ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 41 (2008) 903–913

Template-O-Matic: A toolbox for creating customized pediatric templates

Marko Wilke,[a,b,*] Scott K. Holland,[c] Mekibib Altaye,[c] and Christian Gaser[d]

[a]*Department of Pediatric Neurology and Developmental Medicine, Children's Hospital, University of Tübingen, Germany*
[b]*Section for Experimental MR of the CNS, Department of Neuroradiology, University of Tübingen, Germany*
[c]*Department of Pediatrics, University of Cincinnati, and Imaging Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA*
[d]*Department of Psychiatry, University of Jena, Germany*

Received 2 January 2008; revised 21 February 2008; accepted 25 February 2008
Available online 8 March 2008

Processing pediatric neuroimaging data is a challenge due to pervasive morphological changes that occur in the human brain during normal development. This is of special relevance when reference data is used as part of the processing approach, as in spatial normalization and tissue segmentation. Current approaches construct reference data (templates) by averaging brain images from a control group of subjects, or by creating custom templates from the group under study. In this technical note, we describe a new, and generalized method of constructing such appropriate reference data by statistically analyzing a large sample ($n=404$) of healthy children, as acquired during the NIH MRI study of normal brain development.

After eliminating non-contributing demographic variables, we modeled the effects of age (first, second, and third-order terms) and gender, for each voxel in gray matter and white matter. By appropriate weighting with the parameter estimates from these analyses, complete tissue maps can be generated automatically from this database to match a pediatric population selected for study. The algorithm is implemented in the form of a toolbox for the SPM5 image data processing suite, which we term Template-O-Matic. We compare the performance of this approach with the current method of template generation and discuss the implications of our approach.

Introduction

Magnetic resonance imaging (MRI) has become the imaging method of choice for developmental neuroscience as it offers a non-invasive window into the development of the human brain (Schaer and Eliez, 2007; Wilke and Holland, in press). Despite the difficulties associated with scanning children (Byars et al., 2002), numerous studies have now used MRI to describe normal and abnormal brain development (Castellanos et al., 2002; Giedd et al., 1996; Gogtay et al., 2004; Gothelf et al., 2007; Lenroot et al., 2007; Peterson et al., 2003; Reiss et al., 1996; Schmithorst et al., 2005; Wilke and Holland, 2003; Wilke et al., 2003b).

However, processing of MR images from children poses distinct problems because several steps in image post-processing require implicit or explicit use of reference data derived from adults. For example, routine procedures like tissue segmentation or spatial normalization, if based on adult reference data, have the potential of introducing a severe bias into pediatric imaging data (Hoeksma et al., 2005; Machilsen et al., 2007; Muzik et al., 2000; Wilke et al., 2002, 2003a). Spatial normalization of pediatric imaging data based on adult prior probability maps has been suggested to be sufficiently accurate for coarse-resolution fMRI data, following smoothing (Burgund et al., 2002; Kang et al., 2003), but errors become increasingly important when using structural or functional imaging data with a higher resolution. This emphasizes the importance of using appropriate reference data when processing pediatric imaging data, ideally based on a large sample of subjects reflecting the characteristics of the population under study (Good et al., 2001).

An average template created from a small number of reference subjects may not capture enough variance in the template and may also introduce bias. Acquiring normal, age-appropriate, pediatric brain image reference data is difficult, costly and time consuming, so it is not feasible for every pediatric brain imaging study to construct its own template based on a large normative sample. Until recently, pediatric brain image reference data from a large control population was not readily available to those wishing to construct a template matched to a given study population. This has now changed with the completion of a large-scale MRI study of normal brain development, conducted over several sites in the USA (Evans et al., 2006). It would be straightforward to process this

* Corresponding author. Department of Pediatric Neurology and Developmental Medicine, Children's Hospital, University of Tübingen, Hoppe-Seyler-Str. 1, 72076 Tübingen, Germany. Fax: +49 7071 29 5473.
E-mail address: Marko.Wilke@med.uni-tuebingen.de (M. Wilke).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.02.056

data as described before (Wilke et al., 2002, 2003a) and to generate appropriate average templates from this normal database. However, such an average template created from reference subjects will also bring about unwanted effects, such as not capturing enough variance especially when only few subjects contribute to the template.

As an alternative to using static averages of individual probability maps, here we propose a dynamic statistical approach. This is implemented by analyzing the normal database and then reconstructing appropriate reference data, not from the individual datasets, but from a statistical model that estimates the influence of all variables of interest on the tissue probabilities. For example, if the influence of age is appropriately modeled within each voxel, this information can then be used to construct a prototypical gray matter map for any given age (within the range of available reference data). This transcends previous approaches to provide pediatric (Wilke et al., 2002; Wilke and Holland, 2003) or adult (Hill et al., 2002; Mazziota et al., 1995, 2001) reference data as the influence of specific variables of interest can be isolated and unwanted sources of variance can be removed from the data. We believe this approach represents a significant step forward for brain image data analysis in general and for pediatric neuroimaging studies specifically, where it is imperative to consider explicitly dynamic morphology to avoid biasing results. With this technical note, we describe such an approach.

Methods and subjects

Subjects: origin of the data

Data used in the preparation of this article were obtained from the Pediatric MRI Data Repository created by the NIH MRI Study of Normal Brain Development. This multisite study of typically developing children, from ages newborn through young adulthood was conducted by the Brain Development Cooperative Group and supported by the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute of Neurological Disorders and Stroke (Evans et al., 2006). We used data from objective 1, which included children from age about 5–18 years. MRI data from objective 2 (newborns, infants, and toddlers; Almli et al., 2007) was not yet publicly available at the time of the preparation of this manuscript. Overall, imaging data from 432 subjects from objective 1 was included; as data quality interfered with data processing in 28 subjects (see below), a final 404 subjects were included (see Table 1 for an overview of the age and gender distribution).

Identification of key demographic variables

As part of the NIH study on normal brain development, a comprehensive neuropsychological assessment was completed by all subjects, and numerous demographic variables were recorded and are now made available to registered researchers (Evans et al., 2006). In order to adequately capture and describe normal brain development as attempted here, external variables known to influence brain development must be included in the statistical model. At the same time, it is not meaningful to include variables into such a model that do not substantially contribute to the model (i.e., do not explain enough variance in the model; see below for our procedure to identify such non-contributing variables), or that are not likely to be available for inclusion by the user. From the available demographic details on all subjects, we therefore decided to identify candidate variables that i.) have been suggested to significantly influence brain structure in the past; ii.) can be defined unambiguously, and iii.) are likely to be available for consideration in a model by the user of the algorithm.

Table 1
Age and gender distribution of the overall sample

	Ages 4–5.9	Ages 6–7.9	Ages 8–9.9	Ages 10–11.9	Ages 12–13.9	Ages 14–15.9	Ages 16–18.9	Total
Boys	19	41	31	22	30	21	28	192
Girls	19	51	31	41	23	20	27	212

To this effect, the following candidate variables were identified: age (Castellanos et al., 2002; Giedd et al., 1996; Good et al., 2001; Wilke et al., 2002), gender (Gogtay et al., 2004; Good et al., 2001; Lenroot et al., 2007; Wilke et al., 2007), handedness (Amunts et al., 1996; Hervé et al., 2006; Narr et al., 2007), and cognitive abilities (Colom et al., 2006; Haier et al., 2004; Reiss et al., 1996; Schmithorst et al., 2005; Thompson et al., 2001; Wilke et al., 2003b). Age was measured in months at date of scan (due to the rapid changes occurring during normal brain development, a grading in years was considered too broad [Wilke et al., 2002, 2003a,b, 2007]; this is illustrated by a 6-year-old being "20% older" than a 5-year-old, even if the difference may only be one month). As the global effects of age on brain structure were found to be highly non-linear (Castellanos et al., 2002; Giedd et al., 1996; Lenroot et al., 2007; Peterson et al., 2003; Wilke et al., 2003b; Wilke and Holland, in press), orthogonalized second- and third-order expansions of age were also included in the exploratory analyses. Gender was entered as a binary variable (female = 0, male = 1). Handedness was assessed according to Almli, (1999) and calculated here based on the number of right-handed trials (in older children, >72 months [6 years]) or answers (in younger children, <72 months), employing the approach also used in a classical score of handedness (Oldfield, 1971). This yields a lateralization index ranging from −1 (fully left) to +1 (fully right), with $-.2 < LI < 2$ being considered bimanual. As a proxy for the level of cognitive abilities influencing brain structure, maternal and paternal education were used (see below for a discussion of our rationale). Parental education was graded according to the following scale: "less than 6th grade" (level 1); "less than high school" (level 2), "high school" (level 3), "some college" (level 4), "college" (level 5), "some graduate level" (level 6), or "graduate level" (level 7).

Testing for the influence of key demographic variables

The effect of the demographic variables was tested statistically: all variables were included in a multiple regression model which was run for two tissue classes (gray matter [GM] and white matter [WM]). Cerebrospinal fluid (CSF) maps were generated but were not statistically investigated for this manuscript as they very rarely are of specific interest to researchers; moreover, CSF volume is very low in children (Wilke et al., 2003a), making spurious segmentation results more likely. The variance explained by each factor can be determined by specifying an omnibus *F*-test, testing for the effect of each variable. The sum of *F*-values for each effect can then be expressed as the percentage of overall variance explained by all factors.

As imaging data is commonly smoothed to increase the signal-to-noise ratio (SNR) and to determine the spatial scale at which changes are most sensitively detected (matched filter theorem; Ashburner and Friston, 2000; Salmond et al., 2003; Jones et al.,

2005), we decided to explore three different smoothing widths in order to allow for a variable explaining variance differently on different spatial scales (filter width FWHM [full width at half maximum]=0, 6, and 12 mm). This results in six analyses (two tissue classes by three smoothing widths each). The results from these tests were then used to guide the selection of the variables entering the final analysis in so far as a factor was only considered if it explained more than 5% of the statistical variance in at least half of the analyses conducted.

The following variables were not included for consideration in the analyses as they did not meet the criteria laid out above; they are reported here for the sake of completeness: subject ethnicity (not Hispanic or latino: $n=374$; Hispanic or latino: 30); site location (west: 126; midwest: 150; east: 128); overall household income ($0–$5,000: $n=1$; $5,001–$10,000, $n=2$; $10,001–$15,000, $n=4$; $15,001–$25,000, $n=11$; $25,001–$35,000, $n=20$; $35,001–$50,000, $n=76$; $50,001–$75,000, $n=97$; $75,001–$100,000, $n=96$; $100,001–$150,000, $n=87$; over $150,000, $n=10$).

MR-Imaging details

Children were imaged on 7 standard 1.5 T MR scanners (General Electric Genesis Signa [4 scanners], $n=278$; Siemens Medical Systems Sonata [1 scanner], $n=72$; Siemens Medical Systems Vision [2 scanners], $n=53$; data not available for 1 subject), in each case obtaining a whole-head T1-weighted 3D dataset with the following parameters: TR=22–560 ms, median 24 ms (22 ms, $n=4$; 23 ms, $n=170$; 24 ms, $n=69$, 25 ms, $n=103$; 500 ms, $n=55$; 566 ms, $n=1$; data not available for 2 subjects); TE=8–14 ms, median 10 ms (8 ms, $n=32$; 9 ms, $n=55$; 10 ms, $n=221$, 11 ms, $n=70$; 12 ms, $n=23$; 14 ms, $n=1$; data not available for 2 subjects), pulse angle=30–90 ° (30 °, $n=346$, 90 °, $n=56$; data not available for 2 subjects), matrix=[224–256]×[46–256]×[124–256], median 256×256×124, voxel volume .97–3.09 mm^3, median 1.32 mm^3 (<1 mm^3, $n=3$; 1–1.5 mm^3, $n=339$; 1.5–2.5 mm^3, $n=6$; 2.5–3.5 mm^3, $n=56$).

Imaging data pre-processing

All processing and analyses steps were done using functionality available within the SPM5 software package (Wellcome Department of Imaging Neuroscience, University College London, UK) or using custom scripts and functions, all running within the Matlab programming environment (The Mathworks, Natick, USA). The high-resolution T1-datasets of all children were segmented using the unified segmentation algorithm available within SPM5 (Ashburner and Friston, 2005). However, in order to avoid introducing a systematic bias into the segmentation routine by using the standard adult reference data (Wilke et al., 2003a), we used a novel approach that allows, within the unified segmentation framework, to disregard the prior information maps and generate segmentation results based on voxel intensity alone (Gaser et al. 2007). This approach has already shown to be useful when handling MR imaging data from infants (Altaye et al., 2007). In the current context, segmentation without reliance on *a priori* tissue probability maps means that the algorithm can operate on the source data independent of prior datasets. It should be noted that the removal of the prior tissue information makes the algorithm slightly less robust when confronted with lower quality input data. To avoid spurious results, all segmentation results were screened visually by one experienced rater (MW), and the result was considered inadequate in 28 subjects (6%).

A Hidden Markov Random Field (HMRF; Cuadra et al., 2005) approach with a small prior probability weighting of 0.15 was used to enforce a more consistent labeling of voxels. This approach punishes isolated voxels which are unlikely to be member of a given tissue type (for example, a voxel with a medium probability for GM is more likely gray matter if all surrounding voxels are classified as gray matter with a high probability). This is implemented by calculating the MRF energy U_{26}, from the 26 surrounding voxels which aids in determining the probability of a given voxel to actually belong to this tissue class.

Finally, a "light" cleaning procedure was used that removes extracranial tissue; here, an iterative dilatation/erosion approach within SPM5 is used to remove unconnected tissue voxels which are not likely to be brain tissue. No additional manual skull stripping was done.

In order to further avoid influence from adult reference data during spatial normalization (Hoeksma et al., 2005; Machilsen et al., 2007; Muzik et al., 2000; Wilke et al., 2002), we opted to use an affine-only spatial normalization approach as the overall scaling from this procedure has been shown not to correlate with age in this age range, in contrast to the non-linear effects of spatial normalization (Wilke et al., 2002). Moreover, while the use of non-linear spatial normalization may result in a better overlap between subjects (Ashburner and Friston, 1999), it will introduce a bias by matching regional aspects in each image to a template image. This effect is undesired when creating a template that is meant to retain the regional features of the input sample. We decided against an iterative approach to creating templates where, in subsequent iterations, processing is based on their own average (Wilke et al., 2003a), for two reasons: one, average templates tend to take up a larger space within the bounding box (Thompson et al., 2000; Wilke et al., 2003a), and two, such custom-tailored "child spaces" would make the comparison with results from other imaging studies even more difficult. Additionally, our processing approach was optimized towards minimizing the influence of prior probability maps during segmentation, implying that additional processing iterations would not have improved results.

We developed an optimized affine normalization approach by matching each cleaned native-space tissue map (GM, WM, and CSF) to the respective (adult) normalized tissue map; this has the advantage of using "clean" tissue maps as opposed to the unified segmentation routine where the initial affine mapping is based on the whole head (prior to bias correction and removal of extracranial tissue; Ashburner and Friston, 2005). Consequently, using skull-stripped images has only recently been shown to improve segmentation accuracy within the unified segmentation framework (Acosta-Cabronero et al., 2008). It is important to remember that, using this affine scaling, only the global information from the priors is used, prohibiting an unwanted regional matching of the pediatric tissue maps to the adult priors; this scaling has been shown not to correlate with age (Wilke et al., 2002). The final normalization parameters were then obtained by averaging the three affine transformations from each subject. For this, the parameters for the different tissue maps were weighted as the results from the normalization of CSF must be considered to be less optimal due to the lower volume of CSF in children, and GM must be expected to yield a better fit due to the more distinct structure. Therefore, GM contributed 50%, WM 30%, and CSF 20% to the final set of parameters, resulting in an optimized canonical affine transformation matrix which was applied to all three maps. Images were written out to the "template" bounding box of SPM (182×218×182 voxels) at 1×1×1 mm resolution.

Implementation of the algorithm

The functions of the algorithm were divided into two parts: 1) regression of the reference sample (source population), and 2) template creation for the target population. These steps were separately applied to each of the tissue classes. For the first part, we used a multiple regression model, where each column of the design matrix X is represented by one (demographic) variable. As a result, we can statistically isolate and describe the influence of these crucial variables on structural brain development. As we found that only a few variables significantly explained variance in our model (see below) we restricted the regression to a limited number of variables: for our final analyses, we used age and gender because these variables explained the largest part of the variance in our data. Age was modeled as polynomial regression with up to third order terms; the different age terms were orthogonalized with regard to its preceding column. In order to estimate the parameters of our model we used the well-known equation of the general linear model

$$Y = X\beta + \varepsilon$$

Here, the observed response variable Y is expressed in terms of a linear combination of explanatory variables X and the error term ε. The parameters from the source population were then estimated (step 1) using a least squares approach with

$$\beta = X^{-1}Y$$

To create a new customized template in the second step (based on the given characteristics of the target population), we again utilize the general linear model. The estimated parameters β (obtained in the above regression, step 1) are now weighted in step 2 by the given value X_{new} to obtain the new response variable Y, such that

$$Y_{new} = X_{new}\beta$$

As we apply this equation to each voxel individually, we can now create new tissue maps with respect to any value in the new design matrix (within the range of available values from the source population, as determined in step 1). This new design matrix consists of the values of the three age terms and gender from the target population for which the tissue map should be generated. For each voxel, the linear combination of parameter estimates is then determined and used to generate a final voxel value. This procedure is repeated for the different tissue types and results in simulated maps with regard to the estimated regression of the reference sample. The whole algorithm was implemented in the form of a toolbox for SPM5; due to the user-friendly and automated generation of templates, we termed it Template-O-Matic (TOM).

Approaches to template creation

Two general approaches seem feasible to construct appropriate reference data. First, the average age, gender, etc. is calculated based on the supplied input information (i.e., the demographic variables of the sample under study), and a fitting average template is created accordingly. Here, we term this the "average approach" ($TOM_{average}$). Alternatively, the input sample could be completely matched such that one reference tissue map is generated for each input subject, and these matched reference maps would only be averaged at the end. We term this the "matched pairs approach" ($TOM_{matched\ pairs}$). Here, templates were created using both approaches, as well as using the "classical" way, by simply averaging the tissue maps of the contributing subjects. Differences between the approaches are investigated using a joint histogram, illustrating the correlations between corresponding voxel values between two images (identical images would yield a straight diagonal line; each deviation from this line indicates discordance between the input images).

Scenarios

In order to compare the performance of the new algorithm suggested here with the current standard (using a custom-made pediatric template made up from the contributing subjects), we investigated four different scenarios. Group size (more subjects capture more variance) and age range (a smaller range means less variance) would seem to be most important when assessing the performance of pediatric brain image reference data so we tested the following scenarios: a small sample ($n=12$, a typical scenario just sufficient for a random-effects study of functional MRI data; Friston et al., 1999; Thirion et al., 2007), and a larger sample ($n=48$, typically aimed at investigating structural group differences; Roja et al., 2006; Campbell et al., 2006). Both samples were investigated simulating either a small (36 months) or a large age range (72 months), resulting in four scenarios: large range/small group (LR/SG), large range/large group (LR/LG), small range/small group (SR/SG), small range/large group (SR/LG; see Table 2). Among these scenarios, the group with the small age range and the large group (SR/LG) would be expected to yield the "best" template, while the small group with the large age range (LR/SG) would be expected to yield the "worst" classical template. Both age ranges were centered at 108 months (9 years) of age, as most children are amenable to an MR-examination by that age (Byars et al., 2002). This results in age ranges of 108±18 and 108±36 months. Within these ranges, 94 and 220 subjects, respectively, could potentially be included, and the finally contributing 12 or 48 subjects were selected at random, using functionality available within Matlab.

Results

Demographic details

Of the 404 children that were included, the mean age was 128.24±46.49 months at the date of scan (range, 57–223 months [4.75–18.58 years]); there were 192 boys (47.5%) and 212 girls

Table 2
Demographic details of the four randomly drawn subsamples used for performance testing

Age range	Age	Gender		Scenario	Age	Gender	
		Girls	Boys			Girls	Boys
Large range	103.8±21.6	95	125	Small group LR/SG[a]	102.5±21.2	3	9
				Large group LR/LG	99.1±20.9	30	18
Small range	106.9±10.1	44	50	Small group SR/SG	108.7±9.7	3	9
				Large group SR/LG[b]	105.6±10.1	26	22

Notes: a: scenario expected to produce the "worst" template; b: scenario expected to produce the "best" template; see also Figs. 4–7.

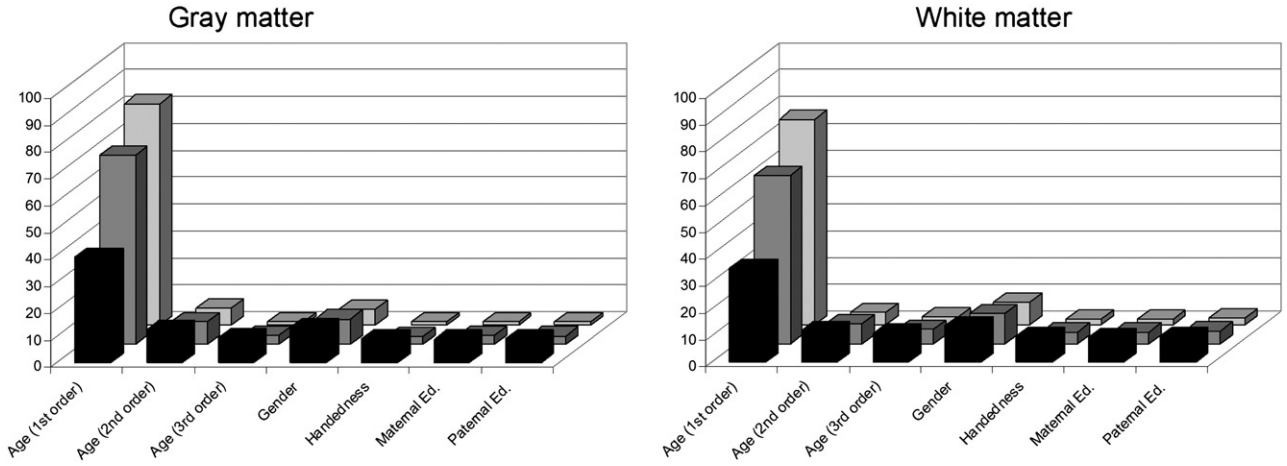

Fig. 1. Variance explained by the included demographic variables; 0 (front row), 6 (middle row), and 12 mm FWHM Gaussian smoothing filter (back row). See also Table 3.

(52.5%), see also Table 1. Right-handedness was present in 360 children; 37 children were considered left-handed, while 5 children were considered bimanual. Data was lacking for 2 children. Parental education was level 1 in $n=0/1$ (maternal/paternal), level 2 in $n=4/9$, level 3 in $n=52/81$, level 4 in $n=123/108$, level 5 in $n=132/107$, level 6 in $n=21/17$, and level 7 in $n=70/79$. No Information (level 0) was available for $n=2/2$ parents. None of the variables correlated significantly with any other variable, with the exception of maternal and paternal education which correlated strongly ($r=.48$, $p<0.001$).

Influence of demographic variables

For both tissue types (GM, WM) and all smoothing widths (0, 6, 12 mm FWHM), age (1st order term) dominated the amount of explained variance. This trend became more apparent with the increase in SNR and spatial scale (see Fig. 1 and Table 3). Of the other variables, the second and third order expansions of age and gender also fulfilled the inclusion criteria. Neither handedness nor maternal or paternal education explained more than 5% of the variance in at least half of the analyses; both variables were thus omitted from all further analyses.

Comparison of new approaches to template creation

Both approaches ($TOM_{average}$ and $TOM_{matched\ pairs}$) yielded visually indistinguishable, high-quality templates for all scenarios. When comparing the two approaches for the two extreme scenarios (LR/SG and SR/LG), it was apparent that the differences between them are minimal (Fig. 2). There was better agreement when a larger group from a smaller range was under study (Fig. 2, right panels). For conceptual reasons (see discussion section), we chose to use the $TOM_{matched\ pairs}$ approach for further comparisons.

Table 3
Percentage of explained variance per retained demographic variable, for each smoothing width and both tissue classes

	FWHM=0		FWHM=6		FWHM=12	
	GM	WM	GM	WM	GM	WM
Age (1st order term)	39.3	35.1	70	62.6	82.5	76.6
Age (2nd order term)	12.1	11.7	8.2	7.5	6.3	4.7
Age (3rd order term)	9.4	10.7	3.3	5.4	1.4	3.1
Gender	13.1	13.5	9.3	11.2	5.9	8.5

See also Fig. 1.

Comparison of classical versus matched pairs approach

A direct comparison of the LR/SG classical gray matter template with the corresponding $TOM_{matched\ pairs}$ template is shown in Fig. 3. When visually assessing the resulting classical templates from the four groups (Figs. 4–7, left panels), it is immediately apparent that the larger groups yield a more distinct classical template. In contrast to this, the $TOM_{matched\ pairs}$ template yielded very distinct tissue maps already in the small group with the large age range (LR/SG, Fig. 4, middle panels) for both GM and WM. When comparing the classical with the new $TOM_{matched\ pairs}$ approach, the differences are substantial when smaller groups are assessed (LR/SG and SR/SG, Figs. 4 and 6) and diminish when larger groups contribute to the classical template (LR/LG and SR/LG, Figs. 5 and 7).

Discussion

In this work, we suggest that, as an alternative to custom (pediatric) template creation by averaging the sample under study, it is both feasible and advantageous to create matched reference data based on the statistical evaluation of a large reference sample.

Influence of key demographic variables

The dominating effect of age on brain structure in this pediatric sample was apparent over all analyses (see Fig. 1 and Table 3), and the linear as well as the second- and third-order expansions of age were therefore included in the model. This is in line with previous studies investigating brain development (Castellanos et al., 2002; Giedd et al., 1996; Good et al., 2001; Wilke et al., 2002). It seems surprising that the non-linear terms did not account for more variance in this sample, but this can be explained by the orthogonalization of these terms (they therefore only accounted for additional variance not already accounted for by the linear term) and the fact

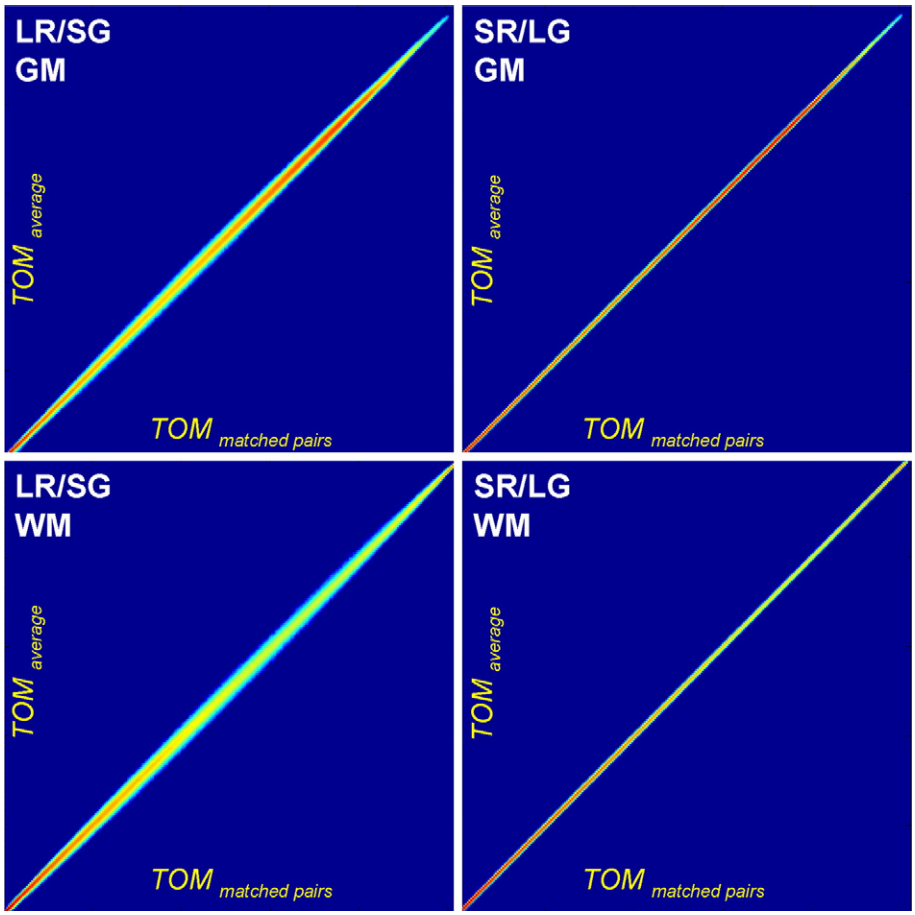

Fig. 2. Comparison of approaches: 2D histogram of the voxel classifications in the templates generated using the toolbox with the "average" (*y*-axis) and the "matched pairs" approach (*x*-axis), for the LR/SG (left panels) and the SR/LG scenario (right panels), for both GM (top panels) and white matter (bottom panels). Note the excellent agreement, more so with increasing number of contributing subjects; also note the disagreement mainly for intermediate voxel probability values, while very low (bottom left) and very high (top right) values correlate well.

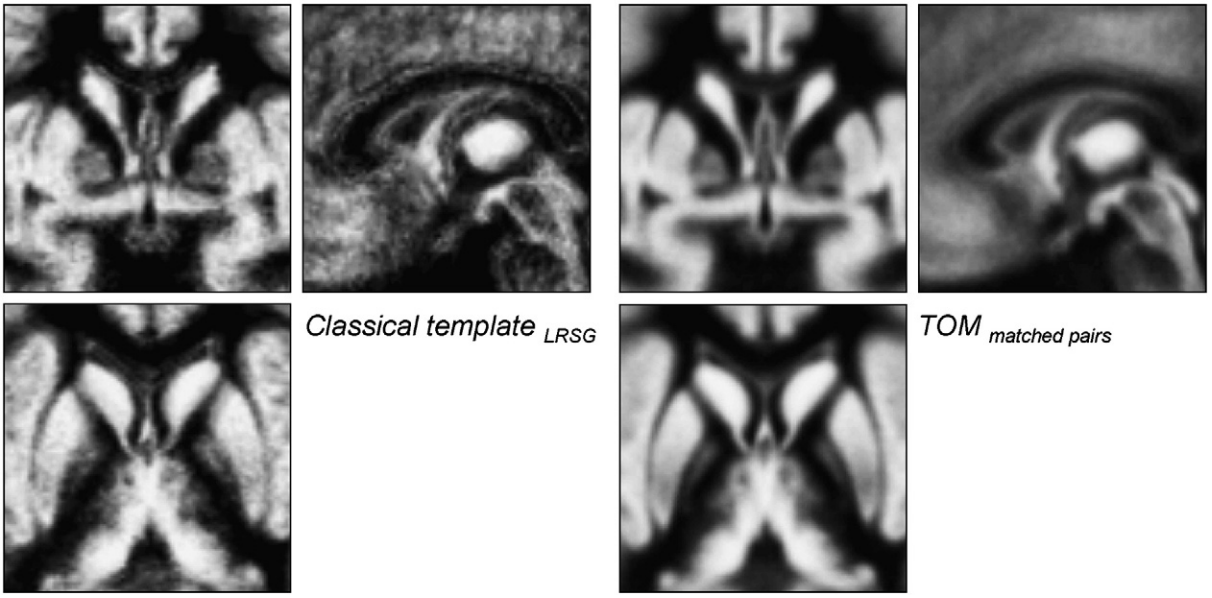

Fig. 3. Scenario large range, small group: Close-up comparison of central gray matter structure representation in templates generated in the classical way (right) and using the toolbox with "matched pairs" approach (left).

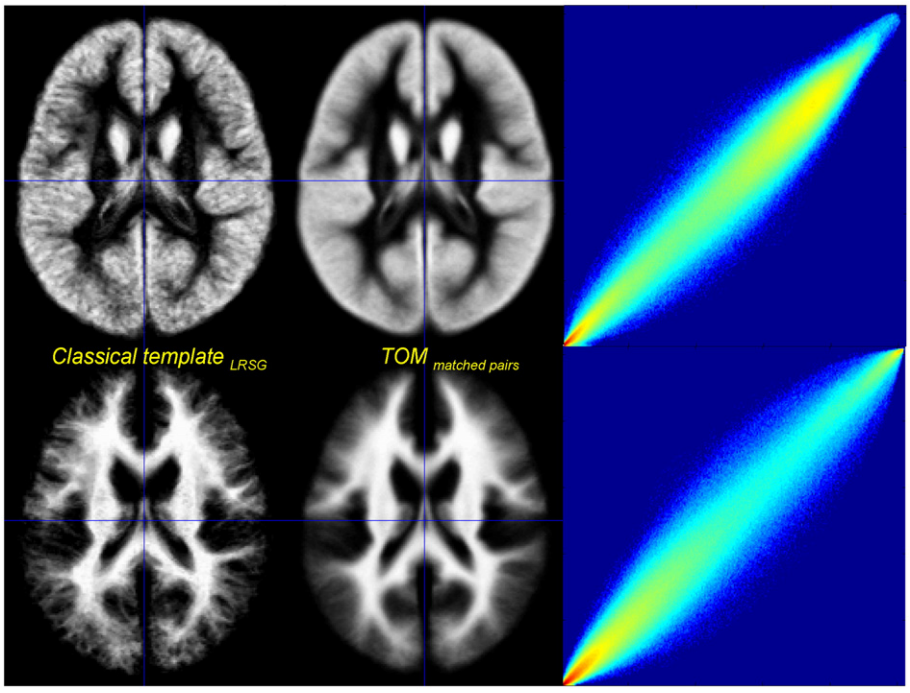

Fig. 4. Scenario large range, small group: Comparison of GM and WM templates generated in the classical way (left), and using the toolbox with the "matched pairs" approach (middle). Note the much more distinct TOM tissue map and the large disagreement in voxel classification probability values between the templates as shown by the wide spread in the histogram.

that we only used a linear spatial normalization and did not investigate regional tissue *volume*, as suggested in so-called optimized voxel-based morphometry studies (Good et al., 2001). The second-most important demographic variable was gender which also explained a substantial amount of variance in all analyses (see Fig. 1 and Table 3). Again, this was expected based on the published

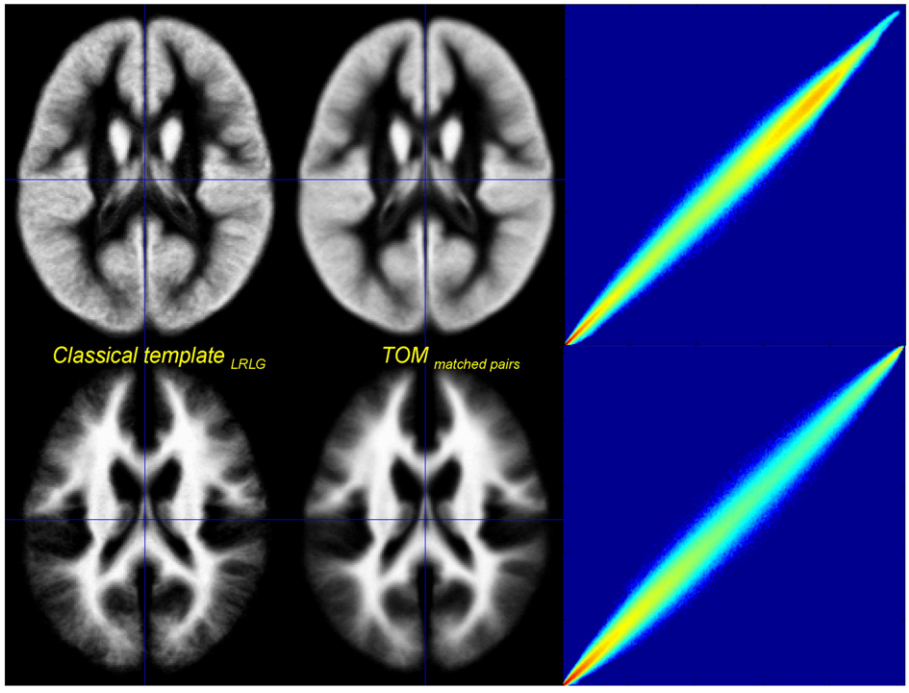

Fig. 5. Scenario large range, large group: Comparison of GM and WM templates generated in the classical way (left), and using the toolbox with the "matched pairs" approach (middle). Note the increasing quality of the classical tissue maps and the diminishing disagreement in voxel classification probability values due to the increase in contributing subjects when compared with Fig. 4. Also note the consistently higher TOM tissue map quality.

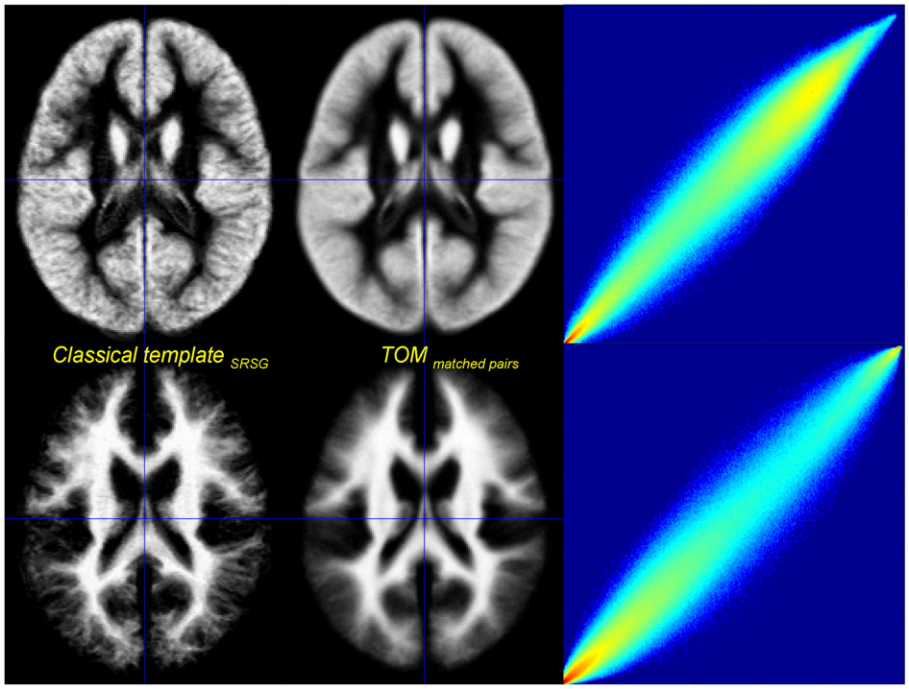

Fig. 6. Scenario small range, small group: Comparison of GM and WM templates generated in the classical way (left), and using the toolbox with the "matched pairs" approach (middle). Note the more distinct classical tissue map due to the decrease in age range when compared to Fig. 4 but the still substantial disagreement in voxel classification probability values as expressed by the wide spread in the histogram.

body of data suggesting a clear influence of gender on brain structure and size in children (Gogtay et al., 2004; Good et al., 2001; Lenroot et al., 2007; Wilke et al., 2007).

None of the other variables fulfilled our inclusion criteria, including handedness: the explanatory power of this variable when assessing brain structure was rather low in this sample. This is

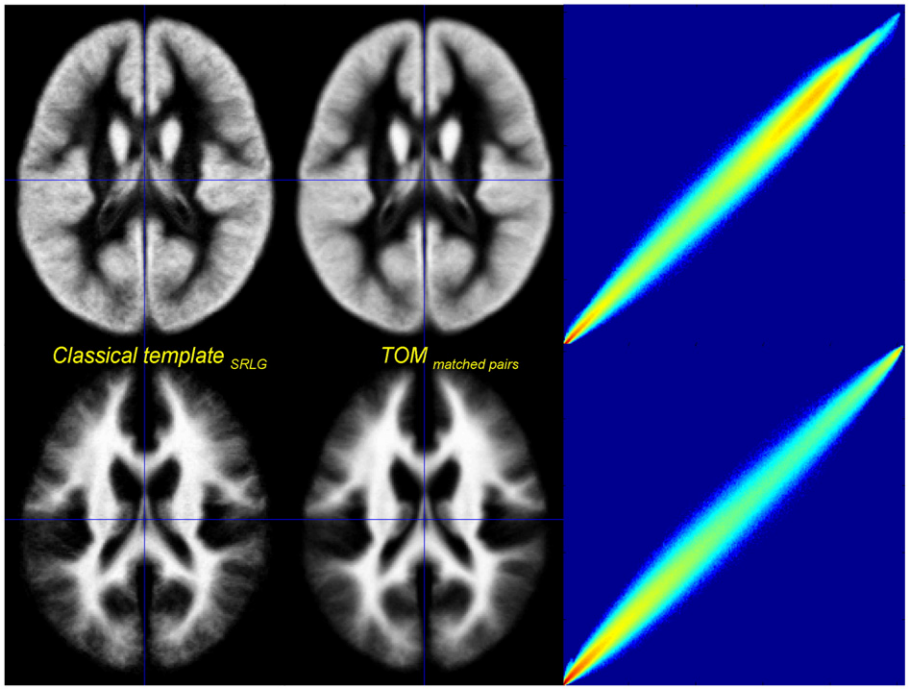

Fig. 7. Scenario small range, large group: Comparison of GM and WM templates generated in the classical way (left), and using the toolbox with the "matched pairs" approach (left). Note the further increase in the classical template quality due to the increasing number of subjects and the diminishing disagreement in voxel classification probability values when compared with Fig. 6. Also note the consistently higher TOM tissue map quality.

somewhat surprising in so far as several studies described an influence of handedness on brain structure (Amunts et al., 1996; Hervé et al., 2006; Narr et al., 2007). However, even large voxel-based studies also failed to find significant effects of handedness (Good et al., 2001), suggesting that the influence of handedness may not easily be captured using voxel-based approaches. As to the method to assess handedness, the simple question "what hand do you write with?" strongly correlated with the more extensive 10-item scoring system (Almli, 1999) used here (r=.91). It is therefore unlikely that other measures of handedness would explain enough of the variance in the selected data to meet our threshold of 5%. For our purposes here, handedness was therefore not considered.

As a proxy for the overall level of cognitive abilities, we used paternal and maternal education instead of the individual subject's IQ. The reason for this was that IQ can be determined in many different ways using a number of established tests, which i.) may not be available to researchers outside of the US, and ii.) may show a strong difference between populations, the reason of which is hotly debated (Suzuki and Aronson, 2005). We therefore decided to use parental education as a less ambiguous measure which also is more likely to be available within another setting or country (Hauser, 1994); additionally, it is known to correlate with household income (Goodman and Whitaker, 2002; White, 2000) which, in a recent study, has been shown to predict neuropsychological test performance in this specific sample (Waber et al., 2007). This correlation of parental education with household income was also present in our sample (r=.46 [maternal], r=.48 [paternal education]). Parental education also has the advantage of being less dynamic than household income (Goodman and Whitaker, 2002), and thus seemed preferable in this setting. However, neither parameter explained enough variance in our analyses to be retained in the final model. Similar to handedness, while specific effects of this variable on brain structure certainly are present (Colom et al., 2006; Haier et al., 2004; Reiss et al., 1996; Schmithorst et al., 2005; Thompson et al., 2001; Wilke et al., 2003b), they may be too regionally specific to influence the whole-brain results assessed here. While both variables were thus excluded from our analyses, there may well be scenarios where the inclusion of either variable (or another variable of interest) may be legitimate and indicated, for example when motor organization or the correlates of cognitive abilities are under study. Our algorithm is therefore set up to allow the inclusion of additional variables of interest when assessing a reference population.

Comparison of approaches to template creation

As apparent from Fig. 2, both approaches to template creation implemented here ($TOM_{average}$ and $TOM_{matched\ pairs}$) work well in that they create very similar templates which are of visually high quality (Figs. 3–7). However, it should be noted that due to the large database, the scenarios all were still rather balanced, as also apparent from Table 3; in the case of an imbalanced reference samples (i.e., older boys and one younger girl), a simple average approach would do injustice to the outlier, which may or may not be a desired effect: it could be argued that the template should capture only the defining characteristics of the input sample, but it could also be argued that the whole variance of the input sample should be present in the template. We believe that the latter will be preferable in most settings and that the $TOM_{matched\ pairs}$ approach might therefore be more suited to provide for such a uniform representation of the input sample in the resulting template. Consequently, we chose to use this approach here. However, as an argument can be made for both approaches, they both are implemented in the final version of the Template-O-Matic toolbox.

Performance of the new approach

When comparing the classical with the new approach to template generation, it is immediately apparent that, especially in the scenarios with a smaller number of subjects, the resulting classical templates are of lower quality when compared with the automatically-generated TOM templates (Figs. 3–4 and 6). The larger samples result in classical templates of clearly higher quality, showing the positive effect of increasing the contributing sample sizes (Figs. 5 and 7). It is remarkable that this results in a lesser deviation from the automatically-generated templates, as apparent in the higher agreement between the classical and the TOM templates (Figs. 5 and 7, left panels). This implies that the better quality of the larger-group classical templates approaches the quality of the TOM templates, suggesting that these may be considered a new gold standard, especially with smaller samples.

It will be interesting to compare the performance of our algorithm to the results from a recently proposed enhancement to the classical processing stream (Ashburner, 2007), suggesting an iterative alternative to the current straight averaging approach used for comparison here. By repeatedly matching all images using a non-linear, diffeomorphic image registration scheme and updating the resulting average image after each iteration, a much more distinct average image is achieved (Ashburner, 2007). However, this procedure still does not address the issues of a small sample size, unwanted variance and the potential bias the unified segmentation routine (using priors) does inject into the processing. Consequently, we believe that there will still be advantages of our algorithm even when the "classical" processing stream is optimized. Should the proposed unbiased approaches to evaluate the accuracy of non-rigid image registration schemes (Christensen et al., 2006) be extended to include pediatric data, it would also be very interesting to assess the impact of using the TOM-constructed maps for spatial normalization as opposed to using the standard adult maps. We hypothesize that the previously-shown bias when using adult data (Wilke et al., 2002, 2003a) may be avoided when custom-made pediatric reference data is used which at the same time provide a more distinct target for nonlinear matching. Further studies are necessary to directly assess performance of TOM reference data under such conditions.

Overall, we believe that the approach taken here, namely to statistically capture the relevant variance of key demographic variables in a large reference sample, represents a critical step forward when constructing reference data. The positive effects of this approach should be most visible in samples that clearly deviate from the "standard samples" that the processing algorithms have been designed upon, namely young and healthy Caucasians as in the case of the MNI standard brain templates (Mazziotta et al., 1995, 2001). The larger the deviation from this standard, the more the sample will benefit from customized priors; this is true for children (Muzik et al., 2000; Wilke et al., 2002, 2003a) but must also be expected for samples representing the other end of the age spectrum, as in studies investigating healthy aging (Smith et al., 2007), mild cognitive impairment (Sandstrom et al., 2006), or Alzheimer's disease (Ishii et al., 2001). Also, with higher resolution and improved processing algorithms, the sensitivity to detect group effects must also be expected to increase substantially, and approaches that were considered "good enough" for a coarse spatial resolution may not be adequate anymore for higher-

resolution data. The number of applications where an unbiased template generation is of value must therefore be expected to increase in the future.

Possible limitations of this study

It must be noted that the main limitation of this study is the fact that not all subjects were imaged at one site with an identical imaging protocol. Several sites contributed, and the imaging parameters, although very similar, still slightly deviated between sites, and different scanners, as a matter of course, cannot be expected to yield identical images even with identical imaging protocols. Also, not all potentially contributing sources of variance were even included in the initial model (e.g., the result of several neuropsychological test batteries) as they were unlikely to be available to the user of the toolbox and would therefore not be modeled. Likewise, ethnicity or race were not considered; although custom priors have been suggested for subjects from a different ethnic background (Lee et al., 2005), we believe that not enough subjects from these backgrounds were included in the sample to allow for this sufficient isolation of the influence of such factors on brain structure. Still, this dataset, due to its size, data quality, and detailed demographic description, is uniquely suited to address the question at hand, and we believe that the main contributing factors have been accurately identified.

Conclusions

Our algorithm allows for the statistical description and modeling of key demographic variables; it yields high-quality tissue maps, matched to the individual input sample. It may be particularly beneficial when smaller groups are investigated as the quality of templates created from such small groups is low. We therefore believe that the Template-O-Matic is a significant improvement over current approaches, allowing for a customized reference data generation and thus aiding in image data processing of "unusual samples", like children or elderly subjects.

Acknowledgments

This work has been supported by the *Deutsche Forschungsgemeinschaft* DFG (SFB550/C4, MW) and by a BMBF research grant "Neuroimaging" (01EV0709, CG). The algorithm is available for free download at http://www.irc.cchmc.org.

This manuscript reflects the views of the authors and may not reflect the opinions or views of the Brain Development Cooperative Group Investigators or the NIH. The contract numbers for the NIH MRI study of normal brain development were N01-HD02-3343, N01-MH9-0002, and N01-NS-9-2314, -2315, -2316, -2317, -2319 and -2320. A listing of the participating sites and a complete listing of the study investigators can be found at the website of the data coordinating center at www.bic.mni.mcgill.ca/nihpd/info/participating_centers.html.

References

Acosta-Cabronero, J., Williams, G.B., Pereira, J.M., Pengas, G., Nestor, P.J., 2008. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. NeuroImage 39, 1654–1665.

Almli, C.R., 1999. Measures of Hand Preference and Use Appropriate for Infants and Children. Washington University.

Almli, C.R., Rivkin, M.J., McKinstry, R.C., Brain Development Cooperative Group, 2007. The NIH MRI study of normal brain development (Objective-2): newborns, infants, toddlers, and preschoolers. NeuroImage 35, 308–325.

Altaye, M., Gaser, C., Mecoli, M., Huie, T., Egelhoff, J., Wilke, M., Holland, S.K., 2007. Infant brain probability templates for segmentation and normalization. NeuroImage 36 (Suppl. 1), S45.

Amunts, K., Schlaug, G., Schleicher, A., Steinmetz, H., Dabringhaus, A., Roland, P.E., Zilles, K., 1996. Asymmetry in the human motor cortex and handedness. NeuroImage 4, 216–222.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry — the methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.

Burgund, E.D., Kang, H.C., Kelly, J.E., Buckner, R.L., Snyder, A.Z., Petersen, S.E., Schlaggar, B.L., 2002. The feasibility of a common stereotactic space for children and adults in fMRI studies of development. NeuroImage 17, 184–200.

Byars, A.W., Holland, S.K., Strawsburg, R.H., Bommer, W., Dunn, R.S., Schmithorst, V.J., Plante, E., 2002. Practical aspects of conducting large-scale functional magnetic resonance imaging studies in children. J. Child Neurol. 17, 885–890.

Campbell, L.E., Daly, E., Toal, F., Stevens, A., Azuma, R., Catani, M., Ng, V., van Amelsvoort, T., Chitnis, X., Cutter, W., Murphy, D.G., Murphy, K.C., 2006. Brain and behaviour in children with 22q11.2 deletion syndrome: a volumetric and voxel-based morphometry MRI study. Brain 129, 1218–1228.

Castellanos, F.X., Lee, P.P., Sharp, W., Jeffries, N.O., Greenstein, D.K., Clasen, L.S., Blumenthal, J.D., James, R.S., Ebens, C.L., Walter, J.M., Zijdenbos, A., Evans, A.C., Giedd, J.N., Rapoport, J.L., 2002. Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. JAMA 288, 1740–1748.

Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H., 2006. Introduction to the non-rigid image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (Eds.), Proceedings of the Third International Workshop on Biomedical Image Registration. Lecture Notes in Computer Science, vol. 4057. Springer-Verlag, Berlin, pp. 128–135.

Colom, R., Jung, R.E., Haier, R.J., 2006. Distributed brain sites for the g-factor of intelligence. NeuroImage 31, 1359–1365.

Cuadra, M.B., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.P., 2005. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Trans. Med. Imaging 24, 1548–1565.

Evans, A.C., Brain Development Cooperative Group, 2006. The NIH MRI study of normal brain development. NeuroImage 30, 184–202.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? NeuroImage 10, 1–5.

Gaser, C., Altaye, M., Wilke, M., Holland, S.K., 2007. Unified segmentation without tissue priors. NeuroImage 36 (Suppl. 1), S68 see also http://dbm.neuro.uni-jena.de/vbm/vbm5-for-spm5/use-of-tissue-priors-experimental/ (current as of November 7, 2007).

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. U. S. A. 101, 8174–8179.

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.

Good, C.D., Johnsrude, I., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. Cerebral asymmetry and the effects of sex and

handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. NeuroImage 14, 685–700.
Goodman, E., Whitaker, R.C., 2002. A prospective study of the role of depression in the development and persistence of adolescent obesity. Pediatrics 110, 497–504.
Gothelf, D., Penniman, L., Gu, E., Eliez, S., Reiss, A.L., 2007. Developmental trajectories of brain structure in adolescents with 22q11.2 deletion syndrome: a longitudinal study. Schizophr. Res. 96, 72–81.
Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., Alkire, M.T., 2004. Structural brain variation and general intelligence. NeuroImage 23, 425–433.
Hauser, R.M., 1994. Measuring socioeconomic status in studies of child development. Child Devel 65, 1541–1545.
Hervé, P.Y., Crivello, F., Perchey, G., Mazoyer, B., Tzourio-Mazoyer, 2006. Handedness and cerebral anatomical asymmetries in young adult males. NeuroImage 29, 1066–1079.
Hill, D.K.G., Hajnal, J.V., Rueckert, D., Smith, S.M., Hartkens, T., McLeish, K., 2002. A dynamic brain atlas. In: Dohi, T., Kikinis, R. (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002. Springer-Verlag, Berlin Heidelberg, pp. 532–539.
Hoeksma, M.R., Kenemans, J.L., Kemner, C., van Engeland, H., 2005. Variability in spatial normalization of pediatric and adult brain images. Clin. Neurophysiol. 116, 1188–1194.
Ishii, K., Willoch, F., Minoshima, S., Drzezga, A., Ficaro, E.P., Cross, D.J., Kuhl, D.E., Schwaiger, M., 2001. Statistical brain mapping of ^{18}F-FDG PET in Alzheimer's disease: validation of anatomic standardization for atrophied brains. J. Nucl. Med. 42, 548–557.
Jones, D.K., Symms, M.R., Cercignani, M., Howard, R.J., 2005. The effect of filter size on VBM analyses of DT-MRI data. NeuroImage 26, 546–554.
Kang, H.C., Burgund, E.D., Lugar, H.M., Petersen, S.E., Schlaggar, B.L., 2003. Comparison of functional activation foci in children and adults using a common stereotactic space. NeuroImage 19, 16–28.
Lee, J.S., Lee, D.S., Kim, J., Kim, Y.K., Kang, E., Kang, H., Kang, K.W., Lee, J.M., Kim, J.J., Park, H.J., Kwon, J.S., Kim, S.I., Yoo, T.W., Chang, K.H., Lee, M.C., 2005. Development of Korean standard brain templates. J. Korean Med. Sci. 20, 483–488.
Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., Thompson, P.M., Giedd, J.N., 2007. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. NeuroImage 36, 1065–1073.
Machilsen, B., d'Agostino, E., Maes, F., Vandermeulen, D., Hahn, H.K., Lagae, L., Stiers, P., 2007. Linear normalization of MR brain images in pediatric patients with periventricular leukomalacia. NeuroImage 35, 686–697.
Mazziotta, J.C., Toga, A.W., Evans, A., Lancaster, J.L., Fox, P.T., 1995. A probabilistic atlas of the human brain: theory and rational for its development. NeuroImage 2, 89–101.
Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Simpson, G., Woods, R., Paus, T., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Schormann, T., Amunts, K., Palomero-Gallagher, N., Parsons, L., Narr, K., Kabani, N., LeGoualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 1293–1322.
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.
Narr, K.L., Bilder, R.M., Luders, E., Thompson, P.M., Woods, R.P., Robinson, D., Szeszko, P.R., Dimtcheva, T., Gurbani, M., Toga, A.W., 2007. Asymmetries of cortical shape: effects of handedness, sex and schizophrenia. NeuroImage 34, 939–948.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Peterson, B.S., Anderson, A.W., Ehrenkranz, R., Staib, L.H., Tageldin, M., Colson, E., Gore, J.C., Duncan, C.C., Makuch, R., Ment, L.R., 2003. Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. Pediatrics 111, 939–948.
Reiss, A.L., Abrams, M.T., Singer, H.S., Ross, J.L., Denckla, M.B., 1996. Brain development, gender and IQ in children. A volumetric imaging study. Brain 119, 1763–1774.
Rojas, D.C., Peterson, E., Winterrowd, E., Reite, M.L., Rogers, S.J., Tregellas, J.R., 2006. Regional gray matter volumetric changes in autism associated with social and repetitive behavior symptoms. BMC Psychiatry 6, 56.
Salmond, C.H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2003. Distributional assumptions in voxel-based morphometry. NeuroImage 17, 1027–1030.
Sandstrom, C.K., Krishnan, S., Slavin, M.J., Tran, T.T., Doraiswamy, P.M., Petrella, J.R., 2006. Hippocampal atrophy confounds template-based functional MR imaging measures of hippocampal activation in patients with mild cognitive impairment. AJNR Am. J. Neuroradiol. 27, 1622–1627.
Schaer, M., Eliez, S., 2007. From genes to brain: understanding brain development in neurogenetic disorders using neuroimaging techniques. Child Adolesc. Psychiatr. Clin. N. Am. 16, 557–579.
Schmithorst, V.J., Wilke, M., Dardzinski, B.J., Holland, S.K., 2005. Cognitive functions correlate with white matter architecture in a normal pediatric population: a diffusion tensor MRI study. Hum. Brain Mapp. 26, 139–147.
Smith, C.D., Chebrolu, H., Wekstein, D.R., Schmitt, F.A., Markesbery, W.R., 2007. Age and gender effects on human brain anatomy: a voxel-based morphometric study in healthy elderly. Neurobiol. Aging 28, 1075–1087.
Suzuki, L., Aronson, J., 2005. The cultural malleability of intelligence and its impact on the racial/ethnic hierarchy. Psychol. Pub. Pol. Law 11, 320–327.
Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B., 2007. Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. NeuroImage 35, 105–120.
Thompson, P.M., Mega, M.S., Narr, K.L., Sowell, E.R., Blanton, R.E., Toga, A.W., 2000. Brain image analysis & atlas construction. In: Fitzpatrick, M., Sonka, M. (Eds.), Handbook of Medical Image Processing and Analysis. SPIE Press, Bellingham, WA, pp. 1073–1117.
Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lönnqvist, J., Standertskjöld-Nordenstam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W., 2001. Genetic influences on brain structure. Nat. Neurosci. 4, 1253–1258.
White, S.H., 2000. Conceptual foundations of IQ testing. Psychol. Pub. Pol. Law 6, 33–43.
Waber, D.P., De Moor, C., Forbes, P.W., Almli, C.R., Botteron, K.N., Leonard, G., Milovan, D., Paus, T., Rumsey, J., 2007. The NIH MRI study of normal brain development: performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. J. Int. Neuropsychol. Soc. 13, 729–746.
Wilke, M., Holland, S.K., 2003. Variability of gray and white matter during normal development: a voxel-based MRI analysis. Neuroreport 14, 1887–1890.
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., Schmithorst, V.J., Holland, S.K., 2003a. Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. Magn. Reson. Med. 50, 749–757.
Wilke, M., Sohn, J.H., Byars, A.W., Holland, S.K., 2003b. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. NeuroImage 20, 202–215.
Wilke, M., Holland, S.K., in press. Structural MR-Imaging studies of the brain in children: Issues and Opportunities. Neuroembryol Aging.
Wilke, M., Krägeloh-Mann, I., Holland, S.K., 2007. Global and local development of gray and white matter volume in normal children and adolescents. Exp. Brain Res. 178, 296–307.