Metric distances between hippocampal shapes indicate different rates of change over time in nondemented and demented subjects

Elvan Ceyhan, PhD[1,2], M. Faisal Beg, PhD[3], Can Ceritoğlu, PhD[2], Lei Wang, PhD[4], John C. Morris, MD[5,6], John G. Csernansky, MD[4,7], Michael I. Miller, PhD[2,7], J. Tilak Ratnanather, PhD[2,7]

June 29, 2009

[1]*Dept. of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey.*
[2]*Center for Imaging Science, The Johns Hopkins University, Baltimore, MD 21218.*
[3]*School of Engineering Science, Simon Fraser University, Burnaby, V5A 1S6, Canada.*
[4]*Dept. of Psychiatry and Behavioral Sciences, Northwestern University, Feinberg School of Medicine, Chicago, IL 60611*
[5]*Dept. of Neurology, Washington University School of Medicine, St. Louis, MO 63110.*
[6]*Alzheimer's Disease Research Center, Washington University School of Medicine, St.Louis, MO 63110.*
[7]*Institute for Computational Medicine, The Johns Hopkins University, Baltimore, MD 21218.*

Address for correspondence:
Elvan Ceyhan,
Dept. of Mathematics, Koç University,
Rumelifeneri Yolu, 34450 Sarıyer,
Istanbul, Turkey

e-mail: elceyhan@ku.edu.tr
phone: +90 (212) 338-1845
fax: +90 (212) 338-1559
short title: Metric distances of hippocampal shapes change in normal aging and dementia.
keywords: morphometry, computational anatomy, Large Deformation Diffeomorphic Metric Mapping (LDDMM), hippocampus, dementia of Alzheimer Type

Abstract

In this article, we analyze the morphometry (shape and size) of hippocampus in subjects with mild dementia of Alzheimer type (DAT) and nondemented controls and how it changes over a two-year time period. Morphometric differences with respect to a template hippocampus were measured by the metric distance obtained from the Large Deformation Diffeomorphic Metric Mapping (LDDMM) algorithm. LDDMM assigns metric distances on the space of anatomical images, thereby allowing for the direct comparison and quantization of morphometric changes. Metric distances are a compound summary measure, but can still provide a notion of how similar or different the shapes are. At baseline, the metric distances in demented subjects are found not to be significantly different from those in nondemented subjects. At follow-up, right metric distances in demented subjects were significantly larger compared to those in nondemented subjects ($p=.0155$). Metric distances in demented subjects increased significantly from baseline to follow-up ($p=.0155$ and $p=.0002$ for left and right sides, respectively) but not in nondemented subjects which suggest that hippocampus in demented subjects undergo more morphometric change over time. Furthermore, combining volume and metric distance can provide a better biomarker for detection of dementia compared to volume and metric distance separately.

1 Introduction

Numerous post-mortem studies have shown that neurofibrillary tangles and amyloid plaques characteristic of Alzheimer's Disease (AD) are prominent within the hippocampus of individuals with mild dementia of the Alzheimer type (DAT) and that the distribution of these neuropathological markers becomes more widespread to include several regions of the neocortex as the disease process progresses [1-7]. The accumulation of neurofibrillary tangles and amyloid plaques characteristic of AD are associated with neuronal damage and death [8]. Furthermore, macroscopic gray matter losses from the accumulation of microscopic scale neuronal destruction are detectable in living subjects using currently available magnetic resonance (MR) imaging. Specifically, volume losses within the hippocampus [9-14] have recently been reported in subjects with mild-to-moderate AD. In a study where the antemortem MR scans and post-mortem material was available for the same subjects, hippocampal volume losses were shown to be powerful antemortem predictors of AD neuropathology [15]. Progressive atrophy of the entire brain has been observed in AD cases [16]. However, due to the complexity of the human brain and the non-uniform distribution of AD neuropathology early in the course of disease, detailed examination of specific brain regions known to be affected early in the AD disease process (e.g., hippocampus) may be preferred for distinguishing preclinical and very mild forms of AD from normal aging [17-19].

Methods developed in the field of Computational Anatomy (CA) that enable quantification of brain structure volumes and shapes between and within groups of individuals with and without various neurological diseases have emerged from several groups in recent years [20-25]. Based on the mathematical principles of general pattern theory [18, 19, 23, 26, 27], these methods combine image-based diffeomorphic maps between MR scans with representations of brain structures as smooth manifolds. Because of their high repeatability and sensitivity to changes in neuroanatomical shapes, they may be especially sensitive to abnormalities of brain structures associated with early forms of AD. Using such methods, we previously demonstrated that the combined assessment of hippocampal volume loss and shape deformity optimally distinguished subjects with very mild DAT from both elder nondemented subjects and younger healthy subjects [10]. These methods also allowed us to demonstrate that hippocampal shape deformities associated with very mild DAT and nondemented aging were distinct [28]. These methods were also extended to quantify changes in neuroanatomical volumes and shapes within the same individuals over time [29]. Other longitudinal neuroimaging analyses of hippocampal structures in individuals with AD have also emerged [30-41].

In particular, morphometric studies of shape have been carried out by metric comparison of anatomical structures via vector field displacements generated by large deformation diffeomorphic metric mapping (LDDMM) between two shapes [42, 43]. The construction of such a metric space allows one to quantify similarities and differences between anatomical shapes [44, 45]. This is the vision laid out by D'Arcy W. Thompson almost one hundred years ago [46].

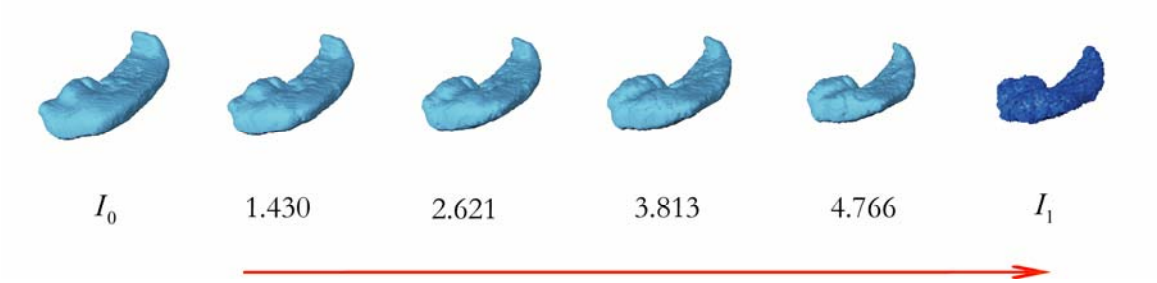


Figure 1 exemplifies the change in the metric distance during the evolution of the diffeomorphic map (1-1 and onto transformations with inverses that are smooth) from a template shape to a target shape. This notion of mathematical biomarker in the form of metric distance can be used in different ways. One is to generate metric distances of shapes relative to a template [42, 43]. Another is to generate metric distances between each shape within a collection [47]. The latter approach allows for sophisticated pattern classification analysis; it is however computationally expensive. We present an analysis based on the former approach which could provide a powerful tool in analyzing subtle shape changes over time with considerably less computational load. This approach may allow detecting the subtle morphometric changes observed in the hippocampus in subjects with DAT in particular for those previously analyzed [29, 48]. These studies compared rates of change in hippocampal volume and shape in subjects with very mild DAT and nondemented subjects matched for age and gender. The change in hippocampal shape over time was defined as a residual vector field resulting from rigid-body motion registration, and changes in patterns along hippocampal surfaces were visualized and analyzed via a statistical measure of individual and group change in hippocampal shape over time and used to distinguish between the subject groups. Hence the motivation to analyze LDDMM generated metric distances between binary hippocampus images at baseline and at follow-up with respect to the same template. That is, the template was compared

again, and not propagated between time points. In this study, we compute and analyze metric distances based on the data used in [29].

In this article, we use a computational anatomy approach to estimate, for two groups of hippocampi, their distances to a standardized prototype hippocampus of a particular healthy elderly individual. The two groups correspond to subjects with mild AD (here, CDR0.5), and no apparent AD (CDR0). A deformation is estimated between subject and prototype that is minimally-deforming in a rigorous sense and well-behaved in biological terms; the magnitude of this deformation is the distance between prototype and subject. These distances are computed on baseline and followup scans in both groups, and differences between groups, timepoints, and sides of the brain are analyzed statistically. We also provide the strengths and weaknesses (i.e., limitations) of LDDMM metric distances as a shape quantifier. We briefly describe the data set in Section 2.1, computation of metric distances via LDDMM in Section 2.2, statistical methods we employ in Section 2.3, and results in Section 3. The discussion concludes the paper.

2 Methods

2.1 Subjects and Data Acquisition

Detailed description of subjects can be found in [29] where 18 very mild DAT subjects with a Clinical Dementia Rating scale of CDR0.5 (11 male, 7 female, 74±4.4 years old) and 26 age-matched nondemented subjects (CDR0, 12 male and 14 female, 73±7.0 years old) were each scanned 2 (range 1-2.6) and 2.2 (range 1.4-4.1) years apart respectively. CDR assessments which detect the severity of dementia symptoms were performed annually in all subjects by experienced clinicians without reference to neuropsychological tests or in-vivo neuroimaging data. The experienced clinician conducted semi-structured interviews with an informant and the subject to assess the subject's cognitive and functional performance; a neurological examination also was obtained. The clinician determined the presence or absence of dementia and, when present, its severity with the CDR. Overall CDR scores of 0 indicate no dementia, while CDR scores of 0.5, 1, 2, and 3 indicate very mild, mild, moderate and severe dementia, respectively [49]. CDR assessments have been shown to have an inter-rater reliability of κ=0.74 (weighted kappa coefficient [50] κ of 0.87) [51], and this high degree of inter-rater reliability has been confirmed in multi-center dementia studies [52]. Elderly subjects with no clinical evidence of dementia (i.e., CDR0) have been confirmed with normal brains at autopsy with 80% accuracy; i.e., approximately 20% of such individuals show evidence of AD [53]. CDR0.5 subjects have subtle cognitive impairment, and 93% of them progress to more severe stages of illness (i.e., CDR > 0.5) and show neuropathological signs of AD at autopsy [53-55]. Although elsewhere the CDR0.5 individuals in our sample may be considered to have MCI [56], they fulfill our diagnostic criteria for very mild DAT and at autopsy overwhelmingly have neuropathologic AD [57].

The scans were obtained using a Magnetom SP-4000 1.5 Tesla imaging system, a standard head coil, and a magnetization prepared rapid gradient echo (MPRAGE) sequence. The sequence (TR/TE - 10/4, ACQ - 1, Matrix - 256×256 , Scanning time - 11.0 min) produced 3D data with a $1mm\times1mm$ in-plane resolution and 1 mm slice thickness across the entire cranium.

A neuroanatomical template was produced using an MR image from an additional elder control (i.e., CDR0 or non-demented) subject (male, age = 69). The choice and a detailed description of the template is provided in [58]. The subject selected to produce this template was obtained from the same source as the other subjects in the study, but was not otherwise included in the data analysis. Expert-produced manual outlines of the left and right hippocampus in the template scan created from using methods previously described [28, 59] were used as binary templates. The left and right hippocampal surfaces of each subject generated at baseline and follow-up were converted to binary hippocampus volumetric images by flood filling the inside of the surface. Each individual hippocampal surface was first scaled by a factor of 2 and aligned with the template surface, which was also scaled by a factor of 2, via a rigid-body rotation and translation before converting to volumetric binary images. In [59] we showed that mapping accuracy could be enhanced at higher resolution because of smaller voxels – voxels at the periphery of the structure (i.e., surface) account for much more of the structural volume at $1\,mm^3$ voxel resolution versus $0.5\,mm^3$. Since then we have adapted this as part of the standard mapping procedure. These surfaces were then converted into binarized image of dimension $64\times112\times64$ with voxel resolutions of $0.5\times0.5\times0.5\,mm^3$, followed by smoothing by a Gaussian filter of $9\times9\times9$ -voxel window and one voxel standard deviation to smooth out the edges for LDDMM analysis.

2.2 Computing Metric Distance via Large Deformation Diffeomorphic Metric Mapping

3. For this journal, mathematical equations on page 5 are probably not appropriate. The mathematical intuition should be summarized without equations.

Metric distances between the binary images and the template image are obtained by computing diffeomorphisms between the images. For any pair $I, J \in \mathcal{I}$, there exists a flow of diffeomorphisms $g_t, t \in [0,1]$ transforming one shape to the other $g \cdot I \sim J$. The metric distance between any pair I, J is given by the length of the shortest or geodesic curve through the space of shapes connecting them. The diffeomorphisms are constructed as a flow of ordinary differential equations $\dot{g}_t = v_t(g_t), t \in [0,1]$ with $g_0 = \mathrm{id}$ the identity map, and associated vector fields $v_t, t \in [0,1]$ [59, 60]. The metric distance between two shapes I, J takes the form

$$d(I,J)^2 = \inf_{v:\, g = \int_0^1 v_t(g_t)\,dt,\, g_0 = id} \int_0^1 \|v_t\|_V^2 \, dt,$$

such that g transforms I to J. The norm $\|\cdot\|_V$ is chosen to ensure that the vector fields are smooth in space (derivatives exist in the squared-energy sense). To calculate the norm we use LDDMM [42] by introducing a cost function measuring correspondence between mapped anatomical objects $C(g \cdot I, J)$ and then computing the geodesic connection to minimize the cost. Here, we solve the inexact matching problem forcing one shape to map onto the other, obtaining a matching cost C which is small but not identically zero. Correspondence is based on the intensity data at the voxel level; reliability has been demonstrated [42]. Vector fields connecting each template-subject pair are then used to compute metric distances, d_k^b, d_k^f $(k = 1, \ldots, 44)$, in each hemisphere at baseline (*b*) and at follow-up (*f*) as illustrated in **Figure 2**. We propose the LDDMM distance measure as a potential marker for AD-associated hippocampus changes; in some sense it measures "shape difference between a given hippocampus and a known healthy one."

2.3 Statistical Methods

We model the metric distances using repeated-measures analysis with diagnosis group, side, and timepoint as main effects and side-by-timepoint (interaction) as the repeated factor, since there is within-subject dependence of metric distances of left and right hippocampus at baseline and follow-up. The metric distances of left hippocampi at baseline for CDR0.5 subjects are labeled as LB-CDR0.5 and those at follow-up are labeled as LF-CDR0.5. CDR0 subjects are labeled as LB-CDR0 and LF-CDR0 accordingly. Similar labeling is done for the metric distances of right hippocampi. Hence, we have four measurements for each subject, so repeated measures analysis can be performed on our data set. We apply four possible variance-covariance structures in our models. The first structure assumes compound symmetry, in which the variances are equal and so are the covariances. The other three structures assume unstructured, autoregressive (AR), and autoregressive heterogeneous (ARH) variances, respectively. In the unstructured model, each variance and covariance term is distinct, in the AR model, the variances are assumed to be equal but the covariances change by time and side, and in the ARH model, the variances are also different and the covariances change by time and side. For more on the variance-covariance (Var-Cov) structures for the repeated measures models, see [61-64] We use various model selection criteria, namely, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Log-likelihood to compare competing models to see which model best fits our data [65].

When repeated measures analysis indicates significant interaction between pairs of diagnosis, side, and timepoint factors, we compare metric distances at baseline and follow-up hippocampi. The LB-CDR0.5 metric distances and LF-CDR0.5 metric distances are dependent as they come from the same person at baseline and follow-up. Likewise for LB-CDR0 and LF-CDR0 pairs. We also calculate and interpret correlation coefficients between metric distances [66, 67]. We also estimate the empirical cumulative distribution functions (cdf) of the metric distances and compare them by Kolmogorov-Smirnov (K-S) test.

We apply logistic discrimination with metric distances and other variables, since the diagnosis has only two levels, namely CDR0 and CDR0.5. We use logistic regression to estimate the risk or probability of having DAT using metric distances, together with side (i.e., hemisphere) and timepoint (baseline vs follow-up) factors (for more on logistic regression and logistic discrimination, see [68] and [69], respectively). We consider the logistic model with the response $\mathrm{logit}\, p = \log[p/(1-p)]$ where $p = P(Y = 1)$ (i.e., the probability that the subject is diagnosed with CDR0.5). First we model $\mathrm{logit}\, p$ with one predictor variable at a time from variables of side, timepoint, and metric distance, etc.; if the

variable is not significant at .05 level, we omit that variable from further consideration. That is, we perform a univariate analysis with just the time variable, just the side variable, etc. as predictors in the model. The remaining variables with all possible interactions among them as the predictor variables constitute the full model. On this full model, we choose a reduced model by a two-step model selection procedure: First an AIC in a stepwise algorithm, and a stepwise backward elimination procedure on the resulting model [65]. We stop the elimination procedure when all the remaining variables are significant at $\alpha = 0.05$ level. That is, first, on the full model we apply the AIC criteria which drops a variable if there is a significant decrease in AIC, and stop when a significant reduction is not possible by removing a variable. On this reduced model, we apply the backward elimination procedure, since although a variable that is kept by AIC might not be significant by the t-test. Based on the final model with significant predictors, we apply logistic discrimination. In general, if the estimated probability is larger than a prespecified probability p_0, the subject is classified as CDR0.5, otherwise the subject is classified as CDR0 (i.e., healthy). This threshold probability p_0 can also be optimized with respect to a cost function which incorporates correct classification rates, sensitivity, and/or specificity [62].

3 Results

3.1 Metric Distances

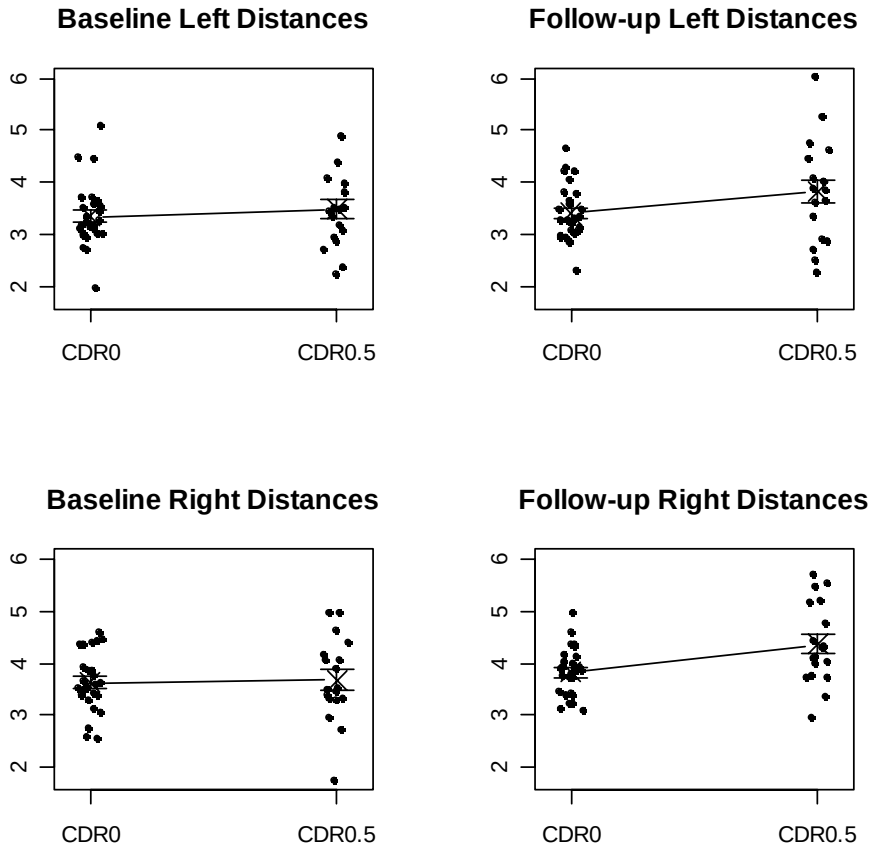


Figure 3 shows the jittered scatter plots of the metric distances by diagnosis group. Let LDB and LDF respectively denote the left metric distances at baseline and follow-up; RDB and RDF are likewise the right metric distances. One-sided and two-sided Wilcoxon tests show that LDB (3.40±0.68) < LDF (3.57±0.77) and RDB (3.65±0.67) < RDF (4.05±0.67) with inequalities being significant at .05 level.

To model metric distances with diagnosis group or side as main effect with compound symmetry in Var-Cov structure, see [62]. To determine, for example, if LB-CDR0 are different from LF-CDR0, we consider a model that

includes the interaction of diagnosis and timepoint. The most promising model for left metric distances is the following model with the compound symmetry in variance-covariance structure based on AIC and BIC.

$$d_{ijk}^{L} = \mu + \alpha_i^{D} + \alpha_j^{T} + \alpha_{ij}^{DT} + \varepsilon_{ijk}$$

where d_{ijk}^{L} is the left metric distance for subject k with diagnosis i at timepoint j, μ is the overall mean, α_i^{D} is the effect of diagnosis level i, α_j^{T} is the effect of timepoint level j=1,2, α_{ij}^{DT} is the diagnosis-by-timepoint interaction, and ε_{ijk} is the error term. The two way interaction of diagnosis-by-timepoint is not significant (F=2.35, df=1,84, p=0.1288). The main effect of diagnostic group is not significant either (F=1.87, df=1,84, p=0.1750). However, the main effect of timepoint is almost significant, i.e., indicates a trend (F=3.91, df=1,84, p=0.0515).

Likewise, the most promising model for right metric distances is the following model with the compound symmetry in variance-covariance structure based on AIC and BIC.

$$d_{ijk}^{R} = \mu + \alpha_i^{D} + \alpha_j^{T} + \alpha_{ij}^{DT} + \varepsilon_{ijk}$$

where d_{ijk}^{R} is the right metric distance for subject k, the other terms as before. The two way interaction of diagnosis-by-timepoint is significant (F=6.49, df=1,84, p=0.0127). The main effect of diagnostic group is not significant (F=2.88, df=1,84, p=0.0934), but the main effect of timepoint is significant (F=17.65, df=1,84, p=0.0001). Figure 4 shows the interaction plots for diagnosis levels over the timepoint levels for left and right metric distances. Although the main effect of group is not significant but main of effect of timepoint is significant , due to significant interaction, the results based on main effect for diagnosis (i.e., group) or timepoint will be misleading; i.e., the distances should be compared at each time point for differences between diagnostic groups.

We apply various post-hoc tests to see which groups are significantly different or significantly change over time and present the corresponding p-values in Error! Reference source not found.. We only use Wilcoxon rank sum test [66] as it requires less assumptions compared to the usual t-test. RF-CDR0.5 mean distances are significantly larger than RF-CDR0 mean distances at .05 level (p=0.0106, one-sided), but LF-CDR0.5 distances are almost significantly larger than LF-CDR0 distances (p=0.0813, one-sided) which is suggestive of a trend. On the other hand, LB-CDR0.5 and LB-CDR0 distances are not significantly different (p=0.6078, two-sided), and likewise for RB-CDR0.5 and RB-CDR0 distances (p=0.9239, two-sided).

For the comparison of dependent pairs such as LB-CDR0.5 and LF-CDR0.5, see also **Error! Reference source not found.**. Again we use Wilcoxon signed rank test (i.e., paired version of Wilcoxon's test). Observe that LB-CDR0.5 metric distances are significantly smaller than LF-CDR0.5 distances at $\alpha = .05$ (p=0.0155, one-sided). The same holds for RB-CDR0.5 vs RF-CDR0.5 metric distances (p=0.0002, one-sided). That is, the morphometric differences of CDR0.5 hippocampi from the template tend to increase by time, which implies that for both left and right distances there is significant change in morphometry of CDR0.5 hippocampi over time. In fact, significant volume reduction over time is detected [29]. The morphometric changes in CDR0.5 right hippocampi from baseline to follow-up is significantly larger than those of CDR0.5 left hippocampi (p = .0445, one-sided). The associated p-value here is obtained by testing the difference sets (LF-CDR0.5)-(LB-CDR0.5) versus (RF-CDR0)-(RB-CDR0.5) using the usual paired t-test. On the other hand, only RB-CDR0 is almost significantly less than RF-CDR0 (p = .0621, one-sided), which implies there is some weak evidence for mild change in healthy right hippocampi as a result of aging.

At $\alpha = 0.05$ level, the cdf of RF-CDR0.5 distances is significantly smaller than the cdf of RF-CDR0 distances (p=0.0259 for K-S test). That is, metric distances RF-CDR0.5 are stochastically larger than RF-CDR0 metric distances. Observe that these results are in agreement with **Error! Reference source not found.**.

Observe here that the post hoc tests suggest that the LDDMM metric distance being a compound measure is not sensitive enough to detect the mild change in morphometry due to DAT in its early period, but detects the morphometric change over time. The repeated-measures analysis, only indicates which variables significantly explain the morphometric changes, i.e., suggests which variables should be compared and how in the post hoc analysis.

3.2 Logistics Discrimination Analysis

In logistic discrimination with metric distances, after the model selection procedure, the resulting model is $\text{logit}\, p_k = \beta_0 + \beta_1 d_{ijk} + \beta_2 d_{ijk}^2$ where p_k is the probability of subject k having very mild dementia and d_{ijk} the distance for subject k with diagnosis i (i=1 for CDR0 and 2 for CDR0.5) at timepoint j (j=1 for baseline and 2 for follow-up), β_0

is the intercept and β_1 and β_2 are the coefficients of the linear and the quadratic terms, respectively. In this logistic discrimination procedure, we treat each data from left, right, baseline or follow-up as a distinct subject. This should potentially increase sensitivity at the expense of specificity, which might be desirable, given the consequences of misidentifying a demented subject as healthy. Each subject has four data (i.e., left and right metric distance and volume measures at baseline and followup), and when one of these is classified as CDR0.5, it would suffice to classify the subject as CDR0.5, while all four data should be classified as CDR0 for the subject to be classified as CDR0. Due to the significant group-by-timepoint interaction, we need to consider diagnosis groups at each time point. When we use d_{ik}^B and d_{ik}^F one at a time in a logistic model, we see that only the model $\text{logit}\, p_k = \beta_0 + \beta_1 d_{ik}^F + \beta_2 \left(d_{ik}^F\right)^2$ has significant coefficients for the distance terms. Using this model in the logistic classifier the optimal threshold value is $p_0 = 0.335$, the correct classification rate is 66%, sensitivity is 89%, and specificity is 50%. With $\text{logit}\, p_k = \beta_0 + \beta_1 d_k^{RF}$ in the logistic classifier the optimal threshold value is $p_0 = 0.38$, the correct classification rate is 70%, sensitivity is 72%, and specificity is 69%. For a more detailed discussion of logistic discrimination with various other models, see [62].

Using volumes obtained from Wang et al. [29], we apply the logistic discrimination using both volume and metric distance as predictors. The stepwise elimination procedure yields $\text{logit}\, p_k = \beta_0 + \beta_1 V_{ijk} + \beta_2 d_{ijk}^9 + \beta_3 V_{ijk} d_{ijk}$ where p_k is the probability of subject *k* having very mild dementia and V_{ijk} the volume and d_{ijk} the metric distance for subject *k* with diagnosis *i* (*i* = 1 for CDR0 and 2 for CDR .5) at timepoint *j* (*j* = 1 for baseline and 2 for follow-up), β_0 is the overall intercept, β_1 is the coefficient for volume, β_2 is the coefficient for nineth power of the metric distance, β_3 is the coefficient for the interaction between volume and distance. When we use baseline or follow-up measures one at a time in a logistic model, we see that the model $\text{logit}\, p_k = \beta_0 + \beta_1 V_{ik}^f + \beta_2 \left(d_{ik}^f\right)^5$ has the most significant coefficients. The optimal threshold value is $p_0 = 0.48$, and the correct classification rate, sensitivity, and specificity are 80%, 61%, and 92%, respectively. Due to significant side-by-timepoint interaction, we use side-by-timepoint combinations one at a time in a logistic model, and see that the following model has the best fit: $\text{logit}\, p_k = \beta_0 + \beta_1 \left(d_k^{rf}\right)^3 + \beta_2 V_k^{lf}$. The optimal threshold value is $p_0 = 0.48$, and the correct classification rate, sensitivity, and specificity are 84%, 83%, and 85%, respectively.

Notice that the logistic regression results with metric distances alone only provide about 70% correct classification of subjects into CDR0.5 and CDR0 categories, which is almost significantly better than chance (i.e., significantly larger than 50%) at .05 level (p=0.0896 one-sided). But the sensitivity measures are significantly better than the chance value (89 % (p=0.0064, one-sided) and 72% (p=0.0578, one-sided)). The specificity values are not significantly different from 50% (p=0.011 and 1). Notice that with volume and distance being used together, there is an increase in correct classification rate, sensitivity, and specificity values. However these increases are slightly significant. Classifying baseline and followup metric distances independently, and calling the result CDR0.5 even if the baseline and followup classifications are divergent is a reasonable approach, since it will increase the chances of detecting a demented subject.

4 Discussion and Conclusion

In this study, we used the Large Deformation Diffeomorphic Metric Mapping (LDDMM) algorithm to generate metric distances between hippocampi in groups of subjects with and without Dementia of Alzheimer type (DAT) in its mild form (labeled as CDR0.5 and CDR0 patients, respectively) at baseline and at follow-up. The subjects in this paper have been previously analyzed using related but different tools. As a single scalar measure, volumes were used for diagnosis group comparisons at baseline and follow-up [29] and displacement, momentum vector fields based on LDDMM were used for discrimination only at baseline [58]. But the metric distances computed from LDDMM has not hitherto been used in diagnosis group analysis. The metric distance gives a single number reflecting the global morphometry (i.e., the size and shape) while volume measurements only provide information on size. So metric distances provide morphometric information not conveyed by volume whereas momentum vector fields also provide local information on shape changes. Furthermore, the morphometric information conveyed by the metric distance depends on the choice of the template, while the morphometric information conveyed by momentum vector fields is independent of the template chosen. That is, although the vector fields change when the template changes, the morphometric information they convey is the same. The results might completely differ if another template was selected, due to the confounding

nature of size for the LDDMM metric distance. The metric distance will provide meaningful shape comparisons if the same template is used for new data as well. The choice of template that yields optimal discrimination results is a topic of ongoing research.

We show that metric differences at baseline are not different between AD and healthy controls. At followup, right-sided metric differences were larger in AD than in controls. Further, metric distances increased in demented subjects but not in controls (see Table 1). Hence we point a major strength of LDDMM distance, it is one single metric to quantify the shape, hence easy to interpret, present, and visualize. However, being a compound summary measure, some information about shape might be lost compared to the vector fields. Although it shows a reduction in distance due to the disease at early stages, this reduction is not statistically significant, so the technique is (statistically) unable to discriminate patients at the earliest stages of AD from controls, but as the disease and atrophy worsens, the technique is able to discriminate patients from controls.

Such differences in morphometry between diagnosis groups or morphometric changes over time can be detected by metric distances computed via LDDMM and could potentially serve as a biomarker for the disease. In regard to statistical analysis, as a compound but brief measure of morphometry, metric distances can thus serve as a first step to identify the morphometric differences, and can be used as a pointer to which direction more detailed analysis on the lines of momentum vector fields could go. The importance of the CDR0 versus CDR0.5 contrast analyzed here is that it tests a necessary but not sufficient condition for the eventual goal of discriminating CDR0 subjects who subsequently convert to CDR0.5, from CDR0 subjects who subsequently stay CDR0.

The morphometric changes in the left hippocampus in CDR0.5 subjects from baseline to follow-up are not significantly different from those in those in the CDR0 subjects ($p=0.134$, two-sided), while the morphometric changes in the right hippocampus in the CDR0.5 group from baseline to follow-up are significantly larger than those in the CDR0 group ($p=0.0070$, one-sided). However, inferences on asymmetry should not be made since metric distances were computed relative to different templates for the left and right hippocampus. Furthermore, metric distance results agree with the volume comparisons [29], hence volume (i.e., scale) might be highly dominating the morphometric changes. In other words, the significant volume reduction might dominate the change in shape, when morphometry is measured by metric distances.

Differences and changes (over time) in morphometry can also be used for diagnostic discrimination of subjects in non-demented (i.e., CDR0) or demented (i.e., CDR0.5) groups. In this study we applied logistic discrimination based on metric distances, as logistic regression not only provides a means for classification, but also yields a probability estimate for having very mild dementia. Furthermore, one can optimize the threshold probability for a particular cost function for the entire training data set, or by a cross-validation technique. The correct classification rate was about 70% in our logistic regression analysis. Metric distances can be used to distinguish AD from normal aging quantitatively; however, to be able to use it for diagnostic purposes, the method should be improved to a greater extent.

Although metric distances increase and volumes decrease by time, the metric distance, being an extremely condensed summary measure gives very similar results as volume. That is, neither volume nor metric distance discriminated left baseline, right baseline, or left followup between CDR0 and CDR0.5; volume reduction and metric distances differences are both significant in CDR0.5 subjects, but neither of them are significant in CDR0 subjects; and ANOVA suggested a significant diagnosis group-by-timepoint interaction for both measures [58, 62]. On the other hand, when volume and metric distances are used together, the classification results improve (correct classification rate is 84%) compared to results based on distance only.

Changes within a subject may produce a more sensitive measure of shape change. However, because the template is different for each longitudinal computation (i.e., each subject's own baseline), parallel transport analysis of the vector fields is required for statistical comparison of group change [70, 71], which is beyond the scope of this work. Also, alternative average template computed from the actual populations as in [72] may also result in more sensitive metric distances. Likewise, metric distances derived from surfaces rather than binary images via LDDMM-Surface [73] may yield better sensitivity. The advent of large scale multi-site neuroimaging studies of AD such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), European-ADNI, Japanese-ADNI and Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL) [74, 75] may provide a useful testbed for a more refined analysis of metric distances. Furthermore, we have not accounted for structural deformations in MRI data, which has likely reduced power. The effects of gradient warping of structural data is also topic of prospective research.

We provide evidence that Large Deformation Diffeomorphic Metric Mapping (LDDMM) can be used to explore structural changes in the hippocampus in AD. In conclusion, metric distances have a useful potential as a biomarker in detecting morphometric changes in the hippocampus in subjects with very mild dementia. Further theory and application are needed to see how metric distances combined with other biomarkers can be used in clinical diagnosis of dementia as well as neurological and psychiatric disorders.

Acknowledgements

We would like to acknowledge support from NIH (P01 AG03991, P50 AG05681, AG05684, P41-RR15241, P50 MH71616, MH 56584), NSF DMS-0456253, NSERC 31-611387 and the Pacific Alzheimer Research Foundation. The authors thank the Clinical Core of the the Alzheimer's Disease Research Center at Washington University for the clinical and cognitive assessments of the participants.

Tables

Comparisons of Metric Distances by Independent Tests			
	p-values for Wilcoxon test		
Groups	2-sided	$1^{st} < 2^{nd}$	$1^{st} > 2^{nd}$
LB-CDR0.5,LB-CDR0	.6078	.7044	.3039
LF-CDR0.5,LF-CDR0	.1625	.9223	.0813
RB-CDR0.5,RB-CDR0	.9239	.5475	.462
RF-CDR0.5,RF-CDR0	.0212*	.99	.0106*
Comparisons of LDDMM Distances by Dependent (i.e., Paired) Tests			
	paired Wilcoxon test		
Groups	2-sided	$1^{st} < 2^{nd}$	$1^{st} > 2^{nd}$
LB-CDR0.5,LF-CDR0.5	.0311*	.0155*	.9861
RB-CDR0.5,RF-CDR0.5	.0005*	.0002*	.9998
LB-CDR0,LF-CDR0	.7127	.3563	.6531
RB-CDR0,RF-CDR0	.1244	.0622	.9409

Table 1: The *p*-values based on independent Wilcoxon rank sum test (top) for both left and right metric distances and *p*-values based on paired Wilcoxon test (i.e., Wilcoxon signed rank test) for both left and right metric distances (bottom).

Figures

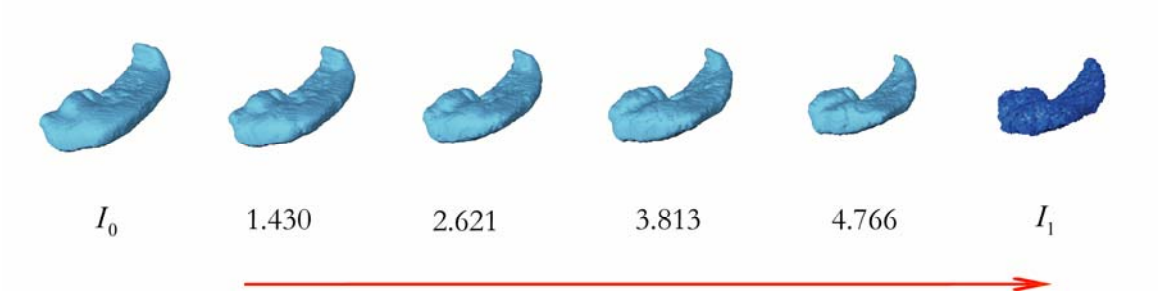


Figure 1: Change in metric distance during diffeomorphic map (g) from template (I_0) to target ($I_1 = gI_0$).

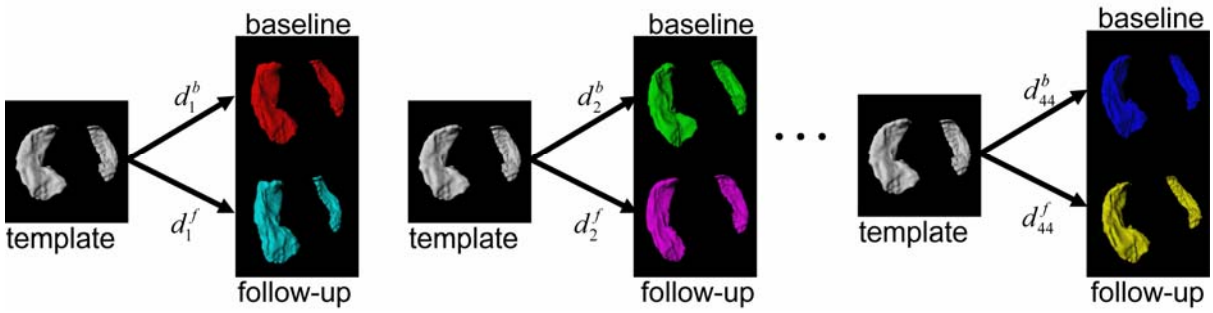


Figure 2: Generation of metric distances $d_k^{\{b,f\}}$ for subjects $k = 1,\ldots,44$ at baseline (*b*) and at follow-up (*f*).

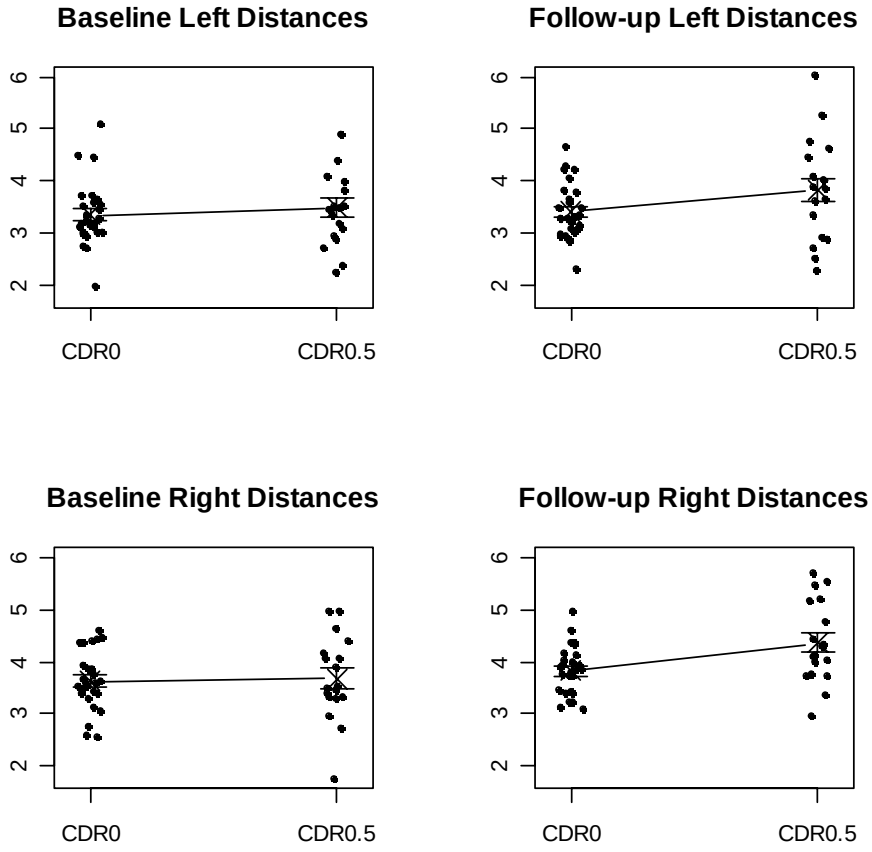


Figure 3: Scatter plots of the metric distances for the left and right distances at baseline and follow-up. The metric distances are jittered for better visualization and the crosses represent the mean distance values. Mean and SEM values by diagnosis group are as follows: LDB for CDR0 and CDR0.5 is 3.34 ± 0.62 and 3.48 ± 0.76, LDF for CDR0 and CDR0.5 is 3.40 ± 0.55 and 3.82 ± 0.98; RDB for CDR0 and CDR0.5 is 3.61 ± 0.57 and 3.68 ± 0.81, RDF for CDR0 and CDR0.5 is 3.87 ± 0.47 and 4.37 ± 0.78.

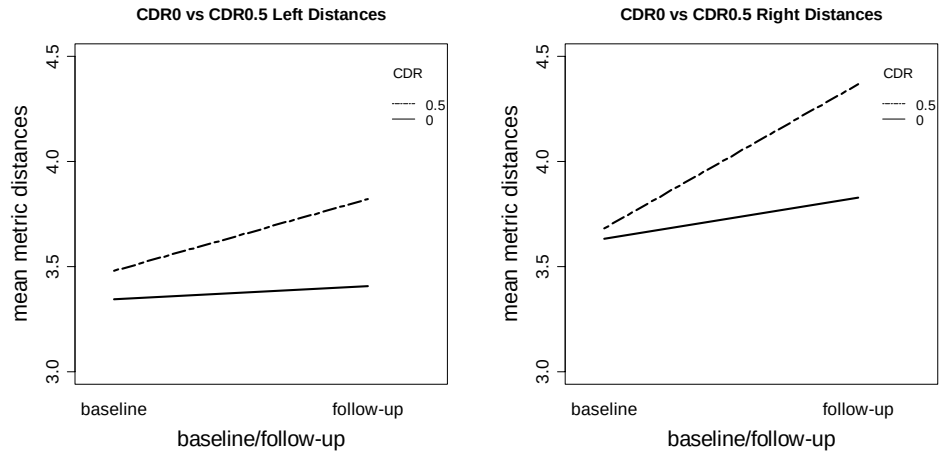


Figure 4: Interaction plots for diagnosis levels over the timepoint levels for left and right metric distances.

References

1 Davis, D.G., F.A. Schmitt, D.R. Wekstein, et al., *Alzheimer neuropathologic alterations in aged cognitively normal subjects.* Journal of Neuropathology and Experimental Neurology, 1999. **58**(4): p. 376-388.

2 Haroutunian, V., D.P. Perl, D.P. Purohit, et al., *Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer Disease.* Archives of Neurology, 1998. **55**(9): p. 1185-1191.

3 Thompson, P.M., M.S. Mega, R.P. Woods, et al., *Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas.* Cerebral Cortex, 2001. **11**(1): p. 1-16.

4 Troncoso, J.C., L.J. Martin, G. Dal Forno, et al., *Neuropathology in controls and demented subjects from the Baltimore longitudinal study of aging.* Neurobiology of Aging, 1996. **17**(3): p. 365-371.

5 Braak, H. and E. Braak, *Staging of Alzheimer's disease-related neurofibrillary changes.* Neurobiol Aging, 1995. **16**(3): p. 271-8; discussion 278-84.

6 Braak, H. and E. Braak, *Staging of Alzheimer-related cortical destruction.* Int Psychogeriatr, 1997. **9 Suppl 1**: p. 257-61; discussion 269-72.

7 Braak, H., E. Braak, and J. Bohl, *Staging of Alzheimer-related cortical destruction.* Eur Neurol, 1993. **33**(6): p. 403-8.

8 Price, J.L., A.I. Ko, M.J. Wade, et al., *Neuron number in the entorhinal cortex and CA1 in preclinical alzheimer disease.* Archives of Neurology, 2001. **58**(9): p. 1395-1402.

9 Convit, A., M.J. de Leon, J. Golomb, et al., *Hippocampal atrophy in early Alzheimer's disease: anatomic specificity and validation.* Psychiatric Quarterly, 1993. **64**(4): p. 371-387.

10 Csernansky, J.G., L. Wang, S. Joshi, et al., *Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus.* Neurology, 2000. **55**(11): p. 1636-1643.

11 Krasuski, J.S., G.E. Alexander, B. Horwitz, et al., *Relation of medial temporal lobe volumes to age and memory function in nondemented adults with Down's syndrome: Implications for the prodromal phase of Alzheimer's disease.* American Journal of Psychiatry, 2002. **159**(1): p. 74-81.

12 Mega, M.S., G.W. Small, M.L. Xu, et al., *Hippocampal atrophy in persons with age-associated memory impairment: Volumetry within a common space.* Psychosomatic Medicine, 2002. **64**(3): p. 487-492.

13 Mu, Q., J. Xie, Z. Wen, et al., *A quantitative MR study of the hippocampal formation, the amygdala, and the temporal horn of the lateral ventricle in healthy subjects 40 to 90 years of age.* American Journal of Neuroradiology, 1999. **20**(2): p. 207-211.

14 Scheltens, P. and F. Barkhof, *Structural neuroimaging outcomes in clinical dementia trials, with special reference to disease modifying designs.* Journal of Nutrition, Health and Aging, 2006. **10**(2): p. 123-128.

15 Gosche, K.M., J.A. Mortimer, C.D. Smith, et al., *Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun study.* Neurology, 2002. **58**(10): p. 1476-1482.

16 Fox, N.C., W.R. Crum, R.I. Scahill, et al., *Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images.* Lancet, 2001. **358**(9277): p. 201-205.

17 Chan, D., N.C. Fox, R. Jenkins, et al., *Rates of global and regional cerebral atrophy in AD and frontotemporal dementia.* Neurology, 2001. **57**(10): p. 1756-1763.

18 Christensen, G.E., R.D. Rabbitt, and M.I. Miller, *Deformable templates using large deformation kinematics.* IEEE Transactions on Image Processing, 1996. **5**(10): p. 1435-1447.

19 Miller, M.I., G.E. Christensen, Y. Amit, et al., *Mathematical textbook of deformable neuroanatomies.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(24): p. 11944-11948.
20 Hogan, R.E., L. Wang, M.E. Bertrand, et al., *MRI-based high-dimensional hippocampal mapping in mesial temporal lobe epilepsy.* Brain, 2004. **127**(8): p. 1731-1740.
21 Miller, M.I., *Computational anatomy: Shape, growth, and atrophy comparison via diffeomorphisms.* NeuroImage, 2004. **23**: p. S19-S33.
22 Thompson, P.M., K.M. Hayashi, E.R. Sowell, et al., *Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia.* Neuroimage, 2004. **23**: p. S2-S18.
23 Grenander, U. and M.I. Miller, *Computational anatomy: An emerging discipline.* Quarterly of Applied Mathematics, 1998. **56**(4): p. 617-694.
24 Toga, A.W. and P.M. Thompson, *Brain atlases of normal and diseased populations.* Int Rev Neurobiol., 2005. **66**: p. 1-54.
25 Toga, A.W., *Computational biology for visualization of brain structure.* Anatomy and Embryology, 2005. **210**(5-6): p. 433-438.
26 Grenander, U., *General Pattern Theory*. 1993, Oxford: Clarendon Press.
27 Grenander, U. and M.I. Miller, *Representations of knowledge in complex systems.* J. R. Statist. Soc. B, 1994. **56**(3): p. 549-603.
28 Wang, L., S.C. Joshi, M.I. Miller, et al., *Statistical analysis of hippocampal asymmetry in schizophrenia.* NeuroImage, 2001. **14**(3): p. 531-545.
29 Wang, L., J.S. Swank, I.E. Glick, et al., *Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging.* NeuroImage, 2003. **20**(2): p. 667-682.
30 Fox, N.C. and P.A. Freeborough, *Brain atrophy progression measured from registered serial MRI: Validation and application to Alzheimer's disease.* Journal of Magnetic Resonance Imaging, 1997. **7**(6): p. 1069-1075.
31 Fox, N.C., P.A. Freeborough, and M.N. Rossor, *Visualisation and quantification of rates of atrophy in Alzheimer's disease.* Lancet, 1996. **348**(9020): p. 94-97.
32 Killiany, R.J., B.T. Hyman, T. Gomez-Isla, et al., *MRI measures of entorhinal cortex vs hippocampus in preclinical AD.* Neurology, 2002. **58**(8): p. 1188-1196.
33 Wang, D., J.B. Chalk, S.E. Rose, et al., *MR image-based measurement of rates of change in volumes of brain structures. Part II: Application to a study of Alzheimer's disease and normal aging.* Magnetic Resonance Imaging, 2002. **20**(1): p. 41-48.
34 Yamaguchi, S., K. Meguro, M. Shimada, et al., *Five-year retrospective changes in hippocampal atrophy and cognitive screening test performances in very mild Alzheimer's disease: The Tajiri project.* Neuroradiology, 2002. **44**(1): p. 43-48.
35 Crum, W.R., R.I. Scahill, and N.C. Fox, *Automated hippocampal segmentation by regional fluid registration of serial MRI: Validation and application in Alzheimer's disease.* NeuroImage, 2001. **13**(5): p. 847-855.
36 Leow, A.D., A.D. Klunder, C.R. Jack, et al., *Longitudinal stability of MRI for mapping brain change using tensor-based morphometry.* Neuroimage, 2006. **31**(2): p. 627-640.
37 Apostolova, L.G., R.A. Dutton, I.D. Dinov, et al., *Conversion of mild cognitive impairment to alzheimer disease predicted by hippocampal atrophy maps.* Archives of Neurology, 2006. **63**(5): p. 693-699.
38 Mungas, D., D. Harvey, B.R. Reed, et al., *Longitudinal volumetric MRI change and rate of cognitive decline.* Neurology, 2005. **65**(4): p. 565-571.
39 Dickerson, B.C. and R.A. Sperling, *Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease.* NeuroRx, 2005. **2**(2): p. 348-360.
40 Ewers, M., S.J. Teipel, and H. Hampel, *Update of structural MRI-based methods for the early detection of Alzheimer's disease*
[Aktuelle entwicklungen der strukturellen MRT zur fru?hdiagnostik der Alzheimer-demenz]. Nervenheilkunde, 2005. **24**(2): p. 113-119.
41 Barnes, J., R.I. Scahill, J.M. Schott, et al., *Does Alzheimer's disease affect hippocampal asymmetry? Evidence from a cross-sectional and longitudinal volumetric MRI study.* Dementia and Geriatric Cognitive Disorders, 2005. **19**(5-6): p. 338-344.
42 Beg, M.F., M.I. Miller, A. Trouve, et al., *Computing large deformation metric mappings via geodesic flows of diffeomorphisms.* International Journal of Computer Vision, 2005. **61**(2): p. 139-157.
43 Miller, M.I., A. Trouve, and L. Younes, *On the metrics and Euler-Lagrange equations of computational anatomy.* Annual Review of Biomedical Engineering, 2002. **4**: p. 375-405.
44 Glaunès, J., A. Qiu, M.I. Miller, et al., *Large Deformation Diffeomorphic Curve Matching.* International Journal of Computer Vision (online), 2008.
45 Trosset, M., C.E. Priebe, Y. Park, et al., *Semisupervised learning from dissimilarity data.* Computational Statistics and Data Analysis, 2008. **52**: p. 4643-4657.

46 Thompson, D.W., *On Growth and Form.* . Canto edition (1st ed., 1917) ed, ed. J.T. Bonner. 1992, Cambridge, UK: Cambridge University Press.
47 Miller, M.I., C.E. Priebe, A. Qiu, et al., *Collaborative Computational Anatomy: An MRI Morphometry Study of the Human Brain via Diffeomorphic Metric Mapping.* Human Brain Mapping, 2008.
48 Wang, L., J.P. Miller, M.H. Gado, et al., *Abnormalities of hippocampal surface structure in very mild dementia of the Alzheimer type.* NeuroImage, 2006. **30**(1): p. 52-60.
49 Morris, J.C., *The Clinical Dementia Rating (CDR): Current version and scoring rules.* Neurology, 1993. **43**(11): p. 2412-2414.
50 Cohen, J., *A coefficient for agreement for nominal scales.* Educational and Psychological Measurement, 1960. **20**: p. 37-46.
51 Burke, W.J., J.P. Miller, E.H. Rubin, et al., *Reliability of the Washington University Clinical Dementia Rating.* Archives of Neurology, 1988. **45**(1): p. 31-32.
52 Morris, J.C., C. Ernesto, K. Schafer, et al., *Clinical dementia rating training and reliability in multicenter studies: The Alzheimer's Disease Cooperative Study experience.* Neurology, 1997. **48**(6): p. 1508-1510.
53 Berg, L., D.W. McKeel Jr, J.P. Miller, et al., *Clinicopathologic studies in cognitively healthy aging and Alzheimer disease: Relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype.* Archives of Neurology, 1998. **55**(3): p. 326-335.
54 Morris, J.C., M. Storandt, D.W. McKeel Jr, et al., *Cerebral amyloid deposition and diffuse plaques in "normal" aging: Evidence for presymptomatic and very mild Alzheimer's disease.* Neurology, 1996. **46**(3): p. 707-719.
55 Price, J.L. and J.C. Morris, *Tangles and plaques in nondemented aging and 'preclinical' Alzheimer's disease.* Annals of Neurology, 1999. **45**(3): p. 358-368.
56 Petersen, R.C., R. Doody, A. Kurz, et al., *Current concepts in mild cognitive impairment.* Archives of Neurology, 2001. **58**(12): p. 1985-1992.
57 Storandt, M., E.A. Grant, J.P. Miller, et al., *Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI.* Neurology, 2006. **67**(3): p. 467-73.
58 Wang, L., M.F. Beg, J.T. Ratnanather, et al., *Large Deformation Diffeomorphism and Momentum Based Hippocampal Shape Discrimination in Dementia of the Alzheimer Type.* IEEE Trans. Medical Imaging, 2007. **26**: p. 462-470.
59 Haller, J.W., A. Banerjee, G.E. Christensen, et al., *Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas.* Radiology, 1997. **202**(2): p. 504-510.
60 Trouve, A., *Diffeomorphisms groups and pattern matching in image analysis.* International Journal of Computer Vision, 1998. **28**(3): p. 213-221.
61 Box, G.E.P., G.M. Jenkins, and G.C. Reinsel, *Time Series Analysis: Forecasting and Control*. 3rd ed. 1994: Holden-Day.
62 Ceyhan, E., C. Ceritoglu, M.F. Beg, et al., *Technical Report # KU-EC-08-3: Analysis of Metric Distances and Volumes of Hippocampi Indicates Different Morphometric Changes over Time in Dementia of Alzheimer Type and Nondemented Subjects*. 2008, Koç University.
63 Littel, R.C., G.A. Milliken, W.W. Stroup, et al., *SAS Systems for Mixed Models*. 1996: SAS Institute.
64 Venables, W.N. and B.D. Ripley, *Modern Applied Statistics with S-PLUS*. 2nd ed. 1997, New York: Springer Verlag.
65 Burnham, K.P. and D. Anderson, *Model Selection and Multi-Model Inference*. 2003, New York: Springer.
66 Conover, W.J., *Practical Nonparametric Statistics*. 3rd ed. 1999, New York: John Wiley & Sons.
67 Zar, J.H., *Biostatistical Analysis*. 1984, New Jersey: Prentice Hall. 718.
68 Dalgaard, P., *Introductory Statistics with R*. 2002: Springer-Verlag.
69 Johnson, D.E., *Applied Multivariate Methods for Data Analysis*. 1998, California: Duxbury Press.
70 Younes, L., *Jacobi fields in groups of diffeomorphisms and applications.* Quart. Appl. Math., 2007. **65**: p. 113-134.
71 Qiu, A., L. Younes, M.I. Miller, et al., *Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type.* Neuroimage, 2008. **40**(1): p. 68-76.
72 Ma, J., M.I. Miller, A. Trouvé, et al., *Bayesian Template Estimation in Computational Anatomy.* NeuroImage, 2008. **42**: p. 252-261.
73 Vaillant, M., A. Qiu, J. Glaunes, et al., *Diffeomorphic Metric Surface Mapping in Subregions of the Superior Temporal Gyrus.* Neuroimage, 2007. **34**: p. 1149-59.
74 Butcher, J., *Alzheimer's researchers open the doors to data sharing* The Lancet Neurology 2008. **6**: p. 480-481.
75 Jack Jr, C.R., M.A. Bernstein, N.C. Fox, et al., *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* J Magn Reson Imaging, 2008. **27**: p. 685-691.