Supplementary Online Content

Wennberg AM, Tosakulwong N, Lesnick TG, et al. Association of Apolipoprotein E ε4 With Transactive Response DNA-Binding Protein 43. *JAMA Neurol.* Published online October 22, 2018. doi:10.1001/jamaneurol.2018.3139

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Structural equation models (SEMs):

SEMs can be considered a type of extension of a regression model. While regression models are discussed using the terms "association" and "coefficient," SEMs use the terms "direct" (unmediated), "indirect" (mediated), and "total" (direct + indirect) effects to describe pathways within the specified model framework. Regression models have two layers, one consisting of an outcome variable, and one consisting of one or more predictor variables. SEMs can extend this by introducing more layers based upon theory or prior information. For example, we can have APOE ε4 predicting Aβ, and Aβ in turn predicting tau, APOE ε4→Aβ→tau. Square boxes contain manifest (measured) variables, and arrows represent regression models joining variables, with the arrowhead pointing to the outcome variable for that regression. Our diagrams flow from left to right. Variables which only appear as predictors are referred to in the SEM literature as exogenous, while all other variables are referred to as endogenous. On the right side, we put the ultimate outcome variable. Variables with arrows both entering and leaving are sometimes called mediators. SEMs thus allow us to address complicated hypotheses, and can provide a clearer picture of processes involved in complex diseases than is possible with classical regression models. We can estimate direct effects (associations joining two variables without any intervening variables), indirect effects (associations joining two variables which pass through intervening or mediating variables), and total (sum of direct and indirect) effects. In the example above, a classical regression model using APOE ε4 and Aβ to predict tau might find that APOE ε4 is not associated with tau after adjusting for Aβ. However, a SEM would allow us to test if APOE ε4 influences tau indirectly through Aβ, and is thus an important part of the process. Logistic regression (using a logit link function based on the inverse of the cumulative logistic

function) and probit regression (using a probit link function based on the inverse of the cumulative normal distribution) are two common choices for dealing with categorical outcomes. Regression coefficients in probit models can be complicated to interpret (they cannot be readily converted to odds ratios as in logistic regression). However, probit models have one distinct advantage over logistic models in SEMs: the direct and indirect effects can be added together to get total effects. This is not true for logistic regressions. We used probit regressions in these SEMs. The equations of the direct effects can be obtained from the tables. The generic equation for the effect on TDP-43 can be written as such:

$$\text{Probit(TDP-43)} = \text{intercept} + \text{APOE } \varepsilon4_{\text{estimate}} + \text{tau}_{\text{estimate}} + \text{age}_{\text{estimate}} + \varepsilon$$

Sensitivity analyses:

We also performed a sensitivity analysis adding hippocampal sclerosis to the SEM. In this model we made the assumption that hippocampal sclerosis is a downstream phenomenon related to TDP-43. This assumption is based on the universal findings of TDP-43 in the absence of hippocampal sclerosis, but the lack, or rarity, of hippocampal sclerosis in the absence of TDP-43. Hippocampal sclerosis was another binary endogenous variable, and we again used probit models in the SEM. We added a small constant to deal with the zero count of hippocampal sclerosis in the absence of TDP-43.[24]

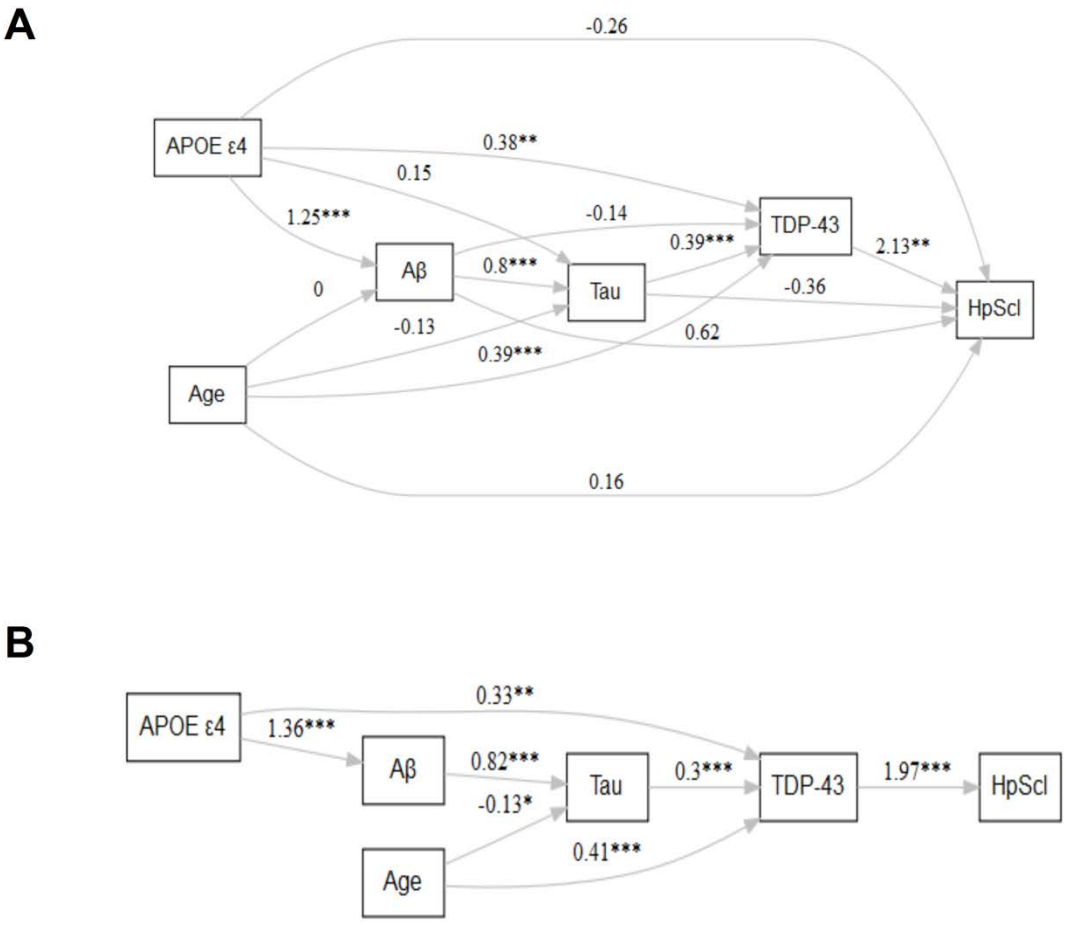
In a secondary analysis we fit an additional SEM to map the associations between APOE ε4, age, Aβ, tau, and TDP-43 using ordinal variables for Aβ and tau. Aβ was classified into four levels using CERAD scores (0=normal; 1=sparse/possible; 2=moderate/probable; 3=frequent/definite). Tau was classified into seven levels (Braak 0-VI). Besides these changes in coding of Aβ and tau, the models remained the same.

eTable 1. Probit regression results from the secondary structural equation model				
	Estimate	Standard Error	z-value	p
Direct effects				
TDP				
APOE ε4	0.319	0.11	2.97	0.003
Tau	0.296	0.04	8.08	<0.001
Age	0.396	0.06	6.37	<0.001
Intercept	-1.650	0.16	10.19	<0.001
Tau				
APOE ε4	0.293	0.08	3.47	<0.001
Age	0.073	0.03	2.23	0.03
Aβ	1.044	0.05	22.09	<0.001
Aβ				
APOE ε4	0.831	0.10	7.96	<0.001
Age	-0.227	0.05	-4.56	<0.001
Indirect effects on TDP				
APOE ε4→Tau	0.087	0.03	3.26	0.001
APOE ε4→ Aβ →Tau	0.257	0.04	5.75	<0.001
Age→Tau	0.022	0.01	2.10	0.04
Age→ Aβ →Tau	-0.070	0.02	-3.97	<0.001
Aβ →Tau	0.309	0.04	7.77	<0.001
Total effects on TDP				
APOE ε4	0.663	0.11	6.15	<0.001
Tau	0.296	0.04	8.08	<0.001
Age	0.347	0.06	5.59	<0.001
Aβ	0.309	0.04	7.77	<0.001
APOE and TDP-43 were coded as positive (1) or negative (0). Aβ was coded as CERAD scores 0-3. Tau was coded as Braak 0-VI.				

eFigure 1. Sensitivity analysis of structural equation models (SEMs) path analyses investigating the effect of APOE ε4 on TDP-43, accounting for hippocampal sclerosis using penalized scores.

eFigure 1a. SEMs path analyses including hippocampal sclerosis downstream of TDP-43. The arrow linking APOE ε4 to hippocampal sclerosis represents a direct effect, and the arrows from APOE ε4 to TDP-43 to hippocampal sclerosis represent indirect effects (mediation by TDP-43) as did the arrows linking APOE ε4 to Aβ to tau to TDP-43 to hippocampal sclerosis (mediation by Aβ, tau, and TDP-43). Numbers on the arrows are estimated regression coefficients on the probit scale. Numbers leading into the same box can be compared, but numbers leading into different boxes cannot because of differing measurement scales. Significant values are marked with asterisks.

eFigure 1b. This SEM was tested in a sensitivity analysis, and included hippocampal sclerosis. It became more parsimonious after eliminating pathways that were not significant at the p<0.10 level. Numbers on the arrows are estimated regression coefficients on the probit scale. Numbers leading into the same box can be compared, but numbers leading into different boxes cannot because of differing measurement scales. Significant values are marked with asterisks.



eFigure 2. Secondary analyses of structural equation models (SEMs) path analyses investigating the effect of APOE ε4 on TDP-43, coding Aβ and tau as ordinal variables.

Significant values are marked with asterisks.