Session 3: PRS calculation and evaluation

Instructors: Meng Lin and Maizy Brasher

Table of contents

Overview: From weights to scores

- Lots of publications on developing PRS models using various methods and training datasets (e.g. your exercise in the previous session)
- Outputs of these models (i.e. weight files) are often publicly available
  - For example, PGS catalog (https://www.pgscatalog.org/) is a centralized resource
  - In a form of summary statistics of risk effect at a selction of variants

chr_name	chr_position	effect_allele	other_allele	effect_weight
1	779322	G	A	-6.211012e-06
1	1005806	T	C	-3.127074e-05
1	1017197	T	C	-3.070467e-05
1	1017587	T	C	-6.693554e-05
1	1018704	G	A	-5.915339e-05
1	1021695	G	A	-0.0002213063
1	1030565	T	C	2.667483e-05
1	1030633	A	G	5.357169e-05
1	1031540	G	A	-0.0001155718

The fundamental calculation of scores

Weight file
(M x 1)

SNP1	-0.3
SNP2	0.2
SNP3	-0.1
…	…
SNPM	0.3

The fundamental calculation of scores

Test cohort genotypes
(N x M)

	SNP1	SNP2	...	SNPM
Sample1	0	2	...	1
Sample2	1	0	...	2
...	...	...	...	...
SampleN	2	1	...	1

Weight file
(M x 1)

SNP1	-0.3
SNP2	0.2
SNP3	-0.1
…	…
SNPM	0.3

The fundamental calculation of scores

Test cohort genotypes
(N x M)

	SNP1	SNP2	...	SNPM
Sample1	0	2	...	1
Sample2	1	0	...	2
...	...	...	...	...
SampleN	2	1	...	1

•

dot product

Weight file
(M x 1)

SNP1	-0.3
SNP2	0.2
SNP3	-0.1
…	…
SNPM	0.3

Score values
(N x 1)

Sample1	1.1
Sample2	0.6
Sample3	0.2
…	…
SampleN	-0.1

The fundamental calculation of scores

Test cohort genotypes
(N x M)

	SNP1	SNP2	...	SNPM
Sample1	0	2	...	1
Sample2	1	0	...	2
...	...	...	...	...
SampleN	2	1	...	1

•

dot product

Weight file
(M x 1)

SNP1	-0.3
SNP2	0.2
SNP3	-0.1
…	…
SNPM	0.3

Score values
(N x 1)

Sample1	1.1
Sample2	0.6
Sample3	0.2
…	…
SampleN	-0.1

Yes, it's that simple!

The fundamental calculation of scores

Test cohort genotypes
(N x M)

	SNP1	SNP2	…	SNPM
Sample1	0	2	…	1
Sample2	1	0	…	2
…	…	…	…	…
SampleN	2	1	…	1

•

dot product

Weight file
(M x 1)

SNP1	-0.3
SNP2	0.2
SNP3	-0.1
…	…
SNPM	0.3

- You can code it in R, python, or any other of your favorite language
- You can use PLINK --score feature
- …

Score values
(N x 1)

Sample1	1.1
Sample2	0.6
Sample3	0.2
…	…
SampleN	-0.1

Yes, it's that simple!

However, harmonization is important in the real world

- Many differences may exist between your dataset (test data) and those used in publications for GWAS and PRS training
- Just imagine trying to merge two very different array datasets
  - Not all variants are found?
  - Genome build difference?
  - Does the test data reflect the counts of the *risk* allele codes?
  - Ambiguous allele code? (MAF matching has limitations)
  - *Effect of a risk allele is relative to a given reference allele (e.g. thinking about GWAS) – all allele code matching*
- Workload increases with larger datasets

Tools for PRS harmonization and calculation

- pgsc_calc: provided by the PGS catalog, https://github.com/PGScatalog/pgsc_calc

- ESCALATOR: https://github.com/menglin44/escalator
  - Container and examples provided in this workshop
  - Optional exercise in Session 3 hands-on markdown

Table of contents

Polygenic scoring can be imperfect

A short list of limitations in example -

Polygenic scoring can be imperfect

- Upstream limitations
  - Genetic discovery limited, are you approaching the heritability "upper bound"?
  - Not capturing/decoding true differences in genetic architectures across ancestries or other context (including GxE, SDOH)
- Uncorrected confounders and bias
  - Baked in from GWAS: ascertainment bias (array+imputation), w/ LD differences
  - Bias furthered in training cohort: LD
  - Methods' limitations
  - Etc.

ARTICLE

Geographic Variation and Bias in
of Complex Diseases and Traits i

Sini Kerminen,[1] Alicia R. Martin,[2,3,4] Jukka Kosl
Ida Surakka,[1,6] Aarno Palotie,[1,2,3,7,8] Markus Perc
Samuli Ripatti,[1,9] and Matti Pirinen[1,9,10,*]

Research Article
Genetics and Genomics

Variable prediction accuracy of polygenic scores within an ancestry group

Polygenic scoring accuracy varies across the genetic ancestry

Yi Ding, Kangc

an K Pritchard, Molly

Demographic history mediates the effect of stratification on polygenic scores

Highly parameterized polygenic scores tend to overfit to population stratification via random effects

ations

Alan J. Aw, Jeremy McRae, Elior Rahmani, Yun S. Song

doi: https://doi.org/10.1101/2024.01.27.577589

This article is a preprint and has not been certified by peer review [what does this mean?].

article 85

r 2020, Pages 4027–4036,

Genetic Risk

https://doi.org/10.1534/g3.120.401658

Published: 01 November 2020 **Article history ▾**

Alicia R. Martin,[1,2]
Benjamin M. Neale,[1,2,3] Simon Gravel,[5,6] Mark J. Daly,[1,2,3] Carlos D. Bustamante,[4]
and Eimear E. Kenny[7,8,9,10,*]

Polygenic scoring can be imperfect

- A limited list of limitations
- Development of PRS is evolving, with improvement in multiple dimensions
  - Bigger GWAS
  - Collecting data from individuals of diverse backgrounds
  - Improvement of methods
  - Etc.

Evaluating PRS performance in the presence of potential confounders

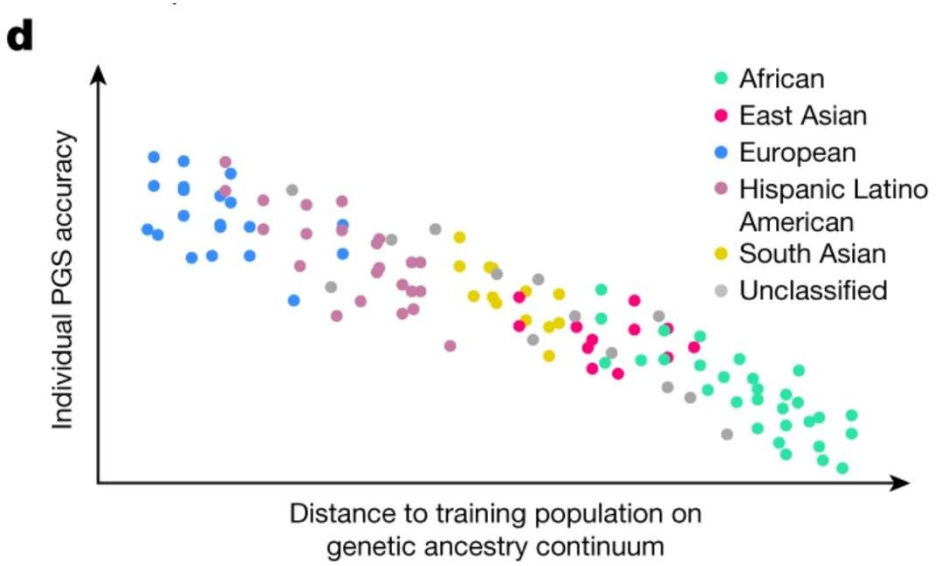
- Should you compare raw PRS values across populations?

Should you compare raw PRS values across populations?

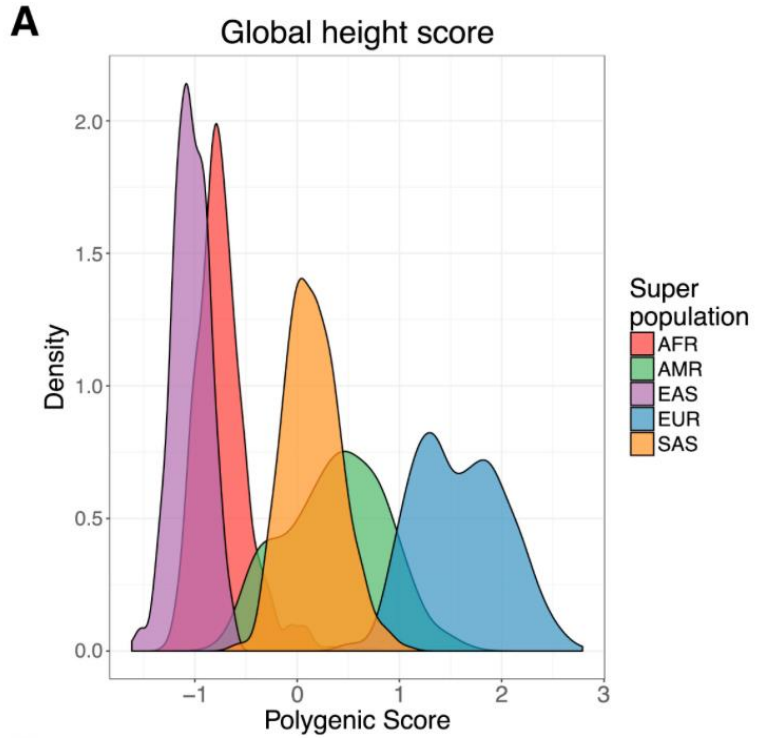
- Estimated PRS is confounded by ancestry
  - Note that this is a different concept than “PRS prediction accuracy decreases/varies by ancestry”

Should you compare raw PRS values across populations?

- Estimated PRS is confounded by ancestry (right)
  - Note that this is a different concept than “PRS prediction accuracy decreases/varies by ancestry” (left)



vs



Ding et al. Nature. 2023

Martin et al. AJHG. 2017.

Should you compare raw PRS values across populations?

- Estimated PRS is confounded by ancestry
  - Note that this is a different concept than “PRS prediction accuracy decreases/varies by ancestry”
- *Why is that?*

Should you compare raw PRS values across populations?

- Estimated PRS is confounded by ancestry
  - Note that this is a different concept than “PRS prediction accuracy decreases/varies by ancestry”
- *Why is that?*
  - *What if causal variants are indeed ancestry-informative? (related to the next section)*

Evaluating PRS performance in the presence of potential confounders

- Should you compare raw PRS values across populations?
- PRS unit (within a presumably homogeneous group)
  - Standardization
  - Stratification
- Control for confounders in your phenotype associations
  - Think of GWAS – “full model”
- Compare with predictions with just non-PRS covariates
  - Think of GWAS – “ null model”
  - Look at incremental improvement in measurement

Evaluating PRS performance in the presence of potential confounders

- Prediction accuracy (commonly used ones):
  - Beta, R^2 (quantitative trait)
  - OR, pseudo-R^2, AUC (dichotomous trait)
  - Individual level (w/ posterior distribution of risk effects available)
  - Etc.

Evaluating PRS performance in the presence of potential confounders

- Metrics you'll use in the hands-on exercise
  - AUC (dichotomous trait)
    - "area under the curve"
    - Measuring performance of a classifier method: if true positive rate is higher than false positive rate
    - Ranges from 0.5 (complete random guess) to 1 (always correctly classify): the higher the better performance

Evaluating PRS performance in the presence of potential confounders

- Metrics you'll use in the hands-on exercise
  - I^2
    - Heterogeneity of a score's effects across groups
    - Similar metrics include Cochran's Q, tau^2, etc.
    - (One of many ways to proxy) portability

Hands-on Exercise: Evaluating PRS performance in the presence of potential confounders

Return to the R Markdown you used for session 1

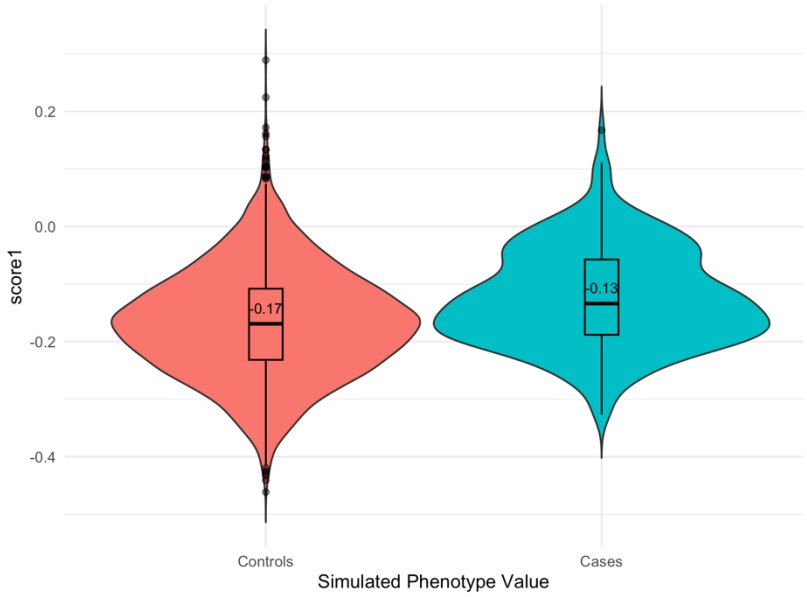
- This part of the tutorial will look at: score distributions between cases and controls, AUC and incremental AUC, and I^2

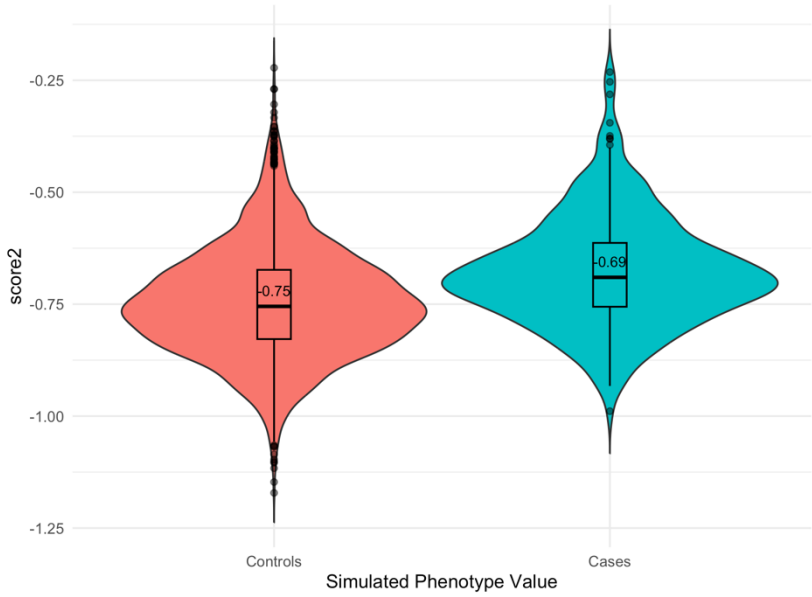
- Work through the code up until "Discussion Break"
  - Note that the section on PRS calculation using ESCALATOR is provided for your reference, but for this tutorial the scores from the PGS catalog have been calculated for you
  - If you can't follow the code or get stuck, DON'T PANIC, figures and tables can be seen in the HTML document
  - Ask a neighbor for help

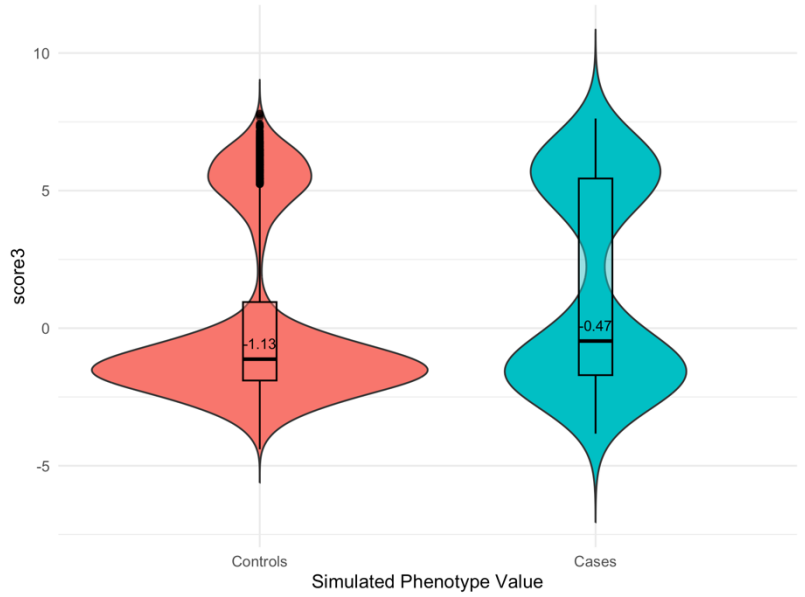
Discussion Question: which of these six scores is the "best"?

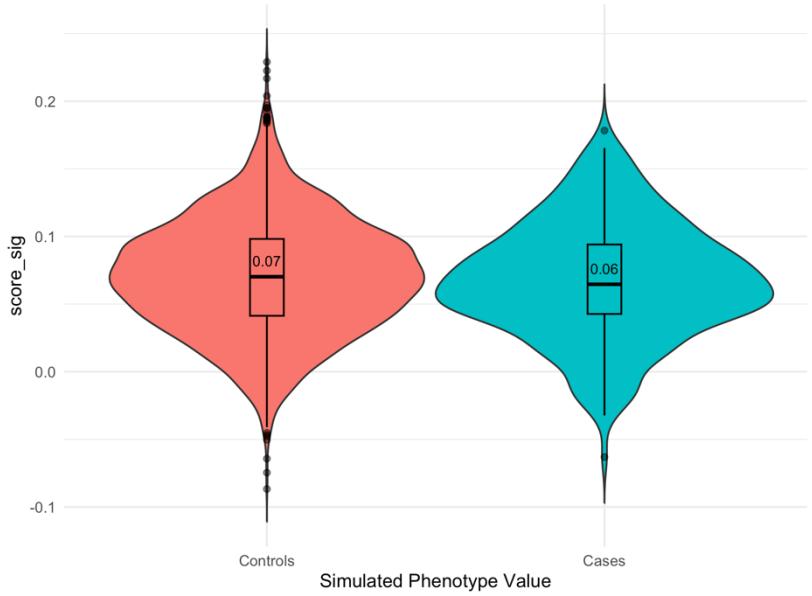
- Consider evidence from the tutorial
- Discuss with the people around you
- We will ask for people to share their thoughts

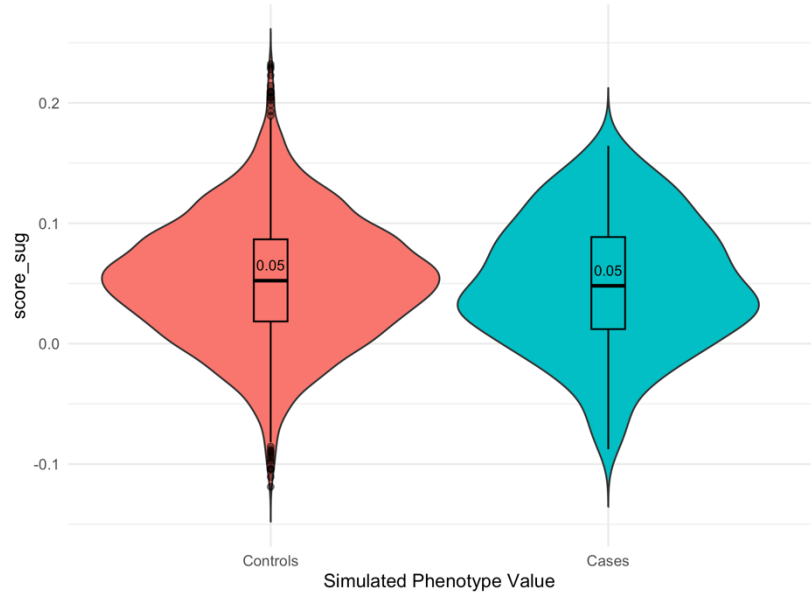
Score distributions by case/control status

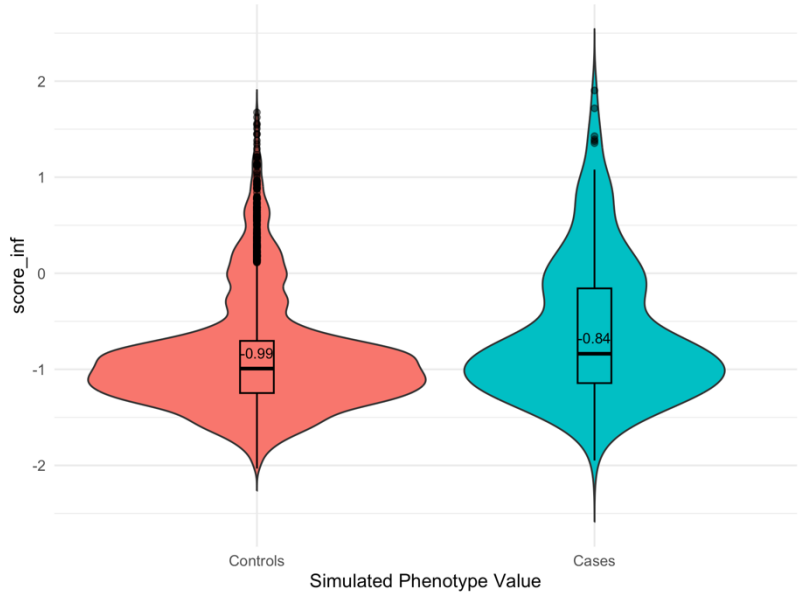




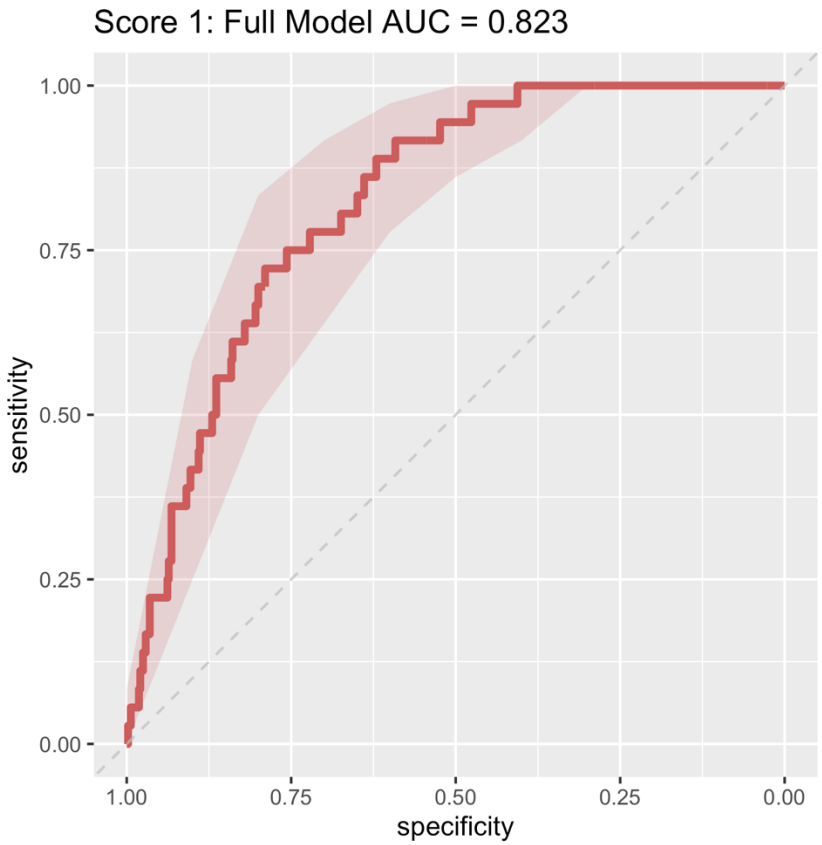


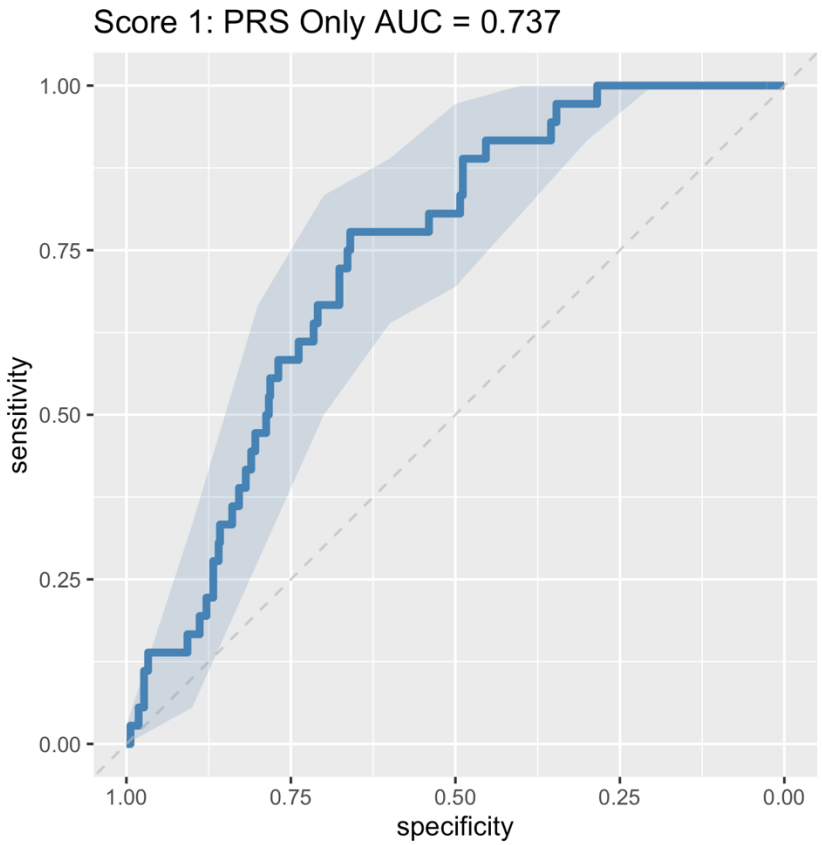






AUC values for European 1000genomes population





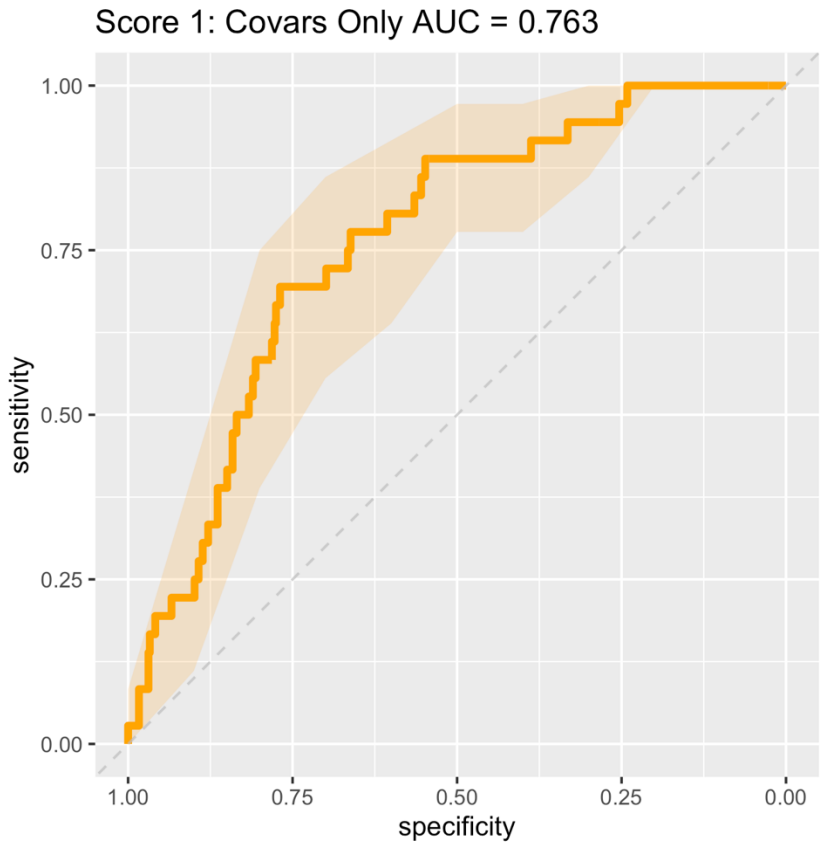


Table of contents

Ancestry calibration after PRS calculations

- Previous exercise: stratified group definitions are used
- But human ancestry is *continuous*: esp. admixed populations with a variation of ancestries. This makes interpretation of risk scores even trickier.
- Recently proposed methods of ancestry calibration:
  - Khera et al, 2019, eMERGE network papers etc. See the full list in the "*recommended readings*" in the end
  - To adjust for ancestry confounding effects from PRS values
  - To move away from using discrete group labels in practical applications

Assumptions behind ancestry calibrations

- Given a trait, the expectation is the true genetic scores across all populations are i.i.d and will follow a normal distribution
  - Based on CLT: PRS= linear combination of lots of causal (common) loci, -> genotypes of the risk loci follow the same binomial sampling
  - The same / similar genetic architectures across populations for the trait
    - Is this always expected to be true?

- Ancestry's effect on PRS is thus only confounding, instead of having true contributions to the genetic values

- Both the mean and variance of an estimated PRS per population is affected by ancestry (proxied by PCs)

Calibration type 1: mean calibration

Khera et al. 2019; Wang et al. 2020

$$\mathrm{PRS} = \alpha_0 + \sum \alpha_i PC_i + \boxed{\varepsilon}$$

Adjusted PRS

- Regression in the dataset as it is, or obtain the coefficients (α_0, α_i) from a separate training dataset, while all the future samples are projected onto that training set to obtain PC in the same space.

You'll practice this method in the hands-on session.

Calibration type 2: mean and variance calibration

Ge et al. 2022; Khan et al., 2022

- Separately regress out PC effects from the mean and the variance

Obtaining parameters

$$\boldsymbol{PRS} = \alpha_0 + \sum \alpha_i \boldsymbol{PC_i} + \boldsymbol{\varepsilon}$$ Mean adjustment

$$\boldsymbol{\delta} = var(\boldsymbol{\varepsilon})$$ Residual variance

$$\boldsymbol{\delta} = \beta_0 + \sum \beta_i \boldsymbol{PC_i} + \boldsymbol{\varepsilon'}$$ Linear adjustment of variance

$$\boldsymbol{PRS_{adj}} = \frac{\boldsymbol{PRS} - (\hat{\alpha}_0 + \sum \hat{\alpha}_i \boldsymbol{PC_i})}{\sqrt{\hat{\beta}_0 + \sum \beta_i \boldsymbol{PC_i}}}$$ Final adjusted ”z-score”

You'll practice this method in the hands-on session.

Calibration type 2.5: mean and variance calibration with non-linear link function

Lennon et al. 2024

- PC-based predicted variance go become negative in a linear model
- Exponential link function to replace the linear model between variance and PCs

$$\boldsymbol{PRS} = \alpha_0 + \sum \alpha_i \boldsymbol{PC_i} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\delta} = var(\boldsymbol{\varepsilon})$$

$$\boldsymbol{\delta} = exp(\beta_0 + \sum \beta_i \boldsymbol{PC_i} + \varepsilon')$$

- Obtaining parameters through maximal likelihood method in a training dataset

Link function to prevent estimated variance becomes negative values.

More details in the recommended readings.

Potential limitation and misuse of ancestry calibration

- Cancelling out true ancestry effects on genetic values
  - think about differential selection pressures and causal variants being ancestry informative
- False assumptions of the genetic architectures across ancestries being homogeneous?
- Does this solve the more upstream problems in PRS model development (i.e. various levels of prediction accuracy across ancestries)?

Hands-on Exercise: Ancestry calibration

Return to the R Markdown

- This part of the tutorial demonstrates that score distribution varies by ancestry group and scores are correlated with PCs, then performs mean and variance calibration based on PCs

- Work through the rest of the code (from "Discussion Break" to the end)
  - All output can be seen in the HTML if you are unable to follow along
  - Talk to your neighbors if you get stuck

- In the final density plots, you should be able to see the by-ancestry distributions shift to almost completely overlap post-calibration

Recommended readings

- Reviews:
  - (tutorial) Choi et al. Nature Protocols. 2020
  - (application) Kachuri et al. Nature Review Genetics. 2024
- Ancestry calibration methods:
  - Khera et al. Circulation. 2019.
  - Wang et al. Journal of the American College of Cardiology. 2020.
  - Khan et al. Nature Medicine. 2022
  - Ge et al. Genome Medicine. 2022
  - Lennon et al. Nature Medicine. 2024