Heterogeneity of cognitive decline in dementia: taking into account variable time-zero severity

Steven J Kiddle[1,2,*], Alice Parodi[3], Caroline Johnston[1,4], Chris Wallace[2,5,6,+], Richard JB Dobson[1,4,7,*,+], for the Alzheimer's Disease Neuroimaging Initiative (ADNI)[a], for the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL)[b], and for the Coalition Against Major Diseases (CAMD)[c]

[1]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, SE5 8AF.

[2]MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK.

[3]MOX, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133, Milan, Italy.

[4]NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation, London, UK, SE5 8AF

[5]Department of Medicine, Box 157, Level 4, Cambridge Biomedical Campus, University of Cambridge, UK, CB2 0QQ.

[6]JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, NIHR Cambridge Biomedical Research Centre, Cambridge Institute for Medical Research, University of Cambridge, WT/MRC Bldg, Addenbrooke's Hospital, Hills Rd., Cambridge CB2 0XY UK

[7]Farr Institute of Health Informatics Research, UCL Institute of Health Informatics, University College London, London WC1E 6BT, UK.

[a]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of

ADNI investigators can be found at: *http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*

[b]Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (*www.loni.usc.edu/*ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at *www.aibl.csiro.au*

[c]Data used in the preparation of this article were obtained from the Coalition Against Major Diseases database (CAMD). As such, the investigators within CAMD contributed to the design and implementation of the CAMD database and/or provided data, but did not participate in the analysis of the data or the writing of this report.

[+]Joint last authors

[*]To whom correspondence should be addressed:

Address: Steven Kiddle, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Box P092, SGDP Building De Crespigny Park London SE5 8AF.

Email address: steven.kiddle@kcl.ac.uk or richard.j.dobson@kcl.ac.uk

Telephone: 0207 848 0924

Fax: 020 7848 0866

Word count: 4046

Abstract

Understanding heterogeneity in Alzheimer's disease (AD) progression is critically important for the optimal design of trials, allowing participants to be recruited who are correctly diagnosed and who are likely to undergo cognitive decline. Current knowledge about heterogeneity is limited by the paucity of long-term follow-up data and methodological challenges. Of the latter, a key problem is how to choose the most appropriate 'time zero' to use in longitudinal models, a choice which affects results. Rather than a pre-specified 'time zero' we propose a novel methodology – Temporal Clustering – that defines a new 'time zero' using individual offsets inferred from the data. We applied this to longitudinal Mini-Mental State Examination (MMSE), where this approach ensures that individuals have similar estimated MMSE scores at this new 'time zero'. Simulations showed that it could accurately predict cluster membership after the application of a filter. Next we applied it to a cohort of 2412 individuals, with large variability in MMSE score at first visit. Temporal Clustering was used to split individuals into two clusters. The group showing faster decline had higher average levels of AD risk factors: cerebrospinal fluid tau and $APOE\,\epsilon 4$. Cluster membership predicted by Temporal Clustering was less affected by individuals' cognitive ability at first visit than was the case for clusters found using Latent Class Mixture Models. Further application and development of this method will help researchers to identify risk factors affecting cognitive decline.

KEYWORDS: MMSE, trajectory, clustering, dementia, cognitive decline, disease progression, Alzheimer's

1 Introduction

Patient heterogeneity in cognitive decline can be modelled to aid in the design of clinical trials for AD [1]. The current FDA-approved model uses a generalised logistic (i.e. sigmoidal) curve and includes only baseline cognitive ability, age, gender and the $APOE\,\epsilon 4$ genetic risk allele [2]. Extensions to the FDA-approved model taking into account additional factors should allow improved design of clinical trials. Additionally, heterogeneity in cognitive decline may be indicative of dementia type, and may therefore help to identify individuals who have been misdiagnosed.

Unfortunately, our understanding of heterogeneity in cognitive decline and the risk factors that affect it are limited. It is challenging to study because of short-term follow-up of individuals in most existing studies. For example, cohort studies such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [3] and clinical trials such as those in the Coalition Against Major Diseases (CAMD) [4] have relatively short-term follow-up over months or a few years, while the course of cognitive decline in individual patients can take a decade or more [5].

As well as having limited temporal resolution, studies of cognitive decline are further complicated by the fact that individuals can be recruited at different times relative to onset of symptoms. A consequence of this is that it can be hard to choose the most appropriate 'time zero' for longitudinal modelling. Commonly used 'time zero' options include a fixed age, dementia diagnosis, death or first visit, but this choice can have large effects on the results and replicability of analyses [6, 7].

For a given 'time zero', a standard approach to analyse rate of change in cognitive ability is the linear mixed model, which include both fixed effects (i.e. covariates) and random effects. This approach has been applied to identify risk factors of cognitive decline, but has been shown to lead to inflated type 1 error rates and confusion between the effect of risk factors on mean level and rates of change [8].

Extensions to the linear mixed model perform better, but are still limited in their ability to model variability in cognitive ability at 'time zero', however defined. For example, Latent Class Mixture Models (LCCMs) extend mixed models to include 'classes', i.e. groups of individuals who follow similar trajectories over time. These classes are 'latent' (i.e. not known in advance) and are therefore inferred from the data. As such they are a natural framework to summarise heterogeneity in cognitive decline. LCMMs have been used to find clusters of individuals with different trajectories of cognitive decline who differ according to psychotic symptoms [5] or time to a dementia diagnosis [9]. However, in some applications the trajectories appear to differ mostly in individuals' scores at 'time zero' (i.e. disease stage), which is problematic as it may not reflect distinct trajectories (e.g. [5]).

Other researchers have taken a different approach, estimating a random change point that can be seen as an inferred 'time zero' (e.g. [10, 11]). These assume that cognitive decline can be divided into two linear sections, representing normal decline and accelerated decline due to dementia respectively. The change point at which they meet is learned from the data. However, the assumption of linearity before and after the change point is probably unrealistic which may bias results.

Rather than use a change point model, other researchers have assumed sigmoidal (i.e. logistic) or exponential trajectories of cognitive decline, with individual offsets (a random effect) used to align each individual to the trajectory model [12, 13]. These models have been used to study the temporal relationship of different biomarkers, leading to models that are only partially consistent with the theoretical model of Jack Jr et al. [14]. Neither of these approaches has been used to develop a clustering approach to help identify risk factors affecting rates of cognitive decline. Outside of dementia research joint clustering and alignment methods have been introduced [15, 16, 17], but these methods depend on densely sampled regular timepoints, which are not available on cognitive decline.

In this work we aimed to study heterogeneity in cognitive decline using data with large

'time zero' variability, a problem that is particularly acute when studying disease progression using Electronic Health Records [18, 19]. However, to allow others to verify our work we used data from open access traditional cohort studies in this analysis, combined so that the total dataset has large 'time zero' variability. We applied a novel method which combines the benefits of clustering and an inferred 'time zero' to improve the analysis of longitudinal heterogeneity. Further, we compare the ability of Temporal Clustering and LCMMs to generate clusters that are related to AD risk factors. To our knowledge this is the second application of such a method to study disease progression [18], and the first attempt to use this type of approach to study cognitive decline. Furthermore, all analyses are done with open source code to allow our work to be verified and extended by others.

2 Methods

All analysis in this paper has been performed in R [20], scripts are freely available at https://github.com/KHP-Informatics/TC.

2.1 Datasets

We combined data including time series of MMSE from two prospective cohort studies, ADNI1 [3] and the Australian Imaging, Biomarkers and Lifestyle study of aging (AIBL) [21], as well as the placebo arms of the two longest clinical trials from the Coalition Against Major Diseases consortium (CAMD; C-1013 and C-1014 [4]). Studies were combined to make a large dataset, all with short follow-up, but with a lot of variability in baseline cognitive ability. They could be combined as MMSE was recorded consistently across all studies. Individuals were classified as having Normal Cognition (NC), Mild Cognitive Impairment (MCI) or AD. In CAMD, where diagnoses were not given, they were inferred based on MMSE in the following way: MMSE 24 – 30 = NC, MMSE 18 – 23 = MCI, or MMSE 0 –

17 = AD [22]. For more details see Supplementary Methods.

For ADNI and AIBL, genomic DNA was extracted from whole blood with *APOE* genotyped using either TaqMan probes for Single Nucleotide Polymorphisms (rs429358, rs7412) or the *Hha1* restriction enzyme, and assessed using Polymerase Chain Reaction [23, 24]. Levels of total tau in cerebrospinal fluid from ADNI participants have been measured using the xMAP Luminex platform, and were downloaded as 'UPENNBIOMARK2.csv' (http://www.adni-info.org/).

2.2 Temporal Clustering

Full technical details are provided in Supplementary Methods and are summarised here. For Temporal Clustering, a parametrised trajectory curve ($\phi(t;\boldsymbol{\theta})$) has to be supplied, where t is time and $\boldsymbol{\theta}$ is a vector of trajectory parameters. For this application we first used a three parameter sigmoidal trajectory (Supplemental Methods), based on the assumptions of Jack Jr et al. [14], but we found that a two parameter exponential decline curve fitted equally well ($\phi(t;\boldsymbol{\theta}) := \theta_1 - exp(\theta_2 t)$), and we therefore used this for inference as it requires fewer parameters. Here, θ_1 represents the maximum MMSE of the trajectory model and θ_2 is an exponential decline rate.

The Temporal Clustering model consists of trajectory parameters ($\boldsymbol{\theta}^k$) for each cluster (k), and offsets (δ_i) for each individual (i). These offsets are used to shift time points to better align individuals to cluster trajectories, i.e. $\phi(t+\delta_i;\boldsymbol{\theta}^k)$ is the expected MMSE score at timepoint t for individual i in cluster k (Figure 1). Using the exponential decline trajectory, the offset ($\hat{\delta}_i$) for an individual (i) is an estimate of the time between first MMSE assessment and the time at which their MMSE score reached one MMSE point below the maximum MMSE of the model, i.e. an MMSE of $\theta_1 - 1$.

A simplifying assumption of Temporal Clustering is that the baseline parameter (θ_1) takes the same values across all clusters, this was found to be necessary to get good performance

and provide identifiability given short follow-up (Supplementary Results).

Temporal Clustering is based on K-means, a commonly used clustering algorithm. K-means finds clusters by initially allocating all individuals to K clusters at random, and then iterating two steps until the model converges. In K-means step (1) involves calculating a mean point over all cluster members, and step (2) involves re-allocating individuals to the cluster whose mean they are closest to.

The difference between K-means and Temporal Clustering is that rather than calculating cluster means in Step (1), instead trajectory parameters ($\hat{\boldsymbol{\theta}}^k$) are inferred for each cluster (k), along with individual offsets ($\hat{\delta}_i$). This iterates with cluster re-assignment (Step 2) until convergence.

In this study, for simplicity, we will seek to split individuals into a groups of slower and relatively faster decliners (i.e. $K = 2$). A discrimination score was calculated to assess the relative quality of fit of each individual's data to their assigned versus unassigned cluster (Supplemental Methods).

2.3 Simulation study

Full details of the simulation study are available in Supplementary Methods but are briefly summarised here. We performed simulations based on the combined cohort structure, i.e. with the same number of individuals, similar time points and with longitudinal MMSE data resembling the real data when plotted. A range of true parameters $\boldsymbol{\theta}$ were used, with 100 simulations performed for each set. For each simulation true offsets δ_i were generated at random, and timepoints assigned to simulate missing data due to death (i.e. missing data when MMSE ≤ 0). The accuracy of clustering was assessed using the Adjusted Rand Index (ARI), which takes values from zero (i.e. no better than chance) to one (i.e. perfect clustering). ARI was calculated using the R package 'mclust'.

2.4 Latent Class Mixture Models

LCMMs were performed using the 'lcmm' package using spline models with default parameters and no covariates (i.e. no fixed effects). 'Time zero' was chosen to be either first visit or 50th birthday. As in Lima-Proust et al. [9] LCMMs were run on normMMSE, that is MMSE scores transformed to have better metric properties [25]. For comparison we also ran LCMM on raw MMSE scores.

2.5 Statistical analysis

Time was coded as days since first visit. We analysed the simulated and real datasets using Temporal Clustering.

Logistic regression was used to build multivariate models of cluster membership using the R package 'glm'. For regression models baseline MMSE points were recoded in units of ten (i.e. divided by 10), and cerebrospinal fluid tau was recoded in units of 100 pg/ml (i.e. divided by 100). Gender, cohort and the number of *APOE* $\epsilon 4$ alleles were coded as categorial variables. CAMD individuals with an age > 89 years have had their age recorded as 999 to reduce identifiability, these 13 individuals were excluded from regression analyses. Individuals with missing *APOE* or cerebrospinal fluid tau data were excluded from the relevant analyses.

'MMSE at first visit' analysis included covariates: age at first visit, gender and cohort. *APOE* analyses were performed on ADNI1 and AIBL individuals and included co-variates gender and cohort as well as age and MMSE at first visit. Cerebrospinal fluid tau analyses included co-variates gender, and the number of *APOE* $\epsilon 4$ alleles as well as age and MMSE at first visit.

3 Results

Demographic information on the individual and combined cohorts are provided in Table 1. The distribution of MMSE differs between these cohorts, generally individuals in the CAMD have lower MMSE scores at first visit (Figure S1). As was our intention, combining these cohorts therefore results in a large variability in MMSE at first visit (interquartile range (15 – 27), Figure 2(a)).

Longitudinal MMSE for individuals with first visit as 'time zero' is shown in Figure 2(b), where a large amount of heterogeneity can be seen. The 2,412 individuals in the combined cohort were followed-up over a median of 1.9 years at a median of 5 visits. At first visit 42% had AD (diagnosed or inferred), whereas at the last visit 60% did (Table 1 and Table S1).

3.1 Accuracy of clustering: a simulation study

After using simulated data to choose default values for Temporal Clustering parameters (Supplementary Results), we then used it to assess clustering accuracy, i.e. to assess whether Temporal Clustering could accurately distinguish between a slower and faster group of cognitive decliners. With no filter applied, the clustering result was only slightly better than would be expected by chance (ARI 0.12; Figure 3).

Clustering accuracy increased as individuals were filtered out on the basis of the discrimination score, a measure of confidence in assignment of the individual to a cluster (Figure 3). A tradeoff can clearly be seen where the more stringent the filter (i.e. the higher the threshold) the higher the clustering accuracy and the lower the number of individuals left after filtering. In our application the risk factors of interest, *APOE* and cerebrospinal fluid tau, are only recorded for 43% and 17% of individuals respectively and so we prioritised retaining individuals above clustering accuracy. For this reason we selected a discrimination score threshold of 2, which led to a median clustering accuracy (ARI) of 0.31 but retained

over half the individuals (median 54%). A much more stringent threshold of 10 led to a clustering accuracy (ARI) of 0.80, but retained fewer than a fifth of individuals on average.

The most obvious characteristics of simulated individuals removed by a discrimination score filter of 2 were that they had higher MMSE scores at first visit (Figure S2a) and/or less than a year of follow-up (Figure S2b).

3.2 Temporal Clustering can distinguish between faster and slower decliners

We next sought to use Temporal Clustering ($K = 2$) to summarise cognitive decline in the combined cohort. Before the discrimination score filter was applied this resulted in one slowly declining cluster containing 1,335 individuals and another which declined faster and contained 1,077 individuals (Figure 2(c)). The estimated maximum MMSE ($\hat{\theta}_1$) of the model was 30. The estimated yearly rate of exponential decline in MMSE score was $\hat{\theta}_2^1 = 0.062$ for the slowly declining cluster and $\hat{\theta}_2^2 = 0.292$ for the faster declining cluster. The trajectory of the faster declining cluster takes approximately 10 years to go from an MMSE of 29 to 0, which fits well with LCMM trajectories learned from AD patients [5], whereas the more slowly declining trajectory is estimated to take approximately 50 years (Figure 2(c)).

As our simulation study suggested it was a sensible step, we filtered out individuals for who we didn't have sufficient confidence in cluster assignment (i.e. those with a discrimination score < 2). This resulted in the removal of 31% of individuals, leaving 969 members (73%) of the slowly declining cluster and 688 members (64%) of the faster declining cluster. As in the simulation study, the filter removed a greater number of individuals with higher MMSE at first visit, presumably as the differences between clusters at that stage are more subtle (Figure 2(a)).

We examined diagnosis at last visit for individuals in ADNI1 or AIBL who remained

after the filter, as these were the only cohorts with diagnostic information provided (instead of inferred). Approximately half of these individuals in the slow declining cluster had AD at last visit (99/191), in comparison to ∼80% (236/301) in the faster declining cluster.

3.3 Association between AD risk factors and AD-like cognitive decline

We next sought to use the filtered Temporal Clustering results to identify risk factors distinguishing the two clusters. Higher MMSE at first visit was associated with membership of the faster declining cluster (log odds ratio (LOD) for 10 point increase = 0.54, p-value = 1.7 $\times 10^{-3}$). Focusing on the subset of individuals with $APOE$ data, we showed a positive and dose dependent association between the number of $APOE$ $\epsilon 4$ alleles and membership of the faster declining cluster (1 allele LOD = 0.67, p-value = 1.3 $\times 10^{-3}$; 2 alleles LOD = 0.81, p-value = 7.8 $\times 10^{-3}$). Finally, we found an association between the fast declining cluster and the level of cerebrospinal fluid tau at first visit in ADNI (LOD for 100 pg/ml increase = 0.84, p-value = 0.014). The association of first visit MMSE, $APOE$ and tau with the faster declining cluster is visualised in Figure 2(d). Results were consistent when no filter was used (Table 2).

3.4 Temporal Clustering produces clusters that are more significantly associated with $APOE$ status than LCMM

We wished to compare the results of Temporal Clustering and LCMMs, specifically which approach produced clusters with higher associations to known AD risk factors ($APOE$ and tau). To make the comparison more straightforward, this was performed on Temporal Clustering results before filtering, and on LCMM models without any co-variates. To make it a fairer comparison for LCMM we used two different choice of 'time zero' (first visit or 50th

birthday) and both raw and normalised MMSE (normMMSE).

Because Temporal Clustering infers a 'time zero' we would expect by design its clusters to be less affected by MMSE at first visit than approaches like LCMM that do not. This was indeed the case, confounding of unfiltered Temporal Clustering clusters by MMSE at first visit was at least three-fold lower in absolute terms than that achieved by LCMM (LOD for a 10 point change = 0.5 for Temporal Clustering versus -1.6 – -10 for LCMM, Table 2). Indeed, MMSE at first visit was by far the most significant predictor of LCMM cluster membership in all cases. The biggest difference in the association of risk factors with unfiltered clusters was for $APOE$, especially the significance of the association of a single $APOE\ \epsilon 4$ allele with cluster membership (p-value = $1.1\ \times 10^{-5}$ for Temporal Clustering versus 0.62 – 0.013 for LCMM, Table 2).

Overall LCMM cluster membersip only had a clear relationship with AD risk factors when MMSE was normalised, which resulted in a large difference in MMSE at first visit between the clusters (Table 2). In contrast Temporal Clustering cluster membership had a clear relationship to AD risk factors, and a lower difference in MMSE at first visit, with and without a discrimination score filter.

4 Discussion

We have introduced a new method - Temporal Clustering - that can model cognitive decline by combining an estimated individual offset with clustering on that new time-scale. We show that this leads to clusters that are less influenced by MMSE at first visit, which we believe will make it easier to identify risk factors of cognitive decline. To this end we show a dose-dependent enrichment of $APOE\ \epsilon 4$ carriers in the faster declining (i.e. AD-like) cluster, a difference that is more significant than for clusters produced by LCMM approaches on this dataset.

There is some inconsistency in the literature about the relationship between *APOE* $\epsilon 4$ and the rate of cognitive decline. Some studies in non-demented individuals have found no relationship, e.g. Winnock et al., [26]. However, the majority of studies have either found a modest relationship (e.g. [27, 28, 29, 30]), or one that depends on other factors such as amyloid beta [31], alcohol [32] and body mass [33]. The inconsistency of these studies may be explained by cohort differences and/or the strong methodological challenges of the study of cognitive decline [7].

In the field of cluster analysis determination of the 'optimal' number of clusters is known to be tricky. For example, Bauer et al. [34] have argued that the optimal number of clusters in a model do not necessarily respond to the number of 'real' subgroups in an application. Instead they argue that clusters can equally well be interpreted as having no meaning beyond being a convenient summary of non-gaussian distributions. Therefore in this study, for simplicity, we have generated models with just two clusters ($K = 2$). By comparing cerebrospinal fluid tau and *APOE* genotype between clusters we showed that it is plausible that clusters summarise genuine heterogeneity.

As with the choice of any 'time zero' other than time in study, Temporal Clustering relies on the convergence assumption, that between-person differences over time are representative of within-person changes [6, 35]. Other assumptions underlying Temporal Clustering include symmetric and independent distribution of errors, as implied by the use of least squares estimation. The bounded nature of MMSE, which takes a minimum of zero and a maximum of 30, means that the true distribution cannot be symmetric. In addition to this Temporal Clustering assumes that data is missing completely at random, a stronger and less realistic assumption than the missing at random assumption of mixed models. However, even with these limitations it is encouraging that reasonable clustering accuracy was achieved in the simulation study after filtering for discrimination score, especially as we explicitly simulated missing data due to death.

While simulations showed the effectiveness of Temporal Clustering at estimating cluster membership, they also showed biased estimation of trajectory parameters, especially for clusters with a slow rate of decline (Supplementary Results). Reducing this bias could be useful in its own right, but may also improve the assignment of individuals to clusters. This bias could be due to overfitting of the individual offsets (δ_i), which could be reduced in the future by penalising unrealistic offsets. Alternatively it could be due to the mixture of between and within individual progression in the model, and the large extrapolation beyond the length of follow-up available (from $\sim$ 2 years to 10 or even 50 years). Therefore, it is hard to know whether the $\sim$50 year trajectory of the slowly declining cluster truely reflects within-individual change, or is an artefact of the model. This could be tested in a datasets with longer individual follow-up.

Despite the biased estimation of trajectory parameters, the more slowly declining cluster is still striking. From a clinical point of view this trajectory appears to decline too slowly to represent AD. Backing this up is the fact that it does include a higher proportion of individuals with NC or MCI at last visit (Table S1). However, the fact that around half of this cluster have a diagnosis of AD at last visit could suggest a problem with the model, perhaps motivating additional clusters. A less likely alternative hypothesis would be that individuals with a diagnosis of AD in the slowly declining cluster are misdiagnosed.

An advantage of mixed or random change point models over Temporal Clustering is that co-variates can be explicitly modelled, rather than considered post-hoc. Extending Temporal Clustering to consider co-variates could allow it to have more flexibility in the baseline of the model, this could get around the current crude assumption that the maximum MMSE in a lifetime is the same for all individuals.

A limitation of this study is the use of MMSE to measure cognitive decline. MMSE is acknowledged to have ceiling and floor effects and to be relatively insensitive to cognitive change before MCI [36]. We concentrated on MMSE within this study as it is one of the

most widely collected measures of cognitive ability in dementia. For example, longitudinal MMSE data is available for thousands of patients at the South London and Maudsley NHS Foundation Trust, where it has been extracted from Electronic Health Records from routine care [19]. However, the method should generalise to other measures of cognitive ability.

In conclusion we have demonstrated that it is possible to model cognitive decline using a combination of clustering and inference of individual offsets. This reduces, but does not eliminate, the effect of baseline MMSE on cluster assignment. Finally, we demonstrated a relationship between clusters and known AD risk factors. We believe that Temporal Clustering and future extensions will be useful for studying progression of dementia biomarkers. To allow others to repeat this analysis and explore extensions the source code is freely available at https://github.com/KHP-Informatics/TC.

5 Acknowledgements

SJK is supported by an MRC Career Development Award in Biostatistics (MR/L011859/1). CW is funded by the Wellcome Trust (107881) and the MRC (MC_UP_1302/5).

The author CJ receives salary support and the authors acknowledge infrastructure support from the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to acknowledge John Todd and Ken Smith (University of Cambridge), as well as Sylvia Richardson (MRC Biostatistics Unit), for hosting SJK during this work and manuscript review. The JDRF/Wellcome Trust Diabetes and Inflammation Laboratory is in receipt of a Wellcome Trust Strategic Award (107212) and receives funding from the JDRF (5-SRA-2015-130-A-N) and the NIHR Cambridge Biomedical Research Centre. The Cambridge Institute for Medical Research (CIMR) is in receipt of a Wellcome Trust Strategic

Award (100140).

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7-/2007-2013) and in-kind contribution from EFPIA companies.

This work was supported by awards to establish the Farr Institute of Health Informatics Research, London, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National

Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Major AIBL funders include CSIRO Flagship Program, Science Industry Endowment Fund, NHMRC, Dementia CRC, CRC-Mental Health, Alzheimer's Association, an anonymous Foundation, GE Healthcare, Janssen, Pfizer, Avid Radiopharmaceuticals, Navidea, Merck, and Astra Zeneca.

We would like to thank Graciela Muniz-Terrera, Rob Howard, Steven Hill, Robert Goudie and peer reviewers for constructive advice.

6 Conflict of interest

SJK has served on a Roche Diagnostics advisory board on blood biomarkers for AD. RJBD receives grant support from Eli Lilly for a research project. AP and CW declare no conflict of interest.

References

[1] Polhamus D, Rogers JA, Gillespie WR, French J, Gastonguay MR. From Evidence Synthesis to Trial Optimization: The adsim Package for Model-based Simulation in Alzheimer's Disease;. Accessed: May 2016. http://www.page-meeting.org/pdf_assets/5515-adsimposter.pdf.

[2] Rogers JA, Polhamus D, Gillespie WR, Ito K, Romero K, Qiu R, et al. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. J Pharmacokinet Pharmacodyn. 2012;39(5):479–498.

[3] Meuller S, Weiner M, Thal L, Petersen R, Jack C, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 2005;1(1):55–66.

[4] Neville J, Kopko S, Broadbent S, Aviles E, Stafford R, Solinksy CM, et al. Development of a unified clinical trial database for Alzheimer's disease. Alzheimers Dement. 2015;11(10):1212–1221.

[5] Wilkosz PA, Seltman HJ, Devlin B, Weamer EA, Lopez OL, DeKosky ST, et al. Trajectories of cognitive decline in Alzheimer's disease. Int Psychoger. 2009;22(2):281–290.

[6] Hoffman L. Considering alternative metrics of time. In: Harring JR, Hancock GR, editors. Advances in Longitudinal Methods in the Social and Behavioral Sciences. Charlotte, NC, USA: IAP Information Age Publishing; 2012. p. 255–287.

[7] Weuve J, Proust-Lima C, Power MC, Gross AL, Hofer SM, Thiebaut R, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. Alzheimers Dement. 2015;11(9):1098–1109.

[8] Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. Misuse of the Linear Mixed Model When Evaluating Risk Factors of Cognitive Decline. Am J Epidemiol. 2011;174(9):1077–1088.

[9] Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. ArXiV. 2015;.

[10] Hall CB, Lipton RB, Sliwinski FE, WF S. A change point model for estimating onset of cognitive decline in preclinical Alzheimer's disease. Stat Med. 2000;19:1555–1566.

[11] Munz Terrera G, van den Hout A, Matthews FE. Random change point models: investigating cognitive decline in the presence of missing data. J Appl Stat. 2011;38(4):705–716.

[12] Yang E, Farnum M, Lobanov V, Schultz T, Raghavan N, Samtani MN, et al. Quantifying the pathophysiological timeline of Alzheimer's disease. International Psychogeriatrics. 2011;26(4):745–753.

[13] Jedynak BM, Lang A, Liu B, Katz E, Zhang Y, Wyman BT, et al. A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. NeuroImage. 2012;63(3):1478–1486.

[14] Jack Jr CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013;12(2):207–216.

[15] Gaffney S, Smyth P. Joint probabilistic curve clustering and alignment. In: Advances in Neural Information Processing Systems 17. MIT Press; 2005. p. 473–480.

[16] Sangalli LM, Secchi P, Vantini S, Vitelli V. k-mean alignment for curve clustering. Computational Statistics & Data Analysis. 2010;54(5):1219–1233.

[17] Kiddle SJ, Windram OPF, McHattie S, Mead A, Beynon J, Buchanan-Wollaston V, et al. Temporal clustering by affinity propagation reveals transcriptional modules in Arabidopsis thaliana. Bioinformatics. 2010;26(3):355–362.

[18] Huopaniemi I, Nadkarni G, Nadukuru R, Lotay V, Ellis S, Gottesman O, et al. Disease progression subtype discovery from longitudinal EMR data with a majority of missing values and unknown initial time points. AMIA Annu Symp Proc. 2014;p. 709–718.

[19] Perrera G, Khondoker M, Broadbent M, Breen G, Stewart R. Factors Associated with Response to Acetylcholinesterase Inhibition in Dementia: A Cohort Study from a Secondary Mental Health Care Case Register in London. PLOS ONE. 2014;9(11):e109484.

[20] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015. Available from: https://www.R-project.org/.

[21] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. International Psychogeriatrics. 2009;21(04):672.

[22] Tombaugh T, McIntyre N. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc. 1992;40(9):922–935.

[23] Frisoni GB, Pievani M, Testa C, Sabattoli F, Bresciani L, Bonetti M, et al. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain. 2007;130(3):720–730.

[24] Gupta VB, Wilson AC, Burnham S, Hone E, Pedrini S, Laws SM, et al. Follow-up plasma apolipoprotein E levels in the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) cohort. Alzheimer's Research & Therapy. 2015;7:16.

[25] Phillips V, Amieva H, Andrieu S, Dufouil C, Berr C, Dartigues JF, et al. Normalized Mini-Mental State Examination for Assessing Cognitive Change in Population-Based Brain Aging Studies. Neuroepidemiology. 2014;43(1):15–25.

[26] Winnock M, Letenneur L, Jacqmin-Gadda H, Dallongeville J, Amouyel P, Dartigue JF. Longitudinal analysis of the effect of apolipoprotein E e4 and education on cognitive performance in elderly subjects: the PAQUID study. J Neurol Neurosurg Psyhiatry. 2002;72(6):794–797.

[27] Albrecht MA, Szoeke C, Maruff P, Savage G, Lautenschlager NT, Ellis KA, et al. Longitudinal cognitive decline in the AIBL cohort: The role of APOE ?4 status. Neuropsychologia. 2015;75:411–419.

[28] Christensen H, Batterham PJ, Mackinnon AJ, Jorm AF, Mack HA, Mather KA, et al. The association of APOE genotype and cognitive decline in interaction with risk factors in a 65–69 year old community sample. BMC Geriatrics. 2008;8(14):1–10.

[29] Gui H, Jiang CQ, Cherny SS, Sham PC, Xu L, Liu B, et al. Influence of Alzheimer's disease genes on cognitive decline: the Guangzhou Biobank Cohort Study. Neurobiol Aging. 2014;35(10):2422:e3–8.

[30] Carrasquillo MM, Crook JE, Pedraza O, Thoma CS, Pankratz VS, Allen M, et al. Late-Onset Alzheimer Risk Variants in Memory Decline, Incident Mild Cognitive Impairment and Alzheimer Disease. Neurobiol Aging. 2015;36(1):60–67.

[31] Lim YY, Villemagne VL, Pietrzak RH, Ames D, Ellis K, Harrington K, et al. APOE e4 moderates amyloid-related memory decline in preclinical Alzheimer's disease. Neurobiol Aging. 2015;36:1239–1244.

[32] Downer B, Zanjani F, Fardo DW. The Relationship Between Midlife and Late Life Alcohol Consumption, APOE e4 and the Decline in Learning and Memory Among Older Adults. Alcohol Alcoholism. 2014;49(1):17–22.

[33] Rajan KB, Skarupski KA, Rasmussen HE, Evans DA. Gene-Environment Interaction of Body Mass Index and Apolipoprotein E ϵ4 Allele on Cognitive Decline. Alzheimer Dis Assoc Disord. 2014;28(2):134–140.

[34] Bauer DJ, Curran PJ. Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. Psychological Methods. 2003;8(3):338–363.

[35] Sliwinski M, Hoffman L, Hofer SM. Evaluating Convergence of Within-Person Change and Between-Person Age Differences in Age-Heterogeneous Longitudinal Studies. Res Hum Dev. 2010;7(1):45–60.

[36] Proust-Limma C, Amieva H, Dartigues JF, Jacqmin-Gadda H. Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. Am J Epidemiol. 2007;165(3):344–350.

Tables

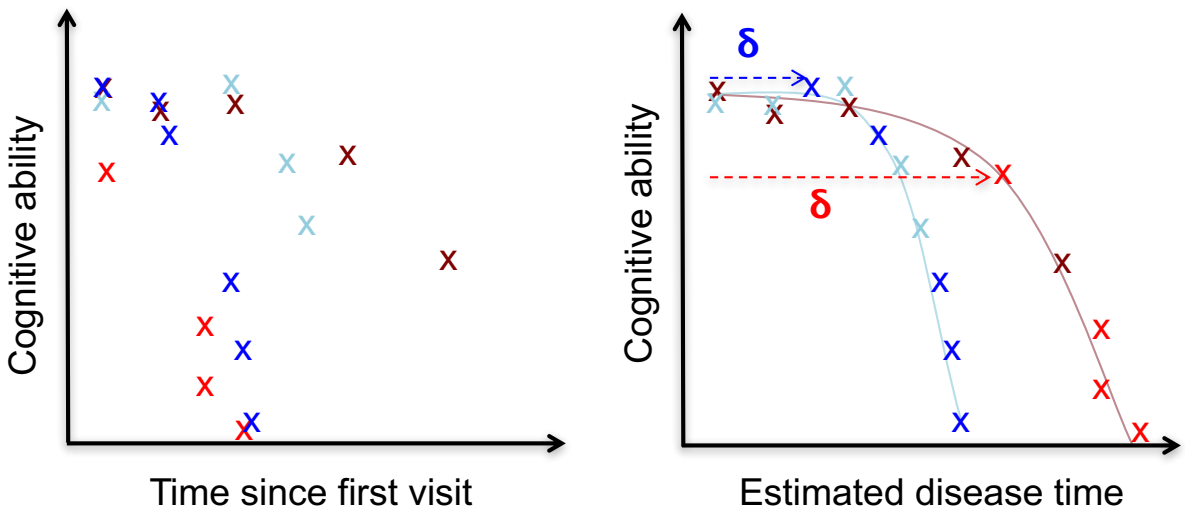
Table 1: Demographic summary of cohorts. Diagnosis for ADNI and AIBL has been recorded, whereas for C-1013 and C-1014 it has been inferred from MMSE scores. Data is summarised using counts or median (interquartile Range). TS = time series, ADNI = Alzheimer's Disease Neuroimaging Initiative, AIBL = Australian Imaging Biomarkers and Lifestyle flagship study of ageing, MMSE = Mini-Mental State Examination, NC = Normal Cognition, MCI = Mild Cognitive Impairment, AD = Alzheimer's disease, and *APOE* $\epsilon 4$ = the $\epsilon 4$ allele of the *apolipoprotein E* gene.

Cohort	**Time series with > 1 time point**	**Number of time points**	**Length of time series in years**	**Diagnoses at first visit (NC/ MCI/ AD)**	**Gender (Female /Male)**	**Age at first visit in years**	**Number of *APOE* $\epsilon 4$ alleles (0/1/2)**
Combined	2412	5 (4 - 6)	1.9 (1.5 - 3.1)	399/920/1093	1186/1226	75 (70 - 80)	
ADNI	785	5 (4 - 6)	3.1 (2.1 - 3.2)	222/382/181	327/458	75 (70 -79)	397/301/87
AIBL	264	3 (3 - 4)	4.4 (3.1 - 4.6)	177/52/35	136/128	72 (66 - 79)	135/110/19
C-1013	719	5 (5 - 6)	1.6 (1.4 - 2.1)	. /247/472	361/358	76 (70 - 80)	Unknown
C-1014	644	5 (4 - 5)	1.5 (0.68 - 1.6)	. /239/405	362/282	76 (70 - 81)	Unknown

Table 2: Table summarising logistic regression analysis, comparing cluster membership to AD risk factors for Temporal Clustering (TC) and LCMM. Four different LCMM models have been run, combining one of two choice for 'time zero' with the choice to use raw MMSE or normalised MMSE (normMMSE). Each line refers to a different logistic regression analysis to better cater for missing risk factor data, except for *APOE* for each clustering method, which were modelled together. Signs for *LOD* have been swapped when appropriate to allow appropriate comparisons, as signs depend on cluster labels which can be swapped arbitrarily. MMSE at first visit is coded in units of ten and cerebrospinal fluid tau is coded in units of 100 pg/ml. N = sample size with complete data, *LOD* = Log Odds Ratio, SE = Standard Error, LCMM = Latent Class Mixture Model, MMSE = Mini Mental State Examination, CSF = cerebrospinal fluid.

Method	**Time zero**	**Variable**	**Risk factor**	**N**	*LOD*	**SE**	**p-value**
TC	Inferred	MMSE	MMSE first visit	2399	0.5	0.14	4.6×10^{-4}
TC	Inferred	MMSE	1 APOE e4	1049	0.61	0.14	1.1×10^{-5}
TC	Inferred	MMSE	2 APOE e4	1049	0.84	0.23	3.6×10^{-4}
TC	Inferred	MMSE	CSF tau	407	0.77	0.24	0.0012
LCMM	First visit	MMSE	MMSE first visit	2399	-1.6	0.23	1.3×10^{-11}
LCMM	First visit	MMSE	1 APOE e4	1049	0.4	0.33	0.22
LCMM	First visit	MMSE	2 APOE e4	1049	0.64	0.41	0.12
LCMM	First visit	MMSE	CSF tau	407	0.38	0.4	0.35
LCMM	Age 50yrs	MMSE	MMSE first visit	2399	-1.7	0.25	3.5×10^{-12}
LCMM	Age 50yrs	MMSE	1 APOE e4	1049	0.17	0.34	0.62
LCMM	Age 50yrs	MMSE	2 APOE e4	1049	0.31	0.45	0.49
LCMM	Age 50yrs	MMSE	CSF tau	407	0.49	0.4	0.21
LCMM	First visit	normMMSE	MMSE first visit	2399	-10	0.72	5.0×10^{-48}
LCMM	First visit	normMMSE	1 APOE e4	1049	0.68	0.27	0.013
LCMM	First visit	normMMSE	2 APOE e4	1049	1.1	0.36	0.0017
LCMM	First visit	normMMSE	CSF tau	407	1.2	0.37	0.0015
LCMM	Age 50yrs	normMMSE	MMSE first visit	2399	-10	0.7	3.7×10^{-47}
LCMM	Age 50yrs	normMMSE	1 APOE e4	1049	0.62	0.27	0.022
LCMM	Age 50yrs	normMMSE	2 APOE e4	1049	1.1	0.35	0.0028
LCMM	Age 50yrs	normMMSE	CSF tau	407	1.1	0.36	0.0023

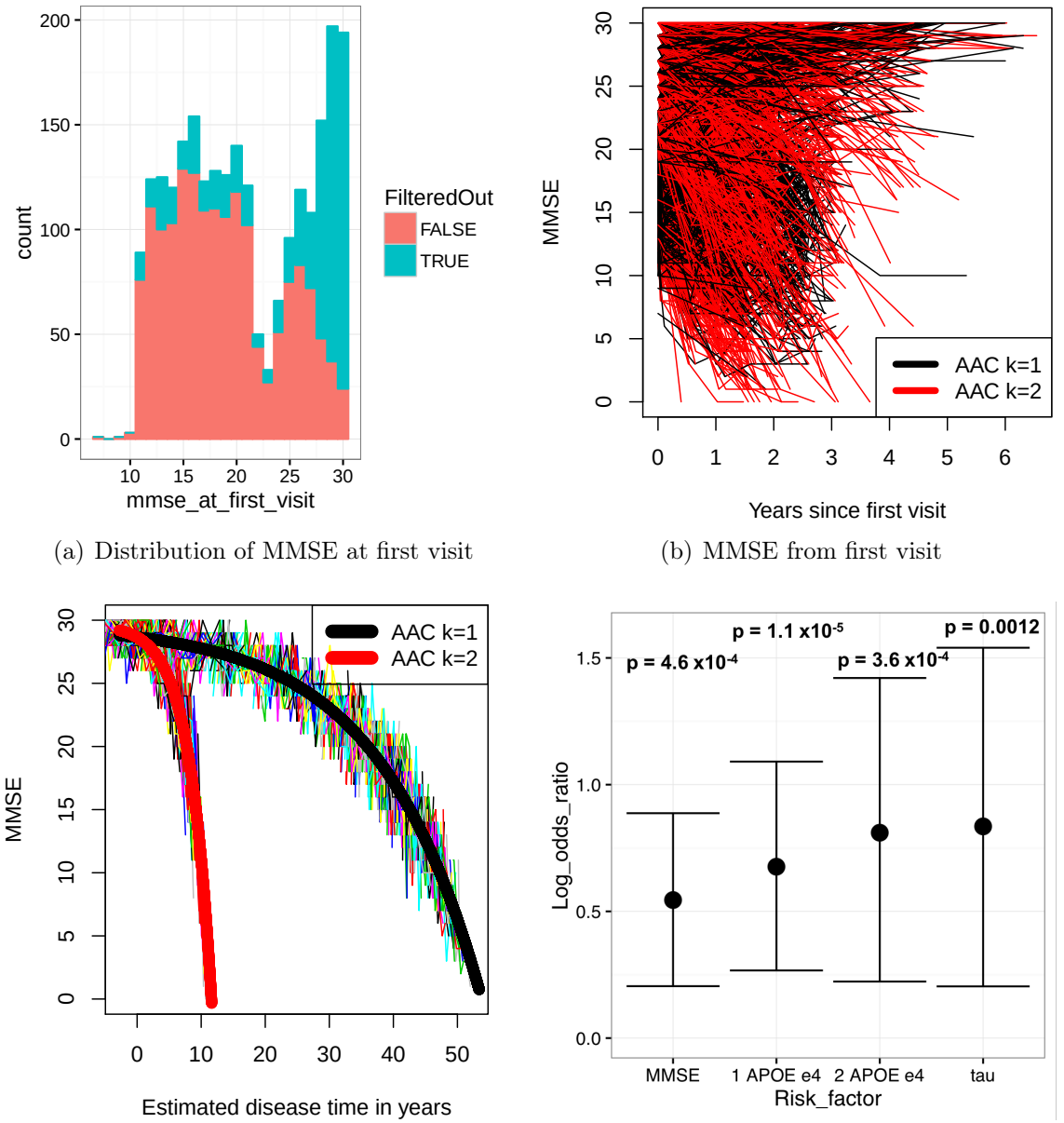
Figures



(a) Illustration of raw data

(b) Illustration of temporal clustering

Figure 1: Illustration of (a) raw longitudinal cognitive ability data and (b) the results of applying temporal clustering ($K = 2$). Different individuals are represented by different colours, with different shades of the same colour used to indicate individuals following the same decline trajectory after alignment. Solid curves in (b) are used to represent the long-term trajectory models ($\phi(t; \boldsymbol{\theta}^k)$) for the blue and red clusters. The individual offsets (δ) applied to dark blue and light red are indicated by dashed lines.



(a) Distribution of MMSE at first visit

(b) MMSE from first visit

(c) MMSE Temporal Clustering results

(d) Association of risk factors with the faster declining cluster

Figure 2: (a) Stacked histogram showing the distribution of MMSE at first visit for the combined cohort (both colours), with the distribution after applying a discrimination score threshold of 2 shown in pink. Spaghetti plots of longitudinal MMSE in the combined cohort with (b) first visit as 'time zero' or (c) a 'time zero' inferred by Temporal Clustering. Unfiltered clusters (k) are represented by either (b) the colour of individual curves or (c) the colour of cluster trajectories (represented by thick curves). (d) Log odds ratios and confidence intervals for risk factors (MMSE at first visit in units of ten, number of *APOE* $\epsilon 4$ alleles or cerebrospinal fluid tau levels in units of 100 pg/ml).

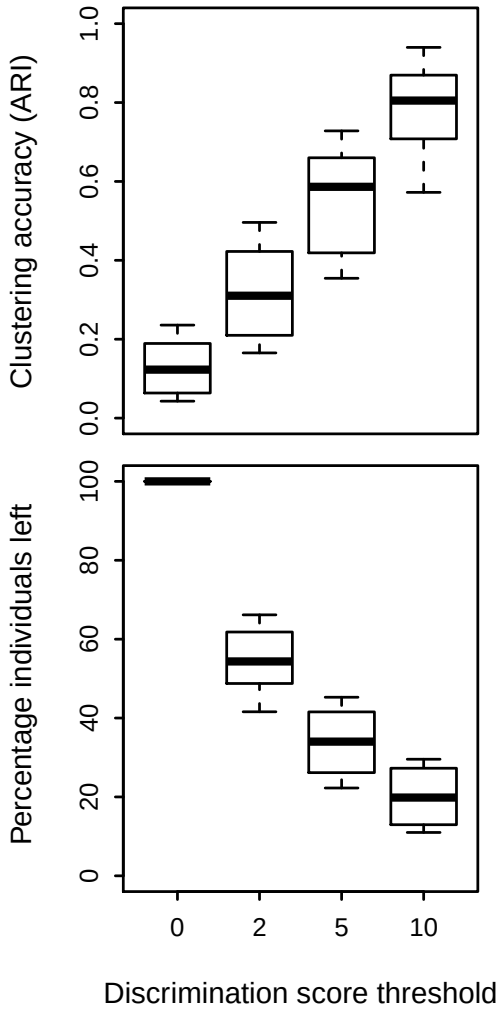


Figure 3: Boxplots of simulation study examining the effect of a discrimination score filter on the accuracy of Temporal Clustering cluster assignment, and number of individuals left after the filter is applied. Discrimination score is measure of relative goodness of fit of each individual to their assigned cluster. Clustering accuracy is measured in Adjusted Rand Index (ARI). Boxplots are over 100 simulations of four different choices of θ_2^1 and θ_2^2.