The Pharma Innovation Journal 2023; 12(5): 4331-4334

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(5): 4331-4334
© 2023 TPI
www.thepharmajournal.com
Received: 22-03-2023
Accepted: 27-04-2023

**Ymberzal Koul**
National Dairy Research Institute, Karnal, Haryana, India

**Gaurav Patel**
National Dairy Research Institute, Karnal, Haryana, India

**Corresponding Author:**
**Ymberzal Koul**
National Dairy Research Institute, Karnal, Haryana, India

# Meta-analysis of GWAS summary statistics: Unveiling the power of collective insights

## Ymberzal Koul and Gaurav Patel

**Abstract**
This review article presents a comprehensive overview of meta-analysis methodologies and their applications in the context of Genome-Wide Association Studies (GWAS). Over the past decade, meta-analysis has emerged as a powerful tool for harnessing the collective power of multiple independent GWAS studies to unravel the genetic architecture underlying complex traits and diseases. Through the integration and analysis of summary statistics from diverse datasets, meta-analysis offers valuable insights into the identification of hidden loci, replication and validation of genetic associations, and the characterization of polygenic effects. In this article, we explore the fundamental principles of conducting meta-analyses in GWAS, including the importance of data harmonization, quality control, and addressing potential sources of heterogeneity. We discuss the various statistical methodologies commonly employed in meta-analysis, such as fixed-effects and random-effects models, as well as novel approaches for accounting for heterogeneity and identifying gene-gene interactions. Additionally, we highlight advancements in pathway and functional analysis, which help elucidate the biological mechanisms underlying the observed associations.

**Keywords:** GWAS, meta-analysis, breed improvement, molecular marker, METAL

## Introduction
Genetic improvement in a herd is a careful and strategic process which requires selection of elite stock that serve as parents of the next generation. Traditionally, selection was primarily based on external appearance, whereby animals exhibiting the desired trait were selected and bred. The rationale behind 'breed the best to the best' was aimed towards increasing the frequency of the desired allele in the population. However, this traditional method of selection could not capture the full extent of genetic potential, especially with regard to traits which are controlled by many genes. Moreover, considerable time is taken for noticeable improvement, especially for quantitative traits, which can slow the genetic progress. In recent years, researchers are focused on identifying specific loci and molecular markers associated with desirable traits. Various types of molecular markers (blood polymorphism, enzyme polymorphism, restriction fragment length polymorphism, amplified fragment length polymorphism, minisatellites, microsatellites, single nucleotide polymorphism, etc.) have extensively been used in genetic mapping and molecular assisted selection. Amongst all molecular markers, SNPs have become a popular tool due to high polymorphism and genome-wide coverage.
Molecular breeding techniques, including genomic selection and marker assisted selection, have enabled animal breeders to make decisions by considering the animal's genetic potential. Marker assisted selection utilises specific genetic markers which are genotyped in a population for breeders to selectively identify animals carrying the favourable alleles. Genomic selection, on the other hand, utilises high-thorough put genotyping technologies to assess many markers and then predict the breeding value for specific traits. These methods allow the selection and retention of elite animal in the breeding herd at a young age. Molecular breeding techniques are complemented by genome-wide association studies which provide insight into the genetic basis of complex traits. GWAS involve analyzing a vast number of genetic markers across the entire genome to identify associations between specific markers and phenotypic traits of interest.
GWAS have played a pivotal role in the identification and validation of common gene variants associated with various common diseases and phenotypes of interest. However, the effect sizes of these genetic variants are generally modest or small in magnitude. Individual GWAS, despite their large sample sizes, often lack sufficient statistical power to detect these

associations with confidence. To overcome this limitation, researchers have turned to meta-analysis as a powerful tool. Meta-analysis involves systematically combining data from multiple independent GWAS to increase the overall sample size and enhance statistical power. By pooling together information from diverse datasets and study populations, meta-analysis enables researchers to identify additional genetic associations that may have been missed in individual studies. Moreover, meta-analysis allows for the exploration of heterogeneity of these associations across different populations and datasets. By evaluating the robustness of associations across diverse study samples, researchers can assess the reproducibility of findings. This information is essential for establishing the credibility and reliability of the identified genetic associations.

This review focuses exclusively on meta-analysis of GWAS summary statistics. Meta-analysis of GWAS summary statistics is a specialized approach that combines and analyses the results from multiple GWAS without accessing individual-level data. By leveraging the summary statistics, such as effect sizes, standard errors, and allele frequencies, researchers can gain valuable insights into the collective genetic architecture of complex traits and diseases.

## Steps involved in meta-analysis of GWAS summary statistics

### 1. Search strategy and data collection

The authors should collectively decide the research question and criteria for inclusion/exclusion of data. Eligibility criteria may be decided based upon the study design (case-control vs cohort), number of samples taken for GWAS, impact factor of the journal, year of study, language etc. At least two authors should independently identify the relevant studies from various databases like PubMed, MEDLINE, Embase, Web of Science, Google Scholar; and any arguments regarding inclusion or exclusion of the study from the meta-analysis should be taken by mutual discussion. The source of data may be research articles, conference proceedings, doctoral or master's thesis submitted to universities, trial registrations, data from open repositories, etc. However, conference proceedings are not peer reviewed and hence may pose error. But they offer the advantage of reducing publication bias (discussed later). Boolean operators 'AND', 'OR'and 'NOT' can be used to refine the search.

After the data is collected, it must be reviewed by the authors and duplicates should be removed. GWAS summary statistics should be collected on Microsoft Excel spreadsheets. Information that must be collected are – name of the study, country where the study was conducted, sample size, breeds of animals, p-value, effect size, standard error, minor allele frequency, reference and alternate allele.

### 2. Publication bias

When conducting a meta-analysis, it is important to consider the potential impact of publication bias on the results. Publication bias refers to the tendency of studies with significant or positive results being more likely to be published, while studies with nonsignificant or negative results may remain unpublished or hidden in the "file drawer." This can lead to a nonrepresentative set of studies in the meta-analysis dataset, which may introduce bias towards significance or positivity. Publication bias is a significant concern, especially when the meta-analysis relies solely on published scientific literature. It is crucial to acknowledge that studies may be suppressed or remain unpublished for various reasons beyond the failure to meet publication criteria. Various methods have been suggested to identify the presence of publication bias in meta-analyses.

- **Funnel Plot:** One widely used approach is the funnel plot, which was initially introduced in 1984 [11]. This graphical method, along with related techniques, allows researchers to visually assess the potential asymmetry in the distribution of study results and identify any missing studies that may indicate publication bias [7].
- **Egger's test:** Egger's test is a statistical test that quantitatively assesses funnel plot asymmetry. It measures the relationship between study size (or precision) and effect size and provides an estimate of the intercept. A significant intercept suggests the presence of publication bias [6].
- **Trim-and-Fill Method:** The trim-and-fill method is used to estimate the potential number of missing studies due to publication bias. It evaluates the asymmetry in the funnel plot and imputes hypothetical missing studies to create a symmetrical distribution. The analysis provides adjusted effect estimates by incorporating the imputed studies [5].
- **Duval and Tweedie's Trim-and-Fill Method:** Similar to the trim-and-fill method, Duval and Tweedie's trim-and-fill method identifies and imputes potential missing studies to assess publication bias. It provides both adjusted effect estimates and a sensitivity analysis that accounts for the potential impact of missing studies [4].
- **Begg's Test and Rank Correlation Test:** Begg's test and the rank correlation test (also known as Kendall's test) are non-parametric methods used to evaluate publication bias. These tests examine the correlation between effect sizes and their corresponding variances or study ranks. Significant results suggest the presence of publication bias [1].

### 3. Heterogeneity Analysis

Heterogeneity analysis is a crucial step in meta-analysis that aims to assess the variability or diversity of effect sizes across studies. It helps determine whether the observed differences in results are beyond what would be expected due to chance alone. Detecting and understanding heterogeneity is important as it can impact the interpretation and generalizability of the meta-analysis findings [9]. There are several methods available to detect and quantify heterogeneity in meta-analyses:

- **Visual Inspection:** One of the initial steps in assessing heterogeneity is visually inspecting the forest plot, which displays the effect sizes of individual studies along with their confidence intervals. If the confidence intervals of the effect sizes overlap widely, it suggests heterogeneity. Conversely, if they mostly overlap or cluster closely together, it indicates homogeneity.
- **Cochran's Q Test:** Cochran's Q test is a statistical test used to evaluate the presence of heterogeneity. It compares the observed differences in effect sizes across studies to the differences expected due to sampling error alone [15]. A significant Q statistic (p-value $< 0.05$) suggests the presence of heterogeneity.
- **$I^2$ Statistic:** The $I^2$ statistic quantifies the proportion of total variation in effect estimates that can be attributed to heterogeneity rather than chance. It ranges from 0% to 100%, where higher values indicate greater

heterogeneity. Generally, $I^2$ values of 25%, 50%, and 75% are considered low, moderate, and high levels of heterogeneity, respectively [9, 2].

- **Sensitivity Analysis:** Sensitivity analysis involves systematically varying the inclusion criteria or analytical methods to assess the robustness of the meta-analysis findings. By excluding studies with high risk of bias, or re-analyzing the data using different statistical models or methods, researchers can evaluate the impact of individual studies or methodological choices on the overall results [3].

**4. Undertaking meta-analysis**

There are two commonly used statistical models for conducting meta-analyses: the fixed-effect model and the random-effects model. In the fixed-effect model, it is assumed that there is a single true effect size which is responsible for all the studies that are included in the meta-analysis. If there is any observed difference in the effect size, that is attributed to sampling error. This model is also known as the common-effect model because it assumes a singular true effect.

The random-effects model allows for the possibility that the true effect sizes may vary across studies. It is possible that some of the studies may share a common effect size, but it is also plausible that different studies may have different effect sizes. In a random-effects meta-analysis model, the effect sizes observed in the conducted studies are assumed to represent a random sample from a specific distribution of these effect sizes. Consequently, the term "random effects" is employed to indicate the presence of multiple true effects.

**5. Softwares for meta-analysis**

There are several software packages available for conducting meta-analysis of genome-wide association study (GWAS) statistics. Here are some commonly used software tools:

- **METAL:** METAL is a popular command-line tool for meta-analysis of GWAS summary statistics. It provides various options for combining p-values, effect sizes, and sample sizes from different studies, allowing for fixed-effect or random-effects meta-analysis. METAL also offers options for genomic control correction and multiple testing adjustment [15].
- **GWAMA:** GWAMA (Genome-Wide Association Meta-Analysis) is another widely used software package for meta-analysis of GWAS results. It supports both fixed-effect and random-effects models, and it can handle a range of effect size measures and study designs. GWAMA also provides options for genomic control correction and correction for sample overlap between studies [12].
- **METASOFT:** METASOFT is a software tool specifically designed for conducting meta-analysis of GWAS summary statistics. It employs a Bayesian framework to combine p-values and effect sizes from multiple studies and provides options for adjusting for sample overlap and accounting for heterogeneity across studies [8].
- **PLINK:** PLINK is a widely used software package for performing genetic association analysis, including meta-analysis of GWAS summary statistics. It provides various commands and options for combining results from different studies and performing meta-analysis on a genome-wide scale [13].

**6. Result Interpretation**

The overall effect size and the magnitude (positive or negative) obtained from the meta-analysis should be analysed. A larger effect size suggests a stronger association between the genetic variant and the trait or disease under investigation. The statistical significance of the overall effect size is reported in form of a p-value which is set at a predefined threshold (such as $p<0.05$). The significant genes may further be studied through gene set enrichment analysis and protein-protein network interaction. Interpretation of the results of a meta-analysis should be done in conjunction with considering the limitations of the included studies, potential biases, and the specific research question being addressed.

**Advantages of conducting a meta-analysis**

Meta-analysis offers several advantages in the field of research.

- Increased statistical power: Meta-analysis combines data from multiple studies, leading to increased sample size and statistical power compared to individual studies. With a larger sample size, meta-analysis can provide more precise estimates of effect sizes and improve the ability to detect smaller or more subtle effects.
- Enhanced generalizability: By pooling data from multiple studies, meta-analysis provides a more comprehensive and representative assessment of the research question at hand. It allows for the inclusion of diverse populations, settings, and study designs, which can enhance the generalizability of the findings to a broader population or context.
- Resolving inconsistencies and heterogeneity: Meta-analysis enables the examination of sources of variation or heterogeneity across studies. It helps identify inconsistencies or conflicting findings among studies and provides a framework to explore potential reasons for the discrepancies. Through subgroup analyses and meta-regression, meta-analysis can uncover factors contributing to heterogeneity and inform further research or intervention strategies.
- Increased precision and confidence intervals: Meta-analysis provides more precise estimates of effect sizes by combining data from multiple studies. The resulting confidence intervals are narrower, providing a more accurate range of plausible effect estimates. This increased precision enhances the confidence in the findings and facilitates decision-making in clinical, policy, or practice settings.
- Identification of small or rare effects: Meta-analysis can detect small or rare effects that may not be easily detectable in individual studies with limited sample sizes. By aggregating data across studies, meta-analysis improves the statistical power to identify subtle or less common effects, providing a more comprehensive understanding of the phenomenon under investigation.
- Synthesis of conflicting evidence: Meta-analysis provides a systematic and quantitative approach to synthesizing conflicting evidence from different studies. It offers a framework for resolving discrepancies, reconciling contradictory findings, and arriving at a more comprehensive and robust conclusion. This can lead to a more reliable and evidence-based understanding of the research question.

## Conclusion

Meta-analysis of GWAS summary statistics is a powerful approach that enables researchers to gain collective insights into the genetic architecture of complex traits and diseases. By combining data from multiple independent GWAS, meta-analysis increases statistical power, identifies additional genetic associations, and explores heterogeneity across populations and datasets.

## References


1. Begg CB. Biases in the assessment of diagnostic tests. Statistics in medicine. 1987 Jun;6(4):411-23.
2. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Research synthesis methods. 2017 Mar;8(1):5-18.
3. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. Biostatistics. 2000 Sep 1;1(3):247-62.
4. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. Journal of the American statistical association. 2000 Mar 1;95(449):89-98.
5. Duval S. The trim and fill method. Publication bias in meta-analysis: Prevention, assessment and adjustments. 2005 Oct 7:127-44.
6. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Bmj. 1997 Sep 13;315(7109):629-634.
7. Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. Statistics in medicine. 1988 Aug;7(8):889-894.
8. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. The American Journal of Human Genetics. 2011 May 13;88(5):586-598.
9. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in medicine. 2002 Jun 15;21(11):1539-1558.
10. Iyengar S, Greenhouse J. Sensitivity analysis and diagnostics. Handbook of research synthesis and meta-analysis. 2009 Feb 5:417-433.
11. Light RJ, Richard J, Light R, Pillemer DB. Summing up: The science of reviewing research. Harvard University Press, 1984, Oct 15.
12. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC bioinformatics. 2010 Dec;11:1-6.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007 Sep 1;81(3):559-575.
14. Tate MW, Brown SM. Note on the Cochran Q test. Journal of the American Statistical Association. 1970 Mar 1;65(329):155-160.
15. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190-2191.