Introduction to Statistical Methods in Meta-Analysis

Jian Wang

October 31, 2013

What is meta-analysis?

- Meta-analysis: the statistical synthesis of information from multiple independent studies.
- Increase powers and reduce false-positive findings
- Advantages:
  - Results can be generalized to a larger population
  - Can use summary data (no sharing individual-level data)
  - The precision and accuracy of estimates can be improved
  - ……
- Pitfalls:
  - Sources of bias are not controlled by the method: a good meta-analysis of badly designed studies will still result in bad statistics
  - Publication bias: studies show negative or insignificant results are less likely to be published
  - ……

What is meta-analysis?

- General steps in meta-analysis:
  - Formulation of the problem
  - Literature search
  - Selection of studies: e.g. selection of specific studies on a well-specified subject
  - Decide summary measures or dependent variables: differences, means, OR, or relative risk
  - Statistical analysis
- History:
  - A historical instance of Meta-analysis dates back to the twelfth century in China, a famous philosopher, Chu Hsi (朱熹, 1130~1200), built up his philosophical theory by summarizing a series of related literatures. He called this research methodology 'Theory of Systematic Rule'(道統論).
  - Karl Pearson analyzed the data from five studies on the correlation between the vaccination for enteristic fever and its mortality.

Fisher's method

- Combine p values from independent tests bearing upon the same overall hypotheses:

$$\chi^2_{2m} = -2\sum_{j=1}^{m} \mathrm{Ln}(p_j)$$

- When the p values tend to be small, the test statistic will be large suggesting that the null hypotheses are not true for every test
- Under null (all null hypotheses are true) and when all p values are independent, it is a chi-squared distribution with 2m degrees of freedom.
- Extend to dependent tests
  - Scaled chi-squared distribution random variable
  - Brown's method: known covariance
  - Kost's method: unknown covariance

Z score method

- Limitations of combining p values
  - Combining p values may be spurious when the direction of effects in the combined studies is not consistent
  - Not straight forward to include weights
- Combine Z scores:

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$$

- w_i is the square root of sample size of the ith study.
- $Zi=\Phi^{-1}(1-pi)$

- Limitations
  - Can not provide an overall estimate of effect size
  - Can not address between-studies heterogeneity

Fixed Effect Model

- Most popular
- Weighted average of effect sizes from a series of studies

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \qquad V_M = \frac{1}{\sum_{i=1}^{k} W_i}$$

- Y_i is effect size of study i, such as logarithm of ORs, beta-coefficients, mean difference or standardized mean difference for a continuous phenotype
- The inverse of the studies' variance is commonly used as study weight, such that larger studies tend to contribute more.
- $w_i=1/v_i$, v_i is the variance of study i.
- $Z=M/\mathrm{sqrt}(V_M)$ is used to test the null hypothesis

Heterogeneity

- Fixed effect model:
  - Assumes all studies in the analysis share a common underlying true effect
  - All observed variance reflects sampling error within study
  - Weights are assigned with the goal of minimizing this within-study error.
- Heterogeneity:
  - Sources of heterogeneity:
    - Some phenotypes are difficult to define and standardize, e.g., behavioral traits
    - Effect size might be higher in studies when individuals are older, or more educated or healthier
    - Genetic studies: different ethnicity groups, different genotyping platform or imputation software
  - In this case, there may be different underlying true effect sizes for different studies

Test for heterogeneity

- Cochran's Q test:

$$Q = \sum_{i=1}^{k} W_i (Y_i - M)^2$$

$$Q = \sum_{i=1}^{k} \left(\frac{Y_i - M}{S_i} \right)^2$$

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i \right)^2}{\sum_{i=1}^{k} W_i}$$

- Under null, it is approximately distributed as a chi-square with k-1 degrees of freedom.
- Not powerful when number of studies is small or within-study variance is large
- It can not be used to estimate the magnitude of true variance

Quantifying heterogeneity

- I^2:

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%,$$

  - Expect value of Q on the assumption that all studies share a common effect size is df
  - Q-df is the excess variation. The part that will be attributed to differences in the true effects from study to study
  - Describes the percentage of total variation across studies that is due to heterogeneity rather than chance.
  - Not directly affect by the number of studies
  - A measure of inconsistency across the findings of the studies and not as a measure of the real variation across the underlying true effects
  - A value of 0% indicates no observed heterogeneity
  - Low, moderate, large and very large for 0-25%, 25-50%, 50-75% and >75%

Random effect model

- Incorporate the between-study variance
- DerSimonian-Laird method for between-study variance

$$T^2 = \frac{Q - df}{C} \qquad C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

- Total variance

$$V_{Y_i}^* = V_{Y_i} + T^2$$

- $w_i^*=1/v_i^*$, v_i^* is the total variance of study i.
- Random effect model

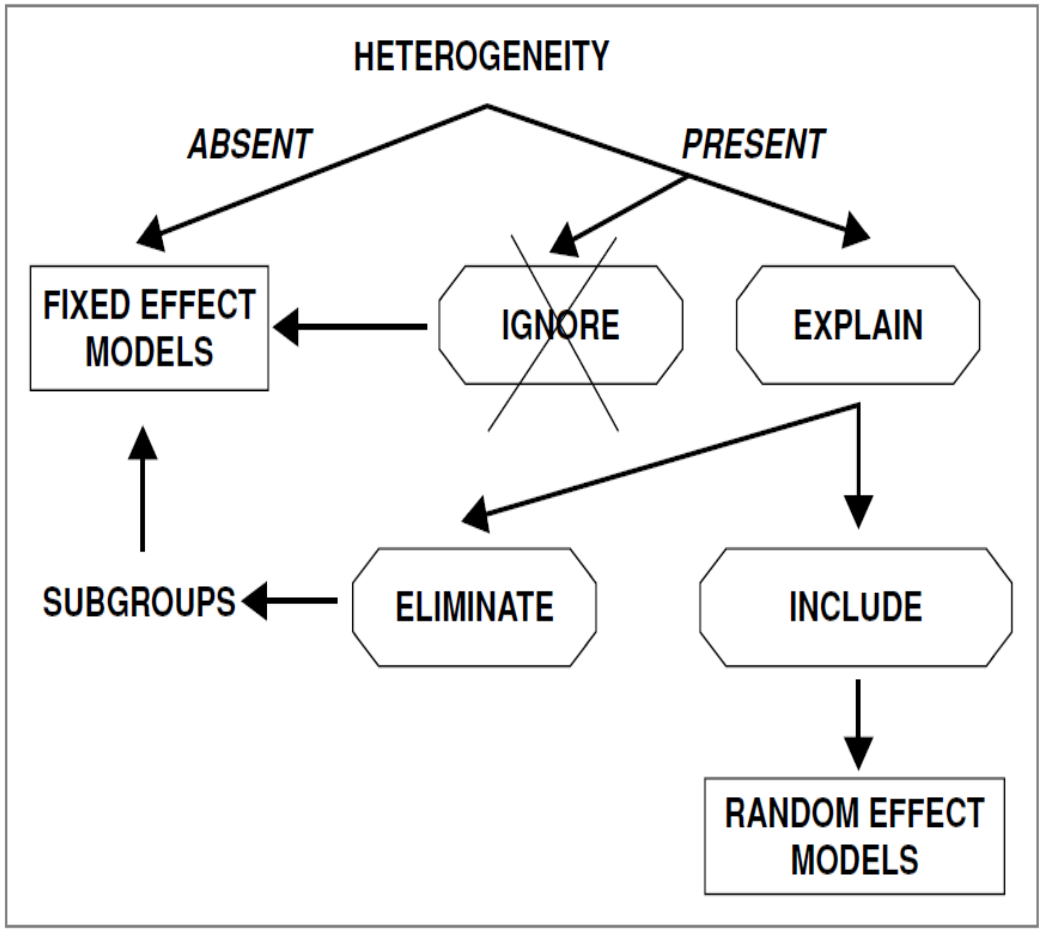
$$M^* = \frac{\sum_{i=1}^{k} W_i^* Y_i}{\sum_{i=1}^{k} W_i^*} \qquad V_{M^*} = \frac{1}{\sum_{i=1}^{k} W_i^*}$$

Fix or random?

- Difference opinions:
  - Random effect model: robust
    - Conservative
    - Between-studies variance have poor precision
    - Heterogeneity test is under powered
  - Fixed effect model: powerful
    - Not realistic
    - False positive increases
  - Bayesian approach for estimating between-studies variance from outside of the current set of studies.
    - Depend on the priors

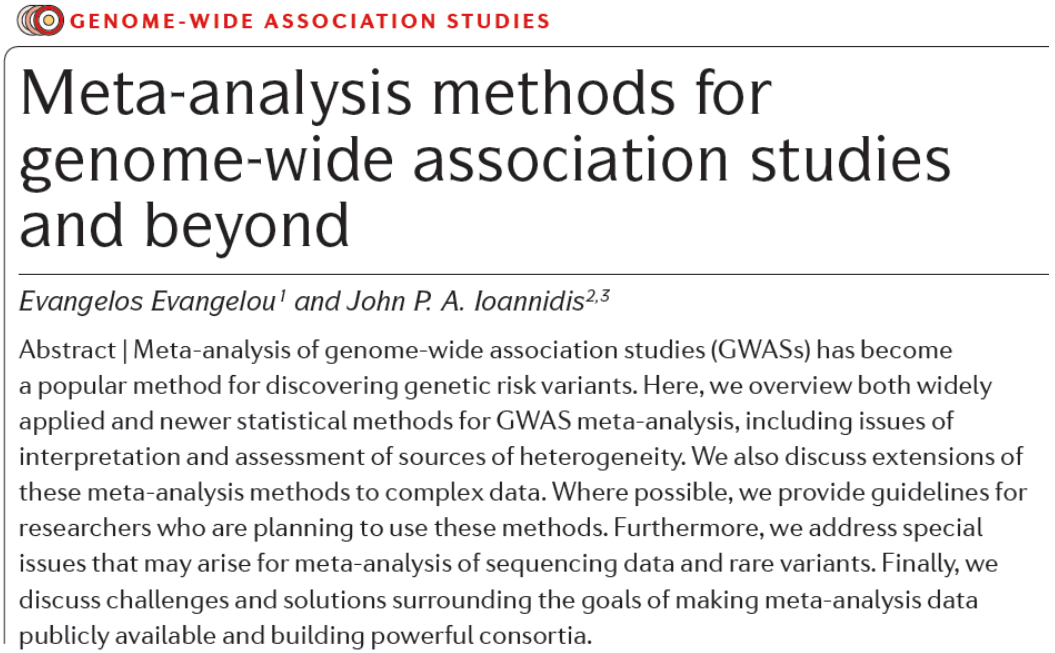
Fix or random?

- Start with a fixed effect model and then switch to a random effect model based on heterogeneity test?
  - Heterogeneity tests often suffers from low powers
  - Decision should be based on our understanding of whether or not all studies share a common effect size, and not on the outcome of a statistical test



Meta-analysis in GWAS

- Genetic effects due to common alleles are small, and detection of signals requires large sample sizes
- Single GWAS are underpowered due to the sample size
- Meta-analysis has become a popular approach for the discovery of new genetic loci for common diseases and phenotypes
- Several hundred GWAS meta-analyses have already been published.



GENOME-WIDE ASSOCIATION STUDIES

Meta-analysis methods for genome-wide association studies and beyond

Evangelos Evangelou[1] and John P. A. Ioannidis[2,3]

Abstract | Meta-analysis of genome-wide association studies (GWASs) has become a popular method for discovering genetic risk variants. Here, we overview both widely applied and newer statistical methods for GWAS meta-analysis, including issues of interpretation and assessment of sources of heterogeneity. We also discuss extensions of these meta-analysis methods to complex data. Where possible, we provide guidelines for researchers who are planning to use these methods. Furthermore, we address special issues that may arise for meta-analysis of sequencing data and rare variants. Finally, we discuss challenges and solutions surrounding the goals of making meta-analysis data publicly available and building powerful consortia.

Box 1 | **Stages in a GWAS meta-analysis**

Setting up an analysis plan

Each genome-wide association study (GWAS) meta-analysis initiative should be based on strong collaborative agreements and should be carefully designed and organized. An analysis team should design and draft a detailed plan that explicitly describes all of the steps of the anticipated analysis. Independent performance of some core statistical analyses by at least two analysts and using different software is not uncommon and may also allow for cross-verification and quality checks of processes and results. The analysis plan should be adopted by all teams, which should try to avoid deviations that introduce unnecessary between-study heterogeneity.

Dealing with heterogeneity

Despite careful planning to avoid heterogeneity, sometimes differences are inevitable, even in prospective designs: for example, when some samples have a family structure or when designs include extreme values[13]. Also, differences in phenotype definition may affect the estimated magnitude of the genetic effects, a factor that needs to be considered in terms of optimizing power for discovering new associations[6]. Ideally, phenotype definitions should be standardized according to stringent definitions applied in all data sets; if perfect standardization is impossible, participating teams should decide what kind of harmonization of definitions is desirable and feasible[14]. Inclusion and exclusion criteria of subjects and variants should be described in detail. Popular exclusion thresholds are >5% missing data, $P < 10^{-5}$ for Hardy–Weinberg equilibrium and quality index <0.3 for imputation metrics (BOX 2 summarizes the challenges of the imputation efforts using 1000 Genome Project panels[2,15]). Also, strand issues should be considered during quality control for the proper alignment of the alleles. Most GWAS meta-analyses to date focus on common variants and exclude variants with minor allele frequency <1%. However, this is likely to change as low-frequency and rare variants become the focus of interest. Statistical methods that account for between-study heterogeneity introduced by various sources are described in the main text.

Data storage

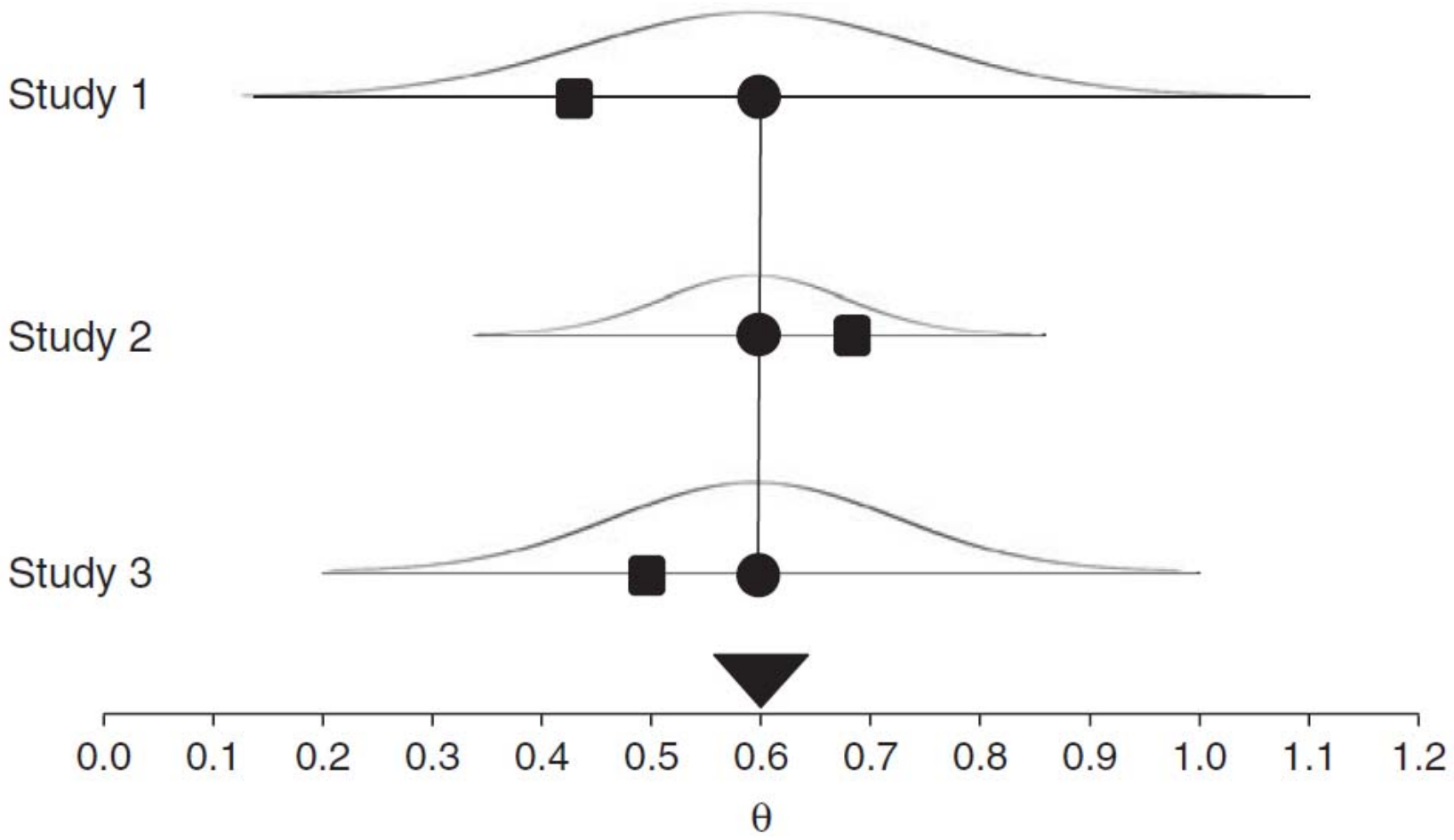
Data storage is an important aspect of meta-analysis as the individual-level data collected by each partner and also single participants' genotypes should be kept secured and unidentifiable. Most collaborative meta-analyses use online storage options to deposit summary data, giving access to members of the analysis team. This enables the partners to retain control of individual-level primary data. In most settings, summary data are statistically as

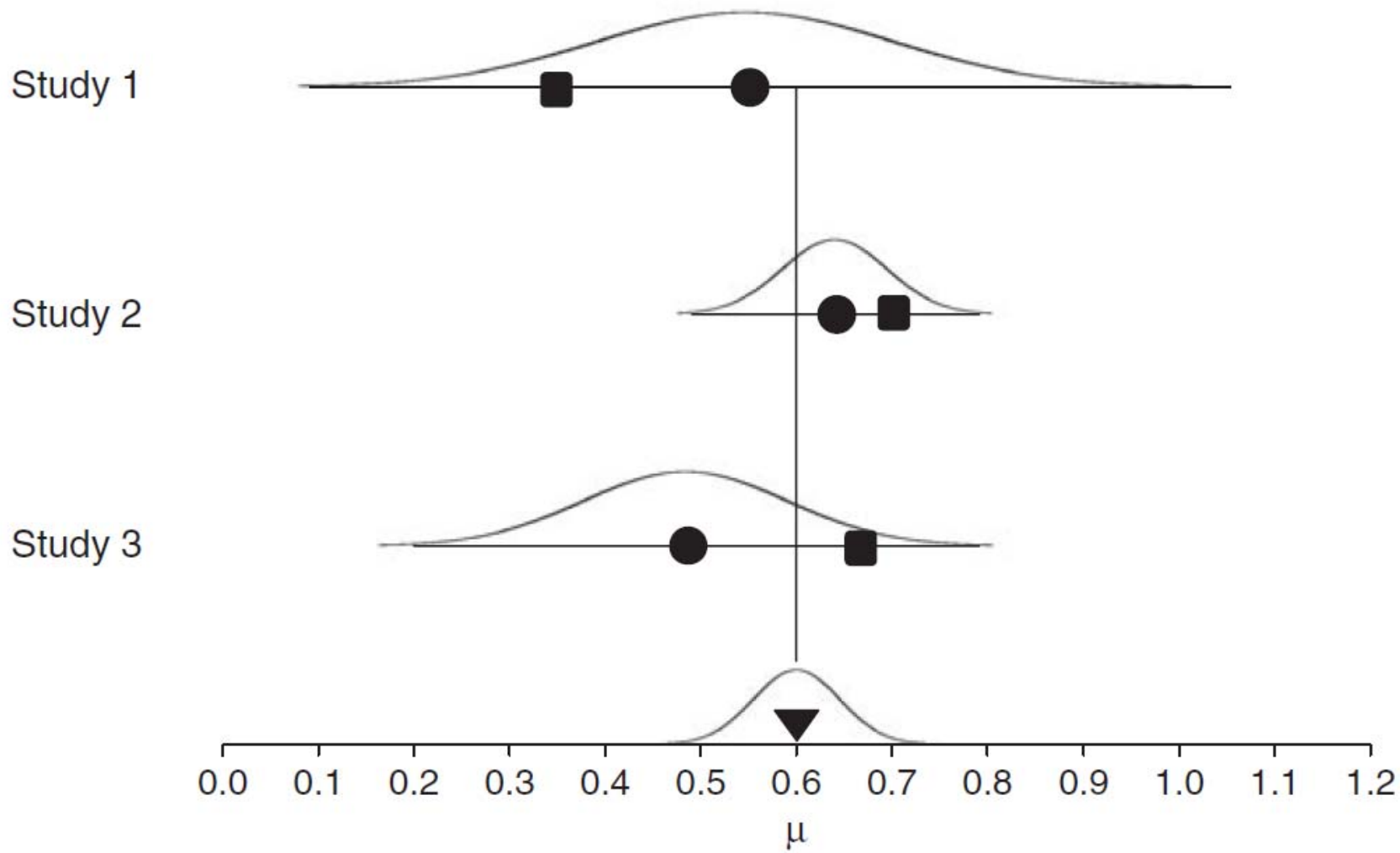
Method	Description	Advantages	Disadvantages	Main software used
P value meta-analysis	Simplest meta-analytical approach	Allows meta-analysis when effects are not available	Direction of effect is not always available; inability to provide effect sizes; difficulties in interpretation	METAL, GWAMA, R packages
Fixed effects	Synthesis of effect sizes. Between-study variance is assumed to be zero	Effects readily available through specialized software	Results may be biased if a large amount of heterogeneity exists	METAL, GWAMA, R packages
Random effects	Synthesis of effect sizes. Assumes that the individual studies estimate different effects	Generalizability of results	Power deserts in discovery efforts; may yield spuriously large summary effect estimates when there are selection biases	GWAMA, R packages
Bayesian approach	Incorporates prior assessment of the genetic effects	Most direct method for interpretation of results as posterior probabilities given the observed data	Methodologically challenging; GWAS-tailored routine software not available; subjective prior information used	R packages
Multivariate approaches	Incorporates the possible correlation between outcomes or genetic variants	Increased power can identify variants that conventional meta-analysis do not reveal using the same data sets	Computationally intensive; software not available for all analyses; some may require individual-level data	GCTA for multi-locus approaches
Other extensions	A set of different approaches that allows for the identification of multiple variants across different diseases	Summary results of previous meta-analyses can be used	May need additional exploratory analyses for the identification of variants; prone to systematic biases	Software developed by the authors of the proposed methodologies

	METAL	GWAMA	MetABEL	PLINK	R packages
Ability to process files from GWAS analysis tools; software used	No	Yes; SNPTEST, PLINK	Yes; ABEL	Yes; PLINK	No
Fixed effects implemented?	Yes	Yes	Yes	Yes	Yes
Random effects implemented?	No	Yes	No	No	Yes
Heterogeneity metrics generated	Q, I^2	Q, I^2	Q, I^2	Q, I^2	Q, I^2
Graphical illustration of meta-analysis results	No	Manhattan and QQ plots	Forest plots	No	Yes

- Most studies so far used fixed effect model: p values are important for the publication
- Start with a fixed effects model but to report the random effects model when heterogeneity is found.

```
CHR       Chromosome code
BP        Basepair position
SNP       SNP identifier
A1        First allele code
A2        Second allele code
N         Number of valid studies for this SNP
P         Fixed-effects meta-analysis p-value
P(R)      Random-effects meta-analysis p-value
OR        Fixed-effects OR estimate
OR(R)     Random-effects OR estimate
Q         p-value for Cochrane's Q statistic
I         I^2 heterogeneity index (0-100)
```


Study 1
Study 2
Study 3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
θ


Study 1
Study 2
Study 3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
μ