Integrating multimodal data with AI for targeted treatment in depression and hypertension

Dr Raquel Iniesta

Department of Biostatistics and Health Informatics

Institute of Psychiatry Psychology and Neurosciences

King's College London

7th February 2025

KING'S College LONDON

The Fair machine learning lab

Lead: Dr Raquel Iniesta

Reader in Statistical Learning for precision medicine

Biostatistics and Health Informatics Department
Institute of Psychology, Psychiatry and Neurosciences
King's College London

Lei Luo
PhD student

Yiyang Ge
Postdoctoral researcher

Nabila Naeem
Research Assistant

Zareena Haque
MSc student

Luxsana Sritharan
MSc student

Rhys Holland
MSc student

Ciel Burguess
MSc student

Laura Fernández
MSc student

*** Postdoctoral Position**
Hiring soon!

Artificial Intelligence *in healthcare*

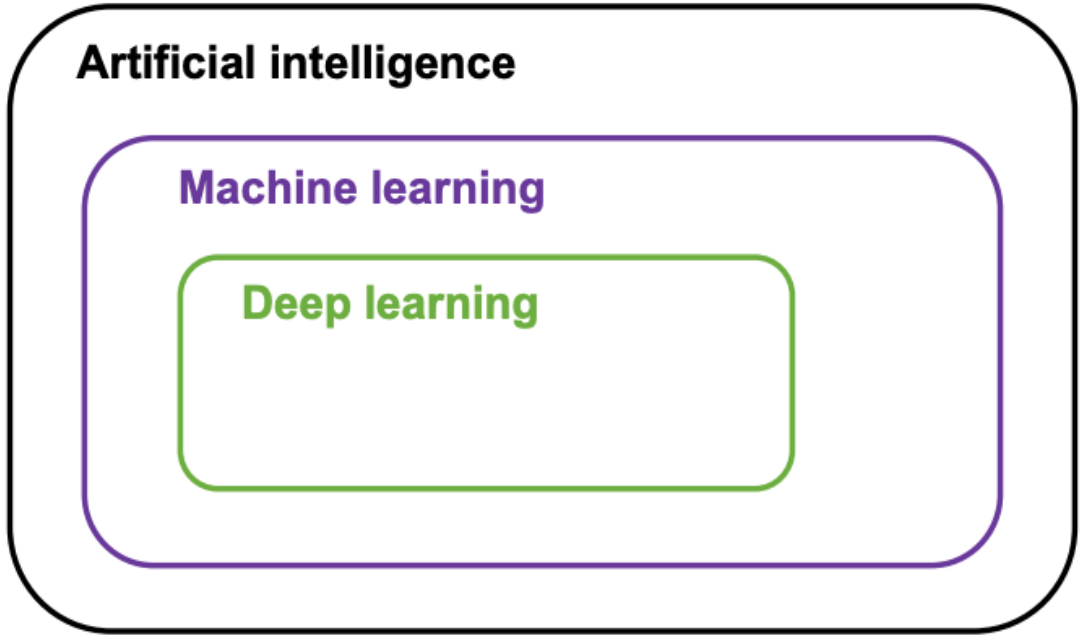
- Healthcare is one of the **most promising** application domains for Artificial Intelligence

- **Visual and sound perception**:
  - AI techniques and their applications can help to detect cancer earlier than before
  - AI systems can potentially surpass human ability as for example in identifying normal and abnormal **chest X-rays**.

- **Decision-making**:
  - AI techniques can make more accurate medical **diagnoses** and **treatment** choices: more challenging, **modest but promising** ability in predicting disease outcomes for complex diseases like depression, hypertension, cancer, or Inflammatory-Bowel-Disease

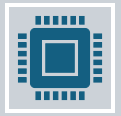
AI, Machine Learning, Deep Learning

- **Artificial intelligence**, machine learning and deep learning are **not synonymous**.
- **Machine learning** is **one approach** to artificial intelligence, but other approaches exist.
- **Deep learning** is a **type of machine learning** which involves neural networks with a large number of hidden layers.



Olivier Colliot. A non-technical introduction to machine learning. Machine Learning for Brain Disorders, Springer, 2023. ffhal-03957125v3f

Machine Learning



Machine learning aims at making a computer **capable** of performing a task without explicitly being programmed for that task.

It means that one will not write a sequence of instructions that will directly perform the considered task. *Example of task*: Predicting the individual response to a particular treatment.

Instead, one will write a program that allows the machine to **learn** how to perform the task by examining data. *Example of data:* Electronic health records, Biological Data, where response to the drug is recorded.

The **output** of this learning process is a computer program itself that performs the desired task (i.e predicting treatment), but this program was not explicitly written. Instead, it has been learned **automatically** by the computer.

Machine Learning *is not* statistical modelling

Traditional modelling culture

1. **Assume** a model «f» for the data. i.e:
   Response = f(predictor,random noise,parameters)
2. **Assess** how good was the model choice:
   goodness-of-fit tests and residual examination
3. Aim is investigating **association**

Traditional ML algorithmic culture

1. Considers the relationship between variables as **unknown**
2. Aims to **build** an algorithmic model based on available data
3. Assess the algorithm by checking how well it performs in new data
4. Aim is investigating **prediction** and **patterns**

Underlying hypothesis on medical AI

"The observed connections in existing data from patients will be reproduced in future patients"

Sub-hypothesis:

"Patients that share similar/dissimilar data on relevant factors will show similar/dissimilar outcomes"

BUT:

- A person is not only made of data: emotions, feelings, personal circumstances (prospect of the **biopsychosocial** model of disease)
- Subjective experience is key for a good response to treatment
- Models **will not capture everything**!

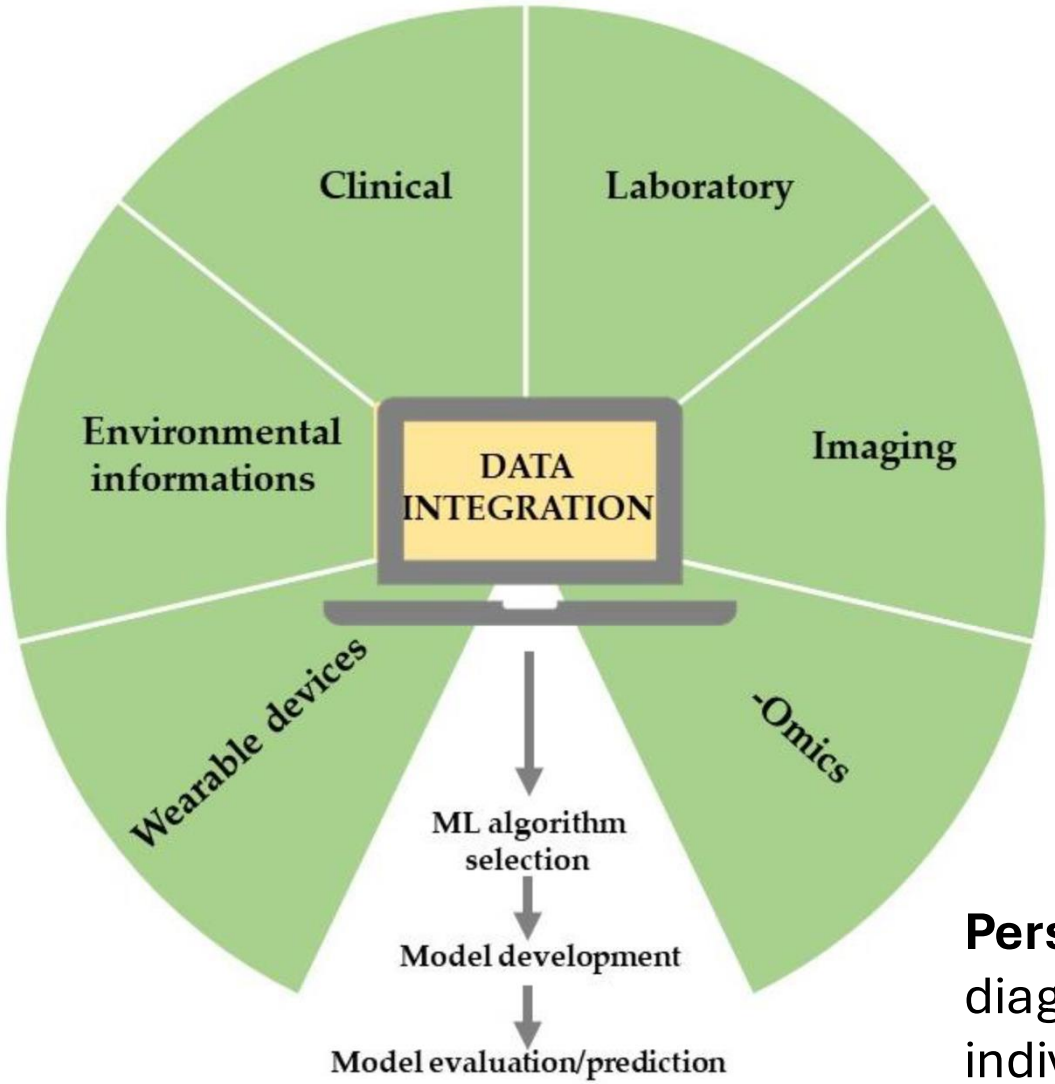
Machine Learning it's all about DATA!

- **Big**: enough sample size
- **Diverse**:
  - The important variables for our investigation are in
  - Enough observations for the relevant variable categories / ranges are included: *ethnicities, gender*
- Ideally, **two datasets**: Training and Testing

Algorithms replicate ***what they saw in data***.
Data should include the relevant information!

Information flow in medical AI

Multimodal Data:
Data captured in multiple different formats

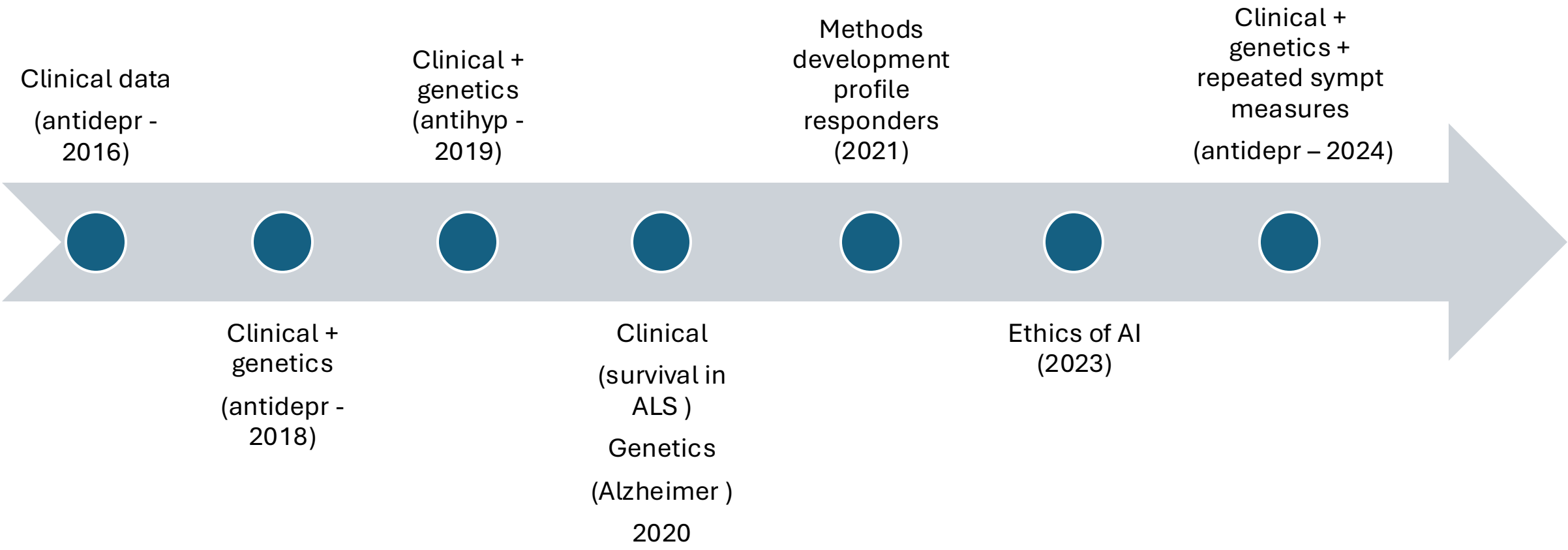


Output information

- **Recommendation for:**
  - Diagnose
  - Treatment choice
- **Alert**
  - Drug adverse reactions

Personalised medicine: targeting diagnosis and treatment choices to individual peoples

For years we have been trying to use AI to predict health outcomes…



Major depression work:
Using machine learning to predict antidepressant treatment response

(1) Machine Learning, Statistical Learning and the Future of Biological Research in Psychiatry

Iniesta, R., Stahl, D. R. & McGuffin, P., Sept 2016, In: **Psychological medicine**. 46, 12, p. 2455-2465 11 p.

(2) Combining clinical variables to optimize prediction of antidepressant treatment outcomes

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., Stahl, D., Dobson, R., Aitchison, K. J., Farmer, A., Lewis, C. M., McGuffin, P. & Uher, R., Jul 2016, In: **Journal of psychiatric research**. 78, p. 94-102 9

(3) Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables

Iniesta, R., Hodgson, K., Stahl, D. R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovšek, M. Z., Souery, D., Dobson, R., Aitchison , K. J., Farmer, A., McGuffin, P., Lewis, C. M. & Uher, R., 3 Apr 2018, In: **Scientific Reports**. 8, 1, 9 p., 5530 .

(4) Optimizing the Prediction of Depression Remission: A Longitudinal Machine Learning Approach

Carr, E., Rietschel, M., Mors, O., Henigsberg, N., Aitchison, K., Maier, W., Uher, R., Farmer, A., McGuffin, P. & Iniesta, R., October 2024, In: **American Journal of Medical Genetics**, Part B: Neuropsychiatric Genetics. 29 October 2024 https://doi.org/10.1002/ajmg.b.33014

Major depressive disorder…

…is a common condition, responsible for a substantial proportion of disability world-wide.

Antidepressant treatment is pharmacological and psychological

Outcomes are unsatisfactory

Some individuals experience dramatic improvements, most **do not benefit** sufficiently

Trial and error → Prolonged disability

Multiple treatment trials → Increased risk of suicide

Diagnosis of depression not sufficient for **treatment selection**

Previous studies analysing **single** clinical and genetic predictors of antidepressant outcome, showed:

- **Multifactorial and polygenic** architecture
- Predictors with **small effect sizes and low predictive ability**

- **Unlikely that a single piece of information** can predict antidepressant treatment outcome **with clinical significance.**
- Need of studying **combinations of variables** rather than single predictors.

Research hypothesis

A combination of baseline demographic, clinical variables, genetic and longitudinal measures of depression severity can predict response to antidepressant treatment at the individual level.

Data from the GENDEP study
(Genome-Based Therapeutic Drugs for Depression)

STUDY

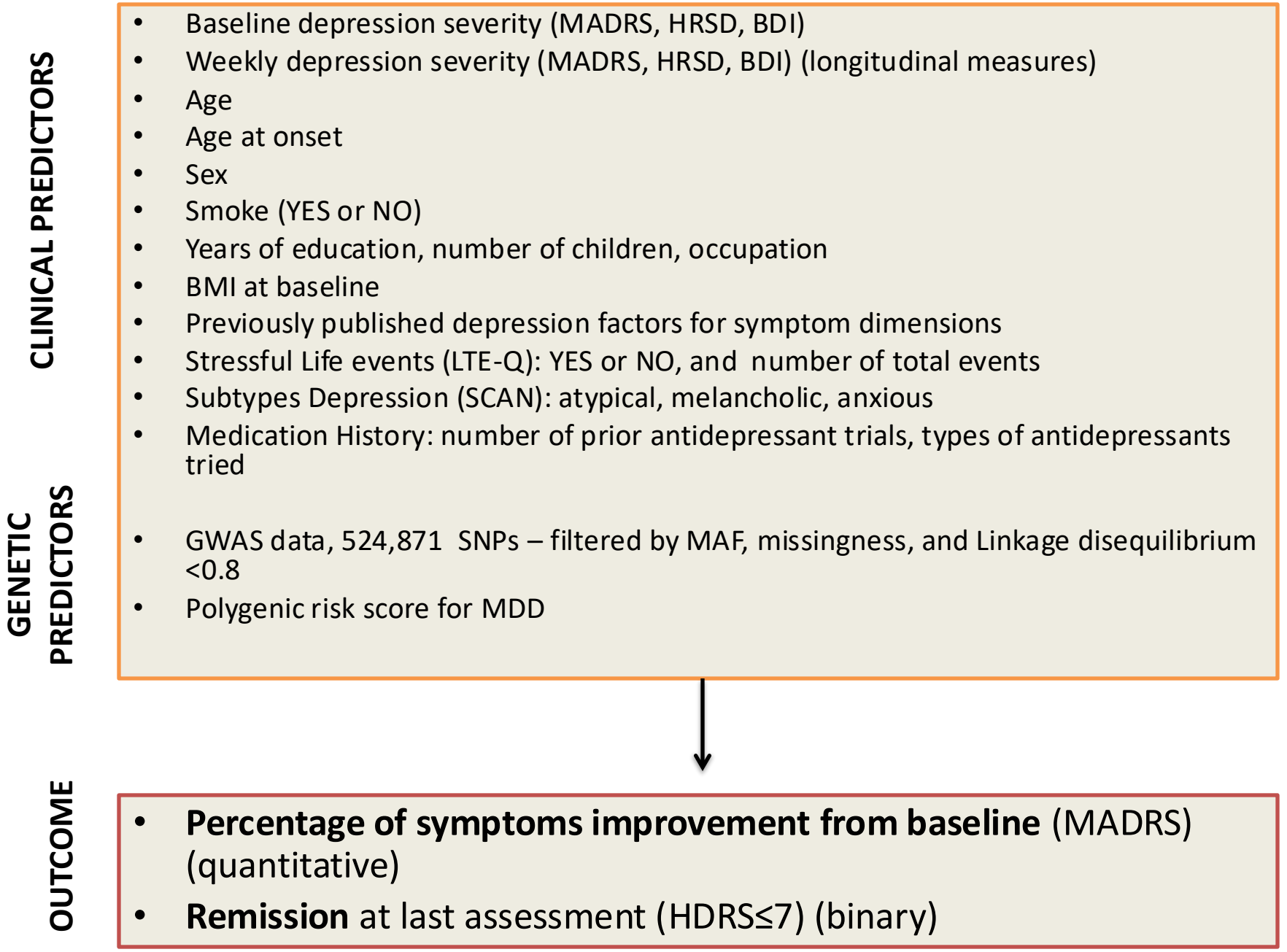
- **Pharmacogenetic** study

SAMPLE

- A total of **714 patients** with **unipolar depression** of at least moderate severity according to ICD-10 or DSM-IV criteria
- Patients were aged **19–72** years. **Caucasian** European parentage.

DESIGN

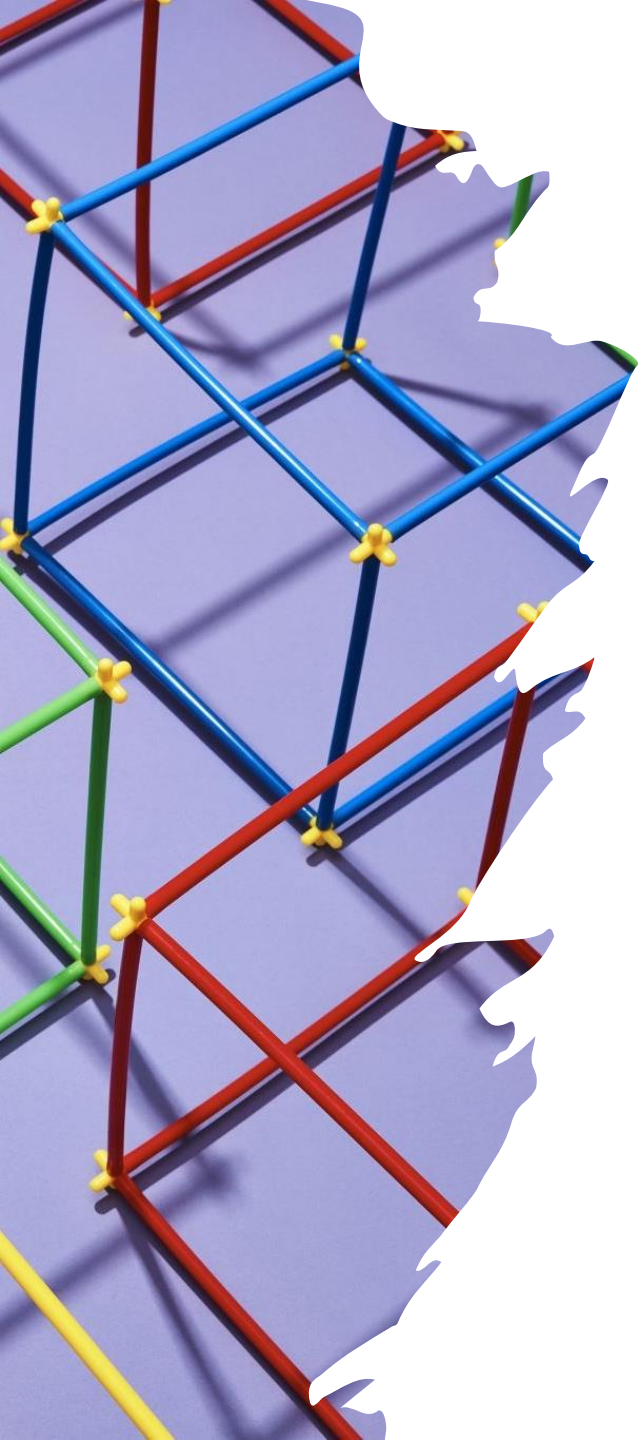
- Randomized to **two active treatment arms**: escitalopram or nortriptyline
- **12-week** follow-up
- **3 main scales of depression severity** (MADRS,HRSD,BDI) assessed weekly


CLINICAL PREDICTORS
GENETIC PREDICTORS
OUTCOME
• Baseline depression severity (MADRS, HRSD, BDI)
• Weekly depression severity (MADRS, HRSD, BDI) (longitudinal measures)
• Age
• Age at onset
• Sex
• Smoke (YES or NO)
• Years of education, number of children, occupation
• BMI at baseline
• Previously published depression factors for symptom dimensions
• Stressful Life events (LTE-Q): YES or NO, and number of total events
• Subtypes Depression (SCAN): atypical, melancholic, anxious
• Medication History: number of prior antidepressant trials, types of antidepressants tried
• GWAS data, 524,871 SNPs – filtered by MAF, missingness, and Linkage disequilibrium <0.8
• Polygenic risk score for MDD
• Percentage of symptoms improvement from baseline (MADRS) (quantitative)
• Remission at last assessment (HDRS≤7) (binary)

Methodology of analysis?

- Total of 525,015 variables for 430 individuals (p>>N) Big data set!
- Traditional statistical methods are efficient to analyse databases with a **small numbe**r of variables.
- When used in high-dimensional data sets, they commonly present problems of **multiple testin**g and **overfitting**
- **Big data sets** absolutely **common** in psychiatric research…

The adoption of **new methods of analysis** is urgently needed!

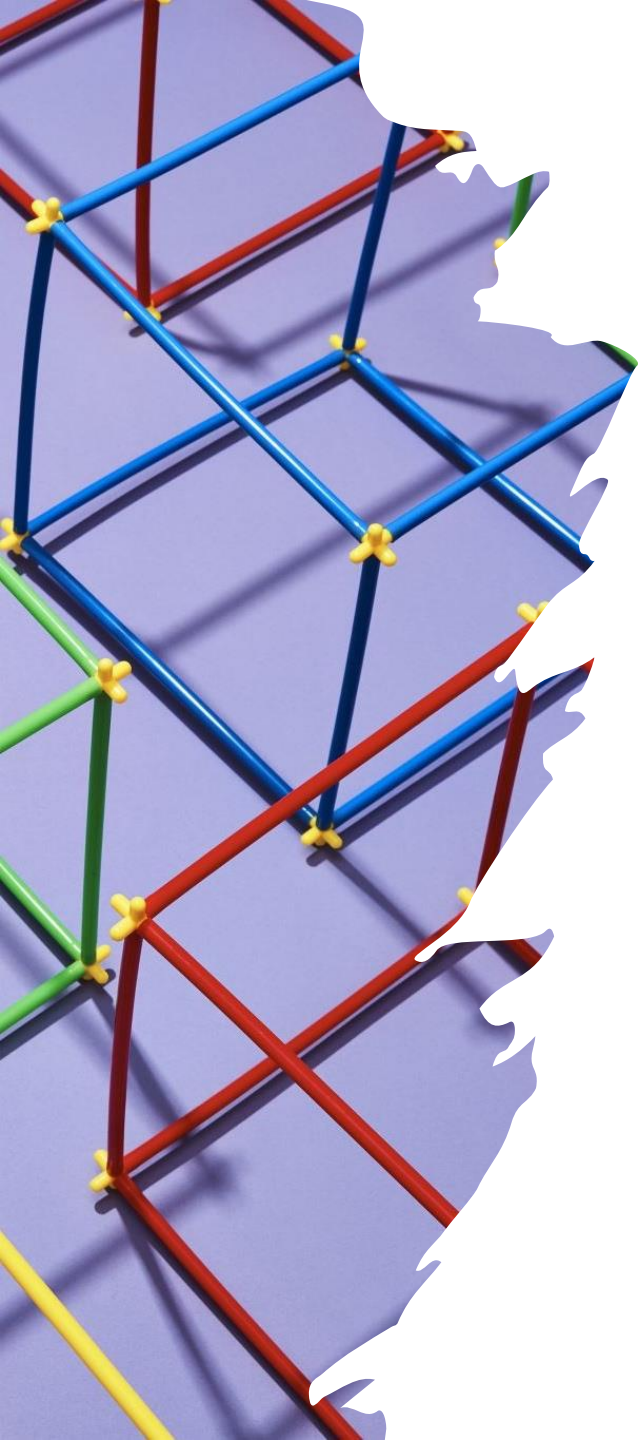


Classical Methods of Machine Learning for medical AI

- *k-Nearest Neighbour Algorithms*
- *Regression-based algorithms: Lasso, Ridge, Elastic-net*
- *Naive Bayes Classifier Algorithms*
- *Support Vector Machines (SVMs)*
- *Tree based algorithms: Random Forest, XGBoost*

Advanced:

- *Neural Networks and Deep Learning*
- *Generative AI*



Classical Methods of Machine Learning for medical AI

- *k-Nearest Neighbour Algorithms*
- ***Regression-based algorithms: Lasso, Ridge, Elastic-net***
- *Naive Bayes Classifier Algorithms*
- ***Support Vector Machines (SVMs)***
- ***Tree based algorithms: Random Forest, XGBoost***

Advanced:

- *Neural Networks and Deep Learning*
- *Generative AI*

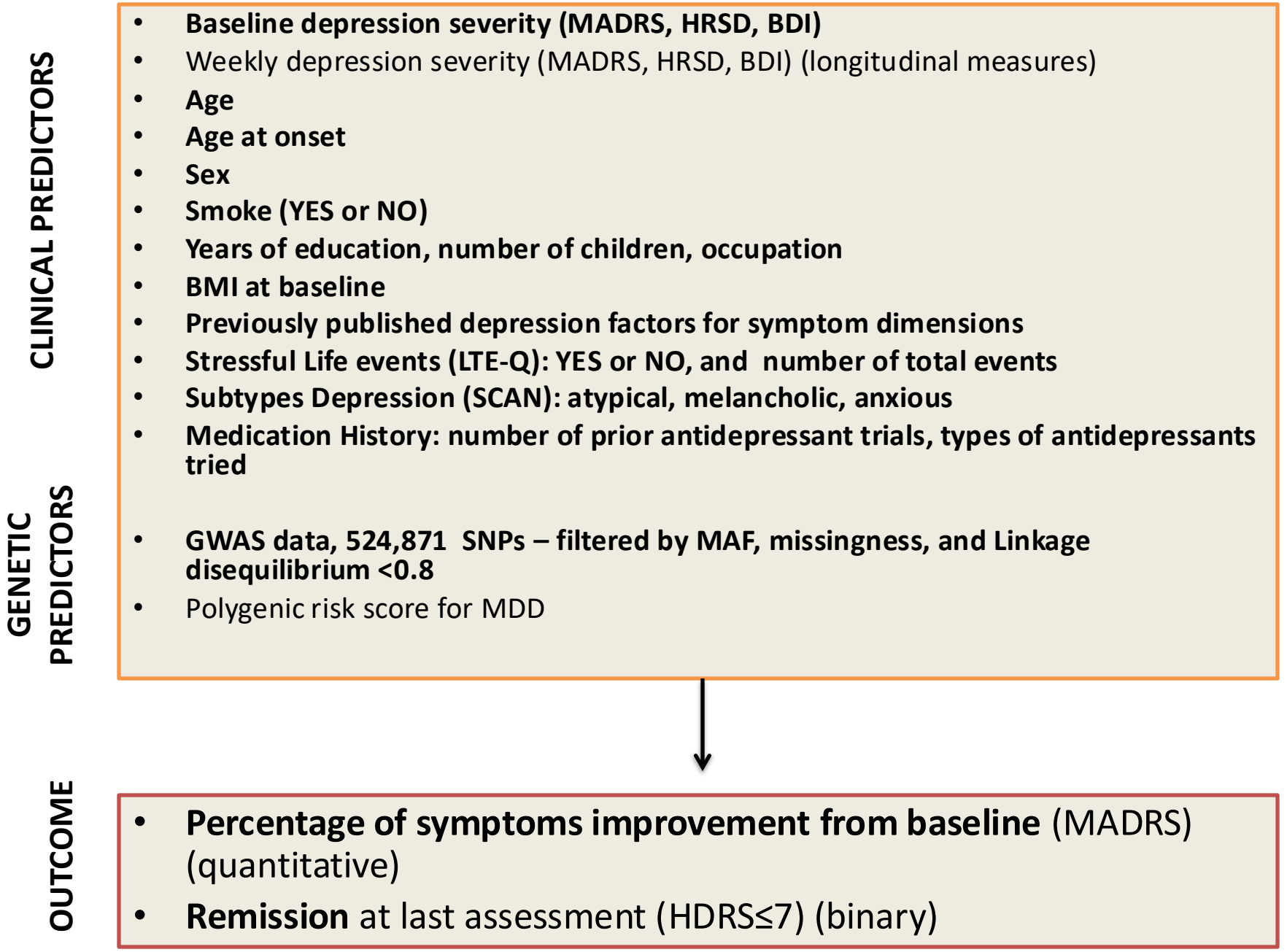
Elastic net are modified regression model

$$logit\,(\pi) = \beta_0 + \beta\, x \qquad \text{where} \quad x = predictors$$

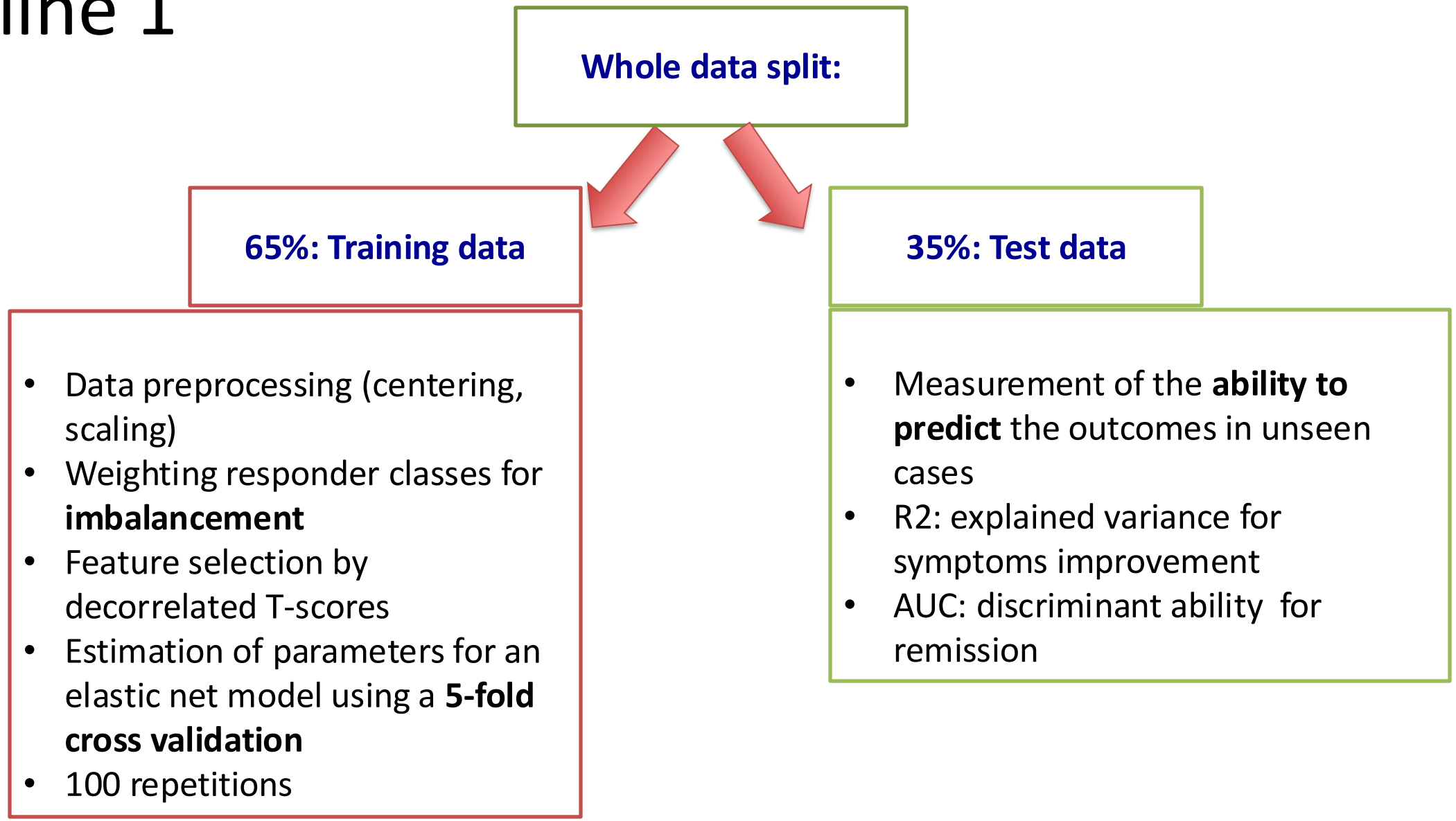
The usual log-likelihood to maximise is **penalised by adding** λ times:

$$P_\alpha(\beta) = \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha \mid \beta_j \mid \right]$$

- Just the most **meaningful predictors are retained** in the model
- Allows **high-dimensional data**: more variables on the models than observations (p>>n)
- **Prevents of overfitting:** when models learn relationship in a sample that do not extrapolate to the population.
- Allowing **correlated variables** to be included or excluded together from the model

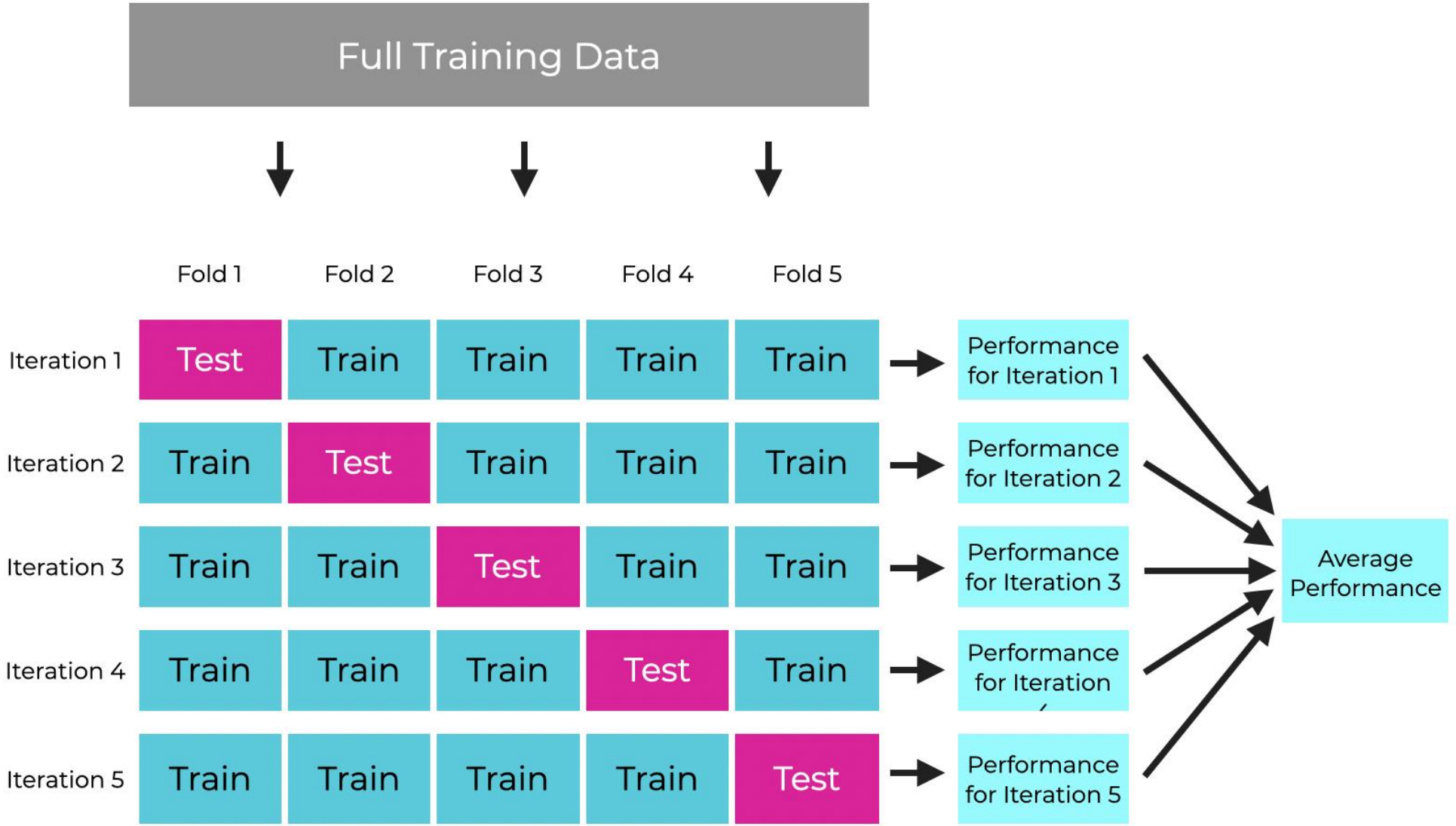

CLINICAL PREDICTORS
GENETIC PREDICTORS
Baseline depression severity (MADRS, HRSD, BDI)
Weekly depression severity (MADRS, HRSD, BDI) (longitudinal measures)
Age
Age at onset
Sex
Smoke (YES or NO)
Years of education, number of children, occupation
BMI at baseline
Previously published depression factors for symptom dimensions
Stressful Life events (LTE-Q): YES or NO, and number of total events
Subtypes Depression (SCAN): atypical, melancholic, anxious
Medication History: number of prior antidepressant trials, types of antidepressants tried
GWAS data, 524,871 SNPs – filtered by MAF, missingness, and Linkage disequilibrium <0.8
Polygenic risk score for MDD
OUTCOME
Percentage of symptoms improvement from baseline (MADRS) (quantitative)
Remission at last assessment (HDRS≤7) (binary)

Pipeline 1

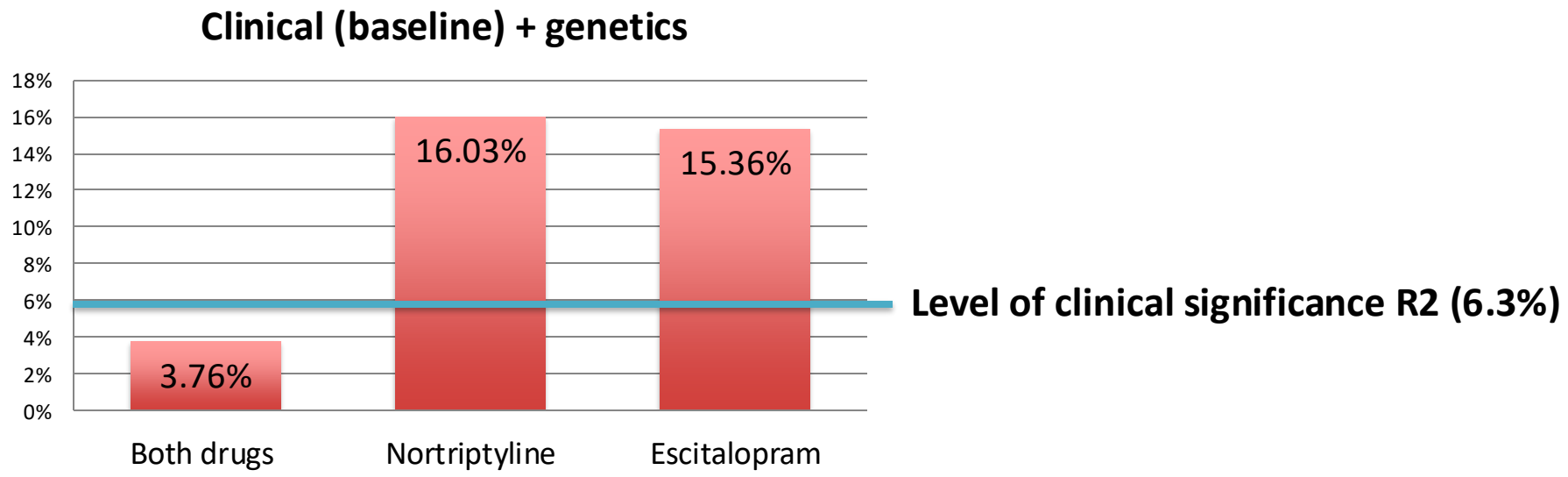


We built models in 3 samples: whole sample, and every drug group.
We predicted two outcomes (percentage improvement and remission) in every sample -> **6 models**

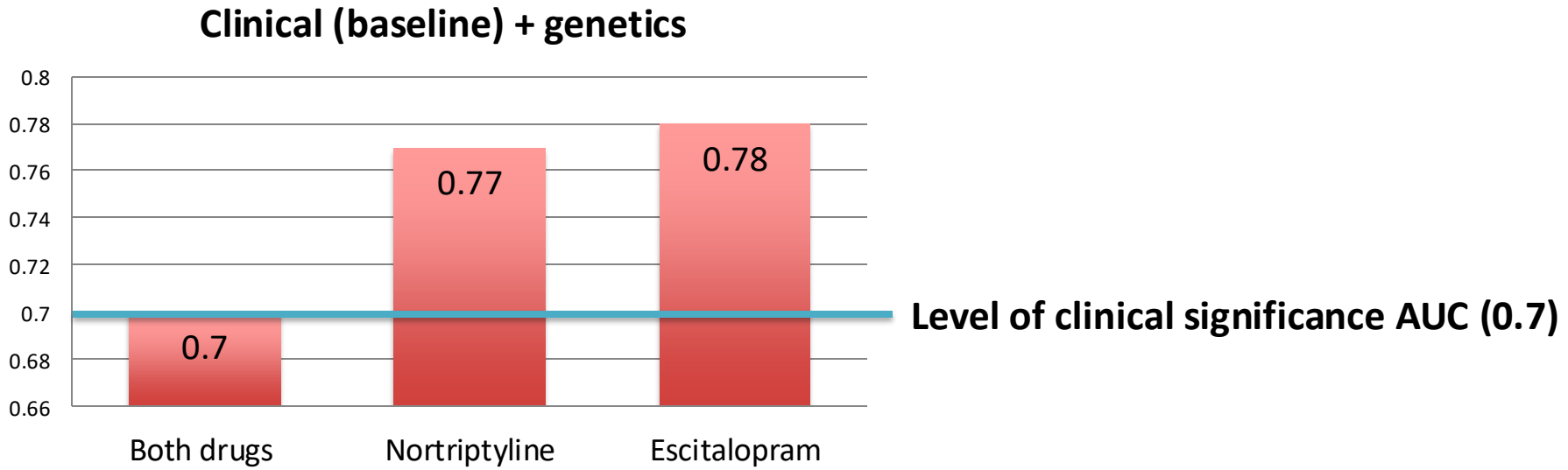
5-fold cross validation



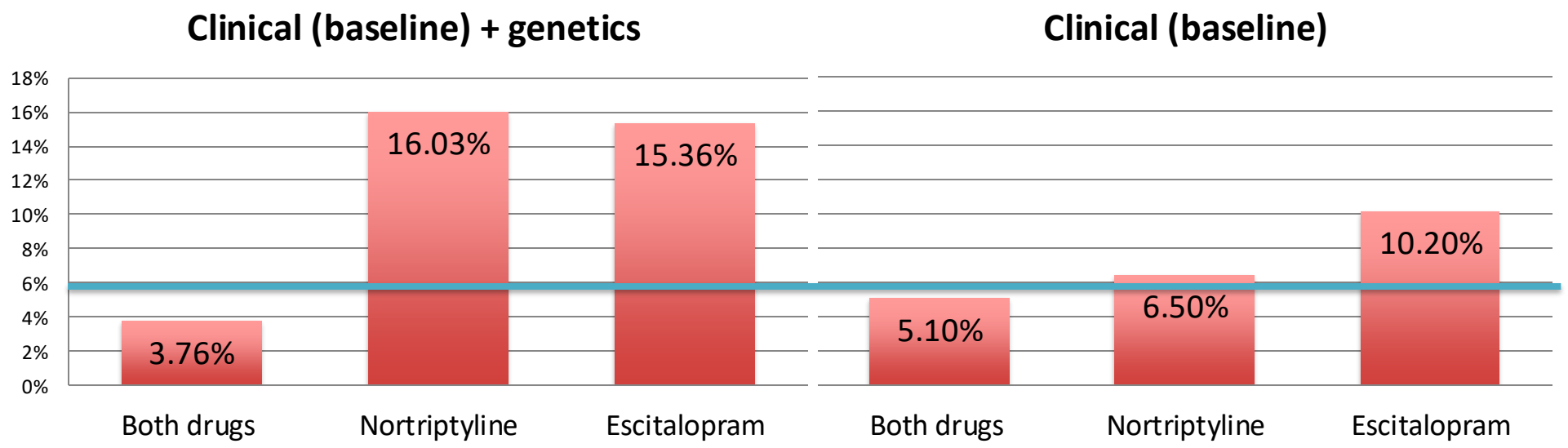
PREDICTION OF SYMPTOMS IMPROVEMENT (% of variance explained)



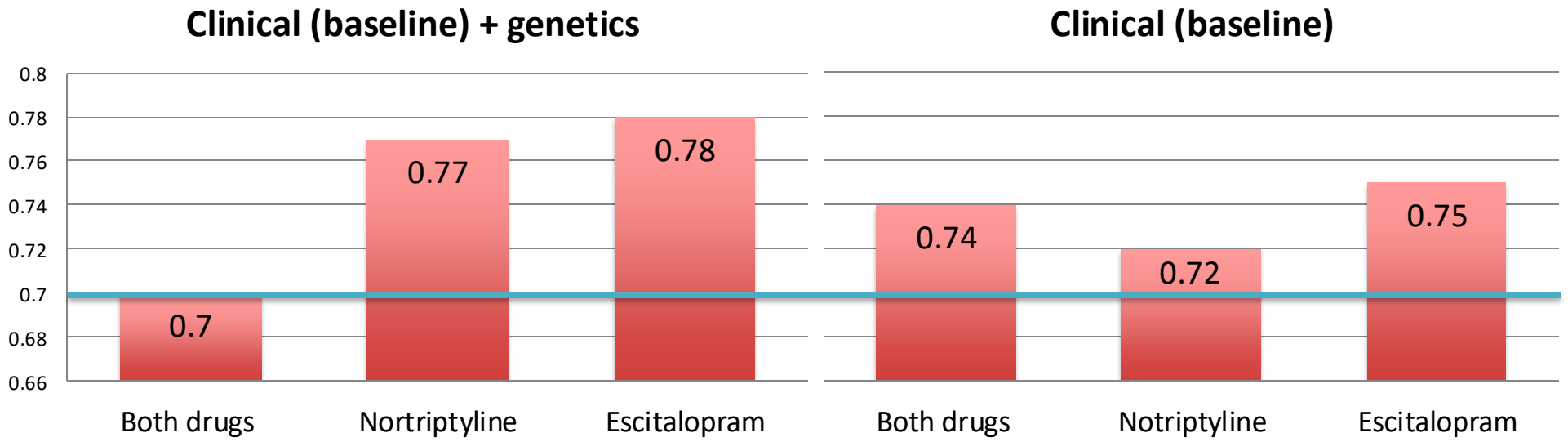
PREDICTION OF REMISSION (Area under ROC curve)



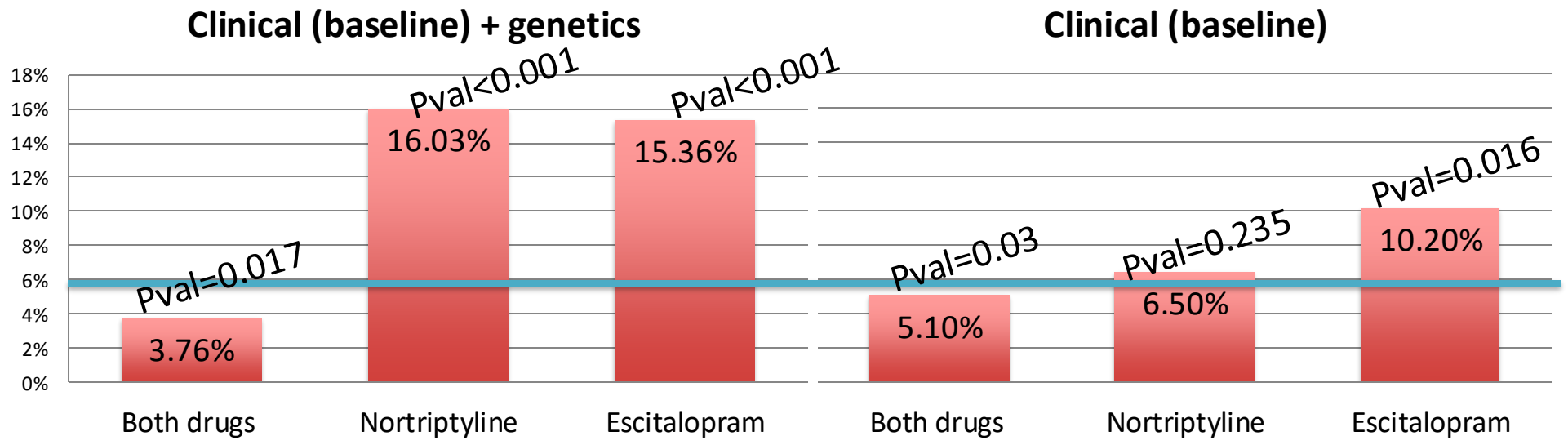
PREDICTION OF SYMPTOMS IMPROVEMENT (Variance explained)



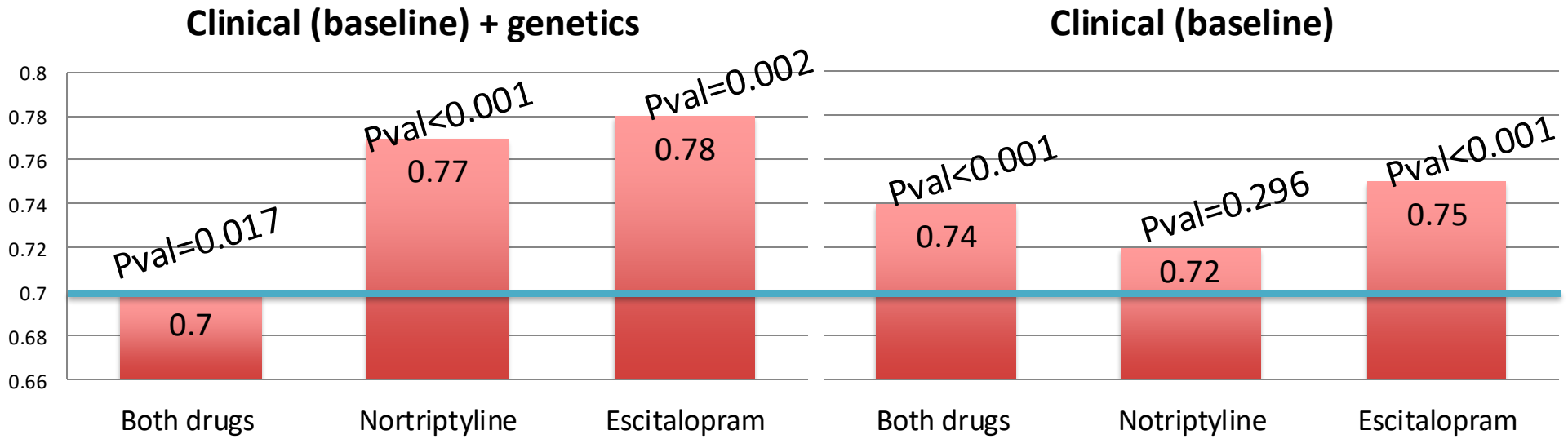
PREDICTION OF REMISSION (Area under ROC curve)



PREDICTION OF SYMPTOMS IMPROVEMENT (Variance explained)



PREDICTION OF REMISSION (Area under ROC curve)



Conclusions

- Some of the SNPs reported here are located in **genes previously associated** with response to lithium in BD (ACCN1), synaptic connectivity (ITGB2), parkinson (TMEM229B), alzheimer's (COL25A1) or coronary risk disease (CFDP1), but not with antidepressant treatment response.

- These SNPs combined with easily obtained **demographic and clinical variables,** can **predict treatment response with clinical and statistical** significance**.**

- The addition of **genetic** predictors to clinical predictors **increased the accuracy** of drug-specific outcomes suggesting a role of genetic variability

- Models were drug-specific, suggesting a **potential for individualized indications** for antidepressant drugs.

- This method is a basis for the **addition of other measures** (longitudinal measures, neuroimaging, biochemical and electrophysiological) that could further improve the prediction.

Addition of longitudinal symptoms measuremment

CLINICAL PREDICTORS

- **Baseline depression severity (MADRS, HRSD, BDI)**
- **Weekly depression severity (MADRS, HRSD, BDI) (longitudinal measures)**
- **Age**
- **Age at onset**
- **Sex**
- **Smoke (YES or NO)**
- **Years of education, number of children, occupation**
- **BMI at baseline**
- **Previously published depression factors for symptom dimensions**
- **Stressful Life events (LTE-Q): YES or NO, and number of total events**
- **Subtypes Depression (SCAN): atypical, melancholic, anxious**
- **Medication History: number of prior antidepressant trials, types of antidepressants tried**

GENETIC PREDICTORS

- GWAS data, 524,871 SNPs – filtered by MAF, missingness, and Linkage disequilibrium <0.8
- **Polygenic risk score for MDD**

OUTCOME

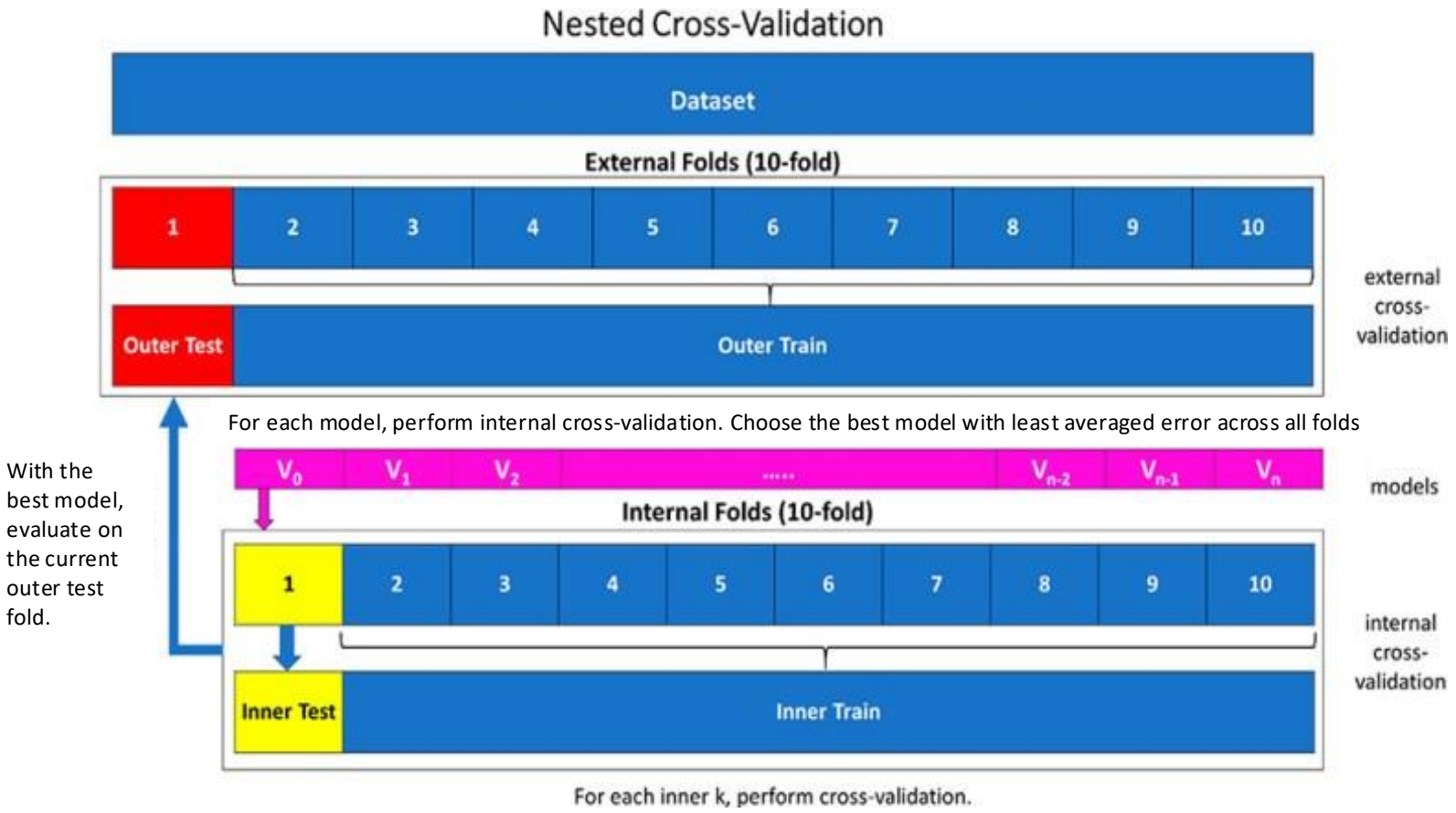
- **Remission** at last assessment (HDRS≤7) (binary)

Pipeline 2 (improved)

Whole data

- Predictive performance for each feature set was evaluated using **repeated, nested k-fold cross-validation** (10-fold repeated 20 times).
- All **modeling steps** were repeated within each fold to avoid data leakage, including:
  - pre-processing (e.g., standardization, removal of features with zero variance)
  - hyperparameter tuning (e.g., alpha)
  - feature selection
  - missing value imputation (imputed using *k*-nearest neighbors ($k = 5$) using the KNNImputer class from scikit-learn (Troyanskaya et al. 2001).)
- Cross-validation metrics were summarized by averaging each repetition (i.e., the mean over 10 folds) and reporting the median of the resulting values.
- Uncertainty intervals were computed based on the outer loop of the repeated, nested cross-validation.

Nested, 10-fold cross validation



There are **two cross-validations** (CVs) in the nested cross-validation (note the word "nested"); an external CV and an internal CV.

The number of folds for both the external and internal CVs can be **defined according to the nature of the project** and the developer

NeCV procedure is **not to get a single**, finalized, robust model that is ready for making predictions, but instead **to get an estimate of the unbiased generalization** of the performance**.**

Pseudocode for Nested, 10-fold cross validation

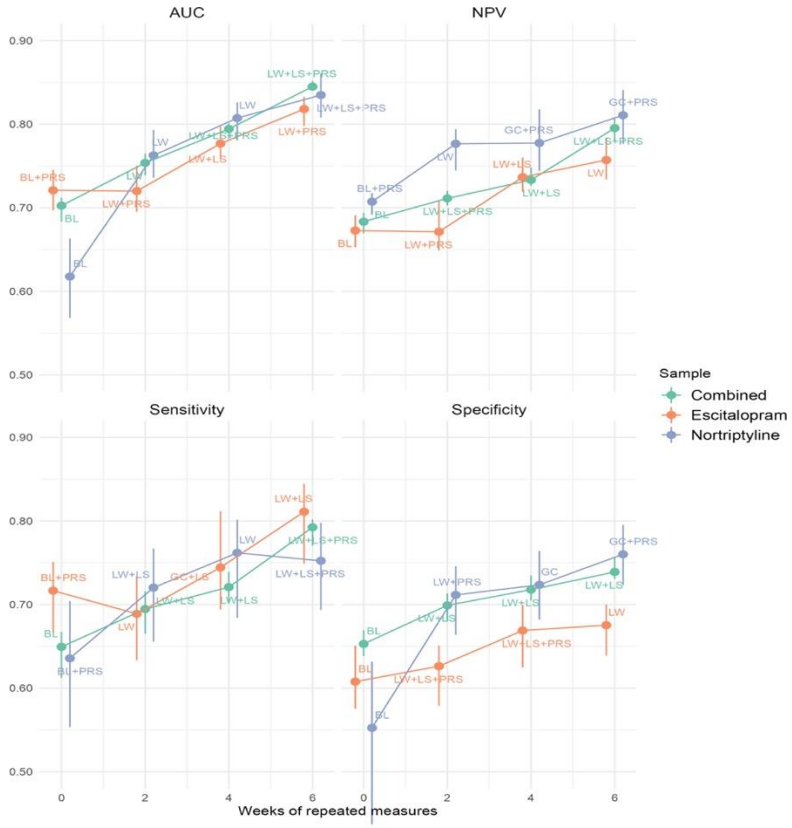
Algorithm of Nested Cross-Validation

1. Split the dataset S into external k (*10*) folds.
2. For i = 1 to external_k */*external loop for evaluation of the model with selected parameters from internal loop*/*
   external_test = fold i
   external_train = all the data except those in i^{th} fold
   For each parameter set p: */*parameter selection*/*
   Randomly split the external_train into internal_k (*10*) folds
   */*for internal cross-validation*/*
   For j=1 to internal_k
   internal_test = fold j
   internal_train = all the data except those in internal_test or external_test
   Train the regression model on the internal_train
   Calculate the test error (MAE) e_j on the parameter internal_test
   Compute internal CV test error $e_{cv} = \frac{1}{inner\_k}\sum_{j=1}^{inner\_k} e_j$
   Select parameter set p with minimum e_{cv}
   Train regression model with selected parameter set on external_train
   Calculate test error e_i on the external_test
3. Calculate external CV test error $e_{cv} = \frac{1}{outer\_k}\sum_{i=1}^{outer\_k} e_i$

Source: https://drzinph.com/nested-cross-validation-cross-validation-series-part-3/

Cross-validation vs. nested cross-validation

- Both native cross-validation (also known as k-fold cross-validation) and nested cross-validation (NeCV) are **useful in working with small datasets**.
- The major pitfall of **cross-validation is that it can give a significantly biased estimate** of the true error because the same CV procedure is used to tune and optimize parameters, to perform model selection, and also to estimate the generalization error.
- In lack of proper EXTERNAL VALIDATION, NeCV is designed to give an almost "**unbiased" estimate of the true error**. These two tasks are split: the internal CV is used for model selection/hyperparameter tuning, while the external CV is used for estimating the generation error.
- The **cross-validation pro is that it is much less computationally** intensive than NeCV.



BL = Baseline only.
LW = Latest weekly measures only*.
GC = Growth curves only*.
+LS = Supplemented with topological landscape variables
+PRS = Supplemented with polygenic risk score
*All models included baseline features.

		0 weeks	2 weeks	4 weeks	6 weeks
Escitalopram	AUC	0.718	0.717	0.777	0.818
	NPV	0.673	0.668	0.737	0.757
	Sensitivity	0.700	0.689	0.744	0.811
	Specificity	0.608	0.625	0.668	0.675
Nortriptyline	AUC	0.618	0.763	0.807	0.834
	NPV	0.706	0.777	0.777	0.809
	Sensitivity	0.630	0.720	0.762	0.747
	Specificity	0.552	0.711	0.724	0.756
Combined	AUC	0.703	0.754	0.794	0.844
	NPV	0.683	0.711	0.733	0.792
	Sensitivity	0.649	0.695	0.721	0.791
	Specificity	0.653	0.699	0.718	0.739

Notes. The classification threshold was set at 40% for the Escitalopram and 'Combined' samples and 30% for Nortriptyline.

Results/conclusions

- **Adding post-baseline weeks** of information suggested incremental benefits in performance over successive weeks
- **No single point showing a marked increase** in performance.
- Our machine learning models predicted depression remission by **week four** with **excellent discrimination of ~0.8** AUC, suggesting a potential to inform treatment decisions
- **Adding a PRS** for MDD did not show an improvement in performance

Hypertension work:

Genetically defined ancestry as predictor of response to antihypertensive treatment

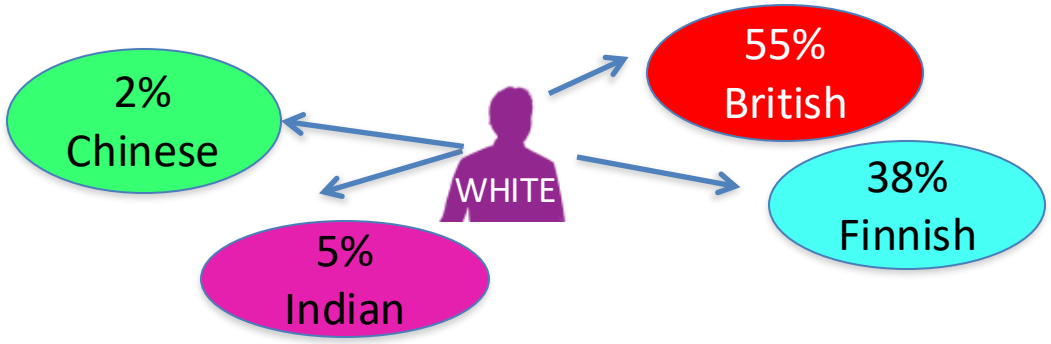
Gene Variants at Loci Related to Blood Pressure Account for Variation in Response to Antihypertensive Drugs Between Black and White Individuals: Genomic Precision Medicine May Dispense With Ethnicity

Iniesta, R., Campbell, D., Venturini, C., Faconti, L., Singh, S., Irvin, M. R., Cooper-DeHoff, R. M., Johnson, J. A., Turner, S. T., Arnett, D. K., Weale, M. E., Warren, H., Munroe, P. B., Cruickshank, K., Padmanabhan, S., Lewis, C. & Chowienczyk, P., 1 Sept 2019, In: Hypertension (Dallas, Tex. : 1979). 74, 3, p. 614-622 9 p.

Estimating the genetic admixture

- Humans are structured, but there are **not discrete races**.
- Today individuals are **ADMIXED**: A single individual presents a percentage of different ancestries.

  Example:



Self defined ethnicity can be misleading important information about ancestry.

Objective of the AIM-HY

Response to antihypertensive drugs is known to differ according to self-defined ethnicity.

Objective: Compare **Genetically Defined Ancestry** (GDA) score to **Self Defined Ethnicity** (SDE) in predicting response to antihypertensive treatment.

AIM-HY Samples

Study	Type of trial	F-up	Drug(s)	Class	Sample size	SDE (%B)
GERA 1	Monotherapy	4w	Hydrochlorothiazide	Thiazide Diuretic	517	48.55
GERA 2	Monotherapy	4w	Candesartan	Angiotensin II receptor antagonist	365	49.04
PEAR 1	Parallel Randomised	9w	Atenolol	Beta-blocker	367	39.51
			Hydrochlorothiazide	Thiazide Diuretic	363	39.39
PEAR 2	Sequential monotherapy	6w	Metoprolol	Beta-blocker	358	45.81
			Chlorthalidone	Thiazide Diuretic-Like	319	44.51
INVEST	Parallel Randomised	6w	Verapamil	Calcium Channel Blocker	198	13.13
			Atenolol	Beta-blocker	250	12.8
GENHAT	Parallel Randomised	6m	Chlorthalidone	Thiazide Diuretic-Like	376	35.37
			Amlodipine	Calcium Channel Blocker	195	33.33
			Lisinopril	Angiotensin Converting Enzyme	238	40.34

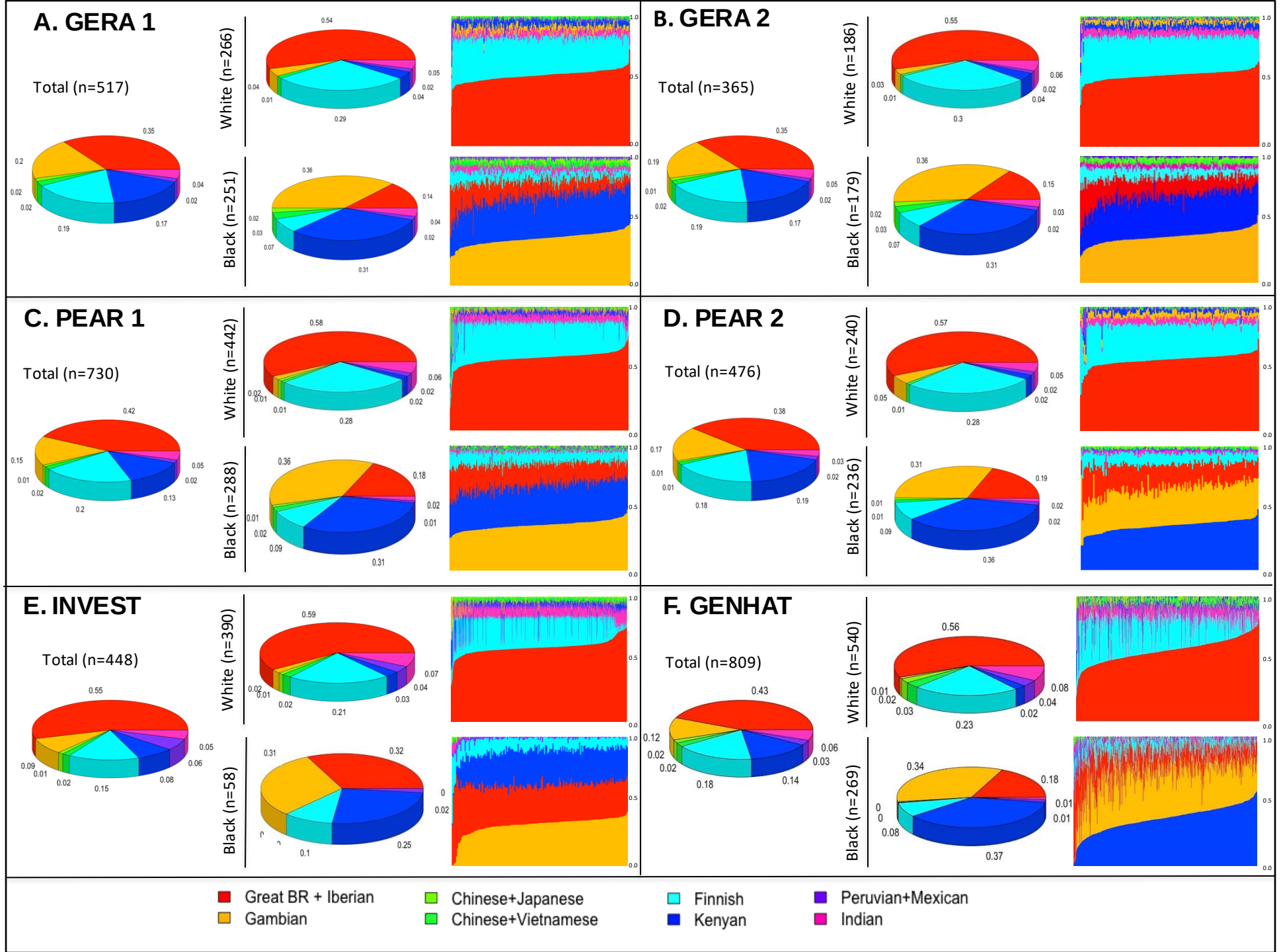
Methodology

Compare Genetically Defined Ancestry **GDA** to Self Defined Ethnicity SDE in **predicting** response to antihypertensive treatment.

1) **GDA estimation:**
   I. Design an algorithm to derive a GDA score of % ancestry attributable to European, African and Asian origin
   II. Use it to compute GDA components for each individual in GERA, PEAR, INVEST and GENHAT datasets.

2) **Modelling**: Investigate the role of SDE and GDA in predicting antihypertensive response in every study-drug sample.

1) GDA estimation:



2) Effect of SDE and GDA on drug response

- **Primary Outcome**:
  Systolic BP change (ΔSBP) defined as pre SBP - post SBP treatment

- **Predictors:**
  SDE: 0 white; 1 black;
  GDA_BLACK: % Gambian + % Kenyan
  baseline SBP, sex, age, BMI

<u>Analysis: focus on prediction, machine learning approach</u>

ii) <u>Machine learning approach</u>: Elastic Linear regression that **selects the best** subsets of predictors and **allow correlated variables to enter to the model** (as SDE and GDA). 5-fold cross-validation for hyperparameters tunning. External validation across studies.

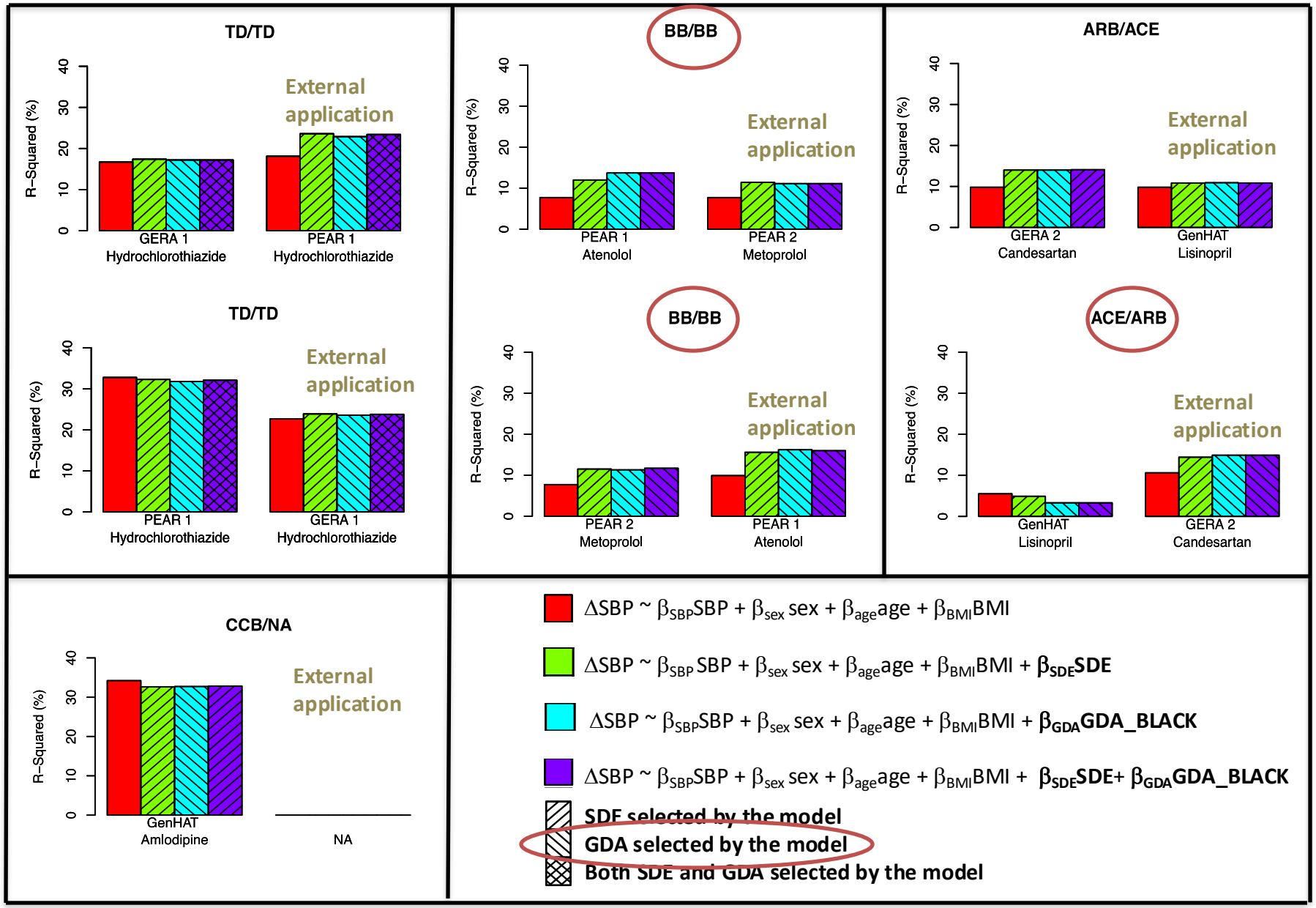
$$\Delta SBP \sim \beta_{SBP}\,\text{baseline SBP} + \beta_{sex}\,\text{sex} + \beta_{age}\text{age} + \beta_{BMI}\text{BMI}$$

$$\Delta SBP \sim \beta_{SBP}\,\text{baseline SBP} + \beta_{sex}\,\text{sex} + \beta_{age}\text{age} + \beta_{BMI}\text{BMI} + \boldsymbol{\beta_{SDE}\text{SDE}}$$

$$\Delta SBP \sim \beta_{SBP}\,\text{baseline SBP} + \beta_{sex}\,\text{sex} + \beta_{age}\text{age} + \beta_{BMI}\text{BMI} + \boldsymbol{\beta_{GDA}\text{GDA\_BLACK}}$$

$$\Delta SBP \sim \beta_{SBP}\,\text{baseline SBP} + \beta_{sex}\,\text{sex} + \beta_{age}\text{age} + \beta_{BMI}\text{BMI} + \boldsymbol{\beta_{SDE}\text{SDE} + \beta_{GDA}\text{GDA\_BLACK}}$$

Results for elastic net linear regression



Conclusions

- In general, **SDE and GDA had a similar** effect on antihypertensive drug response. Both SDE and GDA are associated and predictive of treatment response in HTZ, Betablockers and Candesartan .
- For both **atenolol and metoprolol**, response was **more closely related to GDA** than to SDE. In penalised regression models incorporating both SDE and GDA, GDA was a better predictor of response than SDE.
- Sub-analysis on the **highly admixed** subjects of mainly black ancestry: **GDA was meaningful as predictor** of response for both the beta-blockers, the ACEi and ARB.
- The present results show that **GDA could potentially be used** in admixed peoples.

Overall conclusions

- Genetic data combined with easily obtained **demographic and clinical variables,** can **predict treatment response with clinical and statistical** significance in depression and hypertension**.**
- Studies on predicting treatment response are consistent with a **genetic component** underlying the response to these drugs.
- **Machine learning** was the only methodology capable to deal with high dimensional data
- Further studies involving external sources for validation are required.

Thank you!

Prof Cathryn Lewis
Prof Phil Chowienczyk
Dr Mike Weale
Dr Cristina Venturini

Prof Sandosh Padmanabhan
Dr Desmond Campbell

US COLLABORATORS

Prof Peter McGuffin
Prof Rudolf Uher

Prof Daniel Stahl
Dr Karim Malki
Prof Richard Dobson
Prof Rob Tibshirani
Dr Max Kuhn

Let's make AI **reliable**.
Let's make AI **useful**.
Let's make AI **ethical**.

Dr Raquel Iniesta
raquel.iniesta@kcl.ac.uk