



# **Applications of Longitudinal Machine Learning Methods in Multi-Study Alzheimer's Disease Datasets**

**PhD Dissertation Defense for  
Doctorate Degree in Biostatistics**

Chad Murchison

Department of Biostatistics  
Alzheimer's Disease Research Center

11/10/2021



# Outline

- Overview, objectives and aims
- Approach and design
- Ensemble tree methods
- Sequential neural networks
- Trajectories, forecasting, and imputation
- Final thoughts



# Overview - Inference/Prediction

As a field, the goal of statistics is to learn from our data, commonly using inference and prediction; although intrinsically related there are subtle differences in how these methods are leveraged

## Inference

- Use the model to learn about the data generation process
- Explain the relationship between a response and covariates
- Frequently emphasizes unbiased interpretation of this relationship
- Is often the standard approach for biostatisticians

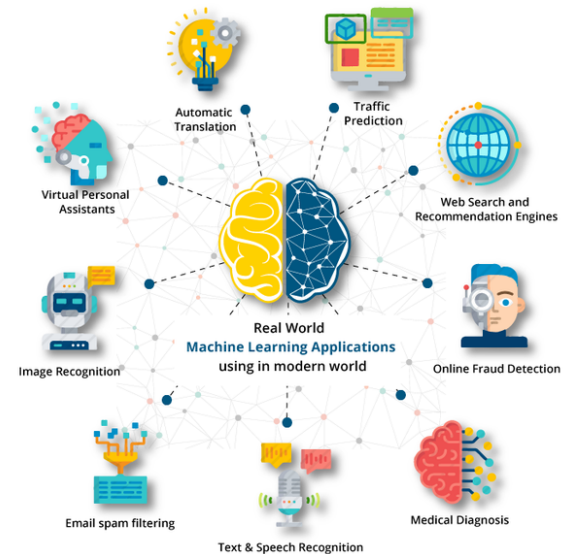
## Prediction

- Use the model to determine likely outcomes from the data
- Role of covariates is to aid in calculating a response
- Emphasis on generalization to unseen data
- Largely the purview of machine/statistical learning (ML/SL)

# Overview - Machine Learning

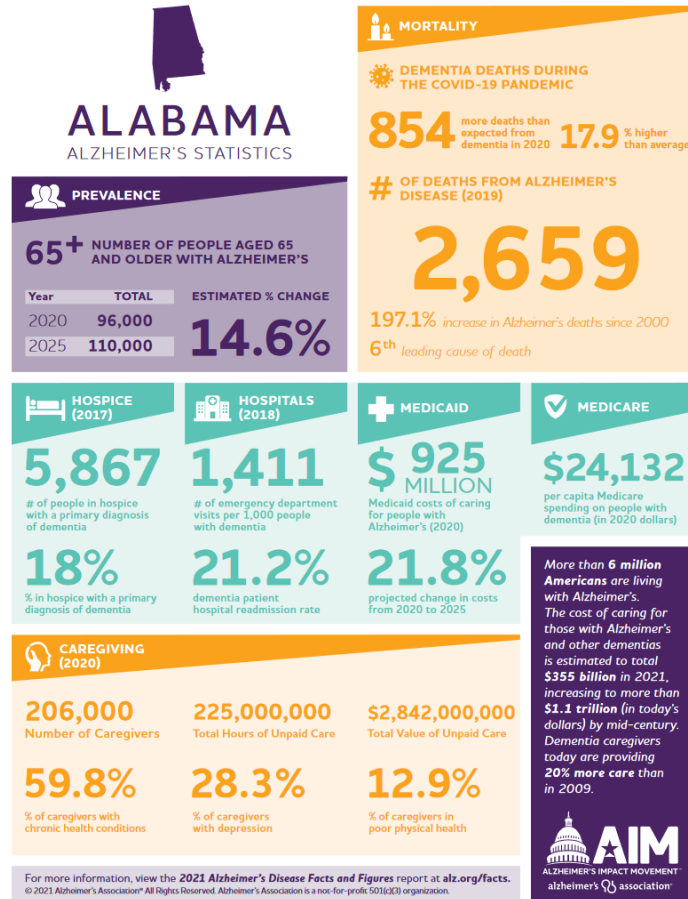
As computational power and data management capacity have increased, machine learning has found greater popularity across a range of applications in a number of different fields such as finance, transportation and medicine

- Image processing
- Identification / Classification
- Intelligent agents
- Natural language processing
- Text mining
- Decision support



One area attempting to utilize its extensive neuroimaging, genetic, biomarker, and clinical/demographic profiling for ML is Alzheimer's disease (AD) and its related dementias

# Overview - Alzheimer's Disease



With the ever-aging population, AD has become an increasingly prominent medical concern at the global, national and local level

Accompanying cost and pervasiveness is a recognition to better identify at risk patients who could benefit from preventative measures

This includes both actual measures of ability across the spectrum of cognitive domains or predicting impairment

Prediction is key as the pathological and biological hallmarks of AD precede cognitive disability substantially

This makes AD ideal for machine learning applications



# Overview - ML/SL and AD

Recognition of the potential of ML/SL in AD is clear; a PubMed search of "machine learning" and "Alzheimer's disease" returned 1355 results with 330 in 2020 alone while arxiv.org cited another 230+ articles. However, substantial limitations remain:

- Models focus on clinical diagnosis as a classification response with little effort to predict neuropsychological outcomes directly; this invariably emphasizes clinician judgment over arguably more objective measures of impairment
- Nearly all ML/SL applications in AD use methods which are cross-sectional in nature and designs that directly leverage or account for subject-specific effects, repeated measures, and serial correlation are comparatively uncommon; a literature review in 2021 found only 8% of 65 reviewed article explicitly used longitudinal methods
- Assessment and evaluation are limited with respect to direct model-to-model comparison as novel paradigms are developed; evaluation metrics such as accuracy are typically reported only for variations on the proposed ML/SL model and reference designs are almost never considered as baselines





# Overview - Objectives

This dissertation sought to remedy these deficiencies by comprehensively comparing the predictive capacity of a variety of longitudinal extensions of ML/SL techniques in a harmonized multi-study dataset of Alzheimer's disease

- Two classes of responses, one continuous and one binary categorical, were considered to contrast model ability for both regression and classification across cognitive outcomes with specific clinical and research contexts
- Several distinct model designs were evaluated including standard inferential methods as reference, well-characterized ensemble methods, and adaptations of sequential deep-learning designs to work with time-dependent data
- Special consideration was given to the development of both whole subject trajectories *de novo* as well as forecasting of final observations specifically utilizing previously observed participant data

Together this research addressed the questions of which longitudinal ML/SL methods give the best performance for AD outcomes, was prediction superior to inferential standards, and what was the impact of subject-specific effects



# Specific Aims

To meet these objectives, the dissertation was composed of three primary aims:

## Aim 1

Assessment of the predictive performance for a variety of discriminative ensemble methods with adaptations specific to longitudinal data when applied to AD specific outcomes. The models under consideration were:

1. Mixed-effects random forest (MERF) with sampling from the feature space
2. Bootstrap aggregated (bagged) mixed-effects trees with a standard non-bagged GLMER tree as reference
3. Sequential boosting of residuals on mixed-effects trees (boosted)

Specifically, what were the regression and classification metrics for these models, were they superior to the inferential references, how did they rank against each other, and were these patterns different when outcomes were predicted as trajectories versus forecasting of observations





# Specific Aims

## Aim 2

To contrast against the predictive ability of the ensemble methods using longitudinal extensions of deep-learning neural network model designs. This was structured around the sequential neural networks with wider use in fields such as image recognition and natural language processing and use their natural extensions to time-dependent data. The methods under consideration are:

1. Long short-term memory recurrent neural networks (LSTM RNN)
2. One-dimensional convolutional neural networks (1D CNN)
3. A multi-layer perceptron / feed-forward neural network as a reference (FNN)

Similar pipelines were carried out for the neural network designs to determine superiority in regression and classification tasks against the inferential references and among themselves and to examine if these patterns were consistent when predicting either whole trajectories or forecasting future observations



# Specific Aims

## Aim 3

Using repeated measures in longitudinal data allows for special consideration of the role of subject-specific effects on prediction which are also expressly used for some longitudinal machine learning methods. Investigation of subject-specific effects on generating whole outcome trajectories and forecasting entailed:

1. Suppression of subject-specific effects to only use population-level covariates
2. Imputation of subject-specific effects based on model parameterizations
3. Directly leveraging subject-specific effects when forecasting future observations based on prior data

Superiority in regression-based prediction metrics was compared among a pre-parameterized reference, an equivalent regression model built *de novo*, and the MERF ensemble method both within and across subject-specific effect designs

This was examined using a 2-year subset of the meta-database as well as an extended 5-year synthetic dataset for validation



# Approach - Outcomes

For a more comprehensive evaluation, both a continuous and categorical outcome were used to assess prediction using regression and classification

**Continuous outcome for regression**  
**Alzheimer's disease assessment scale - cognitive sub-scale (ADAS-Cog)**

- An assessment battery to evaluate memory, reasoning, orientation, and language with a continuous 70 point scale with higher scores indicative of greater levels of impairment
- Commonly used as an outcome in clinical trial settings as a more refined assessment of global and domain specific cognition
- Similar to other neuropsychological assessments like the MocA and MMSE, the ADAS-Cog has higher resolution at greater levels of impairment, giving it greater capacity to distinguish severity among the impaired but loses sensitivity when assessing the cognitively intact

Alzheimer's Disease Cooperative Study				
<b>ADAS – Cognitive Behavior</b>				
SAMPLE FORM – Page 1 of 4				
Center Name	Patient Number	Patient Initials	Examiner Initials	Examination Date
	P R - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
				Month Day Year





# Approach - Outcomes

For a more comprehensive evaluation, both a continuous and categorical outcome were used to assess prediction using regression and classification

**Categorical outcome for classification**  
**Impairment status from the Clinical Dementia Rating (CDR)**

- A five-point scale characterizing six domains of both cognitive and functional performance (Memory, Orientation, Judgment, Community Affairs, Home & Hobbies, Personal Care)
- Used as diagnostic tool in clinical settings and a metric in dementia studies
- Global scores of 0.5 indicate "very mild" or "questionable" dementia with higher scores (1, 2, 3) noting increasing degrees of dementia
- For this research, a score of 0.5 was taken breakpoint to create a binary variables (non-impaired vs impaired) to give a classification emphasizing earlier stages of decline

CLINICAL DEMENTIA RATING (CDR™):					
	0	0.5	1	2	3

Impairment				
None 0	Questionable 0.5	Mild 1	Moderate 2	Severe 3





# Approach - Reference Models

Beyond cross-model comparisons, traditional inferential mixed-effects regression models were used as reference designs with parameterizations taken outside the dataset used for ML/SL model training and testing

## CPAD parameterization for ADAS-Cog

- Simulation model presented in 2012 used for generation of synthetic clinical trial cohorts with a focus on AD intervention
- Used mixed-effects beta regression for ADAS-Cog progression based on baseline age, sex, APOE4 carrier status, and baseline MMSE
- Provides both fixed effect parameters for population covariates and random effects variances for intercept and slope
- Although primarily tuned to predict ADAS-Cog out to two years in linear time has also been used for wider time frames

J Pharmacokinet Pharmacodyn (2012) 39:479–498  
DOI 10.1007/s10928-012-9263-3

### Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis

James A. Rogers · Daniel Polhamus · William R. Gillespie ·  
Kaori Ito · Klaus Romero · Ruolun Qiu · Diane Stephenson ·  
Marc R. Gastonguay · Brian Corrigan





# Approach - Reference Models

Beyond cross-model comparisons, traditional inferential mixed-effects regression models were used as reference designs with parameterizations taken outside the dataset used for ML/SL model training and testing

## **CDR impairment status built ad hoc**

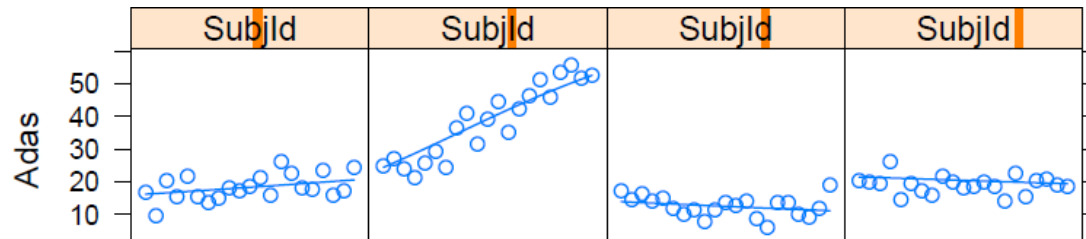
- Built using mixed-effects logistic regression using a holdout subset of the multi-study database
- Follows the same parameterization scheme as the CPAD ADAS-Cog model with population-level covariates and subject-specific random effects
- Exception is unlike CPAD there is no time-MMSE interaction which was found to impede model convergence due to overfitting
- Unstructured random effects design with intercept and slope and a covariance term between the components

# Approach - Reference Models

In Aim 3, the role of pre-determined parameters for random effects were evaluated in the ADAS-Cog to investigate how imputation or suppression of random effects impacted variance and bias when predicting both whole trajectories and observational forecasting

In addition to the CPAD reference model, a *de novo* beta regression model was built directly from the data to update the random effects components and include an unstructured correlation term, as well as the MERF ML model which contains a similar random effects component

This aim was specifically constructed to address a recommendation of the CPAD simulation software to impute random effects when generating synthetic cohorts for feasibility with implications of improved generalization





# Approach - Meta-database

- Assembled from the control participants of 18 AD clinical trials as part of the Alzheimer's Disease Cooperative Study (ADCS) and the four phases of the Alzheimer's Disease Neuroimaging Initiative (ADNI) observational study
- Full dataset includes 8936 unique participants with almost 47000 time points with some data extending out to 12 years of follow-up

ADNI	AP	DHA	EIN	ES	HC	HU	LL	MCI	ML	NL	NS	PI	PIWEB	PR	SIN	SL	VN	VP
4289	149	559	306	136	466	312	446	898	157	122	361	741	60	138	95	341	404	158

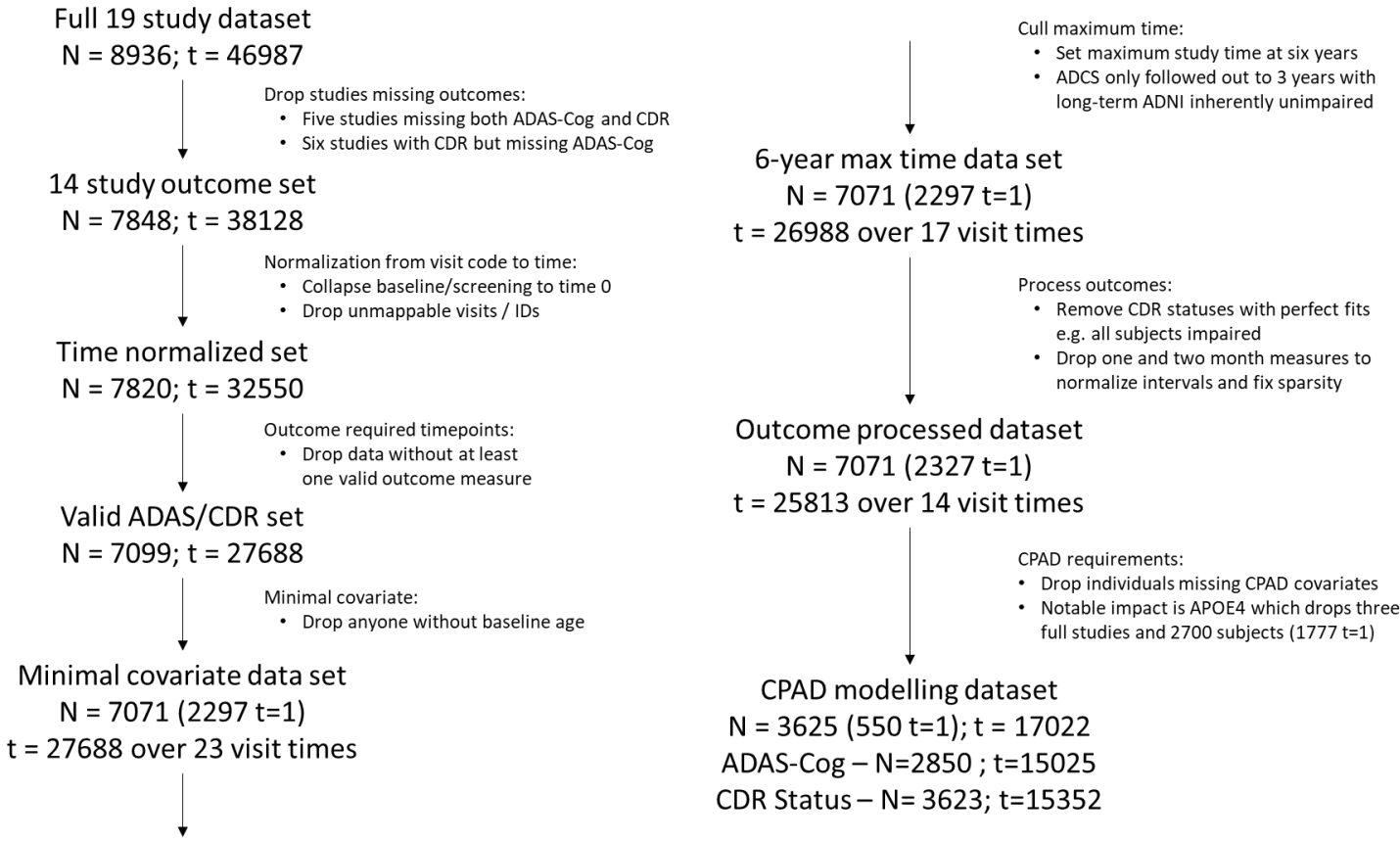
- Harmonization of the dataset mapped disparately coded visits to a continuous temporal variable based on study date and included participants with only baseline data for single visit subjects and screen failures (N=2392)
- Feature space based on minimal dataset required by CPAD and other measures potentially associated with cognitive impairment and high coverage
  - Baseline characteristics: age, sex, race (3 level), ethnicity (2 level), education (5 level), use of anti-dementia medication, APOE4 allele count
  - Time-dependent covariates: MMSE score, weight, blood pressure





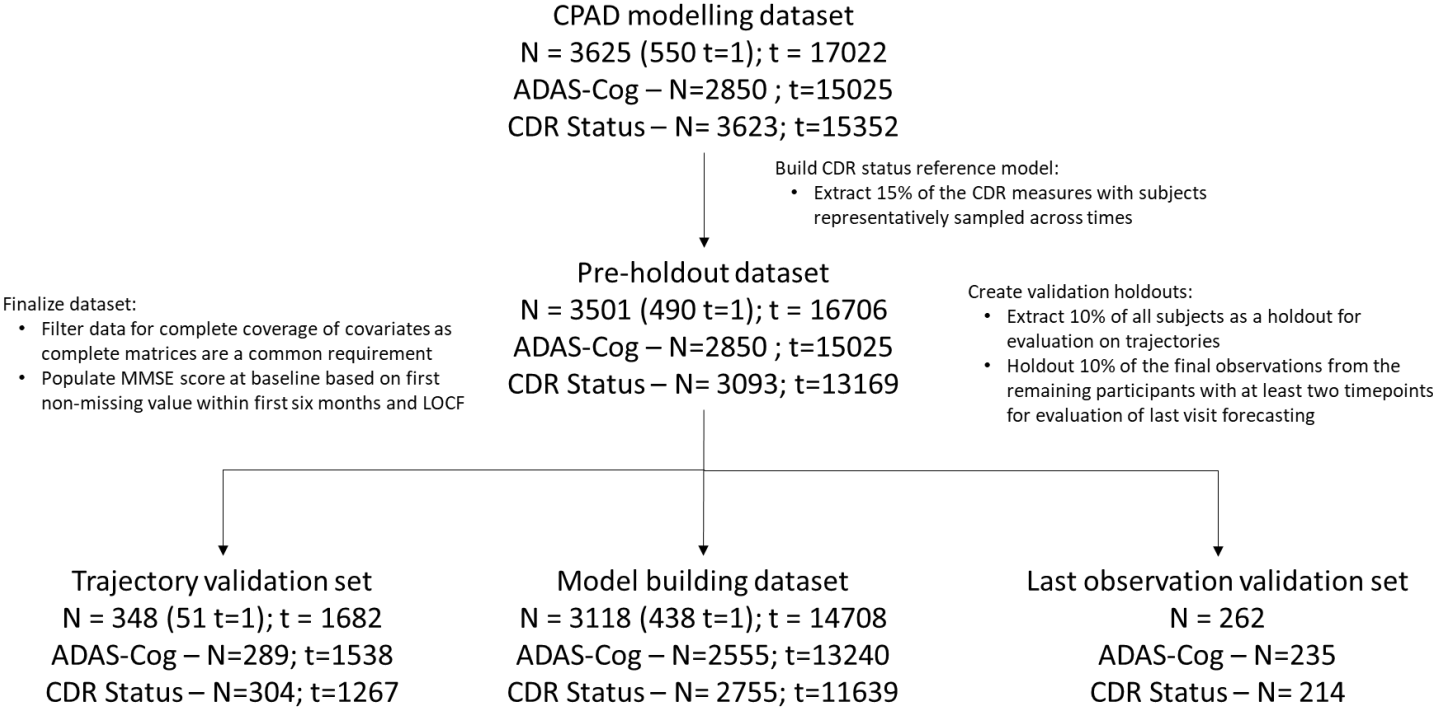


# Approach - Data Preparation





# Approach - Validation Sets

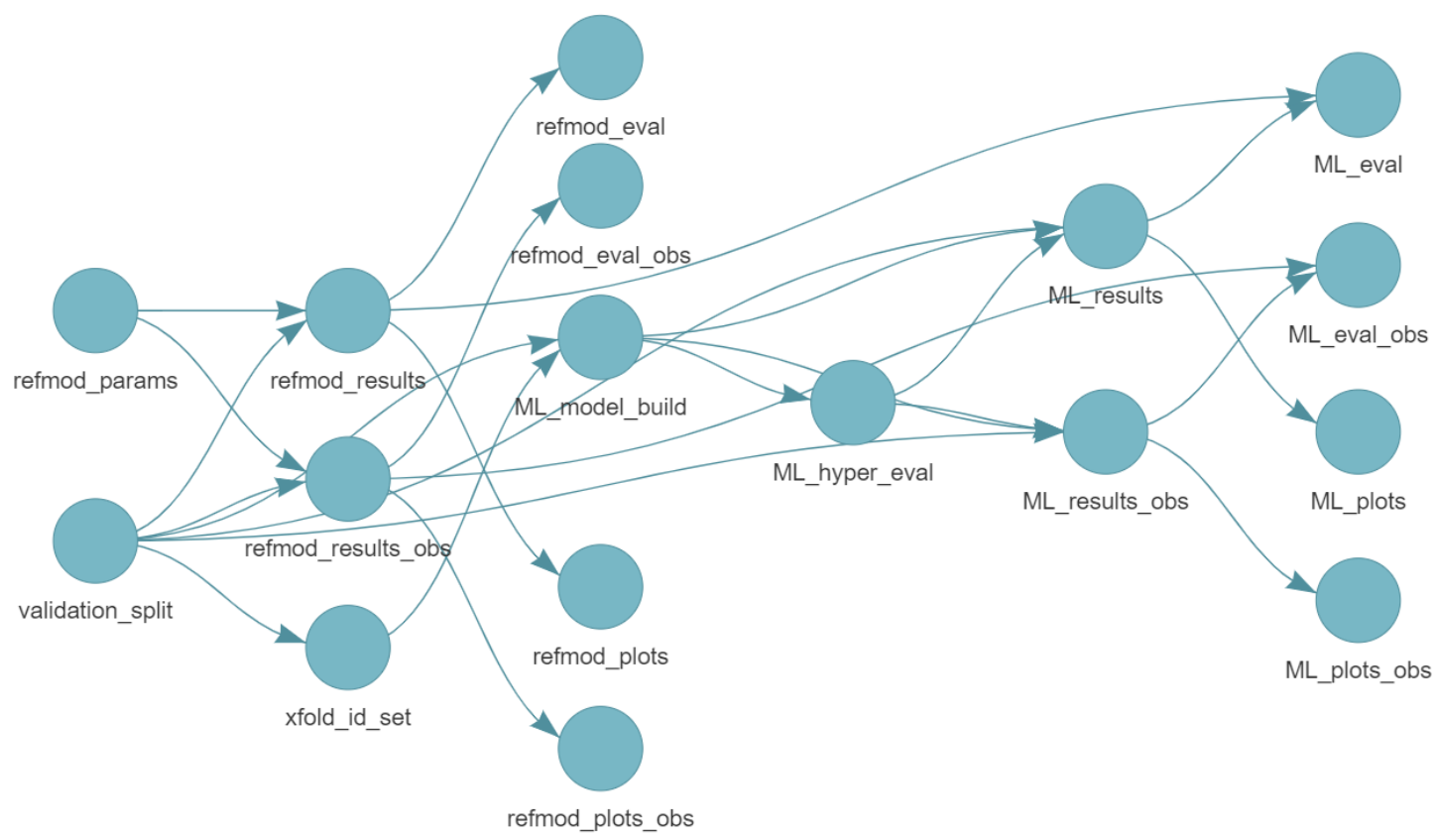


Time	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.5	3	4	5	6
Visits	5985	1232	3570	529	3542	585	1642	110	1912	434	1198	613	375	324





# Approach - Pipeline Flow

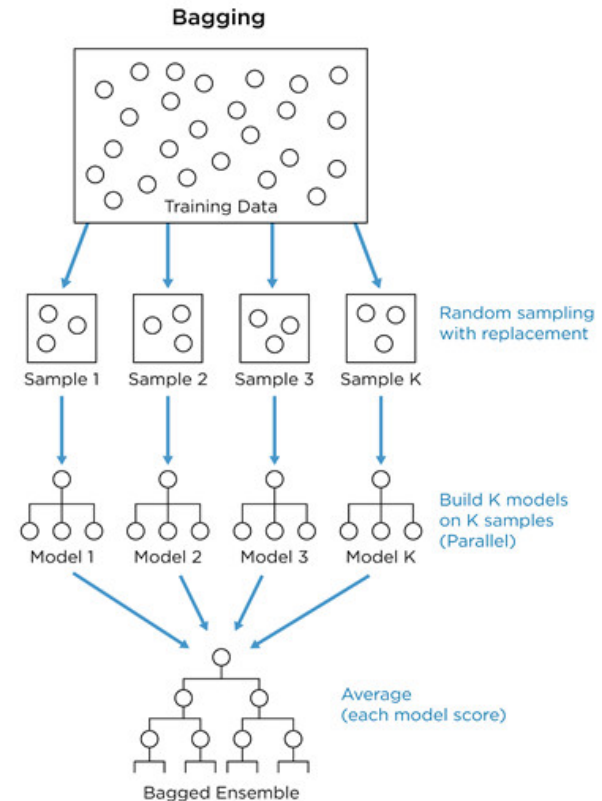




# Approach - Ensemble Methods

## Bagging on GLMER Trees

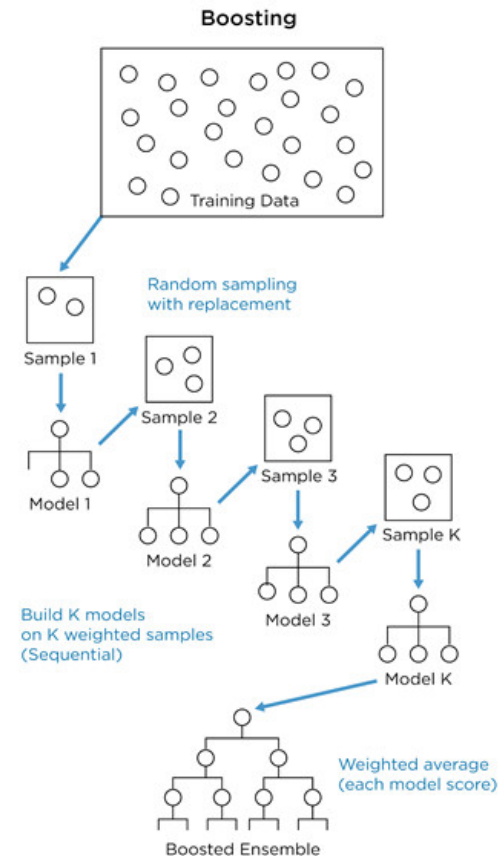
- Randomly samples participants from the training dataset and use out-of-bag samples for validation
- Tree pruning is not necessary, instead results are averaged over the collections trees
- Building of the mixed-effects trees was implemented using the *glmerTree* package in R
- Modified script was developed to conduct the bootstrap aggregation with a single, non-bagged tree as reference



# Approach - Ensemble Methods

## Boosting on Mixed-Effect Trees

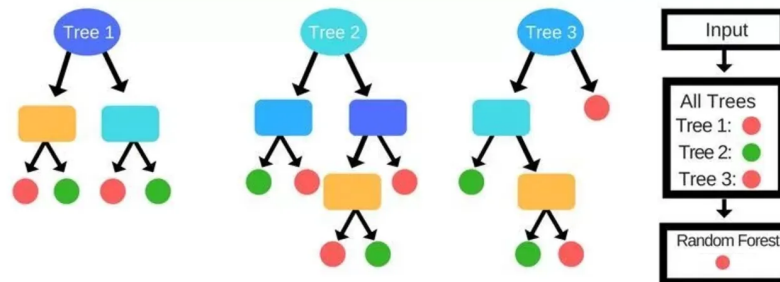
- Instead of sampling from participants during training, the entire training set is used but the data is modified before being iteratively fit to a new tree
- Hyperparameters are important during cross-validation including number of trees, interaction depth, and the rate of learning across tree iterations
- Boosting on the trees was conducted using the *gbm* package in R with support for the clustering of longitudinal design via the *mvtboost* package



# Approach - Ensemble Methods

## Mixed-effect random forest

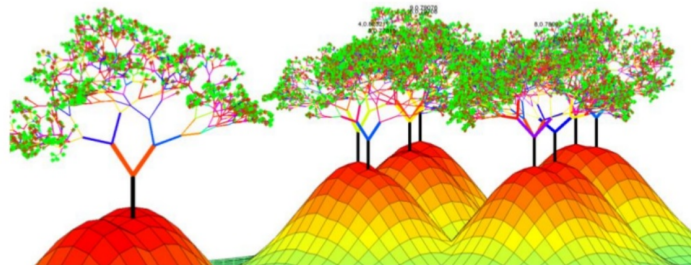
- Instead of randomly sampling the dataset as in bagging, random forests sample from the feature space to make the trees uncorrelated
- For the random forest decision making is similar, taken as the average/majority vote across all trees
- Subject-specific components are implemented in the leaf nodes of the trees with mixed-effects models for clusters
- Implementation used the *LongituRF* package in R with modifications to allow for direct control over how the random effects components were utilized



# Approach - Ensemble Methods

## General ensemble considerations

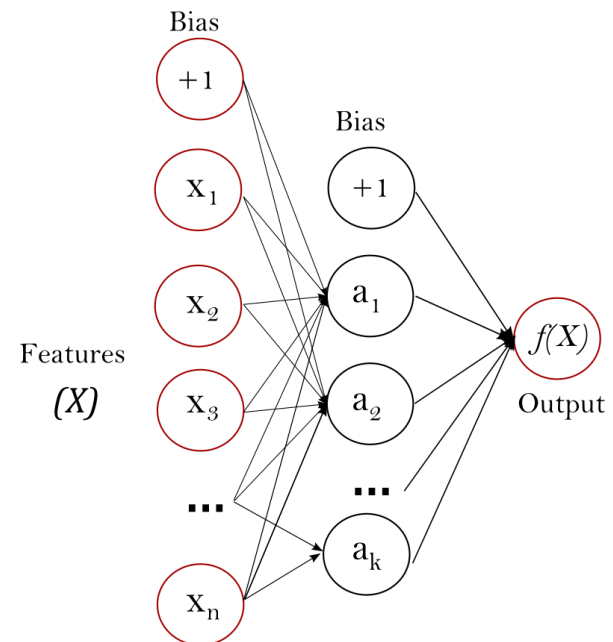
- For all trees, objective functions were optimized using the squared error loss for regression and cross-entropy loss for classification
- Hyperparameter tuning involved features such as number of trees, tree depths, number of observations in leaf nodes, tuning rate (boosting), and sampled proportion size (boosting, bagging) using 10x cross-fold validation
- Modification to the *LongituRF* package was necessary to allow for direct access to the subject-specific components in the leaf nodes to allow for evaluation of suppression, imputation, and fitting of random effects in the MERF model



# Approach - Neural Networks

## Feed Forward Neural Networks

- To compare against the sequential deep learning adaptations, a standard multi-layer perceptron was built as a control design
- Simply considered time as another covariate and views each observation as independent; akin to using an OLS model for longitudinal data
- Hyperparameters were more limited, restricted to number of layers, nodes per layer, and constraint on both the kernel and bias functions



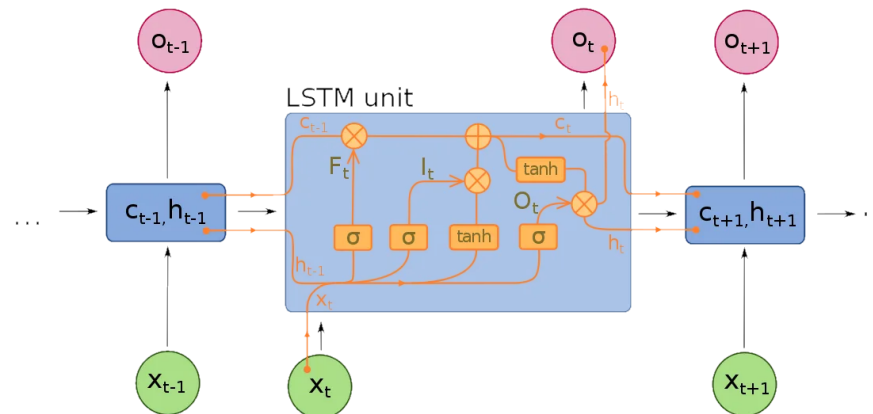




# Approach - Neural Networks

## LSTM Recurrent Neural Networks

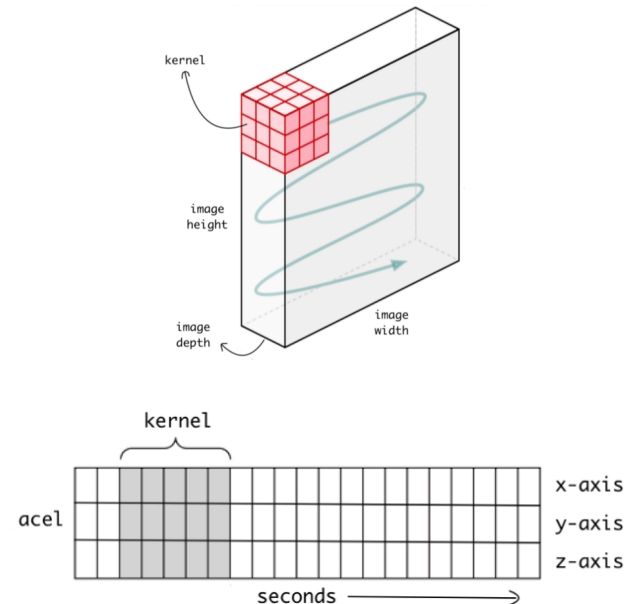
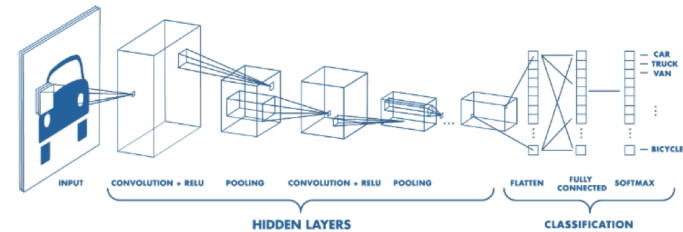
- The LSTM RNN has origins in machine translation and NLP using previously observed words to generate subsequent points
- The unidirectional behavior extends naturally to time-dependent data and uses prior sequence information along with the current input
- Hyperparameters also include the activation weights for the forget/retention gates of the LSTM unit which aids with the "vanishing gradient" problem



# Approach - Neural Networks

## 1D Convolutional Neural Networks

- Instead of retaining varying amounts of prior data, CNNs use a sliding window (kernel) to extract smaller contiguous sequences
- Developed in image processing with sliding windows for feature extraction, 1D extension instead shift along a temporal axis
- Training alternates between convolution (local features) and pooling (combining convolutions) to reduce dimensionality of the feature sets before using a fully connected network for prediction

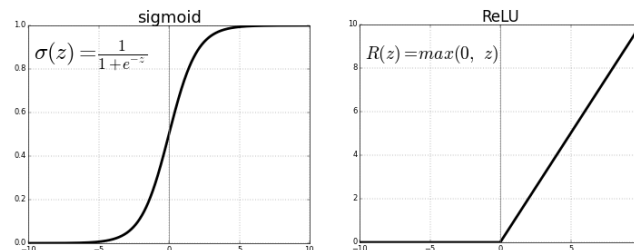




# Approach - Neural Networks

## General network considerations

- Due to the iterative nature of the sequential models, LSTM RNN and 1D CNN had outcomes at time 0 predicted from a FNN built using baseline data prior to sequential prediction when predicting whole trajectories
- All models used standard build methods with connection weights calculated using stochastic gradient descent during back propagation and facilitated by the rectified linear unit activation function with MSE and CE as the loss functions using 10x cross-fold validation
- R does not have native neural network functionality, instead it interfaces with an installation of Python using the *reticulate* package with network building via *TensorFlow* using *keras* as a facilitating API





# Approach - Methods Comparison

After hyperparameter tuning and model selection, predictions were carried out on the validation test sets for whole trajectories and last observation forecasting using the following evaluation metrics:

## Regression on ADAS-Cog

- Root mean square error
- Median absolute error
- Prediction bias  
(absolute value)

## CDR Impairment Classification

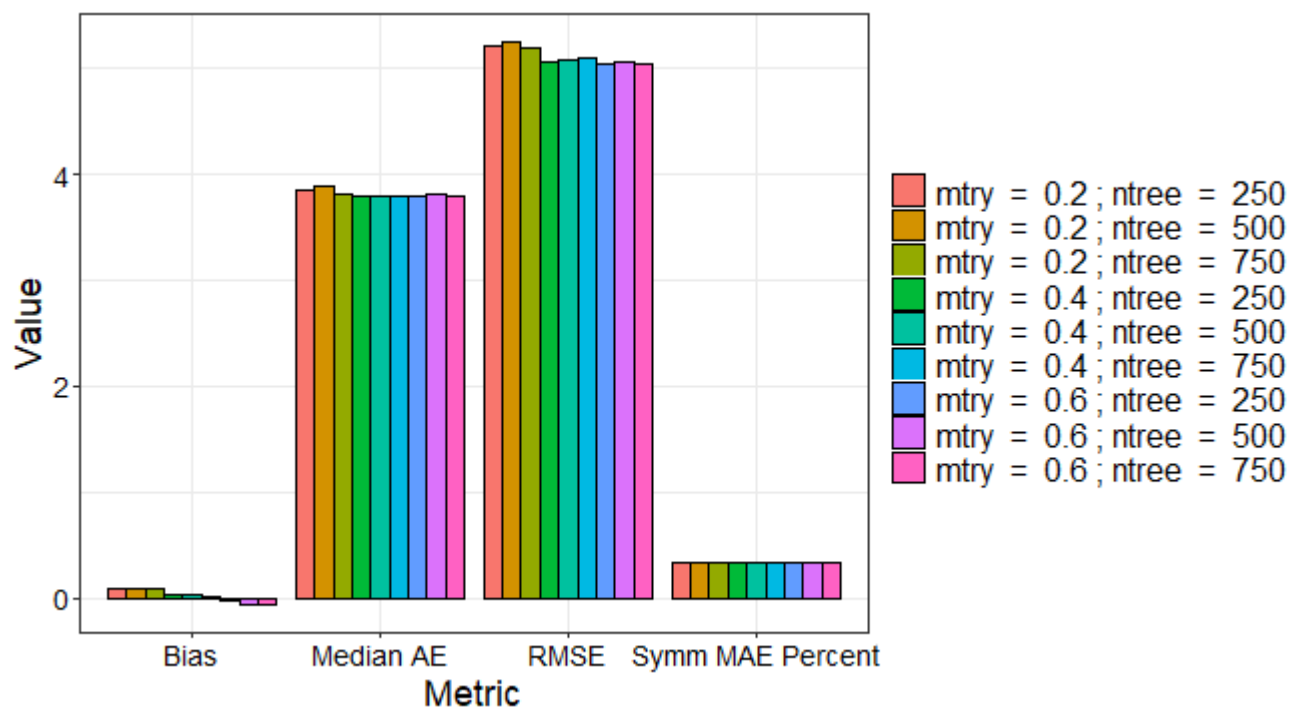
- Accuracy / Precision / Recall
- ROC-based AUC
- Net Reclassification Index  
against reference design

To identify superiority of performance, both absolute differences and percent changes in metrics were calculated compared to the reference models as well as against each other

Additionally, 1000-fold bootstrapping was carried out to build 95% confidence intervals and calculate proportion-based p-values to determine whether the observed changes in metrics were statistically significant



# Approach - Hyperparameter Evaluation



ADAS MERF Example

Metric	Value	Hyperparameters
RMSE	5.0439	mtry = 0.6 ; ntree = 250
Median AE	3.7909	mtry = 0.6 ; ntree = 750
Symm MAE Percent	0.3303	mtry = 0.4 ; ntree = 250
Bias	-0.0116	mtry = 0.4 ; ntree = 750





# Preliminary Results

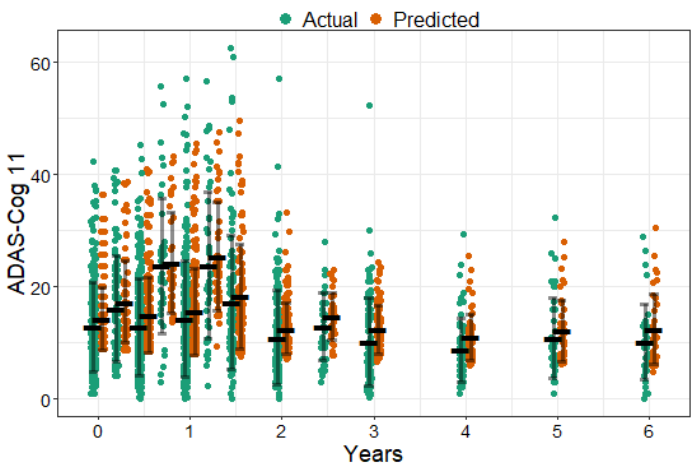
## Reference Models



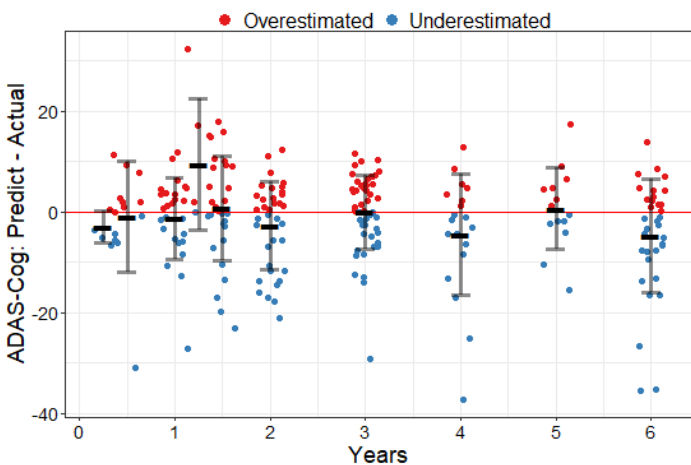
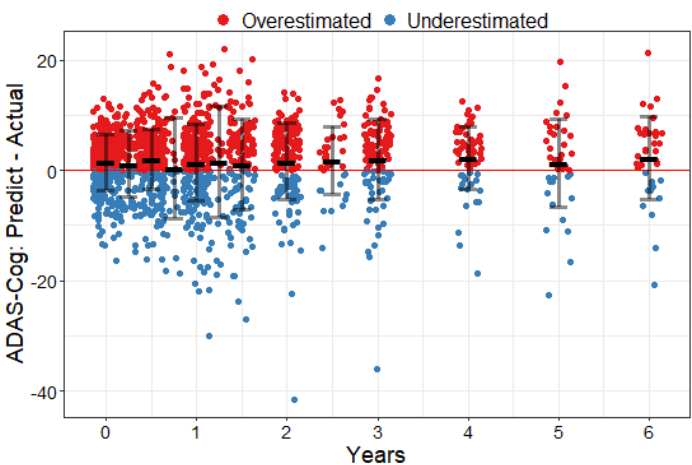
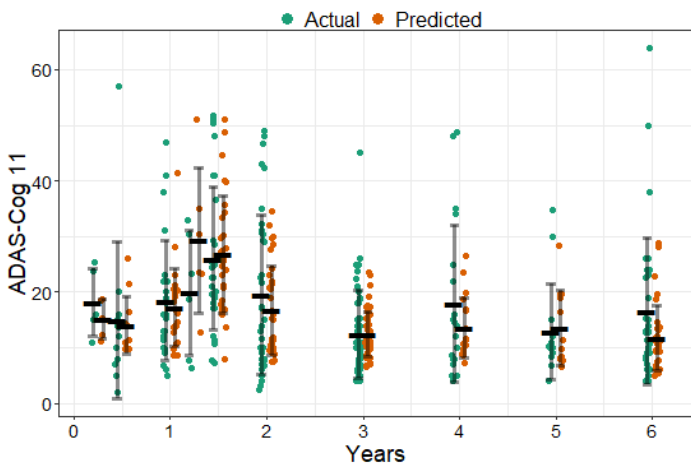
# Reference - ADAS CPAD



Whole Subject Trajectories



Final Observation Forecasts



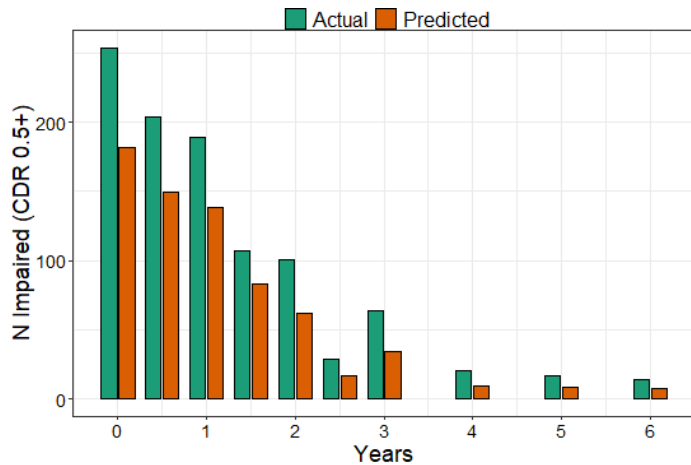
Type	RMSE	MAE	Bias	AV Bias
Trajectories	6.819	5.283	1.602	4.421
Forecasts	9.720	6.859	-1.360	4.799



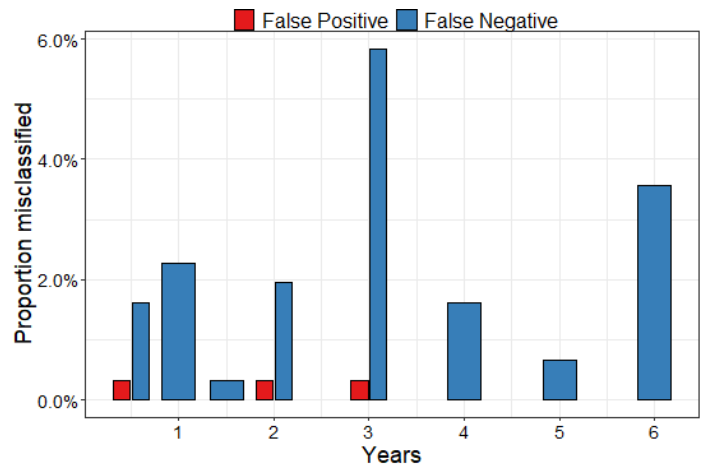
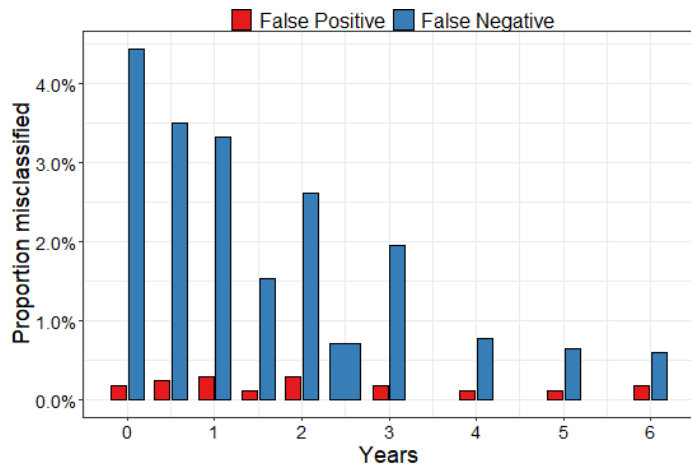
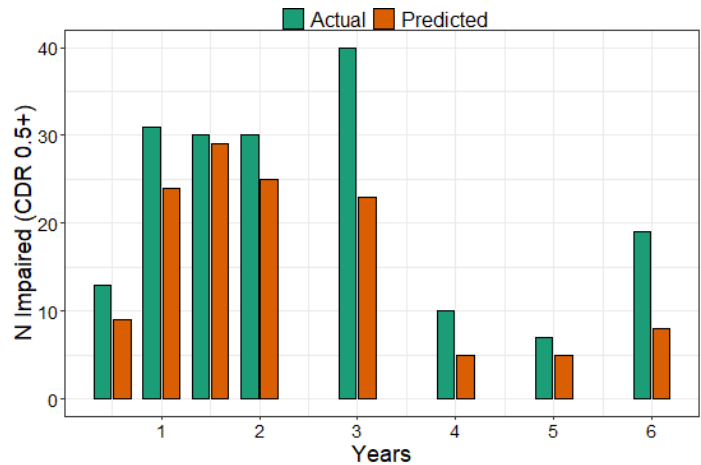
# Reference - CDR Logistic



Whole Subject Trajectories



Final Observation Forecasts



Type	Accuracy	Precision	Recall	AUC
Trajectories	0.7107	0.9579	0.6607	0.8412
Forecasts	0.7302	0.9766	0.6944	0.8246







# Reference Model Observations

- The CPAD model performed best at annual visit time points where data was more common but also exhibited greater variance and bias at later times
- The CPAD model also performed better when predicting entire trajectories compared to final observation forecasting
- The logistic regression model for impairment largely suffered from false negatives, especially at earlier time points although this is also driven by cutpoint selection
- There was a drop in predictive capacity when forecasting final observations for the CDR model but classification was more equitable between trajectories and forecasting in logistic model relative to the variance increase seen in the CPAD model for ADAS-Cog
- Regardless, both models performed reasonably well given their limited covariate sets but clearly displayed room for improvement in predictive ability



# Aim 1 Results

## Ensemble Methods



# Aim 1 - Ensemble ADAS Summary

## Whole Subject Trajectories

Metric	CPAD Reference	MERF	Single GLMER	Bagged GLMER	Boosted
RMSE	6.819	4.928	5.684	5.542	4.943
Median AE	5.283	3.792	4.427	4.323	3.858
Bias	1.602	0.066	0.427	0.423	0.107
AV Bias	4.421	3.162	3.673	3.557	3.227

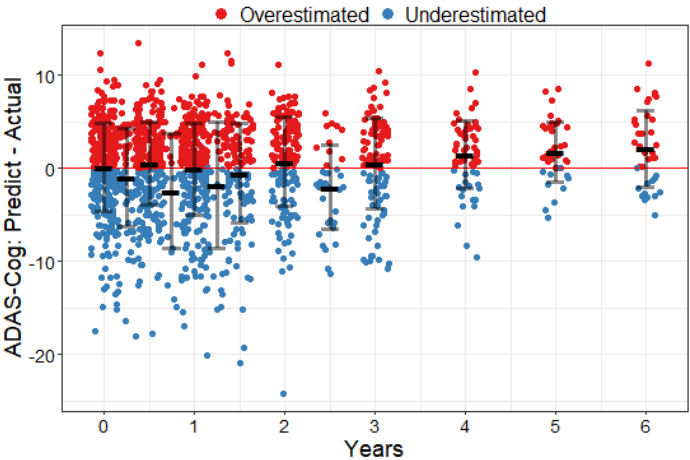
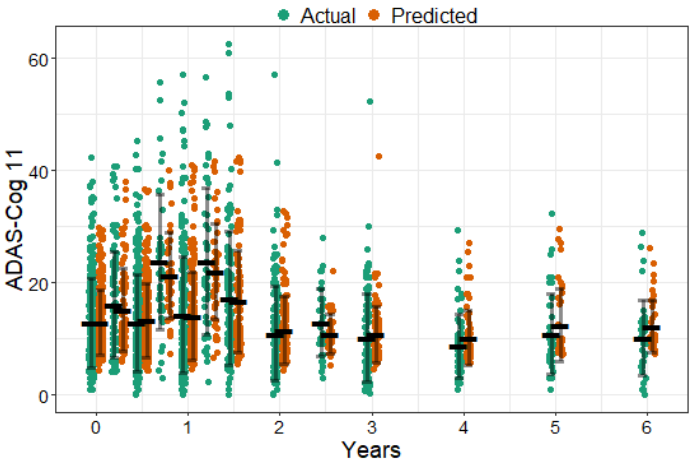
## Final Observation Forecasting

Metric	CPAD Reference	MERF	Single GLMER	Bagged GLMER	Boosted
RMSE	9.720	4.721	4.528	5.142	5.233
Median AE	6.859	3.588	3.357	3.842	3.971
Bias	-1.360	-1.151	-0.593	-0.208	-1.322
AV Bias	4.799	2.820	2.573	2.888	3.158





# Aim 1 - ADAS MERF Trajectories



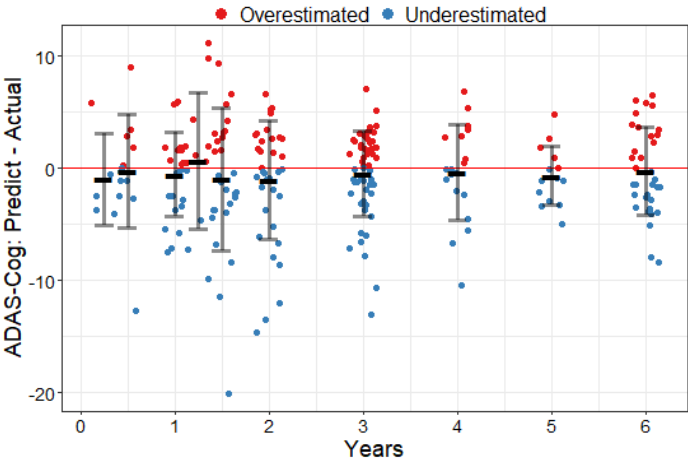
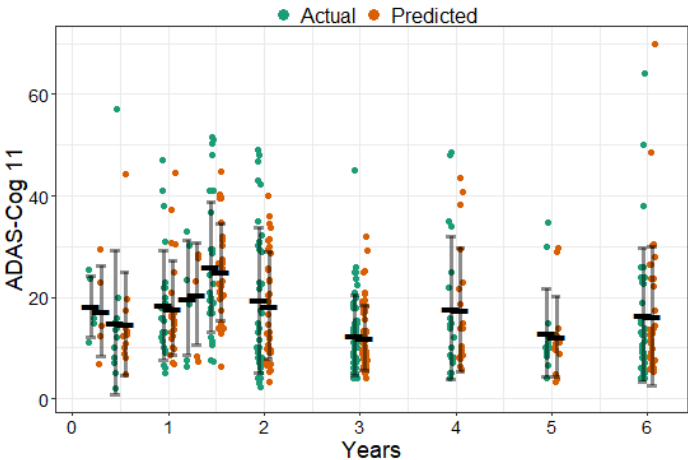
Time	RMSE	MAE	Bias	AV Bias
0	4.749	3.628	0.096	2.885
0.25	5.381	4.211	-1.007	3.407
0.5	4.486	3.439	0.491	2.916
0.75	6.605	5.080	-2.476	3.759
1	4.989	3.820	-0.127	3.226
1.25	6.968	5.668	-1.849	4.412
1.5	5.337	3.902	-0.564	3.059
2	4.830	3.829	0.641	3.490
2.5	4.907	3.869	-2.046	2.766
3	4.856	3.954	0.527	3.668
4	3.880	3.141	1.410	3.004
5	3.561	2.767	1.717	2.133
6	4.541	3.665	2.090	2.968
All	4.928	3.792	0.066	3.162

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
RMSE	4.928	6.819	-1.891	-27.73%	[-2.112, -1.658]
MAE	3.792	5.283	-1.491	-28.23%	[-1.651, -1.321]
Bias	0.066	1.602	-1.536	-95.90%	[-1.797, -1.289]
AV Bias	3.162	4.421	-1.258	-28.46%	[-1.449, -1.036]





# Aim 1 - ADAS GLMER Forecasting



Time	RMSE	MAE	Bias	AV Bias
0.25	3.768	3.336	-1.000	3.704
0.5	4.847	3.153	-0.261	2.158
1	3.662	2.828	-0.554	1.908
1.25	5.609	4.595	0.661	4.452
1.5	6.334	4.789	-1.004	3.253
2	5.290	3.837	-1.073	2.591
3	3.846	2.851	-0.506	2.166
4	4.181	3.148	-0.385	2.759
5	2.630	2.142	-0.697	1.962
6	3.842	3.253	-0.240	2.908
All	4.528	3.357	-0.593	2.573

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
RMSE	4.528	9.720	-4.999	-51.43%	[-5.490, -4.532]
MAE	3.357	6.859	-3.271	-47.69%	[-3.635, -2.899]
Bias	-0.593	-1.360	0.209	-15.37%	[-0.350, 0.796]
AV Bias	2.573	4.799	-1.979	-41.24%	[-2.393, -1.649]





# Aim 1 - Ensemble CDR Summary

## Whole Subject Trajectories

Metric	CDR Logistic	MERF	Single GLMER	Bagged GLMER	Boosted
Accuracy	0.711	0.815	0.803	0.818	0.825
Precision	0.958	0.960	0.967	0.960	0.965
Recall	0.661	0.798	0.776	0.801	0.806
AUC	0.841	0.903	0.892	0.907	0.916

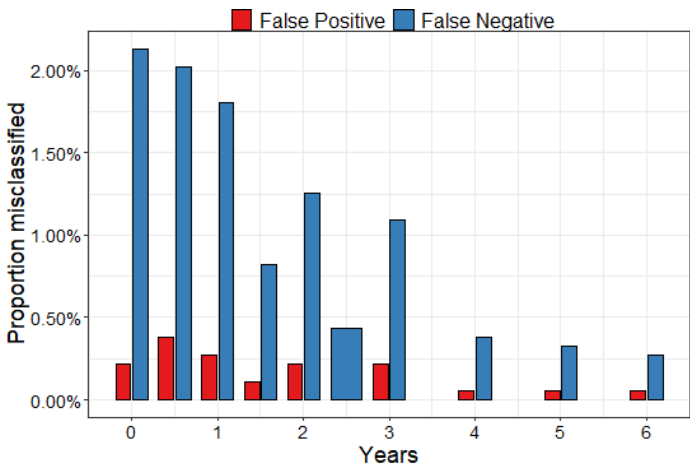
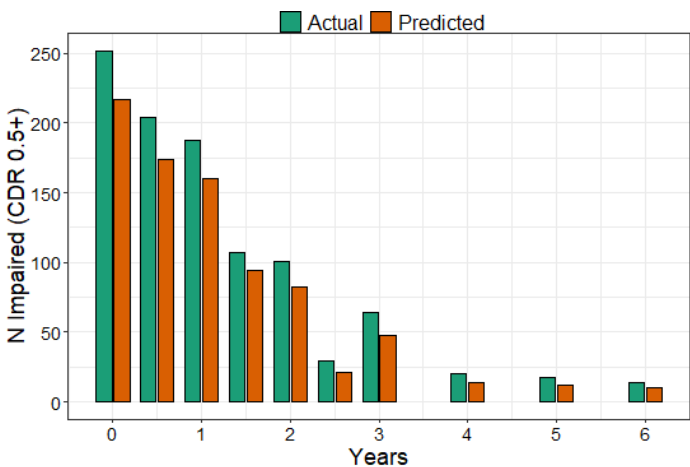
## Final Observation Forecasting

Metric	CDR Logistic	MERF	Single GLMER	Bagged GLMER	Boosted
Accuracy	0.730	0.967	0.967	0.963	0.935
Precision	0.977	0.983	0.989	0.989	0.982
Recall	0.694	0.978	0.972	0.966	0.939
AUC	0.825	0.981	0.981	0.987	0.972





# Aim 1 - CDR Boosted Trajectories



Time	Accuracy	Precision	Recall	AUC	NRI
0	0.8586	0.9816	0.8452	0.9365	0.1236
0.5	0.8233	0.9598	0.8186	0.9050	0.0412
1	0.8333	0.9688	0.8245	0.9110	0.1223
1.5	0.8482	0.9787	0.8598	0.8523	0.1028
2	0.8112	0.9512	0.7723	0.8883	0.2317
2.5	0.7241	1.0000	0.7241	1.0000	
3	0.7209	0.9167	0.6875	0.8388	0.1577
4	0.8431	0.9286	0.6500	0.8355	0.3323
5	0.7879	0.9167	0.6471	0.9375	0.3566
6	0.8125	0.9000	0.6429	0.8452	0.4683
All	0.8248	0.9651	0.8062	0.9157	0.1466

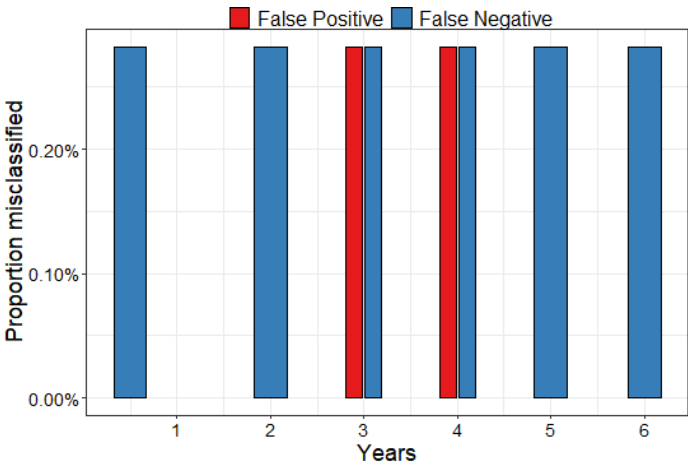
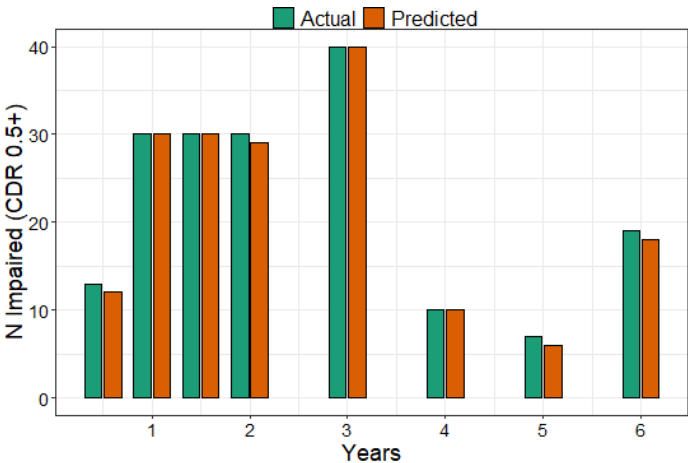
Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
Accur	0.825	0.711	0.105	14.72%	[0.082, 0.125]
Prec	0.965	0.958	0.002	0.23%	[-0.011, 0.015]
Recall	0.806	0.661	0.138	20.82%	[0.111, 0.161]
AUC	0.916	0.841	0.061	7.30%	[0.043, 0.078]

NRI: 0.147





# Aim 1 - CDR Bagged Forecasting



Time	Accuracy	Precision	Recall	AUC	NRI
0.5	0.9286	1.0000	0.9231	1.0000	1.3077
1	1.0000	1.0000	1.0000	1.0000	
1.5	1.0000	1.0000	1.0000	1.0000	
2	0.9722	1.0000	0.9667	1.0000	0.3333
3	0.9565	0.9750	0.9750	0.9500	0.4250
4	0.8667	0.9000	0.9000	0.9800	0.2000
5	0.9167	1.0000	0.8571	0.9714	0.1429
6	0.9677	1.0000	0.9474	0.9825	0.5263
All	0.9626	0.9886	0.9665	0.9868	0.3023

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
Accur	0.963	0.730	0.237	32.46%	[0.209, 0.256]
Prec	0.989	0.977	0.007	0.67%	[-0.016, 0.023]
Recall	0.966	0.694	0.283	40.78%	[0.260, 0.300]
AUC	0.987	0.825	0.157	19.01%	[0.132, 0.172]

NRI: 0.302







# Ensemble Methods Observations

- All models performed notably better with respect to both error and bias metrics when predicting ADAS-Cog when compared to the CPAD model for both whole trajectories and observational forecasting
- The MERF model was notably unbiased compared to the CPAD reference model and, alongside the boosted mixed-effect trees, gave the largest reductions in RMSE and MAE compared to CPAD
- When predicting CDR impairment across whole subject trajectories, overall AUC was found to be higher for all ensemble methods with improvements in accuracy most likely driven by improved recall with fewer false negatives
- Although recall was improved, it was still poorer compared to precision although this could largely be a consequence of this particular dataset, with a preponderance on impaired subject, as well as cutpoint selection
- Critically, when forecasting observations, which leveraged previously observed data, improvements in predictive capacity were striking across all metrics for all ensemble methods



# Aim 2 Results

## Neural Network Methods





# Aim 2 - Neural Net ADAS Summary

## Whole Subject Trajectories

Metric	CPAD Reference	FNN	1D CNN	LSTM RNN
RMSE	6.819	4.660	6.506	6.167
Median AE	5.283	3.560	4.673	4.423
Bias	1.602	0.526	1.075	1.359
AV Bias	4.421	2.830	3.427	3.146

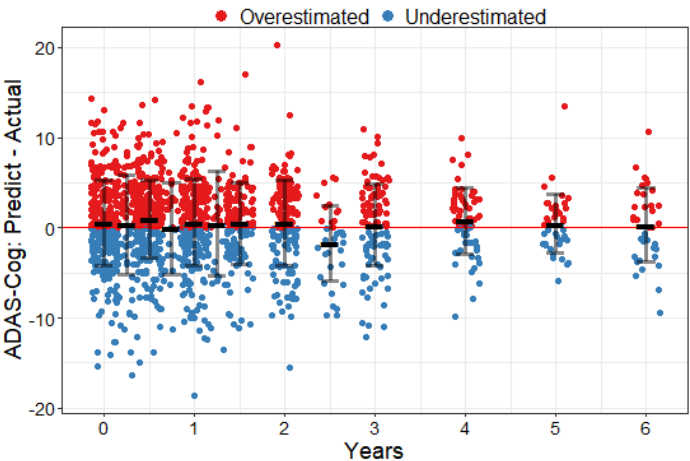
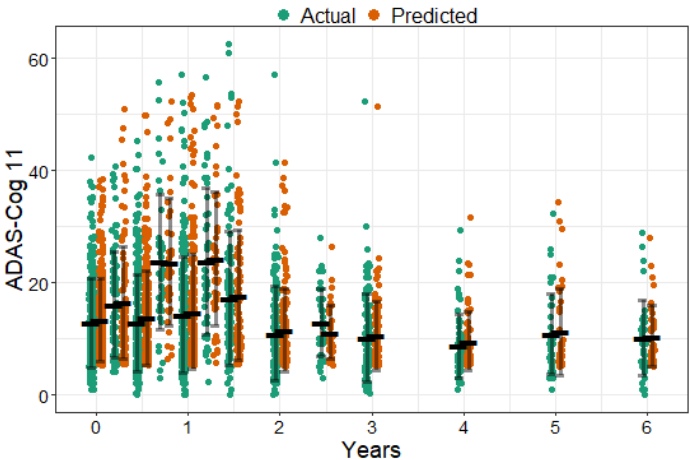
## Final Observation Forecasting:

Metric	CPAD Reference	FNN	1D CNN	LSTM RNN
RMSE	9.720	5.483	5.856	5.650
Median AE	6.859	4.257	3.823	3.727
Bias	-1.360	0.431	-1.068	-1.109
AV Bias	4.799	3.437	2.280	2.438





# Aim 2 - ADAS FNN Trajectories



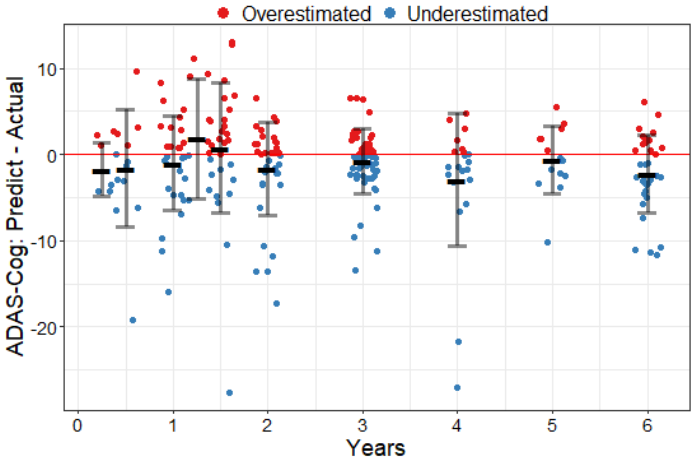
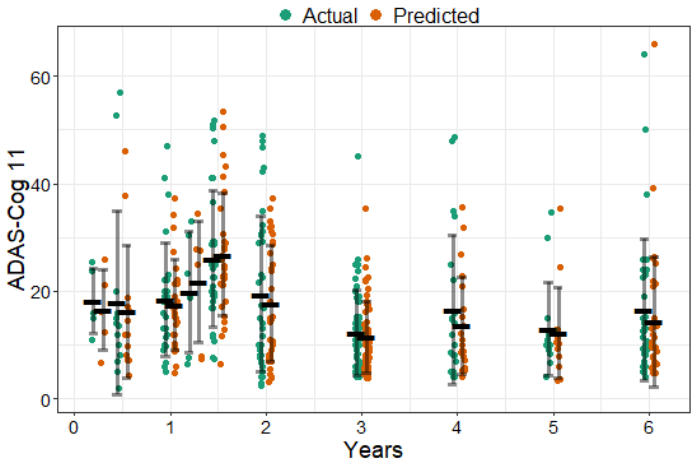
Time	RMSE	MAE	Bias	AV Bias
0	4.709	3.502	0.545	2.648
0.25	5.477	4.344	0.345	3.233
0.5	4.356	3.350	0.971	2.617
0.75	5.008	4.137	-0.088	3.336
1	4.868	3.700	0.603	2.937
1.25	5.781	4.467	0.444	3.682
1.5	4.507	3.365	0.490	2.747
2	4.731	3.684	0.571	3.299
2.5	4.465	3.429	-1.707	2.473
3	4.552	3.591	0.317	2.947
4	3.711	2.929	0.761	2.696
5	3.265	2.386	0.426	1.854
6	4.060	3.284	0.312	2.827
All	4.660	3.560	0.526	2.830

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
RMSE	4.660	6.819	-1.891	-27.73%	[-2.112, -1.658]
MAE	3.560	5.283	-1.491	-28.23%	[-1.651, -1.321]
Bias	0.526	1.602	-1.536	-95.90%	[-1.797, -1.289]
AV Bias	2.830	4.421	-1.258	-28.46%	[-1.449, -1.036]





# Aim 2 - ADAS LSTM Forecasting



Time	RMSE	MAE	Bias	AV Bias
0.25	3.315	3.070	-1.764	3.584
0.5	6.723	4.546	-1.641	2.956
1	5.481	3.950	-1.020	2.971
1.25	6.605	5.394	1.805	5.299
1.5	7.470	5.158	0.759	3.924
2	5.597	3.561	-1.712	1.982
3	3.812	2.565	-0.732	1.777
4	8.071	4.369	-2.978	1.955
5	3.810	2.934	-0.638	2.389
6	4.942	3.774	-2.230	2.824
All	5.650	3.727	-1.109	2.438

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
RMSE	5.650	9.720	-4.999	-51.43%	[-5.490, -4.532]
MAE	3.727	6.859	-3.271	-47.69%	[-3.635, -2.899]
Bias	-1.109	-1.360	0.209	-15.37%	[-0.350, 0.796]
AV Bias	2.438	4.799	-1.979	-41.24%	[-2.393, -1.649]





# Aim 2 - Neural Net CDR Summary

## Whole Subject Trajectories

Metric	CDR Logistic	FNN	1D CNN	LSTM RNN
Accuracy	0.711	0.800	0.836	0.835
Precision	0.958	0.972	0.963	0.958
Recall	0.661	0.768	0.822	0.824
AUC	0.841	0.908	0.855	0.848

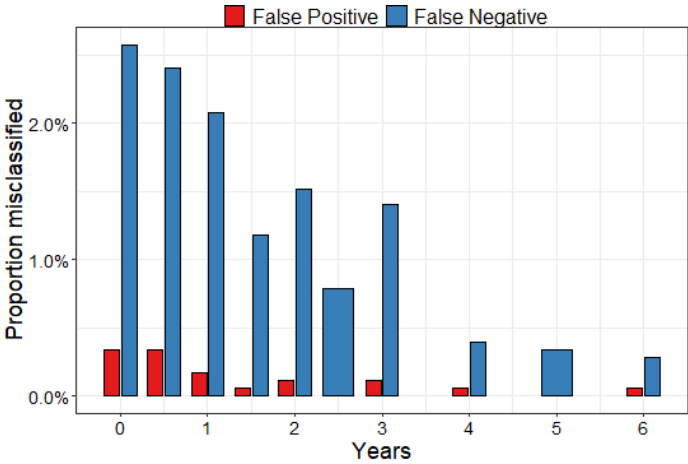
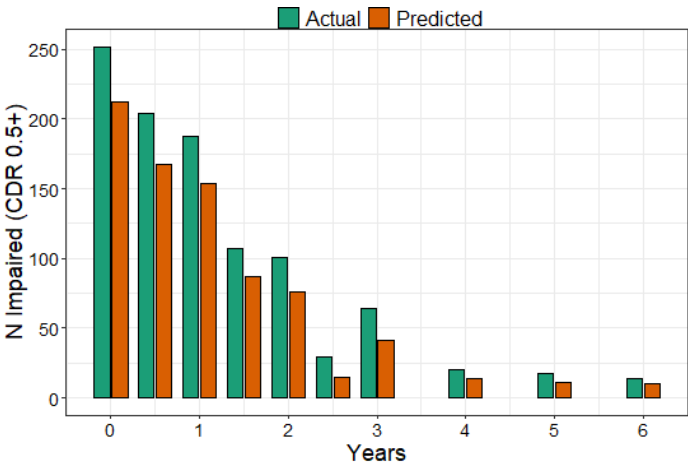
## Final Observation Forecasting:

Metric	CDR Logistic	FNN	1D CNN	LSTM RNN
Accuracy	0.730	0.832	0.949	0.972
Precision	0.977	0.980	0.994	0.983
Recall	0.694	0.816	0.944	0.983
AUC	0.825	0.924	0.958	0.949





# Aim 2 - CDR FNN Trajectories



Time	Accuracy	Precision	Recall	AUC	NRI
0	0.8289	0.9717	0.8175	0.9253	0.0574
0.5	0.8032	0.9641	0.7892	0.8962	0.0340
1	0.8246	0.9805	0.8032	0.9084	0.1511
1.5	0.8036	0.9885	0.8037	0.8449	0.2467
2	0.7972	0.9737	0.7327	0.8951	0.2397
2.5	0.5172	1.0000	0.5172	1.0000	
3	0.6860	0.9512	0.6094	0.8317	0.1705
4	0.8431	0.9286	0.6500	0.8403	0.3323
5	0.8182	1.0000	0.6471	0.8897	0.4191
6	0.8125	0.9000	0.6429	0.8571	0.4683
All	0.8003	0.9720	0.7681	0.9075	0.1343

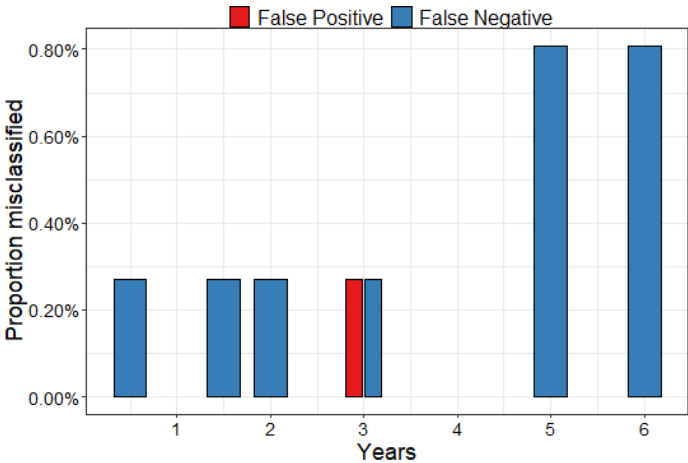
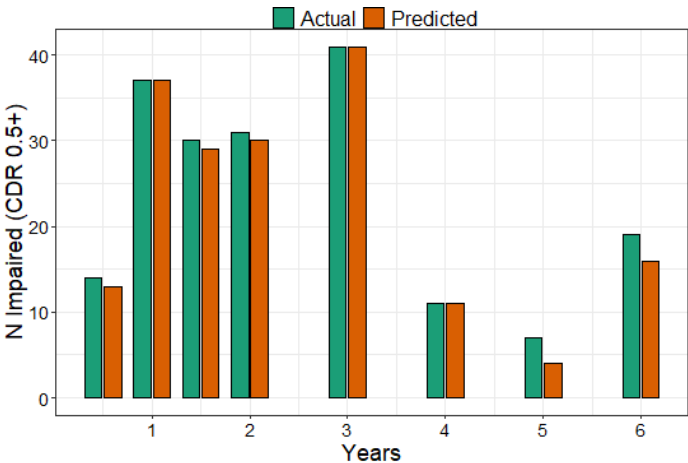
Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
Accur	0.800	0.711	0.105	14.72%	[0.082, 0.125]
Prec	0.972	0.958	0.002	0.23%	[-0.011, 0.015]
Recall	0.768	0.661	0.138	20.82%	[0.111, 0.161]
AUC	0.908	0.841	0.061	7.30%	[0.043, 0.078]

NRI: 0.134





# Aim 2 - CDR CNN Forecasting



Time	Accuracy	Precision	Recall	AUC	NRI
0.5	0.9286	1.0000	0.9231	0.9615	0.2308
1	1.0000	1.0000	1.0000	1.0000	
1.5	0.9667	1.0000	0.9667	1.0000	
2	0.9722	1.0000	0.9667	0.9833	0.8667
3	0.9565	0.9750	0.9750	0.9042	0.6917
4	1.0000	1.0000	1.0000	1.0000	1.0000
5	0.7500	1.0000	0.5714	0.7857	0.6000
6	0.9032	1.0000	0.8421	0.9211	0.5746
All	0.9486	0.9941	0.9441	0.9578	0.7868

Metric	ML Model	Ref Model	Delta Boot	Perc Diff	95% CI Boot
Accur	0.949	0.730	0.237	32.46%	[0.209, 0.256]
Prec	0.994	0.977	0.007	0.67%	[-0.016, 0.023]
Recall	0.944	0.694	0.283	40.78%	[0.260, 0.300]
AUC	0.958	0.825	0.157	19.01%	[0.132, 0.172]

NRI: 0.787







# Neural Network Observations

- Although it made no direct use of sequential data, the FNN model observed strong error and bias performance, most notably when predicting whole trajectories but still performed well with forecasting
- Of interest, MAE was improved in forecasting final observations when using the sequential methods suggesting RMSE and bias were especially impacted by outlier values
- Regardless, all neural network models performed better compared to the CPAD reference for both ADAS trajectories and forecasting
- Impairment classification was comparable for whole subject trajectories across the neural network designs although observational forecasting was much better for the sequential models compared to the feed-forward network
- Issues with trajectory prediction may result from the iterative nature of the sequential neural network model prediction and the reliance on a baseline FNN to seed the trajectories



# Aims 1 & 2

## Cross Model Comparisons

# ML/SL Comparison - ADAS Regression



## RMSE - Whole Trajectories

	Reference	FNN	MERF	Boosted	Bagged	GLMER	LSTM	CNN
Reference	6.819	-2.160 (-31.67%) <0.001	-1.891 (-27.73%) <0.001	-1.876 (-27.51%) <0.001	-1.277 (-18.73%) <0.001	-1.135 (-16.64%) <0.001	-0.652 (-9.56%) <0.001	-0.313 (-4.59%) 0.041
FNN	[-2.351, -1.972]	4.660	-0.269 (-5.45%) 0.004	-0.284 (-5.74%) 0.007	-0.882 (-15.92%) <0.001	-1.025 (-18.03%) <0.001	-1.508 (-24.45%) <0.001	-1.847 (-28.38%) <0.001
MERF	[-2.112, -1.658]	[-0.469, -0.062]	4.928	-0.015 (-0.31%) 0.457	-0.614 (-11.07%) <0.001	-0.756 (-13.30%) <0.001	-1.239 (-20.09%) <0.001	-1.578 (-24.26%) <0.001
Boosted	[-2.097, -1.662]	[-0.481, -0.069]	[-0.247, 0.207]	4.943	-0.598 (-10.80%) <0.001	-0.741 (-13.03%) <0.001	-1.224 (-19.84%) <0.001	-1.563 (-24.02%) <0.001
Bagged	[-1.503, -1.041]	[-1.102, -0.675]	[-0.824, -0.393]	[-0.813, -0.387]	5.542	-0.142 (-2.51%) 0.107	-0.625 (-10.14%) <0.001	-0.965 (-14.83%) <0.001
GLMER	[-1.378, -0.890]	[-1.212, -0.831]	[-0.977, -0.523]	[-0.951, -0.531]	[-0.387, 0.093]	5.684	-0.483 (-7.83%) <0.001	-0.822 (-12.64%) <0.001
LSTM	[-0.996, -0.260]	[-1.699, -1.298]	[-1.463, -1.018]	[-1.429, -1.031]	[-0.846, -0.395]	[-0.718, -0.245]	6.167	-0.339 (-5.21%) 0.034
CNN	[-0.705, 0.050]	[-2.054, -1.655]	[-1.800, -1.341]	[-1.773, -1.370]	[-1.181, -0.740]	[-1.053, -0.572]	[-0.682, 0.013]	6.506

# ML/SL Comparison - ADAS Regression

## RMSE - Observation Forecasts



	Reference	GLMER	MERF	Bagged	Boosted	FNN	LSTM	CNN
Reference	9.720	-5.192 (-53.42%) <0.001	-4.999 (-51.43%) <0.001	-4.577 (-47.09%) <0.001	-4.487 (-46.16%) <0.001	-4.237 (-43.59%) <0.001	-4.070 (-41.88%) <0.001	-3.864 (-39.76%) <0.001
GLMER	[-5.767, -4.665]	4.528	-0.193 (-4.09%) 0.235	-0.615 (-11.95%) 0.019	-0.705 (-13.47%) 0.010	-0.955 (-17.42%) 0.001	-1.122 (-19.86%) <0.001	-1.328 (-22.68%) <0.001
MERF	[-5.490, -4.532]	[-0.754, 0.377]	4.721	-0.421 (-8.20%) 0.049	-0.512 (-9.78%) 0.021	-0.762 (-13.90%) 0.001	-0.929 (-16.44%) 0.001	-1.135 (-19.38%) <0.001
Bagged	[-5.154, -4.026]	[-1.170, -0.030]	[-0.934, 0.095]	5.142	-0.090 (-1.73%) 0.383	-0.340 (-6.21%) 0.120	-0.507 (-8.98%) 0.046	-0.713 (-12.18%) 0.011
Boosted	[-5.073, -3.937]	[-1.289, -0.101]	[-1.004, -0.006]	[-0.677, 0.544]	5.233	-0.250 (-4.56%) 0.190	-0.417 (-7.38%) 0.076	-0.623 (-10.63%) 0.024
FNN	[-4.814, -3.699]	[-1.558, -0.380]	[-1.284, -0.265]	[-0.950, 0.208]	[-0.858, 0.303]	5.483	-0.167 (-2.95%) 0.270	-0.373 (-6.37%) 0.116
LSTM	[-5.099, -3.156]	[-1.682, -0.539]	[-1.418, -0.408]	[-1.131, 0.072]	[-1.019, 0.151]	[-0.748, 0.393]	5.650	-0.206 (-3.52%) 0.304
CNN	[-4.822, -2.844]	[-1.886, -0.728]	[-1.693, -0.611]	[-1.333, -0.118]	[-1.197, -0.014]	[-0.906, 0.218]	[-1.114, 0.699]	5.856



# ML/SL Comparison - ADAS Regression



## AV Bias - Whole Trajectories

	Reference	FNN	LSTM	MERF	Boosted	CNN	Bagged	GLMER
Reference	4.421	-1.591 (-35.99%) <0.001	-1.275 (-28.84%) <0.001	-1.258 (-28.46%) <0.001	-1.194 (-27.01%) <0.001	-0.994 (-22.48%) <0.001	-0.864 (-19.55%) <0.001	-0.748 (-16.92%) <0.001
FNN	[-1.733, -1.429]	2.830	-0.316 (-10.04%) 0.001	-0.332 (-10.51%) <0.001	-0.397 (-12.29%) <0.001	-0.597 (-17.42%) <0.001	-0.727 (-20.43%) <0.001	-0.843 (-22.95%) <0.001
LSTM	[-1.446, -1.092]	[-0.456, -0.140]	3.146	-0.017 (-0.53%) 0.354	-0.081 (-2.51%) 0.120	-0.281 (-8.21%) 0.002	-0.411 (-11.55%) <0.001	-0.527 (-14.36%) <0.001
MERF	[-1.449, -1.036]	[-0.478, -0.161]	[-0.203, 0.167]	3.162	-0.064 (-1.99%) 0.253	-0.265 (-7.72%) 0.008	-0.394 (-11.08%) <0.001	-0.510 (-13.90%) <0.001
Boosted	[-1.424, -1.016]	[-0.526, -0.220]	[-0.254, 0.090]	[-0.247, 0.128]	3.227	-0.201 (-5.85%) 0.017	-0.330 (-9.28%) <0.001	-0.446 (-12.15%) <0.001
CNN	[-1.230, -0.771]	[-0.736, -0.429]	[-0.465, -0.116]	[-0.435, -0.034]	[-0.400, -0.034]	3.427	-0.129 (-3.64%) 0.160	-0.246 (-6.69%) 0.013
Bagged	[-1.079, -0.604]	[-0.867, -0.565]	[-0.597, -0.227]	[-0.576, -0.204]	[-0.560, -0.169]	[-0.361, 0.086]	3.557	-0.116 (-3.17%) 0.170
GLMER	[-0.992, -0.532]	[-0.972, -0.680]	[-0.704, -0.344]	[-0.701, -0.288]	[-0.674, -0.281]	[-0.483, -0.028]	[-0.338, 0.158]	3.673

# ML/SL Comparison - ADAS Regression

## AV Bias - Observation Forecasts



	Reference	CNN	LSTM	GLMER	MERF	Bagged	Boosted	FNN
Reference	4.799	-2.519 (-52.49%) <0.001	-2.361 (-49.20%) <0.001	-2.226 (-46.39%) <0.001	-1.979 (-41.24%) <0.001	-1.911 (-39.83%) <0.001	-1.641 (-34.20%) <0.001	-1.362 (-28.38%) <0.001
CNN	[-2.857, -2.011]	2.280	-0.158 (-6.47%) 0.299	-0.293 (-11.37%) 0.110	-0.540 (-19.14%) 0.008	-0.607 (-21.04%) 0.005	-0.878 (-27.79%) <0.001	-1.157 (-33.66%) <0.001
LSTM	[-2.830, -1.990]	[-0.522, 0.347]	2.438	-0.135 (-5.24%) 0.241	-0.382 (-13.54%) 0.017	-0.450 (-15.57%) 0.007	-0.720 (-22.80%) <0.001	-0.999 (-29.07%) <0.001
GLMER	[-2.464, -1.839]	[-0.678, 0.212]	[-0.552, 0.222]	2.573	-0.247 (-8.76%) 0.102	-0.315 (-10.90%) 0.024	-0.585 (-18.53%) <0.001	-0.864 (-25.14%) <0.001
MERF	[-2.393, -1.649]	[-0.904, -0.096]	[-0.851, -0.025]	[-0.485, 0.058]	2.820	-0.068 (-2.35%) 0.405	-0.338 (-10.70%) 0.030	-0.617 (-17.95%) 0.001
Bagged	[-2.330, -1.677]	[-0.972, -0.103]	[-0.893, -0.093]	[-0.553, -0.010]	[-0.468, 0.290]	2.888	-0.270 (-8.56%) 0.009	-0.549 (-15.98%) <0.001
Boosted	[-2.159, -1.413]	[-1.263, -0.430]	[-1.163, -0.363]	[-0.823, -0.249]	[-0.752, 0.019]	[-0.647, -0.036]	3.158	-0.279 (-8.12%) 0.019
FNN	[-1.962, -0.778]	[-1.524, -0.697]	[-1.442, -0.619]	[-1.102, -0.528]	[-1.031, -0.260]	[-0.926, -0.315]	[-0.824, -0.027]	3.437

# ML/SL Comparison - CDR Classification



## Accuracy - Whole Trajectories

	Reference	CNN	LSTM	Boosted	Bagged	MERF	GLMER	FNN
Reference	<b>0.711</b>	0.126 (17.68%) <0.001	0.124 (17.45%) <0.001	0.114 (16.05%) <0.001	0.107 (15.05%) <0.001	0.105 (14.72%) <0.001	0.093 (13.06%) <0.001	0.090 (12.61%) <0.001
CNN	[0.105, 0.147]	<b>0.836</b>	0.002 (0.20%) 0.439	0.012 (1.40%) 0.120	0.019 (2.28%) 0.048	0.021 (2.58%) 0.022	0.033 (4.09%) 0.002	0.036 (4.50%) <0.001
LSTM	[0.103, 0.144]	[-0.018, 0.021]	<b>0.835</b>	0.010 (1.20%) 0.141	0.017 (2.08%) 0.053	0.019 (2.38%) 0.035	0.031 (3.89%) 0.002	0.034 (4.30%) 0.002
Boosted	[0.090, 0.135]	[-0.010, 0.031]	[-0.010, 0.030]	<b>0.825</b>	0.007 (0.87%) 0.235	0.009 (1.16%) 0.214	0.021 (2.65%) 0.023	0.024 (3.06%) 0.015
Bagged	[0.086, 0.126]	[-0.003, 0.039]	[-0.004, 0.038]	[-0.013, 0.029]	<b>0.818</b>	0.002 (0.29%) 0.422	0.014 (1.77%) 0.096	0.017 (2.17%) 0.062
MERF	[0.082, 0.125]	[0.001, 0.043]	[-0.003, 0.039]	[-0.013, 0.029]	[-0.018, 0.023]	<b>0.815</b>	0.012 (1.47%) 0.147	0.015 (1.87%) 0.084
GLMER	[0.071, 0.114]	[0.011, 0.055]	[0.011, 0.053]	[0.001, 0.042]	[-0.006, 0.035]	[-0.011, 0.034]	<b>0.803</b>	0.003 (0.39%) 0.402
FNN	[0.068, 0.111]	[0.015, 0.057]	[0.014, 0.055]	[0.003, 0.044]	[-0.005, 0.039]	[-0.006, 0.036]	[-0.020, 0.024]	<b>0.800</b>

# ML/SL Comparison - CDR Classification



## Accuracy - Observation Forecasts

	Reference	LSTM	MERF	GLMER	Bagged	CNN	Boosted	FNN
Reference	<b>0.730</b>	0.242 (33.10%) <0.001	0.237 (32.46%) <0.001	0.237 (32.46%) <0.001	0.232 (31.82%) <0.001	0.218 (29.90%) <0.001	0.204 (27.98%) <0.001	0.102 (13.91%) <0.001
LSTM	[0.221, 0.261]	<b>0.972</b>	0.005 (0.48%) 0.242	0.005 (0.48%) 0.254	0.009 (0.97%) 0.162	0.023 (2.46%) 0.020	0.037 (4.00%) <0.001	0.140 (16.85%) <0.001
MERF	[0.209, 0.256]	[-0.016, 0.024]	<b>0.967</b>	0.000 (0.00%) 0.549	0.005 (0.49%) 0.432	0.019 (1.97%) 0.090	0.033 (3.50%) 0.012	0.136 (16.29%) <0.001
GLMER	[0.214, 0.260]	[-0.016, 0.024]	[-0.028, 0.023]	<b>0.967</b>	0.005 (0.49%) 0.406	0.019 (1.97%) 0.100	0.033 (3.50%) 0.008	0.136 (16.29%) <0.001
Bagged	[0.204, 0.256]	[-0.012, 0.028]	[-0.019, 0.023]	[-0.019, 0.023]	<b>0.963</b>	0.014 (1.48%) 0.186	0.028 (3.00%) 0.027	0.131 (15.73%) <0.001
CNN	[0.194, 0.248]	[0.003, 0.043]	[-0.009, 0.042]	[-0.009, 0.042]	[-0.014, 0.037]	<b>0.949</b>	0.014 (1.50%) 0.144	0.117 (14.04%) <0.001
Boosted	[0.172, 0.232]	[0.017, 0.057]	[0.005, 0.056]	[0.009, 0.051]	[0.000, 0.051]	[-0.015, 0.039]	<b>0.935</b>	0.103 (12.36%) <0.001
FNN	[0.045, 0.148]	[0.119, 0.159]	[0.112, 0.159]	[0.107, 0.159]	[0.103, 0.154]	[0.088, 0.146]	[0.070, 0.131]	<b>0.832</b>



# ML/SL Comparison - CDR Classification



## AUC - Whole Trajectories

	Reference	Boosted	FNN	Bagged	MERF	GLMER	CNN	LSTM
Reference	<b>0.841</b>	0.075 (8.87%) <0.001	0.066 (7.89%) <0.001	0.066 (7.80%) <0.001	0.061 (7.30%) <0.001	0.051 (6.09%) <0.001	0.013 (1.60%) <0.001	0.007 (0.84%) <0.001
Boosted	[0.059, 0.090]	<b>0.916</b>	0.008 (0.90%) 0.161	0.009 (0.99%) 0.116	0.013 (1.46%) 0.051	0.023 (2.62%) <0.001	0.061 (7.15%) <0.001	0.068 (7.96%) <0.001
FNN	[0.051, 0.082]	[-0.009, 0.024]	<b>0.908</b>	0.001 (0.09%) 0.471	0.005 (0.55%) 0.245	0.015 (1.70%) 0.042	0.053 (6.19%) <0.001	0.059 (7.00%) <0.001
Bagged	[0.047, 0.082]	[-0.006, 0.025]	[-0.016, 0.017]	<b>0.907</b>	0.004 (0.46%) 0.326	0.014 (1.61%) 0.056	0.052 (6.09%) <0.001	0.059 (6.90%) <0.001
MERF	[0.043, 0.078]	[-0.003, 0.028]	[-0.010, 0.022]	[-0.013, 0.022]	<b>0.903</b>	0.010 (1.14%) 0.136	0.048 (5.61%) <0.001	0.054 (6.41%) <0.001
GLMER	[0.031, 0.068]	[0.007, 0.039]	[-0.002, 0.030]	[-0.003, 0.031]	[-0.009, 0.027]	<b>0.892</b>	0.038 (4.42%) <0.001	0.044 (5.21%) <0.001
CNN	[0.022, 0.066]	[0.044, 0.076]	[0.036, 0.068]	[0.034, 0.069]	[0.030, 0.065]	[0.018, 0.056]	<b>0.855</b>	0.006 (0.76%) 0.001
LSTM	[0.031, 0.072]	[0.050, 0.082]	[0.043, 0.076]	[0.040, 0.076]	[0.036, 0.071]	[0.024, 0.062]	[0.016, 0.058]	<b>0.848</b>



# ML/SL Comparison - CDR Classification



## AUC - Observation Forecasts

	Reference	Bagged	MERF	GLMER	Boosted	CNN	LSTM	FNN
Reference	0.825	0.162 (19.66%) <0.001	0.157 (19.01%) <0.001	0.157 (19.01%) <0.001	0.148 (17.90%) <0.001	0.133 (16.15%) <0.001	0.124 (15.06%) <0.001	0.099 (12.00%) <0.001
Bagged	[0.148, 0.173]	0.987	0.005 (0.55%) 0.217	0.005 (0.55%) 0.195	0.015 (1.49%) 0.028	0.029 (3.02%) 0.001	0.038 (4.00%) <0.001	0.063 (6.84%) <0.001
MERF	[0.132, 0.172]	[-0.011, 0.017]	0.981	0.000 (0.00%) 0.454	0.009 (0.94%) 0.182	0.024 (2.46%) 0.027	0.033 (3.43%) 0.005	0.058 (6.26%) <0.001
GLMER	[0.133, 0.173]	[-0.011, 0.016]	[-0.025, 0.016]	0.981	0.009 (0.94%) 0.176	0.024 (2.46%) 0.025	0.033 (3.43%) 0.007	0.058 (6.26%) <0.001
Boosted	[0.123, 0.168]	[-0.001, 0.026]	[-0.016, 0.025]	[-0.014, 0.025]	0.972	0.014 (1.51%) 0.121	0.023 (2.47%) 0.035	0.049 (5.27%) 0.002
CNN	[0.147, 0.174]	[0.012, 0.040]	[-0.001, 0.039]	[-0.000, 0.039]	[-0.011, 0.035]	0.958	0.009 (0.95%) <0.001	0.034 (3.71%) <0.001
LSTM	[0.133, 0.172]	[0.021, 0.049]	[0.007, 0.048]	[0.008, 0.049]	[-0.004, 0.043]	[0.024, 0.049]	0.949	0.025 (2.73%) <0.001
FNN	[0.058, 0.131]	[0.048, 0.074]	[0.032, 0.074]	[0.035, 0.074]	[0.024, 0.068]	[0.049, 0.075]	[0.034, 0.071]	0.924





# **Aim 3 Results**

## **Random Effects Imputation and Impact on ADAS-Cog Prediction**



# Aim 3 - Meta-DB Imputation

	RMSE			Median AE			AV Bias		
Random Effects Design	CPAD Parameters	De novo BR Model	De Novo MERF Model	CPAD Parameters	De novo BR Model	De Novo MERF Model	CPAD Parameters	De novo BR Model	De Novo MERF Model
Whole Trajectories									
Fixed Effects	6.43 ± 0.196	6.42 ± 0.232	6.32 ± 0.239	4.92 ± 0.125	4.60 ± 0.137	4.64 ± 0.140	1.12 ± 0.312	0.60 ± 0.414	0.21 ± 0.259
Imputed Random Effects	12.54 ± 0.124	7.79 ± 0.191	8.14 ± 0.179	9.19 ± 0.098	5.73 ± 0.117	6.21 ± 0.120	3.03 ± 0.308	0.22 ± 0.286	0.20 ± 0.212
Observation Forecasts									
Fixed Effects	7.70 ± 0.348	7.78 ± 0.385	7.47 ± 0.380	5.79 ± 0.212	5.46 ± 0.223	5.43 ± 0.215	0.57 ± 0.482	1.02 ± 0.529	0.63 ± 0.277
Imputed Random Effects	16.52 ± 0.195	8.93 ± 0.339	10.07 ± 0.291	12.20 ± 0.155	6.46 ± 0.204	7.81 ± 0.192	3.94 ± 0.497	0.58 ± 0.525	0.39 ± 0.454
Fitted Random Effects		4.24 ± 0.238	4.44 ± 0.249		3.00 ± 0.129	3.15 ± 0.137		0.25 ± 0.256	0.27 ± 0.275



# Aim 3 - Meta-DB Design Comparison

## Whole Subject Trajectories

	CPAD FE Only	CPAD Imputed	De Novo BR FE Only	De Novo BR Imputed	De Novo MERF FE Only	De Novo MERF Imputed
CPAD FE Only		-171.0% [-201.4% - -153.5%]	85.8% [52.2% - 141.6%]		434.1% [287.7% - 764.7%]	
CPAD Imputed	95.0% [92.9% - 96.9%]			1284.6% [825.4% - 1929.2%]		1443.7% [1032.4% - 2052.6%]
De Novo BR FE Only	-0.1% [-1.6% - 1.3%]			174.9% [68.4% - 292.9%]	187.4% [98.0% - 370.5%]	
De Novo BR Imputed		-60.9% [-62.3% - -59.6%]	21.3% [19.6% - 23.1%]			11.5% [-38.1% - 84.9%]
De Novo MERF FE Only	-1.8% [-3.3% - -0.3%]		-1.6% [-3.3% - -0.1%]			6.6% [-57.9% - 65.9%]
De Novo MERF Imputed		-54.0% [-55.2% - -52.9%]		4.5% [3.5% - 5.6%]	28.8% [27.0% - 30.6%]	

RMSE comparisons in lower half of matrix with AV bias comparisons in upper half  
Comparisons limited to within the same model type or the same random effects design  
Direction of comparisons is column model relative to row model





# Aim 3 - Meta-DB Design Comparison

## Observation Forecasting

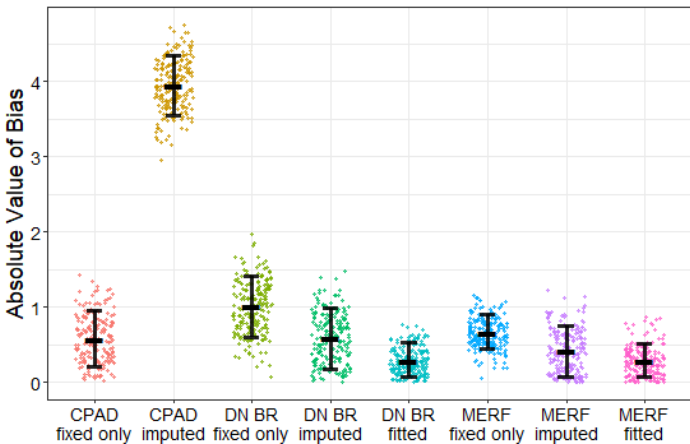
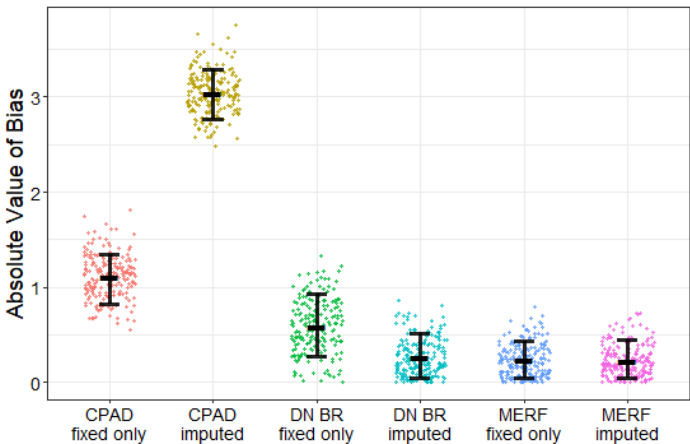
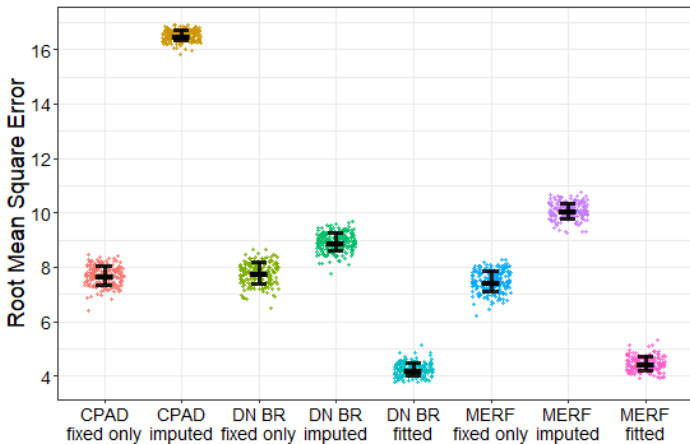
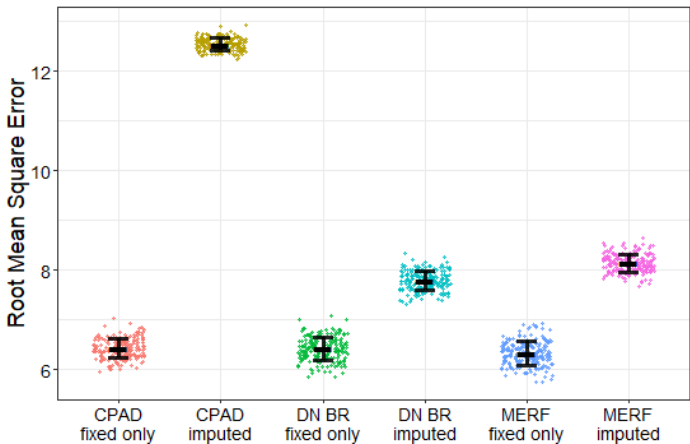
	CPAD FE Only	CPAD Imputed	De Novo BR FE Only	De Novo BR Imputed	De Novo BR Fitted	DN MERF FE Only	DN MERF Imputed	DN MERF Fitted
CPAD FE Only		-593.4% [-862.4% - -448.9%]	-79.0% [-151.3% - -32.7%]			-11.8% [-59.0% - 17.3%]		
CPAD Imputed	114.7% [111.5% - 117.8%]			575.9% [422.7% - 849.0%]			899.3% [625.7% - 1462.8%]	
De Novo BR FE Only	1.0% [-1.0% - 3.2%]			74.5% [25.7% - 143.2%]	305.2% [196.1% - 439.9%]	60.1% [28.2% - 92.2%]		
De Novo BR Imputed		-85.0% [-87.3% - -82.8%]	14.9% [12.6% - 17.2%]		132.3% [50.3% - 238.7%]		47.8% [-3.5% - 152.3%]	
De Novo BR Fitted			-83.4% [-87.5% - -79.3%]	-110.7% [-114.8% - -106.3%]				-8.5% [-52.8% - 35.9%]
DN MERF FE Only	-3.0% [-5.3% - -0.8%]		-4.1% [-6.4% - -1.9%]				61.1% [15.6% - 151.0%]	133.4% [79.0% - 213.8%]
DN MERF Imputed		-64.1% [-65.8% - -62.4%]		12.7% [11.2% - 14.4%]		34.8% [32.5% - 37.4%]		44.9% [-6.3% - 124.1%]
DN MERF Fitted					4.8% [2.2% - 7.4%]	-68.1% [-72.0% - -64.2%]	-126.6% [-130.7% - -122.4%]	

RMSE comparisons in lower half of matrix with AV bias comparisons in upper half





# Aim 3 - Meta-DB Metrics





# Aim 3 - Synthetic Data Imputation

	RMSE			Median AE			AV Bias		
Random Effects Design	CPAD Parameters	De novo BR Model	De Novo MERF Model	CPAD Parameters	De novo BR Model	De Novo MERF Model	CPAD Parameters	De novo BR Model	De Novo MERF Model
Whole Trajectories									
Fixed Effects	12.88 ± 1.552	11.80 ± 1.537	11.54 ± 1.441	9.54 ± 1.105	8.69 ± 1.071	8.84 ± 1.021	1.14 ± 1.677	2.24 ± 2.170	1.11 ± 1.387
Imputed Random Effects	18.77 ± 0.904	15.75 ± 1.230	15.85 ± 1.373	13.89 ± 0.732	11.93 ± 0.942	12.25 ± 1.052	2.34 ± 2.120	1.08 ± 1.421	1.12 ± 1.423
Observation Forecasts									
Fixed Effects	15.15 ± 2.101	13.04 ± 1.926	12.41 ± 1.849	11.28 ± 1.590	9.76 ± 1.432	9.47 ± 1.341	2.01 ± 2.435	2.23 ± 2.241	1.77 ± 2.043
Imputed Random Effects	23.35 ± 1.267	17.03 ± 1.619	17.15 ± 1.685	17.88 ± 1.049	13.05 ± 1.263	13.51 ± 1.322	3.11 ± 2.829	1.19 ± 1.326	1.08 ± 1.353
Fitted Random Effects		2.89 ± 1.548	2.65 ± 1.364		1.28 ± 0.492	1.19 ± 0.441		0.38 ± 0.470	0.44 ± 0.414







# Aim 3 - Synthetic Design Comparison

## Whole Subject Trajectories

	CPAD FE Only	CPAD Imputed	De Novo BR FE Only	De Novo BR Imputed	De Novo MERF FE Only	De Novo MERF Imputed
CPAD FE Only		-106.3% [-175.6% - -48.2%]	-97.5% [-161.1% - -40.2%]		1.9% [-33.3% - 48.1%]	
CPAD Imputed	45.7% [42.2% - 49.5%]			116.5% [70.4% - 194.1%]		108.3% [58.4% - 177.1%]
De Novo BR FE Only	-9.1% [-13.2% - -5.6%]			107.3% [57.8% - 185.7%]	101.2% [58.0% - 172.1%]	
De Novo BR Imputed		-19.1% [-21.4% - -16.9%]	33.5% [29.5% - 37.4%]			-3.9% [-47.2% - 30.1%]
De Novo MERF FE Only	-11.6% [-15.5% - -7.8%]		-2.3% [-5.7% - 1.3%]			-0.9% [-43.2% - 35.8%]
De Novo MERF Imputed		-18.4% [-20.8% - -16.1%]		0.6% [-1.9% - 3.0%]	37.4% [33.0% - 41.9%]	

RMSE comparisons in lower half of matrix with AV bias comparisons in upper half  
Comparisons limited to within the same model type or the same random effects design  
Direction of comparisons is column model relative to row model





# Aim 3 - Synthetic Design Comparison

## Observation Forecasting

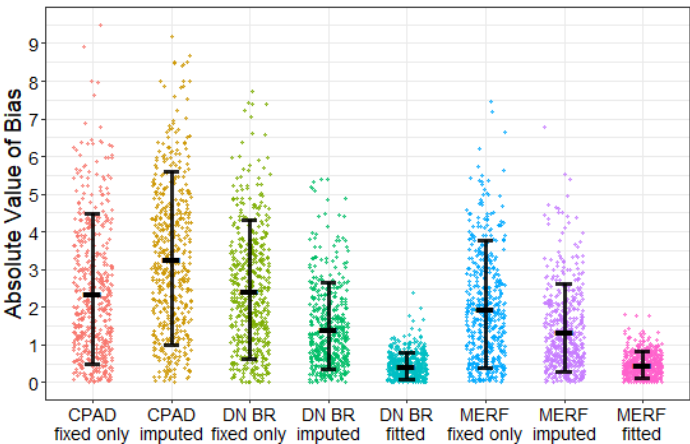
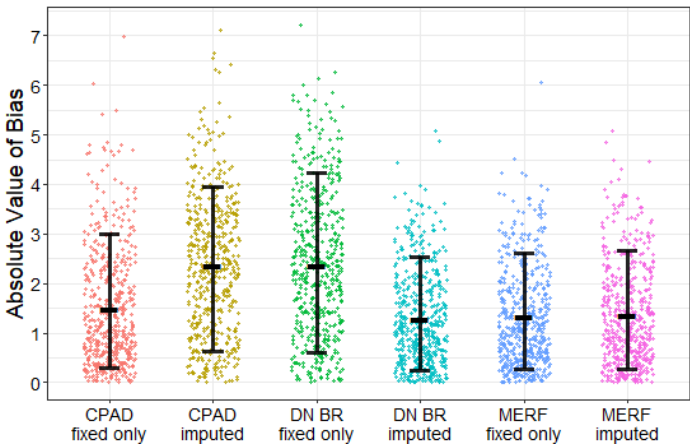
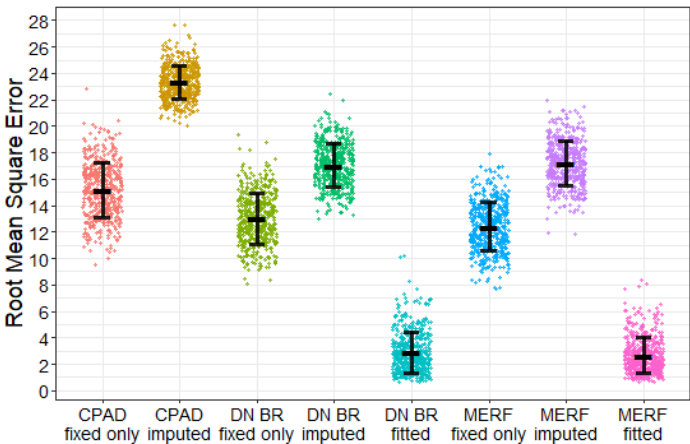
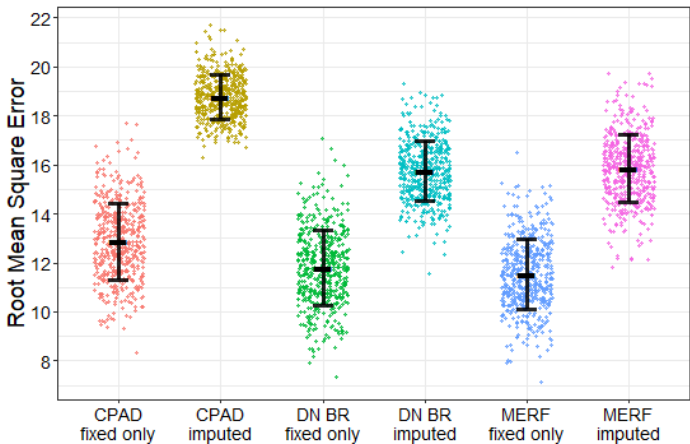
	CPAD FE Only	CPAD Imputed	De Novo BR FE Only	De Novo BR Imputed	De Novo BR Fitted	DN MERF FE Only	DN MERF Imputed	DN MERF Fitted
CPAD FE Only		-54.7% [-106.2% - -17.2%]	-11.2% [-51.9% - 19.7%]			13.6% [-18.8% - 55.0%]		
CPAD Imputed	54.1% [50.0% - 58.7%]			161.2% [109.5% - 247.7%]			186.9% [121.5% - 281.2%]	
De Novo BR FE Only	-16.2% [-20.5% - -11.9%]			87.8% [43.8% - 158.9%]	480.1% [357.3% - 671.1%]	26.3% [-1.7% - 69.6%]		
De Novo BR Imputed		-37.1% [-40.1% - -34.2%]	30.6% [26.4% - 35.1%]		208.9% [133.2% - 315.5%]		9.8% [-25.3% - 47.3%]	
De Novo BR Fitted			-350.6% [-399.7% - -303.9%]	-488.5% [-561.0% - -427.8%]				-14.0% [-46.3% - 15.8%]
DN MERF FE Only	-22.1% [-26.9% - -17.3%]		-5.1% [-9.4% - -0.6%]				63.3% [21.2% - 121.2%]	303.0% [212.6% - 418.8%]
DN MERF Imputed		-36.1% [-39.1% - -33.2%]		0.7% [-2.3% - 3.4%]		38.2% [33.6% - 42.8%]		146.8% [88.0% - 228.5%]
DN MERF Fitted					-9.3% [-27.3% - 5.5%]	-369.0% [-422.1% - -325.0%]	-548.2% [-622.5% - -488.7%]	

RMSE comparisons in lower half of matrix with AV bias comparisons in upper half





# Aim 3 - Synthetic Data Metrics





# Imputation Observations

- While the BR model and MERF models gave the expected patterns of increased error but reduced bias with random effect imputation, the pre-parameterized CPAD model had increases in both error and bias with RE imputation
- Importantly, the increases in bias and error were less pronounced in the synthetic dataset compared to the truncated meta-database dataset suggesting imputation can still generalize well
- This implies utility of imputation with the CPAD model when generating synthetic cohorts for "on average" calculations of the ADAS-Cog (i.e. its intended use) but suggests reliance on only the fixed effect parameterizations is preferred if attempting to predict an ADAS-Cog outcome
- However, there is no substitute of direct leveraging of the data as observed when using the known and fitted subject-specific effects during forecasting



# Final Thoughts

- Whether using the ensemble methods or sequential neural networks, prediction of ADAS-Cog and CDR derived impairment status were uniformly superior compared to the inferential reference models when applied to the meta-database further buttressing the role machine learning can play in AD research
- Importantly, this study specifically addressed ML/SL predictive capacity in longitudinal data; these models generally showed the greatest benefit when they were able to directly leverage prior data when forecasting observations
- The reduction in error and bias were greatly improved in ADAS-Cog prediction compared to the CPAD reference and every model displayed remarkable classification capacity at all time points when prior data could be utilized



# Final Thoughts

- One aspect this research did not address was feature selection or identification of the most powerful predictive covariates; this is an important step when considering the potential role of subject characteristics which have less prevalent coverage such as APOE4 allele status
- Additionally, many of the evaluation metrics for classification are reliant on translating a score value to a binary outcome which further confounds the observation of reduced recall in the ML/SL methods; however, the improvements in AUC, which is agnostic to strict classification decisions, offset this effect
- Trajectories for the sequential neural networks warrant further investigation as use of a standard FNN to seed baseline values prior to sequential generation may have impacted
- Hyperparameter selections, covariate coverage, and even benchmark performance are also all critical considerations when using ML/SL methods in Alzheimer's disease research



# Final Thoughts

- The most important caveat is there was no general consensus in the "best" model for either regression or classification in these designs
- These models ultimately all have their own strengths and weaknesses; for example some give more unbiased predictions with the ADAS-Cog (MERF) while others reduce error (Boosted trees) and the preferred model depends on the goals of the investigator
- This further extends into how subject-specific effects should be utilized with imputation having a role in developing robust synthetic cohorts for feasibility but generally detracting from prediction on a subject-by-subject basis
- Regardless, this research is a critical first step in characterizing predictive performance of longitudinal machine learning methods in Alzheimer's disease, serving as both a current evaluation of the field and a pipeline for future ML/SL paradigms



# Acknowledgements

*Talent wins games, but teamwork and intelligence win championships*

- Michael Jordan

## Committee

- Jeff Szychowski, PhD
- Richard Kennedy, MD, PhD
- Byron Jaeger, PhD
- Erik Roberson, MD, PhD
- Gary Cutter, PhD

Alzheimer's Disease Research Center

Center for Neurodegeneration and  
Experimental Therapeutics

The faculty, staff and students in the  
Department of Biostatistics







# Questions?

