Supplementary Material

Supplementary Statistical methods

The outcome in both hierarchical models, standard deviations of young, healthy volume, was calculated using the following steps. Starting with 87 MRI scans from a cohort of young, cognitively normal subjects between the ages of 35 and 55, we fit regional ordinary least squares (OLS) regressions predicting volume by an intercept term and TIV. Next, we used these OLS models to predict the expected volume in each region of each scan in the autopsy cohort using measured TIV at each time point. We then calculated the residual volume for each region, scan, and subject, and divided these residuals from the autopsy cohort by the standard deviation of the residuals in the young OLS model fits. Thus the regional volumes and rates of atrophy in the older autopsy cohort could then be expressed in terms of standard deviations of regional volumes in the young, healthy cohort. This was useful because it put regions of inherently different sizes on the same scale; a one standard deviation change in the hippocampus is equivalent to a one standard deviation change in any other region and we could then directly compare effects across regions.

Within-person-across-region intercept terms were assumed to follow a multivariate normal distribution centered at zero to allow for nonzero within-person correlation between regional volumes. Within-person early rate modifiers were assumed to follow a normal distribution centered at zero. These random effects account for multiple regions and often multiple scans from a subject.

Algebraically, our model can be represented as

$y_{rij} = \beta_{01r} + \beta_{02r}T_i + \beta_{03r}B_i + \beta_{04r}C_i +$	Intercept terms
$\beta_{05r}T_iB_i + \beta_{06r}T_iC_i + \beta_{07r}B_iC_i +$	
$(\beta_{11r} + \beta_{12r}T_i + \beta_{13r}B_i + \beta_{14r}C_i +$	Early rate terms
$\beta_{15r}T_iB_i + \beta_{16r}T_iC_i + \beta_{17r}B_iC_i)\ ttd_{ij} +$	
$(\beta_{21r} + \beta_{22r}T_i + \beta_{23r}B_i + \beta_{24r}C_i)\ ttd'_{ij} +$	Acceleration terms
$\beta_{30r}\ age\_death_i + \beta_{32r}\ male_i + \beta_{33r}\ FS_{ij} +$	Covariate
$\beta_{34r}\ age\_death_i\ ttd_{ij} + \beta_{35r}\ vascular_i +$	adjustments
$\beta_{36r}\ vascular_i\ ttd_{ij} +$	
$\gamma_{0ri} + \gamma_{1i}\ ttd + \varepsilon_{rij}$	Random effects and

error term

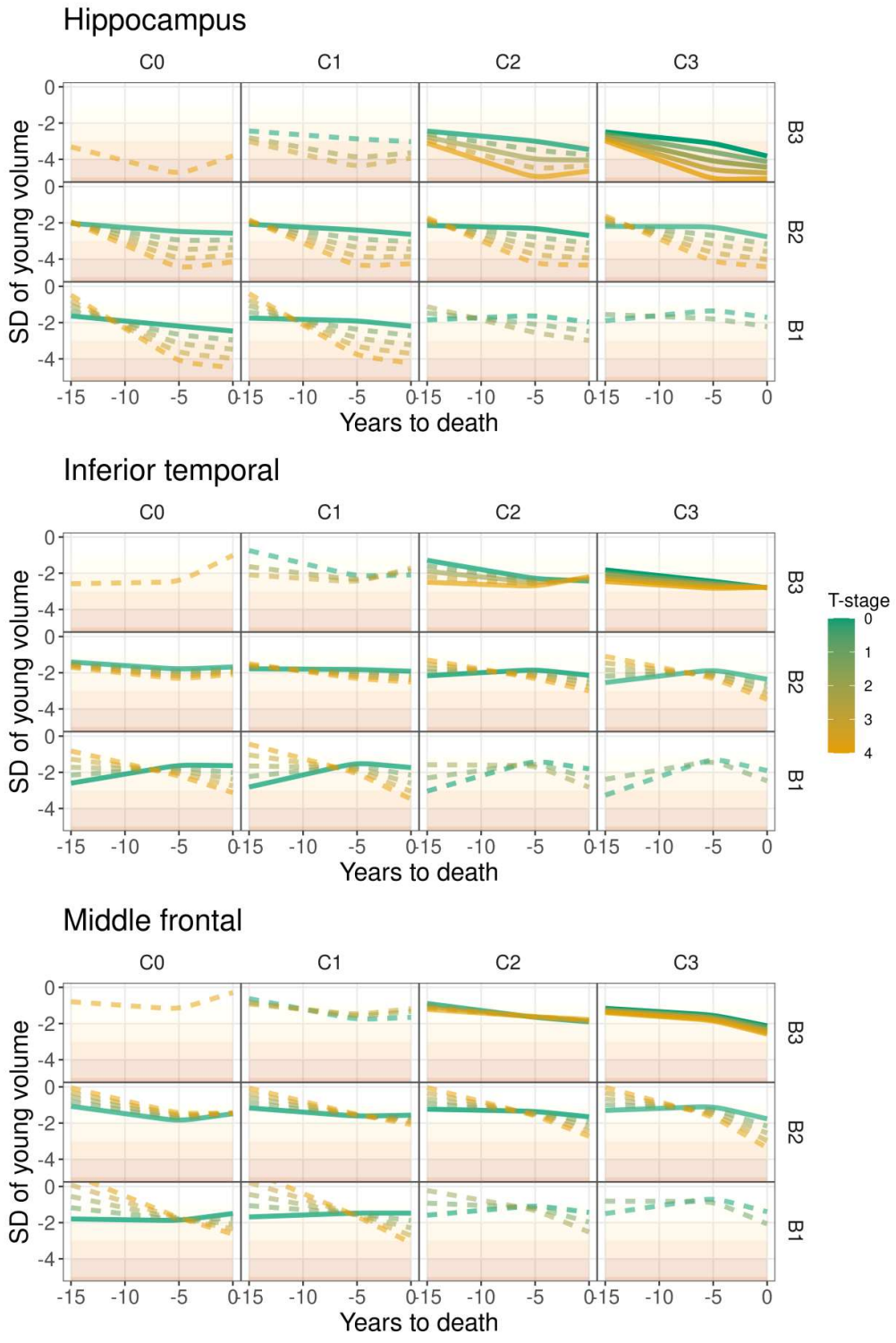
using the following abbreviations: *ttd* = years to death, *ttd'* = restricted cubic spline modification to early rate, *age_death* = age at death, *FS* = field strength, *vascular* = vascular composite score, *r* = region, *i* = subject, *j* = scan, *T* = TDP staging, *B* = Braak staging, and *C* = CERAD staging. We will discuss the results of this model using the following groupings; intercept terms that include an overall intercept per region, β_{01r}, with six modifications depending on protein stage: $\beta_{02r}T_i + \beta_{03r}B_i + \beta_{04r}C_i$ and $\beta_{05r}T_iB_i + \beta_{06r}T_iC_i + \beta_{07r}B_iC_i$, an early rate, including an overall rate per region, β_{11r}, with six modifications that depend on protein stage: $\beta_{12r}T_i + \beta_{13r}B_i + \beta_{14r}C_i$ and $\beta_{15r}T_iB_i + \beta_{16r}T_iC_i + \beta_{17r}B_iC_i$, and acceleration, including an overall acceleration term, β_{21r}, also modified by protein stage: $\beta_{22r}T_i + \beta_{23r}B_i + \beta_{24r}C_i$. The acceleration can be added to the early rate term, resulting in a "late rate" to simplify the interpretation of the results. The rest of the terms include intercept adjustments for demographic and clinical covariates, and a single rate adjustment for age at death. The γ_{0ri} and γ_{1i} terms are the person-specific regional intercept deviations and person-specific rate deviations, and finally, an error term for each observation, ε_{rij}, was included. Age at death was centered at 80 years with a standard deviation of 5 years, time to death was centered so the intercept represents 5 years from death and TIV was centered at 1.4 liters with a standard deviation of 0.2 liters. Continuous protein stage scales were centered at median values, TDP-43 stage at T1, CERAD score at C2, and Braak stage at B2, to improve the estimation and interpretation of intercept terms in the model.

Of primary interest are the associations between regional rates of atrophy and accelerations of atrophy with changes in TDP-43 stage, as well as ADNC stage, the $\beta_{1\bullet}$ and $\beta_{2\bullet}$ terms in the model. The other variable and person-specific terms account for the collinearity in multiple scans per subject and adjustments to volumes and rates to account for the effects of age at death, TIV, sex, field strength, and vascular composite score.

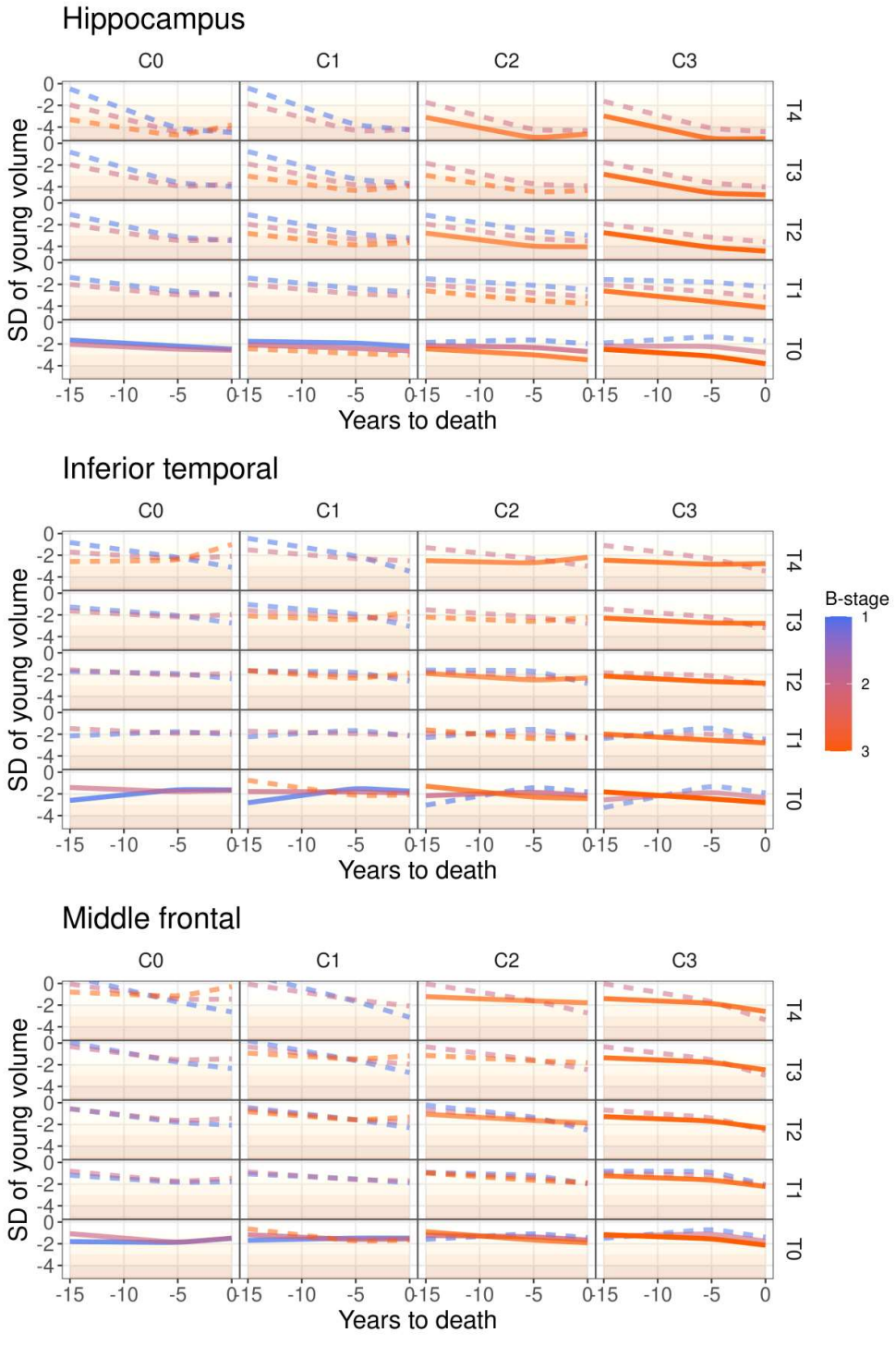
Groups of related parameters were estimated in batches following in the grain of Gelman and Hill [1]. Person-specific intercept terms were assumed to follow a multivariate normal distribution (i.e. the prior distribution of the parameters) allowing for nonzero correlation of volumes across regions within person. We denote these $\gamma_{k\cdot} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma_k})$; $\boldsymbol{\Sigma_k} = \begin{bmatrix} \sigma_{k1}^2 & \phi_k^2 \\ \phi_k^2 & \sigma_{k2}^2 \end{bmatrix} for\ k = 0, 1$, where σ_{k1}^2 represents the variance of the random

person-deviations for the hippocampus and σ_{k2}^2 represents the variance of the random person-deviations for either the inferior temporal or the middle frontal region and ϕ^2 represents the covariance between the hippocampus and second region in each model. The rest of the terms in the model were assumed to follow independent normal prior distributions, denoted $N(\mu, \sigma^2)$, with means following independent standard normal distributions and variances following standard half-normal (strictly positive) distributions. The error term was assumed to follow a normal distribution centered at zero with standard deviation drawn from a half normal prior distribution centered at zero with standard deviation 3, denoted $N(0, 3^2)^+$.

Supplementary Figure 1: Estimated mean volumes (combining volume shift, early rate, and acceleration) by time to death, across TDP-43 stage (T-stage) in each Braak neurofibrillary tangle stage (B-stage) and CERAD plaque score (C0-C3), restricted to combinations of stages present in our data and following the same dashed- and transparent-line scheme used in Figure 4 in the main text to indicate the quantity of data available in each group.



Supplementary Figure 2: Estimated mean volumes (combining volume shift, early rate, and acceleration) by time to death, across Braak stages (B-stage) in each TDP-43 stage (T-stage) and CERAD score (C0-C3), restricted to combinations of stages present in our data and following the same dashed- and transparent-line scheme used in Figure 4 in the main text to indicate the quantity of data available in each group.



References

1. Gelman, A. & Hill J, L. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge University Press, Cambridge, UK, 2006).