1 Supplemental methods

1.1 Datasets

For all datasets we only extracted data for individuals with more than one time point containing non-missing data. For AIBL and ADNI1 age was approximated based on birth year, or both birth year and month respectively.

1.1.1 ADNI1

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu; [3]). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see *http://www.adni-info.org*.

1.1.2 AIBL

AIBL study methodology has been reported previously [21]. Longitudinal MMSE Data was collected by the AIBL group and was obtained from the AIBL database:

http://www.aibl.csiro.au/adni/index.html

1.1.3 CAMD (C-1013 and C-1014)

Data used in the preparation of this article were obtained from the Coalition Against Major Diseases (CAMD) database (*http://c-path.org/programs/camd/*; [4]). The CAMD database provides data only for the placebo arm of AD trials, and provides no further information about the trials or individuals diagnosis at first visit. To ensure sufficient follow-up, we only

used data from the two longest running clinic trials in the CAMD database (trials C-1013 and C-1014).

In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the CAMD. The Coalition brings together patient groups, biopharmaceutical companies, and scientists from academia, the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). The CAMD includes over 200 scientists from member and non-member organizations. The data available in the CAMD database has been volunteered by CAMD member companies and non-member organizations.

1.2 Notation

A dataset contains information on individuals $i \in \{1, 2, \ldots, n\}$. The set of time points at which individual i has been observed is denoted $\boldsymbol{\tau}^i := (\tau^i_1, \ldots, \tau^i_{\#TP(i)})$. The longitudinal univariate observations (in this example MMSE score) for individual i at time t are given by $x_i(t)$, which is only observed for $t \in \boldsymbol{\tau}^i$ (Figure 1a). $\phi(t; \boldsymbol{\theta})$ is a proposed (parametrised) trajectory (Supplementary Methods), which along with an individual's offset δ_i is believed to explain the data, i.e. $x_i(t) \approx \phi(t + \delta_i; \boldsymbol{\theta})$. Define an individual's estimated Disease Time ($\hat{DT}$) at time t as: $\hat{DT}(i, t) := t + \delta_i$ (Figure 1b).

1.3 Estimating disease time

Given $\phi(t, \boldsymbol{\theta})$, $\boldsymbol{\tau}^i$ and $x_i(t)$, for each i define a local loss function using sum of squares:

$$local\_loss(i, \delta_i, \boldsymbol{\theta}) := \| x_i(\boldsymbol{\tau}^i) - \phi\left(\boldsymbol{\tau}^i + \delta_i; \boldsymbol{\theta}\right) \|_2^2 \tag{1}$$

Using this we estimate the 'optimal' offset ($\hat{\delta}_i$), which determines estimated disease time,

using ordinary least squares:

$$\hat{\delta}_i(\boldsymbol{\theta}) := \underset{\delta_i}{\operatorname{argmin}}\ local\_loss(i, \delta_i, \boldsymbol{\theta}) \tag{2}$$

This can be minimised (locally) using the *optimise* function.

1.3.1 Estimating trajectory parameters

For a set of parameters $\boldsymbol{\theta}$ we define a loss function over all individuals:

$$loss(\boldsymbol{\theta}) := \sum_i local\_loss(i, \hat{\delta}_i(\boldsymbol{\theta}), \boldsymbol{\theta}) \tag{3}$$

This involves first estimating $\hat{\delta}_i(\boldsymbol{\theta})$ for all individuals i as above. The $\hat{\boldsymbol{\theta}}$ that minimises *loss* is the one that 'best' fits the data. We can estimate a $\hat{\boldsymbol{\theta}}$ leading to a local minimum in loss using the *optim* function.

1.3.2 Trajectory models

To fit this model we need a parametrised trajectory model $\phi(t, \boldsymbol{\theta})$. In this study we used compared two alternatives:

(a) a sigmoidal trajectory (also used in [13] and similar to that used in [2]).

$$\phi(t; \boldsymbol{\theta}) = \frac{\theta_1}{1 + exp(-\theta_2 t)} + \theta_3 \tag{4}$$

(b) an exponential decline trajectory (similar to that used in [12])

$$\phi(t; \boldsymbol{\theta}) = \theta_1 - exp(\theta_2 t) \tag{5}$$

1.4 Temporal Clustering algorithm

Let K be the number of clusters, and $k \in \{1, \ldots, K\}$ represent an individual cluster. Let $c(i) : \{1, \ldots, N\} \rightarrow \{1, \ldots, K\}$ be the cluster assignment function mapping individuals to clusters. Let $\boldsymbol{\theta}^k$ be the parameters of the model representing the k^{th} cluster (Figure 1b) and $\boldsymbol{\theta}^k_{-1}$ be the all except the first of the trajectory parameters, i.e. excluding θ_1. To perform clustering we introduce an algorithm based on K-means and our trajectory modelling approach (Algorithm 1).

Algorithm 1 Temporal Clustering

Choose number of clusters K.
1). First fix θ_1 by estimating it across all individuals by minimising Equation 3 with respect to $\boldsymbol{\theta}$.
2). Randomly assign each individual to a cluster,
i.e. sample $c(i)$ for all i randomly with replacement from $\{1, \ldots, K\}$.

3). Repeat the following steps up to convergence, defined as no change in $c(i)$.

- For each cluster k, minimise Equation 3 with respect to $\boldsymbol{\theta}^k_{-1}$ using only data from individuals for whom $c(i) = k$.
- For each cluster k and individual i, use $\boldsymbol{\theta}^k$ and the estimated individual offset $\hat{\delta}_i(\boldsymbol{\theta}^k)$ to calculate the $local\_loss$ using Equation 1. Reassign each individual i to the cluster k'_i which results in the smallest $local\_loss$, i.e. $c(i) \leftarrow k'_i$.

For supplementary simulation results only we also used a version of Temporal Clustering without a fixed baseline (θ_1), i.e. a separate baseline version that allowed different clusters to have different baseline (θ_1).

Free parameters in Temporal Clustering include the allowed range for δ (default: $\pm$ 50,000 days), the initial guess at trajectory parameters ($\boldsymbol{\theta}$; default: (28 , 4 $\times 10^{-4}$)), the allowed range for θ_2 if a two parameter trajectory model is being used (default: 0 – 0.005), the number of clusters (K; here fixed at 2) and the maximum number of iterations (default: 30).

1.5 Simulation study

Two types of simulation study were performed, a disease time estimation simulation (equivalent to $K = 1$) and a Temporal Clustering simulation (i.e. $K = 2$). We performed simulations based on both the combined dataset, i.e. with the same number of individuals, similar time points and with longitudinal MMSE data resembling the real data when plotted.

First, vectors of true $\boldsymbol{\delta}$ were uniformly sampled from an interval chosen to generate visually plausible data. The interval chosen depended on the value of true θ_2 (Table S4). To generate time points simulated individuals were randomly assigned the time points of a real individual in such a way that the last time point would have MMSE > 0 (before error was added). This simulated missing data due to 'death' at MMSE > 0, i.e. no individual had data with MMSE < 0, simulating death at MMSE equal to zero. To minimise duplicates of time points used, individuals were sampled without replacement, unless no remaining individual had a last time point with MMSE > 0 in which case they were sampled from the full set of individuals. In other words, almost all the simulated individuals will have different timepoints, but a very small number of simulated individuals will have the same timepoints but different data.

Data was generated based on our model, i.e.

$$x_i(\boldsymbol{\tau}^i) = \phi(\boldsymbol{\tau}^i + \delta_i; \boldsymbol{\theta}) + \boldsymbol{\epsilon}_{i\boldsymbol{\tau}^i} \tag{6}$$

Where $\boldsymbol{\epsilon}_{i\boldsymbol{\tau}^i} := \{\epsilon_{i\tau_1^1}, \ldots, \epsilon_{i\tau^1_{\#TP(i)}}\}$ and $\epsilon_{it} \sim N(0, 1.5)$ i.i.d.

For all the simulations true θ_1 was set to 29. For both the disease time estimation and Temporal Clustering simulations a range of five θ_2 were used to test the methods sensitivity to this parameter. For each θ_2 100 simulations were performed. For the disease time estimation we used $\theta_2 \times 10^4 = 1,\ 3,\ 5,\ 7$ and 9. For the Temporal Clustering simulation we used $\theta_2^1 \times 10^4 = 5, 3, 5, 6, 8$, and $\theta_2^1 \times 10^4 = 5, 1, 1, 10, 10$.

The simulation with $\theta_2^1 = \theta_2^2 = 5 \times 10^4$ was used to test what would happen if there was only one cluster, but the user attempted to find two. This simulation was only used to study trajectory parameter estimation, and was not used in the clustering accuracy calculations as the cluster labels would have been meaningless.

To study the impact of various biases we investigated three different data transformations that led to progressively more realistic and biased data in the disease time estimation simulation . The first transformation simulated a floor effect by rounding the few cases where MMSE < 0, due to the addition of noise, up to zero. The second also rounded each $x_i(t)$ to its nearest integer, as MMSE can only take integer values. The third additionally imposed a ceiling of MMSE $= 30$, the maximum score from the real MMSE test, i.e. all $x_i > 30$ were set to 30. For the Temporal Clustering simulation featured in Figure 3 we only generated simulated data with all three of these transformations applied.

2 Supplementary results

In a simulation the disease time estimation approach worked best with a δ range of $\pm$ 50,000 days (Figures S2 - S3), i.e. using this led to trajectory parameter estimates that were closer to their true value, especially when the true rate parameter was low (e.g. $\theta_1 = 1 \times 10^{-4}$; Figures S2a and S3a). This seemingly high number is probably a sign that overfitting of individual offsets (δ_i) are occurring. For more on this see the Discussion.

In a two-cluster simulation the fixed baseline was better than the separate baseline version of Temporal Clustering at avoiding unrealistic θ_1 estimates, i.e. $\theta_1 > 30$ (Figure S4). In cases where accurate rate estimation is the priority it could make sense to use the seperate baseline version, but the fixed baseline version is preferred here as it leads to both trajectory parameters being interpretable, i.e. within realistic bounds.

Supplementary tables

Table S1: Table comparing diagnosis at last visit with either (a) diagnosis at first visit, or (b) clusters. For (a) a mixture of actual (ADNI & AIBL) and estimated (C-1013 and C-1014) diagnoses are used. For (b) only actual diagnoses for individuals in ADNI and AIBL are used, with clusters filtered (discrimination score > 2). MMSE = Mini-Mental State Examination, NC = Normal Cognition, MCI = Mild Cognitive Impairment, AD = Alzheimer's disease.

	Diagnosis at first visit			**Filtered clusters (ADNI & AIBL only)**	
Diagnosis at last visit	NC	MCI	AD	k = 1	k = 2
NC	361	19	0	32	22
MCI	30	476	75	60	43
AD	8	425	1018	99	236

Table S2: Table of delta ranges used to simulate data. Range used depended on θ_2.

$\theta_2 \times 10^4$	Minimum $\delta \times 10^3$	Maximum $\delta \times 10^3$
1	10	30
3	0	10
5	0	6
6	0	5
7	0	4
8	0	3.5
9	-1	3.5
10	-1.5	3

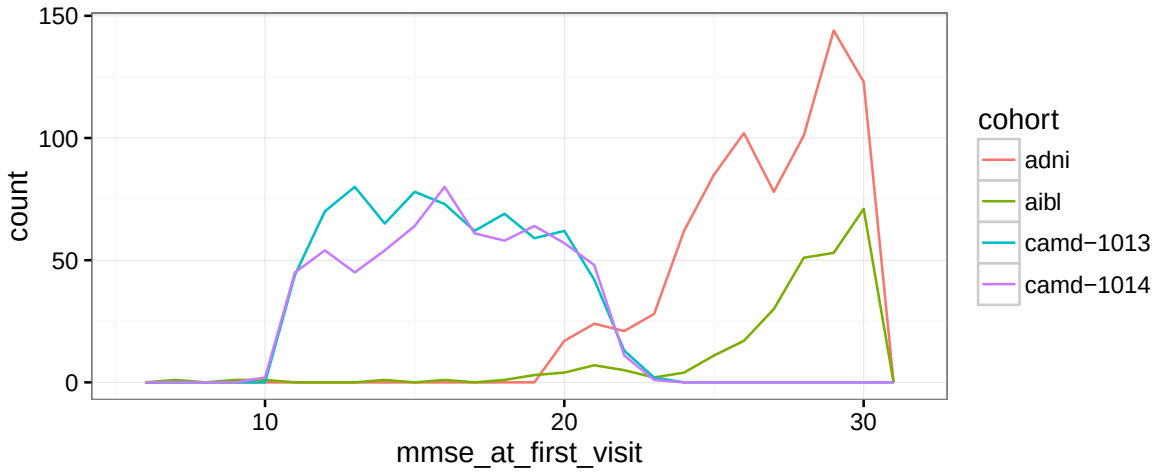
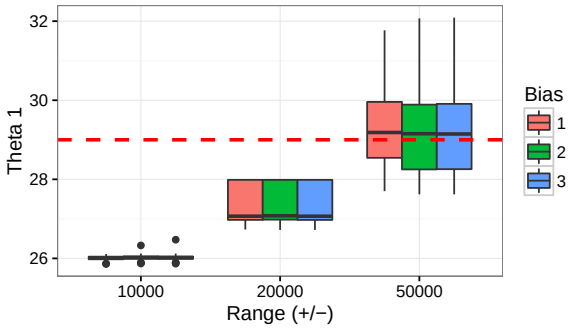
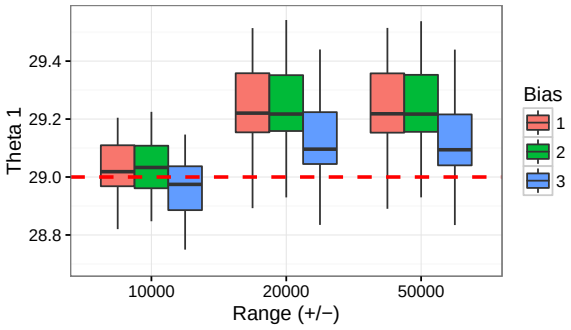


Figure S1: Distribution of MMSE scores at first visit, grouped by cohort.

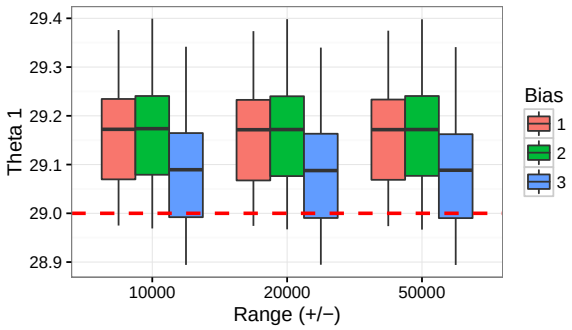
Supplementary Figures



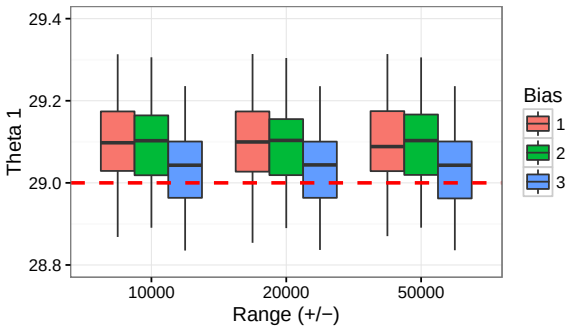
(a) Alignment simulation 1



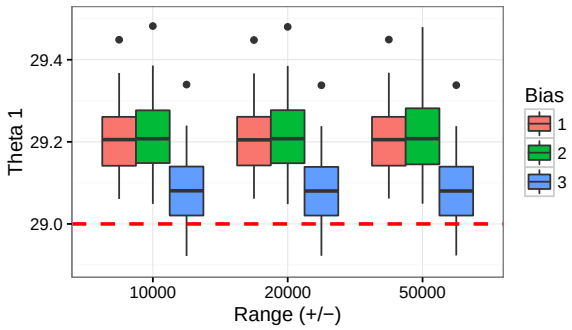
(b) Alignment simulation 2



(c) Alignment simulation 3

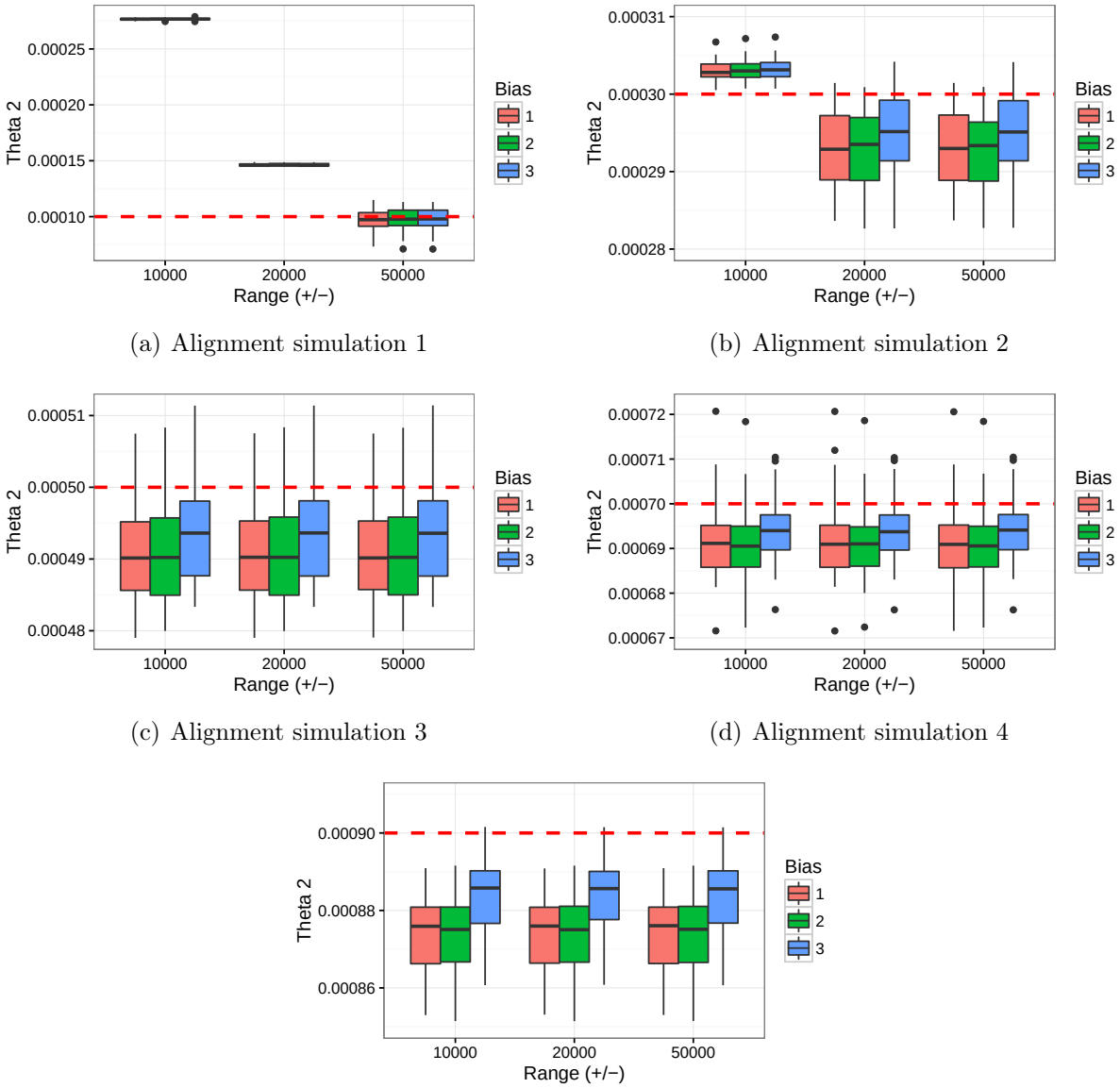


(d) Alignment simulation 4



(e) Alignment simulation 5

Figure S2: Grouped boxplots showing θ_1 estimates from the disease time estimation simulation. Five simulations each use a fixed true $\boldsymbol{\theta}$, with the true θ_1 indicated by dashed red lines. Each simulation differs only in the true value of θ_2: (a) 0.0001, (b) 0.0003, (c) 0.0005, (d) 0.0007 and (e) 0.0009. Simulations are based on the combined cohort, i.e. time points, number of samples and visual appearance. For alignment three different $\boldsymbol{\delta}$ ranges were used: $\pm$ 10000, 20000 or 50000. Three versions of each dataset we generated, each with an increasing amount of bias and plausibility. Bias 1 = Real valued data, but with simulated death (i.e. MMSE < 0 removed). Bias 2 = Integer rounded data with simulated death. Bias 3 = Integer rounded data with simulated death and a ceiling effect (MMSE > 30 set to 30). Outliers excluded in plot for clarity: (b) 2 points ~ 28, (d) 4 points > 30.5 excluded and (e) 1 point > 30.5 excluded.



(a) Alignment simulation 1

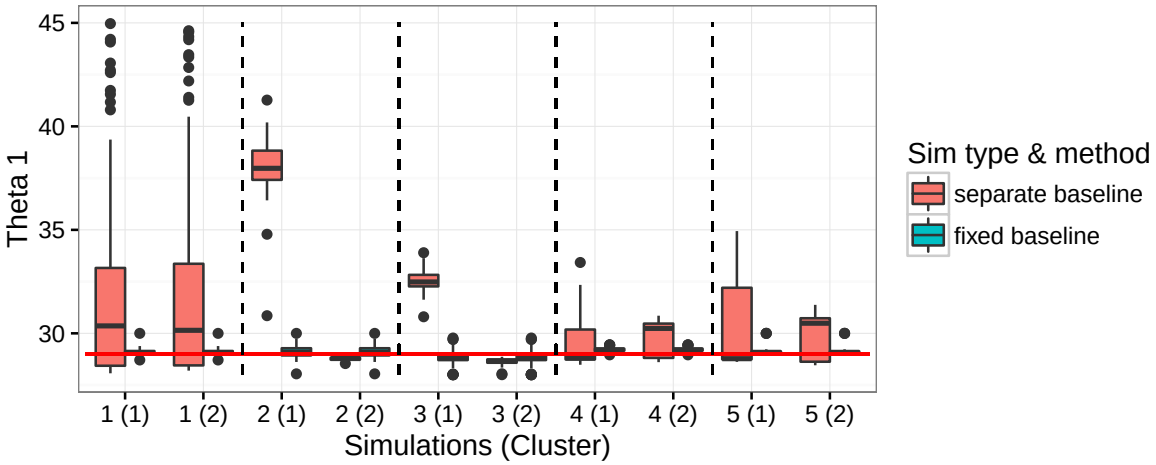
(b) Alignment simulation 2

(c) Alignment simulation 3

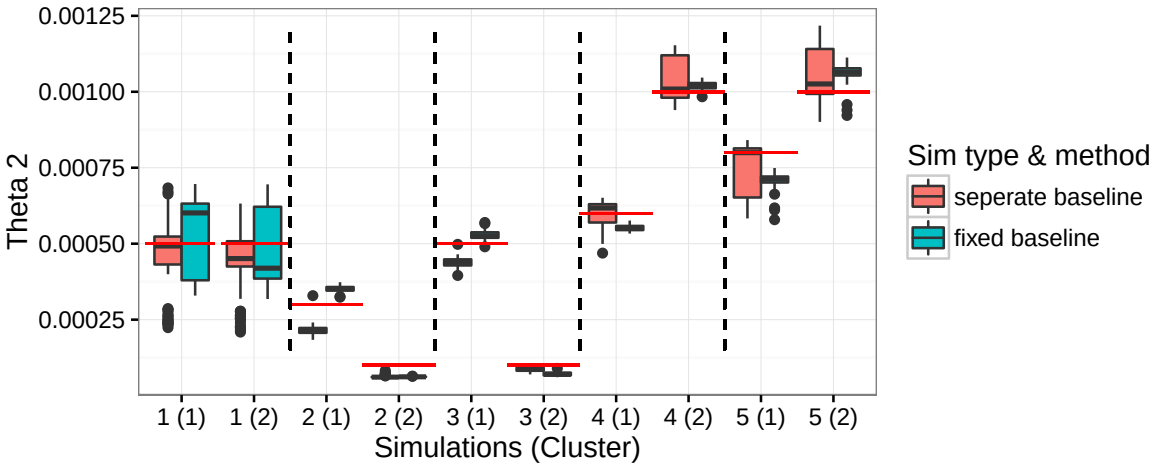
(d) Alignment simulation 4

(e) Alignment simulation 5

Figure S3: Grouped boxplots showing θ_2 estimates from the disease time estimation simulation. Five simulations each use a different fixed true $\boldsymbol{\theta}$. Each simulation differs only in the true value of θ_2, indicated by dashed red lines. Simulations are based on the combined cohort, i.e. time points, number of samples and visual appearance. For alignment three different $\boldsymbol{\delta}$ delta ranges were used: $\pm$ 10000, 20000 or 50000. Three versions of each dataset we generated, each with an increasing amount of bias and plausibility. Bias 1 = Real valued data, but with simulated death (i.e. MMSE < 0 removed). Bias 2 = Integer rounded data with simulated death. Bias 3 = Integer rounded data with simulated death and a ceiling effect (MMSE > 30 set to 30). Outliers excluded in plot for clarity: (b) 2 points > 0.00032, (d) 4 points ~ 0.0006 excluded and (e) 10 point < 0.00085 excluded.

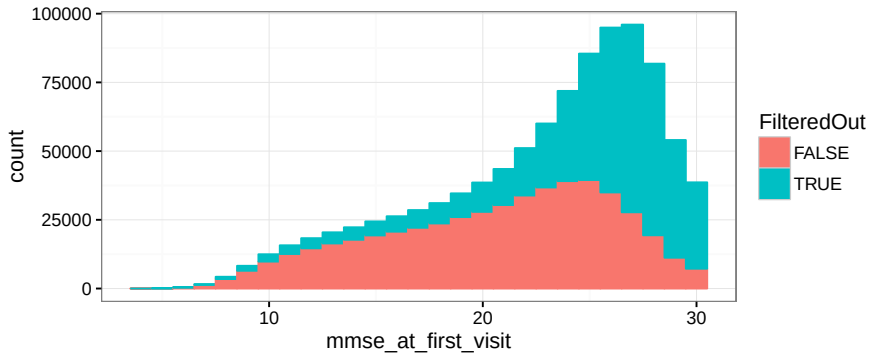


(a) Temporal clustering ($K = 2$) simulation - θ_1 estimates

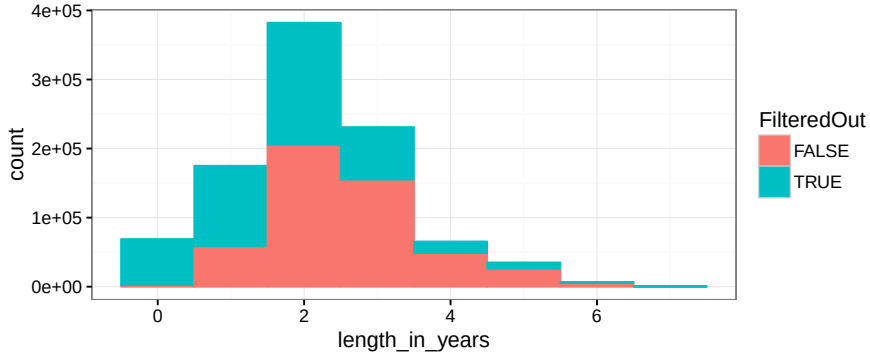


(b) Temporal clustering ($K = 2$) simulation - θ_2 estimates

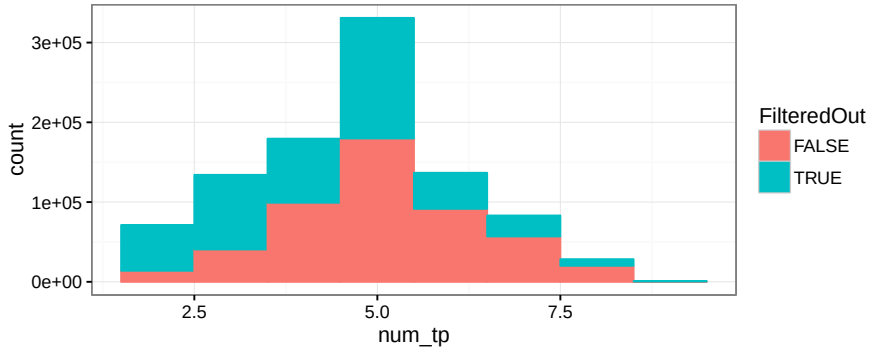
Figure S4: Grouped boxplots showing the result of a simulation study of temporal clustering ($K = 2$). Each simulation differs only in the true value of θ_2^1 and θ_2^2. The results for each simulation are separated by dashed lines. Simulations are based on the combined cohort, i.e. time points, number of samples and visual appearance. True $\boldsymbol{\theta}$ are indicated in red lines. Results are shown for both fixed and separate baseline versions of Temporal Clustering.



(a) MMSE at first visit



(b) Length of follow-up



(c) Number of time points

Figure S5: Characteristics of the simulated individuals removed by a discrimination score filter of two. Stacked histograms are given over 2412 simulated individuals from each 100 iterations of all four simulation types, i.e. different choices of θ_2^1 and θ_2^2 (not including the simulation with $\theta_2^1 = \theta_2^1$). Stacked histograms of the distribution with and without filter of (a) MMSE score at first visit, (b) number of time points and (c) the length of follow-up in days (in year bars).