Evidence of non-normal distributions in brain imaging data from normal subjects: implications for diagnosis of disease

David Alexander Dickie[1], Dominic E Job[1], Joanna M Wardlaw[1], David H Laidlaw[2], and Mark E Bastin[1]
[1]The University of Edinburgh, Edinburgh, United Kingdom, [2]Brown University, Providence, RI, United States

Introduction/target audience: Detection of prodromal brain pathology in neurological diseases, such as schizophrenia and Alzheimer's (AD), may improve diagnoses and outcome[1]. This pathology is often subtle but statistical models of normal ageing brain MRI data may increase sensitivity[1]. The most commonly applied statistical models in brain imaging are parametric, i.e. based on the normal (Gaussian) distribution[2]. Whether or not normal ageing brain MRI volumes are distributed Gaussian, and whether this actually matters, has yet to be determined.

Purpose: This work tested whether or not regional brain MRI volumes were distributed Gaussian in a typically sized adult sample. The impact of distribution shape on the effect size of age group was then determined.

Methods: Coronal T1-weighted volume sequence brain MRI were obtained from 80 subjects (25-64 years; 50% female) using a GE Signa Horizon HDxt 1.5T clinical scanner (General Electric, Milwaukee, WI, USA). All subjects gave written informed consent and were classified as normal via medical histories and a battery of cognitive tests. Subjects were grouped by age (years): 1. 25-34 (n=21); 2. 35-44 (n=23); 3. 45-54 (n=24); 4. 55-64 (n=12).

Non–brain structure was removed from the MRI data by diffeomorphically warping[3] and applying the MNI152 brain atlas to each subject. Errors, e.g. remaining skull, were manually corrected slice-by-slice. Bias field correction was performed and whole brain tissue volumes, grey/white matter (GM/WM), and cerebrospinal fluid (CSF), were calculated using voxel intensity and spatial neighborhood information[4]. The SRI24 regional brain volume atlas was then diffeomorphically warped to each subject so to extract the left and right amygdala, hippocampus, parahippocampal (PH) gyrus, caudate, putamen, and thalamus volumes from whole brain GM volume (Figure 1). Each of these regional volumes were normalized by total intracranial volume (TIV).

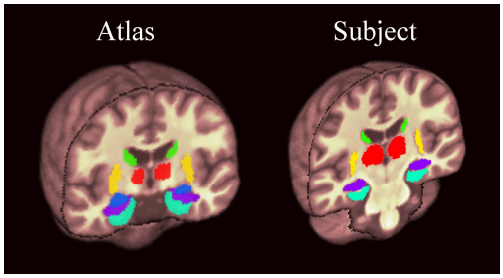

Figure 1. Atlas based segmentation using diffeomorphic registration. The SRI24 regional brain volume atlas (left) was registered to each subject (right). Regional GM volumes were then extracted from each subject.

Parametric (Gaussian) distributions were calculated with the mean and standard deviation (SD) of brain volumes for each age group. In truly Gaussian data, the 2.5th percentile rank value is approximately equal to the mean minus 2 SD, the 16th percentile rank to the mean minus one SD, the 84th percentile rank to the mean plus one SD, and the 97.5th percentile rank to the mean plus 2 SD. We used ±2 SD because ±1.96 SD often underestimated the 95% limits (97.5th–2.5th percentile rank) of simulated Gaussian brain structure data (±2 SD was a closer approximation). Percentile ranks were directly calculated by Equation 1, where n is the number of subjects, for the tth percentile $p=t/100$, j is the integer part of np, g is the fractional part of np, y is the tth percentile, and $x_1, x_2, \ldots, x_n$ are the ordered values of each brain volume. Parametric effect size of age group (SD normalised difference between groups) was calculated with "*Cohen's d*" (Equation 2) where μ_i is the mean of age group i, e.g. 1. 25-34 years, μ_j is the mean of age group j, e.g. 3. 45-54 years, and σ_p is the pooled standard deviation of the groups[5]. If data were Gaussian distributed then the result from Equation 2 would approximate the result from the nonparametric equivalent (Equation 3).

Equation 1	Equation 2	Equation 3	Equation 4
$np = j + g$ $y = 1/2(x_j + x_{j+1})$, if $g = 0$ $y = x_{j+1}$, if $g > 0$	$d = \frac{\mu_i - \mu_j}{\sigma_p}$	$\tilde{d} = \frac{\tilde{\mu}_i - \tilde{\mu}_j}{\tilde{\sigma}_p}$	$\frac{\lvert d_{i,j} - \tilde{d}_{i,j} \rvert}{\lvert d_{i,j} \rvert} \times 100$

In Equation 3, $\tilde{\mu}_i$ is the median of group i, $\tilde{\mu}_j$ is the median of group j, and $\tilde{\sigma}_p$ is the pool of the 50th percentile minus the 16th percentile in each group. Differences in parametric and nonparametric effect sizes were computed using absolute percent error (Equation 4), where where d is the parametrically defined effect size between groups i and j, and $\tilde{d}$ is the nonparametrically defined effect size between groups i and j. A sign was added to percent error if the direction of the effect differed between methods.

Results: The parametric (Gaussian) and nonparametric (actual) distributions of right hippocampal volumes are shown in Figure 2. The Gaussian distribution was often not a good approximation for the shape of the actual data. Similar results were repeated throughout the other regional brain volumes (not shown here due to space limitations). The impact of nonconformance to the Gaussian distribution is shown in Table 1. There were large, unsystematic overestimations/ underestimations of effect size by the parametric method (Table 1).

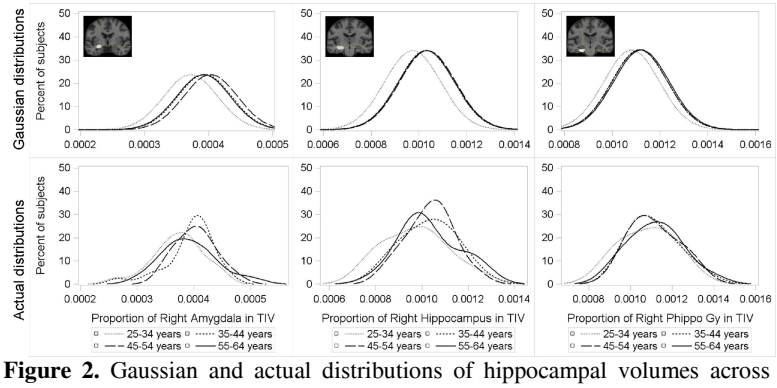

Figure 2. Gaussian and actual distributions of hippocampal volumes across adulthood. Phippo Gy=parahippocampal gyrus; TIV=total intracranial volume.

Table 1. Parametric and nonparametric effect sizes of age group in right hippocampal volumes

Volume	Age	Parametric	Nonparametric	%Err
Hippocampus	1–3	–5.29E–01	–4.41E–01	17
	1–4	–4.99E–01	–9.14E–02	82
	2–4	–2.05E–02	3.82E–01	–1964
	3–4	2.99E–02	3.50E–01	1069
PH gyrus	1–2	–2.79E–01	3.73E–04	–100
	1–3	–2.95E–01	–5.38E–02	82
	1–4	–3.52E–01	–1.34E–01	62

Discussion/Conclusion: The Gaussian distribution did not well approximate the effects of normal ageing on regional brain volumes in a typically sized adult brain imaging sample. This suggests that nonparametric statistical methods will be required to accurately model the effects of normal ageing brain structure and improve detection of prodromal neurological disease pathology.

References: [1] Farrell, C., et al. *Eur. Radiol.* 2009;19:177-183. [2] Ashburner, J., et al. *Neuroimage* 2000;11:805-821. [3] Avants, B.B., et al. *Med. Image Anal.* 2008;12:26-41. [4] Zhang, Y., et al. *IEEE Trans. Med. Imaging* 2001;20:45-57. [5] Cohen, J. *Statistical Power Analysis for the Behavioral Sciencies* 1988.