

Biostatistics for Fluid and Imaging Biomarkers

Michael Donohue, PhD

University of Southern California

Biomarkers for Neurodegenerative Disorders

London, UK

May 2025

Keck School of
Medicine of **USC**

**Alzheimer's Therapeutic
Research Institute**



Course Overview

Topics:

- Hour 1 -- Biostatistics for Fluid and Imaging Biomarkers
- Hour 2 -- Modeling Longitudinal Data (Lars Racket)

Emphases:

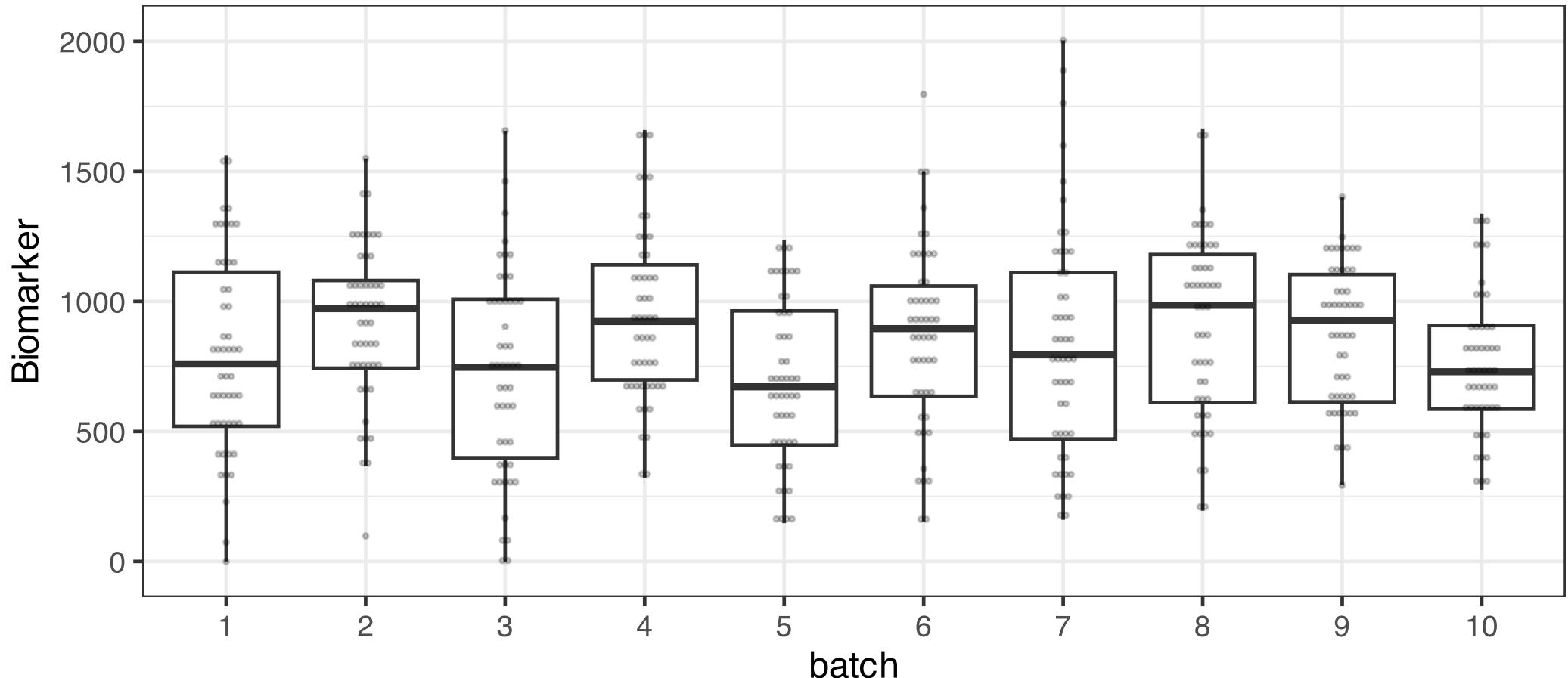
- Visualization
- Demonstrations using R, code available from:
 - <https://github.com/atri-biostats/biomarkers-neuro-disorders>

Session 1 Outline

- Batch Effects
- Experimental Design (Sample Randomization)
- Statistical Models for Assay Calibration/Quantification
- Classification (Supervised Learning)
 - Logistic Regression
 - Binary Trees
 - Random Forest
- Mixture Modeling (Unsupervised Learning)
 - Univariate
 - Bivariate
- Mixture of Experts (Unsupervised Learning with covariates)
- Reference Regions
- Harmonization using:
 - Centroids
 - Empirical Cumulative Distribution Function (ECDF)
 - ComBat

Batch Effects

Batch Effects: Boxplot



Coefficient of Variation

batch	N	Mean	SD	SD/Mean = CV (%)
1	50	790	379	48
2	50	925	299	32
3	50	725	389	54
4	50	951	332	35
5	50	690	312	45
6	50	867	349	40
7	50	837	446	53
8	50	914	348	38
9	50	883	271	31
10	50	763	266	35

- Coefficient of Variation (CV) = SD/Mean
- Often used for quality control (reject batch with $CV > x$)

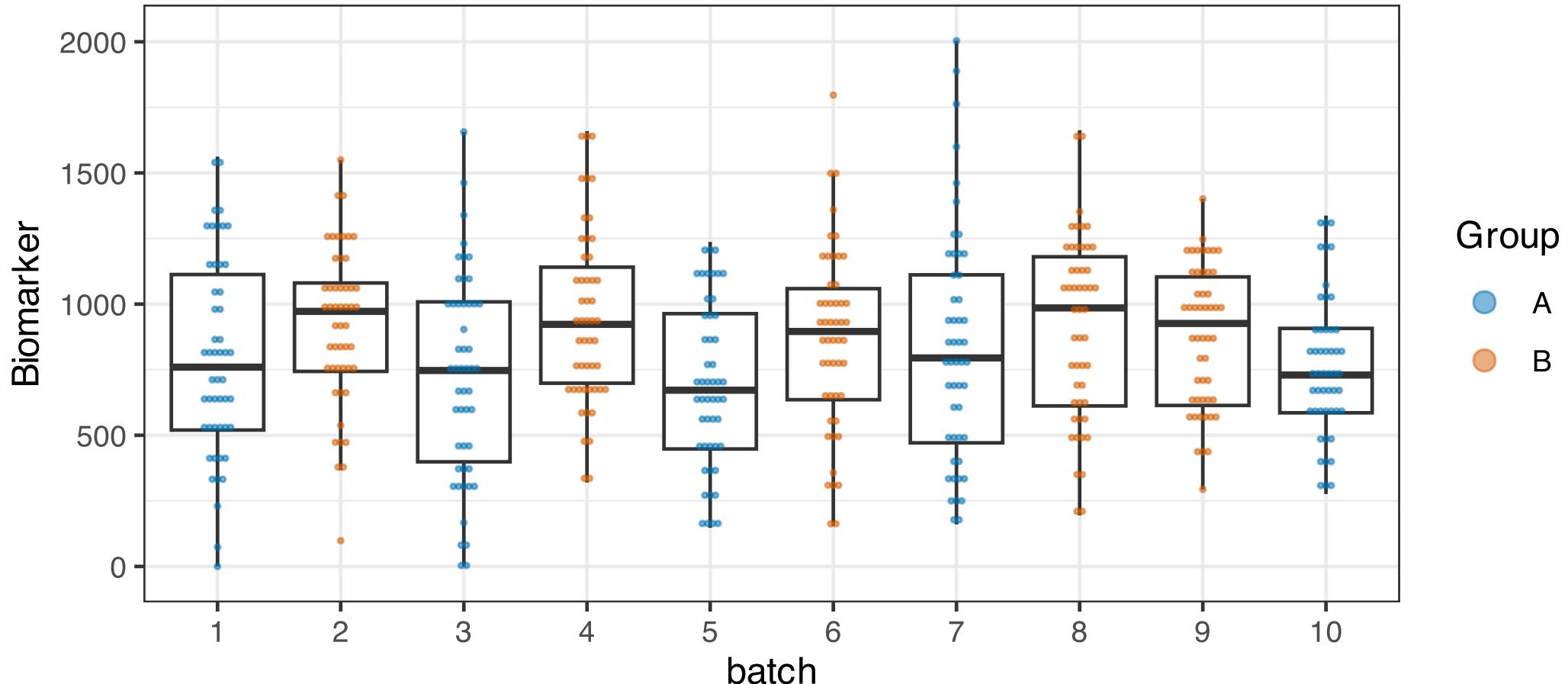
Testing for Batch Effects

```
anova(lm(Biomarker ~ batch, batch_data))
Analysis of Variance Table

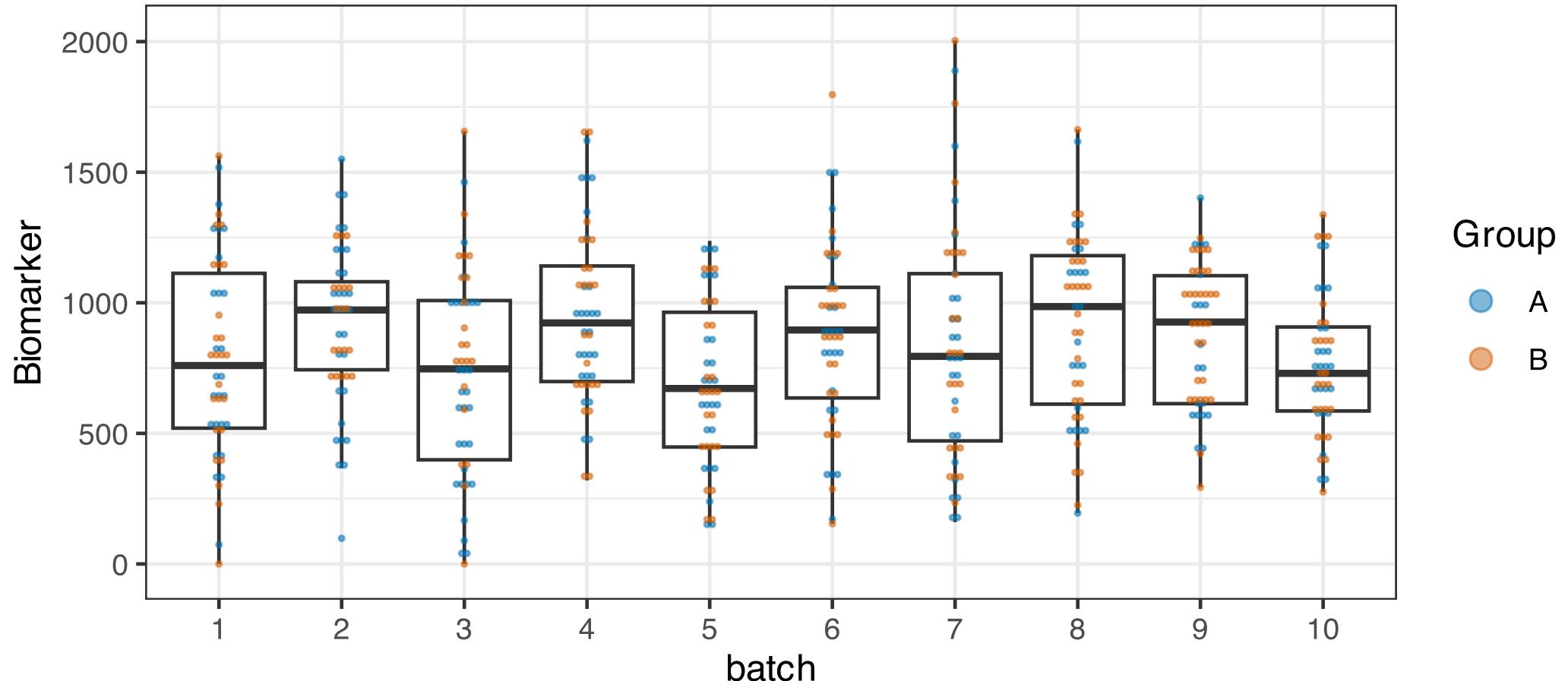
Response: Biomarker
            Df    Sum Sq Mean Sq F value Pr(>F)
batch         9  3573109  397012   3.37 0.00051 ***
Residuals 490  57758046 117874
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Batch explains a significant amount of the variation in this simulated data
- R note: batch variable must be a factor, not numeric (otherwise, you will get a batch slope)

Batch effects: Confounds



Randomized assignment of samples to plates



Experimental Design for Fluid Biomarkers

- Randomize samples to batches/plates
- Longitudinally collected samples (samples collected over time on same individual):
 - If batch effects are expected to be larger than storage effects, consider randomizing *individuals* to batches (i.e. keep all samples from individual on the same plate)
 - However, if storage effects are a concern, timely sample processing might be preferred.
- Randomization can be stratified to ensure important factors (e.g. treatment group, age, APOE $\epsilon 4$) are balanced over batches.

Sample Randomization

We use an R package [SRS](#) ("Subject Randomization System"), which we have modified to deal with the constraints of plate capacity, and keeping samples from the same subject together.

(Note this is different than the SRS package on CRAN)

Subject ID	Num. of samples	Group	Age	Plate
1	5	1	old	11
2	5	1	old	3
3	5	1	old	6
4	5	1	young	4
5	5	1	old	8
6	4	1	young	10
7	5	1	young	2
8	4	1	old	1
9	4	1	young	9
10	5	1	old	12

Sample Randomization

Plate	1	2	3	4	5	6	7	8	9	10	11	12	13
old	3	3	3	3	3	2	3	3	4	4	3	2	3
young	3	3	3	4	4	5	4	3	3	3	3	4	4
Num. samples	27	27	27	29	29	30	30	27	30	29	27	27	30

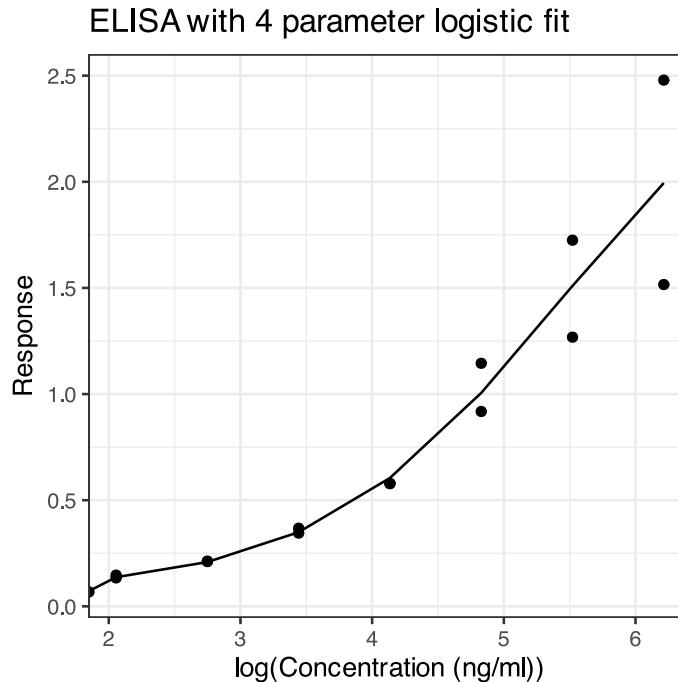
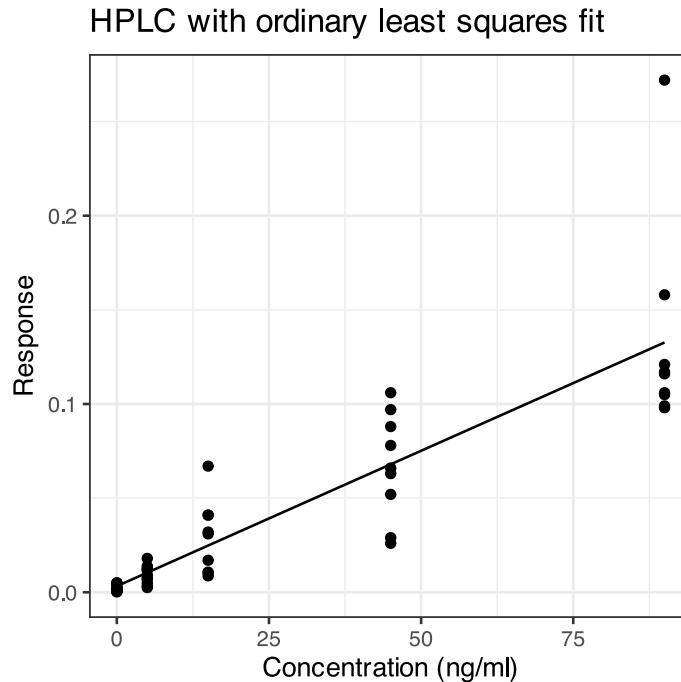
- Number of young and old well balanced across the 13 plates
- Number of samples per plate is also reasonable (plate capacity was set at 30 samples)

Calibration

Calibration

- Calibration: developing a map from "raw" assay responses to concentrations (ng/ml) using samples of *known* concentrations
- We will explore some approaches to calibration with methods from the R package `calibFit` (Haaland et al., 2011; Davidian et al., 1990)
- The package includes some example data:
 - High Performance Liquid Chromatography (HPLC) and
 - Enzyme Linked Immunosorbent Assay (ELISA)
- These examples are taken straight from the package vignette

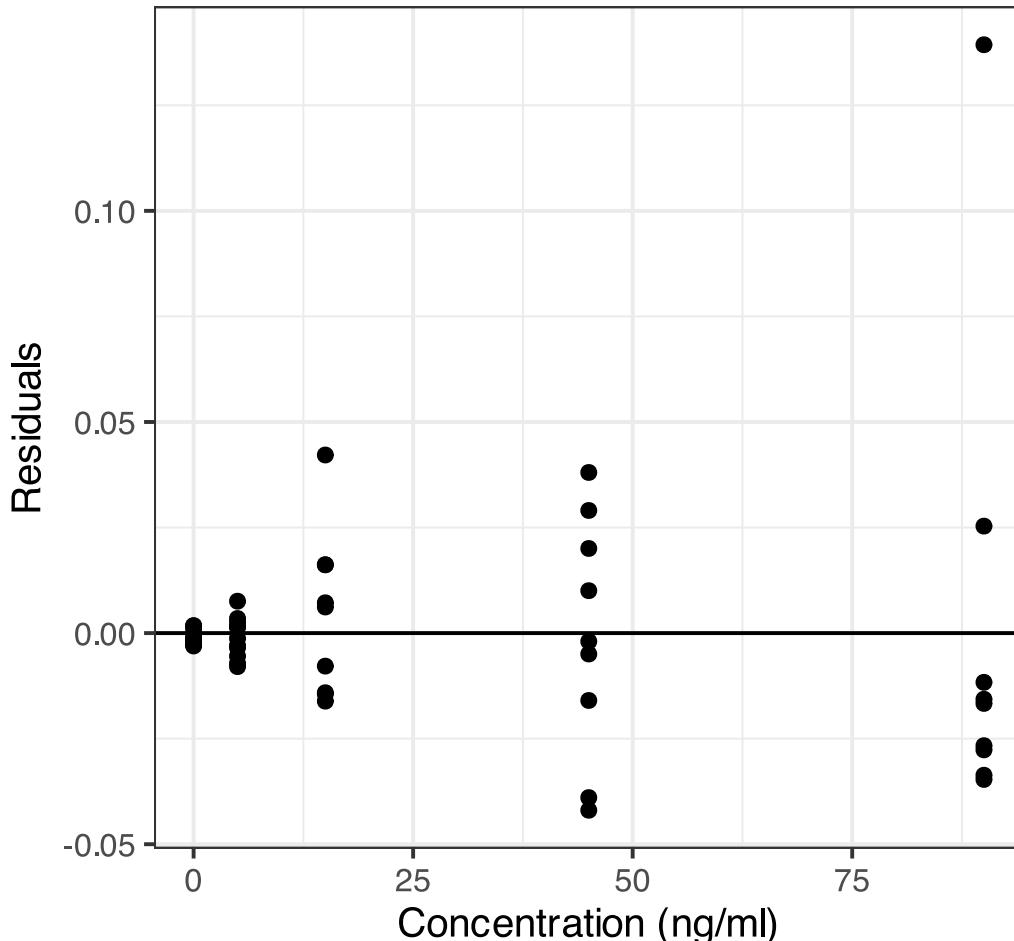
Calibration



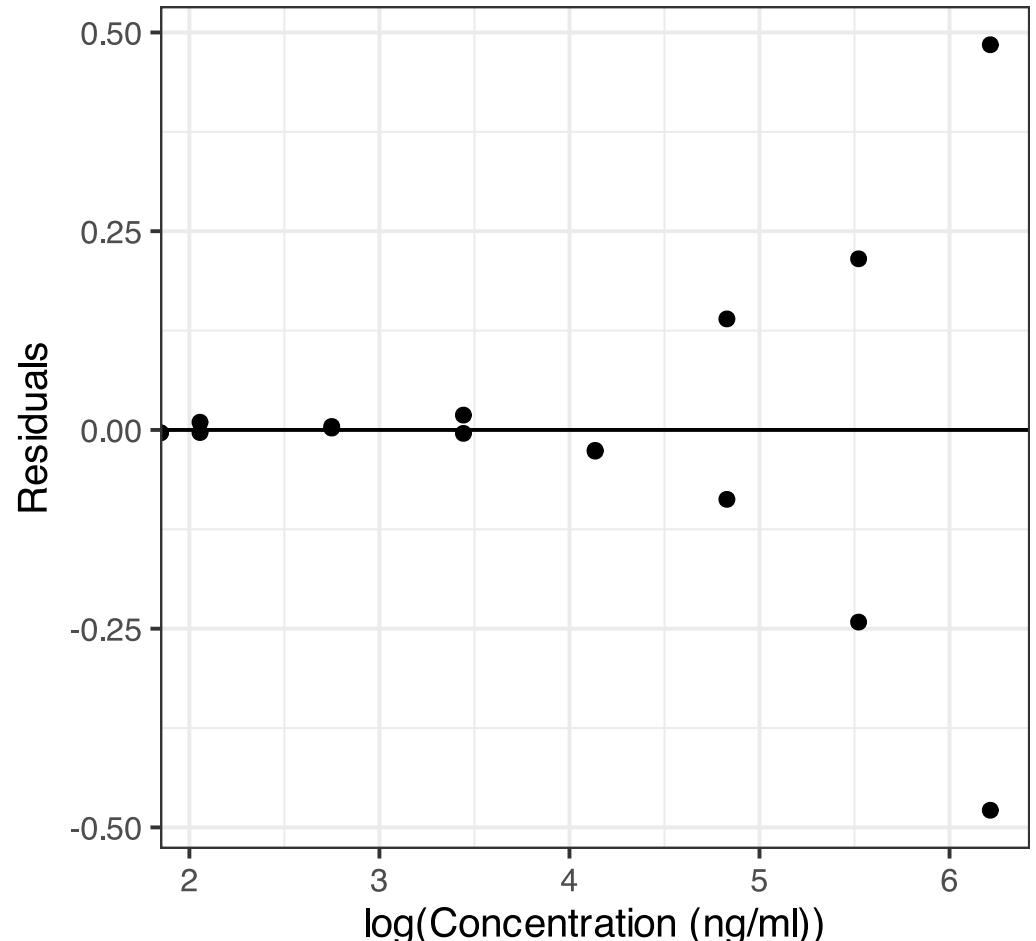
- Calibration curves are estimated using assay responses (vertical axis) from samples of *known* concentration (horizontal axis)
- Curves are subsequently used to map assay responses to estimated concentration values.
- Both fits exhibit *heteroscedasticity*: the error variance is not constant with respect to Concentration
- Most models assume *homoscedasticity*, or constant error variance.

Residuals (difference between response & fitted values)

HPLC with ordinary least squares fit



ELISA with 4 parameter logistic fit



Typical Regression

Typically, regression models are of the form:

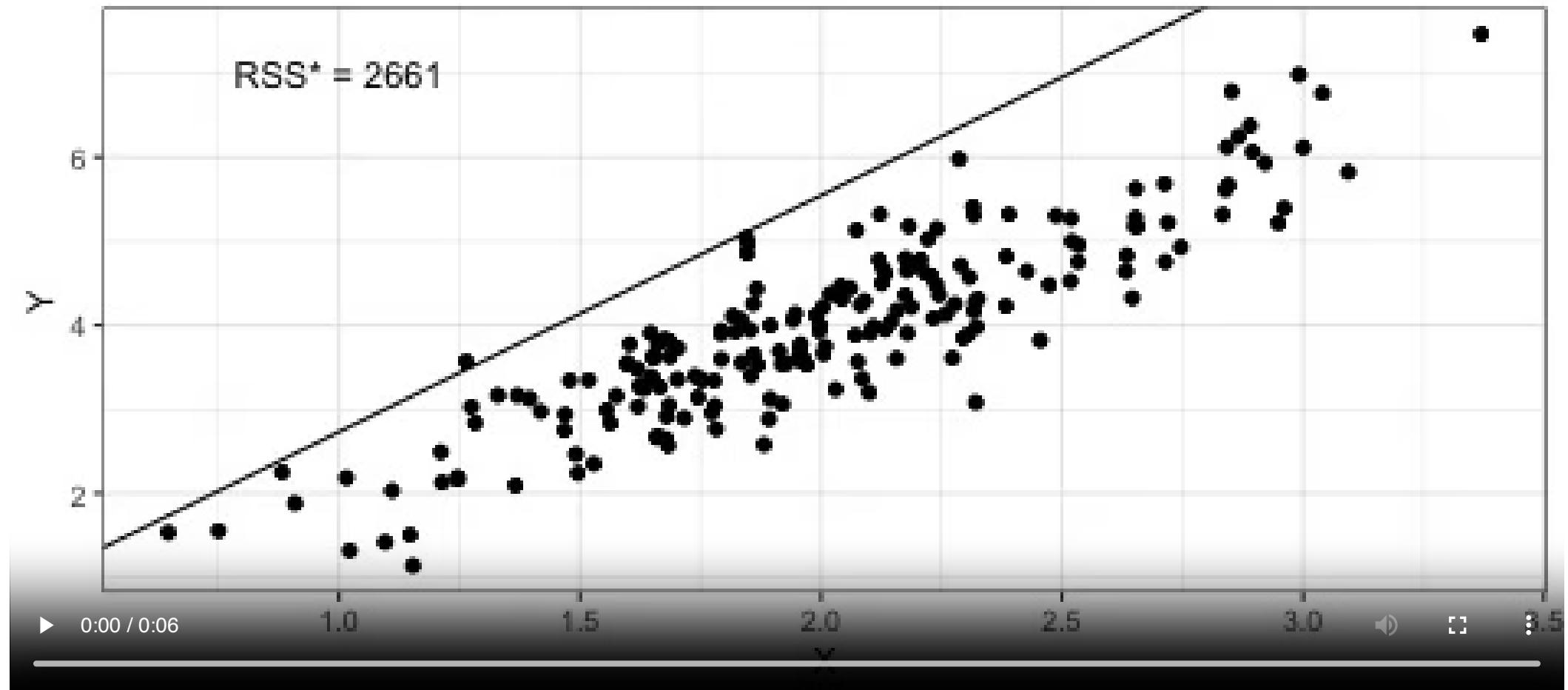
$$Y_i = f(x_i, \beta) + \epsilon_i,$$

where:

- Y_i is the observed response/outcome for i th individual ($i = 1, \dots, n$)
- x_i are covariates/predictors for i th individual
- β are regression coefficients to be estimated
- $f(\cdot, \cdot)$ is the model (assumed "known" or to be estimated)
 - In linear regression $f(x_i, \beta) = x_i\beta$
- ϵ_i is the residual error
- We assume $\epsilon \sim \mathcal{N}(0, \sigma^2)$
- σ is the *constant* standard deviation (*homoscedastic*)

If the standard deviation is not actually constant (*heteroscedastic*), estimates might be unreliable.

Ordinary Least Squares: minimizing the sum of squared residuals



* RSS = Residual sum of squares, or $\sum_i (\text{Observed}_i - \text{Fitted}_i)^2$

Modeling Heteroscedastic Errors

The `calibFit` package includes models of the form:

$$Y_{ij} = f(x_i, \beta) + \sigma g(\mu_i, z_i, \theta) \epsilon_{ij},$$

where,

- Y_{ij} are observed assay values/responses for i th individual ($i = 1, \dots, n$), j th replicate
- $g(\mu_i, z_i, \theta)$ is a function that allows the variances to depend on:
 - μ_i (the mean response $f(x_i, \beta)$),
 - covariates z_i , and
 - a parameter ("known" or unknown) θ .
- $\epsilon_{ij} \sim \mathcal{N}(0, 1)$

In particular, `calibFit` implements the Power of the Mean (POM) function

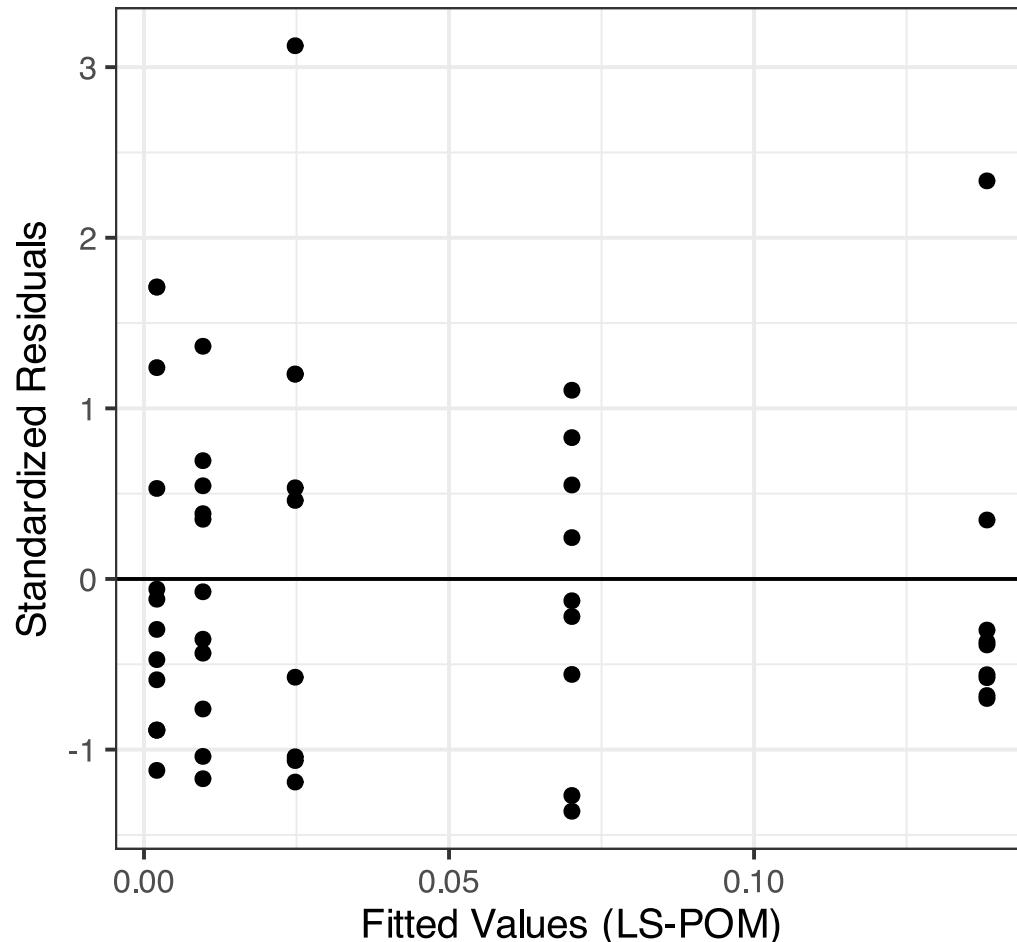
$$g(\mu_i, \theta) = \mu_i^{2\theta}$$

which results in

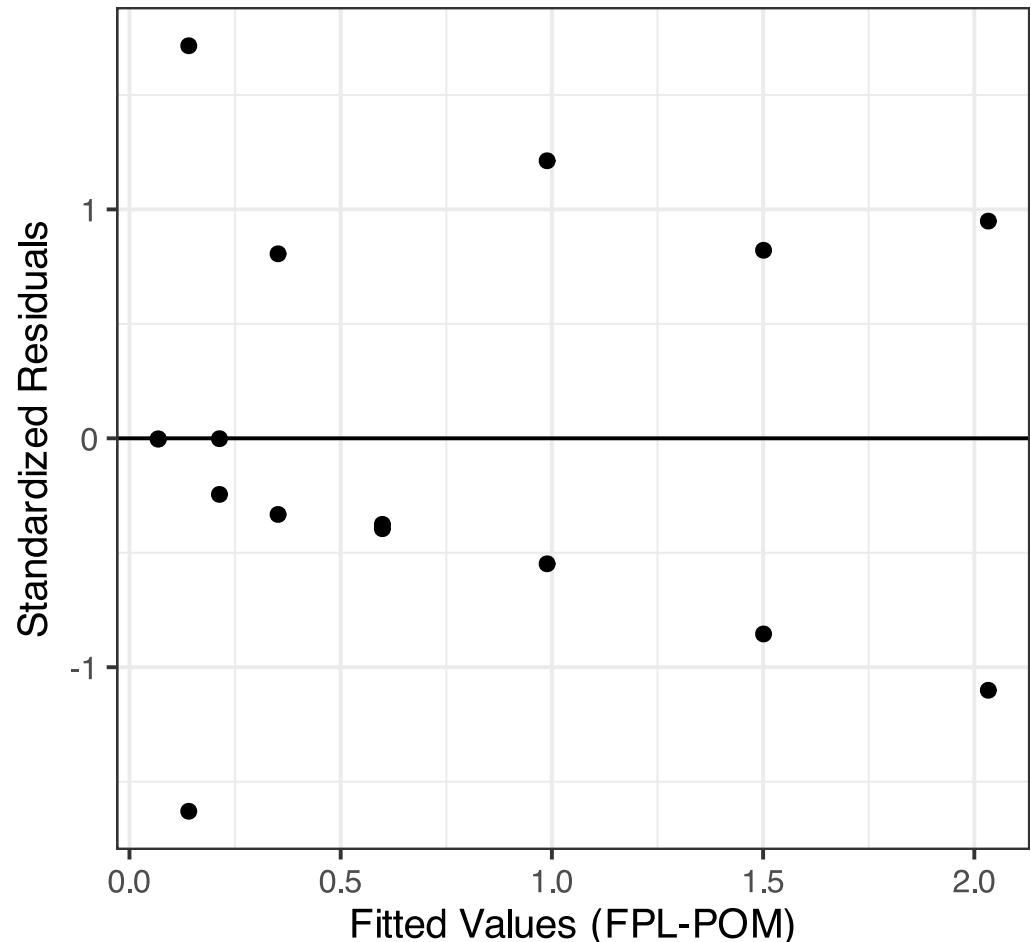
$$\text{var}(Y_{ij}) = \sigma^2 \mu_i^{2\theta}$$

"Homogenized" Residuals From Fits with POM

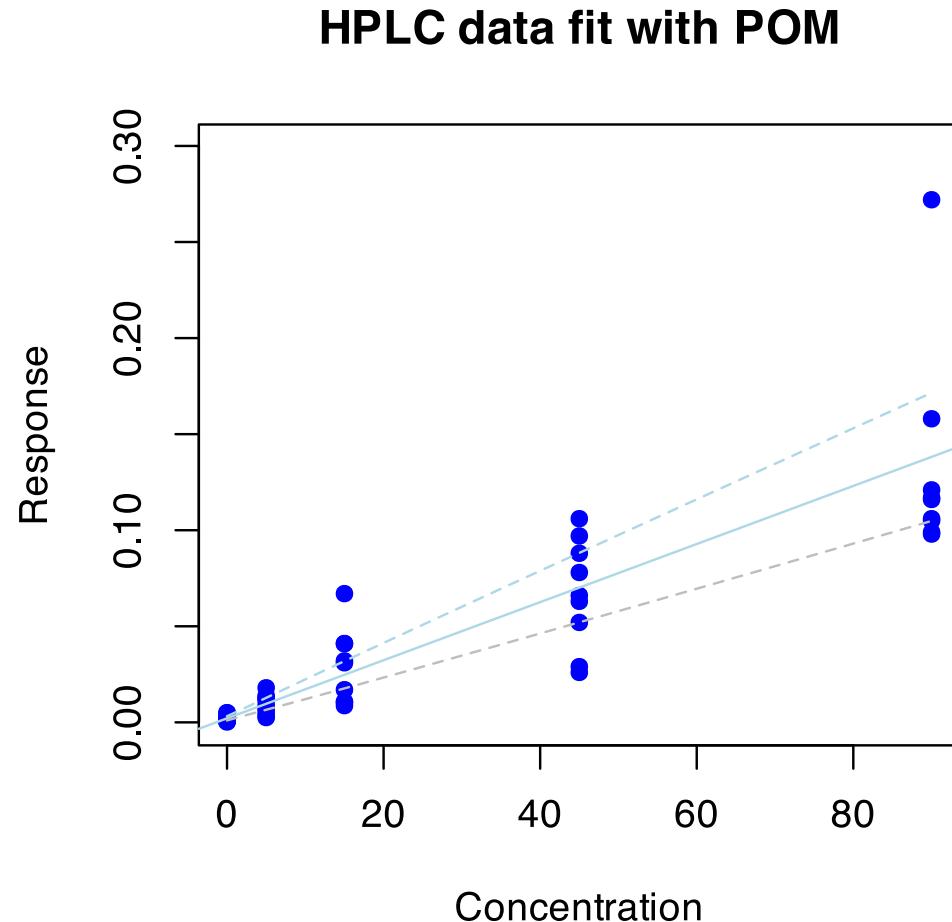
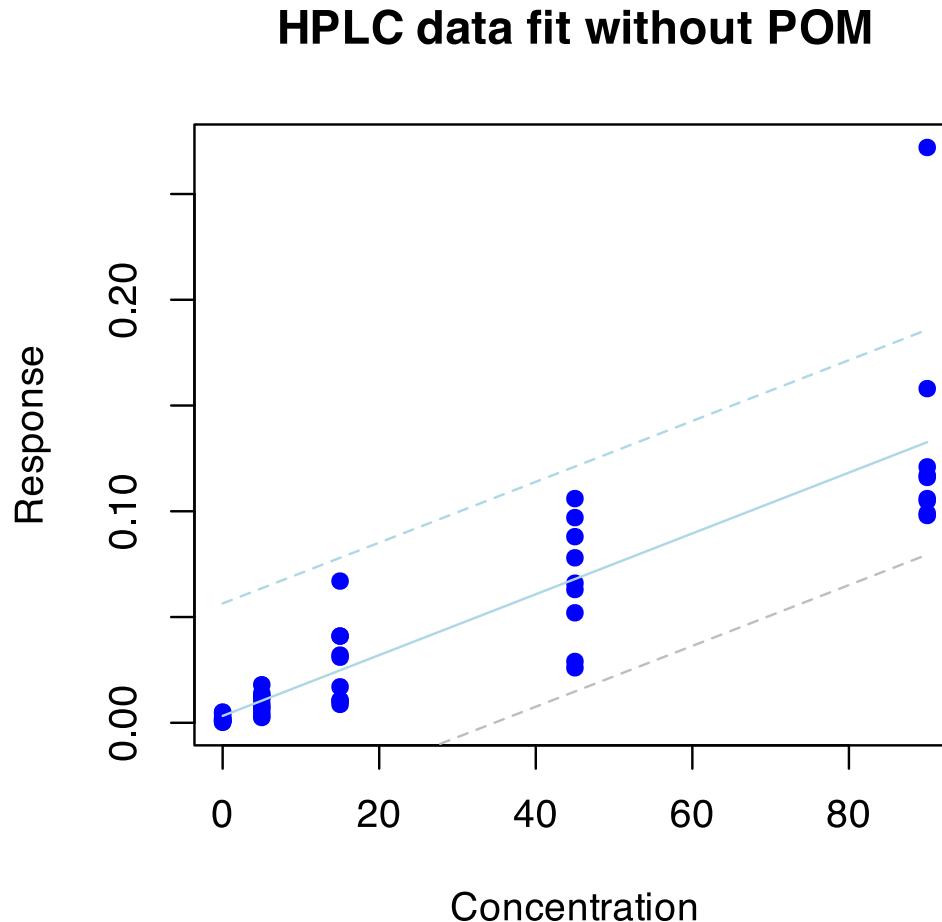
HPLC with least squares POM



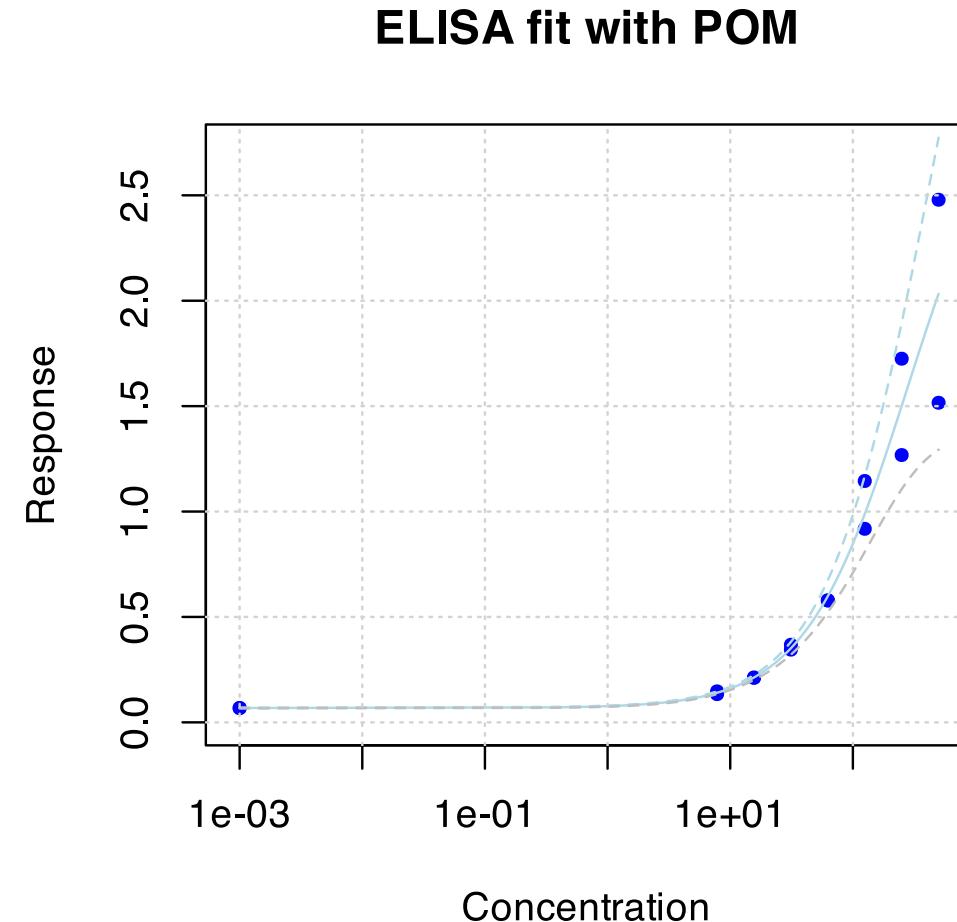
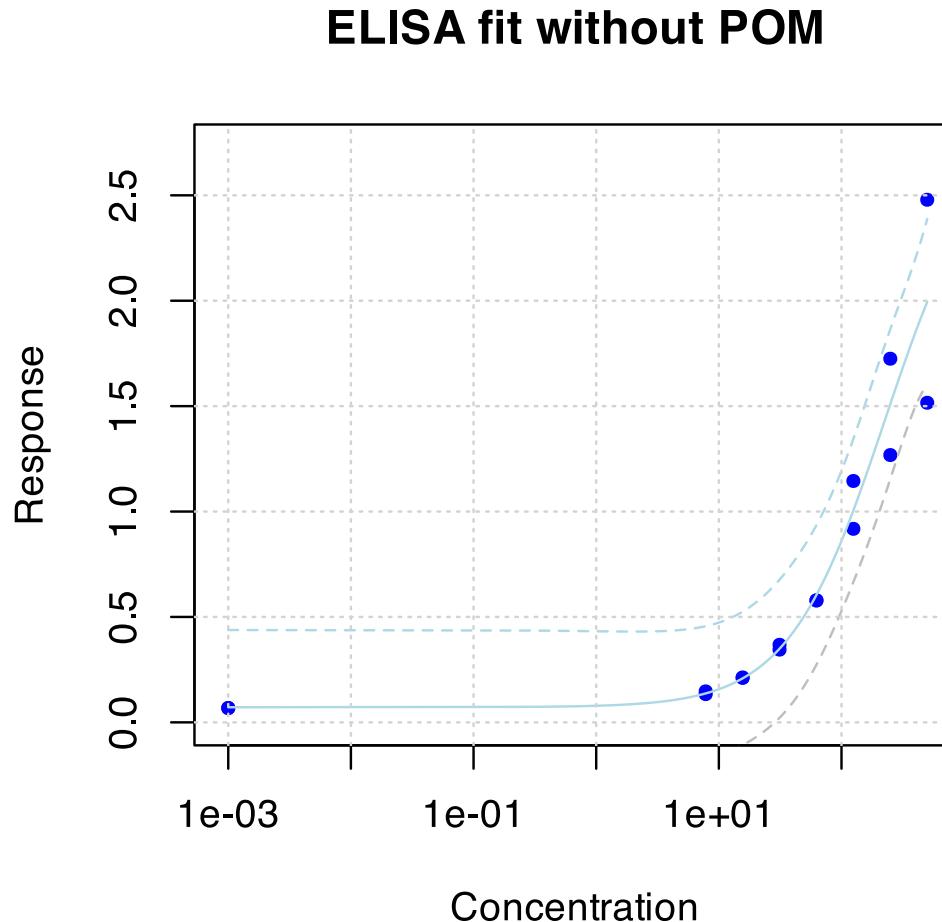
ELISA with 4 parameter logistic POM



HPLC Calibration With/Without POM Variance

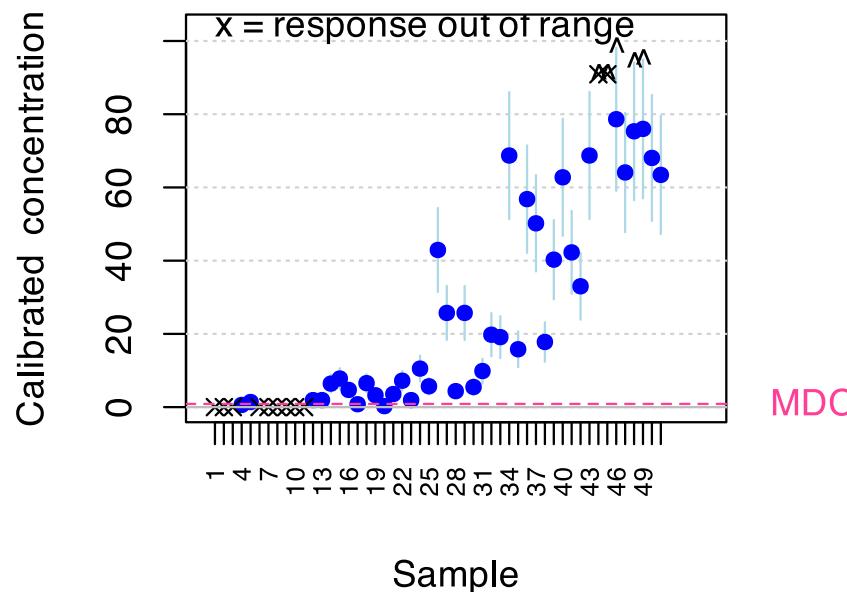


Elisa Calibration With/Without POM Variance

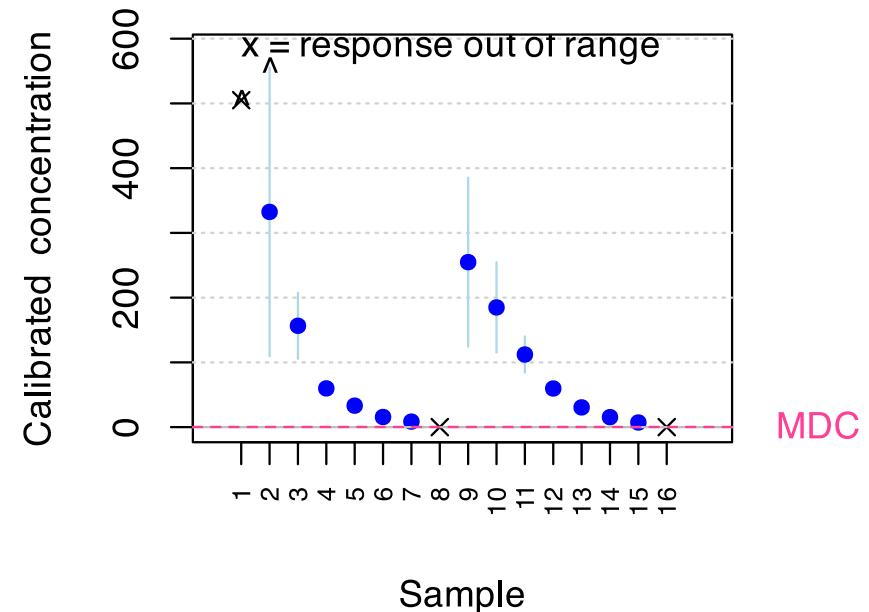


Calibrated Estimates & Minimum Detectable Concentration (MDC) for Each Sample

HPLC calibrated with linear POM



ELISA calibrated with FPL POM



Calibration Statistics

Assuming calibration curve f , mapping concentrations to assay responses, is increasing, we define the following terms.

Minimum Detectable Concentration (MDC): The lowest concentration where the curve is increasing, or:

$$x_{\text{MDC}} = \min\{x : f(x, \beta) > \text{UCL}_0\}$$

where UCL_0 is the upper confidence limit at 0

Reliable Detection Limit (RDL): The lowest concentration that has a high probability of producing a response that is significantly greater than the response at 0, or

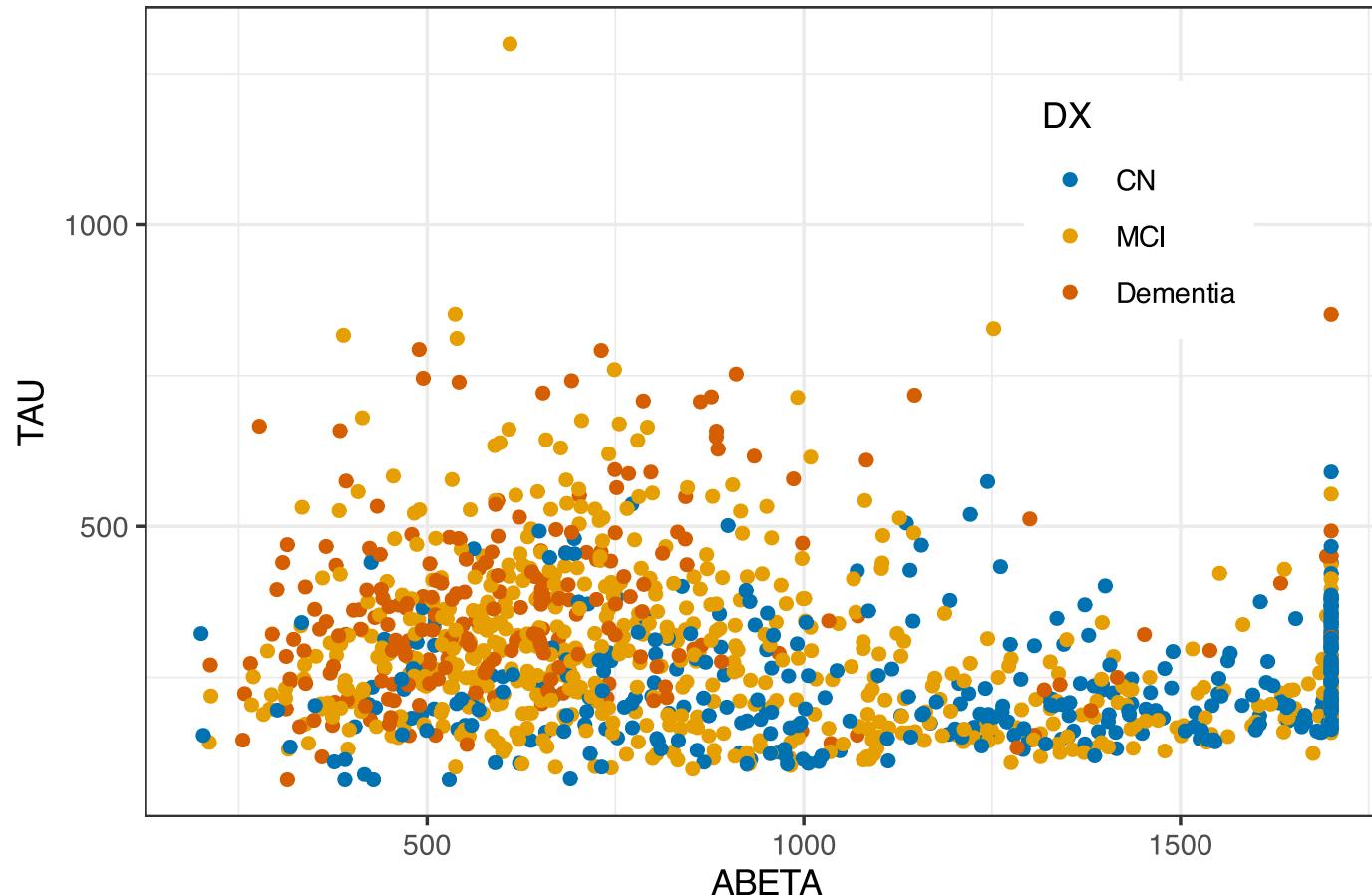
$$x_{\text{RDL}} = \min\{x : \text{LCL}_x > \text{UCL}_0\}$$

Limit of Quantitization (LOQ): The lowest concentration at which the coefficient of variation is less than a fixed percent (default is 20% in the `calibFit` package).

Supervised Learning

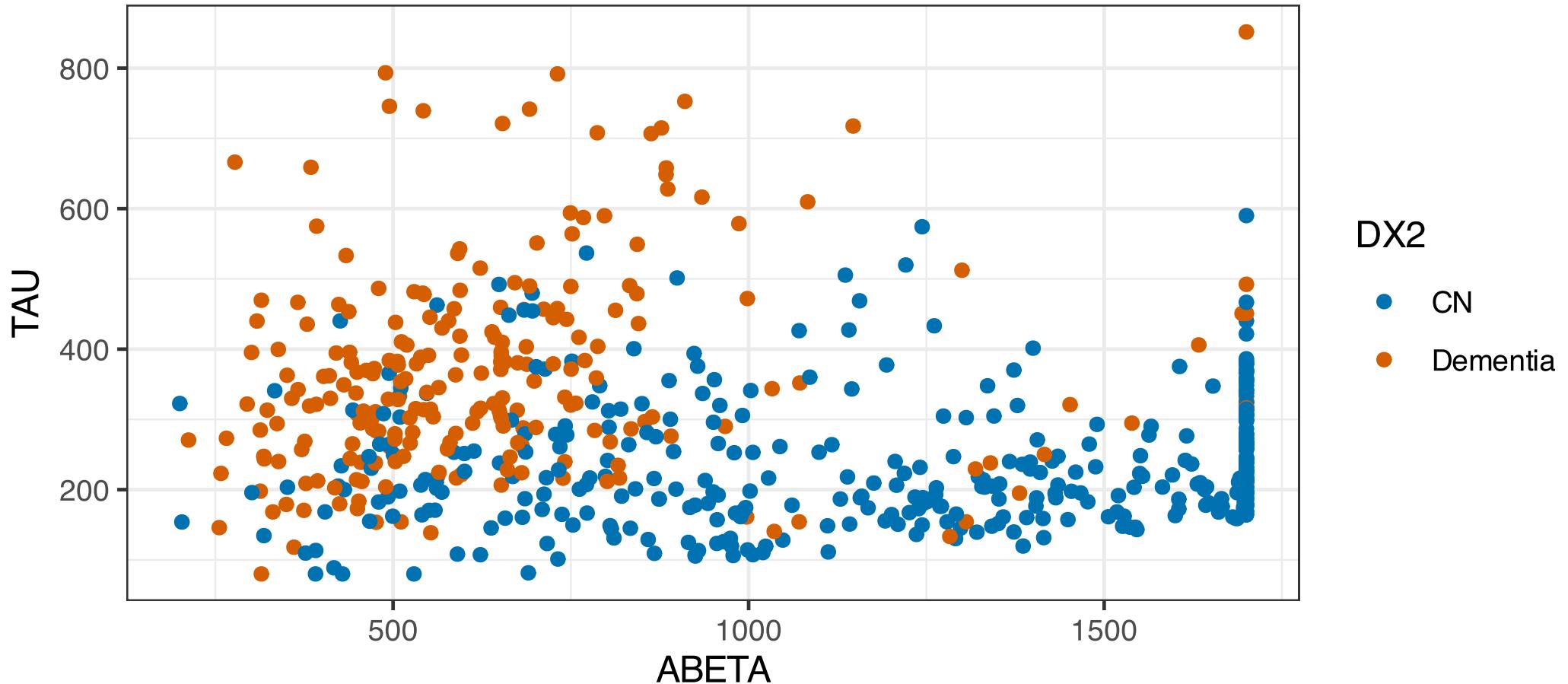
Classification

Classification

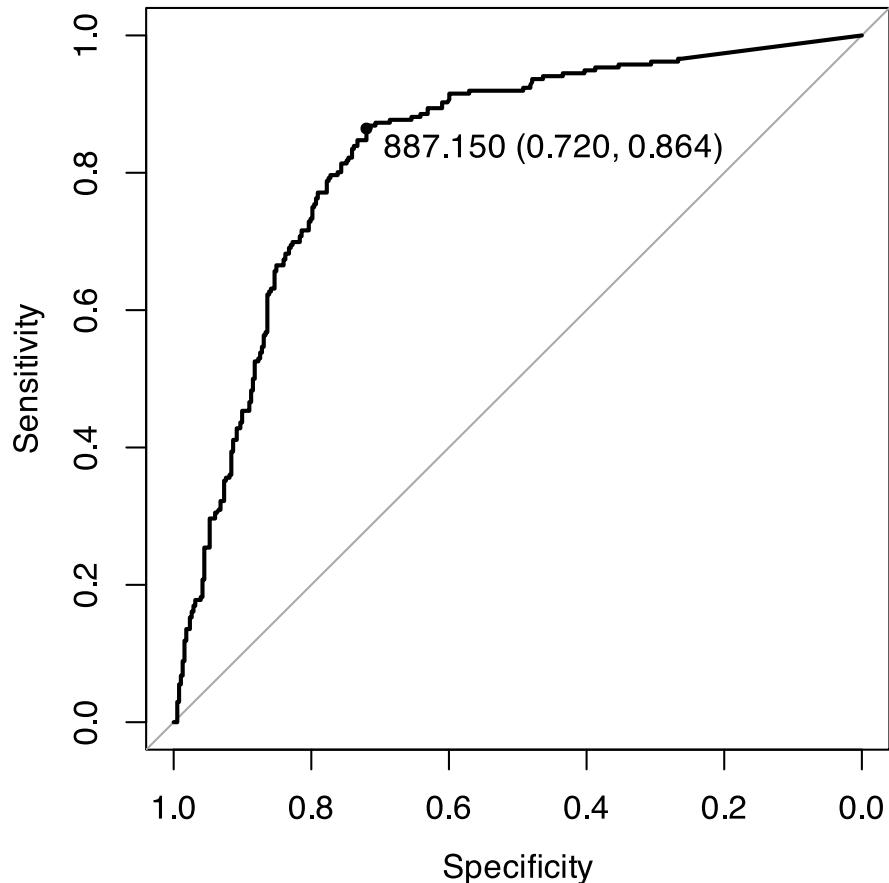


- Data from adni.loni.usc.edu
- CSF Abeta 1-42 and t-tau assayed using the automated Roche Elecsys and cobas e 601 immunoassay analyzer system
- Filter time points associated with first assay, and ignore subsequent time points
- We'll ignore MCI and focus on CN vs Dementia
- Values greater than the upper limit of detection have been assigned the limit

Classification



Reciever Operating Characteristic (ROC) Curves



For each potential threshold applied to CSF A β 42, we calculate:

- Sensitivity: True Positive Rate = TP/(TP+FN)
- Specificity: True Negative Rate = TN/(TN+FP)

This traces out the ROC curve.

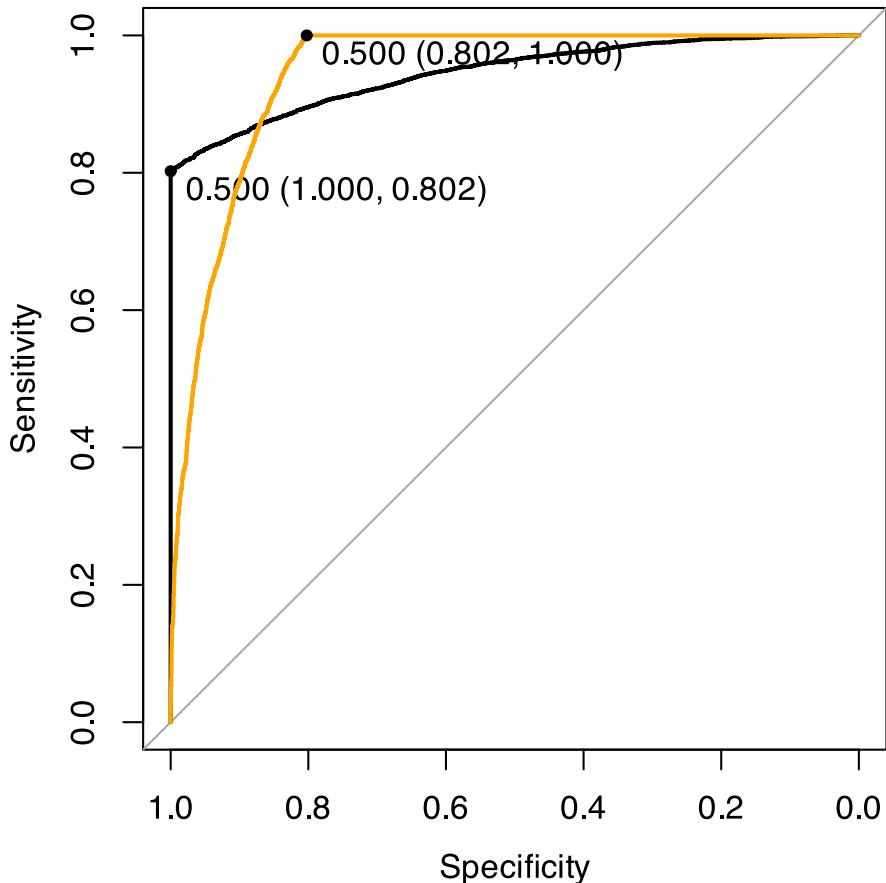
A typical summary of a classifier's performance is the Area Under the Curve (AUC)

AUC=0.83 in this case, with 95% CI (0.8, 0.86)

AUCs close to one indicate good performance.

The threshold shown here maximizes the distance between the curve and the diagonal line (chance) (Youden, 1950)

Same AUCs do not imply same performance



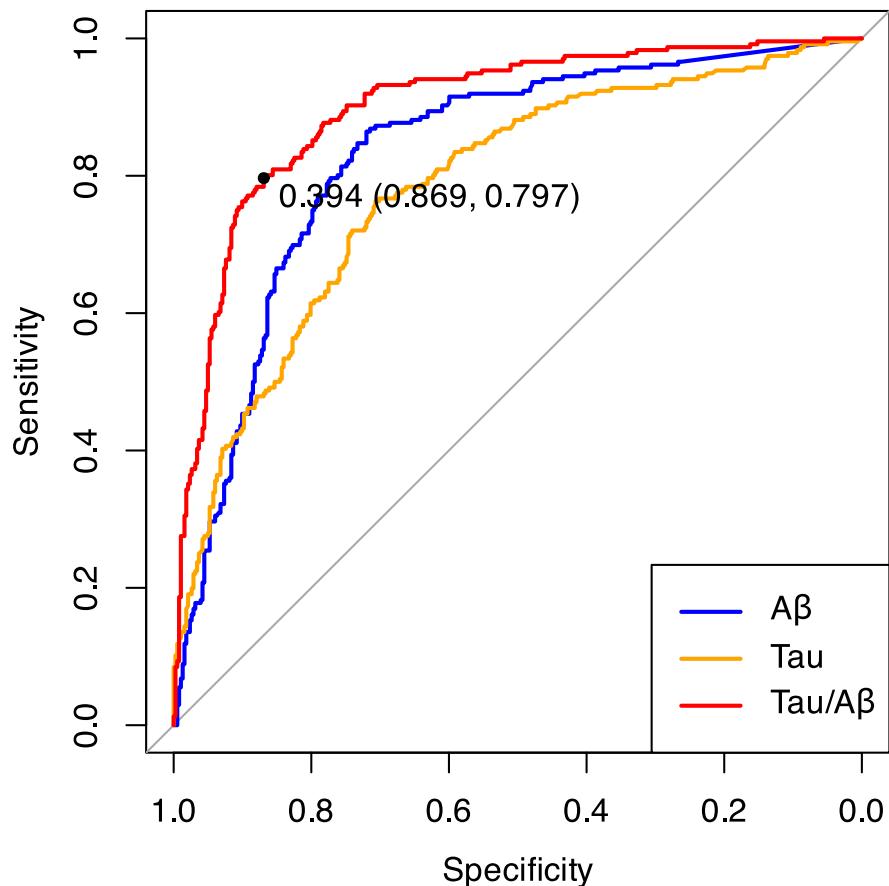
AUC is a somewhat crude metric of classification accuracy.

$AUC_1 = AUC_2 = 0.95$ in this case, with 95% CI (0.94, 0.95)

Both show good overall performance with AUC close to one.

However, one attains 80% sensitivity and 100% specificity, while the other attains the opposite (100% sensitivity and 80% specificity).

Comparing ROC Curves

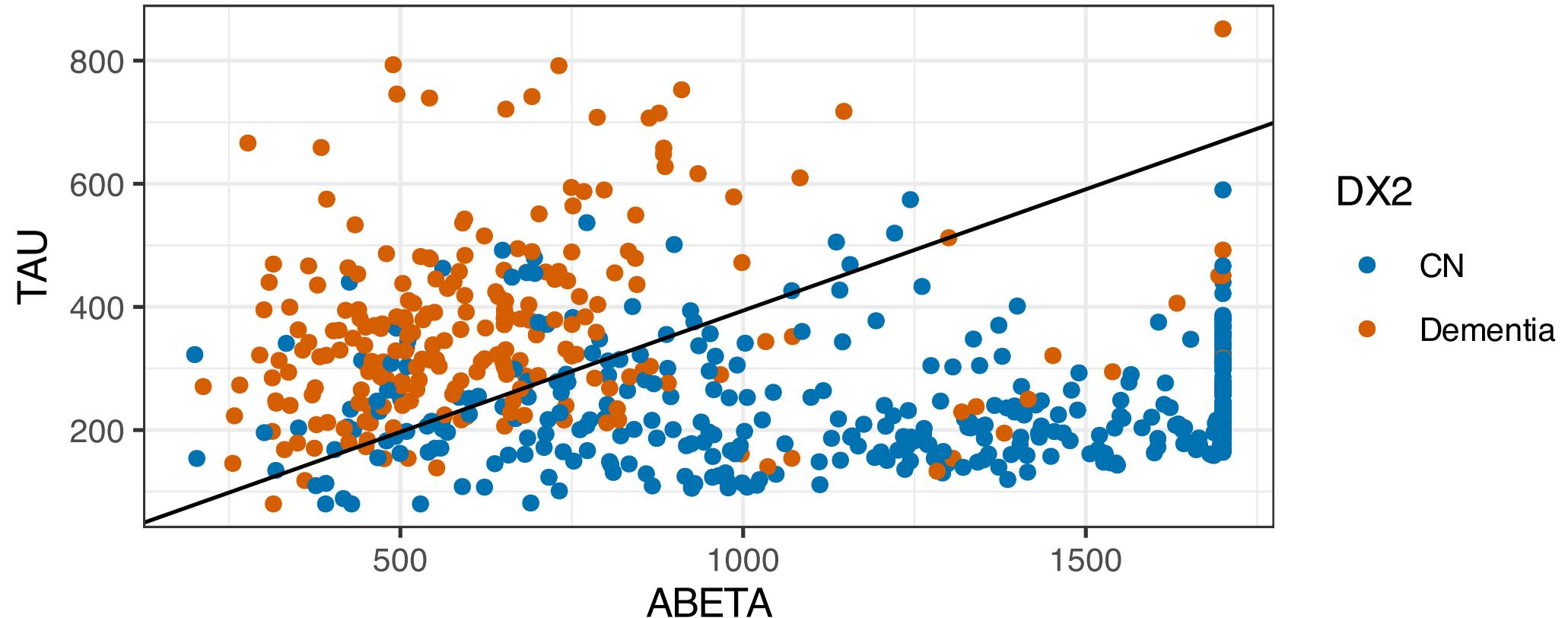


Marker	AUC	95% CI	P-value *
A β	0.83	0.8, 0.86	
Tau	0.78	0.75, 0.82	0.07
Tau/ A β	0.9	0.87, 0.92	<0.001

* Bootstrap test comparing each row to A β (Robin et al., 2011)

So the ratio of Tau / A β shows the best discrimination of NC from Dementia cases.

Youden's Cutoff for Tau / A β Ratio



Line is $Tau = 0.394 \times Abeta$, depicting Youden's cutoff (maximizes sensitivity + specificity)

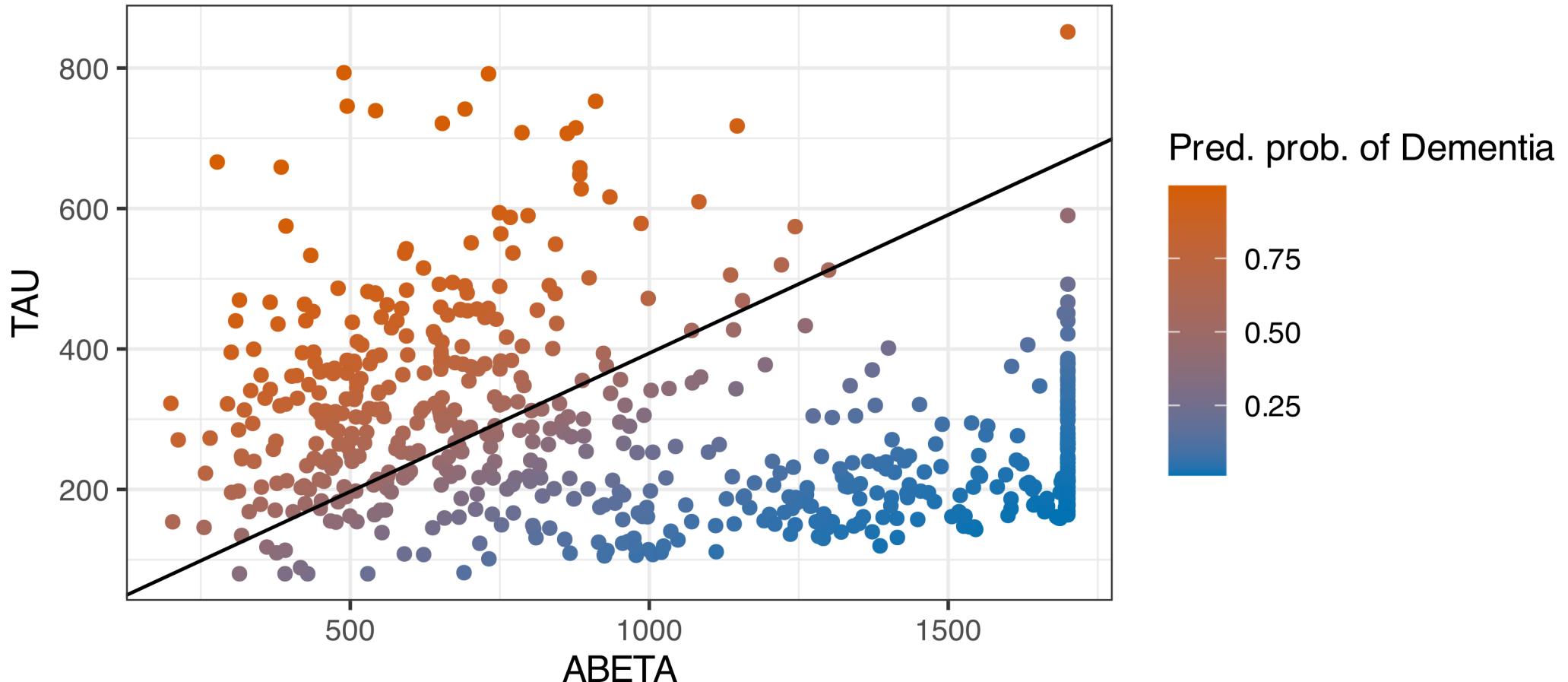
Logistic Regression

Coefficient	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.89	0.13	-6.7	<0.001
scale(ABETA)	-1.59	0.15	-10.6	<0.001
scale(TAU)	1.26	0.14	9.0	<0.001

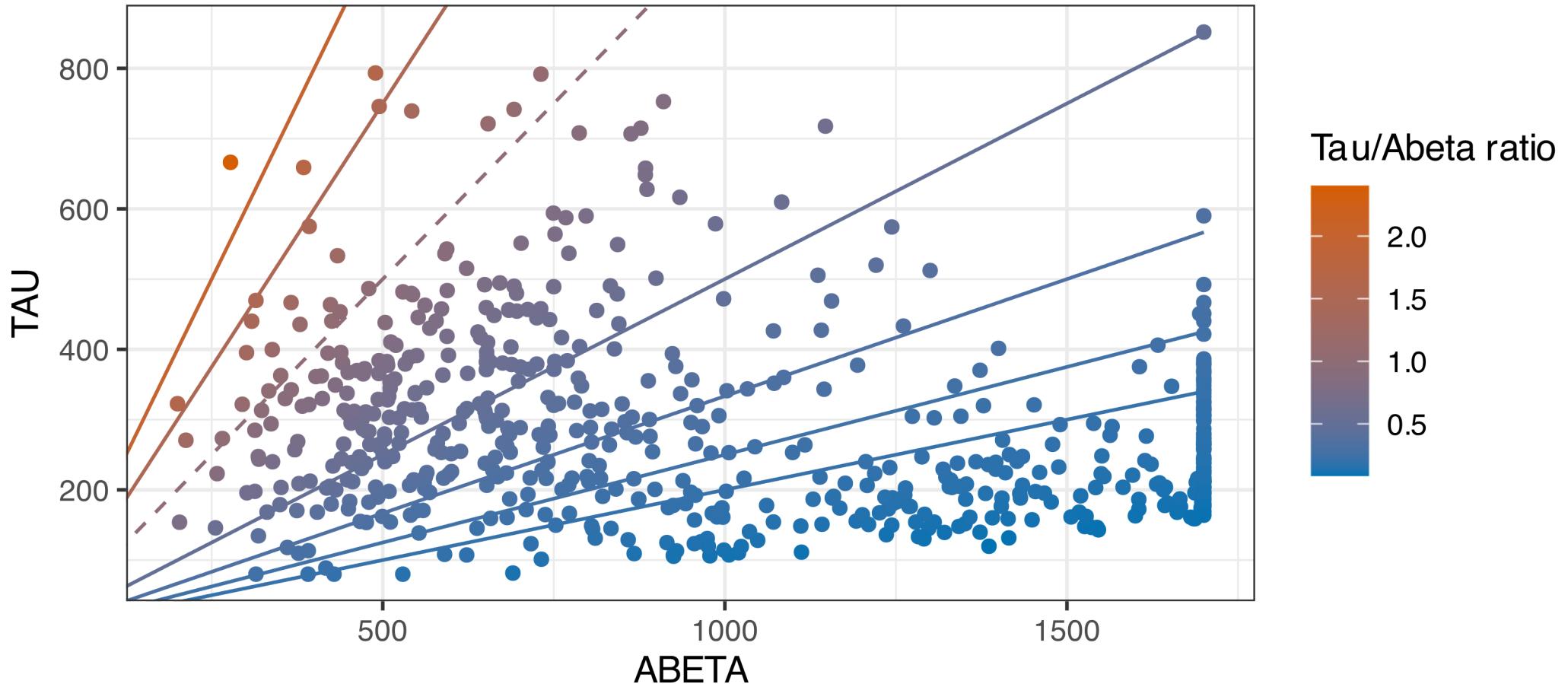
$$\log \left(\frac{p}{1 - p} \right) = \hat{\gamma}_0 + A\beta_z \hat{\gamma}_{A\beta} + \text{tau}_z \hat{\gamma}_{\text{tau}}$$

where $\hat{\gamma}$ are regression coefficients.

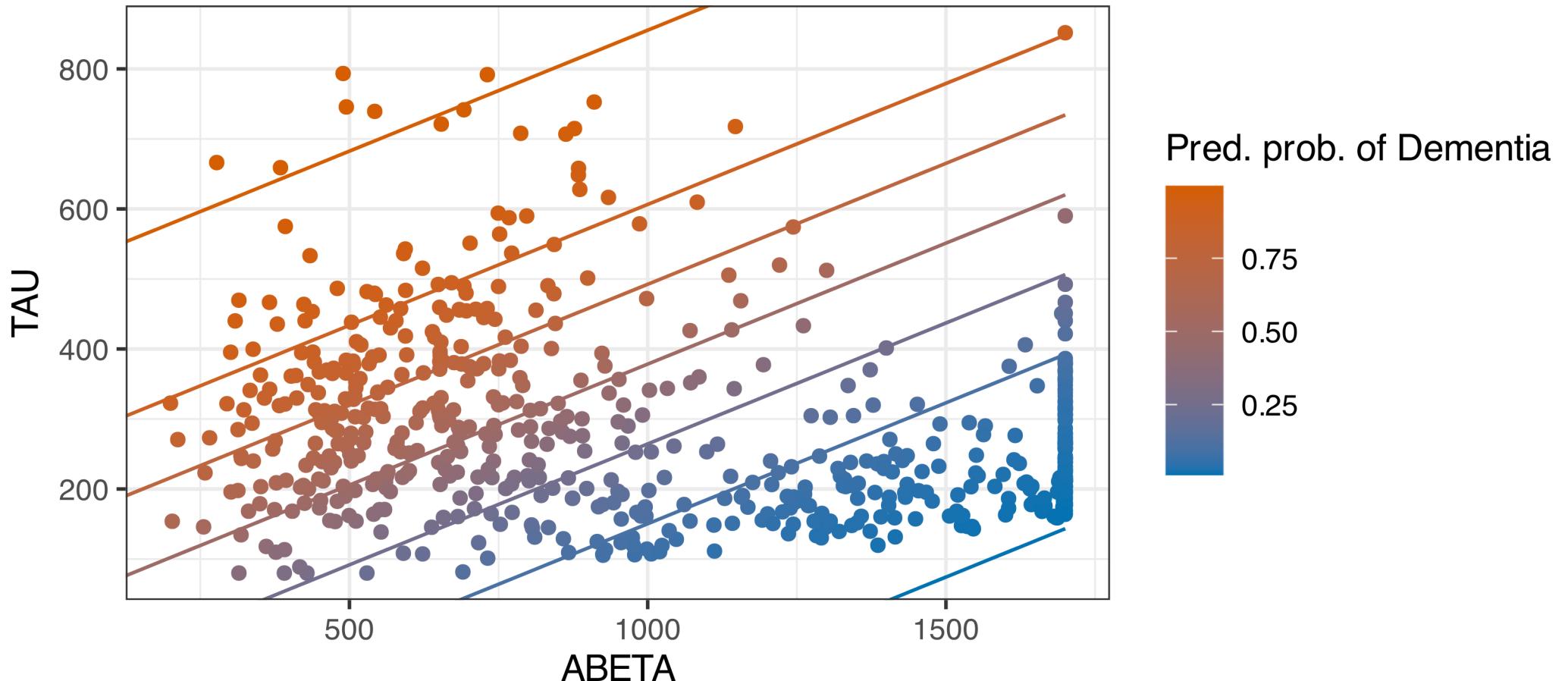
Logistic Regression Predicted Probabilities



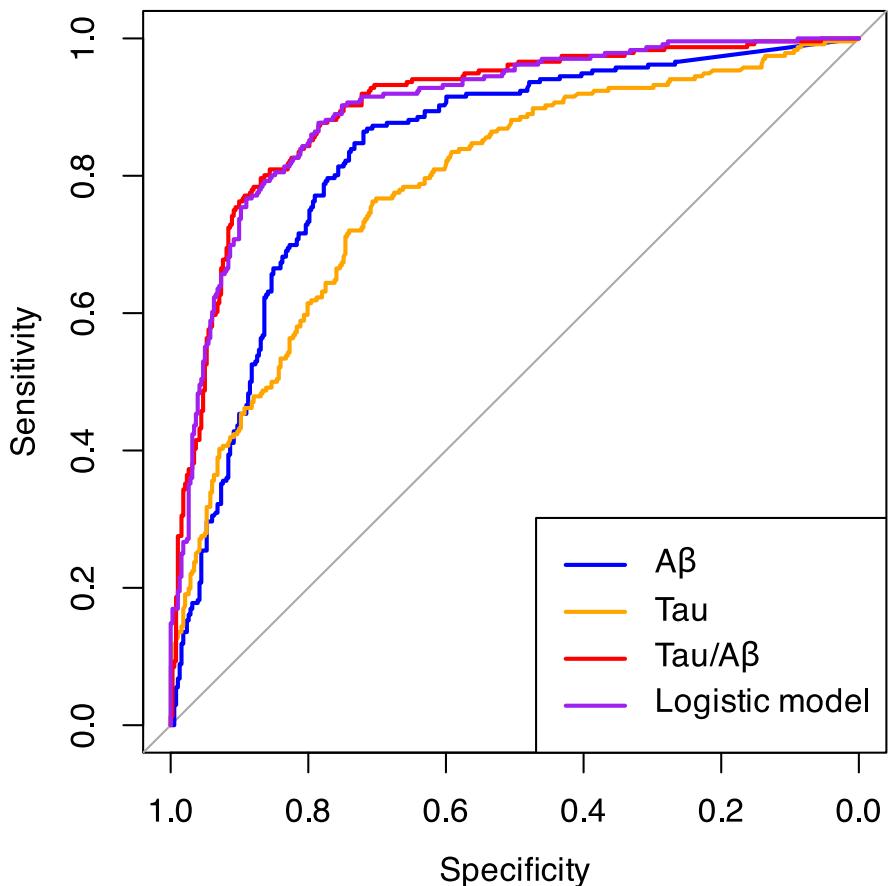
Ratio Contours



Logistic Regression Predicted Probability Contours



Comparing ROC Curves



Marker	AUC	95% CI	P-value *
$A\beta$	0.83	0.8, 0.86	
Tau	0.78	0.75, 0.82	0.07
Tau/ $A\beta$	0.9	0.87, 0.92	<0.001
Logistic model	0.9	0.87, 0.92	<0.001

* Bootstrap test comparing each row to $A\beta$ (Robin et al., 2011)

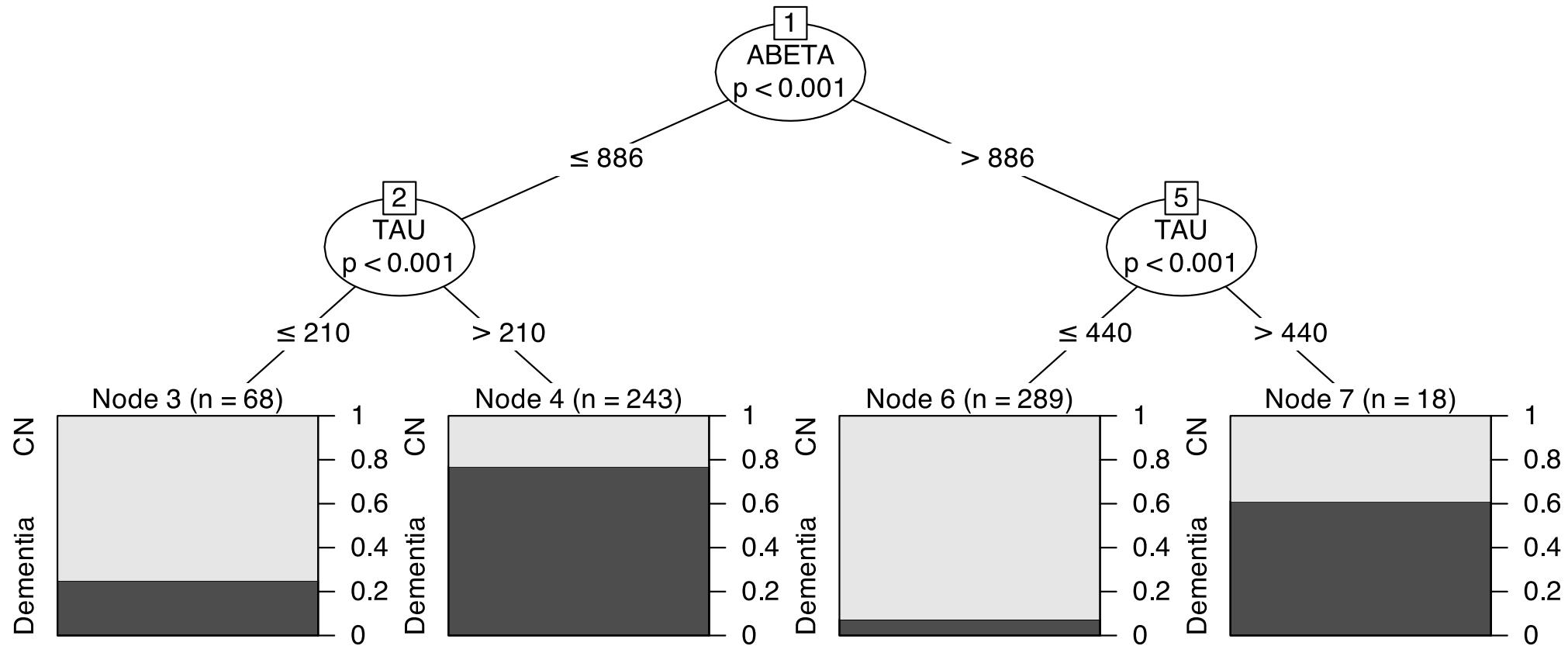
Logistic model ROC is very similar to Tau/ $A\beta$ ratio ROC.

Logistic Regression with Age and APOE

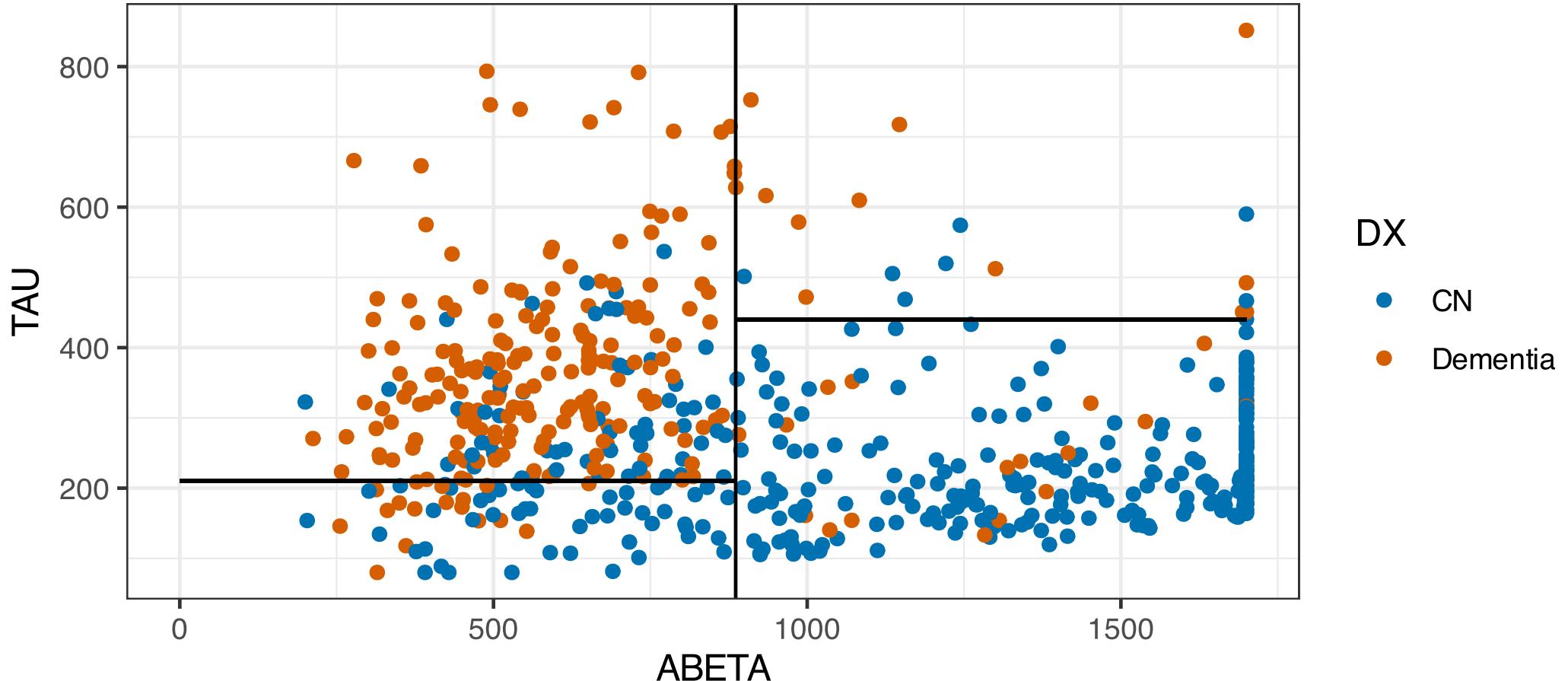
Coefficient	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.12	0.17	-6.5	<0.001
scale(ABETA)	-1.43	0.16	-9.0	<0.001
scale(TAU)	1.19	0.14	8.5	<0.001
scale(I(AGE + Years.bl))	0.14	0.12	1.2	0.230
as.factor(APOE4)1	0.37	0.25	1.5	0.144
as.factor(APOE4)2	1.26	0.45	2.8	0.005

This model does not provide much better ROC, either.

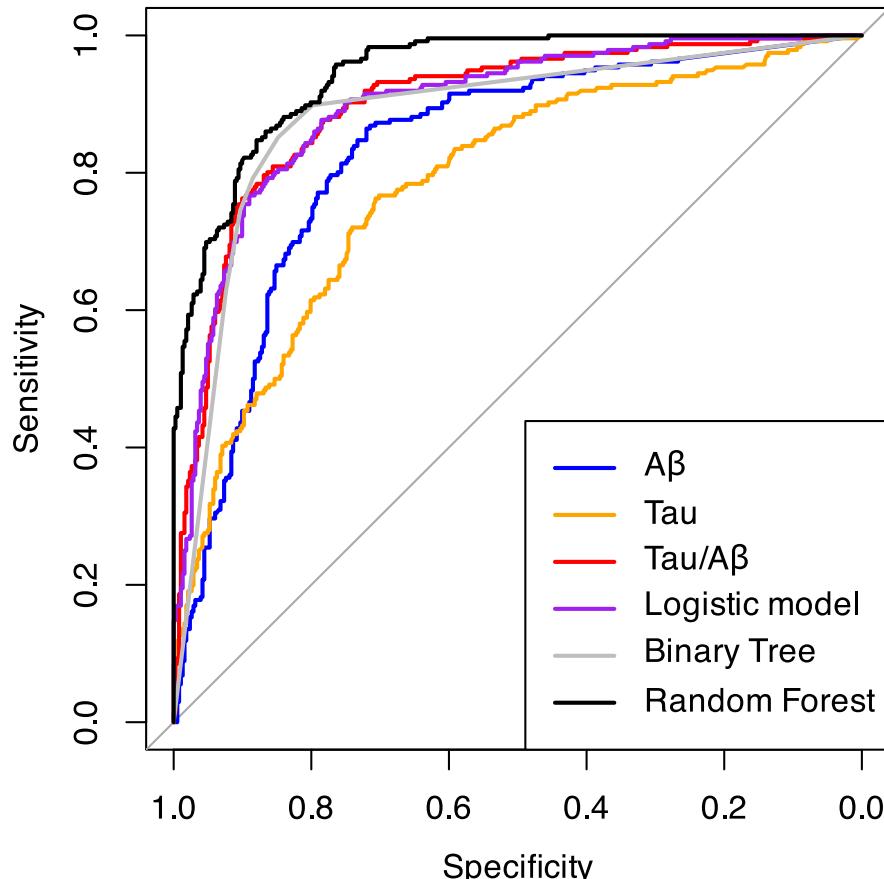
Regression Trees



Tree-based Methods



Comparing ROC Curves



Marker	AUC	95% CI	P-value *
A β	0.83	0.8, 0.86	
Tau	0.78	0.75, 0.82	0.07
Tau/ A β	0.9	0.87, 0.92	<0.001
Logistic model	0.9	0.87, 0.92	<0.001
Binary Tree	0.88	0.86, 0.91	<0.001
Random Forest	0.95	0.93, 0.96	<0.001

* Bootstrap test comparing each row to A β (Robin et al., 2011)

Random Forests (Breiman, 2001; Hothorn et al., 2006) re-fit binary trees on random subsamples of the data, then aggregate resulting trees into a "forest". This results in smoother predictions and a smoother ROC curve.

- All three models should be (cross) validated, since they learn from known Dx.

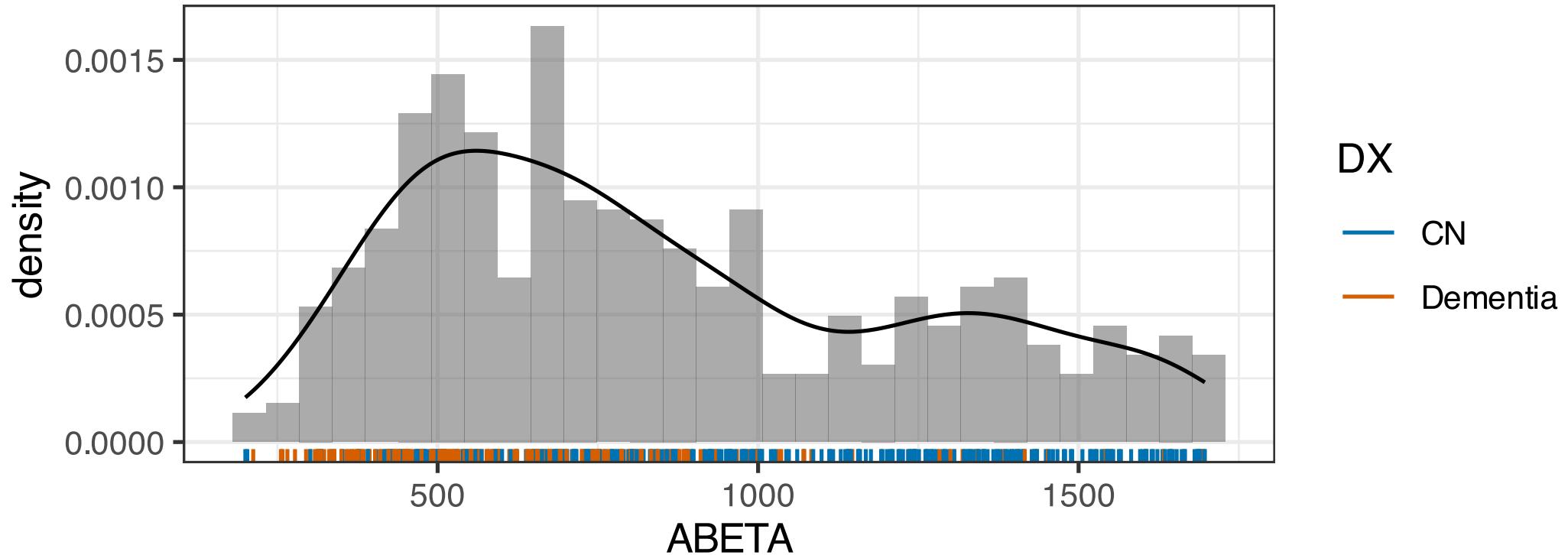
Unsupervised Learning

Mixture Modeling

Unsupervised Learning

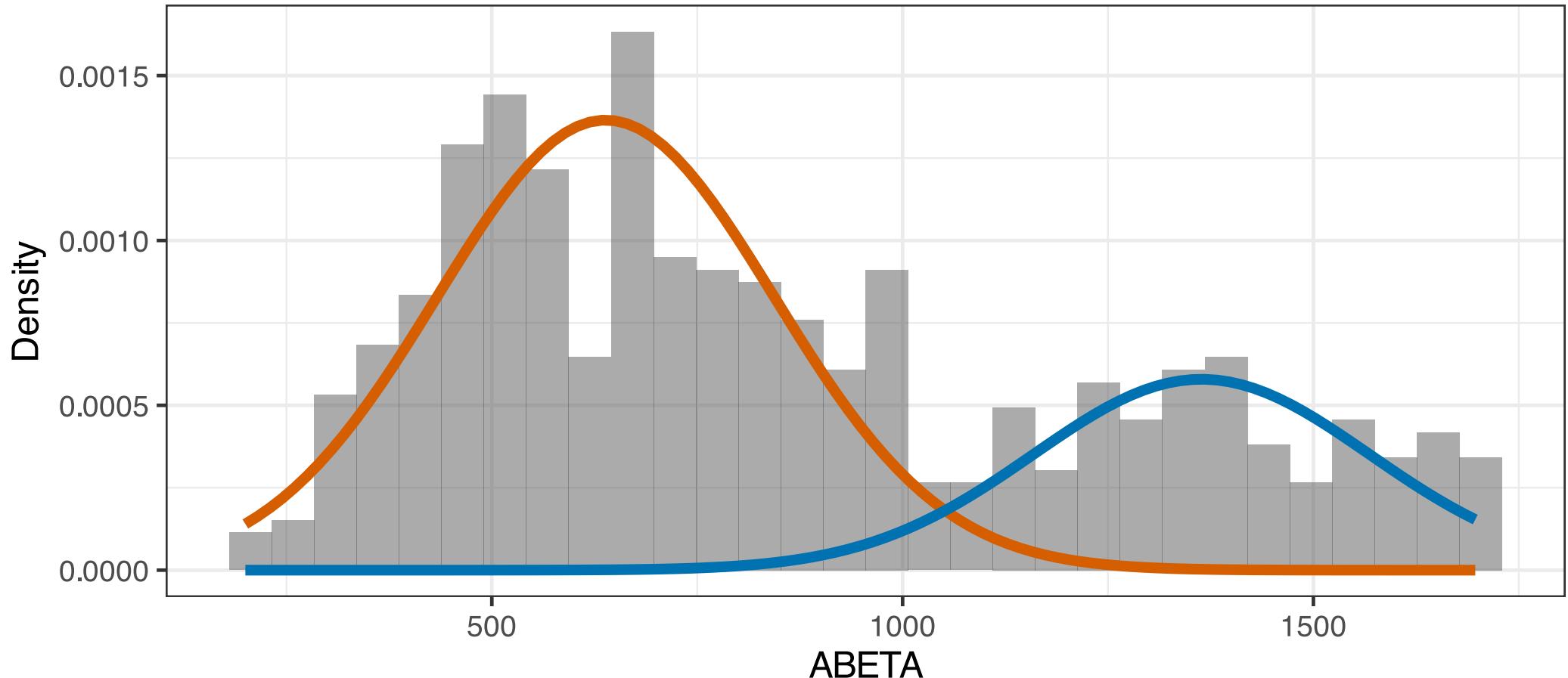
- The classification techniques we just reviewed can be thought of as *Supervised Learning* in which we attempt to learn known "labels" (CN, Dementia).
- *Mixture Modeling* is type of *Unsupervised Learning* technique in which we try to identify clusters of populations which appear to be arising from different distributions
- Don't confuse *Mixture Models* with *Mixed-Effects Models* (which we'll discuss later)
 - Think: "Mixture of Distributions"

Distribution of ABETA



- Distribution is bimodal
- Can we identify the two sub-distributions?
- We'll explore with `mixtools` package (Benaglia et al., 2009)

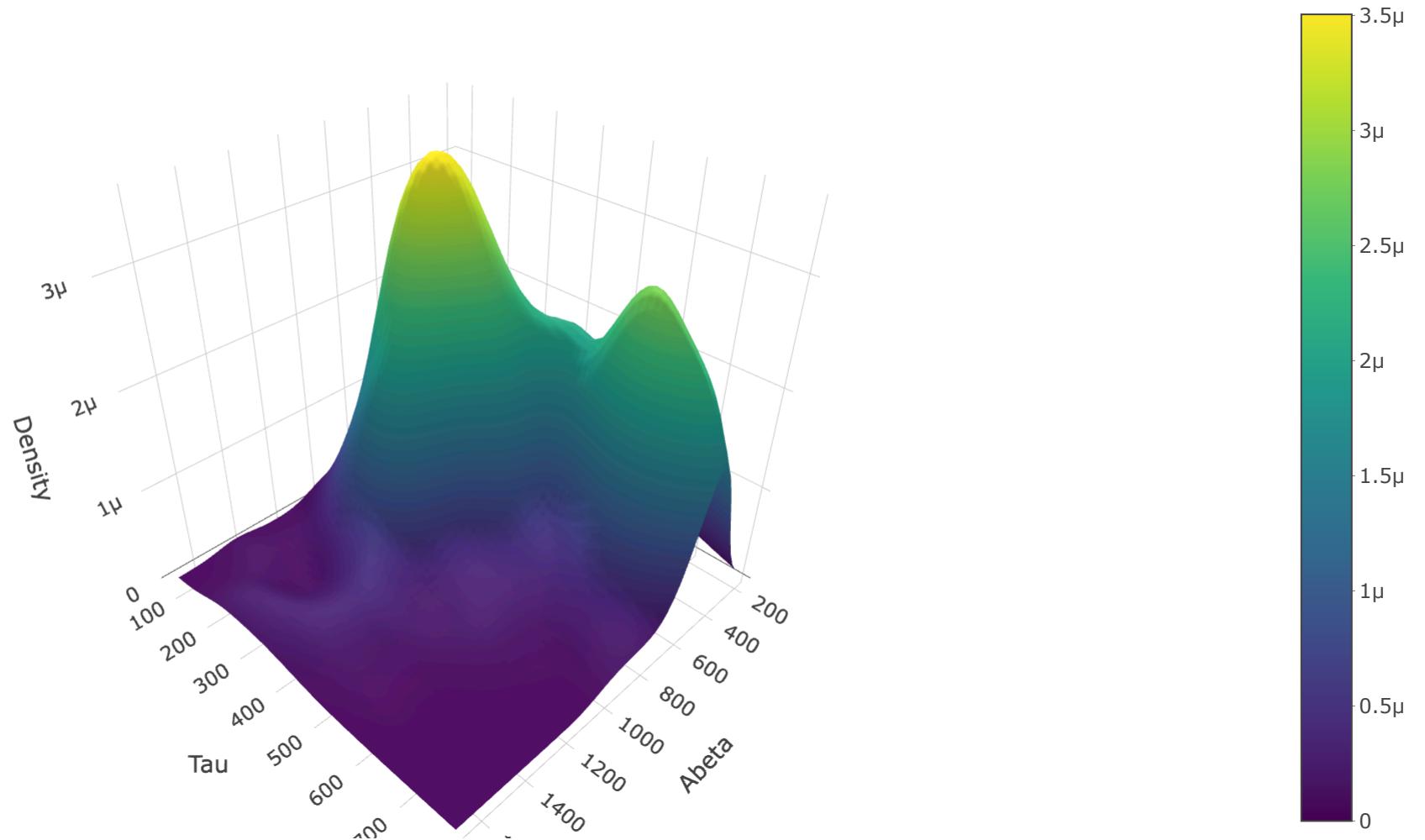
Distribution of ABETA



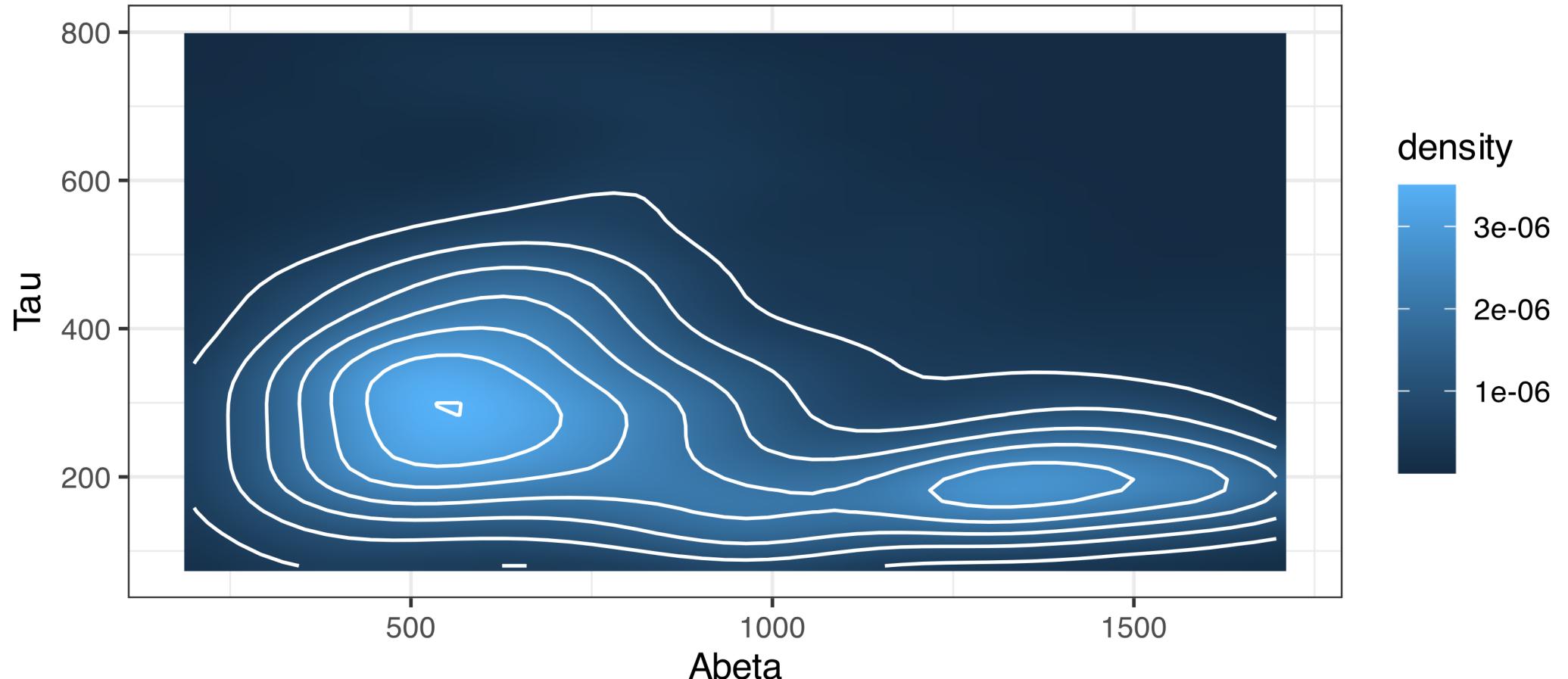
Posterior Membership Probabilities

Abeta	Prob. Abnormal	Prob. Normal
1033	0.58	0.42
1036	0.57	0.43
1044	0.53	0.47
1048	0.52	0.48
1061	0.46	0.54
1071	0.42	0.58
1071	0.42	0.58
1072	0.41	0.59

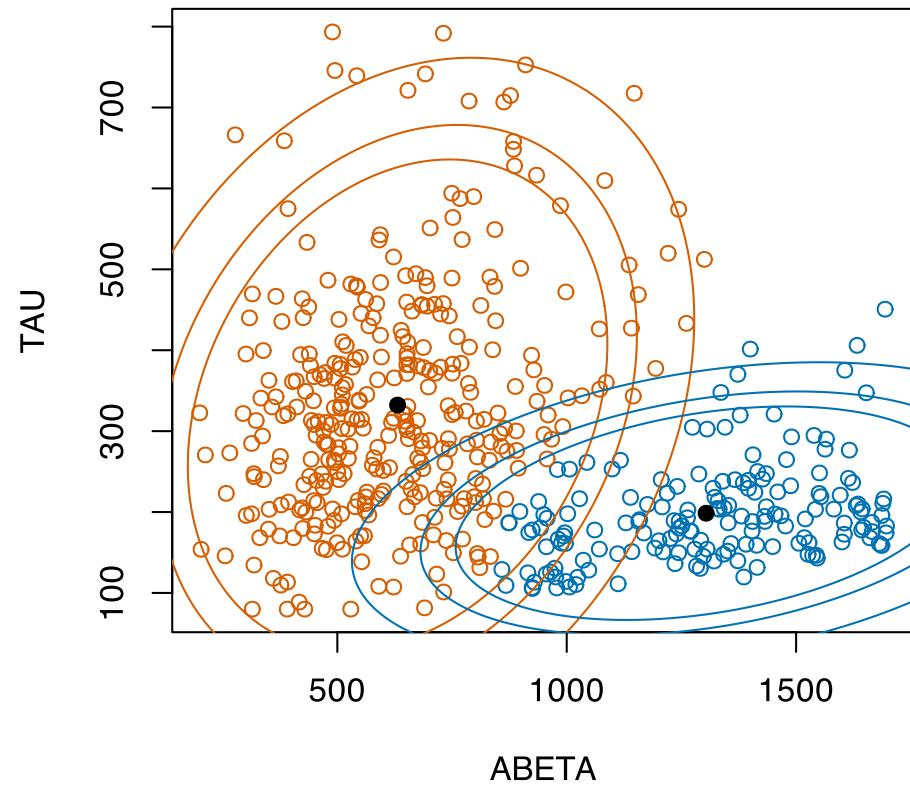
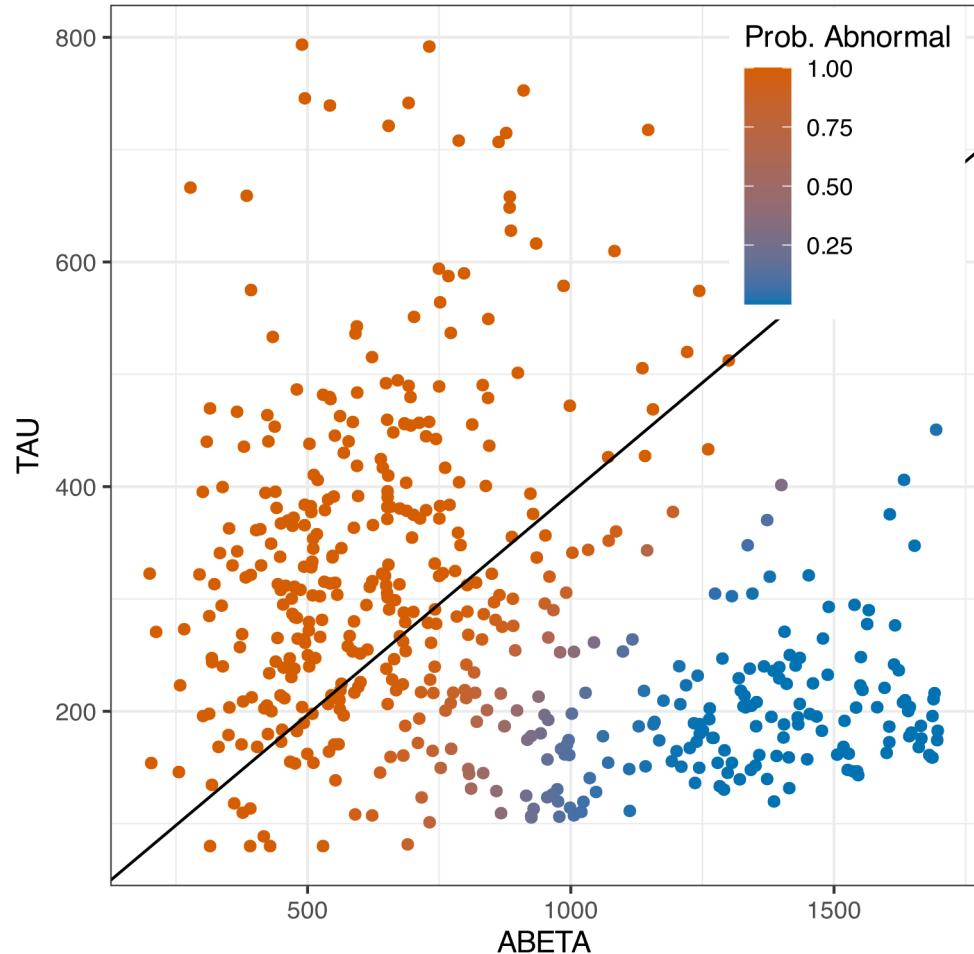
Bivariate Density



Bivariate Density Contour Plot



Bivariate Mixture Model Posterior Probabilities



Mixture of Experts (unsupervised learning with covariates)

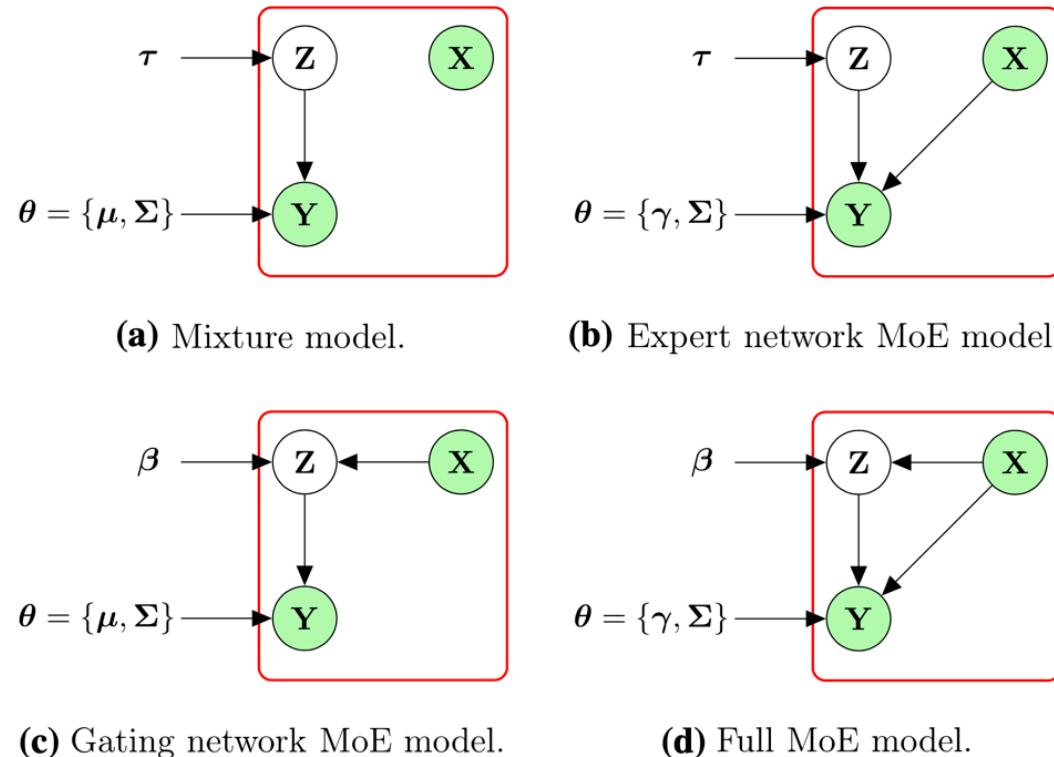
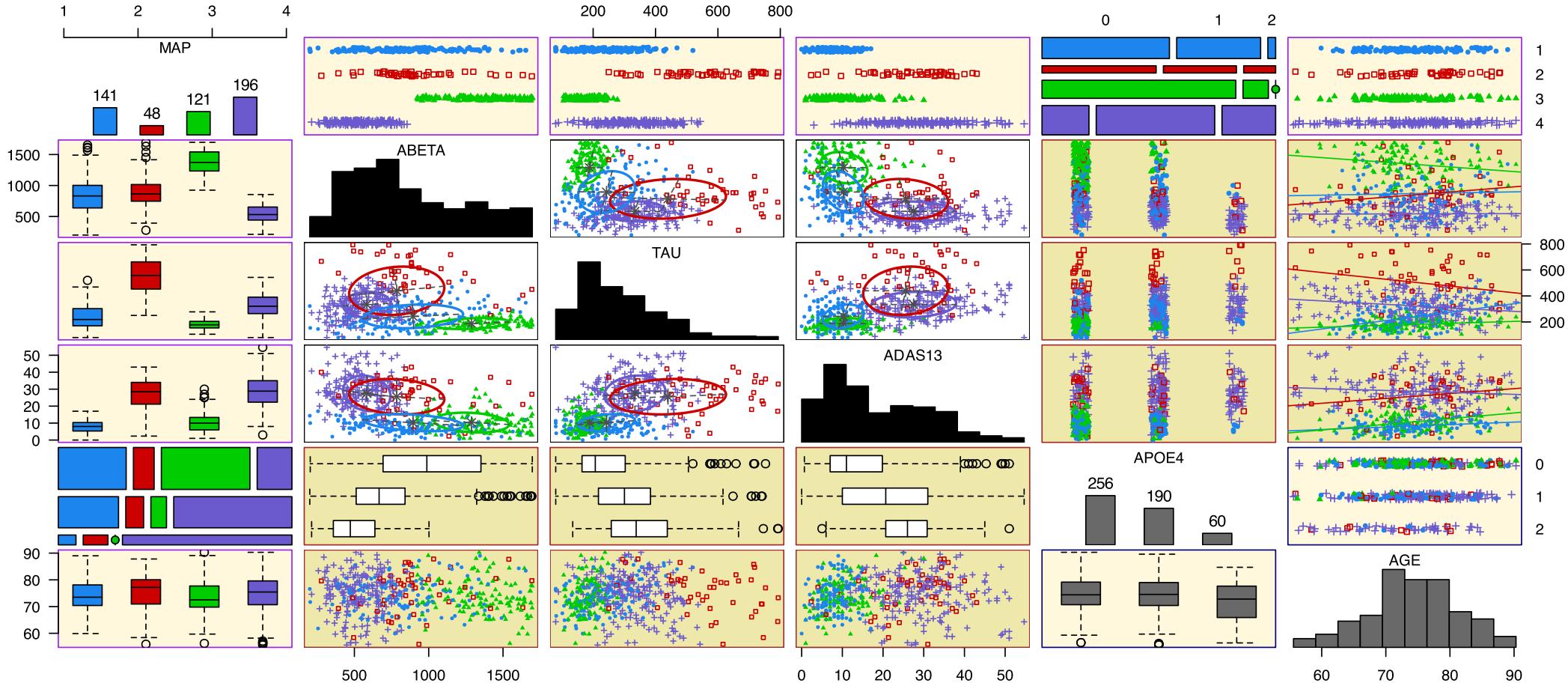


Fig. 1 The graphical model representation of the mixture of experts models. The differences between the special cases are due to the presence or absence of edges between the covariates X and the latent variable Z and/or response variable Y . Note that different subsets of the covariates in X can enter these two different parts of the full MoE model in (d)

Mixture of Experts (unsupervised learning with covariates)



MAP = maximum a posteriori classification

Summary of considerations for fluid biomarkers

- Batch Effects
- Experimental Design (Sample Randomization)
- Statistical Models for Assay Calibration/Quantification
- Classification (Supervised Learning)
 - Logistic Regression
 - Binary Trees
 - Random Forest
- Mixture Modeling (Unsupervised Learning)
 - Univariate
 - Bivariate
 - With covariates (Mixture of Experts)

Imaging: Reference Regions

Reference Regions

A common issue with statistical analyses of numeric summaries derived from imaging data is normalization to a reference region.

Examples:

- Volumetric MRI: regional (e.g. hippocampal) volume relative to Intra-Cranial Volume (ICV)
- Amyloid PET: cortical-to-cerebellum Standardized Uptake Value Ratio (SUVr)

(Also an issue in non-imaging data, e.g. ratio of CSF $\text{A}\beta_{1-42}$ to $\text{A}\beta_{1-40}$)

Beware the "Ratio Fallacy" and "Spurious Correlation"

Spurious correlation refers to the correlation between indices that have a common component. A 'per ratio' standard is based on a biological measurement adjusted for some physical measurement by division. Renowned statisticians and biologists (Pearson, Neyman and Tanner) have warned about the problems in interpretation that ratios cause. This warning has been largely ignored. The consequences of using a single ratio as either the dependent or one of the independent variables in a multiple-regression analysis are described. It is shown that the use of ratios in regression analyses can lead to incorrect or misleading inferences. A recommendation is made that the use of ratios in regression analyses be avoided.

Kronmal (1993) "Spurious correlation and the fallacy of the ratio standard revisited".
Journal of the Royal Statistical Society: Series A.

Examples abound: %ICV, SUVR, BMI, ...

Ratio as Dependent Variable in Regression

```
lm_fit1_sex <- lm(I(Hippocampus/ICV*100) ~ Sex, data=dd2)
```

Coefficient	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.49	0	150.8	<0.001
SexFemale	0.04	0	9.3	<0.001

- Significant association between Sex and Hippocampal Volume (%ICV)!
- (This linear model is essentially just a two-sample t-test)
- Model fit to ADNI CN

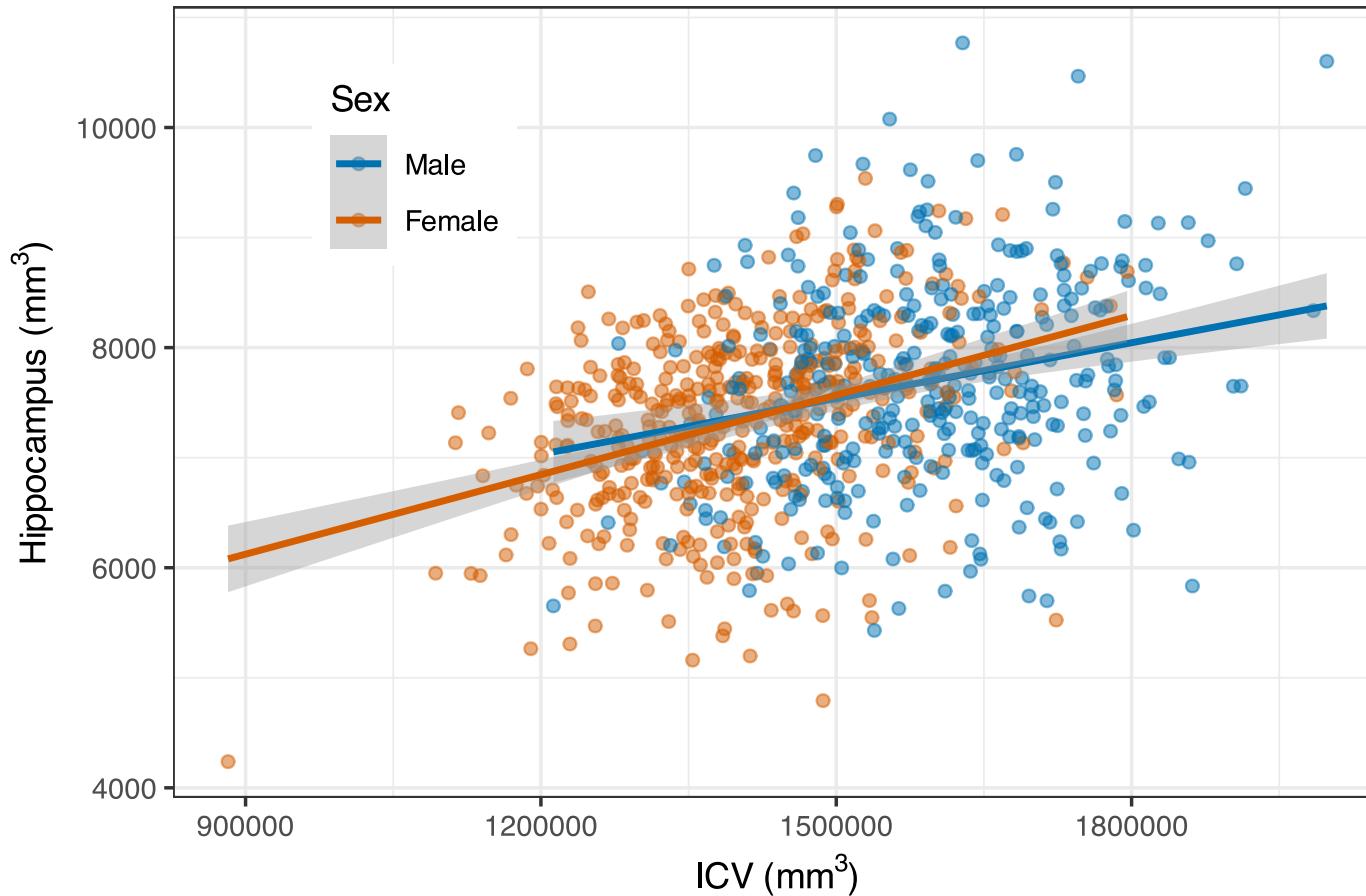
ICV as Covariate Instead of Denominator

```
lm_fit2_sex <- lm(Hippocampus ~ ICV + Sex, data=dd2)
```

Coefficient	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4389	361	12.16	<0.001
ICV	0	0	9.23	<0.001
SexFemale	34	72	0.47	0.64

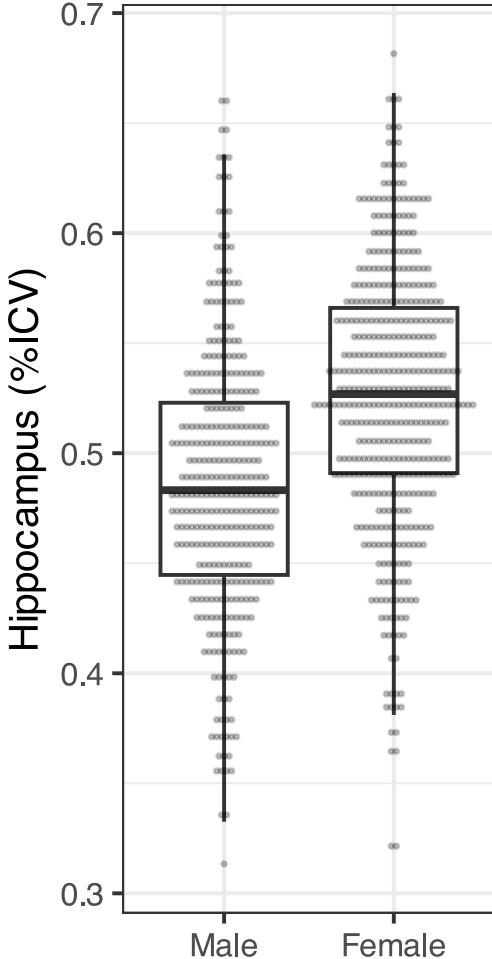
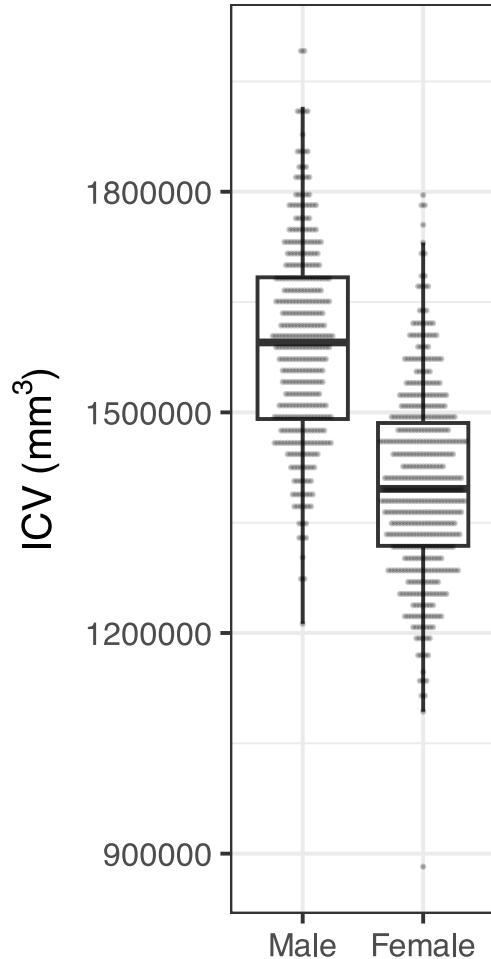
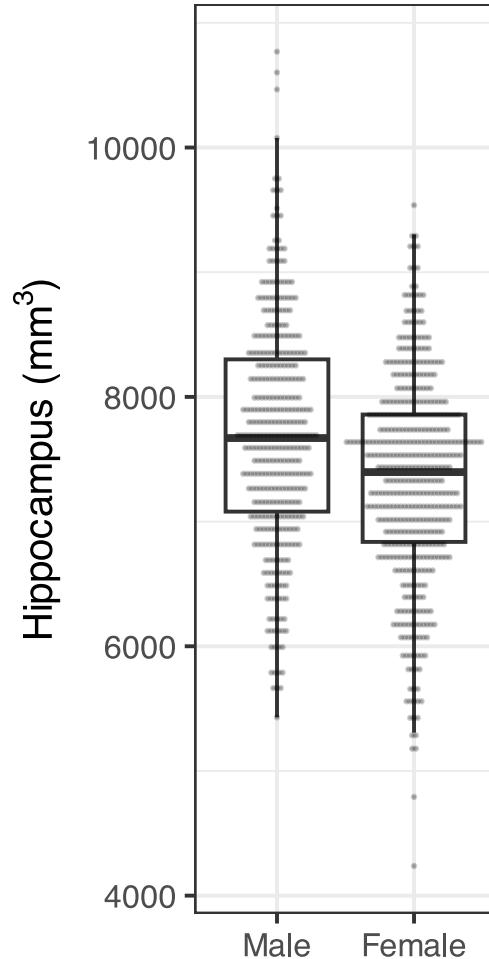
- No association between Sex and Hippocampal Volume (mm³)!
- The **spurious association** in first model is driven by denominator (ICV), not numerator (hippocampal volume)

Hippocampus vs ICV and Sex

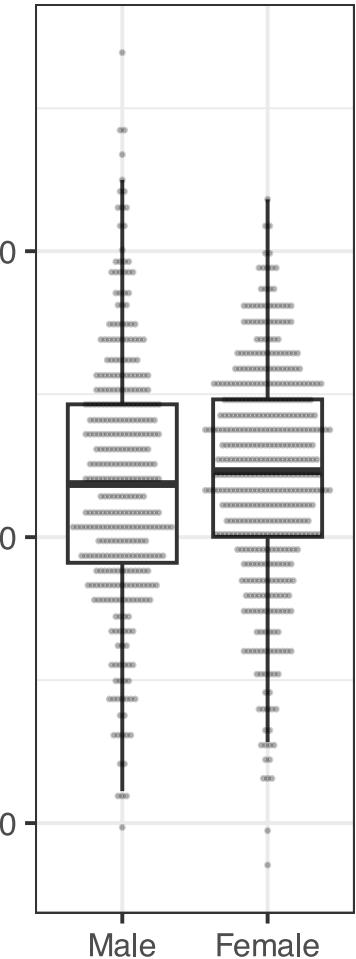


- For a given ICV, there is not much difference between sexes
- These linear fits are more flexible than model above (two slopes for ICV vs one)

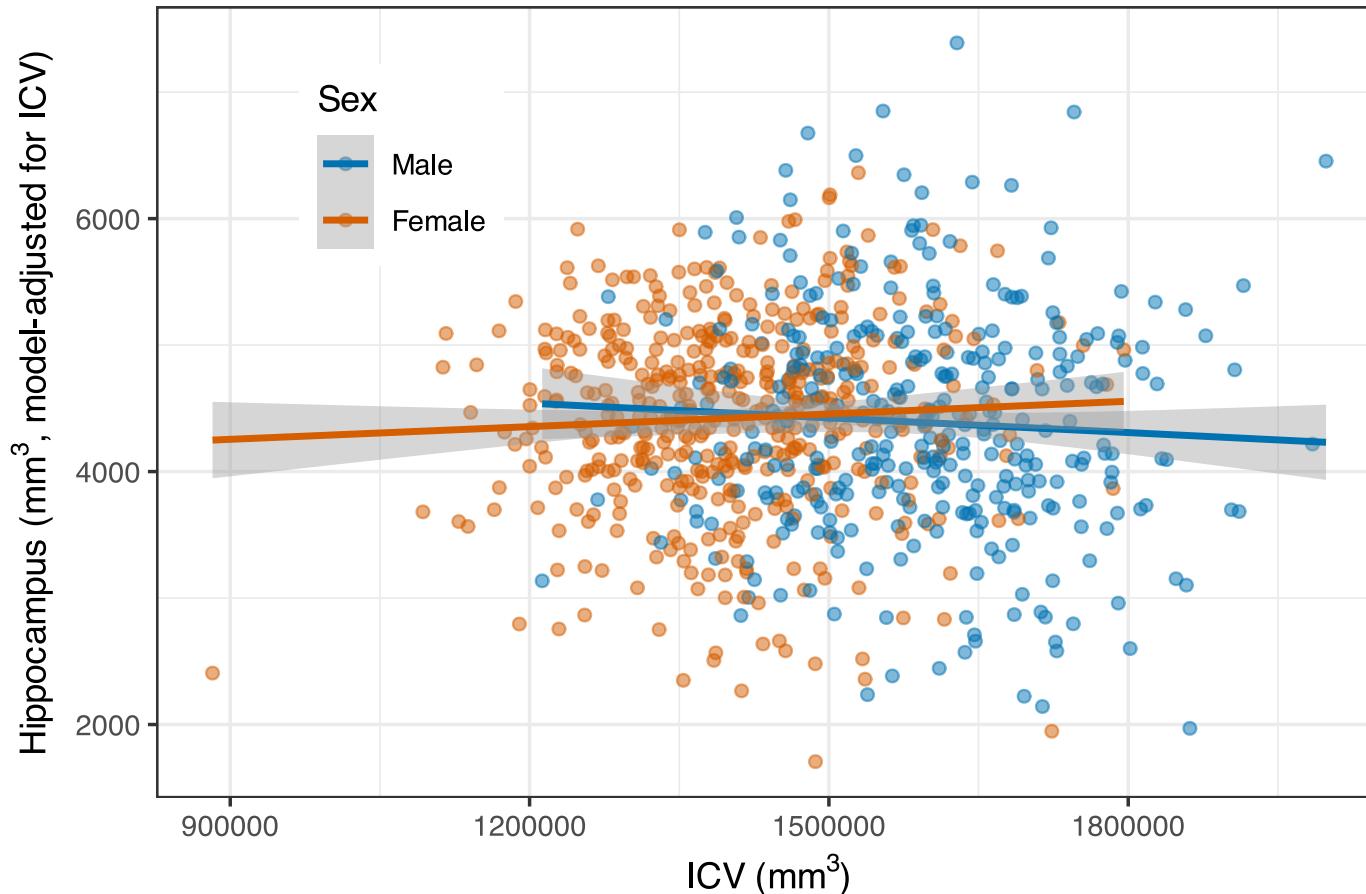
Hippocampal volume (mm³, %ICV, and model-adjusted)



Hippocampus (mm³, model-adjusted for ICV)



Hippocampal volume (mm³, %ICV, and model-adjusted)



- Model:

$$Y_i = \beta_0 + \beta_{\text{ICV}} \text{ICV}_i + \beta_{\text{Male}} \text{Male}_i + \varepsilon_i$$

- Adjusted volume:

$$Y_i - \hat{\beta}_{\text{ICV}} \text{ICV}_i$$

- Adjustment removes slope for ICV

Centiloids

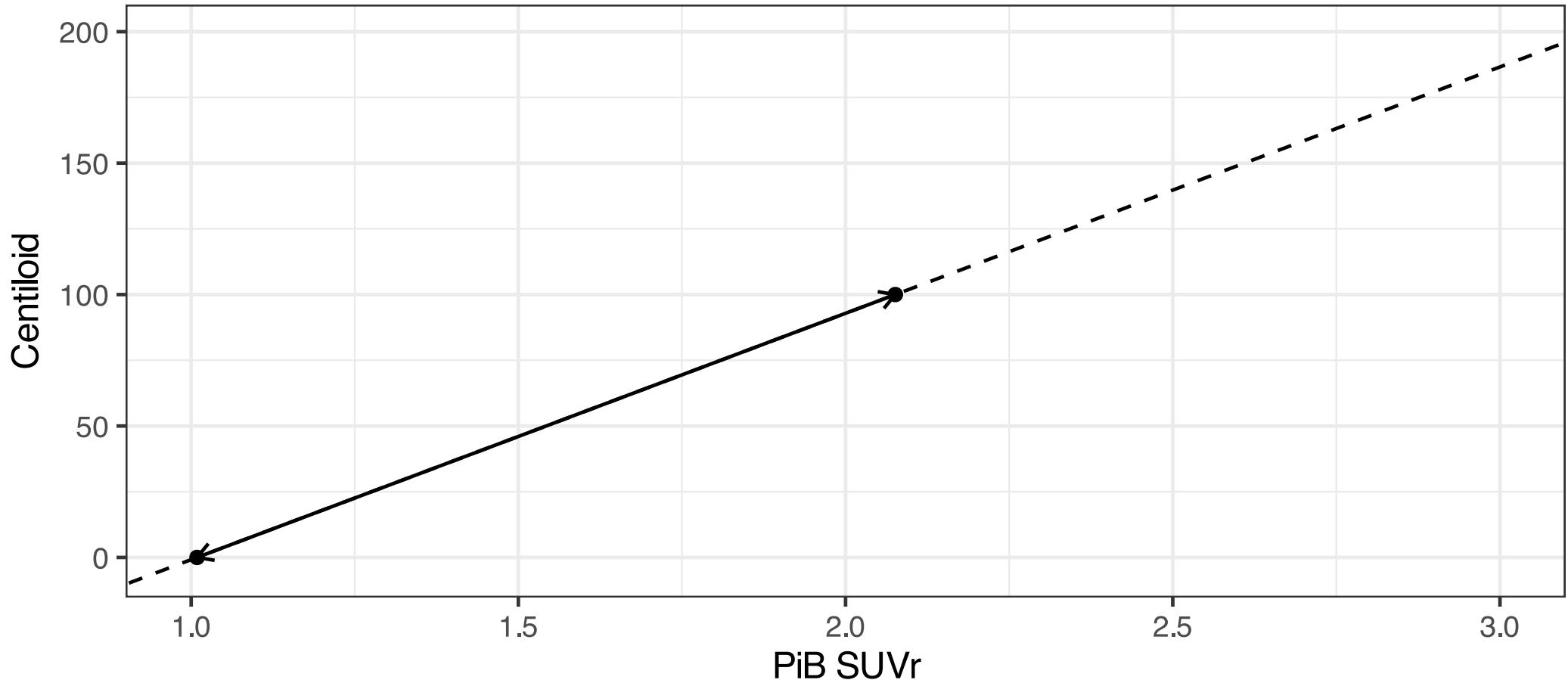
Centiloids

- Centiloids have become the industry standard standardized measure of amyloid PET (Klunk et al., 2015)
- Helpful in multi-site studies where different amyloid PET tracers might be utilized.
- What is a centiloid?

What is a centiloid? Step 1: PiB SUVr to CL

- The original/base PiB SUVr to centiloid map:
- 1.009 PiB SUVr (mean in young controls) → 0 CL
- 2.076 PiB SUVr (mean in AD cases) → 100 CL
- For other PiB SUVr values, draw the line from (1.009 SUVr, 0 CL) to (2.076 SUVr, 100 CL)
- $CL = 100 \times (\text{PiB SUVr} - 1.009)/1.067$ (Klunk et al., 2015)

What is a centiloid? Step 1: PiB SUVr to CL



What is a centiloid? Step 2: Other SUV_r to PiB SUV_r

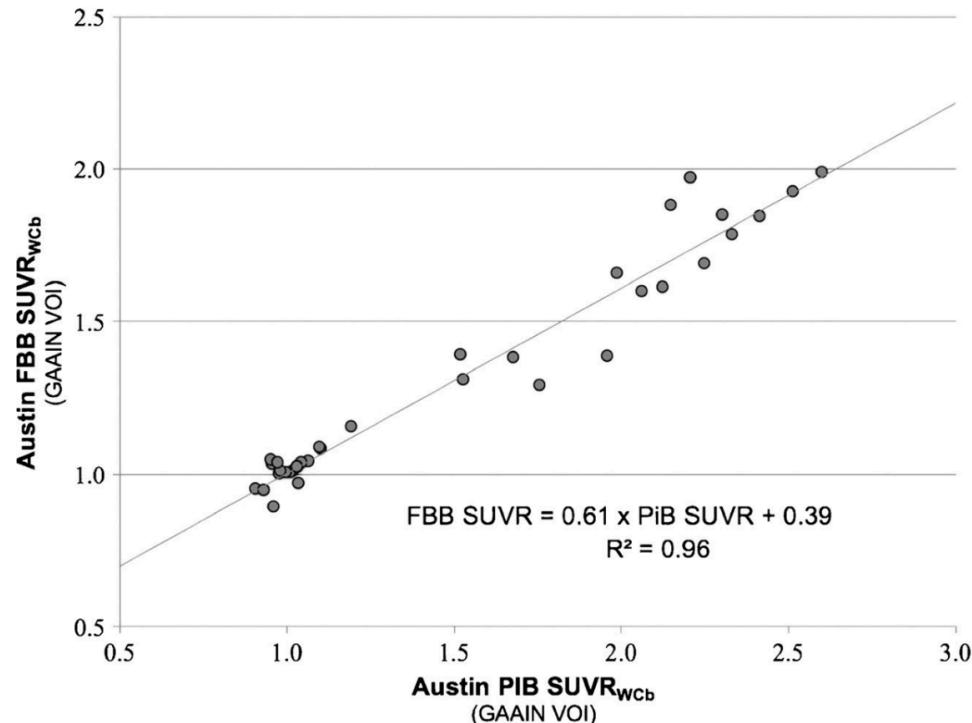
- Step 2 requires *paired* data: each individuals scanned with PiB and the other tracer
- Linear regression (ordinary least squares) is used to create linear transformation
- Transformed data is then mapped to CL using map on prior slide

What is a centiloid? Step 2: Other SUVR to PiB SUVR

Eur J Nucl Med Mol Imaging (2017) 44:2053–2059

2057

Fig. 4 Plot of the paired ^{11}C -PiB SUVR_{WCb} and ^{18}F -FBB SUVR_{WCb} for each subject calculated by the standard Centiloid method with the standard large single cortical region of interest and the whole cerebellum as reference region



$$\text{SUVR}_{\text{FBB}} = 0.61 \times \text{SUVR}_{\text{PiB}} + 0.39$$

Rowe et al. (2017)

Distribution mapping

Properzi et al.

Page 20

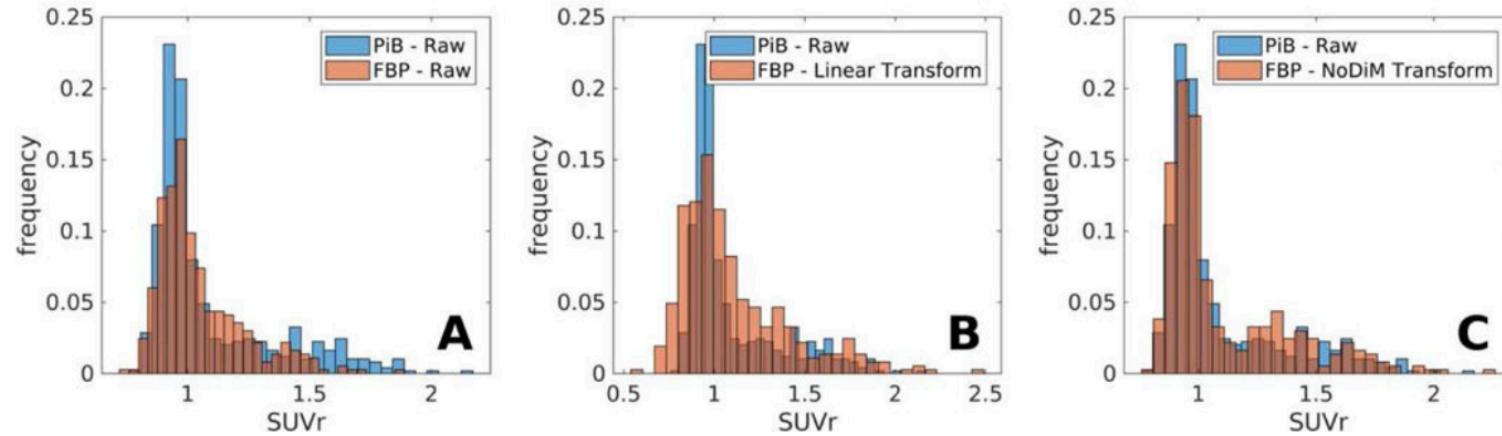
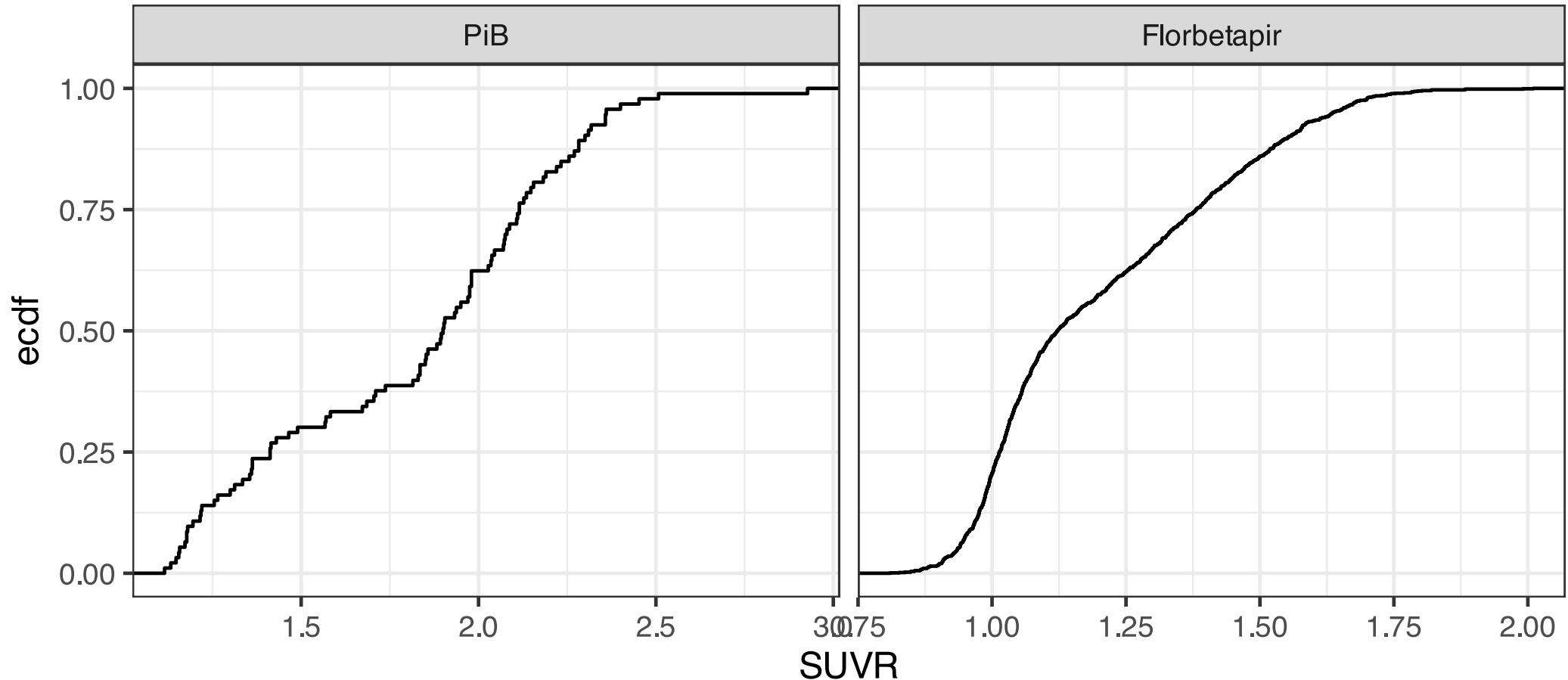


FIGURE 4.
Empirical Histogram Harmonization: Empirical PiB and raw (Panel A), linearly transformed (Panel B), and NoDiM transformed (Panel C) FBP histograms highlight the distributional differences pre and post transformation.

Properzi et al. (2019)

Empirical Cumulative Distribution Function (ECDF)



Weighted ECDFs to correct for sampling differences

N (%) per diagnosis and tracer:

	PiB	Florbetapir
CN	20 (21.5%)	498 (38.5%)
MCI	49 (52.7%)	588 (45.4%)
Dementia	24 (25.8%)	209 (16.1%)

Inverse proportion weights:

	PiB	Florbetapir
CN	4.7	2.6
MCI	1.9	2.2
Dementia	3.9	6.2

Weighted ECDFs to correct for sample differences

```
# Record the sampling adjustment weights in the data
dd <- dd %>% mutate(
  wt = case_when(
    DX == 'CN' & Tracer == 'PiB' ~ invproptab['CN', 'PiB'],
    DX == 'MCI' & Tracer == 'PiB' ~ invproptab['MCI', 'PiB'],
    DX == 'Dementia' & Tracer == 'PiB' ~ invproptab['Dementia', 'PiB'],
    DX == 'CN' & Tracer == 'Florbetapir' ~ invproptab['CN', 'Florbetapir'],
    DX == 'MCI' & Tracer == 'Florbetapir' ~ invproptab['MCI', 'Florbetapir'],
    DX == 'Dementia' & Tracer == 'Florbetapir' ~ invproptab['Dementia', 'Florbetapir']
  )))

```

Weighted ECDFs to correct for sample differences

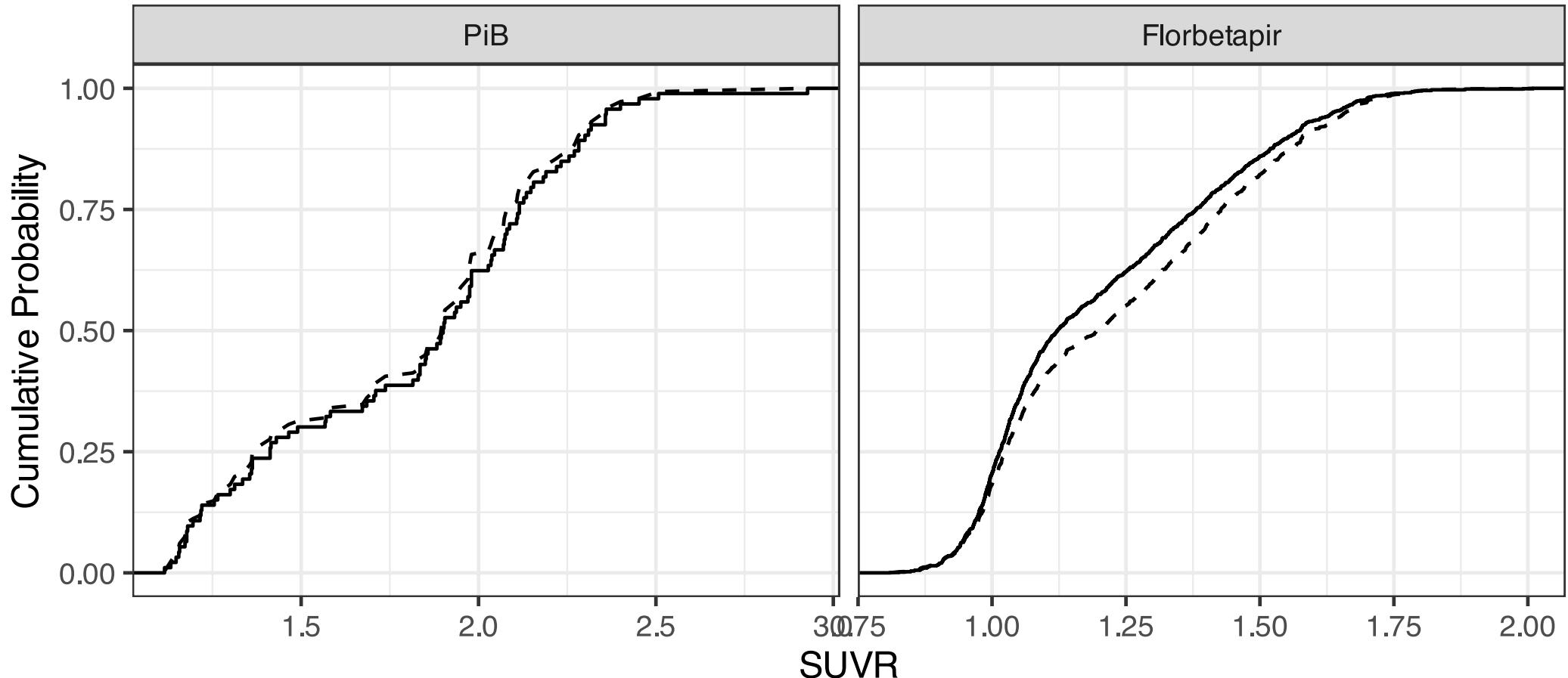
```
# Create adjusted ECDF functions (mapping SUVRs to Cumulative Probabilities)
# Hmisc::wtd.Ecdf returns a data.frame evaluating the ECDF at each observed value
PiB.ecdf <- with(subset(dd, Tracer == 'PiB'), ecdf.func(SUVR, weights=wt))
Fbp.ecdf <- with(subset(dd, Tracer == 'Florbetapir'), ecdf.func(SUVR, weights=wt))

# Create adjusted **inverse** ECDF functions
# mapping Cumulative Probabilities (0 to 1 scale) to SUVRs
PiB.inv.ecdf <- with(subset(dd, Tracer == 'PiB'),
inv.ecdf(SUVR, weights=wt))
Fbp.inv.ecdf <- with(subset(dd, Tracer == 'Florbetapir'),
inv.ecdf(SUVR, weights=wt))
```

Weighted ECDFs to correct for sample differences

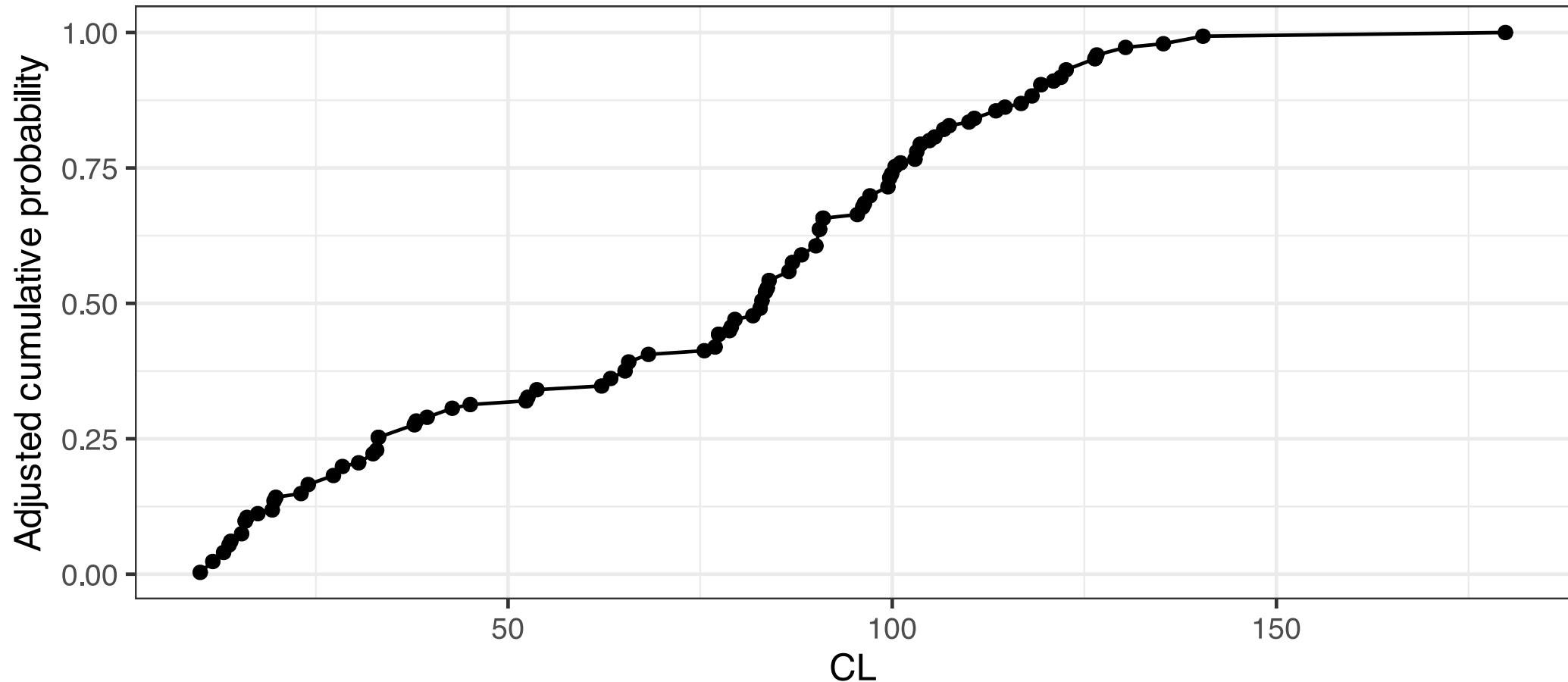
```
dd <- dd %>% mutate(
  `Adjusted cumulative probability` = case_when( #
    Tracer == 'PiB' ~ PiB.ecdf(SUVR),
    Tracer == 'Florbetapir' ~ Fbp.ecdf(SUVR)),
  `Adjusted Z-score` = qnorm(`Adjusted cumulative probability`), # adjusted z-scores
  `Florbetapir to PiB adjusted SUVR` = case_when(
    Tracer == 'Florbetapir' ~ PiB.inv.ecdf(Fbp.ecdf(SUVR))),
  `PiB to Florbetapir adjusted SUVR` = case_when(
    Tracer == 'PiB' ~ Fbp.inv.ecdf(PiB.ecdf(SUVR))),
  CL = case_when(
    Tracer == 'PiB' ~ 100*(SUVR - 1.009)/1.067)) %>%
arrange(Tracer, SUVR)
```

Weighted ECDFs to correct for sample differences

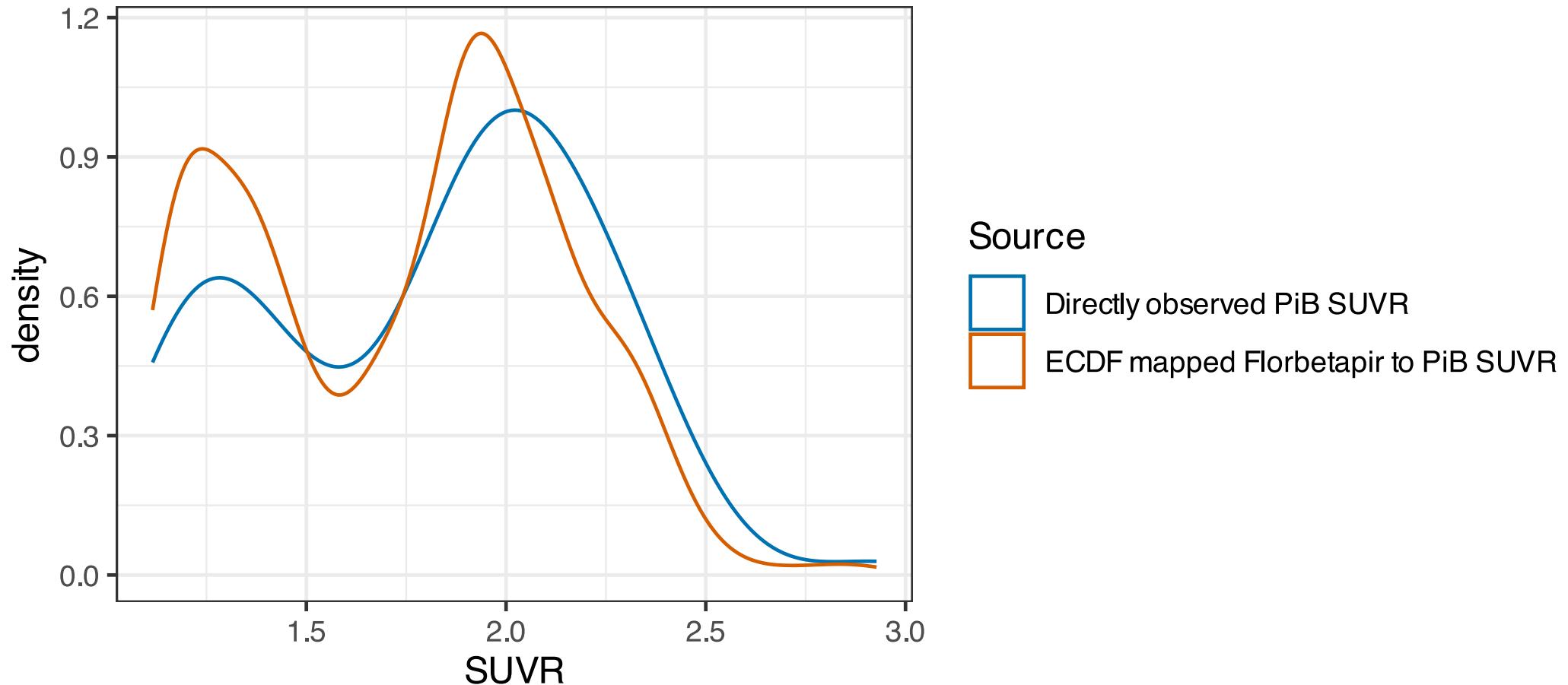


- solid line is un-adjusted ECDF; dashed line is ECDF adjusted for sample differences

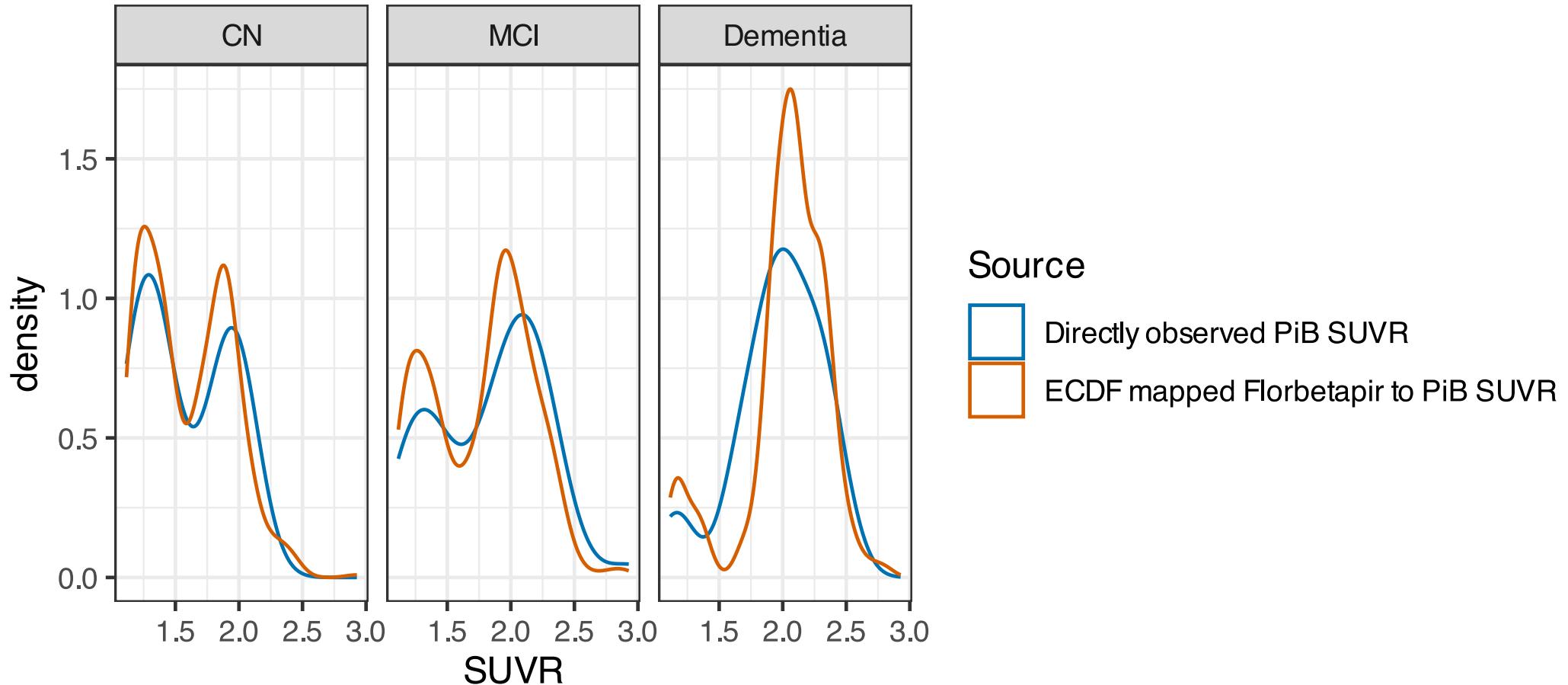
Centiloids vs ADNI Adjusted Percentiles



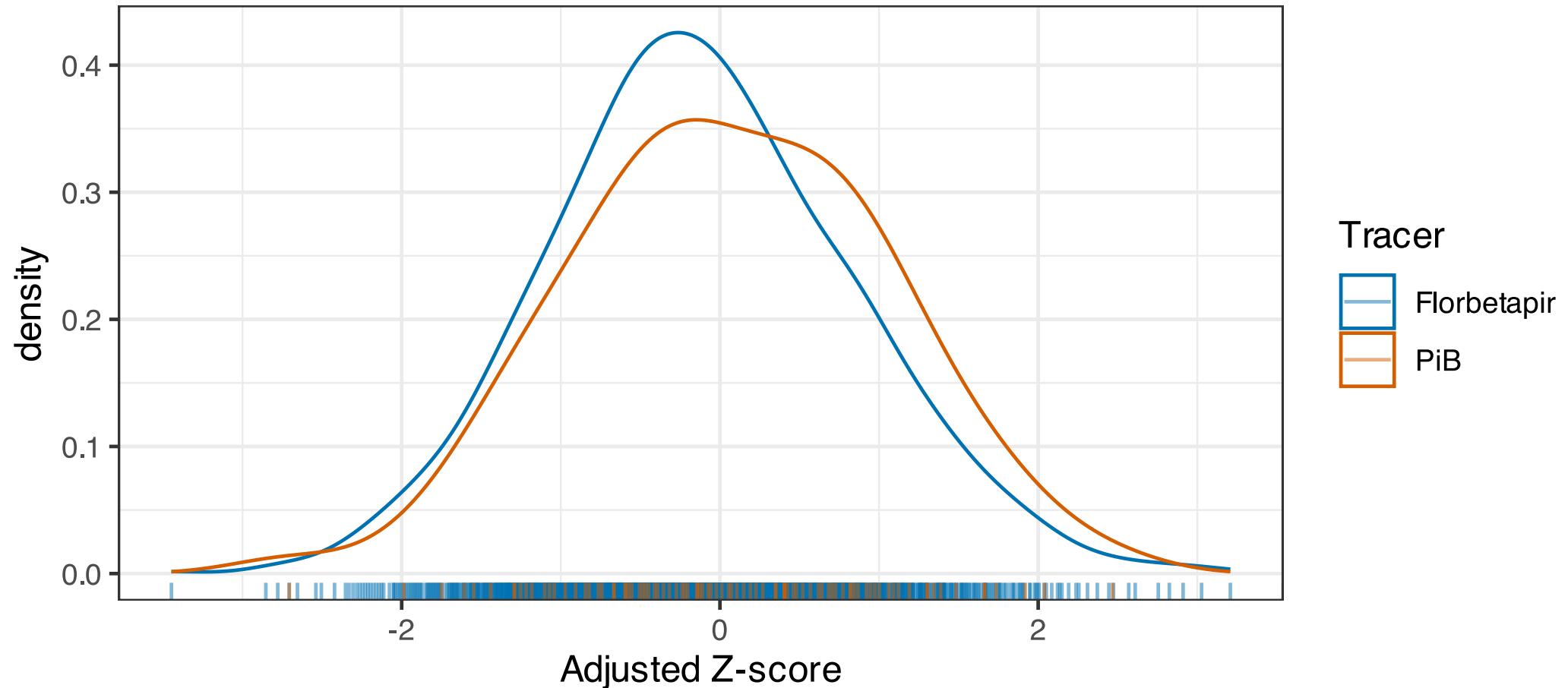
Densities for estimated and actual PiB SUVRs



Densities by diagnosis



Densities for weighted ECDF derived z-scores



Using the ECDF derived z-scores for analysis

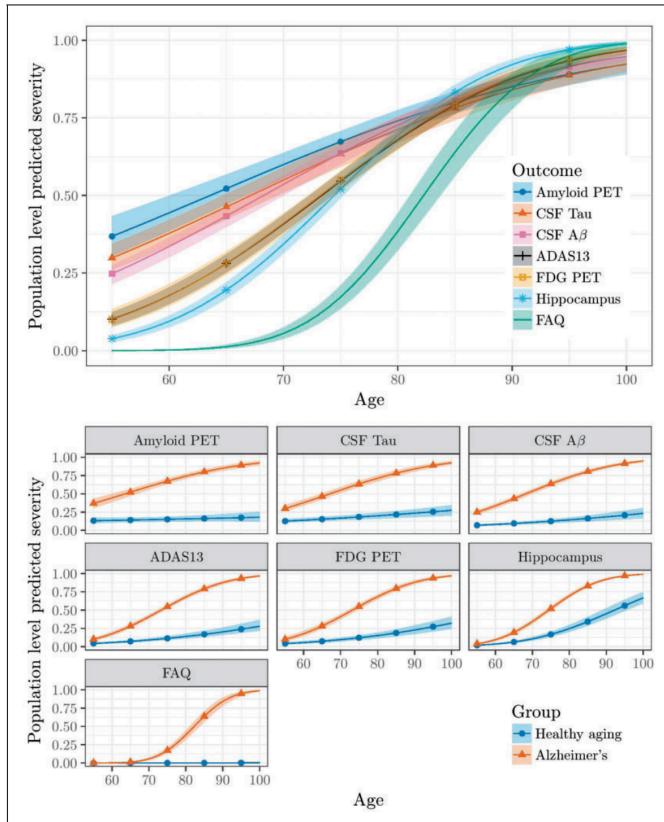
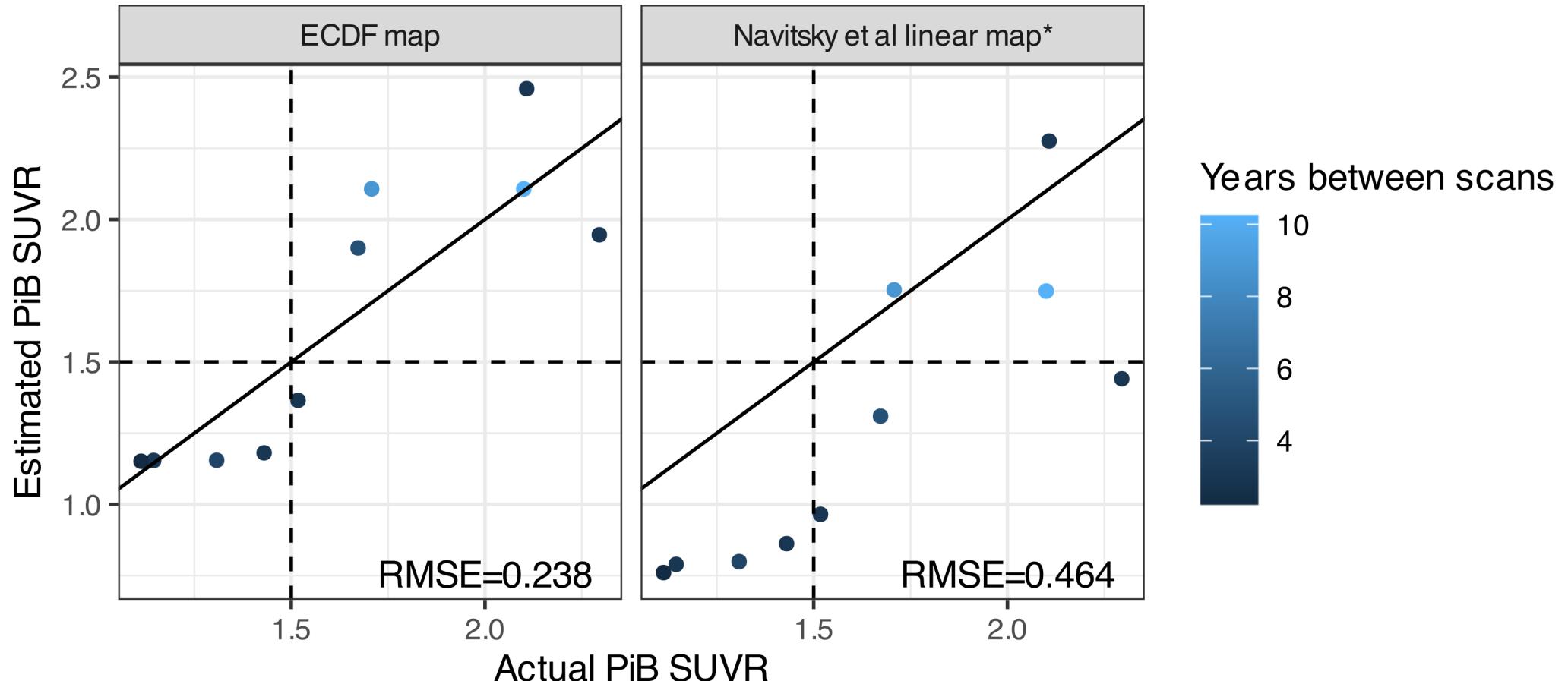


Figure 5. Modeled population-level severity. The top panel shows the modeled average trajectories. The depicted evolution is for female APOE ϵ 4 carriers with the ADNI mean education. Age is calibrated so that the ADAS13 trajectory attains the ADNI mean ADAS13 at the ADNI mean age at baseline. The bottom panel shows the same trajectories for progressive Alzheimer's (red triangles) with contrasting trajectories for healthy aging (blue dots). For the latter, the effect of latent time is forced to be zero to isolate the effect of age. ADAS13: Alzheimer's Disease Assessment Scale (13 Item version); FDG: fluorodeoxyglucose; PET: positron emission tomography; CSF: cerebrospinal fluid; FAQ: Functional Activities Questionnaire; ADNI: Alzheimer's Disease Neuroimaging Initiative.

- ECDF derived z-scores were used in model (assuming Gaussian residuals)
- Posterior estimates (on z-score scale) then back transformed to cumulative probabilities
- Natural interpretation:
 - 0 (most healthy) to 1 (most severe)
 - Comparable across different measures

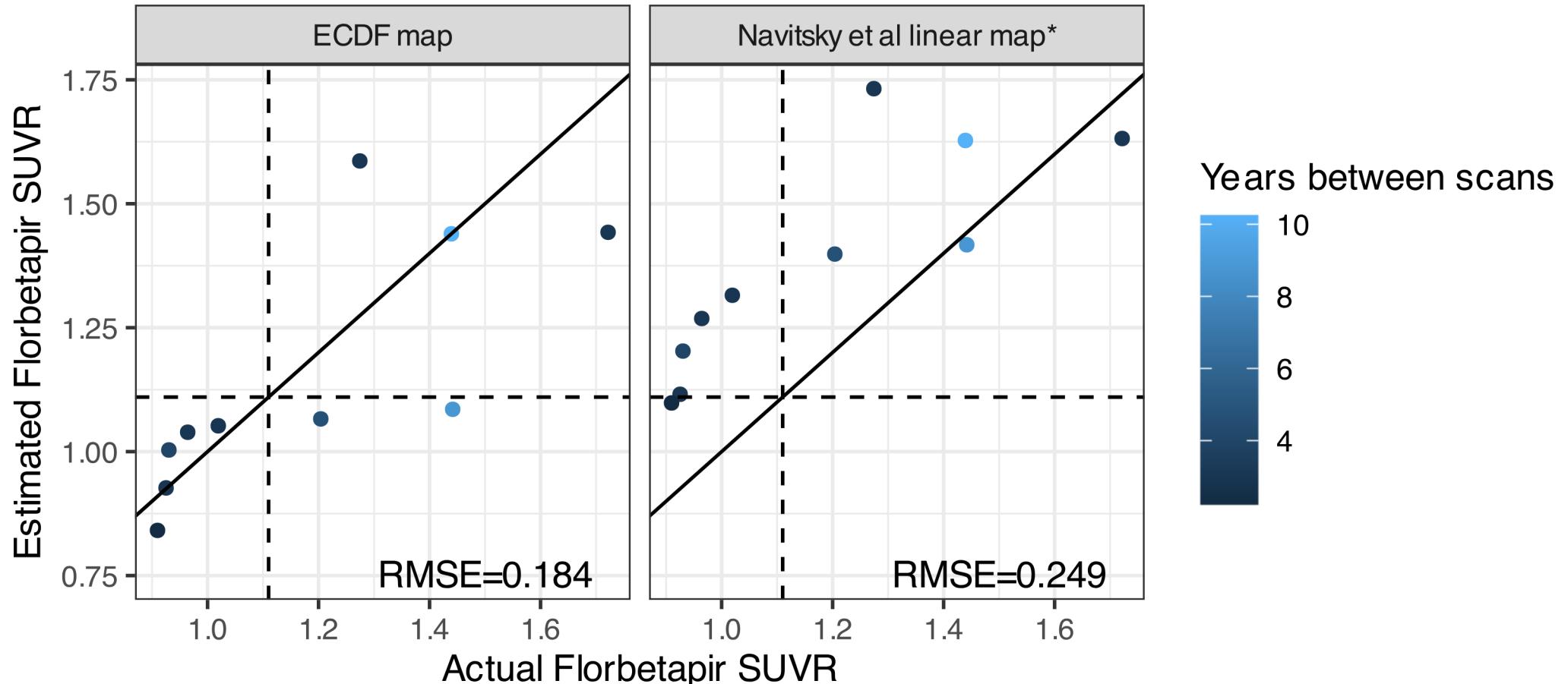
(Li et al., 2019)

Validation on a holdout set with both tracers (FBB → PiB)



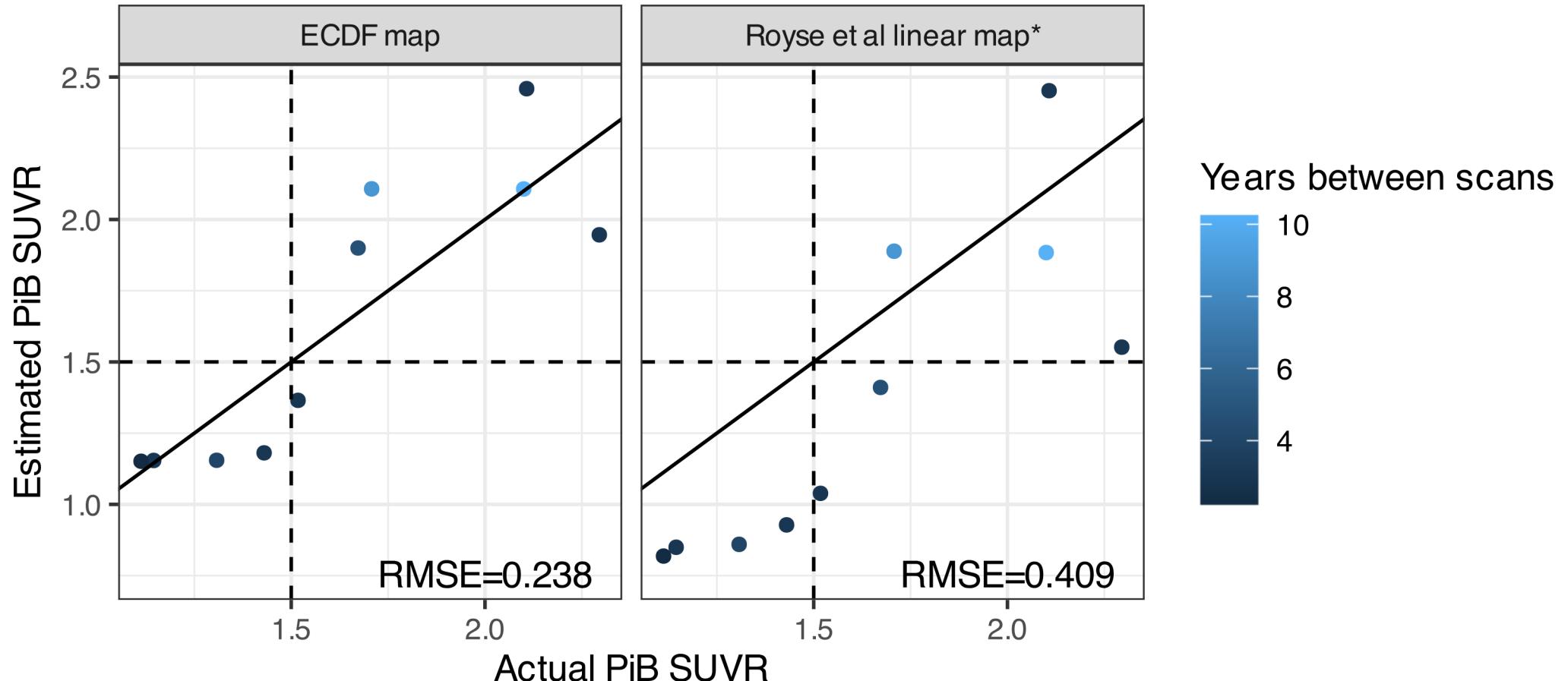
* Using $\text{PiB} = (\text{Florbetapir} - 0.502)/0.536$ from Navitsky et al. (2018)

Validation on a holdout set with both tracers (PiB → FBB)



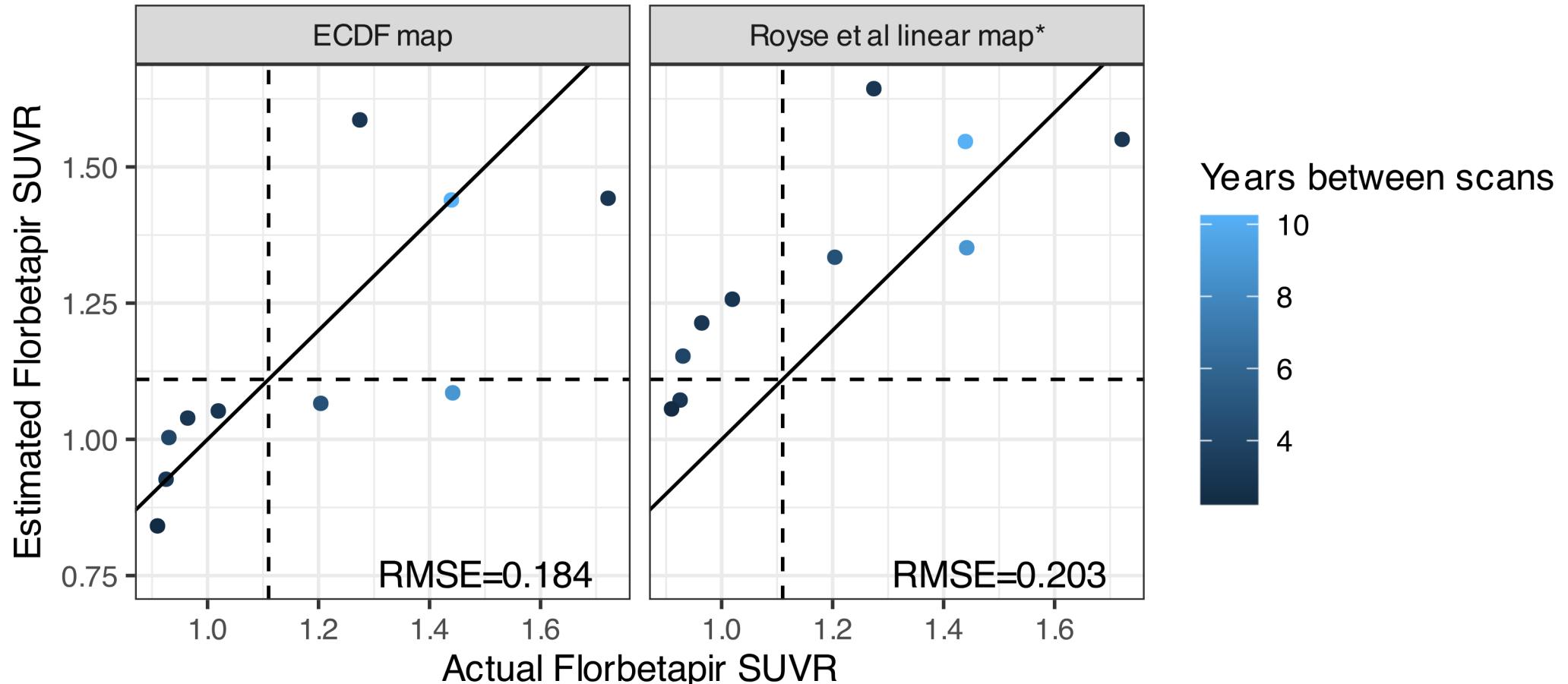
* Using $\text{Florbetapir} = \text{PiB} \times 0.536 + 0.502$ from Navitsky et al. (2018)

Validation on a holdout set with both tracers (FBB → PiB)



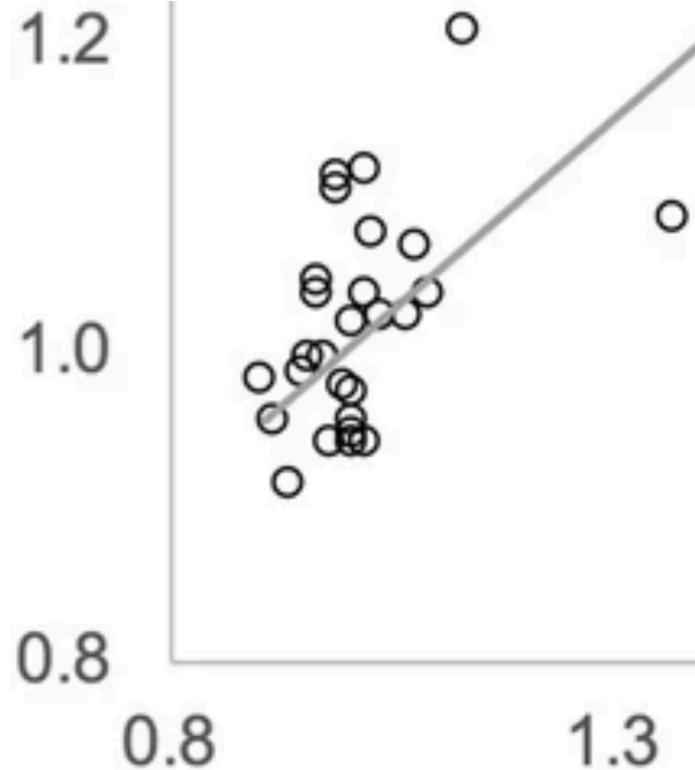
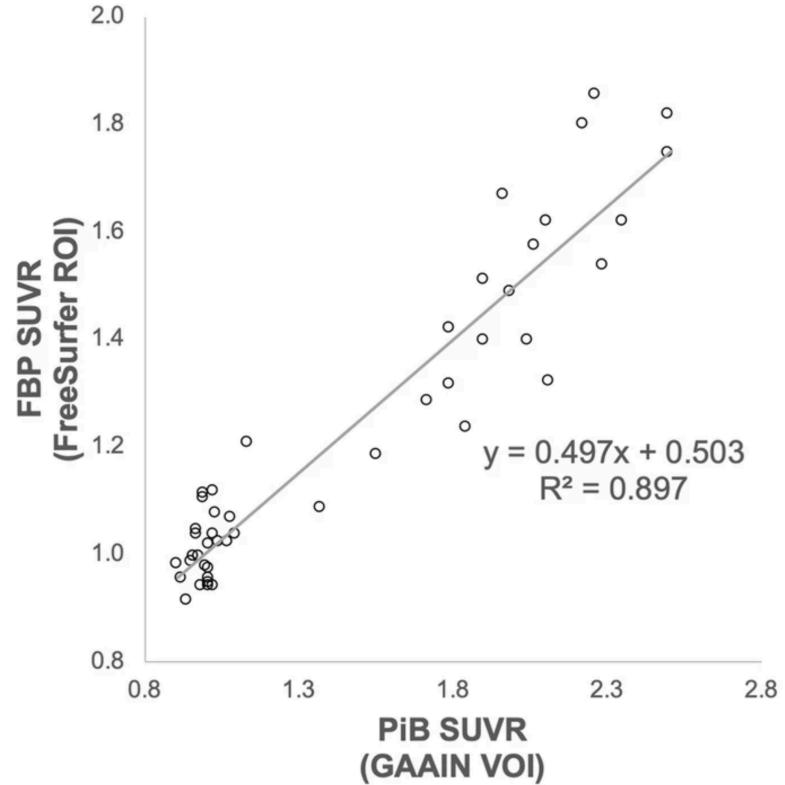
* Using $\text{PiB} = (\text{Florbetapir} - 0.503)/0.497$ from Royse et al. (2021)

Validation on a holdout set with both tracers (PiB → FBB)



* Using $\text{Florbetapir} = \text{PiB} \times 0.497 + 0.503$ from Royse et al. (2021)

Where is linear map struggling?



From Royse et al. (2021)

ComBat Step 1: Estimate effect of tracer on mean & variance

Outcome is (natural) log(SUVR)

	Value	Std.Error	DF	t-value	p-value	
(Intercept)	-3.81e-02	3.31e-02	2.06e+03	-1.15	0.249	
TracerAV45	-3.34e-01	9.98e-03	2.06e+03	-33.43	< 2e-16	***
TracerFBB	-3.32e-01	1.27e-02	2.06e+03	-26.07	< 2e-16	***
APOE41	1.56e-01	8.52e-03	1.72e+03	18.29	< 2e-16	***
APOE42	2.43e-01	1.51e-02	1.72e+03	16.13	< 2e-16	***
DXMCI	1.39e-02	4.23e-03	2.06e+03	3.30	0.001	***
DXDementia	3.75e-02	6.45e-03	2.06e+03	5.81	7.3e-09	***
Age	6.29e-03	3.30e-04	2.06e+03	19.06	< 2e-16	***
MMSE	-3.23e-04	5.74e-04	2.06e+03	-0.56	0.573	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Johnson et al. (2007); Beer et al. (2020)

ComBat Step 1: Estimate effect of tracer on mean & variance

Random effects:

Formula: ~1 | RID
(Intercept) Residual
StdDev: 3.187 1

Variance function:

Structure: Different standard deviations per stratum
Formula: ~1 | Tracer
Parameter estimates:
AV45 PIB FBB
1.00 1.97 0.64

ComBat Step 2: Get fitted values and prediction assuming PiB

```
dd.combat$fitted_log.SUVR <- predict(combat.fit) %>% as.numeric()

dd.combat$pred_log.SUVR_PiB <- predict(combat.fit,
  newdata = dd.combat %>% mutate(Tracer = "PIB")) %>% as.numeric()
```

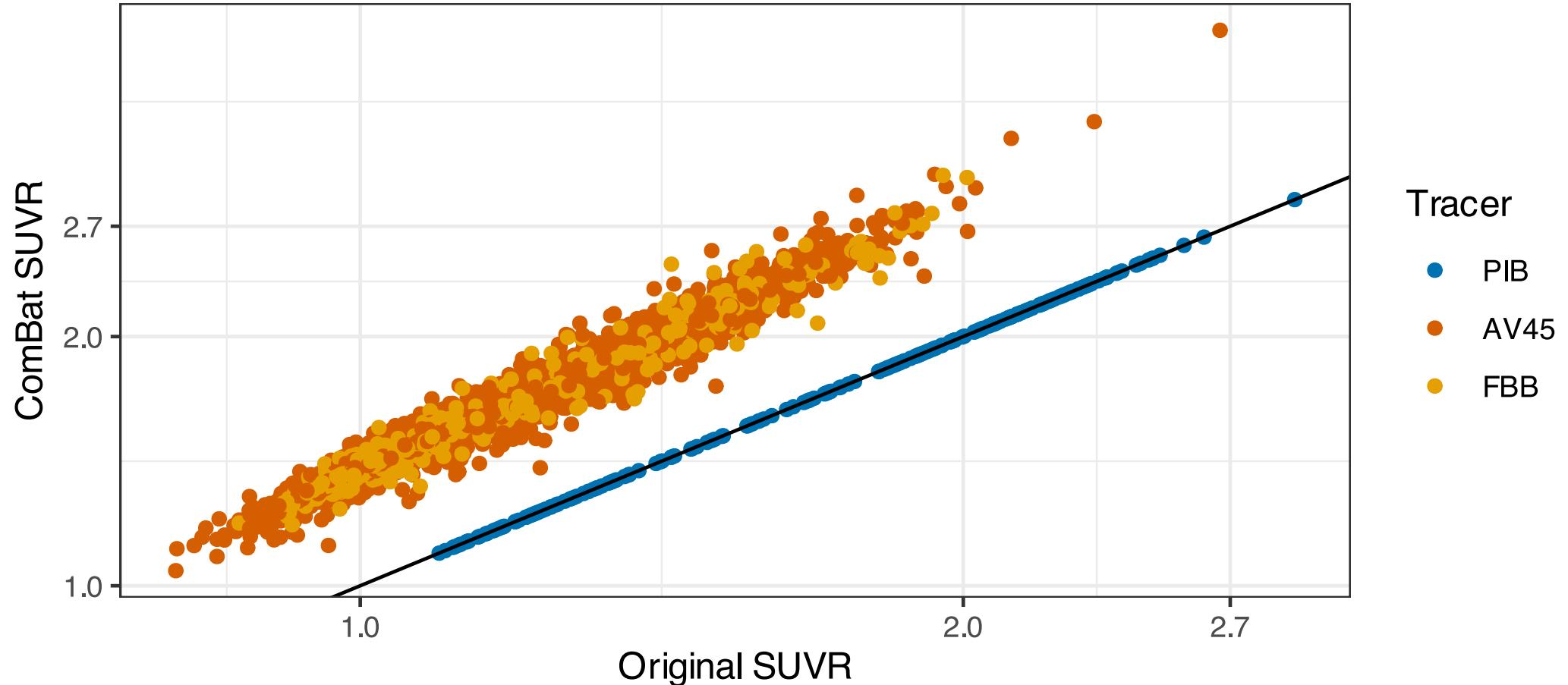
ComBat Step 3: Get residuals, and rescale them to PiB scale

```
(sdWeights <- coef(combat.fit$modelStruct$varStruct, unconstrained = FALSE))  
PIB  FBB  
1.97 0.64  
dd.combat <- dd.combat %>%  
  mutate(  
    resids = log.SUVR - fitted_log.SUVR,  
    homog_resids = case_when(  
      Tracer == 'PIB' ~ resids,  
      Tracer == 'AV45' ~ resids * sdWeights['PIB'] ,  
      Tracer == 'FBB' ~ resids * sdWeights['PIB'] / sdWeights['FBB']),  
    ComBat_log.SUVR_PiB = homog_resids + pred_log.SUVR_PiB,  
    fitted_SUVR = exp(fitted_log.SUVR),  
    ComBat_SUVR_PiB = exp(ComBat_log.SUVR_PiB))
```

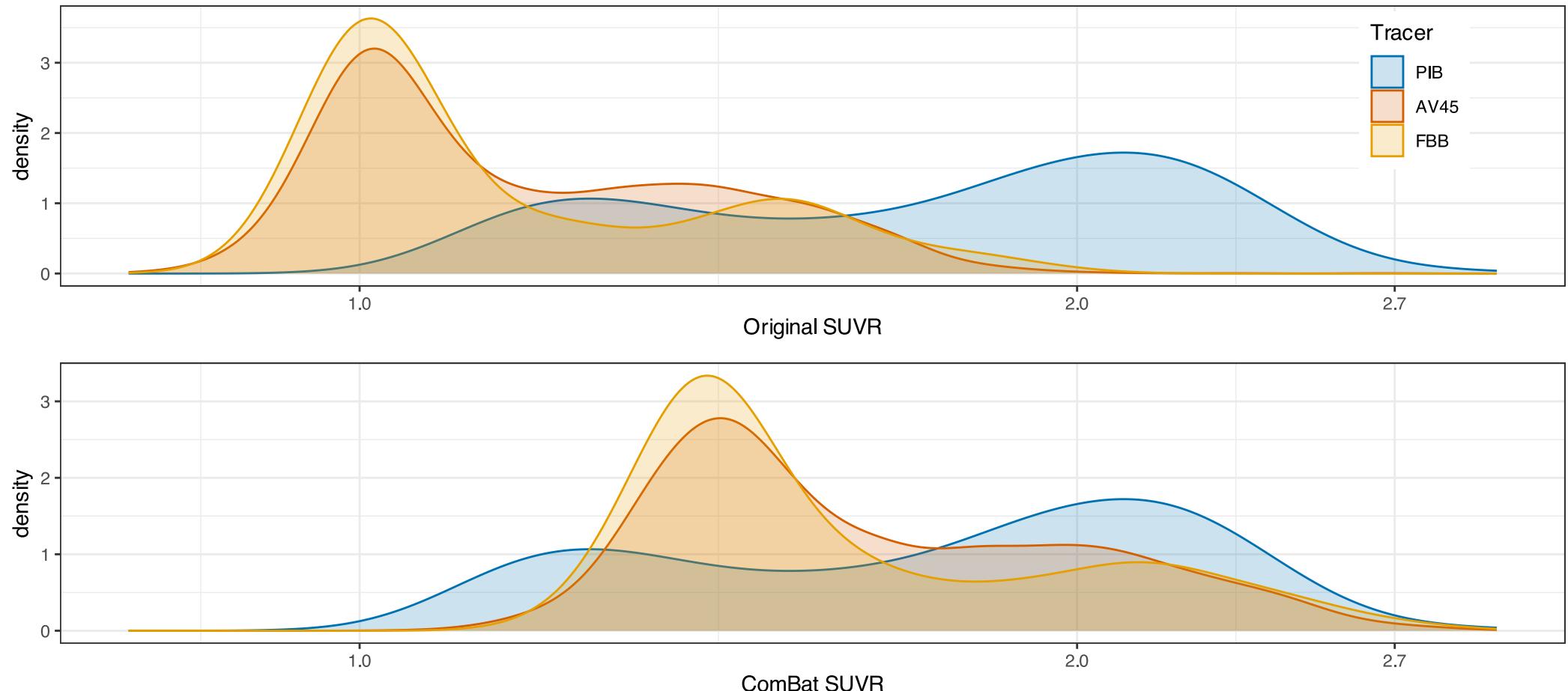
ComBat is the sum of predicted values and residuals "homogenized" to PiB

```
# A tibble: 3 × 3  
Tracer `SD of homog_resids` `SD of resids`  
<fct>          <dbl>        <dbl>  
1 PIB            0.0839       0.0839  
2 AV45           0.0757       0.0384  
3 FBB            0.0574       0.0185
```

ComBat results



ComBat results



Summary of considerations for imaging data

- Avoid using ratios whenever possible. Better to include "denominators" as covariates in regression models.
- The Empirical Cumulative Distribution Function (ECDF) is a useful non-parametric tool for standardization.
 - Can handle non-linear relationships
 - Percentiles are intuitive and familiar
 - ECDF derived maps will not produce out-of-range estimates
 - Percentiles can be mapped to standard normal z -scores to facilitate modelling

Thank you!

References (1/3)

- Beer, J. C., N. J. Tustison, P. A. Cook, C. Davatzikos, Y. I. Sheline, R. T. Shinohara, K. A. Linn, Alzheimer's Disease Neuroimaging Initiative, and others (2020). "Longitudinal ComBat: A method for harmonizing longitudinal multi-scanner imaging data". In: *Neuroimage* 220, p. 117129.
- Benaglia, T., D. Chauveau, D. R. Hunter, and D. Young (2009). "mixtools: An R Package for Analyzing Finite Mixture Models". In: *Journal of Statistical Software* 32.6, pp. 1-29. URL: <http://www.jstatsoft.org/v32/i06/>.
- Breiman, L. (2001). "Random forests". In: *Machine learning* 45.1, pp. 5-32.
- Davidian, M. and P. D. Haaland (1990). "Regression and calibration with nonconstant error variance". In: *Chemometrics and Intelligent Laboratory Systems* 9.3, pp. 231-248.
- Haaland, P., D. Samarov, and E. McVey (2011). *calibFit: Statistical models and tools for assay calibration*. R package version 2.1.0. URL: <https://CRAN.R-project.org/package=calibFit>.
- Hothorn, T., P. Buehlmann, S. Dudoit, A. Molinaro, and M. Van Der Laan (2006). "Survival Ensembles". In: *Biostatistics* 7.3, pp. 355-373.
- Hothorn, T., K. Hornik, and A. Zeileis (2006). "Unbiased Recursive Partitioning: A Conditional Inference Framework". In: *Journal of Computational and Graphical Statistics* 15.3, pp. 651-674.

References (2/3)

- Johnson, W. E. et al. (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics* 8.1, pp. 118-127.
- Klunk, W. E. et al. (2015). "The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET". In: *Alzheimer's & dementia* 11.1, pp. 1-15.
- Kronmal, R. A. (1993). "Spurious correlation and the fallacy of the ratio standard revisited". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 156.3, pp. 379-392.
- Li, D. et al. (2019). "Bayesian latent time joint mixed effect models for multicohort longitudinal data". In: *Statistical methods in medical research* 28.3, pp. 835-845.
- Murphy, K. et al. (2020). "Gaussian Parsimonious Clustering Models with Covariates and a Noise Component". In: *Advances in Data Analysis and Classification* 14.2, pp. 293-325. DOI: [10.1007/s11634-019-00373-8](https://doi.org/10.1007/s11634-019-00373-8). URL: <https://doi.org/10.1007/s11634-019-00373-8>.
- Murphy, K. et al. (2022). *MoEClust: Gaussian Parsimonious Clustering Models with Covariates and a Noise Component*. R package version 1.5.0. URL: <https://cran.r-project.org/package=MoEClust>.

References (3/3)

- Murphy, K. et al. (2020). "Gaussian Parsimonious Clustering Models with Covariates and a Noise Component". In: *Advances in Data Analysis and Classification* 14.2, pp. 293-325. DOI: [10.1007/s11634-019-00373-8](https://doi.org/10.1007/s11634-019-00373-8). URL: <https://doi.org/10.1007/s11634-019-00373-8>.
- Navitsky, M. et al. (2018). "Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the Centiloid scale". In: *Alzheimer's & Dementia* 14.12, pp. 1565-1571.
- Properzi, M. J. et al. (2019). "Nonlinear Distributional Mapping (NoDiM) for harmonization across amyloid-PET radiotracers". In: *Neuroimage* 186, pp. 446-454.
- Robin, X. et al. (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves". In: *BMC Bioinformatics* 12, p. 77.
- Rowe, C. C. et al. (2017). "18 F-Florbetaben PET beta-amyloid binding expressed in Centiloids". In: *European journal of nuclear medicine and molecular imaging* 44.12, pp. 2053-2059.
- Royse, S. K. et al. (2021). "Validation of amyloid PET positivity thresholds in centiloids: a multisite PET study approach". In: *Alzheimer's Research & Therapy* 13.1, pp. 1-10.
- Youden, W. J. (1950). "Index for rating diagnostic tests". In: *Cancer* 3.1, pp. 32-35.