Day 3
Group Projects

Guidelines and Grading Rubic

Itunuoluwa Isewon (Ph.D)

Polygenic Risk Score Analysis, 22 - 27 June 2025,
Makerere University, Uganda

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

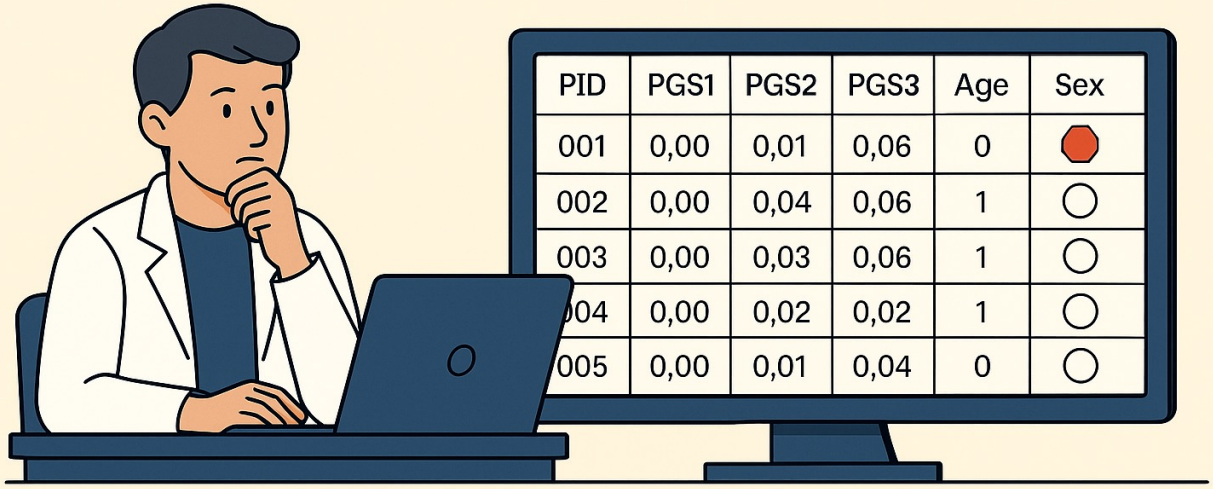
wellcome connecting science

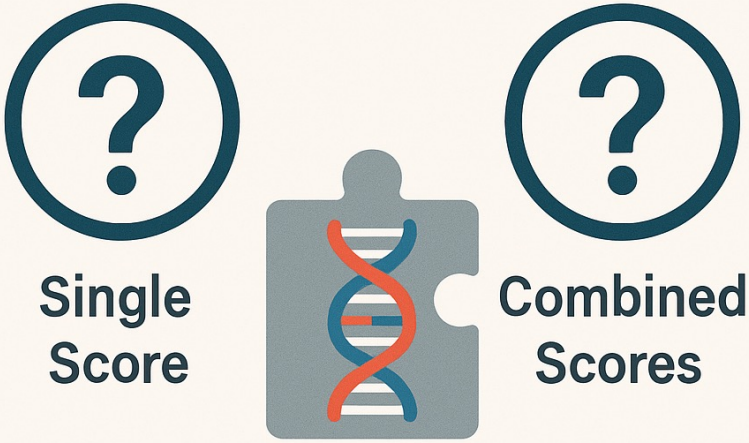
Learning Objectives

- Apply basic principles of polygenic risk scoring using integrated approaches
- Interpret and evaluate PRS performance in African population datasets
- Compare different polygenic risk scores and reason about differences in performance
- Communicate scientific findings clearly and effectively
- Collaborate as a team to solve a data-driven research question

Imagine This

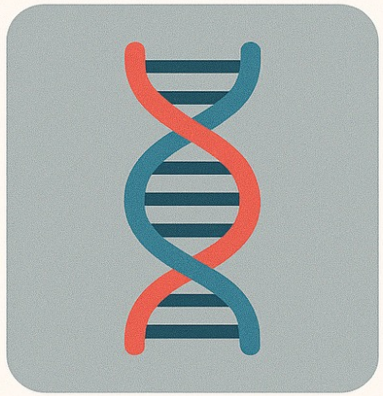
You're a researcher holding a dataset with hundreds of individuals. For each person, you know their genetic risk scores, age, sex, and whether they've developed a condition—maybe high blood sugar, depression, or asthma.





Which genetic risk score is most predictive for a trait? And what happens if you combine them?

Integrated PRS



Combines multiple genetic risk scores into a single model

- More accurate predictions
- Greater insights

Leverages complementary information from different GWAS-derived PRS.

Uses statistical or machine learning models (e.g., elastic net) to assign optimal weights.

Strengths

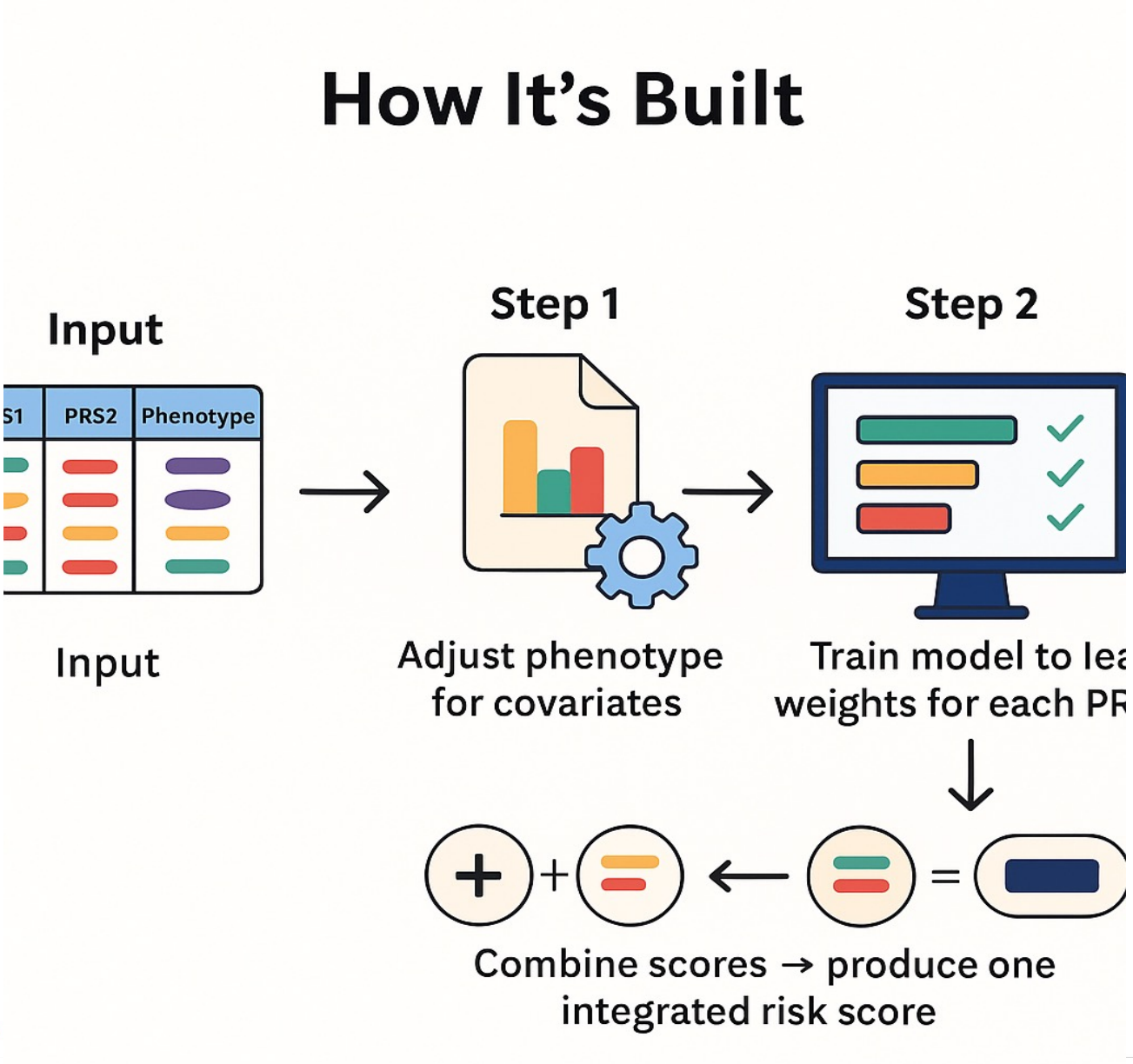
- **Improved prediction**: Outperforms most single PRS by capturing more variance in traits.
- **Trait-specific tuning**: Adapts to the phenotype being analyzed (quantitative or categorical).
- **Robust to noise**: Reduces dependence on any one potentially weak PRS.
- **Customizable**: Can be optimized for ancestry, sex, environment, or disease subtype.

MAKERERE UNIVERSITY | IDI INFECTIOUS DISEASES INSTITUTE | AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

4

Integrative PRS

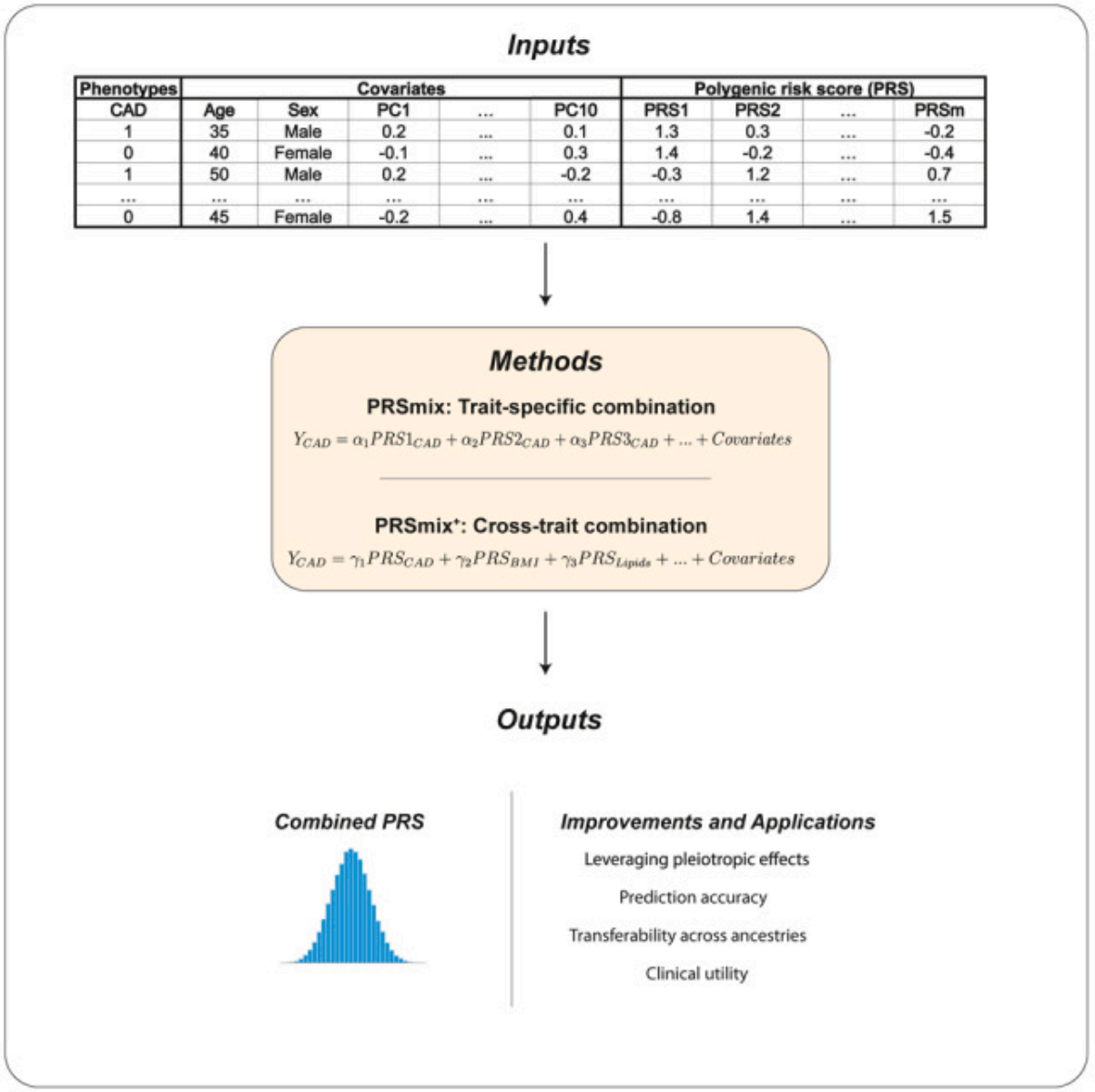


MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

PRSmix / PRSmix+



MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

wellcome connecting science

ELASTIC NETS

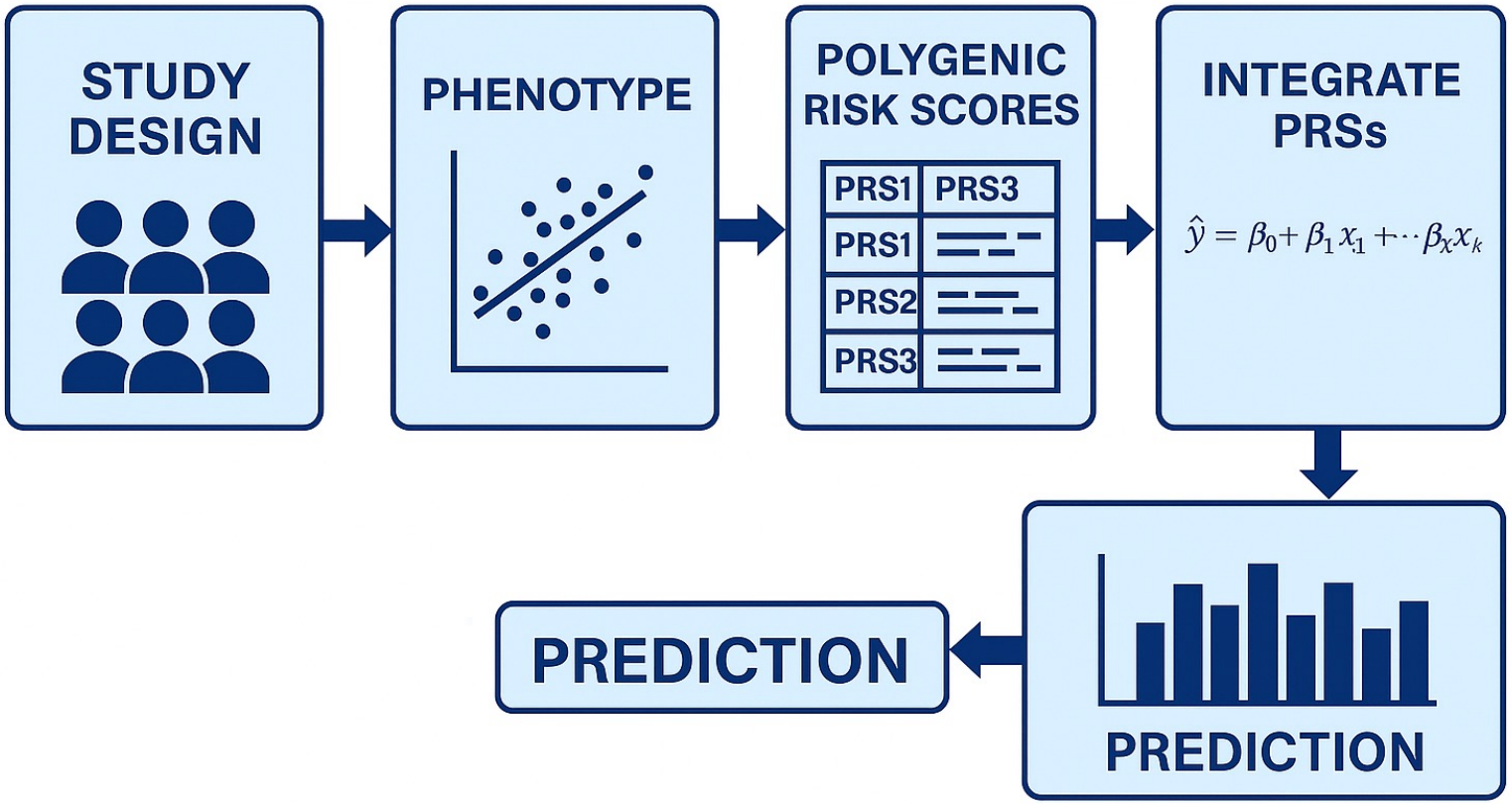
$$\text{minimize} \sum_{i}^{n} (y - \beta_0 - \beta_j) + \lambda_2 \sum \beta_j$$

Lasso (L_1) Ridge (L_2)

Strengths

- **Handle multicollinearity**: which is when some predictors are highly correlated with each other. Elastic net can overcome these problems by selecting a subset of predictors that are correlated, but not redundant.
- **Reduce Overfitting**: Elastic net can balance the bias-variance trade-off by finding a middle ground between underfitting and overfitting.
- **Provide lean set of important predictors**: Elastic net can perform feature selection by setting some coefficients to zero, while keeping others that are significant.

STUDY DESIGN



To do list

- Write a script (markdown) in R or Python for the analysis
- Push to the WCS PRS Course Github Page
- Upload slides to the google drive
- Prepare a project summary report and upload to the google drive
- Prepare slide presentation
- Present your findings (10 mins)
- Answer questions from the panel

MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

9

Datasets – Course Github Day 5

Group 1 Glycated hemoglobin (HbA1C)

Group 2 Insulin reistance (HOMA_IR)

Group 3 OGTT 2hr glucose

Group 4 Subcutaneous adipose tissue (SAT)

MAKERERE UNIVERSITY

INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

PROJECT QUESTIONS

1. **Applicability of PRSmix**
   Do you think the PRSmix approach is likely to perform well in continental African populations? Why or why not?
   - Consider factors such as genetic diversity, linkage disequilibrium patterns, ancestry mismatch in GWAS summary statistics, and trait architecture

2. **Evaluation of PRSmix Performance**
   How well did the PRSmix-derived PRS perform in your assigned dataset?
   - Report and interpret relevant metrics (e.g., AUC, R^2, odds ratios)
   - Comment on the strength of prediction and any subgroup differences

3. **Comparison with PGS Catalog Scores**
   How does the PRSmix-derived PRS compare to the individual PRS scores from the PGS Catalog in terms of predictive performance for your dataset?
   - Which performed better?
   - What factors (e.g., ancestry matach, training data quality, trait polygenicity) might explain the difference?

4. **Interpretation and Limitations**
   What do your results suggest about the utility of polygenic scores for this trait in African populations?

Grading Rubric

- ✓ Introduction/Motivation
- ✓ Analysis
- ✓ Results (Q1 – Q4)
- ✓ Discussion
- ✓ Project Report
- ✓ Markdown on Github
- ✓ Quality of presentation
- ✓ Team Collaboration

wellcome connecting science

11

Suggested Outlines

Slides

- Introduction
- Motivation
- Materials and Methods
- Results
- Insights
- Outlook/Conclusion
- Acknowledgements

Project Summary Report

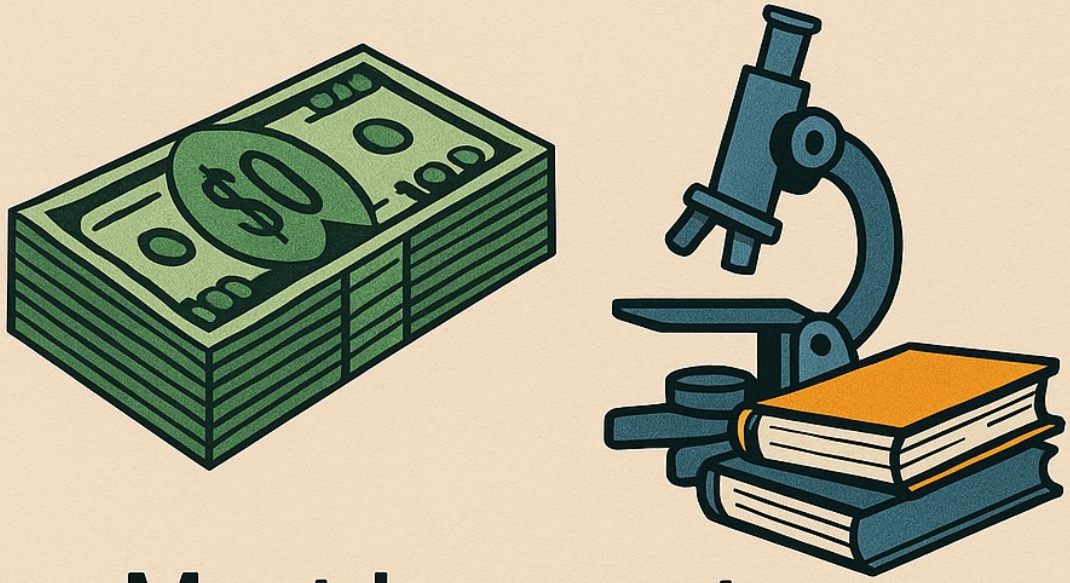
- Title page (Project title, Group Name/Number, Group members, date, location)
- Brief Background including phenotype description, motivation, study objectives
- Dataset Description
- Analysis Pipeline
- Results
- Interpretation (Biological and Clinical)
- Limitations
- Conclusions
- References

MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

12

wellcome connecting science

$100 CASH PRIZE



Must be spent on scientific purposes