Supplementary Information

The effects of microglia on tauopathy progression can be quantified using Nexopathy *in silico* (Nex*is*) models

---

Contents

1 Overview **2**

2 Model Development **2**
2.1 Description of parameters and notation 2
2.2 Derivation of general model in matrix form 3

3 Cost functions and fitting parameter to data: **5**

4 Strategies for parameter consistency **5**

5 Pseudocode for Nex*is*:global and Nex*is*:microglia **5**

6 Gene selection for Nex*is*:microglia **5**

7 Model parameters and fit for *Apoe* **6**

December 20, 2021

1. Overview

Our previous Network Diffusion Model (NDM) can be extended to incorporate the effects of pathology accumulation/clearance rate in addition to pathology diffusion rate. We call this model Nexopathy *in silico* or Nex*is*. This can be further extended to include the potential effects that different cell-types, such as microglia, may have on the progression of neuropathology, particularly in the context of Alzheimer's disease (AD) and other tauopathies. We thus call this model Nex*is*:microglia, as opposed to Nex*is*:global that is posited without the effects of other cellular and molecular mediators of pathology spread. In particular, Nex*is*:microglia seeks to model two different effects of microglia: 1) modulation of intra-regional accumulation/clearance rate, given by the term M_α and 2) modulation of inter-regional spread rate, given by the term M_β. These hypotheses are not mutually exclusive, but they have different biological and mathematical implications that we explore below.

2. Model Development

2.1. Description of parameters and notation

- t: time
- $x_j(t)$: amount of pathology in brain region j (known)
- c_{ij}: connectivity between regions i and j (known)
- u_{jk}: cell-type density/spatial gene expression k at region j (known)
- α: global amplification/depletion term (unknown)
- β: global diffusivity constant (unknown)
- γ: scaling factor for initial pathology i.e. the amount of exogenous tau injected at the seeding site (unknown)
- b_k: parameter on how $\vec{u_k}$ affects transmission (unknown)
- p_k: interaction parameter between cell type k and pathology (unknown)

M_α: Initiation/Clearance terms.

- (i): $\alpha \cdot x_j$ (independent of cell type)
- (ii): $\sum_{k=1}^{n} p_k \cdot u_{jk} \cdot x_j$ (cross-terms between x_j and u_j)

M_β: Inter-regional spread terms.

- (i): $\beta \cdot \sum_i [c_{ij}x_i - c_{ji}x_j]$ (Nex*is*:global)
- (ii): $\beta \cdot \sum_i [c_{ij}(1 + \sum_k b_k u_{ik})x_i - c_{ji}(1 + \sum_k b_k u_{jk})x_j]$ (Nex*is*:microglia, re-weighted by cell types)

We denote vectors with arrows and matrices with capital letters. The Nex*is*:microglia equation takes the form

$$\frac{dx_j}{dt} = \underbrace{M_\alpha(\vec{x}|U, \alpha; \vec{a}, \vec{p}),}_{\text{accumulation/clearance}} + \underbrace{M_\beta(\vec{x}|C, U, \beta, \vec{b})}_{\text{inter-regional spread}} \tag{1}$$

The vector $\vec{x} \in \mathbb{R}^n$ denotes the pathology measured along n regions of interest (ROIs), with an entry x_j denoting the pathology at region j. The vectors $\vec{u_1}, \vec{u_2}, \ldots, \vec{u}_k$ denote the cell-type densities along the same n ROIs and are stored in the matrix U.

$$U = \begin{bmatrix} | & | & & | \\ \vec{u_1} & \vec{u_2} & \cdots & \vec{u_k} \\ | & | & & | \end{bmatrix}$$

The parameters $\{\alpha; \vec{a}, \vec{p}\}$ are associated with initiation/clearance of pathology and $\{C, U, \beta, \vec{b}\}$ are associated with inter-regional spread.

2.2. Derivation of general model in matrix form

We begin by writing all terms explicitly for (i)-(ii) for M_α and (ii) for M_β on a system with 3 ROI given by x_1, x_2, and x_3. We also define the following auxiliary variables:

$$\begin{aligned} s_j^b &= \sum_k u_{jk} \cdot b_k \\ s_j^p &= \sum_k u_{jk} \cdot p_k \end{aligned}$$

The model then writes:

$$\begin{aligned} \dot{x_1} &= \beta\big[c_{21}(1 + s_2^b)x_2 - c_{12}(1 + s_1^b)x_1 + c_{31}(1 + s_3^b)x_3 - c_{13}(1 + s_1^b)x_1\big] + \alpha x_1 + s_1^p x_1 \\ \dot{x_2} &= \beta\big[c_{12}(1 + s_1^b)x_1 - c_{21}(1 + s_2^b)x_2 + c_{32}(1 + s_3^b)x_3 - c_{23}(1 + s_2^b)x_2\big] + \alpha x_2 + s_2^p x_2 \\ \dot{x_3} &= \beta\big[c_{13}(1 + s_1^b)x_1 - c_{31}(1 + s_3^b)x_3 + c_{23}(1 + s_2^b)x_2 - c_{32}(1 + s_3^b)x_3\big] + \alpha x_3 + s_3^p x_3 \end{aligned}$$

Rearranging terms and grouping them as factors of x_1, x_2 and x_3 yields

$$\begin{aligned} \dot{x_1} &= [\alpha + s_1^p - \beta(c_{12} + c_{13})(1 + s_1^b)]x_1 + [\beta c_{21}(1 + s_2^b)]x_2 + [\beta c_{31}(1 + s_3^b)]x_3 \\ \dot{x_2} &= [\beta c_{12}(1 + s_1^b)]x_1 + [\alpha + s_2^p - \beta(c_{21} + c_{23})(1 + s_1^b)]x_2 + [\beta c_{32}(1 + s_3^b)]x_3 \\ \dot{x_3} &= [\beta c_{13}(1 + s_1^b)]x_1 + [\beta c_{23}(1 + s_2^b)]x_2 + [\alpha + s_3^p - \beta(c_{31} + c_{32})(1 + s_3^b)]x_3 \end{aligned}$$

The terms of the RHS of the system of equations can be expressed as a matrix-vector multiplication. Note that the diagonal elements of the matrix have a term of the form $\alpha + s_j^p$. For convenience, we will incorporate them in a diagonal matrix Λ, that easily generalizes for n regions as shown below:

$$\begin{bmatrix} \Lambda_{11} & 0 & 0 & \cdots & 0 \\ 0 & \Lambda_{22} & 0 & \cdots & 0 \\ 0 & 0 & \Lambda_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Lambda_{nn} \end{bmatrix}, \text{where} \quad \Lambda_{jj} = \alpha + s_j^p.$$

To be consistent with previous studies, we will write the remaining matrix as $-\beta L = -\beta(D - C)$, where D is a diagonal matrix. The off-diagonal terms take a simpler form, $L_{ij} = -c_{ij}(1 + s_i^b)$, that easily generalizes for n dimensions. The diagonal terms, however, must be augmented to include $c_{j4} + \cdots + c_{jN}$. In these sums, only the term c_{jj} is missing in line j. Thus, to write the diagonal terms in a more compact form, we introduce the following trick:

$$r_j = -c_{jj} + \sum_i c_{ji}.$$

The RHS of the general model in matrix form becomes

$$\left(\begin{bmatrix} \Lambda_{11} & 0 & 0 & \cdots & 0 \\ 0 & \Lambda_{22} & 0 & \cdots & 0 \\ 0 & 0 & \Lambda_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Lambda_{nn} \end{bmatrix} - \beta \begin{bmatrix} L_{11} & L_{12} & L_{13} & \cdots & L_{1n} \\ L_{21} & L_{22} & L_{23} & \cdots & L_{2n} \\ L_{31} & L_{32} & L_{33} & \cdots & L_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & L_{n3} & \cdots & L_{nn} \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \tag{2}$$

where

$$\begin{aligned} \Lambda_{jj} &= \alpha + s_j^p \\ L_{jj} &= (1 + s_j^b) r_j \\ L_{ij} &= -c_{ij}(1 + s_i^b), \quad \text{when} \quad i \neq j \end{aligned}$$

With the definitions above, the model reads:

$$\frac{d\vec{x}}{dt} = [\Lambda - \beta L]\vec{x} \tag{3}$$

If all $b_k = 0$ or if all $u_{jk} = 0$, $s_j^b = 0$ for all j and the L matrix will reduce to the familiar (unnormalized) Laplacian matrix. It is also worth noting that the degrees of freedom of this system is equal to $3k + 2$, which grows unwieldy when the total number of cell types, m, is large given the low temporal resolution of most pathology data.

3. Cost functions and fitting parameter to data:

From equation (3), we can predict the amount of pathology $\vec{x}$ at any given time t for a given set of parameters Θ. This prediction can be done either numerically (using MATLAB's ODE solvers) or analytically (if $\vec{s}_a = \vec{0}$, then the ODEs can be solved via matrix exponentiation). In practice, we typically have observed data $\vec{y}$ at times $t_1, \ldots, t_N$ and we try to find optimal parameters Θ that will minimize the discrepancies between $\vec{y}(t_i)$ and $\vec{x}(t_i)$. Our Nex*is*:microglia code provides multiple choices of cost functions: (a) Mean Square Error, (b) Mean Absolute Error, (c) Pearson R, and (d) Lin R. For the particular datasets used in this study, the lin. R cost function provided the overall best solutions, although this may vary from problem to problem. The vector of initial pathology, $\mathbf{x_0}$, also merits further explanation. For all studies, we create a binary vector, $\mathbf{x}_{\text{bin}}$, which either indicates the regions in which pathology was injected or was present at the earliest time point. Since the scale of the pathology at the time of injection is unknown and is unrelated to the scale of $\mathbf{x}_{\text{bin}}$, we introduce a parameter, γ, to be fit from the following relation:

$$\mathbf{x_0} = \gamma \cdot \mathbf{x}_{\text{bin}} \, .$$

4. Strategies for parameter consistency

The user should provide plausible lower and upper bounds for all parameters in the model. For instance, the user may first optimize all parameters for the Nex*is*:global model followed by determining the fit of the model after 100 iterations of bootstrapping with 80% resampling of the regions quantified by the study being used. For fitting the Nex*is*:microglia, the mean and 95% confidence intervals determined for the Nex*is*:global parameters α, β, and γ are used as the initial guess and bounds for those parameters, respectively.

5. Pseudocode for Nex*is*:global and Nex*is*:microglia

All model fits presented in the current study were determined by averaging the parameters across 100 iterations of bootstrapping with 80% resampling of the regions quantified by Kaufman, *et al.* When fitting the Nex*is*:microglia model, we used the mean and 95% confidence intervals determined for the Nex*is*:global model parameters α, β, and γ for each dataset as the initial guess and bounds for those parameters, respectively; this helps to minimize the risk of finding degenerate solutions. We describe the procedure more fully in Algorithms 1 and 2 below.

6. Gene selection for Nex*is*:microglia

Based on literature survey we selected eight microglial homeostasis genes (*P2ry12*, *Cx3cr1*, *Fcrls*, *Olfml3*, *Hexb*, *Siglech*, *Sox5*, and *Jun*) as well as *Trem2*, a surface-marker of activated microglia. For convenience of reading, in the remainder of the document, the

model name 'Nex*is*:microglia' will be replaced by Nex*is*:homeostasis in the case of homeostatic microglia or Nex*is*:(*microglial gene name*) in the case of a particular microglial gene being investigated.

The Nex*is*:microglia models were constructed using either the individual microglial gene *Trem2*, the first principal component of all above-mentioned genes, or the first principal component of only the microglial homoestatic signature genes. Individual effect of another important surface-marker of activated microglia, *Cd33*, was not investigated since its effect size is much smaller than *Trem2*. We then compared model fits of the Nex*is*:global to those of the Nex*is*:microglia models using the following metrics: AIC, BIC, and Pearson R-squared of observed and predicted pathology spread collapsed across time-points. We also explored the effect of *Apoe* on the model parameters, since *Apoe*, although not exclusively microglial, is an important AD-risk gene. Since the performance of the model with *Apoe* was worse compared to Nex*is*:*Trem2* and comparable to the Nex*is*:global model fit, we do not include its results in the main manuscript but provide them in this SI instead.

7. Model parameters and fit for *Apoe*

We followed the same strategy of Nex*is*:microglia model fitting and model comparison by using *Apoe* as the target microglial gene. Tables S1, S2, and S3 below include model parameters and fit indices for Nex*is*:global, Nex*is*:*Trem2*, and Nex*is*:*Apoe* for all three datasets, with the intention of comparing Nex*is*:*Apoe* model parameters and fits to the former two. Figure S1 depicts scatter-plots of observed versus modeled pathology using *Apoe* as well as bootstrapped parameters of the Nex*is*:*Apoe* model for dataset DS9.

Algorithm 1: Nexis:global Parameter Fitting

Data: $\mathbf{x_i}$, regional tau pathology for dataset i

Result: $\{\alpha^*_{\text{Nex}is\text{:global}}, \beta^*_{\text{Nex}is\text{:global}}, \gamma^*_{\text{Nex}is\text{:global}}\}$, the optimal set of Nex$is$:global parameters, and $\tilde{\mathbf{x}}_\mathbf{i}^{\text{Nex}is\text{:global}}$, the predicted tau pathology for dataset i under the Nexis:global.

Function `Nexis:global(`$\mathbf{x_{seed}}$,t,α,β,γ`):`

  $\tilde{\mathbf{x}}(t) \leftarrow$ **solve** Equation [1] with $t = t$, $\alpha = \alpha$, $\beta = \beta$, $\gamma = \gamma$, $b = 0$, $p = 0$, $\mathbf{x_0} = \gamma \cdot \mathbf{x_{seed}}$;

  return $\tilde{\mathbf{x}}(t)$;

End Function;

$\alpha_0 \leftarrow 0.5$;

$\beta_0 \leftarrow 1$;

$\gamma_0 \leftarrow \frac{\sum \mathbf{y_i}(t_1)}{n_{seed}}$;

$\mathbf{x_{seed}} \leftarrow$ binary seed vector for dataset i;

$M_i \leftarrow$ the set of quantified regions in dataset i;

$j \leftarrow 1$;

while $j \leq 100$ **do**

  $M_i^j \leftarrow$ the j^{th} random subset of M_i with cardinality $0.8 \times |M_i|$;

  $\mathbf{x_i^j} \leftarrow \mathbf{x_i}|_{M_i^j}$;

  $\alpha^j, \beta^j, \gamma^j \leftarrow \underset{\alpha,\beta,\gamma}{\text{argmin}} \sum_t (1 - R_C(\mathbf{x_i^j}(t), \text{Nex}is\text{:global}(\mathbf{x_{seed}}, t, \alpha, \beta, \gamma))$ initialized at $\alpha = \alpha_0$, $\beta = \beta_0$, $\gamma = \gamma_0$;

  $j = j + 1$;

end

$\alpha^*_{\text{Nex}is\text{:global}} \leftarrow \sum_j \alpha^j/100$;

$\beta^*_{\text{Nex}is\text{:global}} \leftarrow \sum_j \beta^j/100$;

$\gamma^*_{\text{Nex}is\text{:global}} \leftarrow \sum_j \gamma^j/100$;

$\tilde{\mathbf{x}}_\mathbf{i}^{\text{Nex}is\text{:global}} \leftarrow \text{Nex}is\text{:global}(\mathbf{x_{seed}}, t, \alpha^*_{\text{Nex}is\text{:global}}, \beta^*_{\text{Nex}is\text{:global}}, \gamma^*_{\text{Nex}is\text{:global}})$;

Here $R_C(x, y)$ is the Lin's concordance correlation coefficient between vectors x and y. We find that our heuristic choice of initial α, β, and γ does not impact the results of Algorithm 2.

Algorithm 2: Nex*is*:microglia Parameter Fitting

Data: $\mathbf{x_i}$, regional tau pathology for dataset i

Result: $\{\alpha^*_{\text{Nex}is\text{:microglia}}, \beta^*_{\text{Nex}is\text{:microglia}}, \gamma^*_{\text{Nex}is\text{:microglia}}\}$, the optimal set of Nex*is*:microglia parameters, and $\tilde{\mathbf{x}}_\mathbf{i}^{\text{Nex}is\text{:microglia}}$, the predicted tau pathology for dataset i under the Nex*is*:microglia.

Function `Nexis:microglia(`$\mathbf{x_{seed}}$`,`t`,`α`,`β`,`γ`):`

  $\tilde{\mathbf{x}}(t) \leftarrow$ **solve** Equation [1] with $t = t$, $\alpha = \alpha$, $\beta = \beta$, $\gamma = \gamma$, $b = b$, $p = p$, $\mathbf{x_0} = \gamma \cdot \mathbf{x_{seed}}$;

  return $\tilde{\mathbf{x}}(t)$;

End Function;

$\alpha_0 \leftarrow 0.5$;

$\beta_0 \leftarrow 1$;

$\gamma_0 \leftarrow \frac{\sum \mathbf{y_i}(t_1)}{n_{seed}}$;

$\mathbf{x_{seed}} \leftarrow$ binary seed vector for dataset i;

$M_i \leftarrow$ the set of quantified regions in dataset i;

$j \leftarrow 1$;

while $j \leq 100$ **do**

  $M_i^j \leftarrow$ the j^{th} random subset of M_i with cardinality $0.8 \times |M_i|$;

  $\mathbf{x_i^j} \leftarrow \mathbf{x_i}|_{M_i^j}$;

  $\alpha^j, \beta^j, \gamma^j \leftarrow \underset{\alpha,\beta,\gamma}{\text{argmin}} \sum_t (1 - R_C(\mathbf{x_i^j}(t))$;

  $j = j + 1$;

end

Algorithm 1 is called to obtain initial guesses for α, β, and γ

$\alpha^*_{\text{Nex}is\text{:microglia}} \leftarrow \sum_j \alpha^j/100$;

$\beta^*_{\text{Nex}is\text{:microglia}} \leftarrow \sum_j \beta^j/100$;

$\gamma^*_{\text{Nex}is\text{:microglia}} \leftarrow \sum_j \gamma^j/100$;

$\tilde{\mathbf{x}}_\mathbf{i}^{\text{Nex}is\text{:microglia}} \leftarrow \text{Nex}is\text{:microglia}(\mathbf{x_{seed}}, t, \alpha^*_{\text{Nex}is\text{:microglia}}, \beta^*_{\text{Nex}is\text{:microglia}}, \gamma^*_{\text{Nex}is\text{:microglia}})$;

Table S1: **Comparison of model fits for Nex*is*:global and Nex*is*:microglia for dataset DS4.** The Nex*is*:microglia fits using *Trem2* and *Apoe* are compared to Nex*is*:global fits in dataset DS4 using Fisher's R-to-Z test. AIC: Akaike's Information Criteria, BIC: Bayesian Information Criteria, R^2: model fit parameter, α: intra-regional tau accumulation rate; β: inter-regional tau spread rate; p: microglia-mediated intra-regional tau accumulation rate; b: microglia-mediated inter-regional tau spread rate. The parameters are presented as mean with 95% confidence intervals. *** $p = 0.000$, ** $p < 0.005$.

Dataset	**DS4**		
Models compared	**Nex*is*:global**	**Nex*is*:*Trem2***	**Nex*is*:*Apoe***
Model fit indices			
$\boldsymbol{R^2}$	0.17	0.20	0.18
AIC	-3030	-3047	-3031
BIC	-3012	-3021	-3005
Fitted parameters			
$\boldsymbol{\alpha}$	0.25 [0.05 - 0.30]	0.18 [0.12 - 1.41]	0.32 [0.27 - 0.32]
$\boldsymbol{\beta}$	1.89 [0.05 - 3.84]	1.41 [1.32 - 2.46]	1.36 [1.32 - 2.46]
$\boldsymbol{p}$	-	0.75 [0.38 - 0.89]	-0.14 [-0.17 - -0.01]
$\boldsymbol{b}$	-	4.30 [0.72 - 5.93]	0.75 [-0.69 - 1.32]

Table S2: **Comparison of model fits for Nex*is*:global and Nex*is*:microglia for dataset DS6.** The Nex*is*:microglia fits using *Trem2* and *Apoe* are compared to Nex*is*:global fits in dataset DS6 using Fisher's R-to-Z test. AIC: Akaike's Information Criteria, BIC: Bayesian Information Criteria, R^2: model fit parameter, α: intra-regional tau accumulation rate; β: inter-regional tau spread rate; p: microglia-mediated intra-regional tau accumulation rate; b: microglia-mediated inter-regional tau spread rate. The parameters are presented as mean with 95% confidence intervals. *** $p = 0.000$, ** $p < 0.005$.

Dataset	**DS6**		
Models compared	**Nex*is*:global**	**Nex*is*:*Trem2***	**Nex*is*:*Apoe***
Model fit indices			
$\boldsymbol{R^2}$	0.47	0.59***	0.50
AIC	-3250	-3404	-3289
BIC	-3233	-3379	-3263
Fitted parameters			
$\boldsymbol{\alpha}$	0.21 [0.14 - 0.23]	0.26 [0.22 - 0.27]	0.27 [0.27 - 0.27]
$\boldsymbol{\beta}$	2.09 [0.29 - 2.61]	1.46 [1.46 - 1.46]	1.46 [1.46 - 1.46]
$\boldsymbol{p}$	-	-0.35 [-0.49 - -0.06]	-0.14 [-0.17 - -0.09]
$\boldsymbol{b}$	-	0.57 [-0.24 - 1.66]	1.17 [0.38 - 1.54]

Table S3: **Comparison of model fits for Nex*is*:global and Nex*is*:microglia for dataset DS9.** The Nex*is*:microglia fits using *Trem2* and *Apoe* are compared to Nex*is*:global fits in dataset DS9 using Fisher's R-to-Z test. AIC: Akaike's Information Criteria, BIC: Bayesian Information Criteria, R^2: model fit parameter, α: intra-regional tau accumulation rate; β: inter-regional tau spread rate; p: microglia-mediated intra-regional tau accumulation rate; b: microglia-mediated inter-regional tau spread rate. The parameters are presented as mean with 95% confidence intervals. *** $p = 0.000$, ** $p < 0.005$.

Dataset	**DS9**		
Models compared	**Nex*is*:global**	**Nex*is*:*Trem2***	**Nex*is*:*Apoe***
Model fit indices			
$\boldsymbol{R^2}$	0.44	0.52**	0.46
AIC	-3899	-3981	-3917
BIC	-3881	-3954	-3891
Fitted parameters			
$\boldsymbol{\alpha}$	0.19 [0.16 - 0.20	0.23 [0.19 - 0.24]	0.24 [0.24 - 0.24]
$\boldsymbol{\beta}$	3.01 [2.19 - 3.35]	2.10 [2.10 - 2.10]	2.10 [2.10 - 2.10]
$\boldsymbol{p}$	-	-0.31 [-0.45 - -0.11]	-0.13 [-0.17 - -0.10]
$\boldsymbol{b}$	-	0.41 [-0.21 - 0.92]	0.90 [0.19 - 1.13]

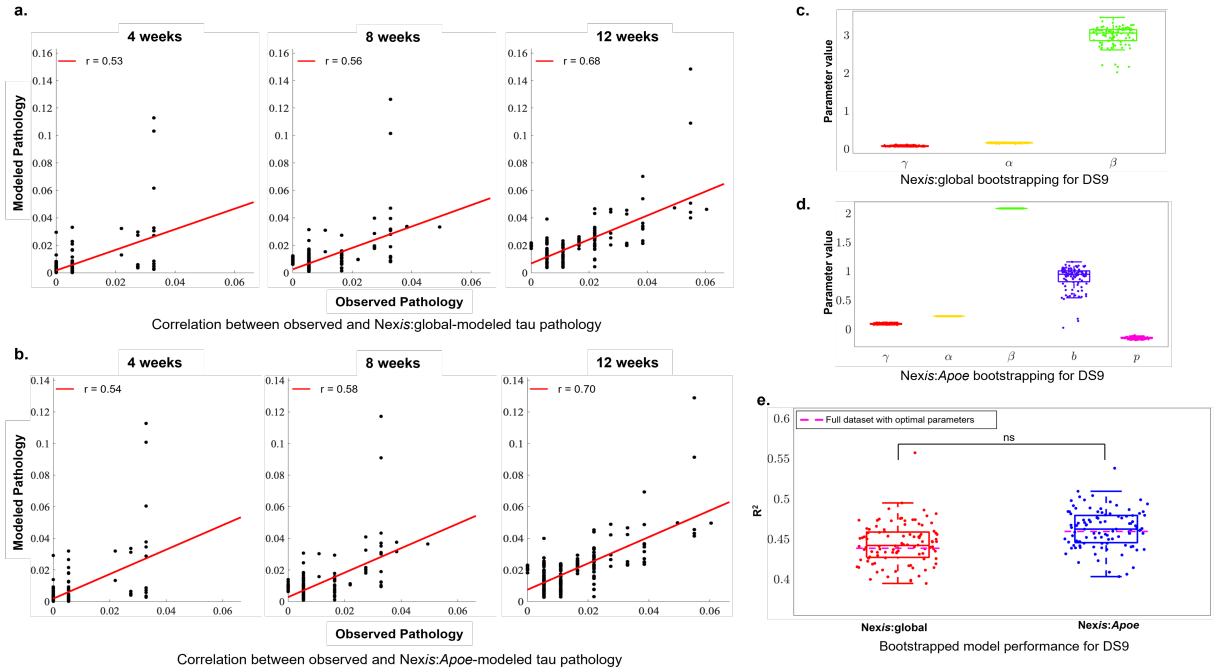

Figure S1: **Nex*is*:global and Nex*is*:*Apoe* model fits for dataset DS9.** Panels a and b depict the Nex*is*:global and Nex*is*:*Apoe* modeled tau pathology progression, respectively, in DS9 dataset at 4, 8, and 12 weeks post hippocampal seeding. Panels c and d depict the 100-iteration 80% resampled bootstrapped values of both Nex*is* model parameters for DS9. Panel e depicts the comparison between model fits of the 80% resampled bootstrapped dataset with the full dataset. Global parameters: γ (seed rescale), α (intra-regional tau accumulation rate), and β (inter-regional tau spread rate). Microglial parameters: p (microglia-mediated intra-regional tau accumulation rate) and b (microglia-mediated inter-regional tau spread rate). Model fit is depicted by R^2.