

## Linear normalization of MR brain images in pediatric patients with periventricular leukomalacia

Bart Machilsen,<sup>a</sup> Emiliano d'Agostino,<sup>b</sup> Frederik Maes,<sup>b</sup> Dirk Vandermeulen,<sup>b</sup> Horst K. Hahn,<sup>c</sup> Lieven Lagae,<sup>d</sup> and Peter Stiers<sup>a,e,\*</sup>

<sup>a</sup>Laboratorium voor Neuropsychologie, K.U.Leuven Medical School, Leuven, Belgium

<sup>b</sup>Medical Image Computing (Radiology-ESAT/PSI), Faculties of Medicine and Engineering, University Hospitals K.U.Leuven, Leuven, Belgium

<sup>c</sup>Center for Medical Diagnostic Systems and Visualization-MeVis, Bremen, Germany

<sup>d</sup>Department of Pediatrics, University Hospitals K.U.Leuven, Leuven, Belgium

<sup>e</sup>Neurocognition Department, Maastricht University, Maastricht, The Netherlands

Received 17 March 2006; revised 9 December 2006; accepted 19 December 2006

Available online 17 January 2007

**The feasibility of linear normalization of child brain images with structural abnormalities due to periventricular leukomalacia (PVL) was assessed in terms of success rate and accuracy of the normalization algorithm. Ten T1-weighted brain images from healthy adult subject and 51 from children (4–11 years of age) were linearly transformed to achieve spatial registration with the standard MNI brain template. Twelve of the child brain images were radiologically normal, 22 showed PVL and 17 showed PVL with additional enlargement of the lateral ventricles. The effects of simple modifications to the normalization process were evaluated: changing the initial orientation and zoom parameters, masking non-brain areas, smoothing the images and using a pediatric template instead of the MNI template. Normalization failure was reduced by changing the initial zoom parameters and by removing background noise. The overall performance of the normalization algorithm was only improved when background noise was removed from the images. The results show that linear normalization of PVL affected brain images is feasible.**

© 2007 Elsevier Inc. All rights reserved.

### Introduction

Spatial normalization of MR images to a common stereotaxic brain template is a pivotal step in many clinical and research applications. All automated intersubject comparisons of brain images require a spatial alignment of the images prior to the comparison. Normalization to a standard brain template is also required for many automated single-subject analysis tools (e.g. automatic tissue segmentation algorithms and labeling methods). The quality of the normalization step strongly affects the results of

subsequent analyses (e.g. voxel-based, deformation-based and tensor-based morphometry, automated volumetry, functional MRI) (Ashburner and Friston, 2000; Bookstein, 2001; Muzik et al., 2000; Shen and Davatzikos, 2003; Wilke et al., 2003).

Automated intensity-based normalization algorithms have been developed to achieve a spatial match between a brain image and a standard template (e.g. Collignon et al., 1995; Friston et al., 1995; Maes et al., 1997; Wells et al., 1996; Woods et al., 1992). These algorithms calculate spatial deformations by maximizing a measure of similarity between a floating and a reference image. In the deformation process a distinction is made between linear deformations (translation, rotation, zoom and shear), which are applied in the same manner throughout the brain image, and non-linear deformations, which invoke different transformations on different parts of the image. These two steps are often performed in succession (Hill et al., 2001), because the accuracy of non-linear normalization depends on the quality of the linear alignment.

Particularly in the context of morphologic NMR studies of brain disorders, questions have been raised regarding the effect of spatial normalization on the validity of the analysis. The reference image used to guide the normalization process presents the first problem. The most widespread digital atlas – the ICBM152 template created at the Montreal Neurological Institute (MNI) – is derived from 152 coregistered brain images from healthy adult subjects (Evans et al., 1994). Consequently, it might match better the images of neurologically healthy adults than images of patient or child brains (Bookstein, 2001; Karas et al., 2003; Muzik et al., 2000). Population-specific (Wilke et al., 2002) and even study-specific (Karas et al., 2003) atlases have been proposed to circumvent this problem.

Another issue concerns the presence of gross structural abnormalities in patient brain images (lesions, malformations, tumors, etc.). These affect the signal intensity distribution in the patient's image to such an extent that normalization algorithms may introduce undesirable transformations in order to maximize the similarity with a

\* Corresponding author. Neurocognition Department, Maastricht University, Faculty of Psychology, P. O. Box 616, 6200 MD Maastricht, The Netherlands. Fax: +31 43 388 41 25.

E-mail address: peter.stiers@psychology.unimaas.nl (P. Stiers).

Available online on ScienceDirect (www.sciencedirect.com).

template. Manually masking the lesion site improves the normalization in case of spatially confined abnormalities (Brett et al., 2001).

The aim of the present study was to evaluate the quality of linear spatial normalization of patient brain images with gross structural abnormalities. Periventricular leukomalacia (PVL) was chosen as an exemplary operationalization for gross structural abnormalities. PVL is one of the most common brain injuries in preterm-born children. It consists of ischemic damage to the white matter around the lateral ventricles which appears as regions of hyperintensity in T2-weighted and FLAIR MR images (Baker et al., 1988; Flodmark et al., 1989; Maalouf et al., 1999). Often, the damage results in white matter reduction, which in turn leads to enlarged and irregularly shaped ventricles.

In the present study, several manipulations to the images were applied prior to the linear normalization. We considered changes in the initial position of the floating image, masking of non-brain voxels, smoothing the floating images and the use of a pediatric instead of an adult brain template. The impact of these manipulations on normalization quality was assessed both in terms of the ability of the algorithm to compute a mathematically acceptable solution and the accuracy of the successful solutions. We further investigated whether the effects of these simple preprocessing steps differed between adult and child brain images and whether these effects were dependent on the amount of white matter signal alteration and ventricle enlargement in the images.

## Methods

### Subjects

Fifty-one children with a history of prematurity and ten healthy adults with no known neurological history participated in this study. Their characteristics are summarized in Table 1. The pediatric patients were recruited from the population of preterm children followed at the Pediatrics Department of the University Hospital Gasthuisberg, Leuven, Belgium. They were diagnosed by an experienced radiologist from multi-channel MR data sets. In 12 children, the MR-scan was radiologically normal. Thirty-nine were diagnosed with periventricular leukomalacia, and in 17 of these the lateral ventricles were labeled enlarged. These pediatric groups will be referred to as the Normal group, the PVL group and the PVL+ group, respectively. Patients with additional neurological complications (e.g. intracranial hemorrhage, focal infarctions, brain malformation, postnatal traumatic or infectious brain lesions) were excluded from the study. No

subject was sedated for the image acquisition. The study was approved by the local ethics committee and informed consent was obtained from all participants and/or their parents.

### MR image acquisition

High-resolution T1-weighted MR images were collected from all subjects. All images were acquired on the same 1.5-T Philips Gyrosan NT Intera scanner, using a 3D MPRAGE sequence (TR 22 ms, TE 4.6 ms, flip angle 30°). The resolution of the images was 0.86×0.86×1.00 mm, comprising 150 axial slices with 256×256 in-plane dimensions.

For all children additional proton-density, T2-weighted and FLAIR images were acquired on the same scanner. A dual echo STIR sequence (TR 4000 ms, TE1 17.2 ms, TE2 86 ms, TI 160 ms, flip angle 90°) yielded a proton density and a T2-weighted image. The images comprised 44 adjacent 3 mm thick axial slices of 256×256 voxels with an in-plane resolution of 1.17×1.17 mm. The 3D FLAIR sequence (TR 5000 ms, TE 100 ms, TI 2000 ms, flip angle 90°) yielded images with a resolution of 1.21×1.21×1.50 mm, comprising 100 axial slices with 256×256 in-plane dimensions. For each child, the multi-channel images were automatically aligned by applying rigid-body transformations (translation in 3 directions and rotation around each orthogonal axis) with the MIRIT software (Medical Image Computing, KULeuven, Belgium), implementing the algorithm described in Maes et al. (1997).

### Segmentation

The T1-weighted images were segmented into tissue-specific density maps using the EMS software (Expectation-Maximization Segmentation; Van Leemput et al., 1999). To improve segmentation quality for pediatric images with gross morphological abnormalities, the anatomic template was matched with the brain image using a viscous fluid deformation algorithm (D'Agostino et al., 2003). The resulting deformation fields were applied to the template density maps. The deformed template density images guided the subsequent segmentation step.

### Volumetry

To investigate the effect of lateral ventricle enlargement on linear normalization, the volume of the lateral ventricles was quantified on the T1-weighted images, using a three-dimensional interactive

Table 1  
Characteristics of groups studied

Groups <sup>a</sup>	n	Gender	Age (years)	Gestation (weeks)	Cerebral palsy	LVI (%) <sup>b</sup>	T2HI (%) <sup>c</sup>
		F/M	M (SD) range	M (SD) range	n (%)	M (SD) range	M (SD) range
Adult	10	4/6	24.7 (3.2) 19.0–30.0	Not available	0 (0.0)	1.400 (0.577) 0.357–2.108	Not available
Normal	12	7/5	7.4 (2.9) 4.3–13.4	33.4 (2.0) 31–36	0 (0.0)	0.753 (0.485) 0.301–2.042	0.268 (0.101) 0.037–0.377
PVL	22	11/11	7.2 (1.6) 4.6–12.1	32.5 (2.7) 26–36	15 (68.2)	1.199 (0.446) 0.394–2.297	0.479 (0.154) 0.207–0.742
PVL+	17	6/11	7.8 (2.2) 5.0–12.8	31.1 (3.5) 25–36	11 (64.7)	3.039 (1.295) 1.546–6.445	0.554 (0.250) 0.123–1.227

<sup>a</sup> Patient categorization based on radiological evaluation of available MR data.

<sup>b</sup> Lateral ventricle index (%) = 100 × ventricle volume / total brain volume.

<sup>c</sup> T2-hyperintensity index (%) = 100 × volume of white matter voxels with hyperintense MR signal / total brain volume.

watershed-transform (IWT) algorithm (Hahn and Peitgen, 2003). IWT volumetry has been shown reliable for both normal and pathological anatomy and for children as well as adults (Hahn et al., 2004). Importantly, this semi-automatic procedure does not require a prior normalization of the images. Hence, the effect of ventricular abnormality could be evaluated independent of linear normalization. The volumes of the lateral ventricles and of the whole brain were calculated from the original brain images. For each subject the lateral ventricle index (LVI) was defined as the ratio of the volume of the lateral ventricles over the total brain volume, multiplied by 100.

The amount of white matter voxels exhibiting MR signal hyperintensity was determined for each child with a fully automated model-based tissue classification algorithm (Van Leemput et al., 2001) from multi-channel MR images comprising the proton-density-weighted, T2-weighted and FLAIR images. The algorithm performs intensity-based tissue classification by estimating tissue-specific intensity models from the data and simultaneously detecting outlier voxels whose intensity values are not well explained by the model. The lesion class was set to white matter, with the constraint that FLAIR signal intensity should be larger than the average FLAIR gray matter intensity. A custom-made periventricular mask was applied to the resulting lesion images to quantify the number of outlier voxels. This number, multiplied by the voxel-size, yielded the volume of periventricular voxels showing white matter hyperintensity. The T2-hyperintensity index (T2HI) was calculated as the ratio of the volume of the hyperintense periventricular voxels over the total brain volume, multiplied by 100. Fig. 1 visualizes the relationship between the T2HI and the LVI for the three radiologically defined pediatric groups.

#### Linear normalization

The MIRIT software was used to spatially normalize the T1-weighted MR images with a standard brain template. MIRIT uses mutual information as a measure of statistical dependence between intensities of corresponding voxels on two images. This dependence is assumed to be maximal if the two images are

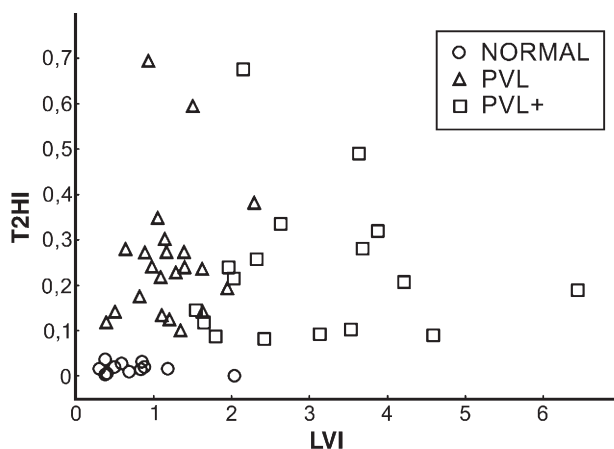


Fig. 1. Relationship between the lateral ventricle index (%) [LVI=100×ventricle volume/total brain volume] and the T2-hyperintensity index (%) [T2HI=100×volume of periventricular white matter voxels with hyperintense MR signal/total brain volume] for the three radiologically defined pediatric groups (Normal=radiologically normal children; PVL=children showing periventricular leukomalacia on multi-channel MR-images; PVL+=PVL with additional enlargement of the lateral ventricles).

geometrically aligned. As such, mutual information is applied as an objective criterion to find the linear transformation that spatially normalizes two images (Maes et al., 1997).

A linear transformation between two 3D data sets is fully described by 12 parameters: translation, rotation, zoom and shear, each along three orthogonal dimensions. To correct for different head positions in the scanner, the MIRIT algorithm was initiated with approximations for 3 of the 12 linear transformation parameters: rotation around the superior–inferior and the anterior–posterior axis, and translation around the left–right axis. These approximations were derived by manually aligning the midplane of each image with the template. This was identical for all methods, as were the MIRIT optimization parameters which were set to their default values (Maes et al., 1997). Contrary to Ashburner et al. (1997), no prior constraints about the distribution of zoom and skewness parameters were incorporated in the algorithm.

#### Modifications to the normalization procedure

Four different modifications to the normalization procedure were employed. Their effect on the success rate and the accuracy of the linear normalization was assessed.

##### Modification to initial position and volume

The information about the starting position of a floating image was manipulated by providing the algorithm with manual approximations for some of the twelve normalization parameters that the algorithm has to compute in order to maximize mutual information.

In the *baseline* method no prior information beyond the initial correction for head position described above was provided. In the *slant correction* method, the rotation angle around the left–right axis, necessary to align the anterior commissure (AC) and posterior commissure (PC) in the axial plane in the floating and reference image, was included as an additional ‘initial guess’ for the algorithm. In the *standard space* method, in addition to the parameters in the slant correction method, the AC was co-localized with its position in the reference image. This amounts to providing additional estimates of translation along the superior–inferior and the anterior–posterior dimensions. Finally, in the *volume correction* method, the starting estimates for the zoom parameters were modified depending on the ratio of the reference over the floating image volumes. Total brain volume for the template was calculated from the corresponding gray and white matter priors. Total brain volume for the original brain images was calculated from the segmented gray and white tissue maps obtained with the EMS software, by summing the probabilities in each map over all voxels. The IWT volume estimates were not used for this purpose because the IWT algorithm is not fully automated and because it could not be applied to the smooth template. The agreement between the two methods was very high, however. The IWT estimates of total brain volume were on average 0.47% smaller ( $SD=0.023$ ) than the estimates based on the segmented tissue density images. The correlation between the two estimates was 0.98. For each brain image, an isotropic volume correction factor was calculated as the cube root of the ratio of the floating image brain volume over the reference brain volume. This factor was used as the starting estimate for each of the three zoom parameters.

##### Masking of background image areas

MR images differ considerably in the amount of background noise. The *background exclusion* method assessed the contribution

of the background area on linear normalization by excluding background voxels from the analysis. For this purpose, individual binary masks were generated by thresholding each image at 10% of its intensity range. This mask image was then repeatedly smoothed with a 7-mm full-width at half-maximum (FWHM) Gaussian kernel and dichotomized at a 50% threshold. Ten iterations were sufficient to fill the ‘holes’ in the mask and to remove all background image

areas. The resulting mask was applied to the original floating image, resulting in the removal of background voxels while leaving most of the skull intact. The same procedure was applied to create a background-deleted MNI template.

In the *skull stripping* method, the BET utility (Smith, 2002) was used to remove background and skull from all original images. The skull-stripped ICBM152 template (available from

Table 2  
Description of landmarks and position of landmarks in MNI and CCHMC template

ID	Name	Plane	Description	MNI template			CCHMC template		
				x	y	z	x	y	z
1	Extreme left	Sagittal	Find slice with last cerebral voxels. Determine midpoint of cluster.	<b>-71.4</b>	-30.0	5.2	<b>-73.7</b>	-31.9	8.8
2	Extreme right	Sagittal	Find slice with last cerebral voxels. Determine midpoint of cluster.	<b>70.6</b>	-31.9	7.3	<b>73.8</b>	-33.3	9.4
3	Extreme anterior	Coronal	Find slice with last cerebral voxels. Determine midpoint of cluster.	0.6	<b>71.9</b>	1.1	1.0	<b>73.3</b>	5.7
4	Extreme posterior	Coronal	Find slice with last cerebral voxels. Determine midpoint of cluster.	0.6	<b>-106.2</b>	2.2	1.0	<b>-109.9</b>	0.2
5	Extreme superior	Axial	Find slice with last cerebral voxels. Determine midpoint of cluster.	0.6	-27.1	<b>79.1</b>	1.3	-31.1	<b>87.9</b>
6	Inferior temporal L	Axial	Find the most inferior slice still showing left temporal lobe voxels. Determine midpoint of cluster.	-23.4	-8.3	<b>-51.7</b>	-23.4	-7.6	<b>-47.9</b>
7	Inferior temporal R	Axial	Find the most inferior slice still showing right temporal lobe voxels. Determine midpoint of cluster.	24.6	-6.3	<b>-51.8</b>	24.1	-5.4	<b>-48.4</b>
8	Inferior frontal L	Coronal–axial intersection	In the sagittal plane just left of midline, seek coronal slice through most anterior curving of cingulate s. Seek sagittal slice through center of left eyeball. Determine lowest point of frontal lobe.	-32.4	46.0	<b>-19.9</b>	-33.3	46.6	<b>-15.8</b>
9	Inferior frontal R	Coronal–axial intersection	In the sagittal plane just right of midline, seek coronal slice through most anterior curving of cingulate s. Seek sagittal slice through center of right eyeball. Determine lowest point of frontal lobe.	34.6	46.0	<b>-19.9</b>	35.7	46.8	<b>-16.6</b>
10	Temporocerebellar junction L	Sagittal through 8	Most anterior point in the slice where cerebellum and temporal lobe part.	-32.4	<b>-33.5</b>	<b>-32.7</b>	-33.5	<b>-33.0</b>	<b>-30.0</b>
11	Temporocerebellar junction R	Sagittal through 9	Most anterior point in the slice where cerebellum and temporal lobe part.	34.6	<b>-34.5</b>	<b>-32.6</b>	35.8	<b>-33.8</b>	<b>-30.8</b>
12	Inferior occipital	Axial	Find most inferior slice where occipital lobe voxels are visible posterior of cerebellum. Determine point in midline and through middle of cluster.	0.6	-91.2	<b>-22.4</b>	0.7	-91.9	<b>-25.2</b>
13	Putamen L	Coronal–axial intersection	In the sagittal plane, find center of thalamus. In axial slice through center, find coronal slice through fornix. At the intersection, determine most lateral point of left putamen in basal ganglia.	<b>-32.6</b>	0.0	6.0	<b>-33.4</b>	-0.6	11.1
14	Putamen R	Coronal–axial intersection	In the sagittal plane, find center of thalamus. In axial slice through center, find coronal slice through fornix. At the intersection, determine most lateral point of right putamen in basal ganglia.	<b>32.6</b>	0.0	6.0	<b>34.0</b>	-0.3	10.5
15	Splenium of corpus callosum	Coronal	Find slice in which corpus callosum no longer connects the cerebral hemispheres. For other coordinates, go one slice anteriorly and determine midpoint of cluster.	0.6	<b>-43.6</b>	14.8	1.0	<b>-46.1</b>	18.1
16	Tectum of corpus callosum	Axial	Find slice in which corpus callosum no longer connects the cerebral hemispheres. For other coordinates, go one slice down and determine midpoint of cluster.	0.6	-7.1	<b>28.3</b>	1.1	-8.4	<b>34.2</b>
17	Superior pons	Coronal	Find slice showing the dip between the two hemispheres of the pons at its superior junction with the brainstem, just before it starts expanding.	0.6	<b>-21.0</b>	<b>-20.2</b>	0.8	<b>-20.6</b>	<b>-17.1</b>
18	Hippocampus L	Coronal through 17	Determine intersection of lines touching lateral and superior extent of left hippocampal gyrus.	<b>-38.4</b>	-18.6	<b>-10.3</b>	<b>-39.5</b>	-19.2	<b>-6.8</b>
19	Hippocampus R	Coronal through 17	Determine intersection of lines touching lateral and superior extent of right hippocampal gyrus.	<b>38.6</b>	-18.6	<b>-10.3</b>	<b>40.1</b>	-18.8	<b>-7.4</b>
20	Anterior commissure	Axial	A few mm below the genu of the cc, find the slice showing the small connection between the cerebral hemisphere, immediately posterior of the head of the nucleus caudatus.	0.6	<b>3.6</b>	<b>-6.1</b>	0.9	<b>3.8</b>	<b>-2.0</b>
21	Posterior commissure	Axial	A few mm below the splenium of the cc, and a few mm above the colliculus superior, find the slice showing the small connection between the thalami.	0.6	<b>-24.4</b>	<b>-5.0</b>	0.9	<b>-25.0</b>	<b>-1.6</b>
22	Nucleus caudatus L	Coronal through 20	Determine intersection of lines touching lateral and superior extent of left caudate nucleus head.	<b>-19.4</b>	4.7	<b>21.8</b>	<b>-19.8</b>	3.9	<b>28.0</b>
23	Nucleus caudatus R	Coronal through 20	Determine intersection of lines touching lateral and superior extent of right caudate nucleus head.	<b>21.6</b>	4.8	<b>24.8</b>	<b>22.9</b>	4.1	<b>30.8</b>

The reported coordinates are in MNI space.



[http://www.mrc-cbu.cam.ac.uk/Imaging/Common/no\\_skull\\_norm.shtml](http://www.mrc-cbu.cam.ac.uk/Imaging/Common/no_skull_norm.shtml)) was used as the reference image.

### Smoothing

In this method, the images were smoothed with an isotropic Gaussian kernel because it has been shown that this decreases the likelihood that the normalization algorithm gets caught in a local minimum or maximum (Ashburner and Friston, 1997). Within this smoothing method, three different kernel widths were explored: 4 mm, 8 mm and 12 mm FWHM.

### Pediatric template

As a final modification to the linear normalization procedure, we used a pediatric brain template instead of the MNI template. A population-specific template resembles more the population under study and may therefore reduce the intersubject variability after normalization. The updated CCHMC template (Cincinnati Children's Hospital Medical Center; Wilke et al., 2002), based on MRI data from 200 children aged 5–19, was chosen as the reference image. We repeated the baseline and background exclusion methods with the CCHMC template as the reference image. The background excluded CCHMC template was created in the same way as the background excluded MNI template.

### Accuracy evaluation

#### Predefined Landmarks

To evaluate the quality of linear normalization, 23 anatomic landmarks were identified in all original images (Table 2). The choice of landmarks was based on the method proposed by Grachev et al. (1999), with some modifications based on landmarks used in other studies (Arndt et al., 1996; Salmond et al., 2002). We ensured that the landmarks were widely distributed across the brain. All landmarks were quickly and easily identified by defining anatomic reference planes: First, the image was manually reoriented to align the midsagittal plane with an orthogonal reference frame. Next, the AC and PC were placed in the same axial plane, and the image origin was set at the AC in the midline. Three-dimensional landmark coordinates were then defined relative to this reference position of the image. Inter-observer differences in the location of landmarks in this way have been shown to be relatively small (Arndt et al., 1996). The mean absolute deviation between landmark coordinates located by two of the authors (P.S. and B.M.) on an arbitrary subset of 22 images was less than 1 mm (0.946 mm) in the present study (60% were localized within 1 mm; over 90% within 2 mm).

Homologous landmarks were defined for the MNI and CCHMC templates. To avoid localization bias between the two templates, the 23 landmarks were only manually localized on the MNI template. The position of these landmarks was transposed to the CCHMC template by calculating the linear and non-linear parameters to normalize the MNI template to the CCHMC template using SPM5 (Wellcome Department of Imaging Neuroscience, London, <http://www.fil.ion.ucl.ac.uk/spm>). The resulting parameters were applied to the MNI landmarks to obtain the coordinates for the CCHMC landmarks. Table 2 presents the coordinates for both templates.

#### Accuracy index

Although three-dimensional coordinates were obtained for all landmarks, the accuracy calculation was based on only one or two

coordinates for each landmark. This was done because not all coordinates are equally relevant for describing normalization accuracy. For instance, the inferior–superior dimension is irrelevant for assessing the left outermost boundary of the brain. The relevant coordinates are indicated in bold in Table 2.

Normalization accuracy was defined as the difference between the set of landmarks on a normalized floating image  $F$  and the corresponding landmarks on a reference image  $R$ . The absolute accuracy index AAI,

$$AAI = \frac{1}{n} \sum_{i=1}^n w_{ix}(f_{ix} - r_{ix})^2 + w_{iy}(f_{iy} - r_{iy})^2 + w_{iz}(f_{iz} - r_{iz})^2,$$

was calculated as the squared Euclidean distance, averaged over all  $n$  landmarks and weighted for the relevant dimensions. Relevant coordinates had a weight  $w$  of one, irrelevant coordinates had a weight of zero.

#### Optimal linear transformation

Given the landmark coordinates and their associated weights, the most optimal linear transformation  $M$  that maps the landmarks from a floating image to a particular reference image could be derived by minimizing the sum of squared differences between the two landmark sets.

Let  $F$  be the  $(n \times 4)$  matrix of  $n$  floating landmarks, with the first three columns representing the  $x$ ,  $y$  and  $z$  coordinates, and the fourth being a column vector consisting of only ones. Let  $R$  and  $W$  be similar  $(n \times 4)$  matrices for the reference landmarks and the associated weights. The elements of the optimal linear transformation can then be calculated as

$$\begin{bmatrix} m_{k1} \\ m_{k2} \\ m_{k3} \\ m_{k4} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n w_{ik} f_{i1}^2 & \sum_{i=1}^n w_{ik} f_{i1} f_{i2} & \sum_{i=1}^n w_{ik} f_{i1} f_{i3} & \sum_{i=1}^n w_{ik} f_{i1} \\ \sum_{i=1}^n w_{ik} f_{i2} f_{i1} & \sum_{i=1}^n w_{ik} f_{i2}^2 & \sum_{i=1}^n w_{ik} f_{i2} f_{i3} & \sum_{i=1}^n w_{ik} f_{i2} \\ \sum_{i=1}^n w_{ik} f_{i3} f_{i1} & \sum_{i=1}^n w_{ik} f_{i3} f_{i2} & \sum_{i=1}^n w_{ik} f_{i3}^2 & \sum_{i=1}^n w_{ik} f_{i3} \\ \sum_{i=1}^n w_{ik} f_{i1} & \sum_{i=1}^n w_{ik} f_{i2} & \sum_{i=1}^n w_{ik} f_{i3} & \sum_{i=1}^n w_{ik} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n w_{ik} f_{i1} r_{ik} \\ \sum_{i=1}^n w_{ik} f_{i2} r_{ik} \\ \sum_{i=1}^n w_{ik} f_{i3} r_{ik} \\ \sum_{i=1}^n w_{ik} r_{ik} \end{bmatrix},$$

where  $k$  stands for the row index of the  $(4 \times 4)$  transformation matrix  $M$ .

The resultant transformation  $M$  provides the optimal linear registration between the configurations of selected landmark coordinates on the floating and on the reference image. No other linear transformation brings the landmark coordinates on the two images in closer correspondence.

#### Relative linear transformation

The relative accuracy index (RAI) for a particular floating and reference image was defined as the difference between the optimal AAI for these images and the AAI obtained by the linear normalization algorithm by maximizing the mutual information between the same floating and reference images. As such, the RAI indicates how far the solution offered by the intensity-based algorithm deviates from the optimal linear normalization solution.

#### Statistics

Group-wise comparisons between the three pediatric groups on descriptive characteristics involved chi-square and  $t$ -tests. For  $t$ -

tests, randomization methods were used to avoid underlying parametric assumptions (Manly, 1991). These involve 100,000 randomizations of the data. The associated  $p$ -values will be indicated as randomization probabilities ( $rp$ ).

The effect of prior modifications on linear normalization was evaluated with analysis of variance (ANOVA) on the relative accuracy indices. For each modification approach, a separate bifactorial split-plot ANOVA was computed with the methods as a within-subjects factor and the radiologically defined groups as the between-subjects factor. Accuracy indices were Box–Cox transformed ( $\lambda=-0.5$ ) prior to the analyses to circumvent violations of the normality of residuals (Shapiro–Wilk test) and the homogeneity of variances (Levene's test) assumptions. For each ANOVA, outliers within any method were omitted from all methods involved in that particular analysis. Planned comparisons contrasting the adult and the pooled pediatric groups on the one hand and the different pairs of pediatric groups on the other hand were then used to elaborate on the effects observed in the ANOVAs. Pairwise contrasts of methods within individual subgroups were assessed as Bonferroni-corrected post hoc comparisons.

Finally, multiple regression analyses were applied at the image voxel level to investigate the relationship between the objective measures of pathology and tissue distribution mismatch after linear normalization with the best method as indicated by the ANOVAs. SPM5 was used to assess the independent contribution of the LVI and T2HI measures to the probability distribution of the two main tissue classes. The analysis was performed separately for the white matter and gray matter tissue maps smoothed with an isotropic 8-mm FWHM Gaussian kernel. Statistical parametric maps were computed for both positive and negative partial correlations of gray matter and white matter probabilities with the LVI and the T2HI measures. The  $t$ -maps were evaluated at the significance level 0.05, corrected for multiple comparisons (Friston et al., 1991; Worsley et al., 1992). Only clusters of 10 or more voxels were considered.

## Results

### Characteristics of the pediatric groups

Characteristics of the three radiologically defined pediatric groups are compared in Table 1. The groups did not differ in age (largest difference:  $t(37)=1.08$ ;  $rp>0.05$ ) or gender ( $\chi^2(2)=1.633$ ,  $p>0.05$ ). There was, however, a small but significant difference in gestation age between the Normal and the PVL+ groups ( $t(27)=2.03$ ;  $rp<0.05$ ). No difference in gestation age was found between either of these two groups and the PVL group.

The average percentage of voxels showing hyperintense T2-weighted MR signals (T2HI) was 0.268 ( $SD=0.101$ ) for the radiologically normal children. This was significantly higher in the PVL group ( $M=0.479$ ,  $SD=0.154$ ,  $t(32)=4.25$ ;  $rp<0.001$ ) and in the PVL+ group ( $M=0.554$ ,  $SD=0.250$ ,  $t(27)=3.74$ ;  $rp<0.001$ ), although the difference between the PVL and PVL+ groups was not significant ( $t(37)=1.16$ ;  $rp>0.05$ ). As expected, the lateral ventricle volume index (LVI) was significantly higher in the PVL+ group ( $M=3.04$ ,  $SD=1.30$ ) compared to the PVL group ( $M=1.20$ ,  $SD=0.45$ ,  $t(37)=6.23$ ;  $rp<0.001$ ). However, the LVI was also significantly larger in the PVL group than in the normal children ( $M=0.75$ ,  $SD=0.49$ ,  $t(32)=2.70$ ;  $rp=0.01$ ), even though the radiologist had categorized these children as having PVL without ventricle enlargement. The T2-hyperintensity index and the lateral ventricle index in the 51 pediatric brain

images were only moderately correlated ( $r(49)=0.27$ ,  $p=0.06$ ; see Fig. 1).

### Optimal linear normalization

ANOVA on the optimal accuracy indices (based on the best possible linear transformation) with the adult or pediatric brain template as the within-subjects factor and the radiologically defined groups as the between-subjects factor, detected a main effect of groups ( $F(3,57)=4.43$ ,  $p<0.01$ ) and of templates ( $F(1,57)=19.34$ ,  $p<0.001$ ). The interaction subject groups  $\times$  templates was significant as well ( $F(3,57)=5.73$ ,  $p<0.01$ ). Planned comparisons showed that the main effect of subject groups was not due to the difference between adult and pediatric subjects ( $F(1,57)=2.88$ ,  $p>0.05$ ). As Fig. 2 illustrates, there was however a difference between the three pediatric groups: Landmarks on the PVL images deviated more from the templates than the two other pediatric groups (PVL-Normal:  $F(1,57)=6.22$ ,  $p<0.05$ ; PVL-PVL+:  $F(1,57)=6.90$ ,  $p<0.05$ ; Normal-PVL+:  $F(1,57)=0.02$ ,  $p>0.05$ ). The observed interaction could be explained by the fact that the landmarks from the adult group fitted better the MNI landmarks than the CCHMC landmarks (post hoc  $F(1,57)=23.17$ ,  $p<0.05/4$ ), whereas no significant difference between templates was observed for any of the three pediatric groups (smallest  $p>0.05/4$ ). The optimal normalization solution for our pediatric images is thus not systematically influenced by the choice of template.

The above results indicate that the optimal linear solution differs between subject groups. Therefore, the AAI is an inappropriate measure to evaluate normalization accuracy. For example, images from the PVL group are more likely to yield higher absolute accuracy indices than images from the PVL+ group, since the optimal solution for PVL+ children is superior to the optimal solution for PVL children. The AAI cannot discriminate between the performance of the linear normalization algorithm and the best possible linear fit between two sets of landmarks based on their spatial constellations. Therefore, the evaluations will be based on relative accuracy indices.

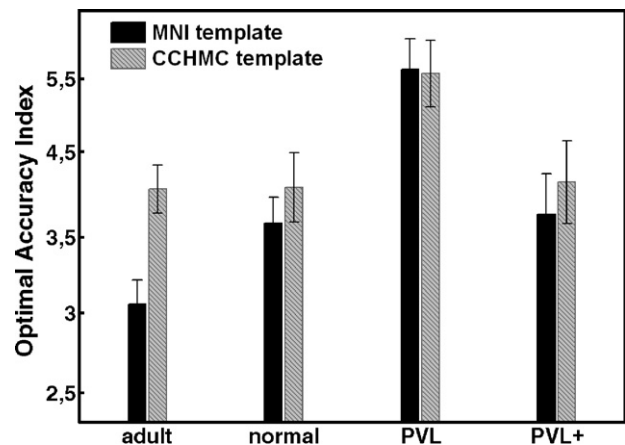


Fig. 2. Average optimal accuracy indices for each subject group for both the MNI and CCHMC templates. These indices are mean squared distances between landmark coordinates on a template and on the optimally normalized floating image. Optimal normalization is achieved by minimizing the sum of squared distances between the landmark coordinates on floating and reference image. Whiskers denote 95% confidence intervals.

### Baseline results

Fig. 3A summarizes the relative accuracy indices obtained with the baseline method. It illustrates the superior result obtained with adult compared to pediatric brain images (median test:  $\chi^2=11.57$ ,  $p<0.001$ ). It also shows that the baseline method produces some unsuccessful solutions. Although no gold standard is available to assess normalization accuracy (Woods et al., 1998), we considered a linear normalization unsuccessful if the RAI exceeded Tukey's outlier criterion, defined as  $1.5(Q_3-Q_1)+Q_3$  (Tukey, 1977). Calculated on the relative accuracy indices obtained with the baseline method on all subjects, this cut-off value equals 33 (horizontal line in Fig. 3A). Outliers only occurred in the pathological pediatric groups. Similar results were obtained for all normalization methods in this study (Table 3).

### Effects of modifications

To evaluate the performance of the linear normalization methods, all methods within a particular modification approach were first compared by ANOVA. Subsequently, the best methods from each approach were compared in a separate analysis.

For the initial position approach (Fig. 4A), only the effect of subject groups was significant ( $F(3,51)=25.04$ ,  $p<0.001$ ). Planned comparisons reveal that this main effect was due to the difference between adults and all three pediatric groups ( $F(1,51)=72.55$ ,  $p<0.001$ ), not between any of the pediatric groups (highest

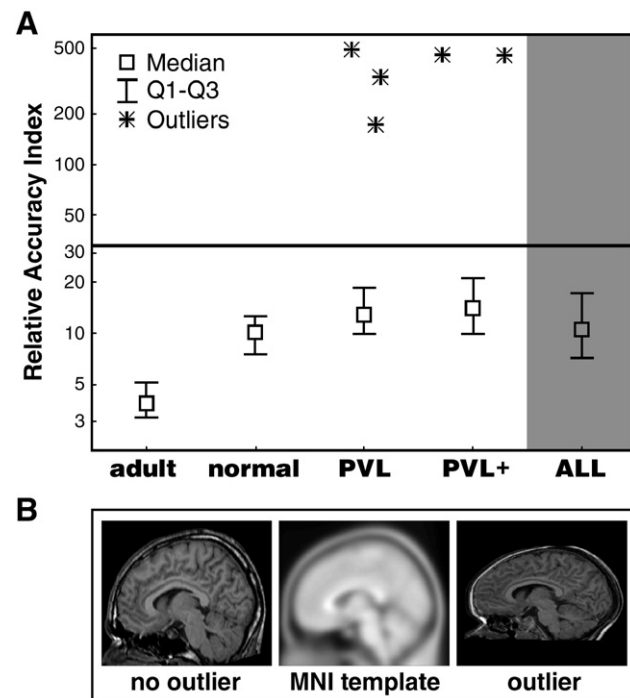


Fig. 3. (A) Whisker plots for each subject group summarize the relative accuracy indices obtained for the baseline method. The squares represent the median values for each subject group, the whiskers the first and third quartiles. The plot in the shaded area is based on all 61 subjects. The horizontal line represents the cut-off for successful normalization. The scale of the ordinate is logarithmic to enable the graphical presentation of all outliers. (B) Sagittal slices from a successfully normalized image (left), a normalization failure (right) and the MNI template (middle).

Table 3

Number of unsuccessful linear normalizations (relative accuracy index exceeds 33)

Methods	Groups				Total
	Adults	Normal	PVL	PVL+	
<i>Initial position</i>					
Baseline	0	0	3	2	5
Slant correction	0	0	3	1	4
Standard space	0	0	3	1	4
Volume correction	0	0	0	0	0
<i>Background</i>					
Background exclusion	0	0	0	0	0
Skull stripping	0	1	0	1	2
<i>Smoothing</i>					
Smoothing 4 mm	0	0	2	2	4
Smoothing 8 mm	0	2	1	2	5
Smoothing 12 mm	0	1	3	4	8
<i>Pediatric template</i>					
CCHMC baseline	0	0	3	1	4
CCHMC background exclusion	0	1	0	2	3

$F$ -value between Normal and PVL group,  $F(1,51)=0.89$ ,  $p>0.05$ ).

For the background area approach (Fig. 4B), only the interaction between subject groups and methods reached significance ( $F(3,55)=7.36$ ,  $p<0.001$ ). Post hoc tests revealed that the background exclusion method outperformed the skull stripping method for adult subjects ( $F(1,55)=9.85$ ,  $p<0.05/4$ ). The opposite was true for the children from the PVL group ( $F(1,55)=10.57$ ,  $p<0.05/4$ ). No difference was found for the normal and PVL+ groups (smallest  $p>0.05/4$ ).

The smoothing approach (Fig. 4C) revealed a main effect of subject groups ( $F(3,45)=20.549$ ,  $p<0.001$ ) and of methods ( $F(2,90)=25.04$ ,  $p<0.001$ ). The group effect was caused by the superior results obtained for adult compared to pediatric brain images (planned comparisons  $F(1,45)=57.04$ ,  $p<0.001$ ). Planned comparisons revealed no difference between any of the three pediatric groups (smallest  $p>0.05$ ). The main effect of methods was due to the inferior normalization quality obtained when smoothing the images with a 12-mm kernel compared to a 4-mm (post hoc  $F(1,45)=26.50$ ,  $p<0.05/3$ ) or an 8-mm kernel ( $F(1,45)=36.97$ ,  $p<0.05/3$ ). No significant difference was found between the 4- and 8-mm kernels ( $F(1,45)=0.80$ ,  $p>0.05/3$ ). No smoothing (baseline), smoothing with 4 mm, 8 mm and 12 mm kernels resulted in average relative accuracy indices of 9.82, 11.27, 11.29 and 13.99, respectively.

Fig. 4D directly compares the best method from each approach, as indicated by the above analyses. In the absence of a main effect of methods or whenever post hoc tests could not decide between two methods, the most robust method (i.e. the method that produced the smallest amount of outliers over all subject groups) was chosen. As such, the volume correction, background exclusion and 4-mm smoothing methods were compared. ANOVA revealed a significant main effect of subject groups ( $F(3,53)=18.61$ ,  $p<0.001$ ) and of methods ( $F(2,106)=104.71$ ,  $p<0.001$ ). Planned comparisons show again that adult brain images were better aligned with the MNI template than pediatric brain images ( $F(1,53)=53.19$ ,  $p<0.001$ ). No significant difference was found between the three

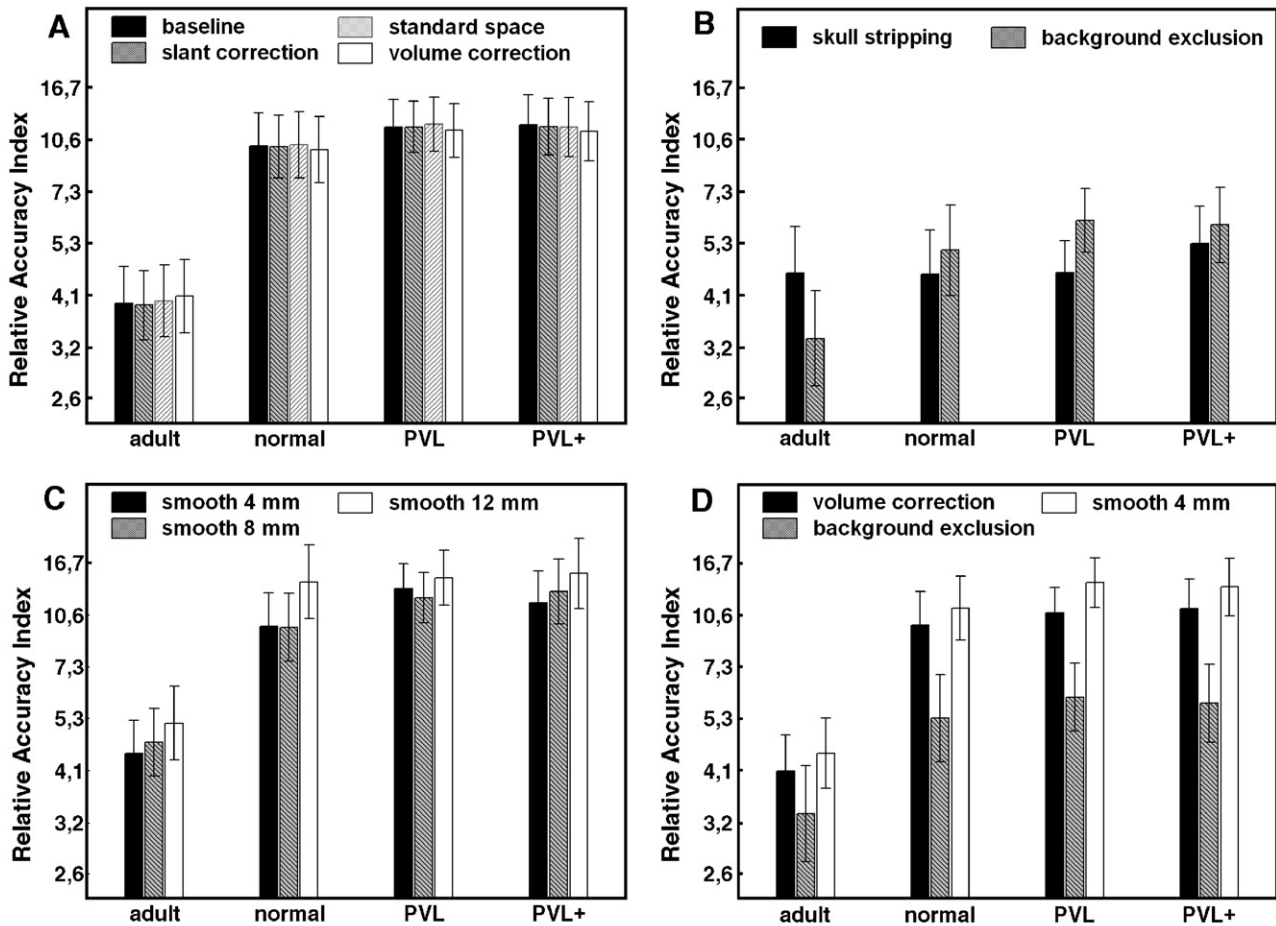


Fig. 4. Mean relative accuracy indices (RAI) for each subject group. (A) The four methods from the initial position approach are being compared: baseline, slant correction, standard space and volume correction. (B) Masking of non-brain areas: skull stripping and background exclusion. (C) Smoothing with three different kernel widths. (D) Comparison between the best methods from each approach group. Error bars represent  $\pm 1$  SD. Images causing a normalization failure ( $RAI > 33$ ) for any of the methods compared within each plot were omitted from that particular analysis.

pediatric groups (highest  $F$ -value between Normal and PVL group:  $F(1,53) = 0.94$ ,  $p > 0.05$ ). Post hoc tests revealed that background exclusion outperformed volume correction ( $F(1,53) = 91.73$ ,  $p < 0.05/3$ ). Volume correction in turn was more accurate than smoothing with a 4-mm kernel ( $F(1,53) = 35.59$ ,  $p < 0.05/3$ ). No additional gain in accuracy is to be expected from a combination of these three methods, since only the background exclusion method improved the landmark co-localization compared to the baseline method.

It should be noted that the lack of a significant difference between our three pediatric groups might be a mere corollary of the radiological classification. This classification did not fully correspond with the objective pathology measures (Fig. 1). To investigate the possibility of a relationship between the performance of the normalization algorithm and our objectivation of ventricle enlargement (LVI) and white matter abnormality (T2HI), we did a multiple regression analysis on all pediatric subjects with the background exclusion RAI as the dependent variable and the two pathology indices as predictor variables. Partial correlations between the background exclusion RAI and the two predictor variables were not significant (LVI:  $t(47) = 0.26$ ,  $p > 0.05$ ; T2HI:  $t(47) = 0.46$ ,  $p > 0.05$ ). This implies that the performance of the background exclusion method is not only

comparable for the three radiologically defined pediatric groups, but is also not related to the objective amount of ventricle enlargement or white matter signal abnormality.

#### Normalizing to a pediatric template

Fig. 4 shows that even with the background exclusion method a significant difference remained between the adult and pediatric subjects ( $F(1,53) = 16.61$ ,  $p < 0.001$ ). Since our results are based on relative accuracy indices, this effect cannot be attributed to the fact that the landmarks on the adult images are in closer correspondence with the MNI template than the landmarks on the pediatric images. The difference can only be due to an inferior performance of the normalization algorithm for pediatric brain images.

To investigate whether the use of a pediatric template would further improve linear normalization of pediatric brain images, we repeated the baseline and background exclusion methods with the CCHMC template as the reference image (Fig. 5). ANOVA on the relative accuracy indices revealed a significant main effect of methods ( $F(3,141) = 26.61$ ,  $p < 0.001$ ) and of subject groups ( $F(3,47) = 5.30$ ,  $p < 0.01$ ). The interaction subject groups  $\times$  methods was significant as well ( $F(9,141) = 5.06$ ,  $p < 0.001$ ).



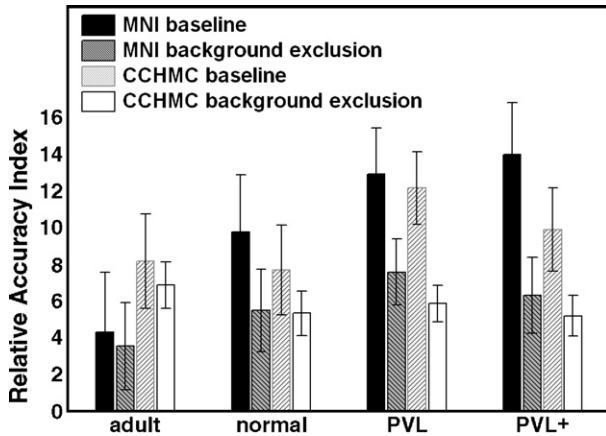


Fig. 5. Mean relative accuracy indices for each subject group for the baseline and background exclusion methods, and for normalization to the MNI and to the CCHMC templates. Error bars represent  $\pm 1$  SD.

Removing the background noise resulted in better accuracy levels for both templates, as indicated by planned comparisons (MNI:  $F(1,47)=38.64$ ,  $p<0.001$ ; CCHMC:  $F(1,47)=50.02$ ,  $p<0.001$ ). Planned comparisons did not reveal a significant effect of templates for the baseline method ( $F(1,47)=1.61$ ,  $p=0.21$ ), nor for the background exclusion method ( $F(1,47)=0.02$ ,  $p=0.43$ ). For both methods, however, a significant interaction between templates and subject groups was observed (baseline:  $F(1,47)=17.56$ ,  $p<0.001$ ; background exclusion:  $F(1,47)=10.69$ ,  $p<0.01$ ). For the baseline method, the CCHMC template outperformed the MNI template for the pooled pediatric groups ( $F(1,47)=12.10$ ,  $p<0.01$ ), whereas the opposite was true for the adult images ( $F(1,47)=8.66$ ,  $p<0.01$ ). For the background exclusion method, the MNI template was again better than the CCHMC template for adult images ( $F(1,47)=7.90$ ,  $p<0.01$ ). However, no significant difference was now present between using the CCHMC or the

MNI template when normalizing pediatric images ( $F(1,47)=2.85$ ,  $p=0.10$ ).

#### Relationship between pathology measures and normalization accuracy

The predefined landmarks yield only a single measure for the quality of linear normalization: the relative accuracy index. This index provides no information on the spatial distribution of remaining mismatches after linear normalization. In this final section, the relationship between the brain pathology and spatial distribution of tissue density maps will be explored with multiple regression analyses at the image voxel level. The pediatric gray and white probability maps, normalized to the MNI template using the background exclusion method, were modeled as dependent variables in the regression. The LVI and T2HI were independent regressors. Of the eight maps computed (2 tissue types  $\times$  2 directions of effect), only three revealed significant voxel clusters. All three pertained to the lateral ventricle index. The results are summarized in Fig. 6.

The largest classification mismatch that was linearly related to the size of the lateral ventricles (corrected for dependency on T2HI) was a reduction of white matter probability along the anterior horn, body, atrium and occipital horns of the lateral ventricles. Particularly near the body and the anterior horn of the ventricle, this was complemented by a reduction in gray matter, reflecting a lateral–inferior shift of the caudate nucleus and the superior thalamus. A second large region of ventricle volume dependent tissue distribution mismatch was seen in the basal ganglia. Here, reduced probability of gray matter was complemented by increased white matter probability. Lastly, smaller clusters of voxels showing increased white matter probability with increasing ventricle volume were found at the ventral side of the posterior temporal lobe. This probably reflects the more downward positioning of the ventral posterior brain.

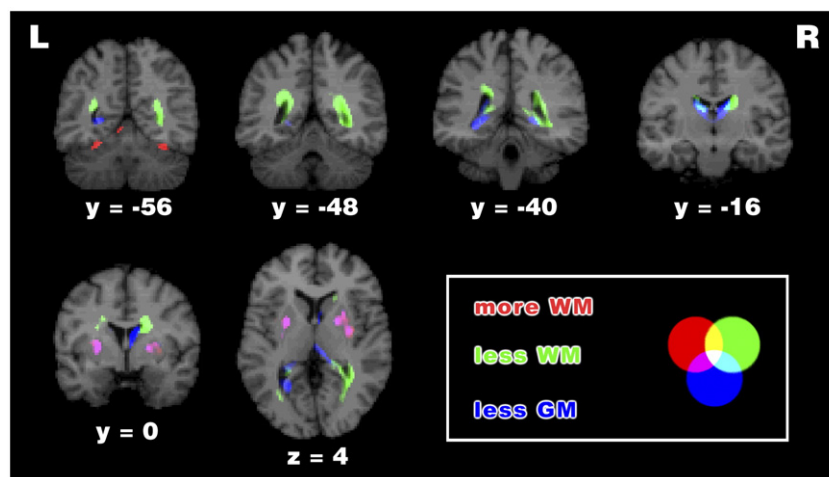


Fig. 6. Significant clusters of partial correlations between the lateral ventricle volume (corrected for the volume of hyperintense WM voxels) and smoothed (8 mm kernel) white matter and gray matter probability maps, as indicated by voxel-wise regression analysis using statistical parametric mapping. The analysis was performed on all pediatric images normalized to the MNI template using the background exclusion method. Clusters of ten or more significant voxels at the corrected alpha level of 0.05 are overlaid on the ICBM high-resolution single-subject template (LONI). Five coronal slices and one axial slice are presented. Red indicates clusters of positive partial correlations between white matter maps and lateral ventricle volume (LVI); green indicates a negative partial correlation between white matter maps and LVI; blue indicates a negative partial correlation between the gray matter maps and LVI (L=left; R=right).

## Discussion

Even without any preprocessing, linear normalization of adult and child brain images proved highly robust against outliers. Minor changes in procedure were needed to improve the robustness for pediatric MR images with ventricle enlargement and white matter abnormalities. No more outliers were observed for these abnormal images when correcting for total brain volume or removing background noise. Without any preprocessing, the successfully normalized adult images were more accurately normalized than the successfully normalized pediatric images. An overall gain in accuracy was achieved by eliminating the weight of the image background from the computations. Nonetheless, the accuracy remained lower in radiologically normal child images as well as in images evidencing white matter disease. This result cannot be fully explained by biological differences between pediatric and adult brains because we evaluated the performance of the algorithm against the best possible coregistration that can be obtained with the specific landmark configuration of each individual image. The result is also not entirely due to differences in signal intensities and distribution of different tissue classes because the normalization accuracy did not significantly improve when using a child template instead of the adult MNI template.

Although the volume correction method solved the outlier problem, it did not result in improved accuracy for the successfully normalized images. Correcting the slant of the brain or co-localizing the AC in floating and reference image did not influence the robustness or the accuracy of the results. These findings indicate that the normalization algorithm is very well able to overcome initial differences in position and size between floating and reference images.

Smoothing did not reduce the amount of outliers and even decreased the landmark co-localization for the successfully normalized images compared to the baseline method. While smoothing increases the similarity in intensity distributions between floating and reference images, it also reduces the clustering of voxels within the distribution, which corresponds to the anatomic differentiations in the image. Consequently, many more possible matching solutions between the two fuzzy images are acceptable for the algorithm. This increases the chance that the algorithm gets stuck in the wrong local maximum while seeking convergence between a floating and a reference image.

Removing the background noise from the MR images yielded the best results, for both pediatric and adult images, and regardless of the neurological characteristics of the subject. Although the intensity of background noise is usually very low compared to the intensity of brain voxels, the disproportionately large volume of background compared to brain voxels in MR images lends a substantial weight to background noise in the assessment of mutual information. By removing the background, the algorithm is no longer influenced by these low intensity signals and is more focused on brain voxels. For the same reason, the skull stripping method performed well, although this method did yield some unsuccessful solutions. For adult images, skull stripping was less accurate than removing background noise. Conversely, for children in the PVL group it outperformed the background exclusion method. For the two other pediatric groups no difference between skull stripping and background exclusion was found.

The present study also allows for a comparison between adult and pediatric templates. The landmarks defined on the adult brain images fitted better the MNI template than the CCHMC template.

Since pediatric brains have been shown to differ from adult brains in size and morphology (Giedd et al., 1996; Sowell et al., 1999, 2002), one would expect pediatric brains to match better the CCHMC than the MNI template. Remarkably, the mathematically optimal normalization solution for pediatric images did not depend on the choice of template. This finding might be due to more heterogeneity in the pediatric template. The CCHMC templates are divided into three age groups: young (5–10 years), medium (10–13 years) and old (13–19 years). Neither of these age ranges fully covered any of our three pediatric groups (Table 1). Therefore, the CCHMC template based on the entire data set of 200 children was chosen in the present study. This template represents the variability in normal age-related brain morphology for children from 5 to 19 years old. This might explain why the landmarks on this template do not correspond better with the child brains than the MNI template. More age-appropriate pediatric templates may yield a mathematically better correspondence between the two sets of landmarks. The fact that all pediatric subjects in our study were born prematurely might also explain why the optimal landmark fit did not improve with a pediatric template. Even in the absence of overt white matter damage very preterm infants can show reduced gray matter (Ajayi-Obe et al., 2000; Inder et al., 2005), in the cortex as well as in deep gray matter and in myelinated white matter (Inder et al., 2005), although these abnormalities are much more pronounced with increasing prematurity and with additional white matter abnormality.

Without any modifications to the normalization procedure, pediatric images were more accurately normalized to the CCHMC template than to the MNI template. Removing the background noise resulted in even better accuracy levels when normalizing these pediatric images to the MNI template. This effect of background noise was also observed for the CCHMC template. However, with background noise removed from the images, the use of the CCHMC template did not significantly improve normalization accuracy over the use of the MNI template. Together, these findings imply that the removal of background noise is the most important modification for accurate normalization of pediatric images. With background noise removed, considerations beyond accuracy may determine the choice between the MNI and the CCHMC template.

Three crucial factors might explain the results obtained with the CCHMC template. Firstly, the CCHMC template reflects the large variability in normal age-related brain morphology. This might allow more variable solutions to the normalization of individual child brain images. Secondly, the CCHMC template was created from 3.0 Tesla images, while our images were acquired on a 1.5-T scanner. The present study cannot evaluate the contribution of the difference in magnetic field strengths, but this difference might have corrupted the intensity-based normalization algorithm. Lastly, all our pediatric subjects were born prematurely. As noted above, this fact might lead to a different tissue composition and hence might influence the intensity-driven linear normalization algorithm.

The present study also investigates the difference in normalization accuracy between the three pediatric subgroups. No method included in this study revealed a differential effect on the relative accuracy indices for the three subgroups. This implies that the intensity-based linear normalization algorithm is not influenced by lateral ventricle enlargement or white matter abnormalities in terms of co-localizing two sets of landmarks. In terms of optimal accuracy

indices, some remarkable results were obtained. On the one hand, no difference in optimal accuracy indices was found between the radiologically normal and the PVL+ children. This implies that the conspicuous and usually extensive lateral ventricle enlargement has no impeding effect on the best possible linear normalization solution. This finding may be explained by the fact that the predefined landmarks did not directly encode the position of the ventricles because it was obvious that affine normalization cannot and should not normalize such large structural abnormalities. The regression analyses on the tissue probability maps show that this goal is achieved. The main regions of misalignment were located in periventricular and subcortical regions, while the boundaries between cerebrospinal fluid and gray matter and between gray and white matter in the more peripheral brain were well aligned with the template. On the other hand, a remarkable difference in optimal accuracy indices was found between the PVL group and the two other pediatric groups. The landmarks on the PVL affected images without ventricle enlargement could not be brought equally close to the landmarks on the two templates than the landmarks for the two other pediatric groups. This unexpected finding might be due to the fact that our sample of PVL children is accidentally atypical. Or, it might be due to unknown characteristics of periventricular leukomalacia that could alter the brain in such a way that the predefined landmarks deviate more from the template. There is still no good explanation why PVL results in enlarged ventricles in some cases and not in others.

## Conclusion

Our results show that the coregistration of patient images with PVL and lateral ventricle dilatation can be considerably improved without increased sophistication of the algorithms and procedures. Simple modifications as eliminating background noise, and/or performing a global volume correction suffice for successful normalization. Moreover, the accuracy of the normalization solution seems relatively unaffected by the occurrence of extensive structural deformations, at least for the outer brain structures. Of course, the effect of the structural abnormality on non-linear normalization needs yet to be established. But this effect cannot be evaluated fully without the knowledge that the starting point of the non-linear analysis – i.e. the linearly normalized image – is robust against the amount of structural abnormality. The present study is not a comprehensive validation of linear spatial normalization procedures for patient brain images. Nonetheless, it shows that with minor modifications, at least the linear normalization step is feasible for clinical as well as research purposes with respect to children with PVL.

## Acknowledgments

Supported by the K.U. Leuven Research Fund (K.U. Leuven Onderzoeksfonds) grants nr. OT/01/43, PDM/01/156 and PDM/03/251. L. Lagae is holder of the “UCB chair in cognitive dysfunctions in childhood” at the K.U. Leuven.

## References

Ajayi-Obe, M., Saeed, N., Cowan, F.M., Rutherford, M.A., Edwards, A.D., 2000. Reduced development of cerebral cortex in extremely preterm infants. *Lancet* 356, 1162–1163.

- Arndt, S., Rajarethinam, R., Cizadlo, T., O’Leary, D., Downhill, J., Andreasen, N.C., 1996. Landmark-based registration and measurement of magnetic resonance images: a reliability study. *Psychiatry Res.: NeuroImaging* 67, 145–154.
- Ashburner, J., Friston, K.J., 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* 6, 209–217.
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—The methods. *NeuroImage* 11, 805–821.
- Ashburner, J., Neelin, P., Collins, D.L., Evans, A.C., Friston, K.J., 1997. Incorporating prior knowledge into image registration. *NeuroImage* 6, 344–352.
- Baker, L.L., Stevenson, D.K., Enzmann, D.R., 1988. End-stage periventricular leukomalacia: MR evaluation. *Radiology* 168, 809–815.
- Bookstein, F.L., 2001. “Voxel-based morphometry” should not be used with imperfectly registered images. *NeuroImage* 14, 1454–1462.
- 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.
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multimodality image registration using information theory. In: Bizais, Y., Barillot, C., Di Paola, R. (Eds.), *Information Processing in Medical Imaging*. Kluwer Academic Publishers, Dordrecht, pp. 263–274.
- D’Agostino, E., Maes, F., Vandermeulen, D., Suetens, P., 2003. A viscous fluid model for multimodal non-rigid image registration using mutual information. *Med. Image Anal.* 7, 565–575.
- Evans, A.C., Collins, D.L., Neelin, P., MacDonald, D., Kamber, M., Marrett, T.S., 1994. Three-dimensional correlative imaging: applications in human brain mapping. In: Thatcher, R.W., Hallett, M., Zeffiro, T., John, E.R., Huerta, M. (Eds.), *Functional Neuroimaging: Technical Foundations*. Academic Press, San Diego, pp. 145–162.
- Flodmark, O., Lupton, B., Li, D., Stimac, G.K., Roland, E.H., Hill, A., Whitfield, M.F., Norman, M.G., 1989. MR imaging of periventricular leukomalacia in childhood. *Am. J. Roentgenol.* 152, 583–590.
- Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S.J., 1991. Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab.* 11, 690–699.
- Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.-B., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalization of images. *Hum. Brain Mapp.* 3, 165–189.
- Giedd, J.N., Snell, J.W., Lange, N., Rajapakse, J.C., Casey, B.J., Kozuch, P.L., Vaituzis, A.C., Vauss, Y.C., Hamburger, S.D., Kaysen, D., Rapoport, J.L., 1996. Quantitative magnetic resonance imaging of human brain development: ages 4–18. *Cereb. Cortex* 6, 551–560.
- Grachev, I.D., Berdichevsky, D., Rauch, S.L., Heckers, S., Kennedy, D.N., Caviness, V.S., Alpert, N.M., 1999. A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. *NeuroImage* 9, 250–268.
- Hahn, H.K., Peitgen, H.O., 2003. IWT—Interactive Watershed Transform: a hierarchical method for efficient interactive and automated segmentation of multidimensional gray-scale images. *Proc.-SPIE Med. Imag.* 5032, 643–653.
- Hahn, H.K., Millar, W.S., Klinghammer, O., Durkin, M.S., Tulipano, P.K., Peitgen, H.O., 2004. A reliable and efficient method for cerebral ventricular volumetry in pediatric neuroimaging. *Methods Inform. Med.* 43 (4), 376–382.
- Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J., 2001. Medical image registration. *Phys. Med. Biol.* 46, R1–R45.
- Inder, T.E., Warfield, S.K., Wang, H., Hüppi, P.S., Volpe, J.J., 2005. Abnormal cerebral structure is present at term in premature infants. *Pediatrics* 115, 286–294.
- Karas, G.B., Burton, E.J., Rombouts, S.A.R.B., van Schijndel, R.A., O’Brien, J.T., Scheltens, Ph., McKeith, I.G., Williams, D., Ballard, C., Barkhof, F., 2003. A comprehensive study of grey matter loss in patients with Alzheimer’s disease using optimized voxel-based morphometry. *NeuroImage* 18, 895–907.
- Maalouf, E.F., Duggan, P.J., Rutherford, M.A., Counsell, S.J., Fletcher, A.M.,

- Battin, M., Cowan, F., Edwards, A.D., 1999. Magnetic resonance imaging of the brain in a cohort of extremely preterm infants. *J. Pediatr.* 135, 351–357.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* 16 (2), 187–198.
- Manly, B.J.F., 1991. *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- Muzik, O., Chugani, D.C., Juhász, C., Shen, C., Chugani, H.T., 2000. Statistical parametric mapping: assessment of application in children. *NeuroImage* 12, 538–549.
- Salmund, C.H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2002. The precision of anatomical normalization in the medial temporal lobe using spatial basis functions. *NeuroImage* 17, 507–512.
- Shen, D., Davatzikos, C., 2003. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *NeuroImage* 18, 28–41.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Sowell, E.R., Thompson, P.M., Holmes, C.J., Batth, R., Jernigan, T.L., Toga, A.W., 1999. Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage* 9, 587–597.
- Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L., 2002. Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. *Dev. Med. Child Neurol.* 44, 4–16.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* 18, 897–908.
- Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imaging* 20, 677–688.
- Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multimodal volume registration by maximization of mutual information. *Med. Image Anal.* 1, 35–51.
- Wilke, M., Schmithorst, V.J., Holland, S.K., 2002. Assessment of spatial normalization of whole-brain magnetic resonance images in children. *Hum. Brain Mapp.* 17, 48–60.
- Wilke, M., Kassubek, J., Ziyeh, S., Schulze-Bonhage, A., Huppertz, H.J., 2003. Automated detection of gray matter malformations using optimized voxel-based morphometry: a systematic approach. *NeuroImage* 20, 330–343.
- Woods, R.P., Cherry, S.R., Mazziotta, J.C., 1992. Rapid automated algorithm for aligning and reslicing PET images. *J. Comput. Assist. Tomogr.* 16, 620–633.
- Woods, R.P., Grafton, S.T., Watson, J.D.G., Sicotte, N.L., Mazziotta, J.C., 1998. Automated Image Registration: II. Intersubject validation of linear and nonlinear models. *J. Comput. Assist. Tomogr.* 22, 153–165.
- Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* 12, 900–918.