Text S1:

Primary measures of the healthy cohorts

Fig. S1.1, S1.2, and S1.3 show the MRI derived measures of average cortical thickness, total grey matter surface area, and exposed grey matter surface area, respectively. As in the main text, we show each hemisphere as a separate data point. We show the data separately for both genders together (first column in each figure), females (second column in each figure), and males (third column in each figure). All measures are also shown dependent on age, to ensure that any systematic changes between the datasets are not due to age or gender effects.

Systematic and significant changes in offset can be seen between all three healthy cohorts (NKI, OASIS, HCP) in their average cortical thickness.

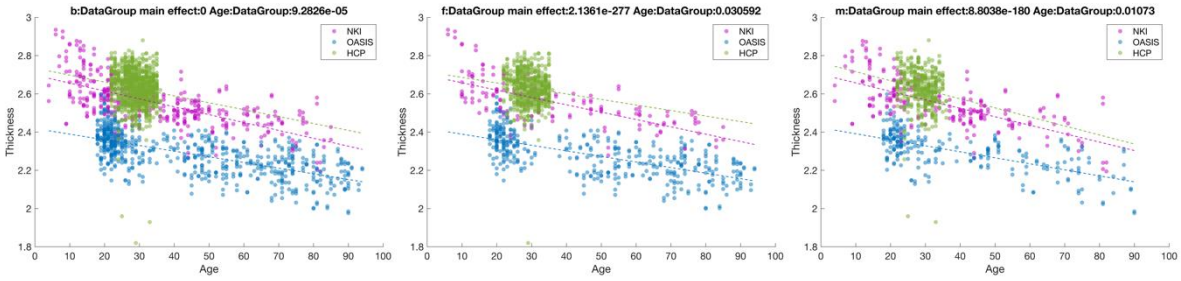


Fig. S1.1: Average cortical thickness for all three health cohorts shown over age for both genders, females, and males, respectively.

No systematic changes in offset can be seen visually between all three healthy cohorts (NKI, OASIS, HCP) in their total grey matter surface area, although statistically there may be a marginally significant difference in offset.

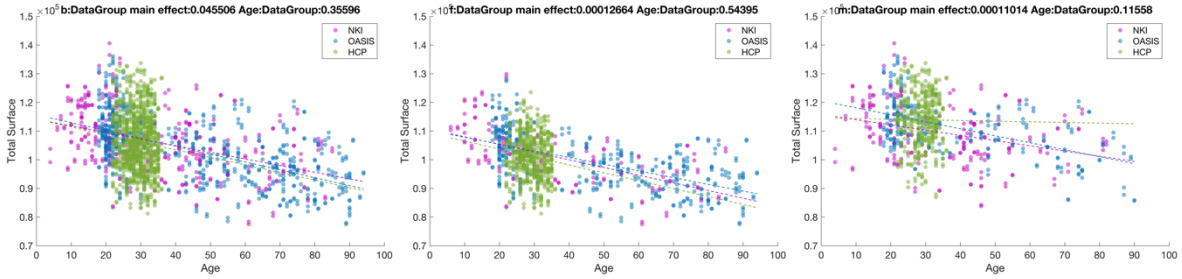


Fig. S1.2: Total grey matter surface area for all three health cohorts shown over age for both genders, females, and males, respectively.

No systematic changes in offset can be seen visually between all three healthy cohorts (NKI, OASIS, HCP) in their exposed grey matter surface area, although statistically there may be a marginally significant difference in offset.

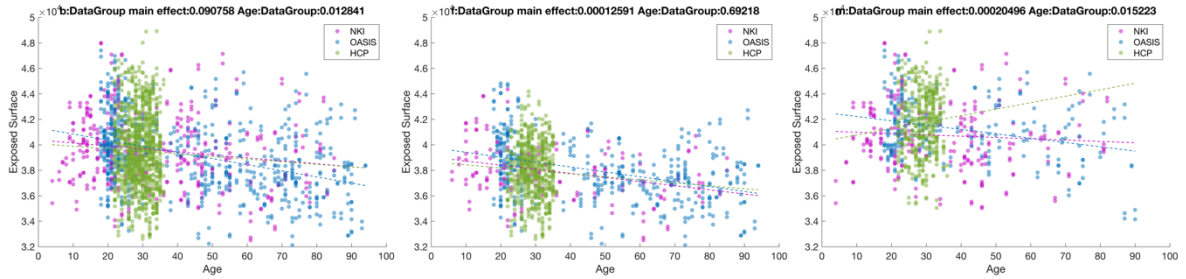


Fig. S1.3: Exposed grey matter surface area for all three health cohorts shown over age for both genders, females, and males, respectively.

Overall, it is also worth noting that the average cortical thickness decreases notably with age in a, what appears to be, linear fashion. The rate of this decrease is fairly consistent between all three cohorts (which is surprising given that the HCP dataset only covers a small age range, but further supports the notion of a linear decrease). The total and exposed grey matter surface areas also show a slight trend of decrease over age, which again appears fairly consistent across cohorts.

Text S2:

Gender differences in scaling behaviour

Gender differences in scaling relationship were investigated by subdividing each dataset into separate age groups and testing the following linear model y ~ x + Gender + x*Gender, where x= log10(A_E) and y = log10(sqrt(Thickness)*A_T). The age groups were chosen to be decades of life starting with 5 years. (I.e. 5-14 year olds are grouped together, 15-24 year olds are grouped together, etc.) This is to ensure a reasonable sample size for each age group. Finally, we set the significance level at 5% (i.e. $p<0.05$ we call significant).

HCP. No significant main or interaction effects were identified in HCP data for the following age groups: 15-24, 25-35. As the HCP data has a very high sample size in this age range, we further tested the original age groups as suggested by the HCP database: 21-25, 26-30, 31-35. Also here, we found no significant main or interaction effects.

NKI. No significant main or interaction effects were identified in NKI data for the following age groups: 0-15, 45-54, 65-74, 75-85.

A significant main (p=0.00009) and interaction effect (p=0.014) with gender was identified in the 16-24 age group. However, a t-test revealed a significant difference in male/female age distribution within this group (male, female mean= 20.9,19.7 p=0.029, cohen's d=0.482). This may explain these effects, as age is known to have an effect on the scaling relationship.

A significant interaction effect was identified in the 35-44 group (p=0.016). Again, a t-test revealed a significant difference between the age distribution of males and females within this age group (male, female mean=41.2, 38.4, p=0.001, cohen's d=1.24).

A significant main gender effect was identified in the 55-64 group (p=0.013). This age group contained a low number of data points (Male N=12, Female N=14) which may explain this discrepancy (Fig. S2.1).

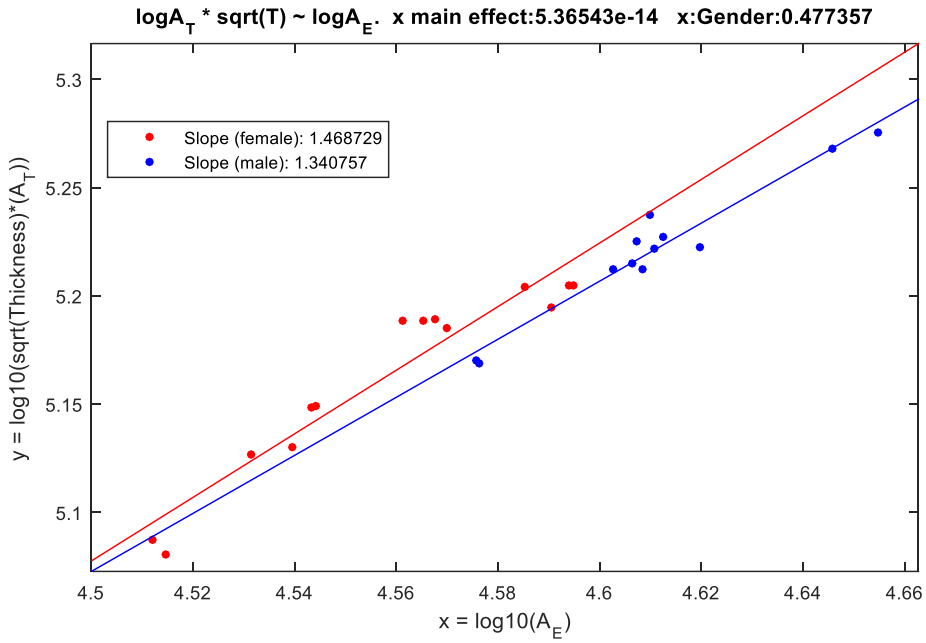


Fig. S2.1: NKI 55-64 age group

OASIS. No significant main or interaction effects were identified in NKI data for the following age groups: 16-24, 25-34, 55-64 ,65-74, 75-84, 85+.

A significant main gender effect was found in the 16-24 group (p=0.03). This effect was marginal and we shall discuss if perhaps the significance level has to be adjusted due to multiple comparisons later.

A significant interaction effect was found in the 35-44 group (p=0.0001). This again may be due to the low and disproportionate sample size. Also there may be some outliers distorting the regression.

When we only tested the regression in a range where both datasets have datapoints this interaction effect was no longer present (essentially removing 3 female subjects and 1 male subject, see arrows in Fig S2.2).

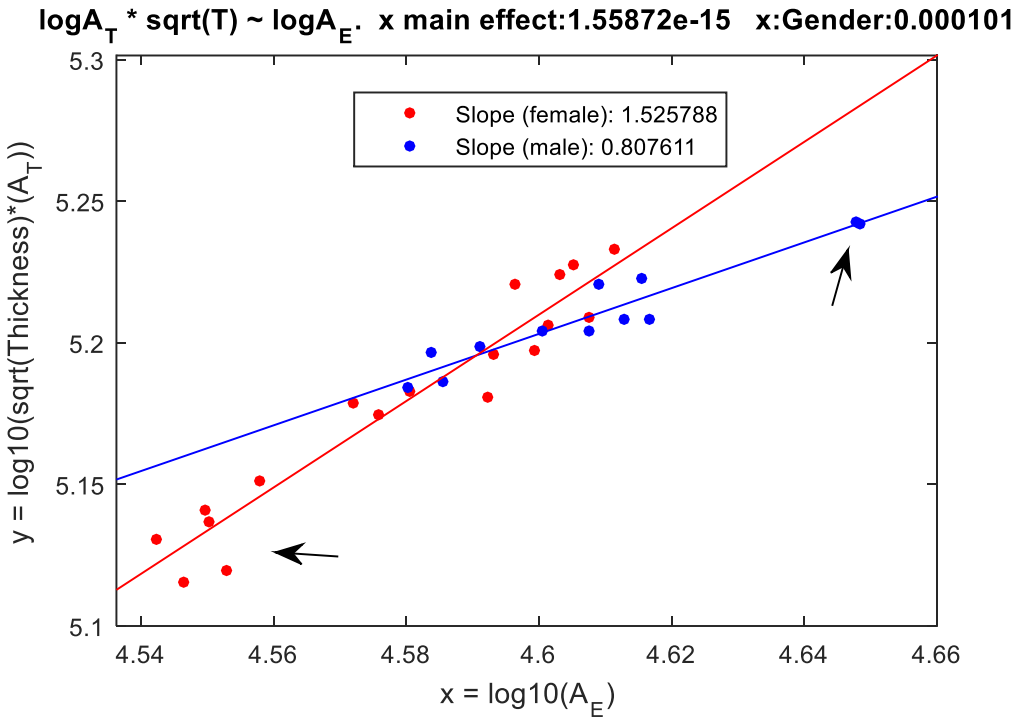


Fig. S2.2: OASIS 35-44 group.

A significant main effect was found in the 45-54 group (p=0.00006). In this age group the number of female subjects outweighed the number of males (M=24, F=52). A large number of females aged 54 were included in this age range. Following narrowing of the age band to 45-53 in order to exclude these individuals and random sample matching, this main gender effect remained significant (p=0.0001). We further tested the age group 40-49 and 50-59, but found no significant main or interaction effect except for a main effect (p=0.04) in the 40-49 group. Hence it is inconclusive if the effect in the 45-54 group is genuine.

Summary. Most age groups in all three healthy cohorts did not exhibit main or interaction effects with gender supporting the idea that there is no difference in the scaling relationship between gender types. In some cases, effects were identified, however, in most of those case these could be attributed to differences in the age distributions, outliers, or low sample sizes. Next, we will perform a multiple linear regression with age as continuous variable to validate our findings in this section.

Text S2.1: Age as a continuous regressor

To further confirm that there is no main or interaction effect in the scaling law with gender, we also tested all three healthy cohorts (NKI, OASIS, HCP) with age as a continuous regressor (rather than the analysis above, where we divided subjects into age groups). The effect of on the scaling relationship was examined by including age

as a continuous regressor in the linear model: y ~ x + Gender + Age + x*Gender + x*Age + x*Age*Gender, where x= log10(A_E) and y = log10(sqrt(Thickness)*A_T).

HCP. A significant main effect of age was found (p<0.001). No interaction effects with age were found. **No main or interaction effects with gender** were found.

NKI. A significant main effect of age was found (p<<0.001). No interaction effects with age were found. A significant main gender effect was found (p=0.0008); although this was not due to a significant difference in mean age (t-test, p=0.5). However, the age distributions show a large proportion of males in the age range 40-50, and there were more male than female subjects, which may have led to this effect (Fig. S2.3). No x:gender interaction effect was identified (p=0.097)

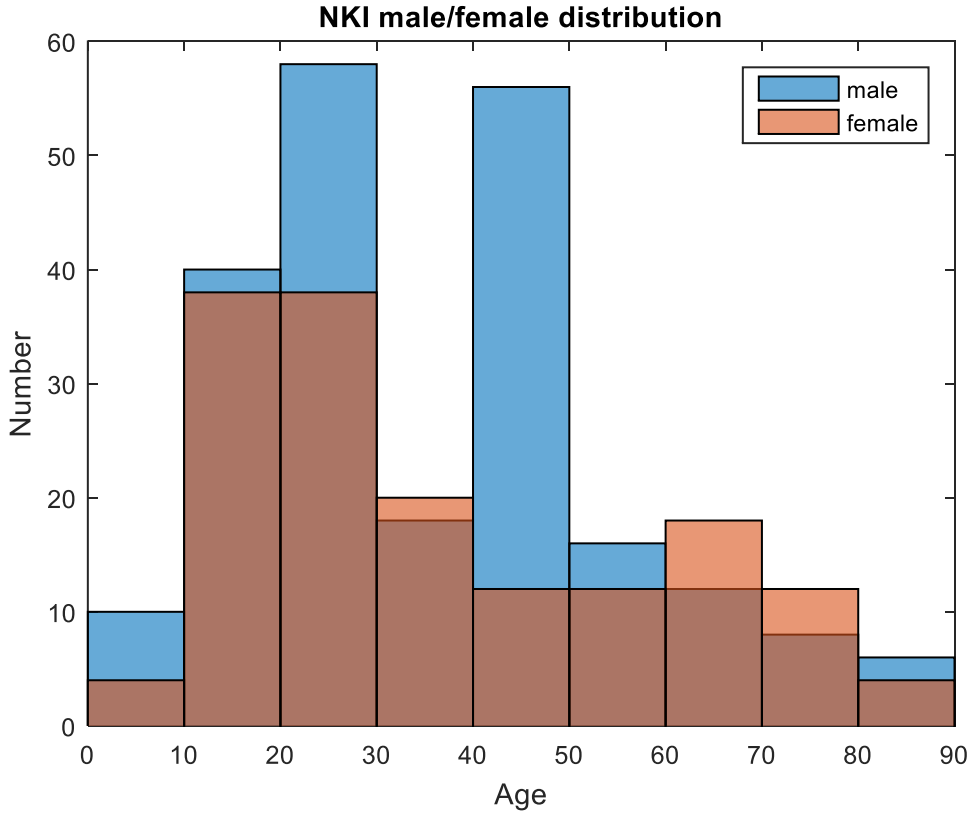


Fig. S2.3: Age distribution of males and females in the NKI dataset.

OASIS. A significant main effect of age was found (p<0.001). A slightly significant interaction effect with age was found (p=0.03); it is unclear what the cause of this may be, we shall discuss if perhaps the significance level has to be adjusted due to multiple comparisons later. **No main or interaction effects with gender** were found.

We also tested if the slope of the scaling law was significantly different to 1.25 when including age as a continuous regressor. At 95% confidence intervals, the slope was not different to 1.25 in any of the three datasets.

In conclusion, although some age groups in some cohorts showed some significant effects, we could mostly demonstrate that it was due to disproportionate age distributions between the male and female groups. The multiple linear regression (MLR) using age as continuous variable mostly confirms that there is no main of interaction effect with gender. Another point worth noting is that most of the p-values for “significant effects” are fairly high (between 0.05 and 0.0008). By the argument of multiple comparisons, the significance level should be adjusted to 0.05/81=0.0006 (with over 20 regressions, and 3 p-values each, and 21 p-values for the continuous age MLR), and these p-values should be discarded as not significant. Overall, we conclude that male and female brains follow exactly the same scaling law (without gender main or interaction effects) in all three cohorts.

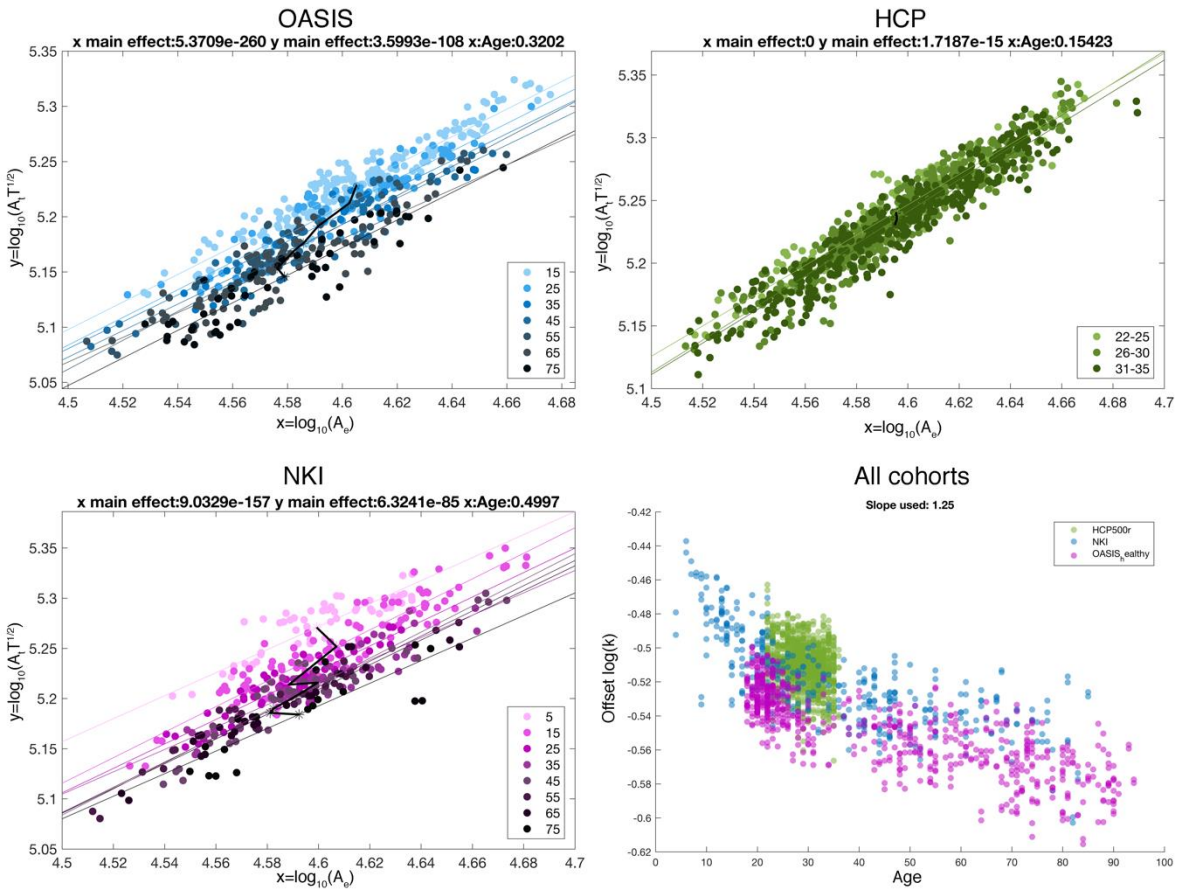


Fig. S3: Scaling law for all three datasets throughout age. Scaling law for different age groups in the OASIS, HCP & NKI dataset using both males and females. Regression lines are show in the corresponding colour as the age group. The age group average is shown as a black solid line. Bottom right: Using the predicted slope of 1.25, the estimated offset is shown for all three datasets over age (see Methods for details).

Text S4:

NKI 35-45 slope estimate

The scaling law slope estimate for the NKI 35-45 age group fell short of the predicted value of 1.25. Using the regress function in Matlab, the slope was estimated to be 1.1061 with a 95% CI ranging from 0.9821 - 1.2301.

A t-test revealed a significant difference between the age distribution of males and females within this age group (p=0.001, Cohen's d=1.24). The age distribution within this group reveals a large number of older males relative to females (Fig. S4.1).

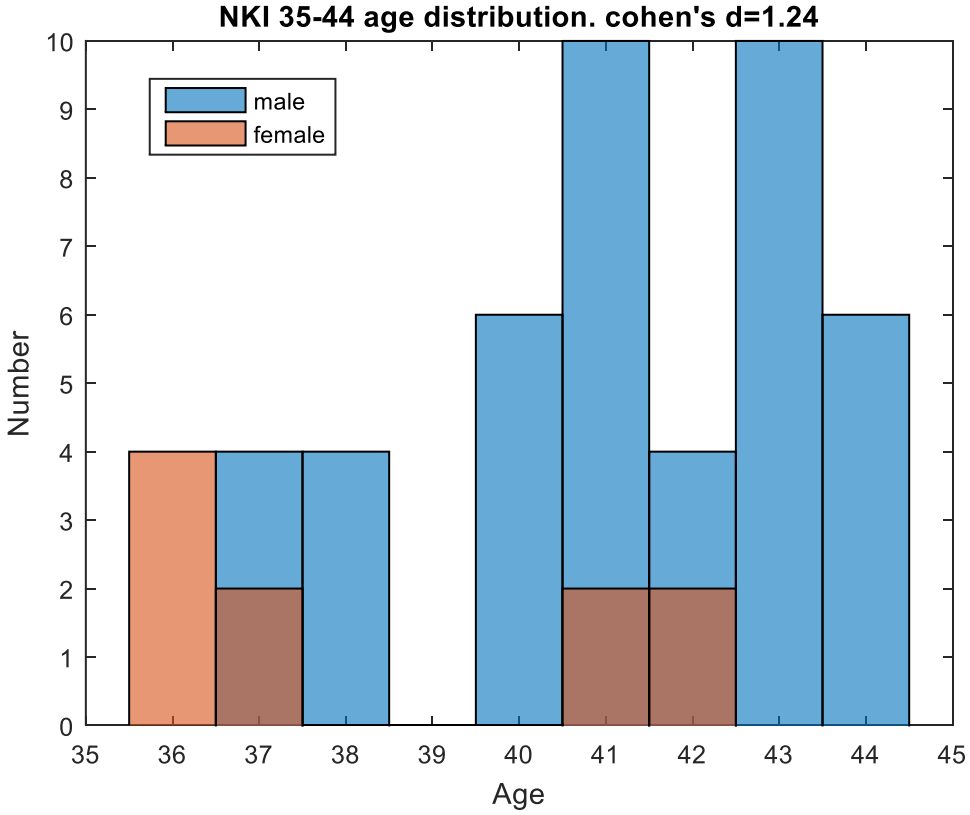


Fig. S4.1: NKI 35-45 age distribution.

When the age boundaries are narrowed to 35-42 in order to remove these older males, a regression analysis estimates the scaling law slope to be 1.1596 with a 95% CI ranging from 1.0207-1.2985. This encompasses the value 1.25 as predicted by the proposed scaling law. Therefore, it seems that this spurious result may have been due to an imbalanced age distribution within the sampling group.

Text S5:

Gender differences in ADNI data.

ADNI Alzheimer's patients

No main gender or x:gender interactions effects were seen in the age groups 55-64 and 85-91.
A significant x:gender interaction was identified within the 65-74 age group (p=0.0002, Fig. S5.1). This effect was probably not due to different age distributions between the male and female group (p=0.258 Cohen's d=0.208).

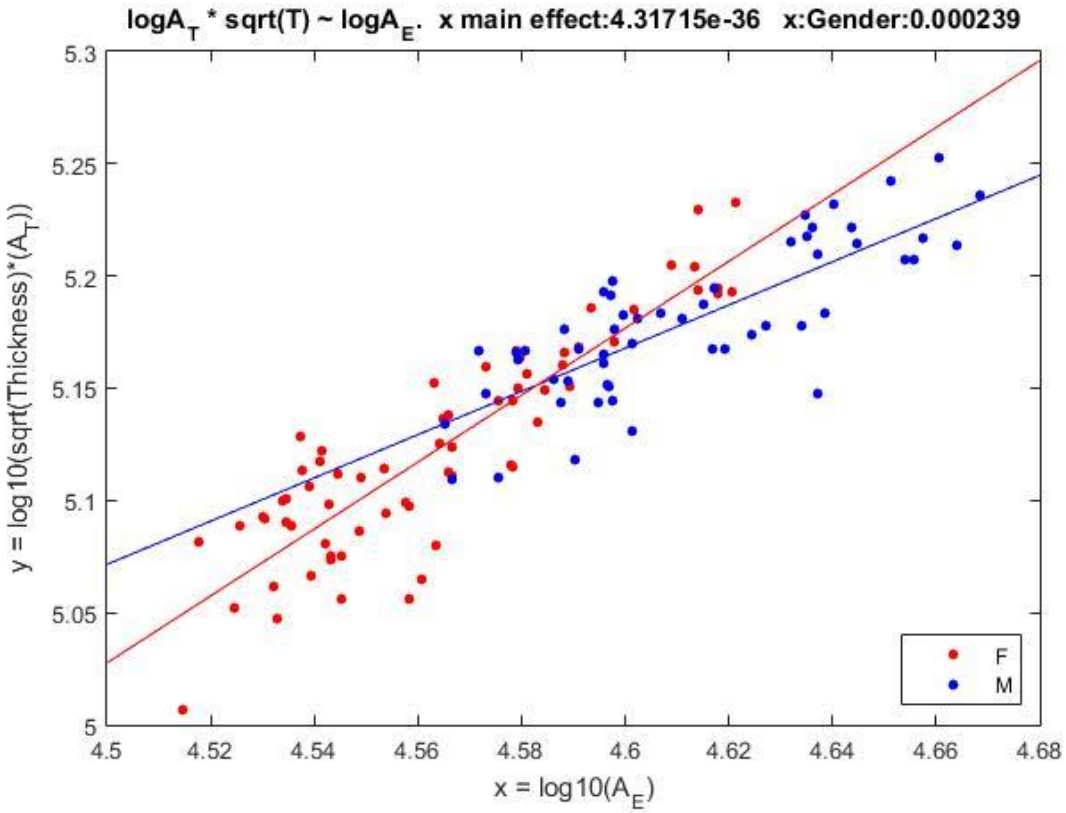


Fig. S5.1: ADNI Alzheimer's patients age 65-74, comparing male and female scaling.

A significant main gender effect (p=0.046) and trend towards significant x:gender interaction effect (p=0.061) were identified within the 75-84 age group (Fig. S5.2). This effect may be caused by a slightly different age distribution between the male and female group (p=0.0194 Cohen's d=0.3814). Hence we narrowed down the age groups and split the group into 74-79: main effect (p=0.002), x:gender (p=0.2); and 80-84, main (p=0.113), x:gender (p=0.041). However, there is still indication that the male and female groups are possibly different.

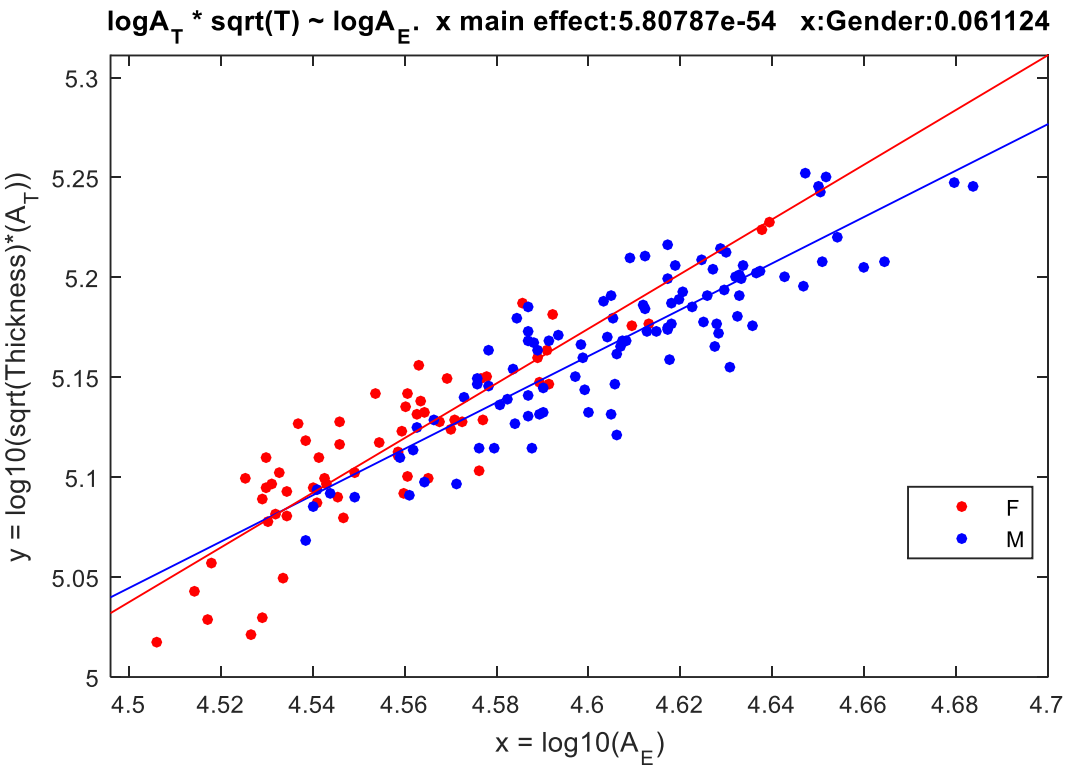


Fig. S5.2: ADNI Alzheimer's patients age 75-84.

For the ADNI Alzheimer's patients, we conclude that there is possibly a difference in scaling between males and females. Two out of four age groups showed significant differences in scaling, and the remaining two age groups had relatively low sample sizes. Hence, to ensure that we are not confusing gender effects, we proceeded to treat males and females separately in the main manuscript.

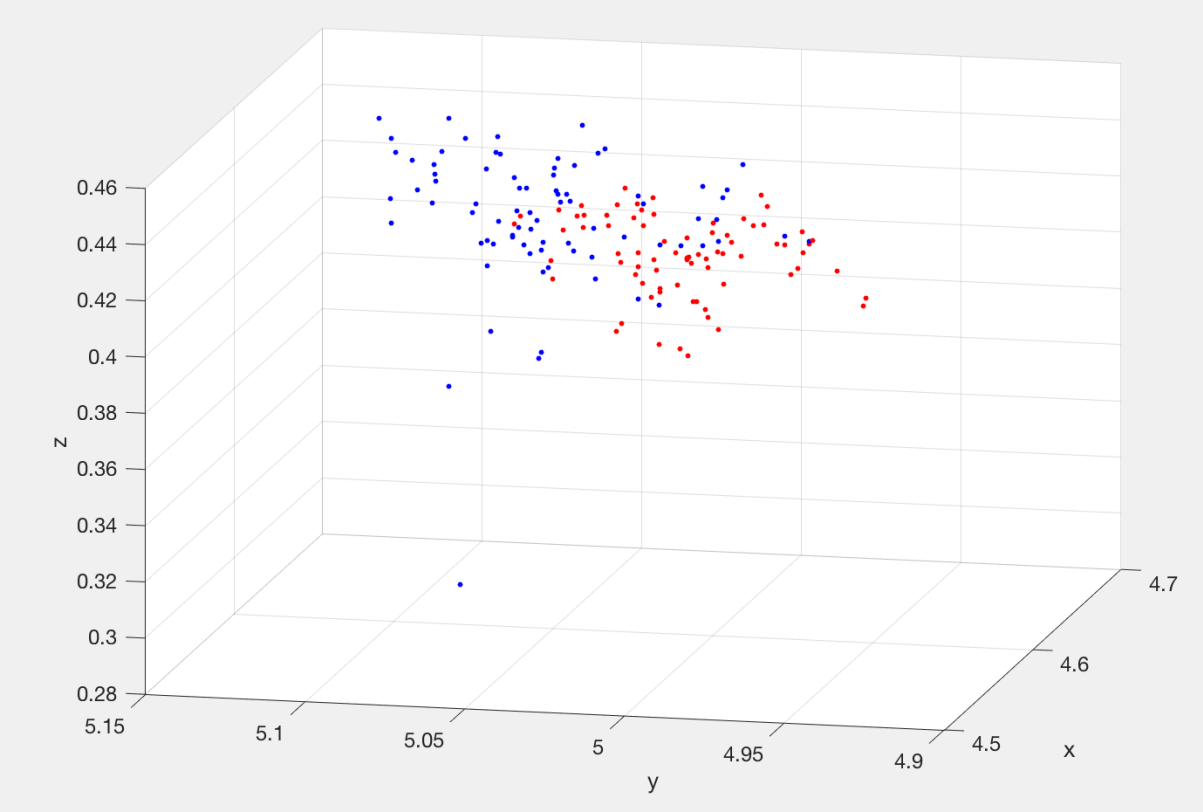
ADNI control data

No significant gender effects or x:gender interaction effects were present in the ADNI control data for the following age groups 55-64, 65-74, 75-84, 85-95.

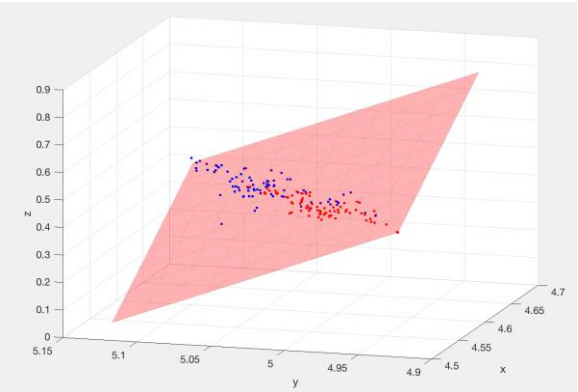
Text S6:

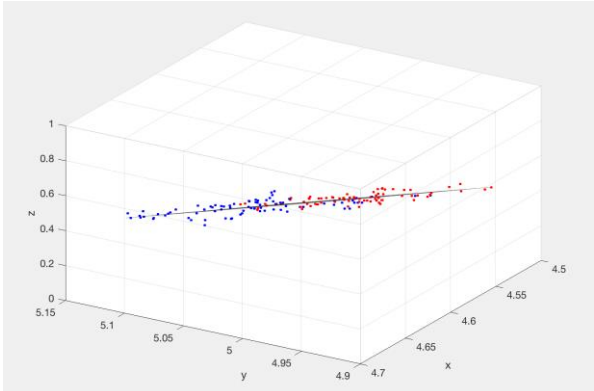
Understanding the proposed scaling law as coordinate transformation

The goal is to define a new coordinate system in the 3D space spanned by $x=\log(A_E)$, $y=\log(A_T)$, and $z=\log(T)$, and understanding the proposed scaling law in that context. To motivate this, we can plot our data (HCP, 20-25 y.o., red=female, blue=male) in this space:



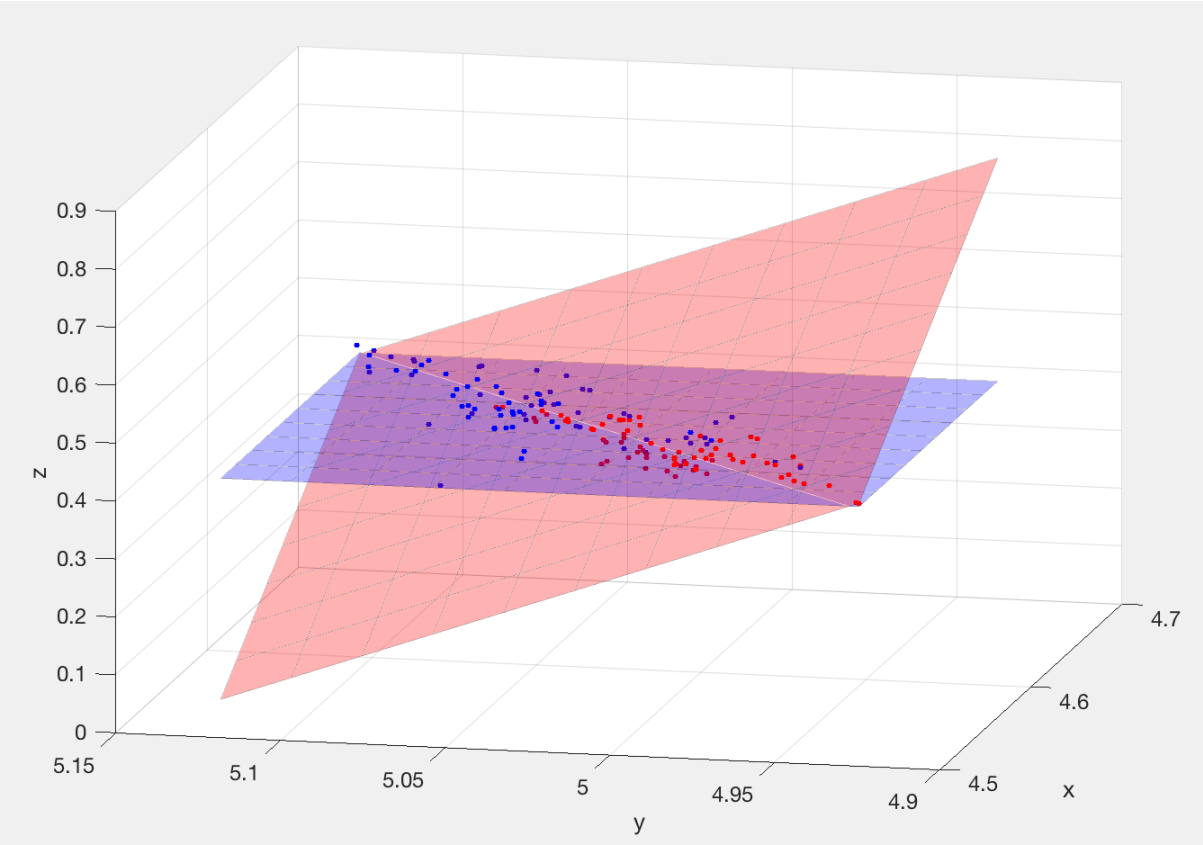
We can see that the data has an intrinsic structure, which supports the idea that the three quantities are related by some physical law/principle, as previously proposed [Mota & Herculano-Houzel 2015]. Indeed, when we plot the proposed scaling law (**k1=-1.25*x+y+0.5*z**, which defines a plane) in this space, it appears to capture the trend:



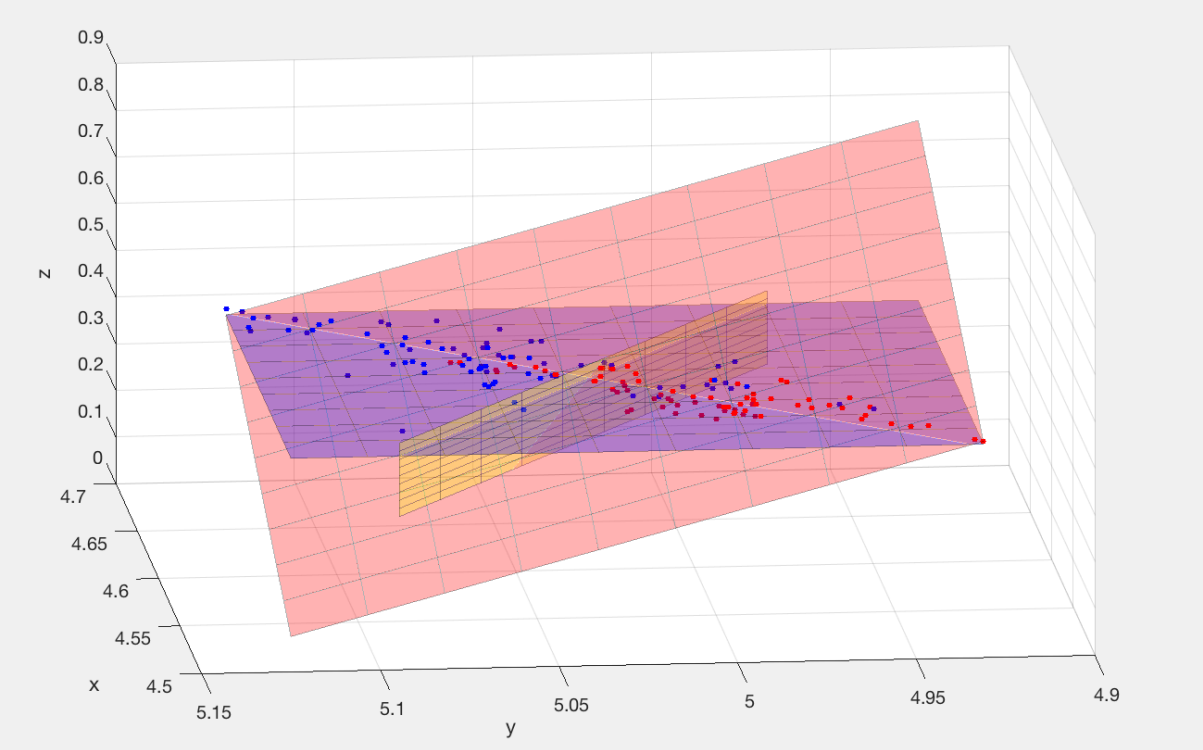


Note that k1 is set to the mean of the dataset in the plots. Note that k1 does not change between males & females (as shown in the main manuscript).

As shown in the main manuscript, a second quantity that does not change in this dataset is cortical thickness. We define **k2=z** (blue plane):

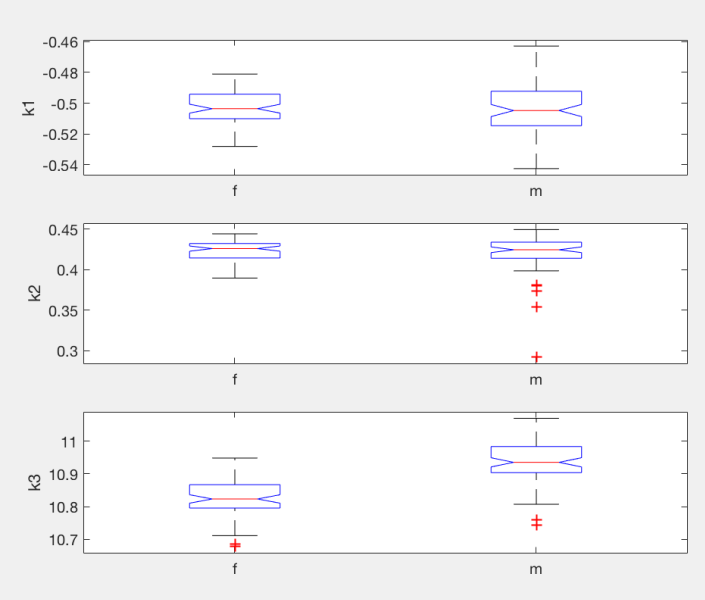


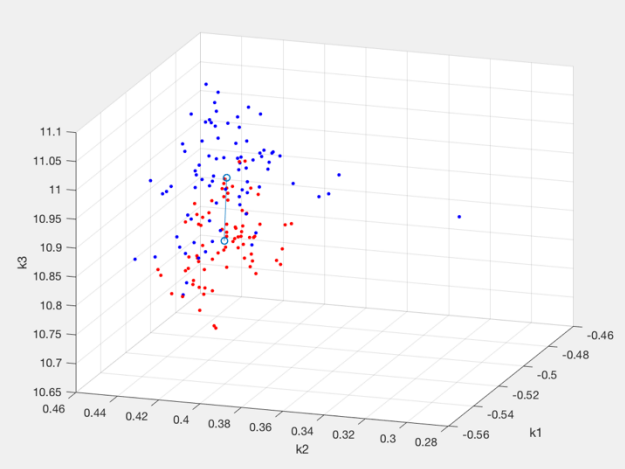
At this point, we can essentially see that k1 and k2 define a new set of coordinates that is more natural to capturing the data. To complete the linear (but not affine) coordinate transformation, we need to define k3. We chose **k3=x+1.25y** (yellow plane):

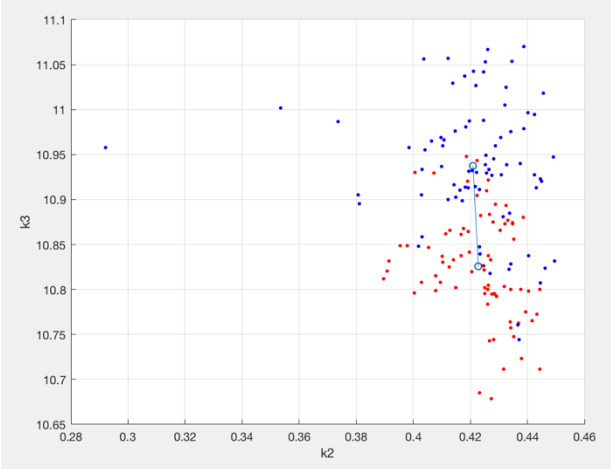


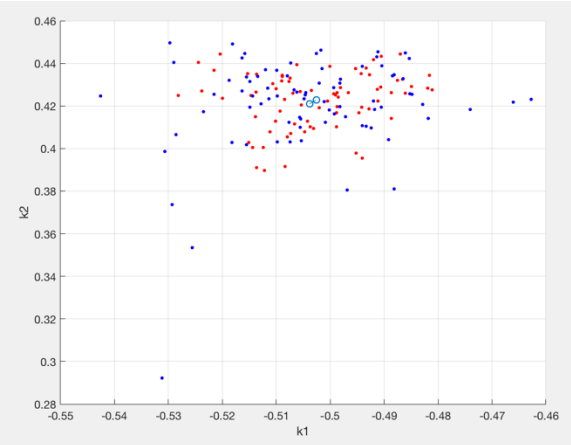
The reasoning behind this choice is to find a perpendicular plane to the intersection of the k1 (red) and k2 (blue) plane.

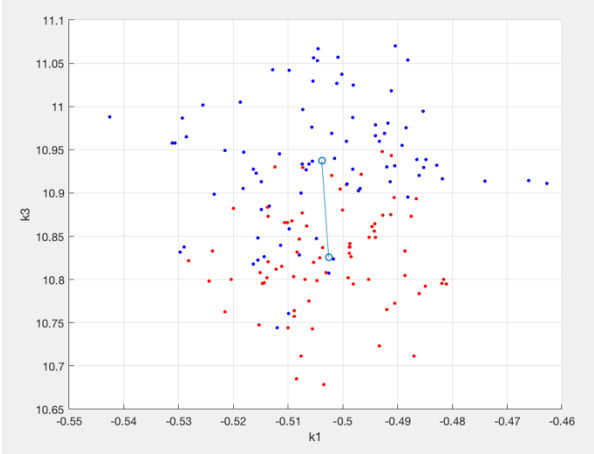
Not surprisingly, the data can now be transformed into k1, k2, k3 space, where the gender difference is not seen in the first 2 coordinates, but only in the third coordinate:





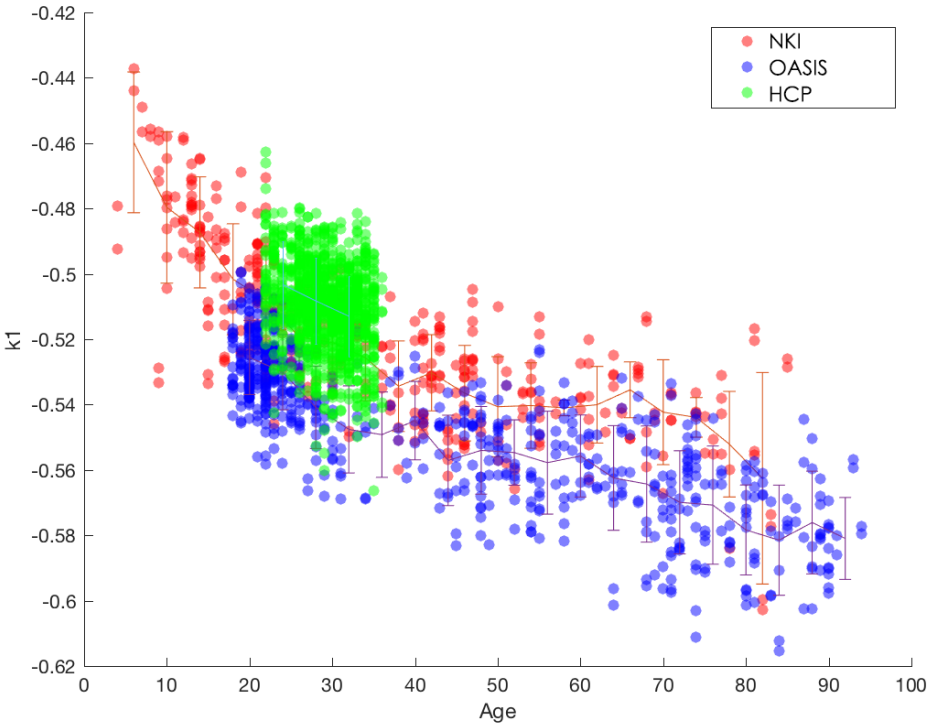


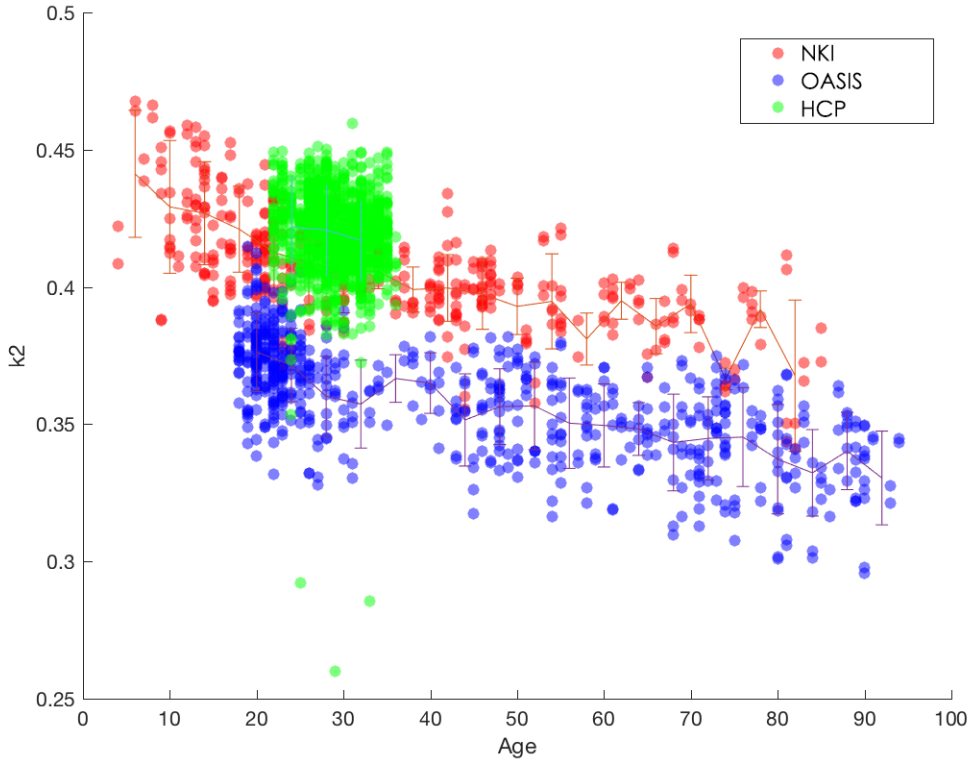


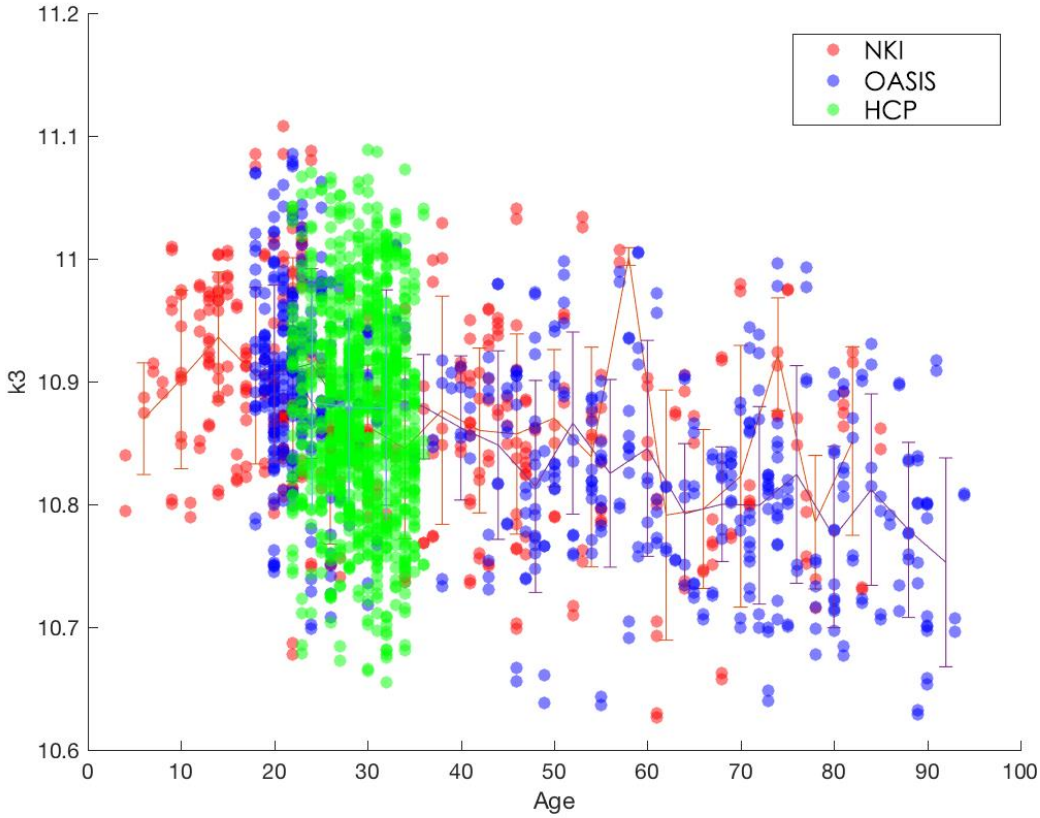


In other words, following this coordinate transformation, changes due to gender only express themselves in the k3 direction.

We can now turn to see how this new coordinate system highlights the changes with age described in the manuscript.

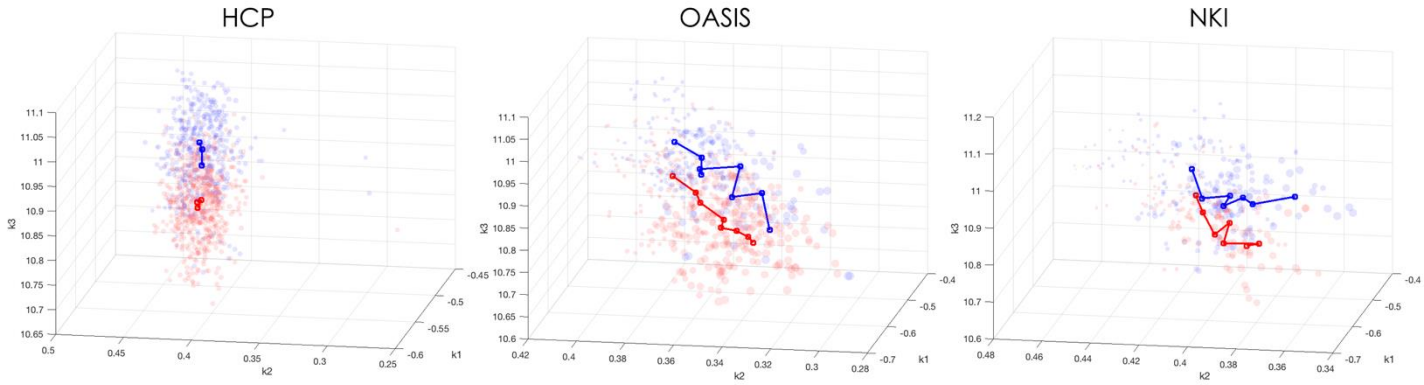

NKI
OASIS
HCP
k1
Age
-0.42
-0.44
-0.46
-0.48
-0.5
-0.52
-0.54
-0.56
-0.58
-0.6
-0.62
0
10
20
30
40
50
60
70
80
90
100


NKI
OASIS
HCP
k2
Age
0.5
0.45
0.4
0.35
0.3
0.25
0
10
20
30
40
50
60
70
80
90
100



We see changes in all three coordinates, but the strongest in k1, and only fairly slight in k3. It's also worth noting that all three dimensions show complimentary information (see especially 0 – 20 year olds). This also shows that the offset drift in k1 is not purely due to the drift in k2 (thickness).

Finally, we show the data actually transformed into the k1, k2, and k3 space:



Each hemisphere of a subject is shown in k1, k2, k3 space with a semi-transparent marker. The marker size indicates the age of the subject (bigger marker = older subject), and the colour indicates the gender (red=female). Age and gender group averages are shown with the thick solid line. As described before, the gender effect stays the same throughout age and only expresses itself as an k3 offset. The age effect is seen as a movement oriented diagonally in this space.

Biological interpretation:

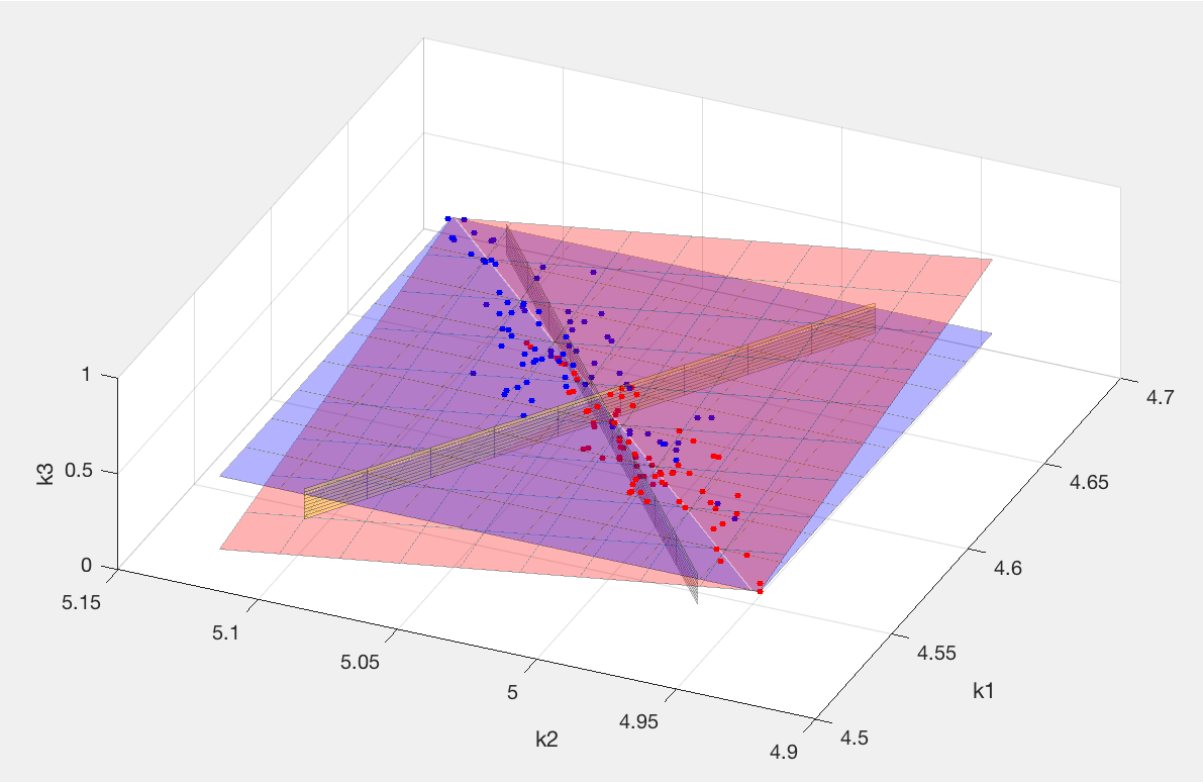
k1, derived from the theory, describes the negative tension on the grey matter surface, which we hypothesise to be white matter tension, or CSF pressure.

k2 is straight-forward, and corresponds to cortical thickness.

k3 actually has a straight-forward interpretation too: it describes the brain volume related changes in total and exposed surface area. In other words, if two brains have the same value of k3, but one brain is bigger in total volume than the other (i.e. bigger exposed surface area), then the bigger brain must have a smaller total surface area. In yet other words, big values of k3 generally indicate a big brain volume, and small values of k3 generally indicate a small brain volume.

In this context, it is worth noting that the gender induced changes are detectable only in k3, meaning that the only difference between genders is their volume and the related change in total and exposed surface area. Also worth pointing out is the age related change in k3, which between 0-20 years appear to be increasing initially, but decreasing again thereafter. Interestingly, this agrees well with the reported rise and fall of white matter volume in a healthy lifespan [Ge *et al.* 2002, Yeatsman *et al.* 2014].

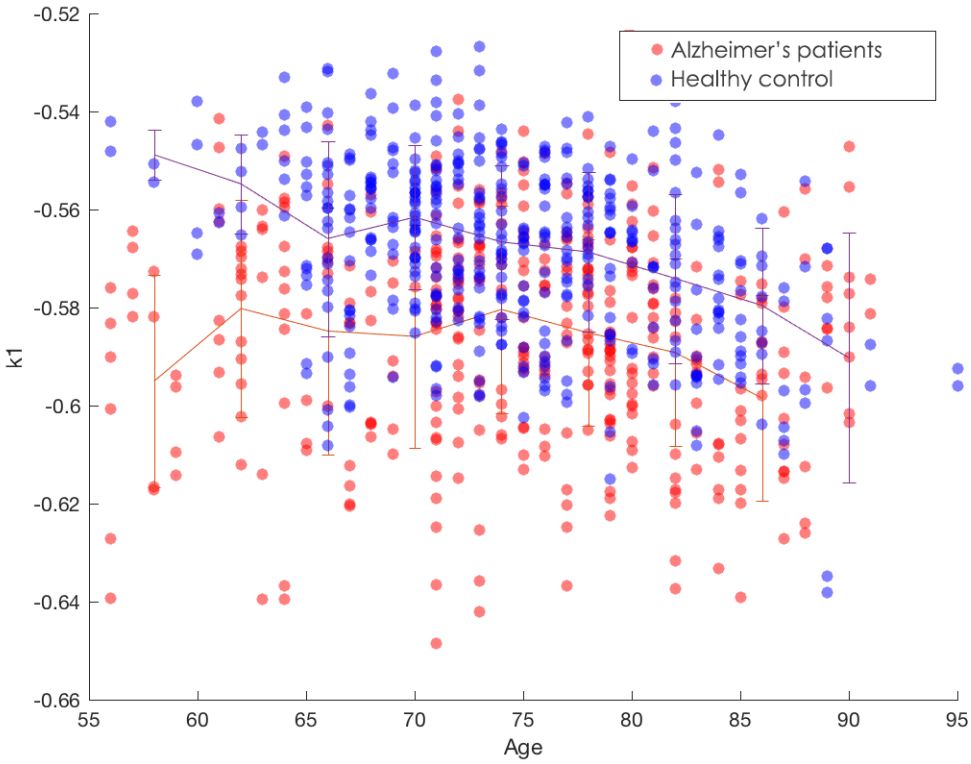
Finally, we want to clarify that k3 is very different to the Gyrification Index:

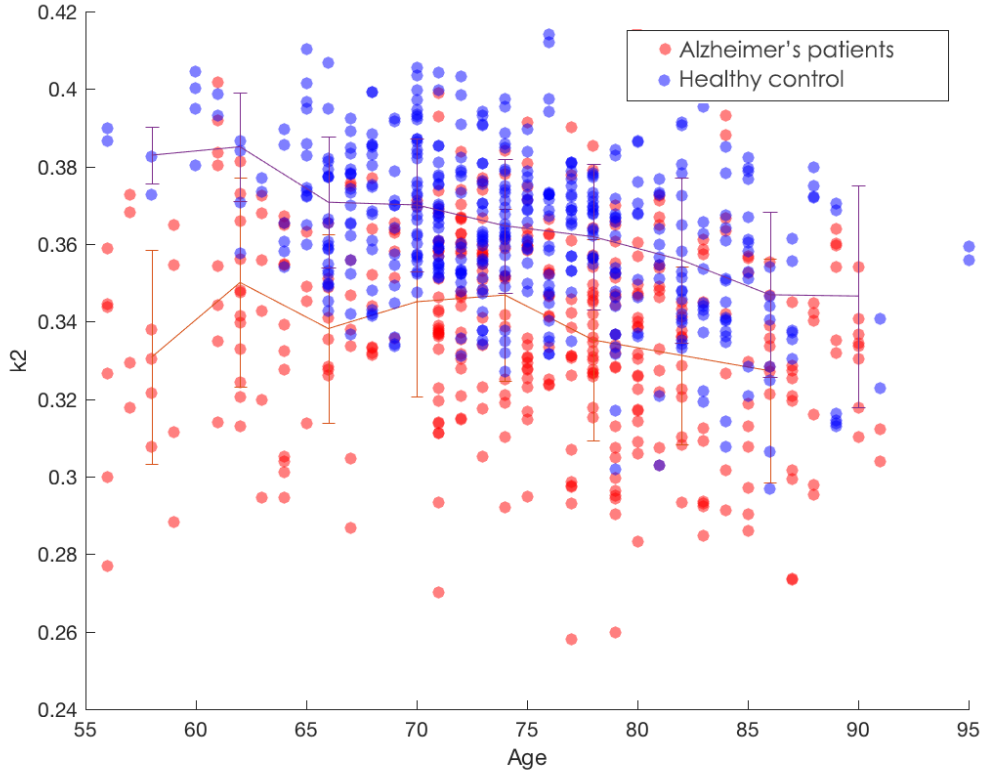


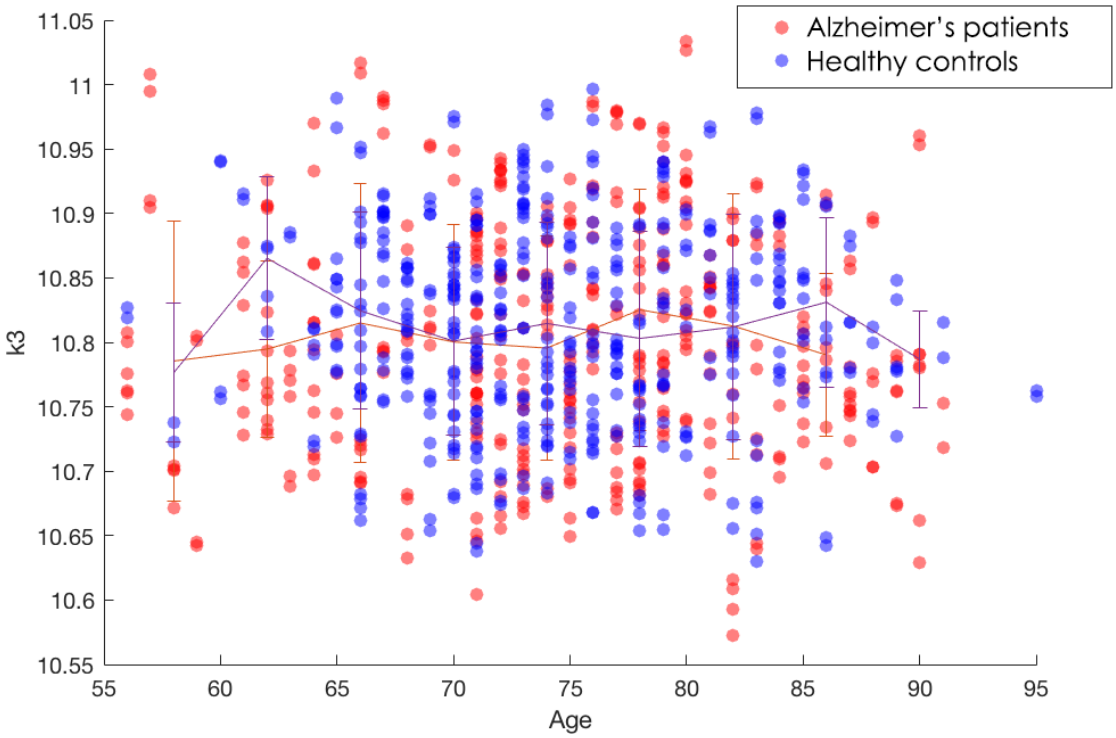
The plane corresponding plane to the gyrification index is the grey plane. As immediately visible, the gyrification index plane is not a natural plane that describes the trend in the data.

Alzheimer's Disease

Repeating the mapping of the ADNI data (red is Alzheimer's Disease, blue is Control) to the k1, k2, k3 coordinates gives:

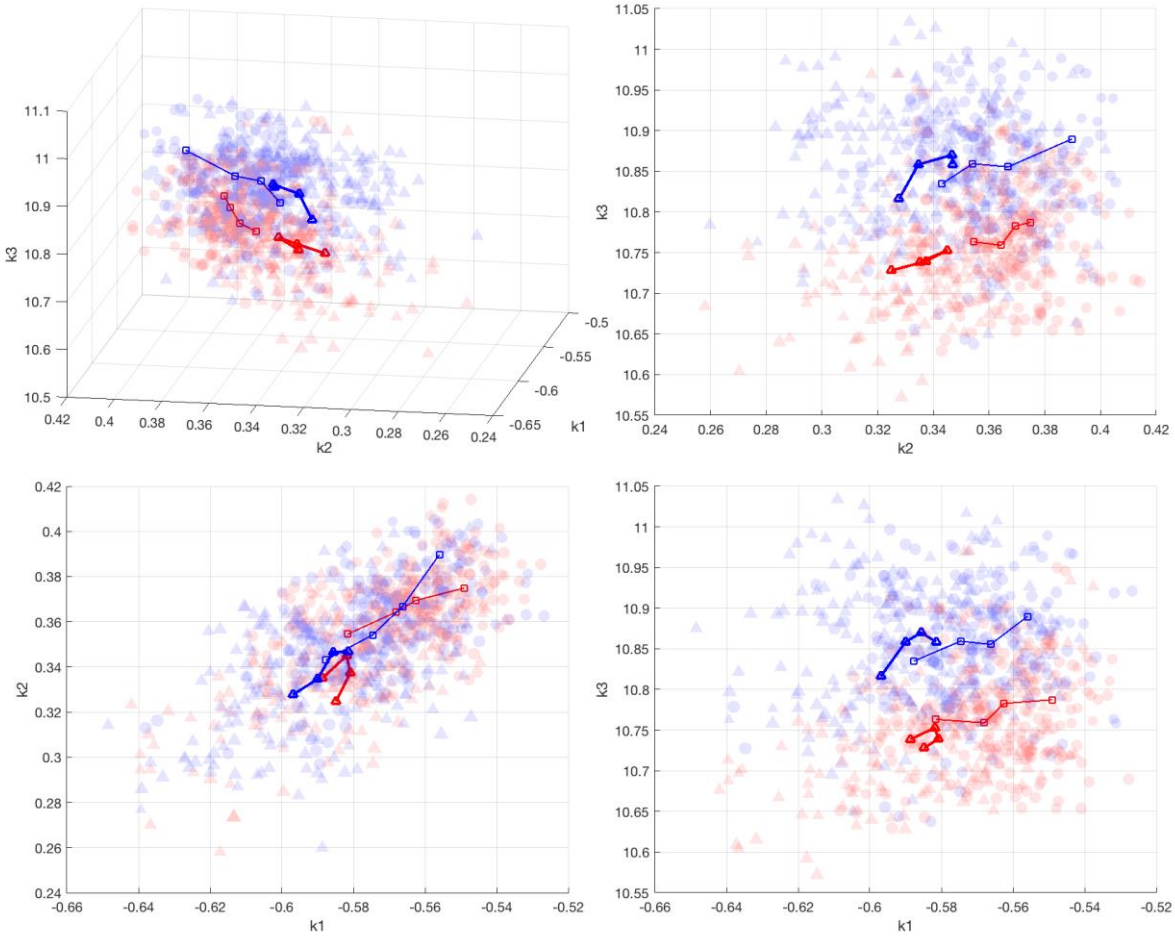






In k3, there is no difference between control and patients. The change in k1 and k2 appear redundant, and hence the difference between control and patients appears to be driven by thickness changes alone.

Finally, we also show the ADNI subjects in the k1, k2, k3 space directly. As before each hemisphere is shown by a marker. Females are red and males as blue. The healthy subjects are shown in round markers and the AD subjects in triangle markers. The age and gender group averages are shown as a solid line (thin for controls, thick for AD).



As before we see the clear difference between genders in k3 direction, and the progression diagonally for age in the healthy subjects. The AD subjects are clearly separated in direction of k1 and k2. Interestingly, although the gender difference is preserved in k3 direction, the AD subjects (especially females) do not show the typical drift in the age direction.

References:

[Ge *et al.* 2002] Brain Development Cooperative Group. Total and regional brain volumes in a population-based normative sample from 4 to 18 years: the NIH MRI Study of Normal Brain Development. Cereb. Cortex 22, 1–12 (2012).

[Mota & Herculano-Houzel 2015] Mota, B. & Herculano-Houzel, S. BRAIN STRUCTURE. Cortical folding scales universally with surface area and thickness, not number of neurons. Science 349, 74–77 (2015).

[Yeatsman *et al.* 2014] Yeatman, J. D., Wandell, B. A. & Mezer, A. A. Lifespan maturation and degeneration of human brain white matter. Nat Commun 5, 4932 (2014).

Text S7:

Supplementary methods

Data sources

We used four publicly available datasets in this study. The **HCP** dataset has been obtained from 500 subjects release from the Human Connectome Project (www.humanconnectome.org). Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

The **OASIS** data has been obtained from the Open Access Series of Imaging Studies project (www.oasis-brains.org). We used the cross-sectional dataset [Marcus *et al.* 2007], and only included the subset of healthy subjects. OASIS project was supported by P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

The **NKI** data [Nooner *et al.* 2012] has been obtained from Nathan Kline Institute (NKI) Rockland Sample (http://fcon_1000.projects.nitrc.org/indi/pro/nki.html), where funding for key personnel was provided in part by the New York State Office of Mental Health and Research Foundation for Mental Hygiene. Additional project support provided by the NKI Center for Advanced Brain Imaging (CABI), the Brain Research Foundation (Chicago, IL), the Stavros Niarchos Foundation, and NIH grant P50 MH086385-S1.

The **ADNI** data has been obtained from the Alzheimer's Disease Neuroimaging Initiative (http://adni.loni.usc.edu/). Data collection and sharing for the ADNI project was supported by ADNI (National Institutes of Health Grant U01 AG024904).
We included subjects from ADNI1, ADNI GO, and ADNI2 [Jack *et al.* 2010], and selected all 3T sessions. For all subjects we only included one scan point (first 3T scan), to be consistent with the previous three cross-sectional datasets. We also only included the healthy controls and the Alzheimer's patients in this study. Details (e.g. field strength, voxel size etc.) for all the datasets can be found in Table S7.1.

As we only used publicly available datasets, no informed consent procedure was required. We confirm that we complied with all the data usage policies of each of the datasets we used. A general ethical approval was granted by Newcastle University to carry out the described research under the reference number 1186/2015.

Data processing

For all datasets we used Freesurfer (http://surfer.nmr.mgh.harvard.edu/) for processing. We ran the Freesurfer recon-all pipeline on all subjects in the NKI and ADNI datasets. For the HCP dataset, we used the preprocessed package provided by the human connectome project,

which already includes the Freesurfer subjects. Details regarding the preprocessing can be found in [Glasser *et al.* 2013]. For the OASIS dataset we used the Freesurfer subjects provided.

To find the exposed surface, we ran the local gyrification pipeline in Freesurfer with the standard settings, and use the surfaces *h.pial-outer-smoothed produced by that pipeline. For some subjects this pipeline failed (subjects with zero entries in Dataset S8); as it is a very small minority of subjects, we have simply excluded them from our analysis.

Cortical thickness measurement

We loaded the grey matter surface produced by Freesurfer (*h.pial) and the cortical thickness (*h.thickness) into Matlab, and matched each vertex in the surface mesh to their corresponding thickness. We then determined triangles in the surface mesh with zero thickness (i.e., with at least one vertex with zero thickness) and excluded them from subsequent analyses. As a last step we calculated the average vertex thickness for each remaining triangle, and the total average cortical thickness as an average of all the triangle thicknesses. The conversion from vertex based thickness to triangle based thickness is to enable comparisons to the surface area (which is triangle based) for present and future analyses.

To double check the validity of our results based on this method, we repeated all analyses with the average cortical thickness provided by Freesurfer; all results remain substantially unchanged.

Total and exposed surface area measurement

The total exposed surface area was calculated by loading the *h.pial surface into Matlab. Again, we removed triangles with zero thickness (corpus collosum) and took the sum of the remaining triangle areas. This is in agreement with the approach taken by previous comparative neuroanatomical studies [Mota *et al.* 2015].

The total exposed surface area was calculated by loading the *h.pial-outer-smoothed into Matlab. We find and remove the triangles on the exposed surface corresponding to the previously-removed zero-thickness triangles on the pial surface, and take the sum of the remaining triangle surface areas. We use a heuristic to find the corresponding triangles, which essentially determines the top three nearest triangles (by triangle center distance) within 5mm radius for each zero-thickness triangles on the pial surface. Visual inspection of the results shows a very good agreement in general. Additionally, Fig. 1 in the main manuscript shows that our data agrees well with the previous comparative neuroanatomy data obtained by manual labelling. For each subject, we thus obtain a datapoint for each hemisphere, which, as in the previous studies, we treat as individual datapoints. As each subject has exactly two datapoints, this should not distort the results. Finally, we also calculated the **gyrification index** as the ratio of the total grey matter surface area over the exposed grey matter surface area.

Supplementary Table S7.1: Details of the public datasets we used. For all the references, the links are provided.

Dataset	HCP	NKI	OASIS (cross sectional)	ADNI control	ADNI AD
Data sample origin	https://db.humanconnectome.org/app/template/Index.vm	http://fcon_1000.projects.nitrc.org/indi/pro/nki.html	http://www.oasis-brains.org/	http://adni.loni.usc.edu/	
Primary publication	Glasser *et al.* 2013 http://www.sciencedirect.com/science/article/pii/S1053811913005053	Nooner *et al.* 2012 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3472598/	Marcus *et al.* 2010 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2895005/	Jack *et al.* 2010 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2886577/	
Number of subjects	526	207	316	235	199
Female	315	87	197	124	93
Male	211	120	119	111	106
Age range	22-36 (mean: 29.21)	4-85 (mean: 35.01)	18-94 (mean: 45.09)	56-95 (mean: 74.60)	56-91 (mean: 74.91)
Voxel size	0.7 mm isotropic	1 mm isotropic	1 mm isotropic	varied	
Field strength	3 Tesla	3 Tesla	1.5 Tesla	3 Tesla	3 Tesla
Scanning protocol	MP-RAGE	MP-RAGE (http://fcon_1000.projects.nitrc.org/indi/pro/nki/NKI_MPRAGE_PROTOCOL.pdf)	MP-RAGE	MP-RAGE (http://adni.loni.usc.edu/methods/documents/mri-protocols/)	
Scanner type	Siemens (Skyra modified)	Siemens (Magnetom)	Siemens (Vision Scanner)	Differed across sites	
Pre-processed data provided	Freesurfer subjects created from T1 and T2 images	As Lim *et al.* 2013 http://cercor.oxfordjournals.org/content/25/6/1477	Freesurfer subjects created from 3-4 T1 images	Preprocessed (see http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/)	
Freesurfer version	5.3 modified version	5.0	dev 20061005	5.3	

Data S8

Microsoft Excel 97 - 2004 Worksheet

The study brought together existing data obtained upon request and subject to licence restrictions from a number of different sources. Full details of data available in the documentation at http://dx.doi.org/10.17634/122519-1

The dataset is also available on the author's website (http://ywang.co.uk/) under the category resources, or on Zenodo under the link https://zenodo.org/record/61348, or upon email request (yujiang.wang@ncl.ac.uk).