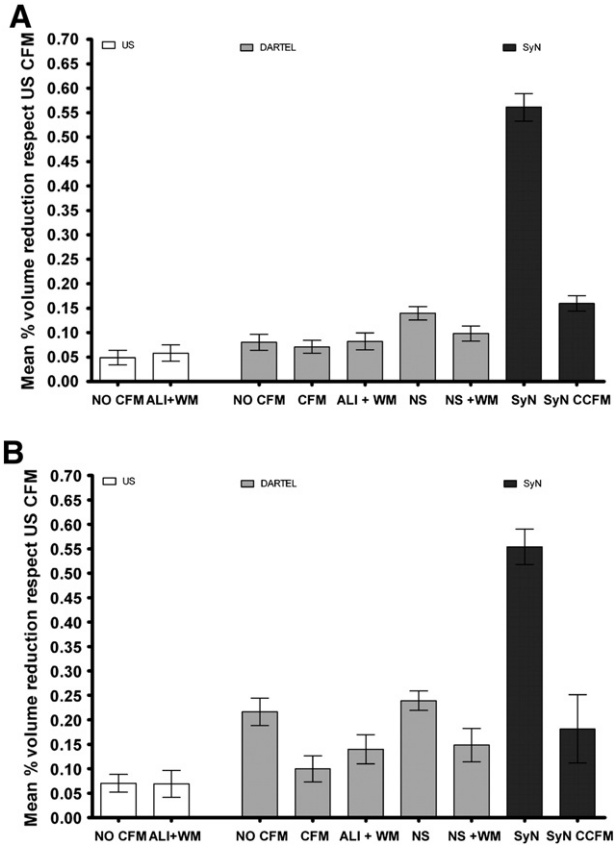


Fig. 7. Mean (and SEM) percentage of reduction in lesion volume compared to Unified Segmentation with CFM for resected (A) and stroke datasets (B).

those computed with cost function masking) do not provide such good RMSD.

It is important to comment that the volume measure reported in this manuscript (which is extracted from manual depiction of normalized lesions) can convey two possible mixture effects: the effect of the normalization to the lesion definition itself and the effect that the normalization could have on the ability of an expert to delineate the lesion. Following this logic, the automatic lesion specification, extracted from the NLDs and reported on Table S6 of the supplementary material, could only describe the effect of normalization on the lesion definition. However, given that there were no significant differences between the volumes of manually depicted lesions and automatic obtained NLDs, the effect of the normalization in the ability of an expert to delineate a lesion is minimal for all the algorithms analyzed.

Another important issue in the evaluation of the different automatic methods is to what extent lesioned brains might yield to inconsistencies in the normalizations. Visual inspection of segmented GM and WM images demonstrated how large sections of non-brain tissue were included in the GM segmentation in both resections and strokes for Unified Segmentation without CFM, Unified Segmentation with CFM and ALI toolbox with (WM + CSF)/2. Poor segmentation resulting in the inclusion of non-brain tissue has been previously reported (Fein et al., 2006); however, that study used an SPM2 normalization process (Ashburner and Friston, 1999). These segmentation errors have been associated with the misregistration of individual brain images with the reference template. To check this assumption, manual correction to the anterior commissure for the images which were erroneously segmented was performed. All the algorithms which failed to provide good segmentations for these images yielded optimum results once the lesioned images were corrected. It is important to notice that all the New Segment-based approaches always provided good segmentations for all the 500 artificially generated lesioned images without any type of manual correction.

[1] It is important to remember that DARTEL was always used with four times the default regularization. This was done as the results of the small parameter search (see Table 3) showed that there were not statistical differences in the RMSD score for DARTEL with Unified Segmentation plus CFM used to provide for the GM and WM segmentations needed, and that of Unified Segmentation with CFM which has proven to be the gold-standard within SPM lesioned brain normalization (Andersen et al., 2010).

When looking at the normalized patient images, an *out-of-brain distortion (OoBD)* effect was discovered (Fig. 3) which consists in a displacement of the lesioned zone out of the borders of where normalized brain tissue should be. This effect is most notorious in lesions close to the skull such as those in the epileptic resected group (Fig. 1A). Only Unified Segmentation with CFM, SyN based methods and both DARTEL with New Segment based approaches were free from this OoBD (see Fig. 3). Although it is quite clear that CFM and CCFM are the reason why this effect does not appear when using them with Unified Segmentation and SyN respectively, it is difficult to know why New Segment based algorithms do not suffer from this *OoBD*. The more accurate way of segmenting into GM and WM provided by New Segment might account for the absence of OoBD effect, as better normalization is expected if images are better segmented. Also the amount of regularization used by DARTEL might also help in avoiding OoBD.

All these results seem to suggest that that the three best normalization algorithms are Unified Segmentation with CFM, DARTEL with New Segment using (WM ± CSF)/2 as a prior and SyN with CCFM. All of them have some advantages and drawbacks. A general advantage is that none of them suffers from the OoBD described above and, although Unified Segmentation failed to provide good segmentations for 9 out of the 500 artificially generated lesioned images, the problem was solved when the images were manually corrected to the anterior commissure. First of all, it is clear than SyN was the best method normalizing healthy subjects. For the RMSD measure it yielded the second lowest score, being not significantly different than Unified Segmentation with CFM in the stroke dataset and also achieved the third lowest score for the resected dataset. Nevertheless, its mean lesion volume post-normalization was quite lower when compared to the one obtained using Unified Segmentation with CFM. Another drawback is that it is not fully automated, as masks must be depicted. On the other hand, DARTEL with New Segment using (WM + CSF)/2 showed the third highest NCC scores for the healthy dataset and the lowest RMSD scores for both lesioned datasets. However, its mean lesioned post-normalization volume was much lower than the one obtained by Unified Segmentation with CFM. The main advantage of this algorithm is that it is fully automatic. Finally, Unified Segmentation with CFM was worse than the other algorithms at normalizing healthy subjects, but it showed the second lowest RMSD score for both lesioned dataset and the highest mean post-normalization lesion volume. Its greatest disadvantage is that it is very time consuming, as lesion masks have to be drawn and besides, sometimes the images must be corrected to the anterior commissure.

Finally, it is important also to consider the limitations of the present investigation. First, the pool of 500 simulated lesions created has increased the amount of covariance in the analysis, mostly due to the fact that the same brain with different lesions is repeated twenty times. Although Greenhouse–Geisser correction and paired t-tests take the covariance of the data into account, this could affect the statistical results. Second, a bias may have been introduced in the study by performing a parameter search to tune the regularization parameter for DARTEL. This parameter search was justified by the lack of previous studies analyzing the optimal parameters to be used in lesion datasets, in contrast to all the other algorithms analyzed in the present paper. Unified Segmentation with CFM is considered in SPM the gold standard in order to normalize brains with extensive injuries. Besides, it has its own parameter search performed in previous studies (Andersen et al., 2010; Crinion et al., 2007). ALI toolbox is an evolution of Unified Segmentation and in Seghier et al. (2008) original manuscript, a small parameter search to tune the number of iterations which could improve segmentation was also carried out. SyN normalization performance with lesioned brains has also already been compared (at least qualitatively) to other normalization algorithms (Kim et al., 2007). Besides, it has also been used to normalize a dataset of stroke patients (Schwartz et al., 2009) and in ANTS manual (Avants et al., 2011b) some indications for the parameters to be used when normalizing injured brains can be found. However, there is still a risk of bias, as the parameter search is done on some of the same data that is later used to test the performance of the normalization algorithms. Third, a different type of regularization could have been used with DARTEL, as in this study we used only four times the linear elastic default regularization. The effects of using more regularization or membrane or bending energies (allowed by DARTEL) should be carefully investigated to determine if these modifications improve the results of the normalization process. In addition, SyN is a very flexible algorithm with multiple parameters that could be adjusted. A different similarity metric rather than cross-correlation, more iterations, higher gradient step, more or Bspline regularization and higher radius of the correlation window could be used to improve the normalization performance over lesioned brains. Although volumes for SyN were corrected to make them comparable to the other methods normalized to the ICBM452 template, they should be carefully chosen as the correction might not be perfect. Parameters such as number of Gaussians per class, Bias FWHM or warp frequency cut-off could also have been changed for Unified Segmentation and New Segment. The number of iterations in the ALI toolbox could also be increased to evaluate if this factor affects the normalization results. Fourth, the specific template created for SyN may have given this method an advantage over the other algorithms. Nevertheless, DARTEL also uses a subject specific template, as it alternates between computing this template and warping all subjects' TPMs into a better alignment. It is also important to note that ANTS environment provides a fully automated program to compute the subject specific template needed for SyN to work properly, just as DARTEL needs New Segment (or an equivalent method) to provide the GM and WM TPMs it needs as input. Finally, another highly detailed MRI template as Colin27 could have been used as template for the SyN based algorithms instead of creating a subject-specific one. However, these limitations notwithstanding, we are confident that our results give a general idea of the advantages and disadvantages of these ten different ways of normalizing brains with large lesions.

Much research is still needed to create or improve fully automated algorithms for the normalization of lesioned brains. For example, increasing regularization or tuning the parameters under DARTEL could eventually lead to DARTEL not reducing the lesioned volume while getting the lowest RMSD score and being fully automated. Another possible approach would be to provide an automatic way to create the masks of the lesions needed by Unified Segmentation with CFM and SyN with CCFM. One option is to use the ALI toolbox that, besides normalization, can also automatically identify and delineate lesions and has been recently tested against manual methods with very promising results (Wilke et al., 2011).

Conclusion

Our results show that, the large-deformation model and the improved segmentations provided by DARTEL and New Segment with (WM + CSF)/2 as a prior, provide a more accurate normalization than Unified Segmentation with CFM, but with shrinkage in lesion volume. The process of building a lesion binary mask can take from 30 min to 8 h for depicting the injury in a rough or a well-defined manner (Andersen et al., 2010); thus, an automated procedure may always be preferred. In this respect, DARTEL plus New Segment (WM + CSF)/2 not only is fully automated but also achieves the lowest significant root mean squared displacement score for every group of interest, irrespective of the lesion type. Nevertheless, before the abandonment of manual normalization procedures such as Unified Segmentation with CFM, further research is needed, first of all to solve the lesion volume reduction problem and also to take into account the effect of the normalization method on functional MRI data (Crinion et al., 2007).

Acknowledgments

ALI toolbox was kindly provided by Mohamed Seghier. We want to thank M. Seghier and E. Cámara for their suggestions and comments on a previous version of the present article. This work was funded by an Obra Social La Caixa grant to P. Ripollés and supported by Grants from the Spanish Government PSI2008-03885 to Ruth de Diego-Balaguer, PSI2009-09101 to Josep Marco-Pallarés and PSI2008-03901, La Marato de TV3 (Neuroscience Program) and the Catalan Governments (SGR 2009 SGR 93) to Antoni Rodriguez-Fornells. Josep Marco-Pallarés is supported by the Ramon y Cajal program of the Spanish Department of Science. We want to specially thank Diana Lopez-Barroso, David Cucurell, Nuria Rojo, Julià Amengual and Cesar Garrido for their help scanning the patients analyzed in the present study. Finally, we want also to thank the very constructive comments of the reviewers of the present manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2012.01.094.

References

Andersen, S.M., Rapcsak, S.Z., Beeson, P.M., 2010. Cost function masking during normalization of brains with focal lesions: still a necessity? Neuroimage 53, 78–84.
Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.
Ashburner, J., 2009. Computational anatomy with the SPM software. Magn. Reson. Imaging 27, 1163–1174.
Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.
Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.
Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41.
Avants, B.B., Yushkevich, P., Pluta, J., Gee, J.C., 2009. The optimal template effect in studies of hippocampus in diseased populations. Neuroimage 49, 486–500.
Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011a. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044.
Avants, B.B., Tustison, N.J., Song, G., 2011b. Advanced Normalization Tools (ANTS) manual. http://picsl.upenn.edu/ANTS/ants.pdf 2011.
Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F., 2003. Voxel-based lesion-symptom mapping. Nat. Neurosci. 6, 448–450.
Brett, M., Leff, A.P., Rorden, C., Ashburner, J., 2001. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 14, 486–500.
Cheung, M.C., Chan, A.S., Lam, J.M., Chan, Y.L., 2009. Pre- and postoperative fMRI and clinical memory performance in temporal lobe epilepsy. J. Neurol. Neurosurg. Psychiatry 80, 1099–1106.
Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., Friston, K., 2007. Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. Neuroimage 37, 866–875.
Dice, L.R., 1945. Measures of the amount of ecological association between species. Ecology 26, 297–302.
Fein, G., Landman, B., Tran, H., Barakos, J., Moon, K., Di, S.V., Shumway, R., 2006. Statistical parametric mapping of brain morphology: sensitivity is dramatically increased by using brain-extracted images as inputs. Neuroimage 30, 1187–1195.
Immonen, A., Jutila, L., Muraja-Murro, A., Mervaala, E., Aikia, M., Lamusuo, S., Kuikka, J., Vanninen, E., Alafuzoff, I., Ikonen, A., Vanninen, R., Vapalahti, M., Kalviainen, R., 2010. Long-term epilepsy surgery outcomes in patients with MRI-negative temporal lobe epilepsy. Epilepsia 51, 2260–2269.
Kim, J., Avants, B., Patel, S., Whyte, J., 2007. Spatial Normalization of Injured Brains for Neuroimaging Research: An Illustrative Introduction of Available Options. http://www.ncrrn.org/papers/methodology_papers/sp_norm_kim.pdf 2007.
Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46, 786–802.
Nachev, P., Coulthard, E., Jäger, H.R., Kennard, C., Husain, M., 2008. Enantiomorphic normalization of focally lesioned brains. Neuroimage 39, 1215–1226.
Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. Behav. Neurol. 12, 191–200.
Rorden, C., Karnath, H.O., Bonilha, L., 2007. Improving lesion-symptom mapping. J. Cogn. Neurosci. 19, 1081–1088.
Schwartz, M.F., Kimberg, D.Y., Walker, G.M., Faseyitan, O., Brecher, A., Dell, G.S., Coslett, H.B., 2009. Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. Brain 132, 3411–3427.
Seghier, M.L., Ramlackhansingh, A., Crinion, J., Leff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation–normalization models and fuzzy clustering. Neuroimage 41, 1253–1266.
Tahmasebi, A.M., Abolmaesumi, P., Zheng, Z.Z., Munhall, K.G., Johnsrude, I.S., 2009. Reducing inter-subject anatomical variation: effect of normalization method on sensitivity of functional magnetic resonance imaging data analysis in auditory cortex and the superior temporal region. Neuroimage 47, 1522–1531.
The FIL Methods Group, 2010. SPM8 Manual. http://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf 2010.
Wilke, M., de Haan, B., Juenger, H., Karnath, H.O., 2011. Manual, semi-automated, and automated delineation of chronic brain lesions: a comparison of methods. Neuroimage 56, 2038–2046.
Yassa, M.A., Stark, C.E., 2009. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. Neuroimage 44, 319–327.
Yogarajah, M., Focke, N.K., Bonelli, S.B., Thompson, P., Vollmar, C., McEvoy, A.W., Alexander, D.C., Symms, M.R., Koepp, M.J., Duncan, J.S., 2010. The structural plasticity of white matter networks following anterior temporal lobe resection. Brain 133, 2348–2364.