Supplementary material: Disease Knowledge Transfer across Neurodegenerative Diseases

1 Parameter Estimation

We estimate the model parameters using a two-stage approach. In the first stage, we perform belief propagation within each agnostic unit and then within each disease model. In the second stage we jointly optimise across all agnostic units and disease models using loopy belief propagation. An overview of the algorithm is given in Figure 1. Given the initial parameters estimated from the first stage (line 1), the algorithm continuously updates the biomarker trajectories within the agnostic units (lines 4-5), dysfunction trajectories (line 8) and subject-specific time shifts (line 10) until convergence. The cost function for all parameters is nearly identical, the main difference being the measurements (i, j, k) over subjects i, visits j and biomarkers k that are selected for computing the measurement error. For estimating the trajectory of biomarker k within agnostic unit $\psi(k)$, measurements are taken from Ω_k representing all measurements of biomarker k from all subjects and visits. For estimating the dysfunction trajectories, $\Omega_{d,l}$ represents the measurement indices from all subjects with disease d (i.e. $d_i = d$) and all biomarkers k that belong to agnostic unit l (i.e. $\psi(k) = l$). Finally, Ω_i (line 10) represents all measurements from subject i, for all biomarkers and visits.

1 Initialise $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{\beta}^{(0)}$
2 **while** $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ *not converged* **do**
; // Estimate biomarker trajectories (disease agnostic)
3 **for** $k = 1$ *to* K **do**
4 $\theta_k^{(u)} = \arg\min_{\theta_k} \sum_{(i,j)\in\Omega_k} \left[y_{ijk} - g\left(f(\beta_i^{(u-1)} + m_{ij}; \lambda_{d_i}^{\psi(k),(u-1)}); \theta_k\right)\right]^2 - log\ p(\theta_k)$
5 $\epsilon_k^{(u)} = \frac{1}{|\Omega_k|} \sum_{(i,j)\in\Omega_k} \left[y_{ijk} - g\left(f(\beta_i^{(u-1)} + m_{ij}; \lambda_{d_i}^{\psi(k),(u-1)}); \theta_k^{(u)}\right)\right]^2$
; // Estimate dysfunction trajectories (disease specific)
6 **for** $d = 1 \in \mathbb{D}$ **do**
7 **for** $l = 1 \in \Lambda$ **do**
8 $\lambda_d^{l,(u)} = \arg\min_{\lambda_d^l} \sum_{(i,j,k)\in\Omega_{d,l}} \left[y_{ijk} - g\left(f(\beta_i^{(u-1)} + m_{ij}; \lambda_d^l); \theta_k^{(u)}\right)\right]^2 - log\ p(\lambda_d^l)$
; // Estimate subject-specific time shifts
9 **for** $i = 1 \in [1, \ldots, S]$ **do**
10 $\beta_i^{(u)} = \arg\min_{\beta_i} \sum_{(j,k)\in\Omega_i} \left[y_{ijk} - g\left(f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k),(u)}); \theta_k^{(u)}\right)\right]^2 - log\ p(\beta_i)$

Fig. 1: The algorithm used to estimate the DKT parameters, based on loopy belief-propagation.

2 Generation of synthetic dataset

We tested DKT on synthetic data, to assess its performance against known ground truth. More precisely, we generated data that follows the DKT model exactly, and tested DKT's ability to recover biomarker trajectories and subject time-shifts.

We generated the synthetic data as follows, using parameters from Table 1:

- We simulate two synthetic diseases, "synthetic PCA" and "synthetic AD"
- We define 6 biomarkers that we allocate to agnostic units l_0 and l_1 (Table 1 top)
- Within each agnostic unit, we define the parameters $theta_0$ – $theta_5$ corresponding to biomarker trajectories within the agnostic unit.
- For each disease, we define the parameters λ corresponding to trajectories of dysfunction scores.
- We then sample data from 100 synthetic AD subjects and 50 PCA subjects with β_i as given in Table 1 bottom using the model likelihood (Eq. 2 from main paper). For each subject, we generate data for 4 visits, each 1 year apart.

	Trajectory parameters
Biomarker allocation	$l_0 : \{k_0, k_2, k_4\}$, $l_1 : \{k_1, k_3, k_5\}$
Agnostic unit l_0	$\theta_0 = (1, 5, 0.2, 0)$, $\theta_2 = (1, 5, 0.55, 0)$, $\theta_4 = (1, 5, 0.9, 0)$
Agnostic unit l_1	$\theta_1 = (1, 10, 0.2, 0)$, $\theta_3 = (1, 10, 0.55, 0)$, $\theta_5 = (1, 10, 0.9, 0)$
"Synthetic AD"	$\lambda_0^0 = (1, 0.3, -4, 0)$ and $\lambda_0^1 = (1, 0.2, 6, 0)$
"Synthetic PCA"	$\lambda_1^0 = (1, 0.3, 6, 0)$ and $\lambda_1^1 = (1, 0.2, -4, 0)$
	Subject parameters
Number of subjects	100 (synthetic AD) and 50 (synthetic PCA)
Time-shifts β_i	$\beta_i \sim U(-13, 10)$ years
Diagnosis	$p(control) \propto Exp(-4.5)$, $p(patient) \propto Exp(4.5)$
Data generation	4 visits/subject, 1 year apart, $\epsilon_k = 0.05$

Table 1: Parameters used for synthetic data generation, emulating the TADPOLE and DRC datasets.

3 Demographics of test sets

The cohort from the Dementia Research Center UK used for validation had the following demongraphics: 10 subjects diagnosed with Posterior Cortical Atrophy, with a mean age of 59.4, 40% females, as well as 10 age-matched controls with a mean age of 59.3, 50% females.

When splitting the TADPOLE subjects into three populations with different progressions, the resulting groups had the following demographics:

Cohort	Nr. subjects	Nr. visits	Age (baseline)	Gender (%F)
Controls (Hippocampal)	31	2.3 ± 1.8	74.4 ± 6.9	38%
AD (Hippocampal)	21	1.5 ± 0.8	74.5 ± 5.5	42%
Controls (cortical)	21	2.3 ± 1.3	70.9 ± 5.4	42%
AD (cortical)	35	1.7 ± 0.9	72.8 ± 7.4	28%
Controls (subcortical)	28	3.0 ± 1.5	73.7 ± 6.5	42%
AD (subcortical)	27	1.6 ± 0.9	73.7 ± 7.5	33%

Table 2: Demo

References