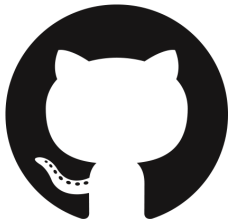
Meta-analysis of GWAS for LDL cholesterol

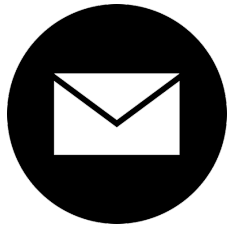
SMED8020

2 June 2022

Brooke Wolford, PhD

 @bnwolford

 https://github.com/bnwolford

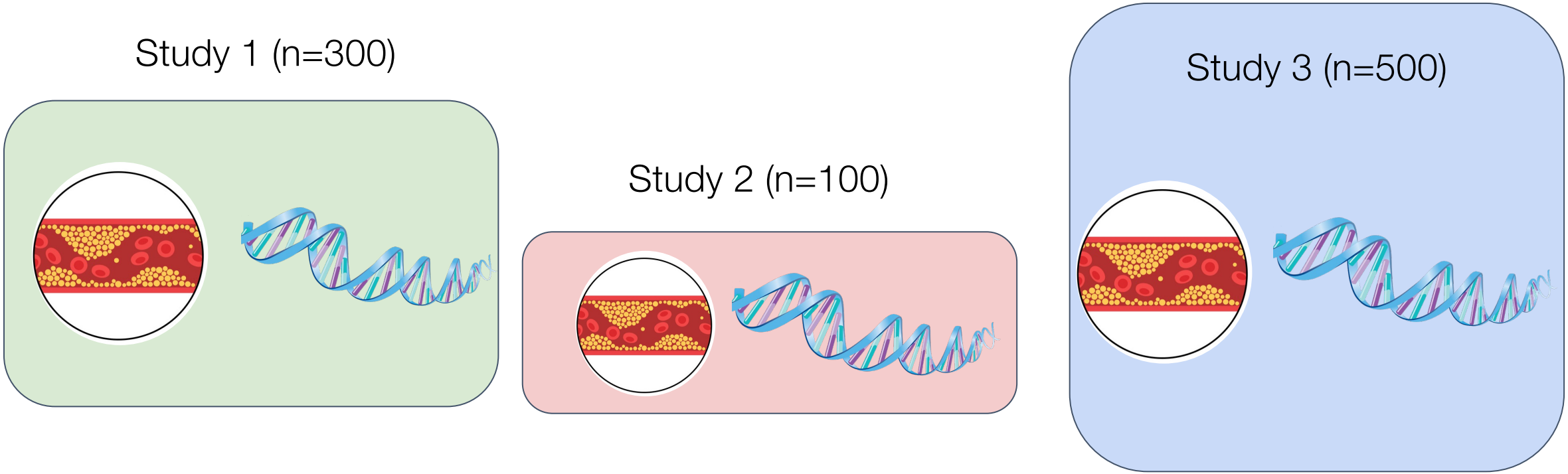
 brookewo@ntnu.no

NTNU
K.G. Jebsen Center for Genetic Epidemiology

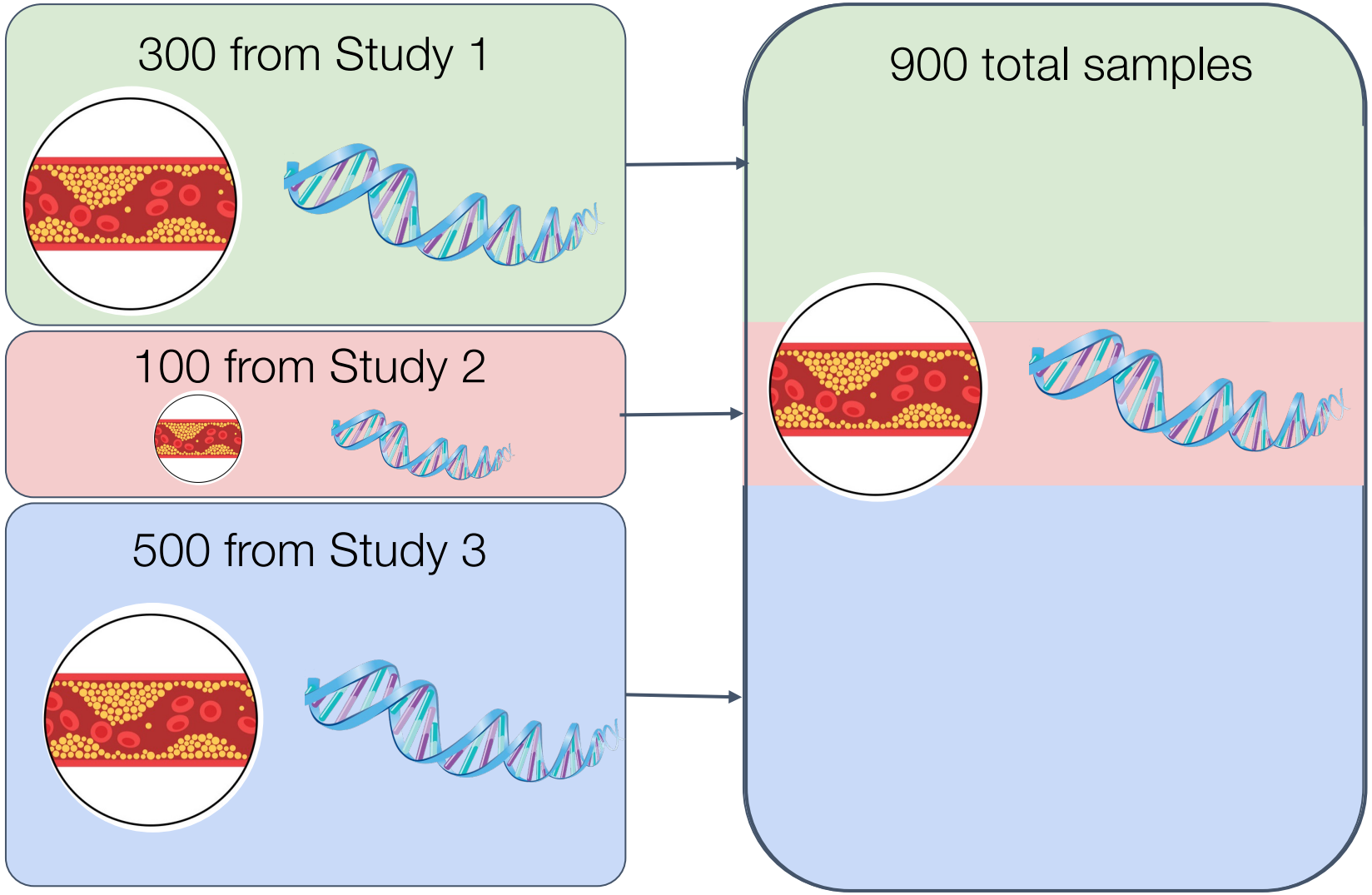
Slides revised from Sarah Hanks

Motivating example: LDL cholesterol and genetic variation

- 3 studies of various sample sizes
- Each collected same variables (LDL cholesterol, age, sex, genotypes)
- How do we use data from multiple studies to answer our research question?



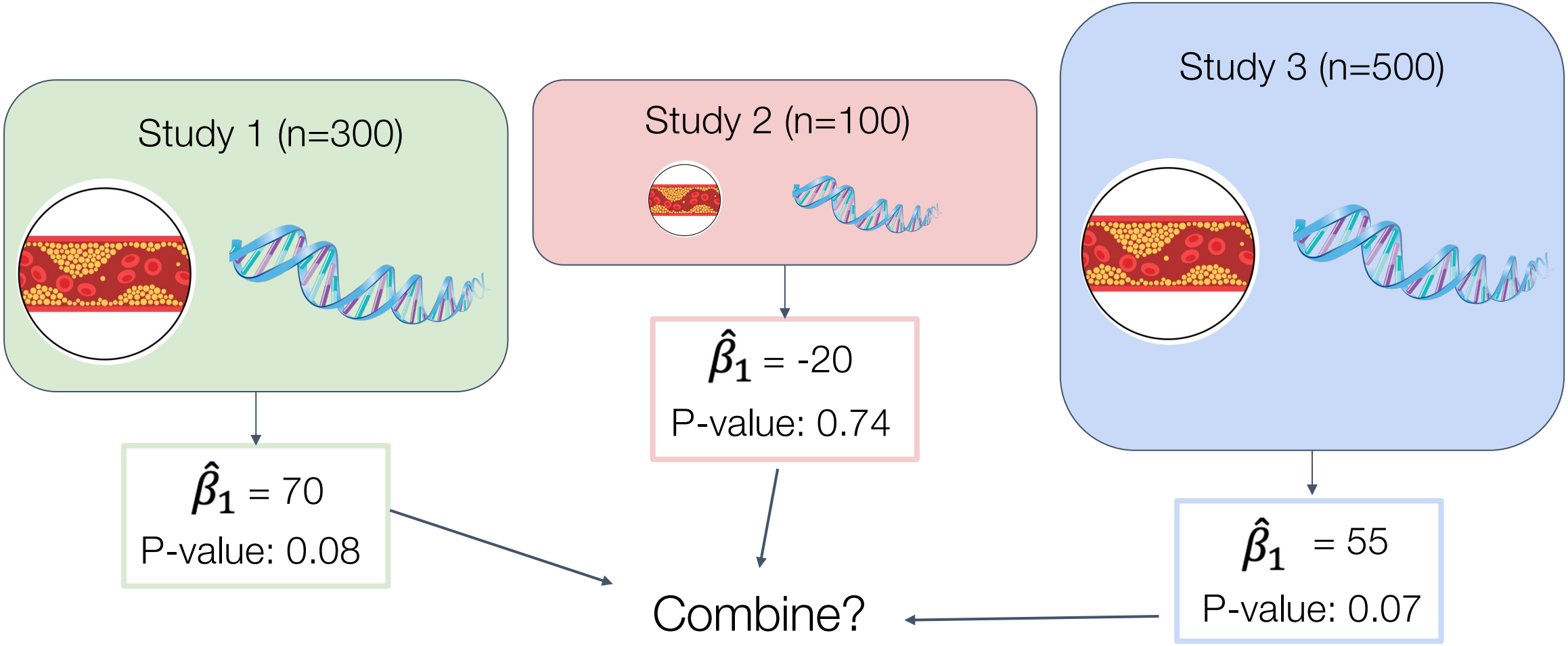
Mega-analysis pools individual-level data across studies



Fit model and estimate/ make inferences about parameters in the usual way

Factors we consider:

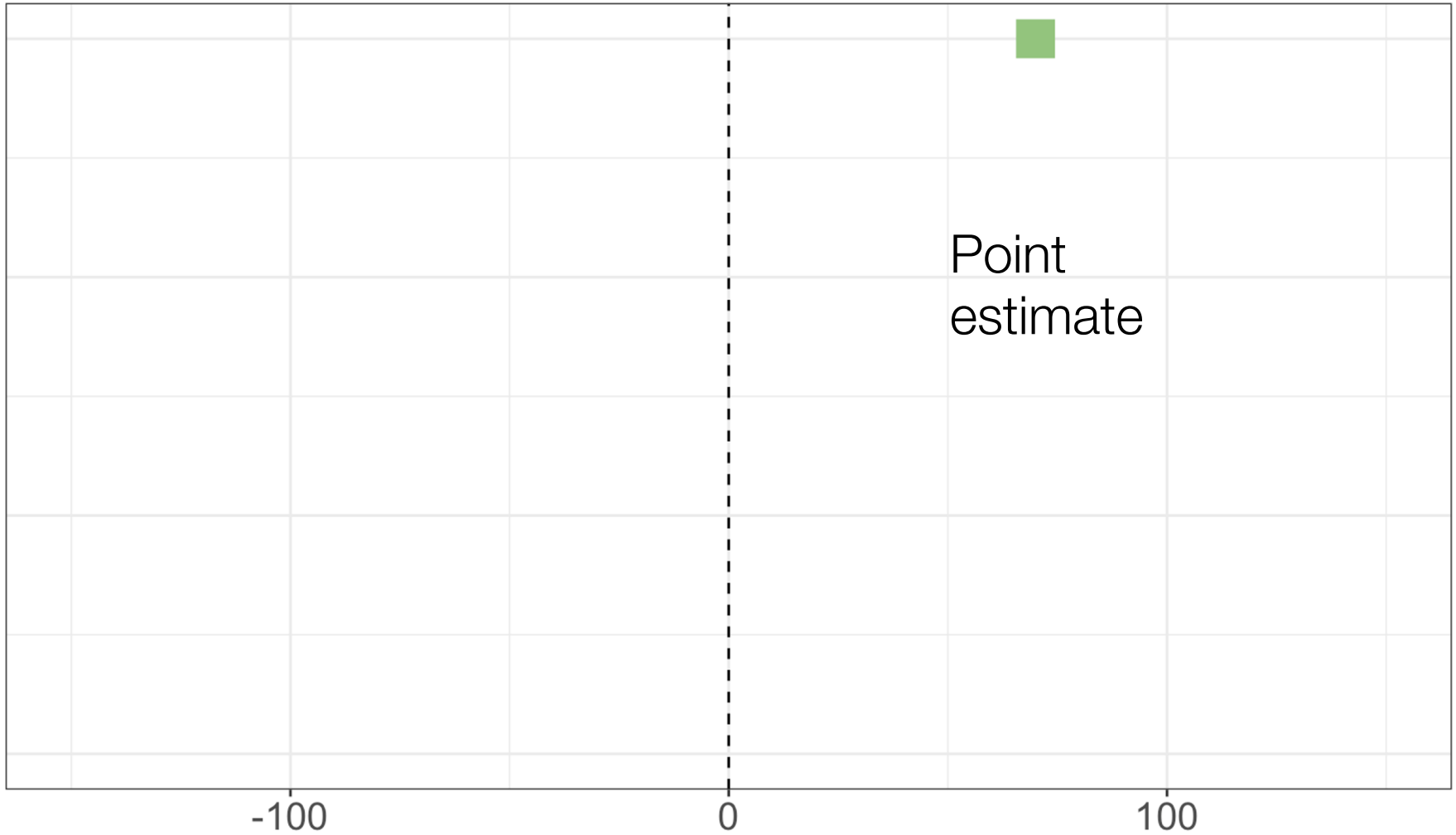
- Direction of effect and magnitude of $\boldsymbol{\beta}$ in each study
- Strength of associations (p-value) in each study
- Sample size



Why perform a meta-analysis?

- Resolve uncertainty between studies
- Improve effect size ($\boldsymbol{\beta}$) estimates
- Increase statistical power by increasing sample size (as effecitient as pooling individual left data)
- Identify patterns among the studies
- Analyze different ancestries separately before combining to avoid population stratification
- Avoid privacy concerns: meta-analysis is performed with summary statistics ($\boldsymbol{\beta}$, CI/P-value) and does not require access to individual level data

Forest plots help us visualize results from multiple studies

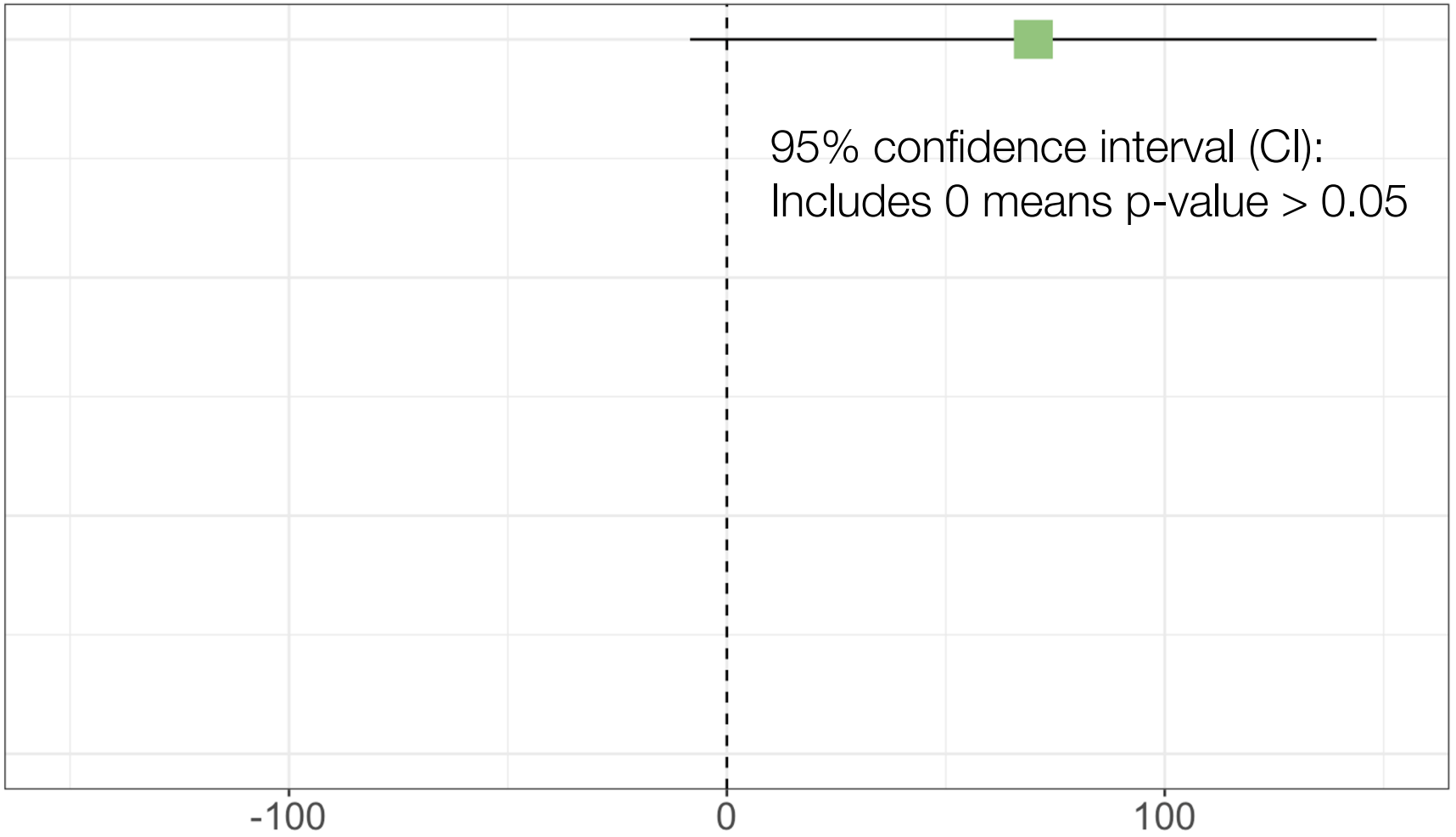


Study 1

β_1<0: Allele 1 associated with lower cholesterol

β_1>0: Allele 1 associated with higher cholesterol

Forest plots help us visualize results from multiple studies

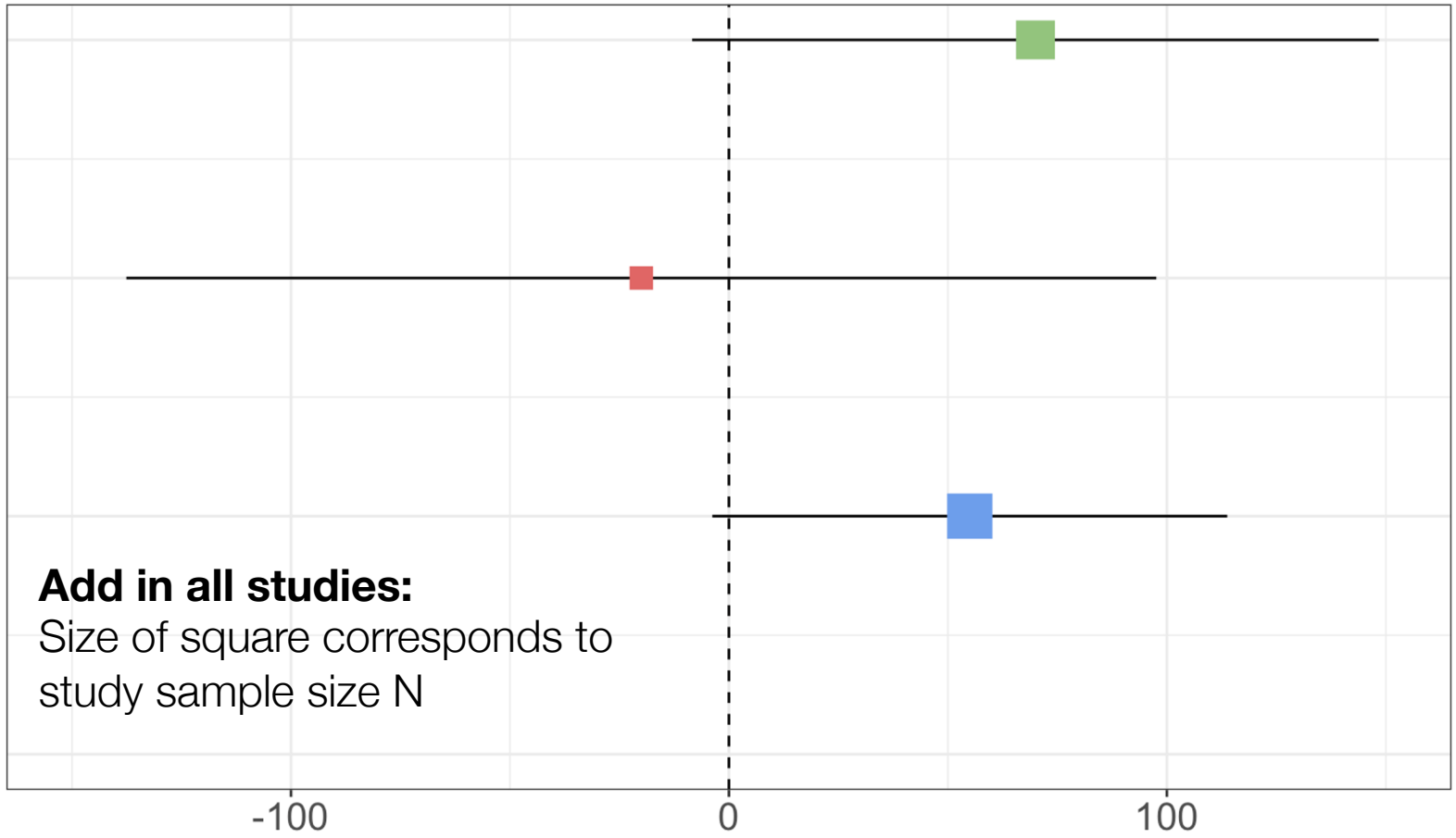


Study 1

β_1<0: Allele 1 associated with lower cholesterol

β_1>0: Allele 1 associated with higher cholesterol

Forest plots help us visualize results from multiple studies



Study 1

Study 2

Study 3

β_1<0: Allele 1 associated with lower cholesterol

β_1>0: Allele 1 associated with higher cholesterol

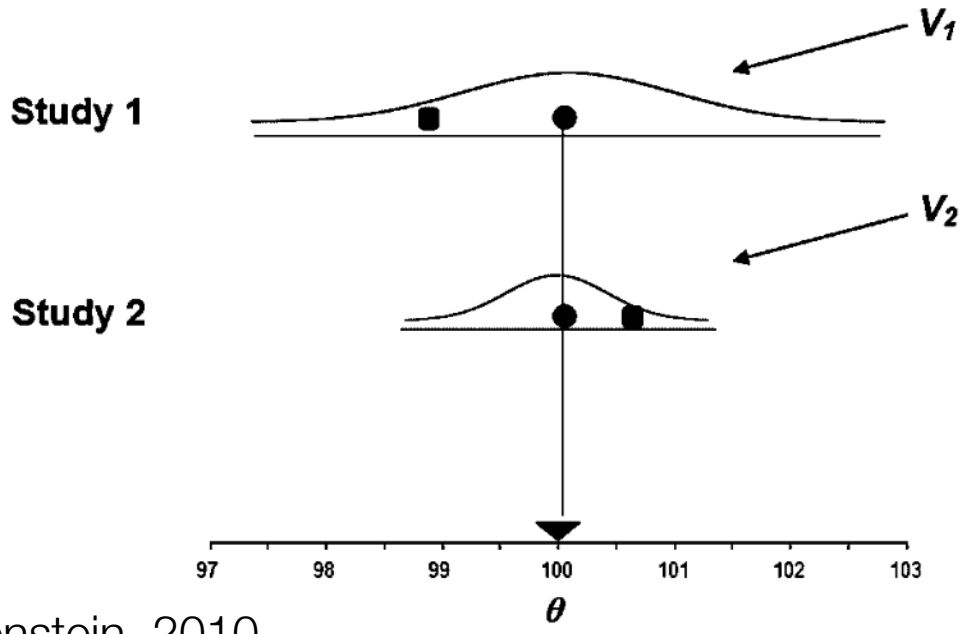
Meta-analysis assumptions

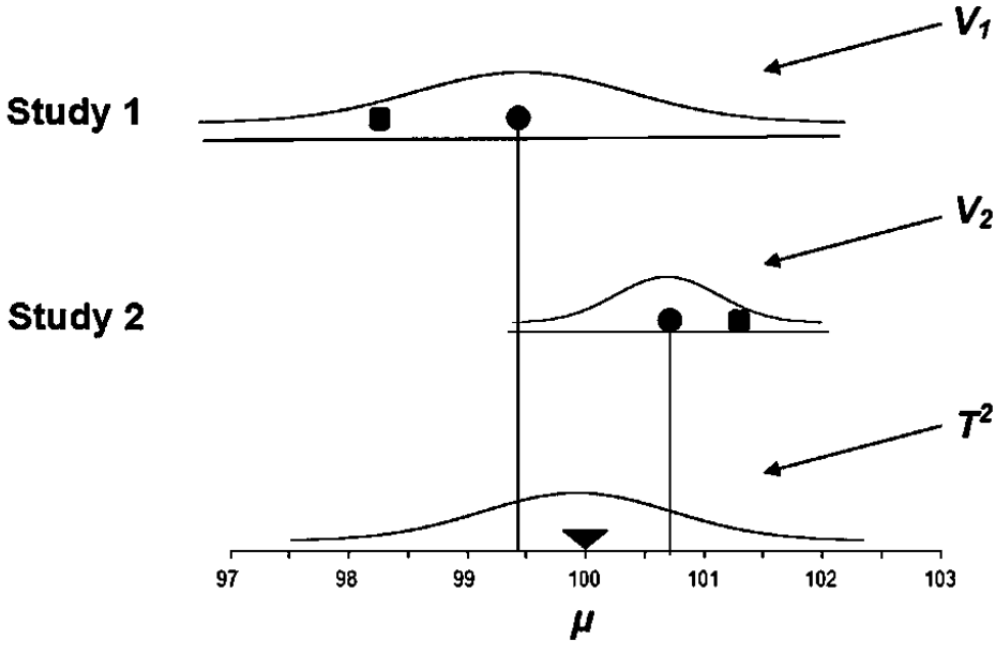
- Each sub-study has controlled type 1 error and proper QC
- No sample overlap between studies
  - Methods in development for this situation (https://genome.sph.umich.edu/w/images/7/7b/METAL_sample_overlap_method_2017-11-15.pdf)
- Fixed or random effects meta-analysis have their own assumptions
- Estimated beta value is in terms of the same allele, consistent direction

Zeggini, Pharmacogenomics, 2009

Fixed vs random effects

Fixed: one common true effect size (β) that underlies all the studies in the analysis
Random: distribution of true effect sizes, effect size in each study is different with means assumed to be chosen from Gaussian





Borenstein, 2010

Fixed effect model: calculation of weights

- Most common strategy to calculate weights for fixed effect meta-analysis is called inverse variance weighting
- Key idea: give larger weights to studies with more precise $\boldsymbol{\beta}$ estimates (usually larger sample size)

Weight assigned to i^{th} study:

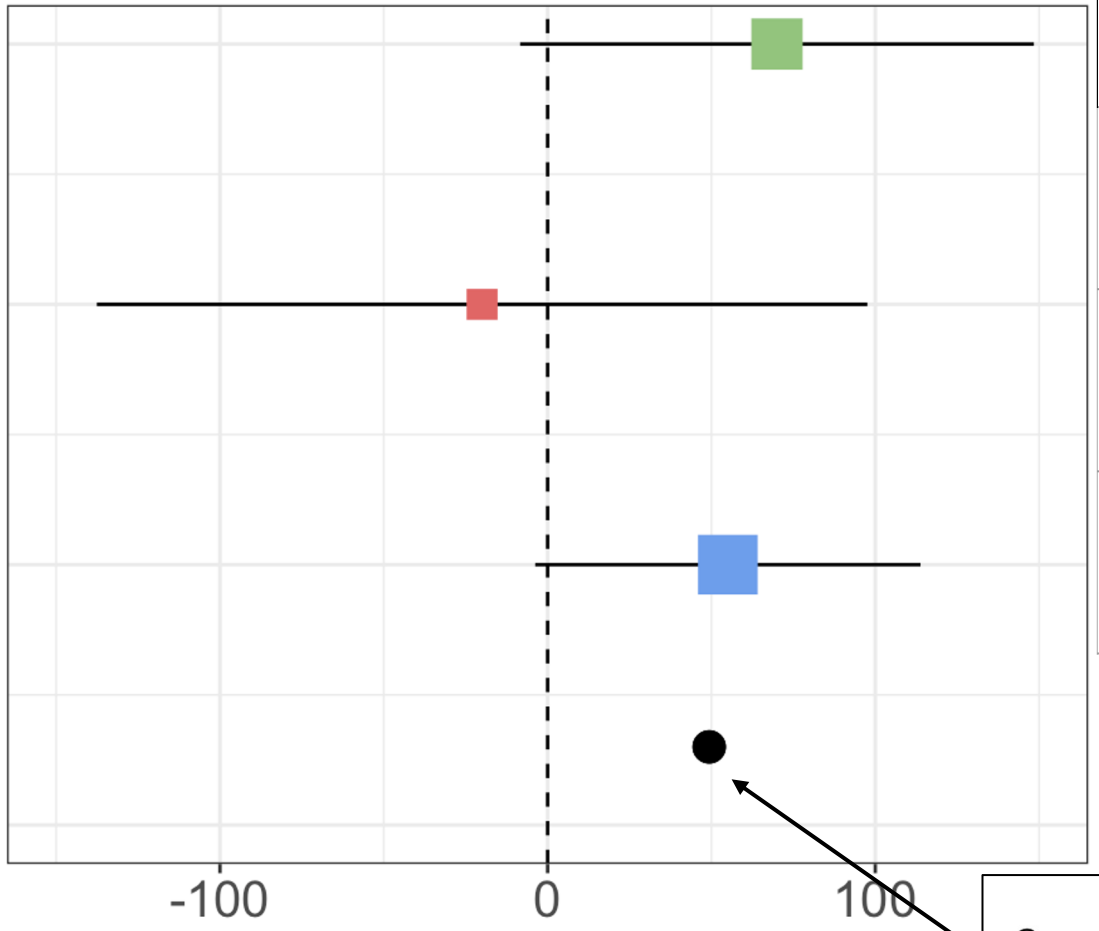
$$w_i = \frac{1}{var(\hat{\beta}_i)}$$

Weighted average:

$$\hat{\beta}_{pooled} = \frac{\sum w_i \hat{\beta}_i}{\sum w_i}$$

Fixed effect model: calculation of pooled estimate

Study	$\hat{\beta}_i$	$var(\hat{\beta}_i)$
1	70	1,600
2	-20	3,600
3	55	900



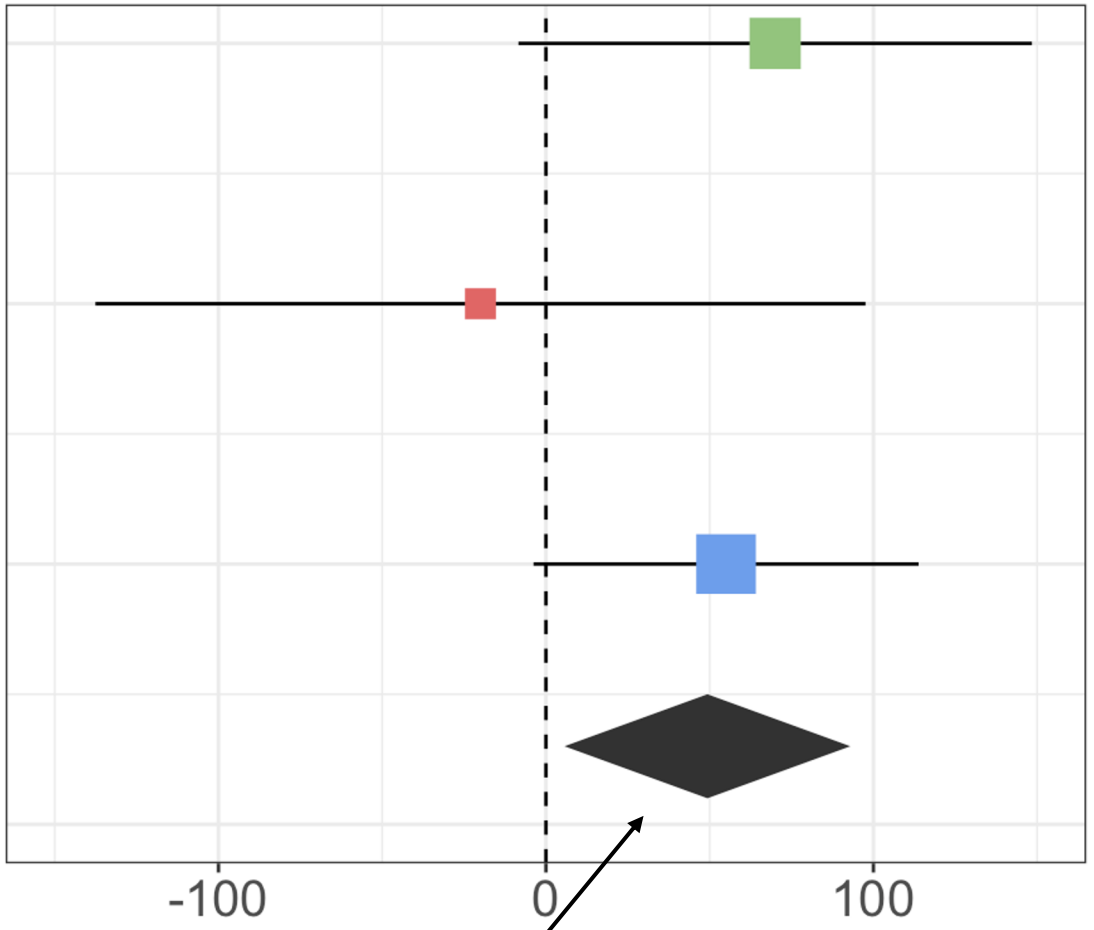
w_i	$w_i\hat{\beta}_i$
0.00063	0.044
0.00028	-0.0056
0.0011	0.061

$\sum w_i$ =0.002 $\sum w_i\beta_i$ =0.099

$$\beta_{pooled} = \frac{\sum w_i\beta_i}{\sum w_i} = \underline{49.31}$$

Fixed effect model: calculation of pooled confidence interval

$$se(\hat{\beta}_{pooled}) = \sqrt{\frac{1}{\sum w_i}}$$



None of these studies had significant evidence of an association alone

Combining evidence across studies shows significant positive association between Allele 1 and cholesterol

Represent pooled 95% CI as diamond

95% CI = (5.6, 93.0)

METAL

Table 1. Formulae for meta-analysis

	Analytical strategy	
	Sample size based	Inverse variance based
Inputs	N_i - sample size for study i P_i – P-value for study i Δ_i - direction of effect for study i	β_i- effect size estimate for study i se_i - standard error for study i
Intermediate Statistics	$Z_i = \Phi^{-1}\left(1-\frac{pi}{2}\right) * \text{sign}(\Delta_i)$ $w_i = \sqrt{N_i}$	$w_i = 1/\text{SE}_i^2$ $se = \sqrt{1/\sum_i w_i}$ $\beta = \sum_i \beta_i w_i / \sum_i w_i$
Overall Z-Score	$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$	$Z = \beta/SE$
Overall P-value	$P = 2\Phi(\lvert -Z \rvert)$	

Willer et al, 2010

Quantifying between-study heterogeneity

Calculated by METAL

- Cochran's Q statistic: test of homogeneity of the effect across cohorts
  - k=number of cohorts
  - χ^2_{k-1} test statistic

$$Q = \sum_{i=1}^{k} w_i (\beta_i - \beta_{pooled})^2$$

- I^2 statistic: percent of total variation across studies due to heterogeneity instead of chance
  - Ranges from 0-100%
  - I^2>50%: moderate heterogeneity
  - I^2>75%: high heterogeneity

$$I^2 = \max\left(0, \frac{Q - (k-1)}{Q} * 100\right)$$

Random effects meta-analysis

- In fixed effect meta-analysis, we assume that each study is estimating the same effect size
- Sometimes, studies address the same research question but still have important differences that influence effect sizes
  - Variation in outcome or covariate measurement
  - Variation in study populations
- Random effects meta-analysis allows for differences in the true effect sizes between different studies

Fixed effect vs random effects

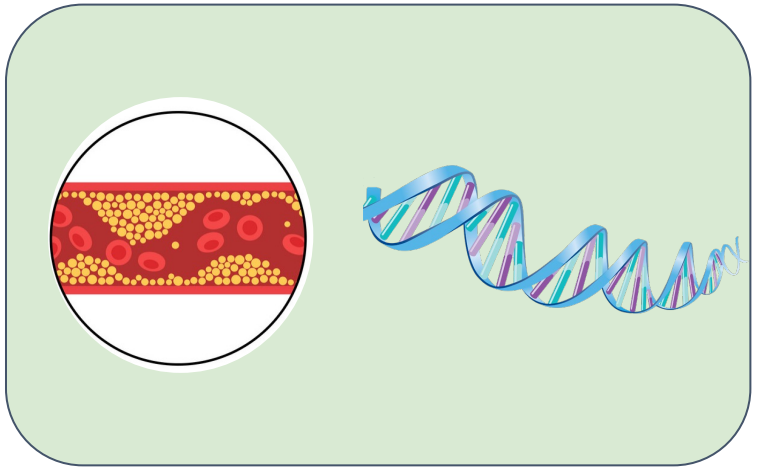
	Fixed effect	Random effects
Assumption	One true common effect across all studies	Study-specific effects come from a common distribution
Pooled effect estimate	True common effect	Mean of common distribution
Weights	Based on study variance	Based on both study variance and variance of common distribution

Other GWAS meta-analysis software

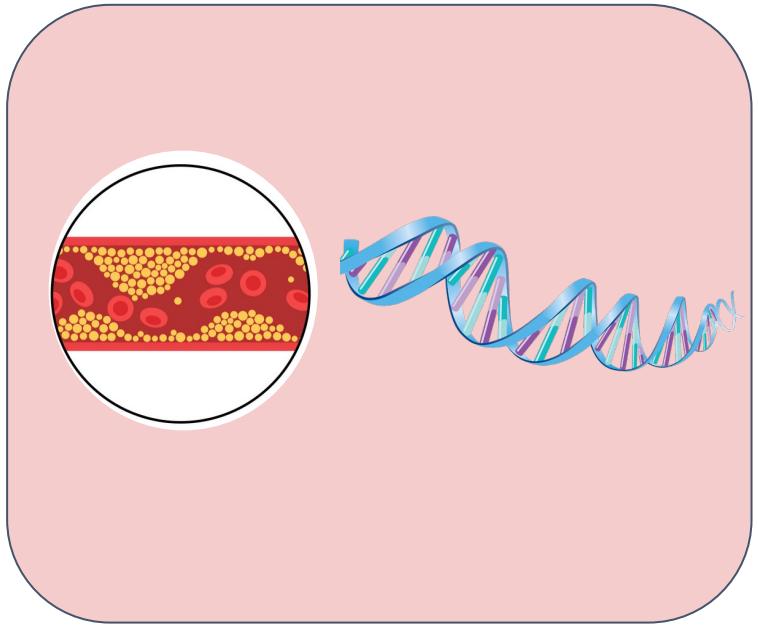
- GWAMA (Genome-Wide Association Analysis)
  - Mägi et al, BMC Bioinformatics, 2010
- METASOFT
  - Han et al, AJHG, 2011
- MANTRA (Meta-Analysis of Trans-ethnic Association Studies)
  - Morris et al, Genetic Epidemiology, 2011
- MR-MEGA (Meta-Regression of Multi-Ethnic Genetic Association)
  - Mägi, Hum Mol Gen, 2017
- SMetABF
  - Sun et al, PLOS,2022

Practical Day 3: meta-analysis of LDL cholesterol GWAS

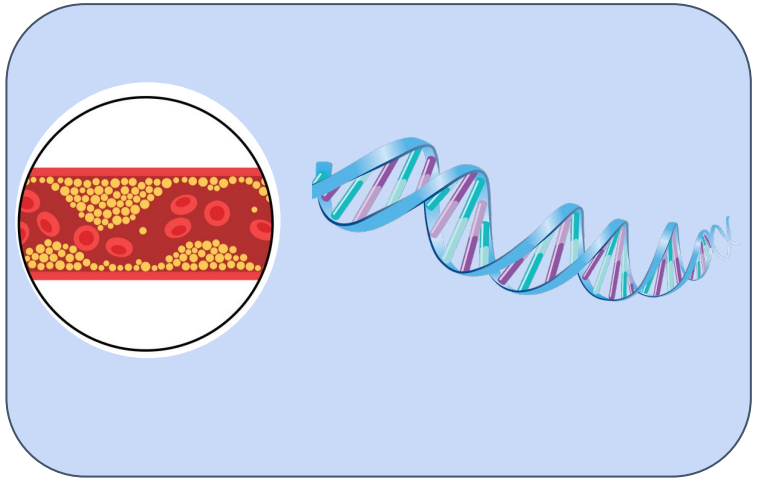
Biobank Japan (BBJ)
(n=72,866)



Global Lipids Genetics
Consortium (GLGC)
(n=89,138)



Trøndelag Health Study (HUNT)
(n=67,429)

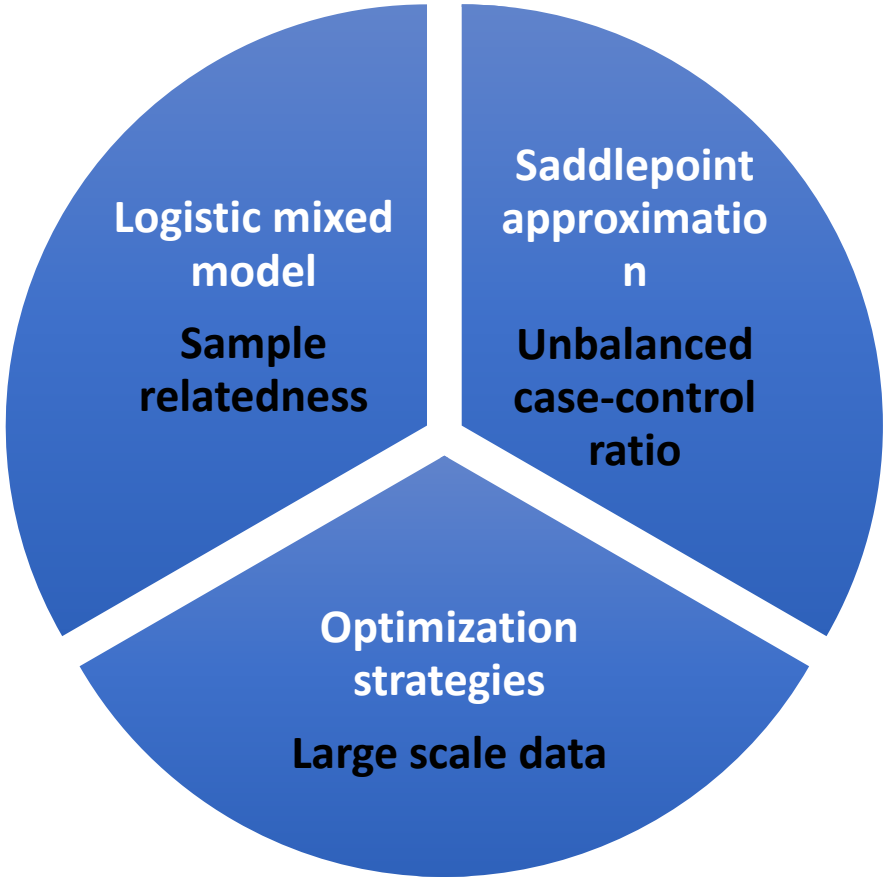


In choosing software for GWAS consider…

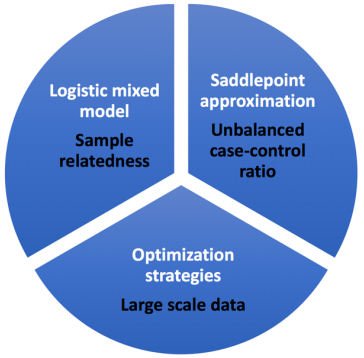
- Is your trait binary or quantitative?
  - Logistic vs linear regression
- Is your sample size large and what computational capability do you have?
  - Memory and time constraints
- Is your sample related?
  - Subjects are not independent
- Is your sample geographically/ancestrally homogeneous?
  - Population stratification

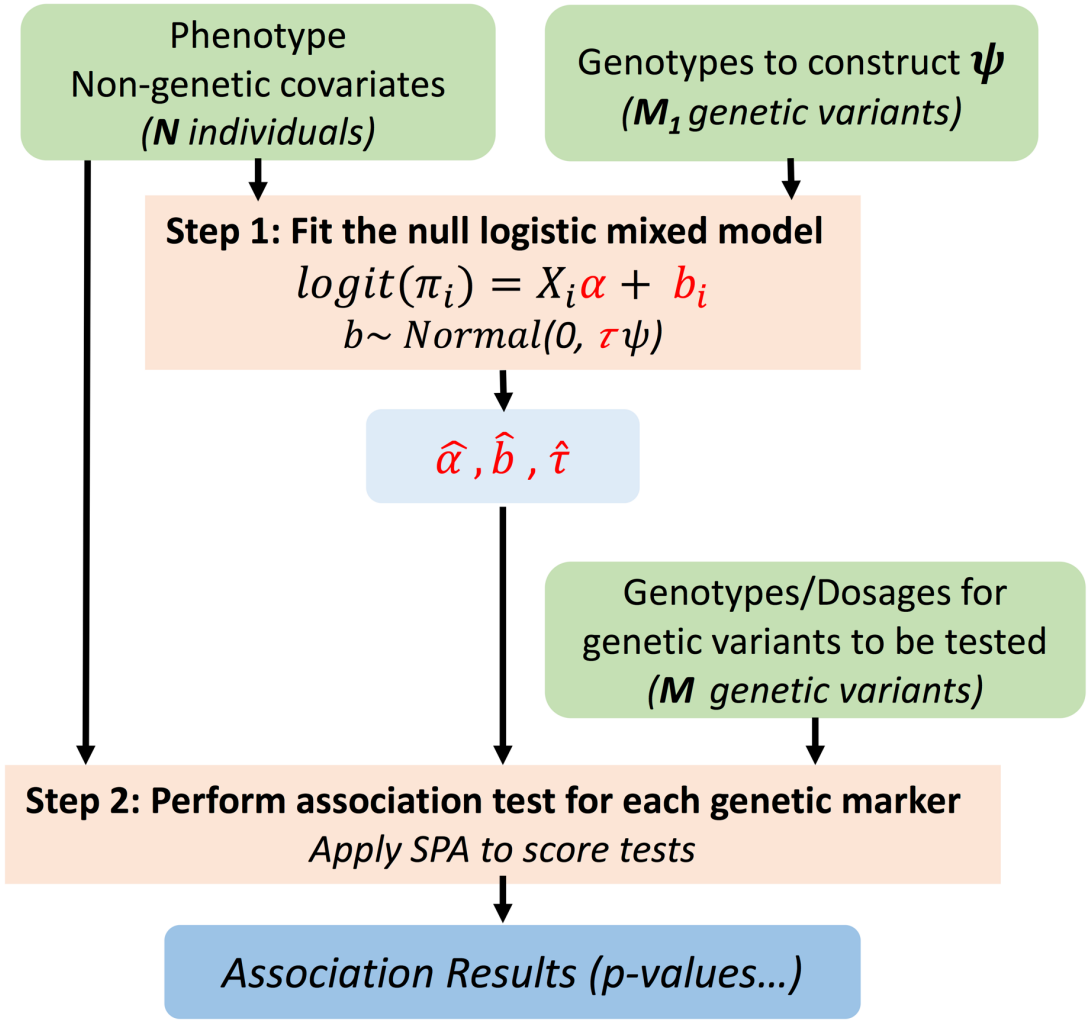
SAIGE

Scalable and Accurate Implementation of GEneralized mixed model



Wei Zhou, Nature Genetics, 2018





https://github.com/weizhouUMICH/SAIGE

Note: If a linear mixed model (e.g. BOLT-LMM) is used to test for association on an all-or-none trait, the estimate must be adjusted before meta-analysis with other logistic regression
Lloyd-Jones, et al 2018 Genetics