



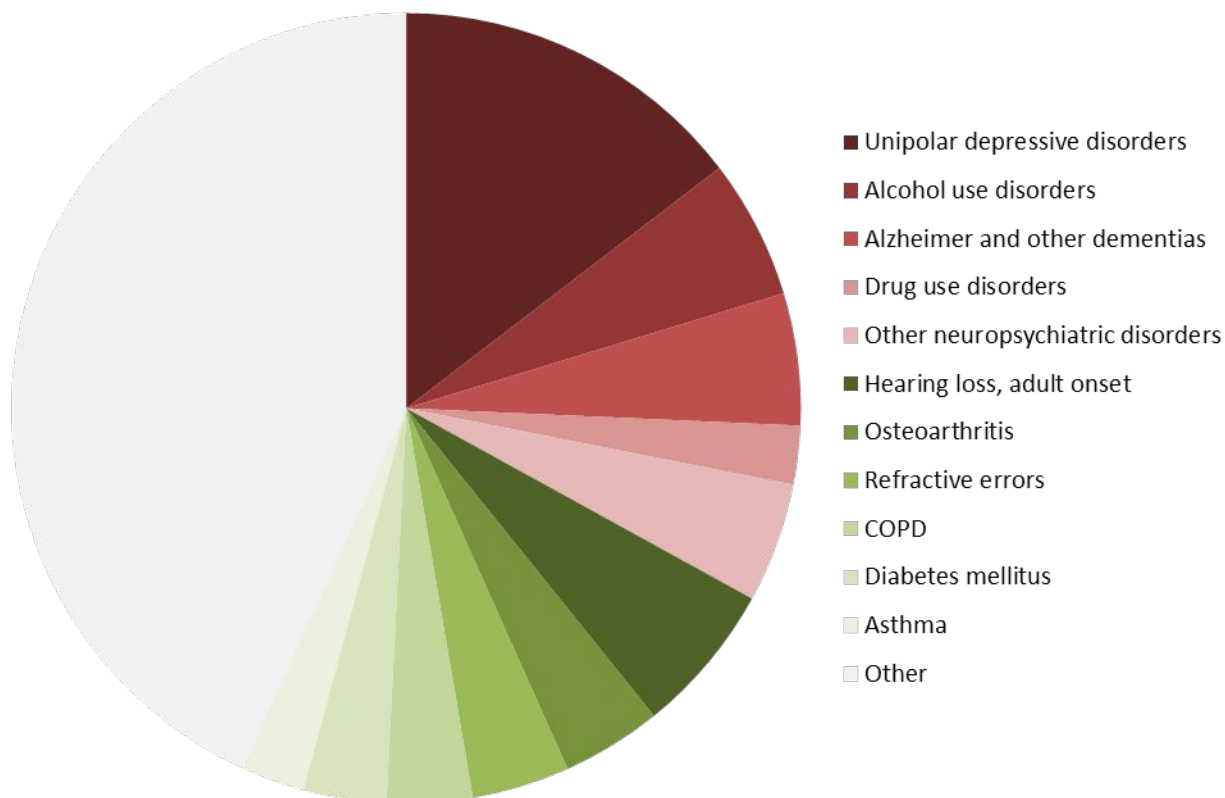
Machine Learning for Clinical Neuroscience

Paul Zhutovsky

Department of Psychiatry
Amsterdam UMC, Location AMC
University of Amsterdam

Disability due to brain disorders

WHO: years lost due to disability



Costs of brain disorders

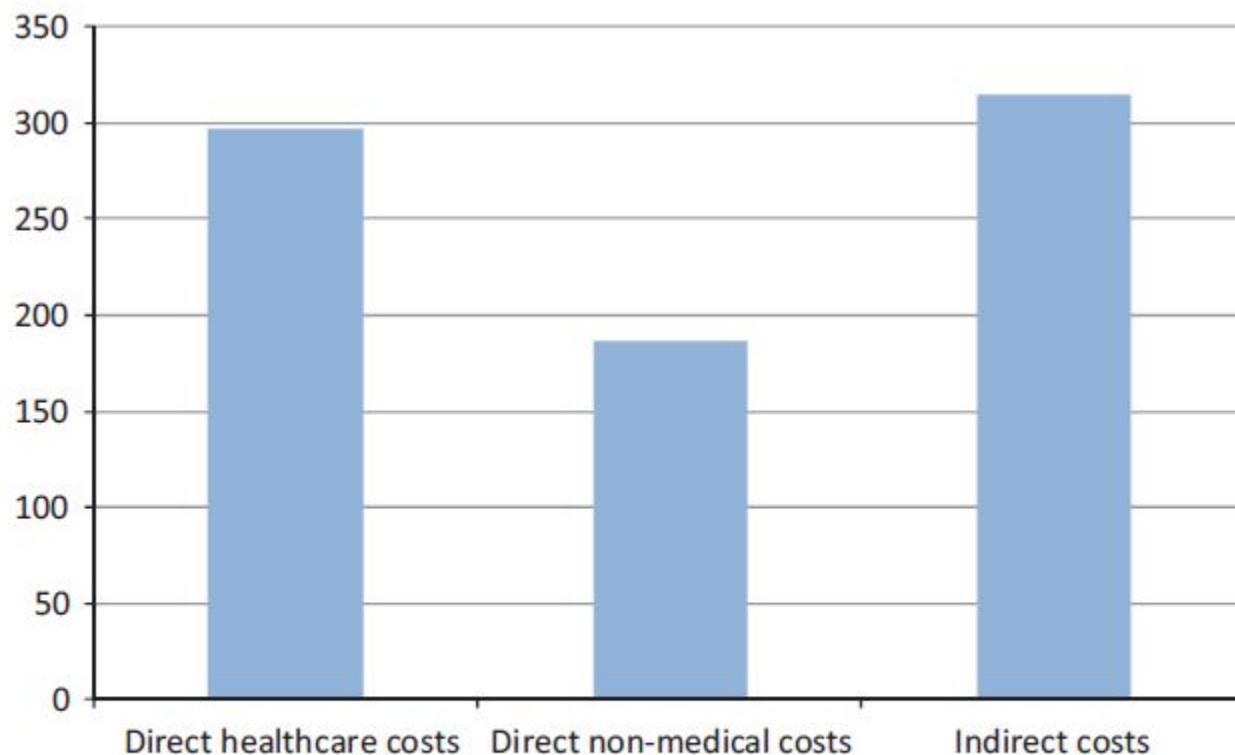


Figure 3 Total cost of brain disorders in Europe (billion €PPP 2010).

- Aggregated cost: € 719 billion per year
- 24% of the total direct health care cost in Europe in 2010

Problem 1: diagnosis

Article

DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses

Darrel A. Regier, M.D., M.P.H.

William E. Narrow, M.D., M.P.H.

Diana E. Clarke, Ph.D., M.Sc.

Helena C. Kraemer, Ph.D.

S. Janet Kuramoto, Ph.D., M.H.S.

Emily A. Kuhl, Ph.D.

David J. Kupfer, M.D.

Objective: The DSM-5 Field Trials were designed to obtain precise (standard error <0.1) estimates of the intraclass kappa as a measure of the degree to which two clinicians could independently agree on the presence or absence of selected DSM-5 diagnoses when the same patient was interviewed on separate occasions, in clinical settings, and evaluated with usual clinical interview methods.

Method: Eleven academic centers in the United States and Canada were selected, and each was assigned several target diagnoses frequently treated in that setting. Consecutive patients visiting a site during the study were screened and stratified on the basis of DSM-IV diagnoses or symptomatic presentations. Patients were randomly assigned to two clinicians for a diagnostic interview; clinicians were blind to any previous diagnosis. All data were entered directly via an Internet-based software system to a secure central server. Detailed research design and statistical

methods are presented in an accompanying article.

Results: There were a total of 15 adult and eight child/adolescent diagnoses for which adequate sample sizes were obtained to report adequately precise estimates of the intraclass kappa. Overall, five diagnoses were in the very good range (kappa=0.60–0.79), nine in the good range (kappa=0.40–0.59), six in the questionable range (kappa=0.20–0.39), and three in the unacceptable range (kappa values <0.20). Eight diagnoses had insufficient sample sizes to generate precise kappa estimates at any site.

Conclusions: Most diagnoses adequately tested had good to very good reliability with these representative clinical populations assessed with usual clinical interview methods. Some diagnoses that were revised to encompass a broader spectrum of symptom expression or had a more dimensional approach tested in the good to very good range.

(*Am J Psychiatry* 2013; 170:59–70)

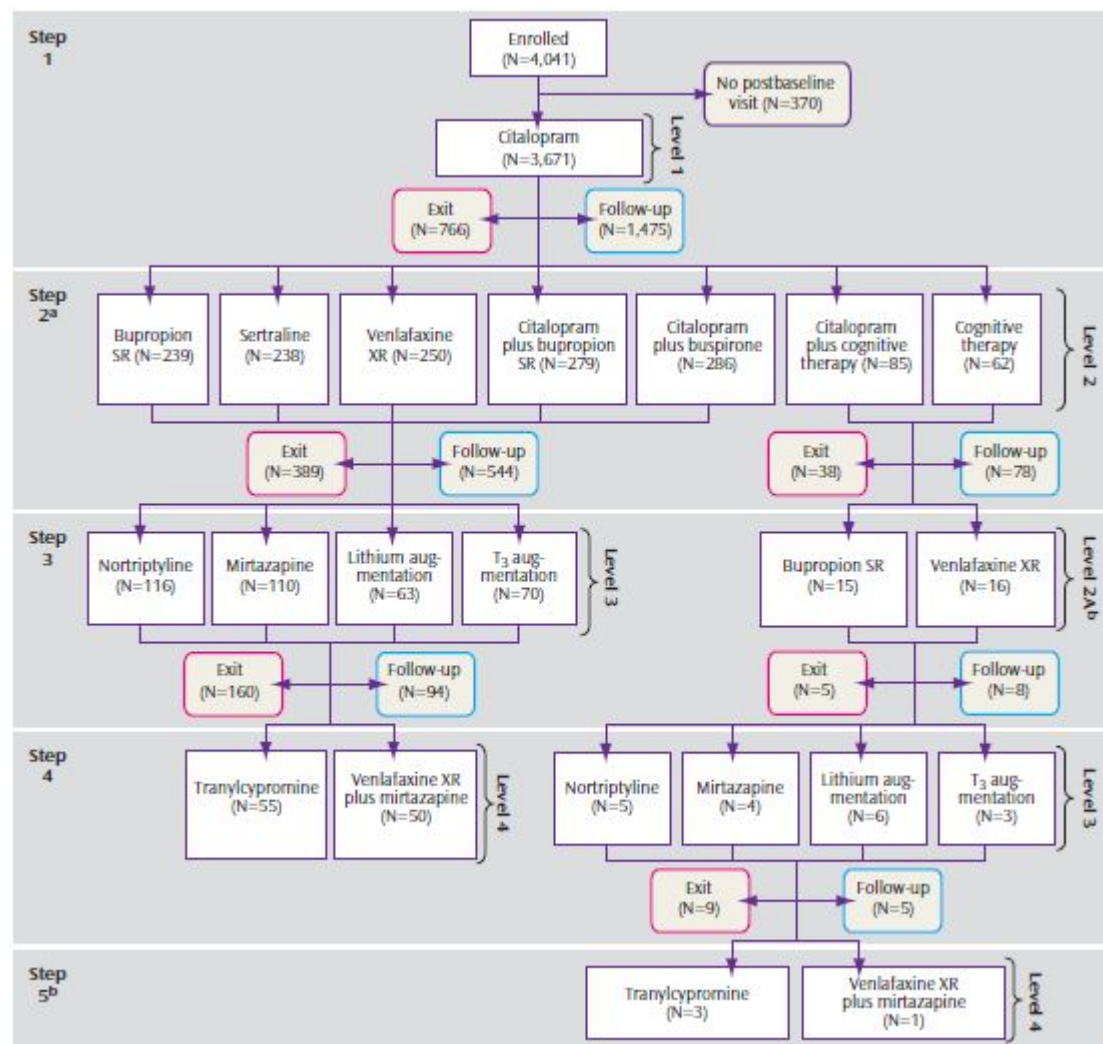
Problem 2: treatment response

37%

31%

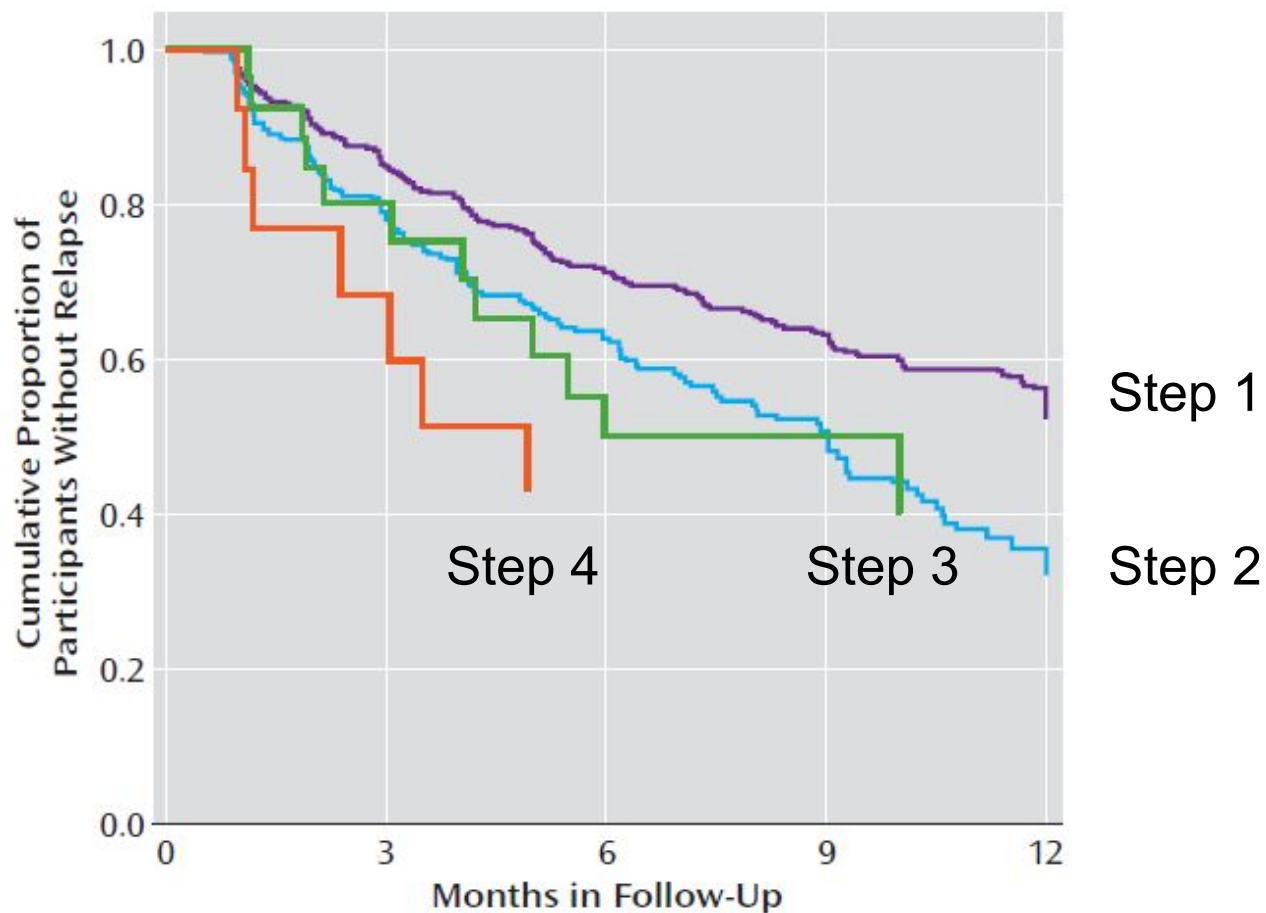
14%

13%

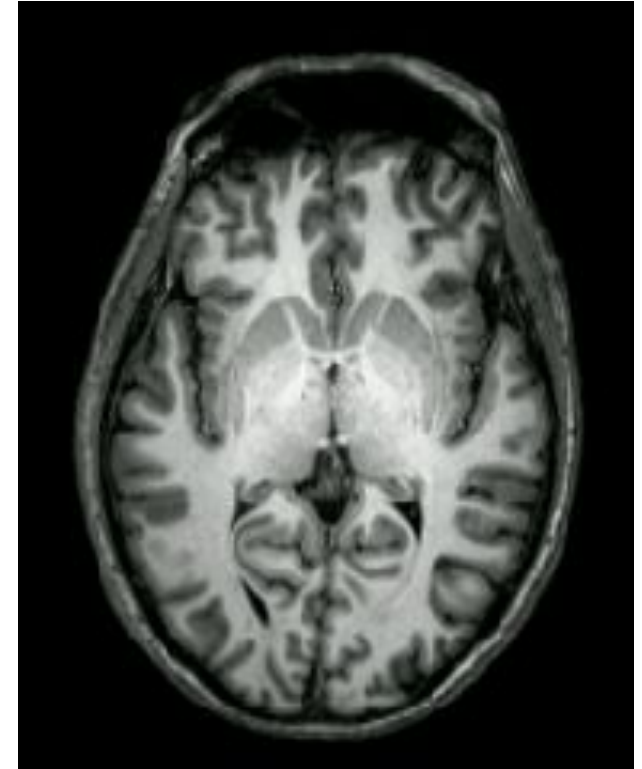
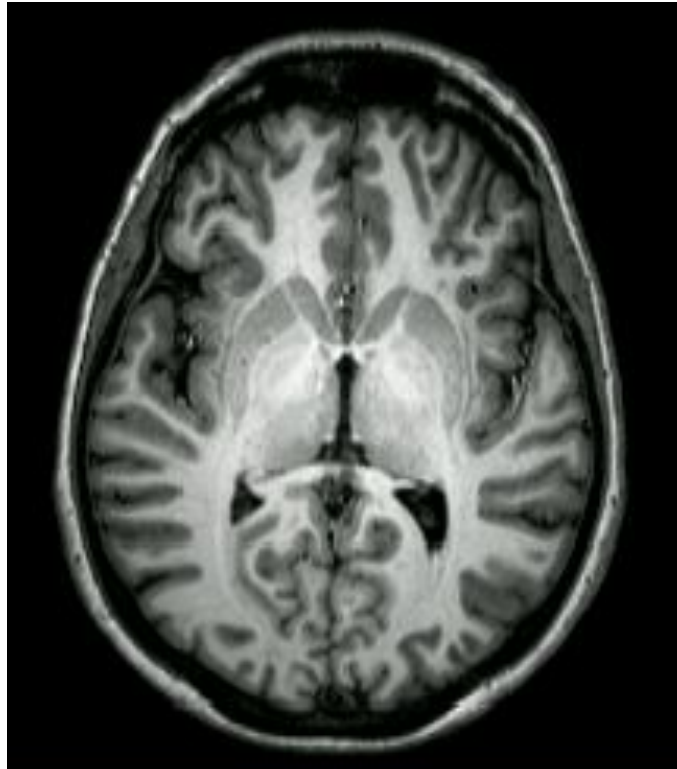


Total 67% of patients recover

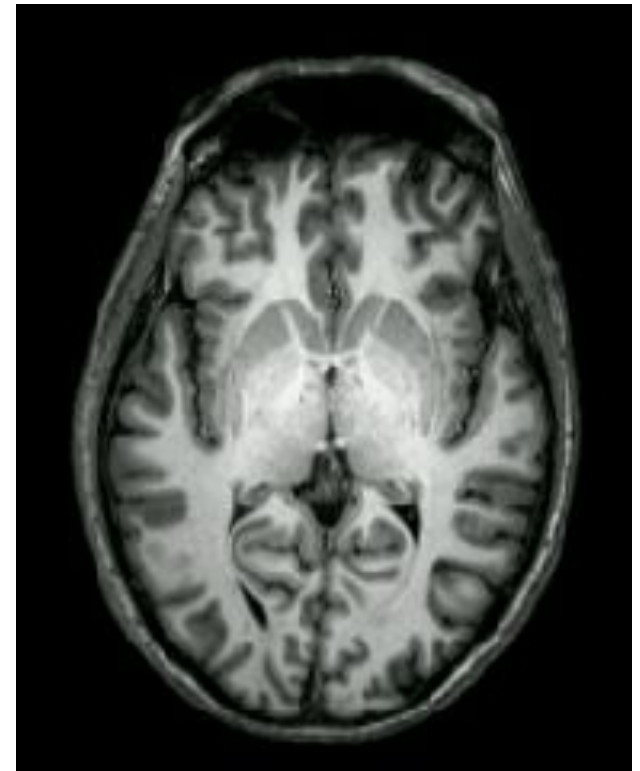
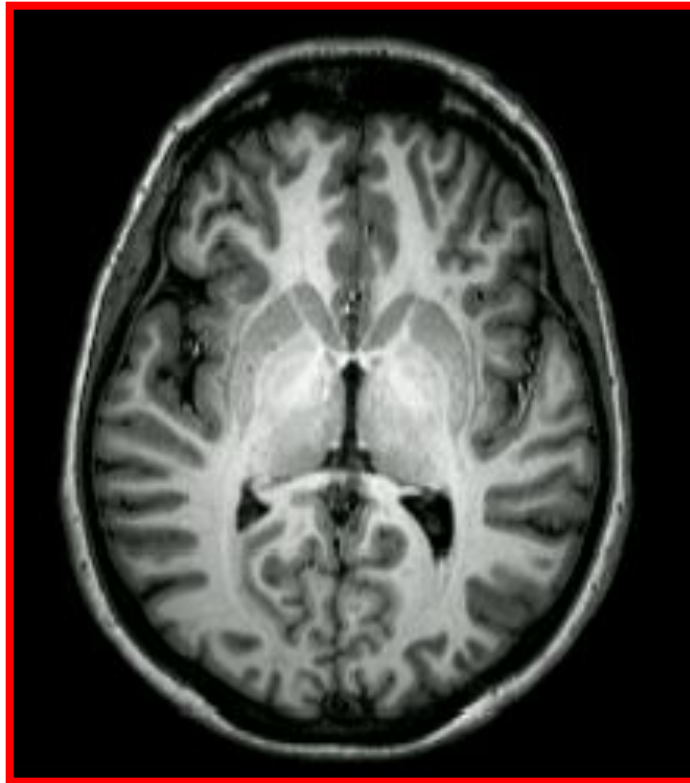
Problem 3: Relapse



Which scan is from a depressed patient?

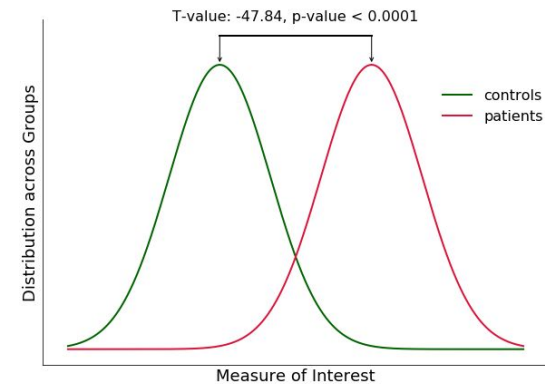
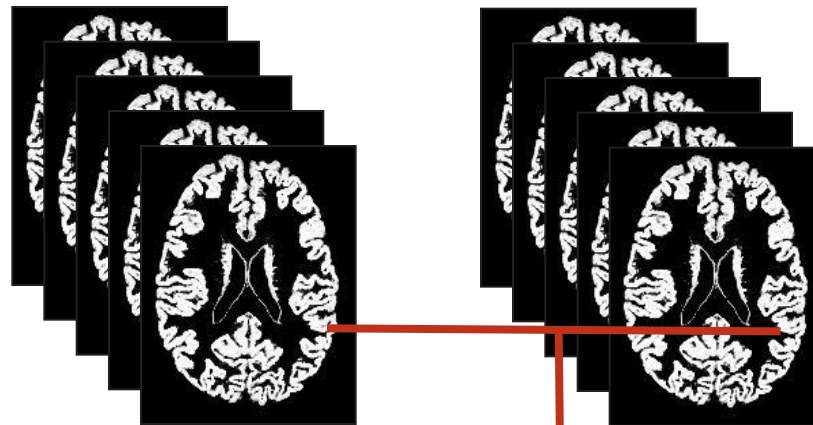


Which scan is from a depressed patient?



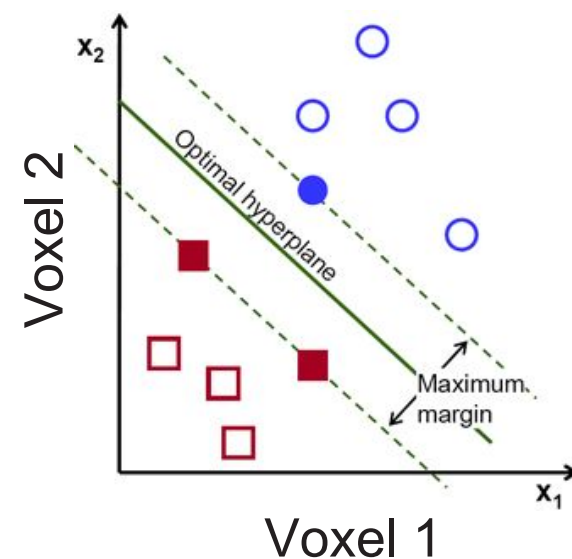
Standard MRI analysis

- Inference on group level
- **Consistency across subjects**
- Mass-univariate analysis
- Good anatomical localization



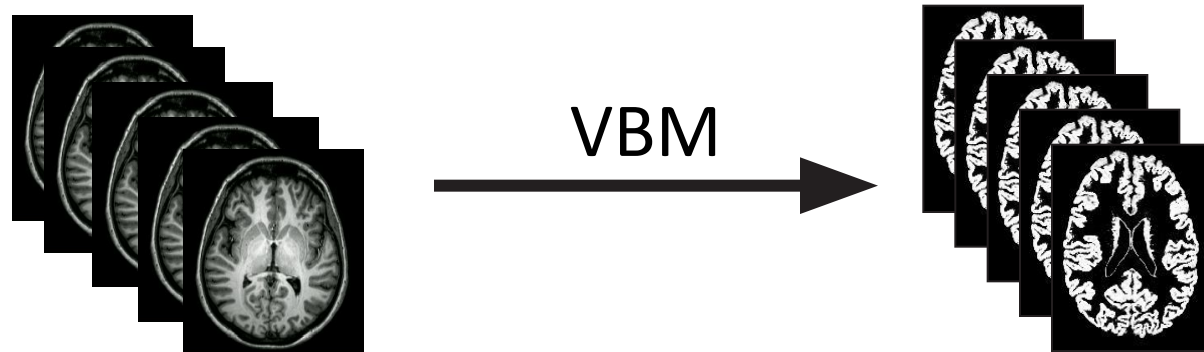
Machine learning analysis

- Inference on individual level
- **Variability across subjects**
- Multivariate pattern analysis
- Poor anatomical localization



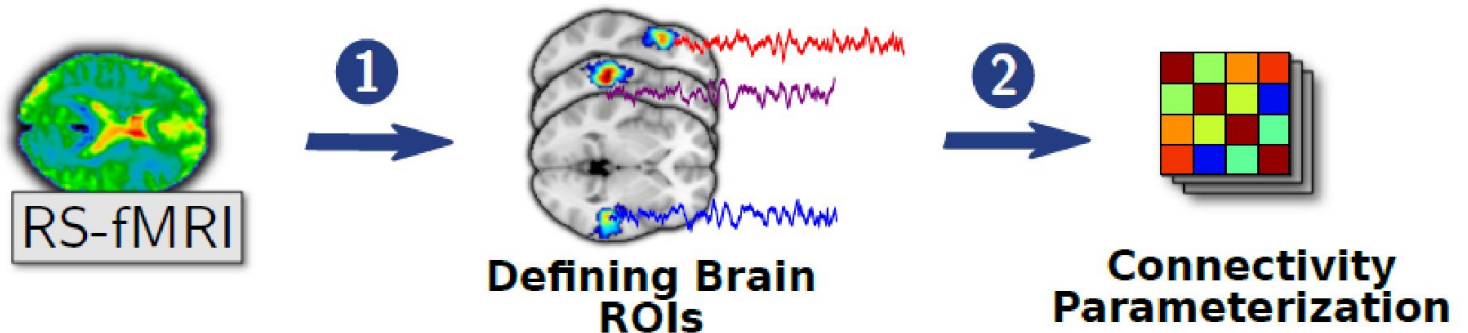
Which data can be used?

1. Structural MRI (Grey Matter Volumes)



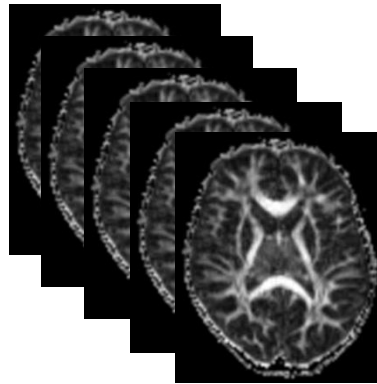
2. Functional MRI

Resting-State:



Which data can be used?

3. Diffusion Tensor Imaging



FA maps

... and many many more

Decision of imaging modality and data representation should (ideally) be based on **prior knowledge!**

Which questions are clinically relevant?

1. Diagnosis

- Simplest case: patient vs. control
- Most interesting case: differential diagnosis

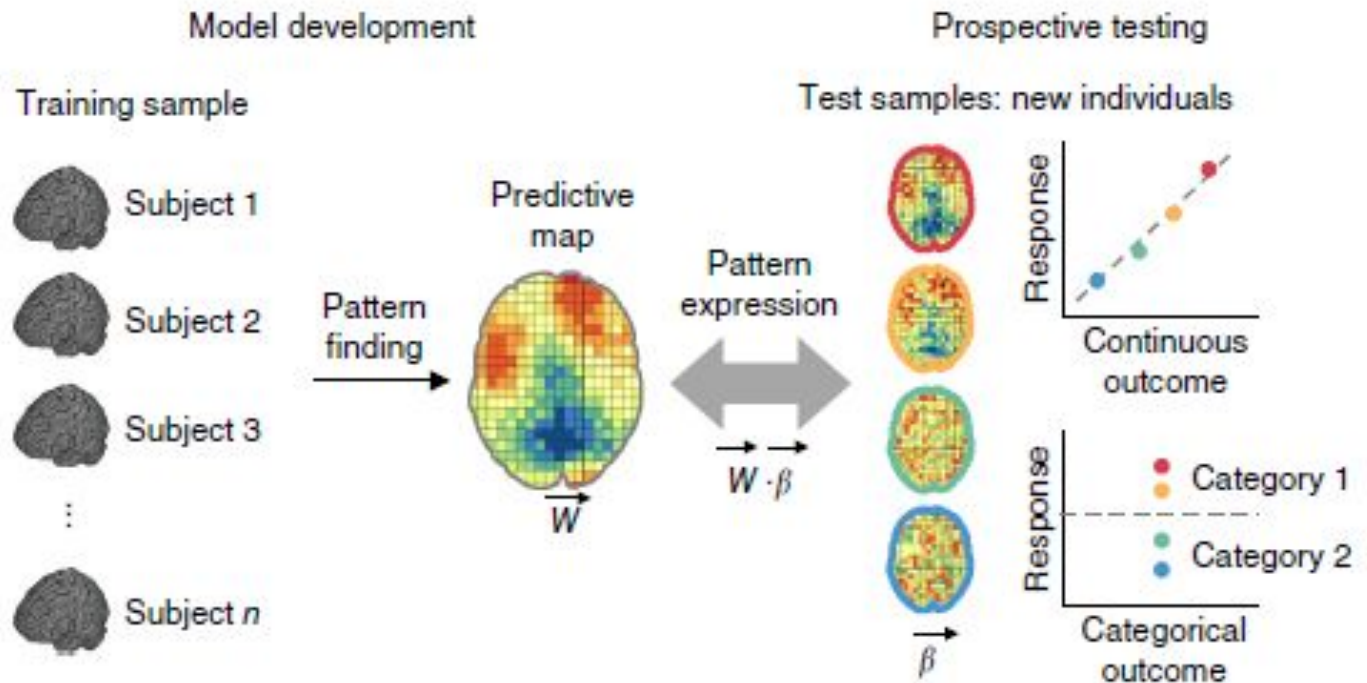
2. Treatment outcome prediction

- Will a patient benefit from a certain treatment?
- Which treatment would be the best for a patient?

3. Prognosis

- Will an at-risk patient develop a certain disorder?
- Will a patient suffer a relapse after treatment?

Summary of the procedure



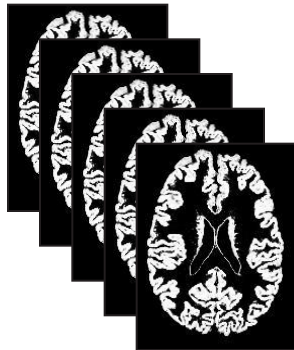
Example application: Prognosis of bvFTD

Patients with
**Late-Onset Behavioral
Changes** (n = 73)

Developed **Psychiatric**
disorders (n = 27)

Developed **bvFTD**
disorders (n = 18)

Developed **Neurological**
disorders (n = 28)



baseline

follow-up
after 2 years

Can we **predict** diagnostic group of our patients after a
2-year follow given baseline **VBM** data?

Prognosis of bvFTD: Which classifications?



Disorder
information
after 2 years



bvFTD vs.
Neurology + Psychiatry

bvFTD vs. Neurology

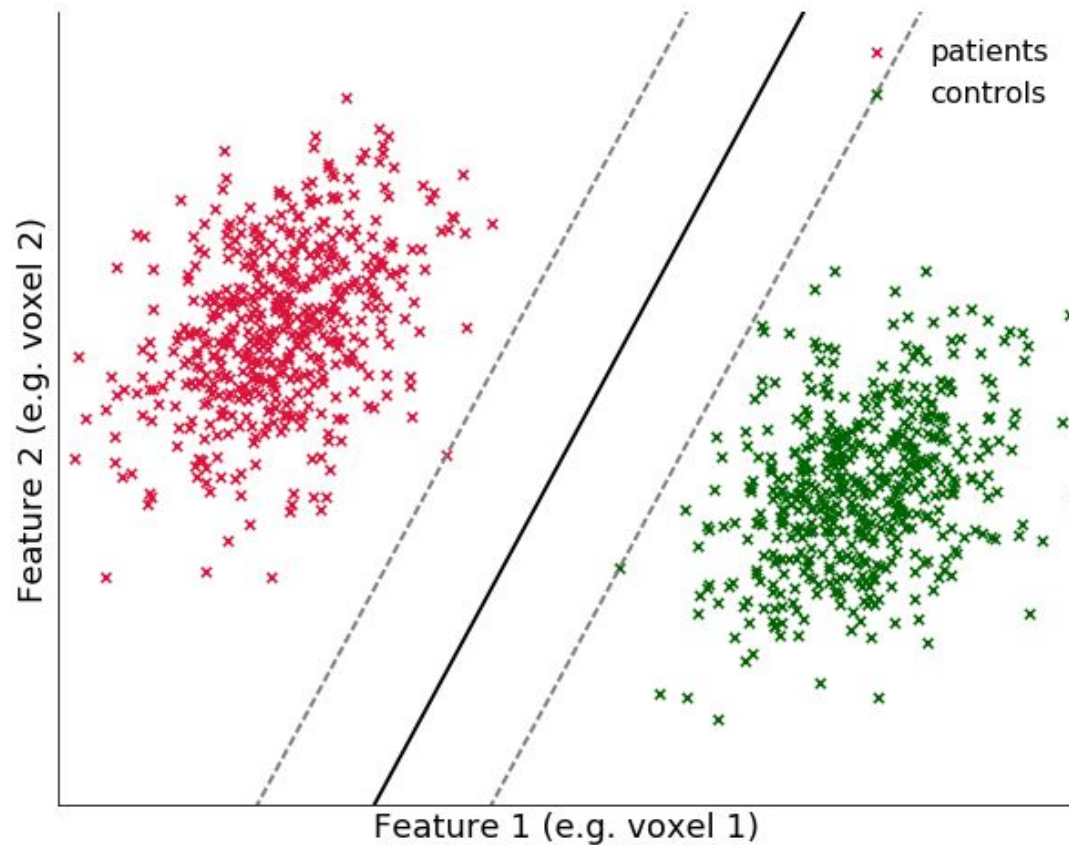
bvFTD vs. Psychiatry

Binary
Classification

bvFTD vs. Neurology vs.
Psychiatry

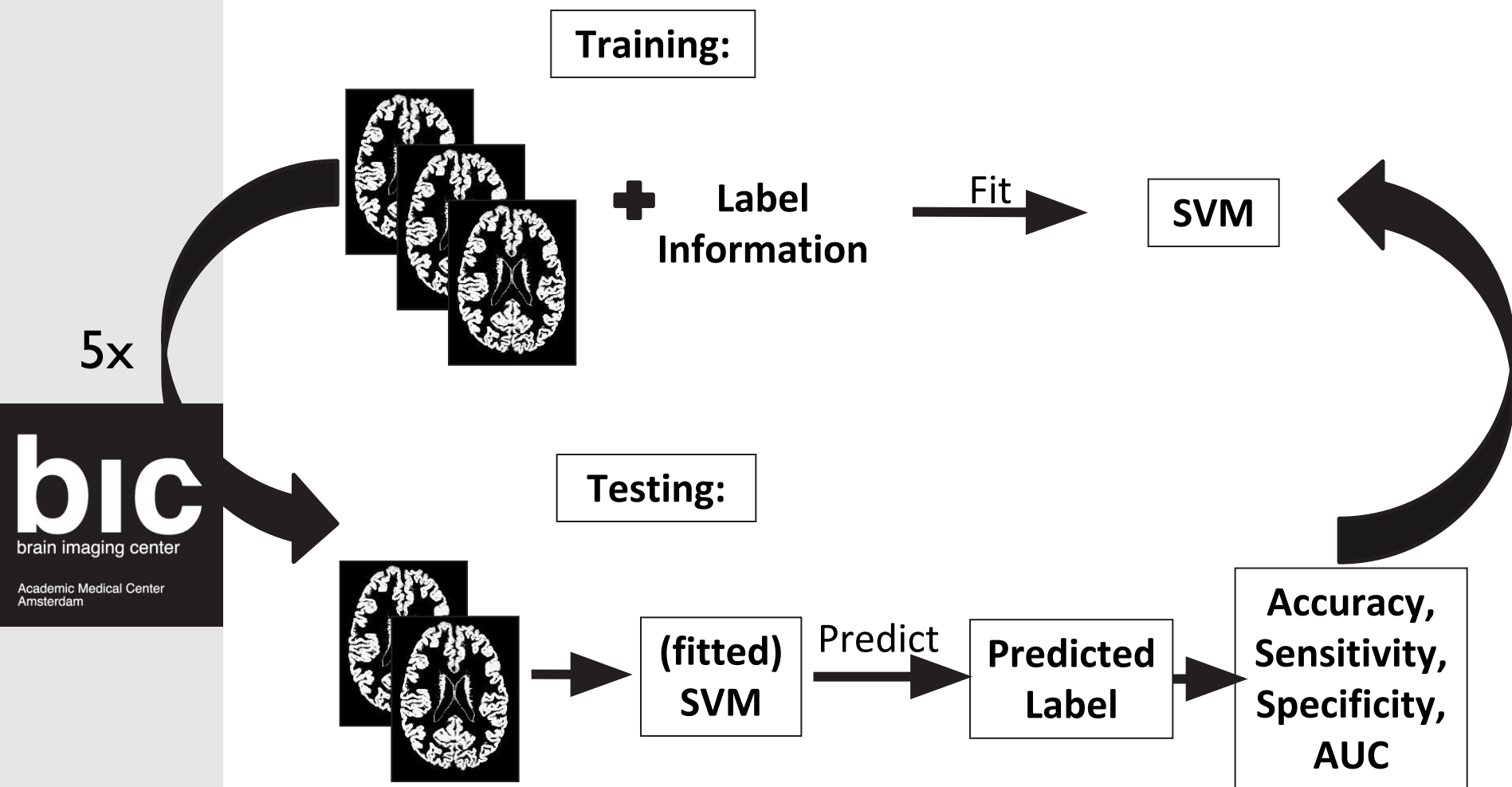
Multiclass
Classification

Prognosis of bvFTD: How to classify?



Linear SVM

Prognosis of bvFTD: How to estimate performance? Cross-validation!



Prognosis of bvFTD: What are those performance metrics?

		Diagnosis after 2 years	
		bvFTD	Other
Predicted Diagnosis	bvFTD	True positive (TP)	False positive, (FP)
	Other	False negative (FN)	True negative (TN)

Accuracy: $(TP + TN) / (TP + FP + FN + TN)$

Sensitivity: $TP / (TP + FN)$ ~ How many patients with **bvFTD** did our algorithm correctly identify?

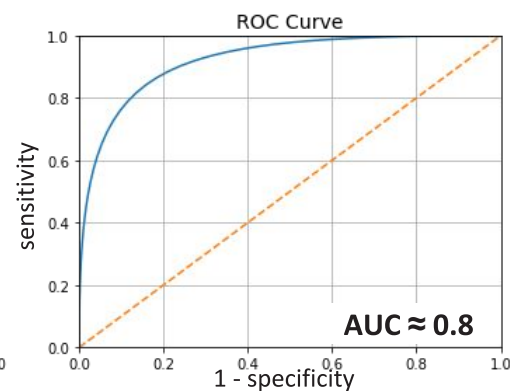
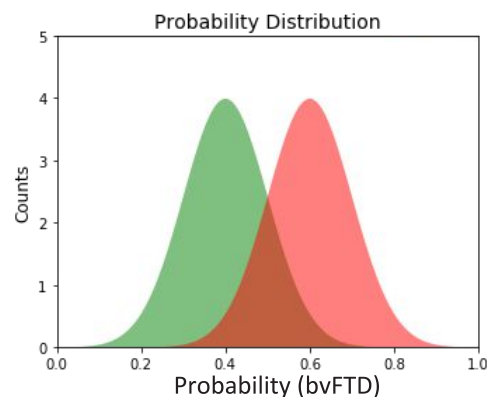
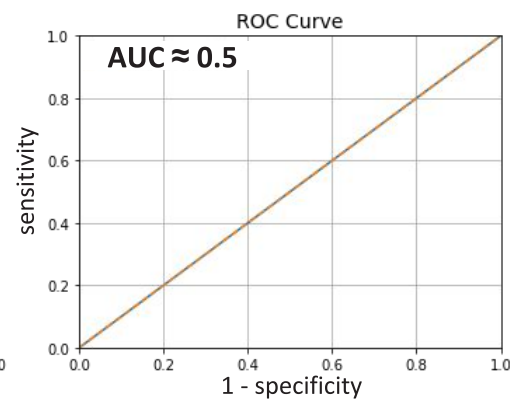
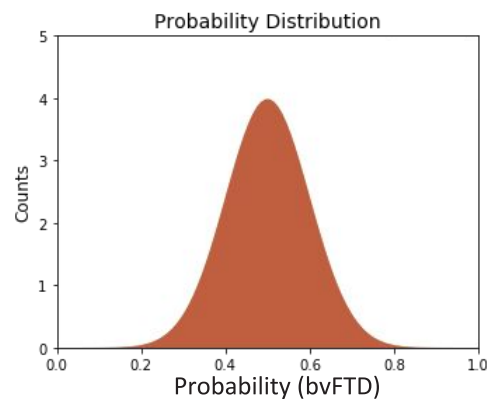
Specificity: $TN / (TN + FP)$ ~ How many of the **other** patients did our algorithm correctly identify?

Balanced Accuracy: $(\text{Sensitivity} + \text{Specificity}) / 2$

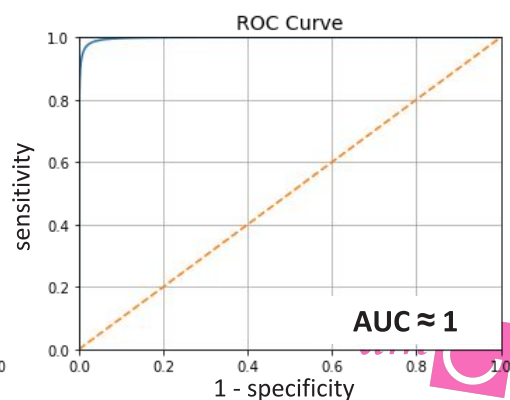
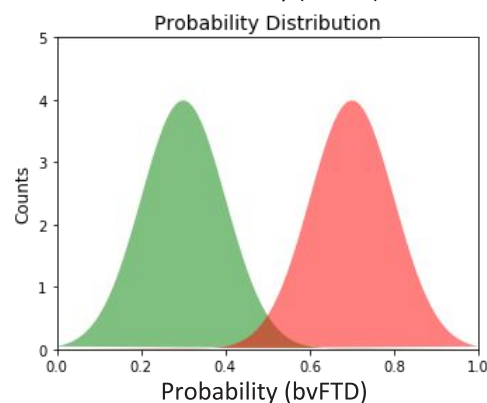
Prognosis of bvFTD: And what is AUC?

Given a **classifier** which provides **probabilistic output**

Can we estimate how well the **predicted classes** are **separated**?

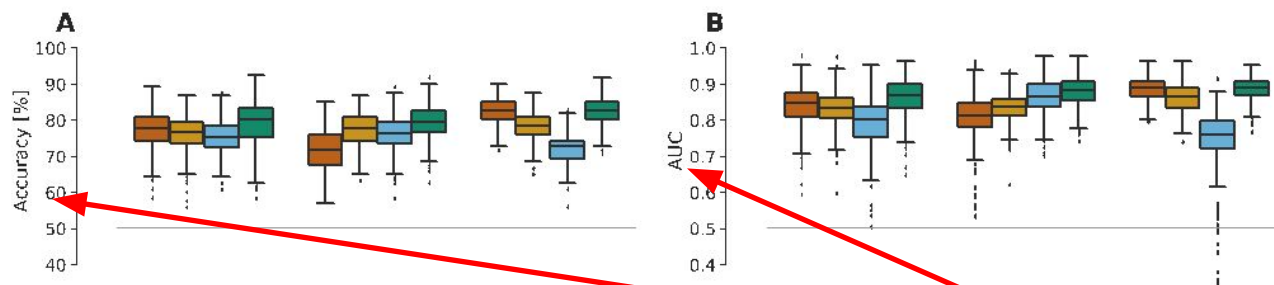


■ bvFTD
■ other



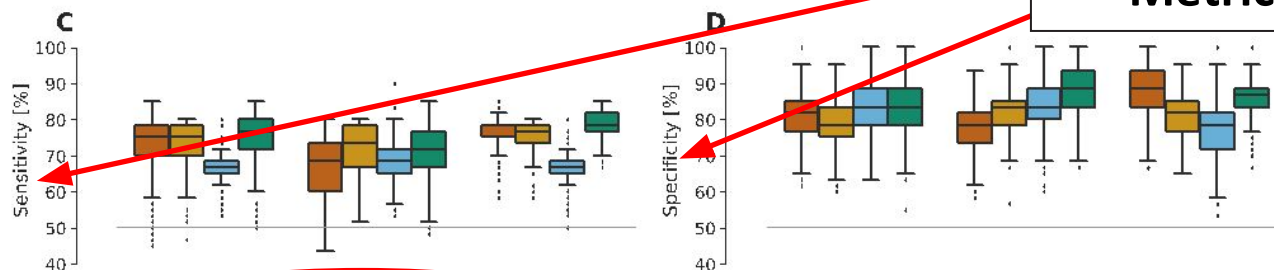
AUC/AUROC:
Area under (ROC) curve

Prognosis of bvFTD: Outcome



Binary Classifications

Performance Metrics



bvFTD vs. Neurology+Psychiatry

bvFTD vs. Neurology

bvFTD vs. Psychiatry

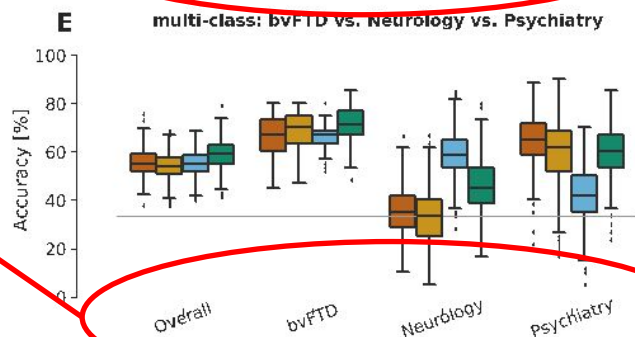
bvFTD vs. Neurology+Psychiatry

bvFTD vs. Neurology

bvFTD vs. Psychiatry

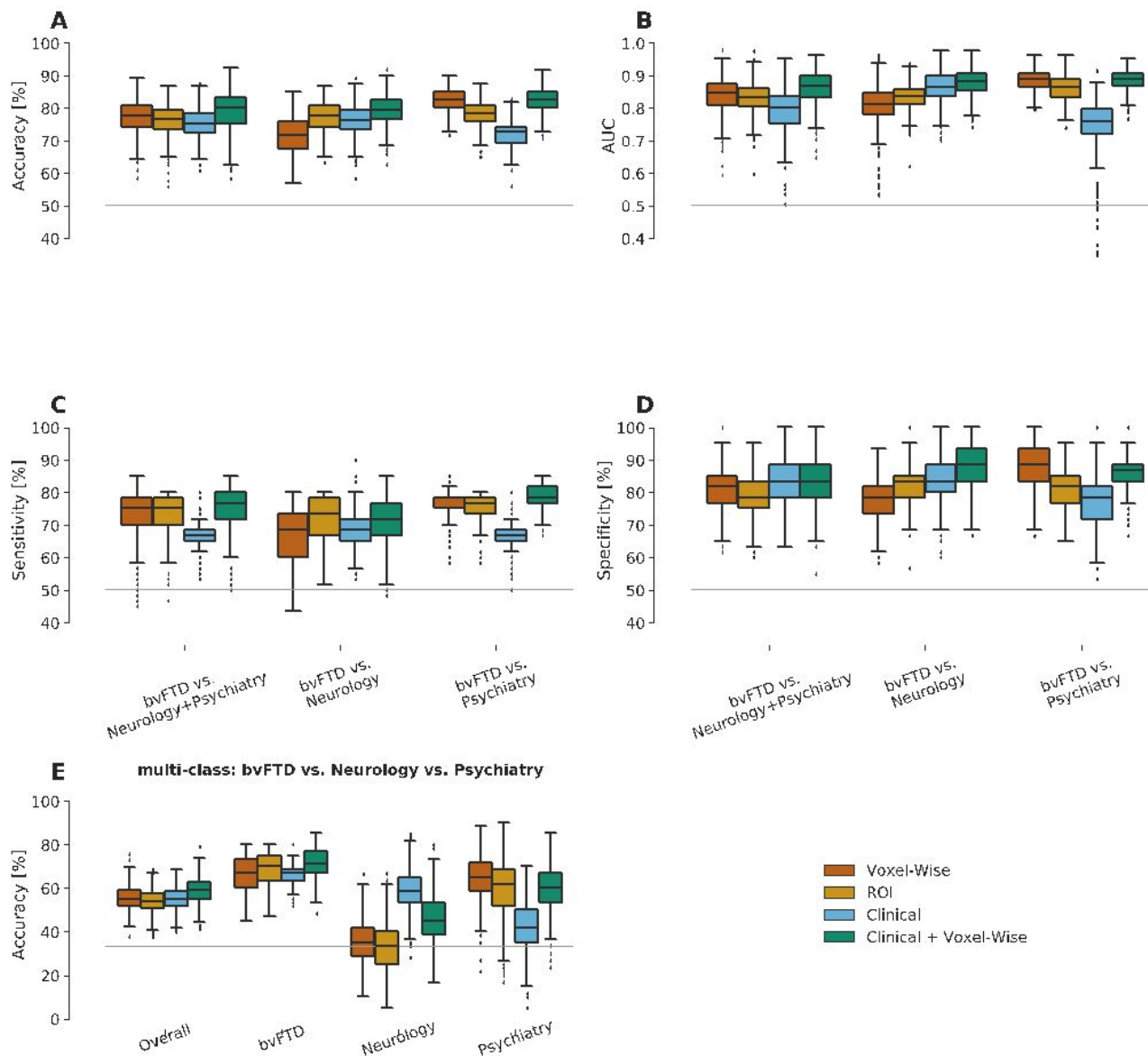
Data Used

Voxel-Wise
 ROI
 Clinical
 Clinical + Voxel-Wise

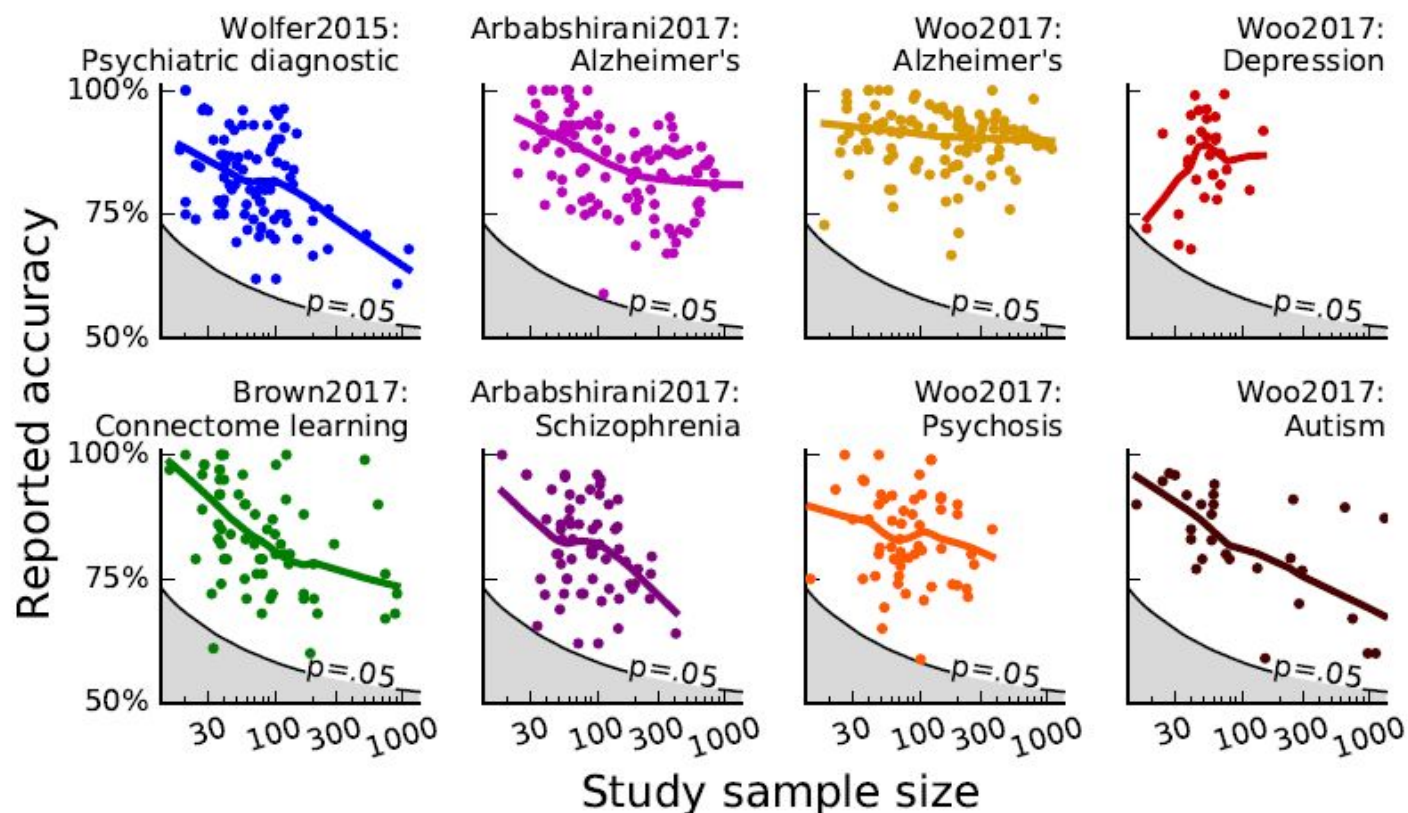


Multiclass Classification

Prognosis of bvFTD: Outcome



Sample size: Caveat



- Large variability in cross-validated measure in small sample size
- Even for studies with 100 participants estimation errors of $\pm 10\%$ can be found!

ML in Clinical Neuroscience: Promises and Caveats

Potential to revolutionize the way we treat patients!

- **Biomarker** development
- **Personalized medicine**

But: many problems yet to solve

- Most studies have data of **very few subjects**
How well can this **generalize**? With how much **heterogeneity** can we deal?
- Most studies are focused on **predicting diagnosis**
(patients vs. controls)
Is that really **clinically relevant**?

Questions?

Thank you for your attention!

