

# Meta-analysis for genome-wide association studies using case-control design: application and practice

Sungryul Shim<sup>1</sup>, Jiyoung Kim<sup>2</sup>, Wonguen Jung<sup>2</sup>, In-Soo Shin<sup>3</sup>, Jong-Myon Bae<sup>4</sup>

<sup>1</sup>Institute for Clinical Molecular Biology Research, Soonchunhyang University Hospital, Seoul; <sup>2</sup>Department of Radiation Oncology, Ewha Womans University School of Medicine, Seoul; <sup>3</sup>Department of Education, Jeonju University, Jeonju; <sup>4</sup>Department of Preventive Medicine, Jeju National University School of Medicine, Jeju, Korea

This review aimed to arrange the process of a systematic review of genome-wide association studies in order to practice and apply a genome-wide meta-analysis (GWMA). The process has a series of five steps: searching and selection, extraction of related information, evaluation of validity, meta-analysis by type of genetic model, and evaluation of heterogeneity. In contrast to intervention meta-analyses, GWMA has to evaluate the Hardy–Weinberg equilibrium (HWE) in the third step and conduct meta-analyses by five potential genetic models, including dominant, recessive, homozygote contrast, heterozygote contrast, and allelic contrast in the fourth step. The ‘genhwcci’ and ‘metan’ commands of STATA software evaluate the HWE and calculate a summary effect size, respectively. A meta-regression using the ‘metareg’ command of STATA should be conducted to evaluate related factors of heterogeneities.

**KEY WORDS:** Meta-analysis, Reviews, Genome-wide association study, Polymorphism, Genetic models

## INTRODUCTION

Malignant neoplasm, or cancer, is one of the most prevalent chronic diseases, which develops as a result of a somatic mutation. Advancing from this theory, a personalized medicine is currently gaining traction for the diagnosis and treatment of cancer [1], and such trends call for the synthesis of evidence related to genome-wide epidemiology [2].

With the advances in genetic technologies, the subjects of analyses in studies aiming to discover disease-related genomes have changed into chromosomal abnormalities, allelic heterogeneity, and single nucleotide polymorphisms (SNPs). According to these changes, linkage analysis studies, genetic association studies (GTAS), and genome-wide association studies (GWAS) has been

currently ongoing [2,3].

However, a phenomenon known as the “winner’s curse,” which is characterized by low replicability of results, has been appearing in follow-up studies on genes that were previously associated with a particular disease through genome-wide epidemiology studies [4-6]. Population stratification, diverse testing methods, and insufficient sample sizes have been implicated in this phenomenon [7-9], all of which constitute the rationale for the meta-analysis of genome-wide epidemiology studies [10-12].

This review introduces the process of a genome-wide meta-analysis (GWMA), which involves a meta-analysis of findings of GWAS that investigate the SNPs associated with a particular disease [13]. Particularly, this study presents an example of a meta-analysis in practice, in an attempt to inspire further GWMA studies in Korea.

## PROCESS OF GENOME-WIDE META-ANALYSIS

The general procedures of a GWMA introduced by previous studies [10,12-17] could be divided into five steps as shown in Table 1. Two features that distinguish GWMA from traditional systematic reviews are the Hardy-Weinberg equilibrium (HWE) test in step 3 for a quality evaluation of the selected literature

Correspondence: Jong-Myon Bae

Department of Preventive Medicine, Jeju National University School of Medicine, 102 Jejudaehak-ro, Jeju 63243, Korea  
Tel: +82-64-755-5567, Fax: +82-64-725-2593, E-mail: [jmbae@jejunu.ac.kr](mailto:jmbae@jejunu.ac.kr)

Received: Dec 10, 2016, Accepted: Dec 18, 2016, Published: Dec 18, 2016

This article is available from: <http://e-epih.org/>

© 2016, Korean Society of Epidemiology

© This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Table 1.** Five steps of conducting a genome-wide meta-analysis

Actions	
Step 1	Searching and Selection
Step 2	Extraction of related information
Step 3	Evaluation of validity
Step 4	Meta-analyses by types of genetic model
Step 5	Evaluation of heterogeneity

and the use of genetic models for meta-analyses in step 4.

Here, we present the study by Song et al. [18], which examined the association between Fc receptor-like 3-169 C/T polymorphism and rheumatoid arthritis in Asians, to describe the process of HWE testing and summary effect size calculating using a statistical program. The study selected 15 articles with a pooled sample of 22,312 individuals (11,170 cases + 11,142 controls). The selected articles were divided into three races (Asians, Europeans, and Native North Americans) for subgroup analysis. The polymorphic genotypes for the meta-analysis were CC, CT, and TT. We introduce the commands used on STATA version 14.2 (StataCorp, TX, USA) and interpret the results.

### Step 1: searching and selection

The search for GWAS articles involves different sources and keywords from those used for a search of general systematic reviews. We recommend the use of data sources on the organized tables by Casado-Vela et al. [19], Ramasamy et al. [20], and Wallace et al. [21]. Keywords such as 'genetics, alleles, and polymorphisms' are some medical subject headings regarding genome-wide epidemiology [22].

We recommend the use of the flow chart suggested by Sagoo et al. [12] for the literature selection process following the electronic search.

### Step 2: extraction of related information

The sets of information extracted from the selected GWAS articles are needed for the evaluation of the validity of each article in the next step. Items for evaluating the validity of GWAS articles have been suggested by Attia et al. [4], de Bakker et al. [14], Ramasamy et al. [20], and Khoury et al. [23]. Considering that GWMA results are applied to patient treatments, we strongly recommend the use of the items suggested by Attia et al. [4]. The organization of tables is recommended by the suggestions of Sagoo et al. [12].

If the quality of each of the selected genetic epidemiology studies must be assessed, the assessment checklist provided as supplementary data in the study by Thakkinstian et al. [24] or the checklist suggested on the "Strengthening the Reporting of Genetic Association Studies" by Little et al. [25] may be used.

### Step 3: evaluation of validity

One critical aspect of validity assessment for GWMA findings is the satisfaction of HWE assumption. HWE states that the frequencies of genes and genotypes remain in equilibrium over generations under limited conditions [3]. For example, given that the frequencies of two alleles, called A and a, of a gene are p and q, respectively, where  $p+q=1$ , the frequencies of the genotypes AA, Aa, and aa are  $p^2$ ,  $2pq$ , and  $q^2$ , respectively, where  $p^2+2pq+q^2=1$ . Using this equation, we can predict the frequency of a genotype with a known allele frequency.

The subjects of HWE testing depend on the study design. In a cohort study or cross-sectional study, HWE should be tested on the entire study population. On the other hand, HWE is only tested on the control group in a case-control study because the case group may not confirm to the HWE if the genotype is associated with a disease. Studies that deviate from the HWE should be excluded from step 4, and their meanings should be investigated in step 5 through a sensitivity analysis.

The most popular test to verify the HWE is the chi-squared test [26], a statistical technique that compares the observed values from a group with estimated values based on the assumption of HWE. In other words, it assesses the degree of deviation of observed values from the estimated values. A p-value of less than 0.05 is considered statistically significant and is interpreted to be a violation of the HWE.

For HWE analysis of case-control studies in the STATA software, genotypic counts of the case and control groups should be listed following the <genhwcci> command. For example, in Table 1 of the article by Song et al. [18], the genotypic counts for TT, TC, and CC in one of the 15 studies (Han et al. [27]) were 132, 180, and 65 in the case groups and 51, 133, and 114 in the control groups, respectively. Figure 1 shows the results of entering <genhwcci 132 180 65 51 133 114, binvar label (TT, TC, CC)> into the software. 'binvar' requests that standard errors from a binomial distribution are reported, and 'label' requests that results are presented according to the genotype. The p-value in the chi-square test for the control group was 0.257, which indicates that it does not violate the HWE.

### Step 4: meta-analyses by types of genetic model

In a C/T polymorphism where C is dominant and T is recessive, there are five possible types of genetic models: dominant (CC+CT vs. TT), recessive (CC vs. CT+TT), homozygote contrast (CC vs. TT), heterozygote contrast (CC vs. CT), and allelic contrast (C vs. T) [17,18,28,29].

Add the frequencies for the case and control groups of each article according to each model before performing the meta-analyses. For example, in the study by Han et al. [27], multiply CC and TT by two and add TC to each value for an allelic contrast (C vs. T) (Figure 1). In other words, the C for the case group

```

genhwcci 132 180 65 51 133 114, binvar label (TT, TC, CC)
  Genotype |          Case          Control | Total
-----+-----+-----+-----
      TT, |          132          51 |    183
      TC, |          180          133 |    313
      CC |           65          114 |    179
-----+-----+-----+-----
    total |          377          298 |    675
-----+-----+-----+-----
      Case
      Allele | Case Frequency Std. Err.
-----+-----+-----+-----
      T | 444 0.5889 0.0179 (binomial)
      C | 310 0.4111 0.0179 (binomial)
-----+-----+-----+-----
    total | 754 1.0000
-----+-----+-----+-----
Estimated disequilibrium coefficient (D) = 0.0034
                                         SE = 0.0125

Hardy-Weinberg Equilibrium Test:
  Pearson chi2 (1) = 0.073 Pr= 0.7865
likelihood-ratio chi2 (1) = 0.073 Pr= 0.7866
Exact significance prob = 0.8315

      Control
      Allele | Control Frequency Std. Err.
-----+-----+-----+-----
      T | 235 0.3943 0.0200 (binomial)
      C | 361 0.6057 0.0200 (binomial)
-----+-----+-----+-----
    total | 596 1.0000
-----+-----+-----+-----
Estimated disequilibrium coefficient (D) = 0.0157
                                         SE = 0.0139

Hardy-Weinberg Equilibrium Test:
  Pearson chi2 (1) = 1.283 Pr= 0.2573
likelihood-ratio chi2 (1) = 1.279 Pr= 0.2580
Exact significance prob = 0.2751

Test H0: cases under HWE: (given controls under HWE)
likelihood-ratio chi2 (2) = 50.807 Pr= 0.0000

```

**Figure 1.** Results of Hardy-Weinberg equilibrium testing using the STATA 'genhwcci' command of Han et al. [27].

becomes 310 ( $= 65 [CC] \times 2 + 180 [TC]$ ), and T becomes 444 ( $= 132 \times 2 + 180$ ). By the same method, the C for the control group becomes 361 ( $= 114 \times 2 + 133$ ), and T becomes 235 ( $= 51 \times 2 + 133$ ). Apply this method to the remaining 14 articles, and perform the meta-analyses.

For a frequency-based meta-analysis on STATA, use the <metan> command. Refer to Shim et al. [30] for creating a forest plot, calculating summary effect size, calculating the I-squared value for an evaluation of heterogeneity, creating a funnel plot to assess publication bias, and applying options for the Egger or Begg test. Figure 2 is a forest plot obtained from a meta-analysis of an allelic contrast model with the data from Song et al. [18], using the command <metan case\_C case\_T control\_C control\_T, or randomi by(ethnicity)>.

### Step 5: evaluation of heterogeneity

If heterogeneity is present, difference of race should be first considered [15,29], as differences in genetic pools may lead to heterogeneity among genome-wide epidemiology studies [4,31].

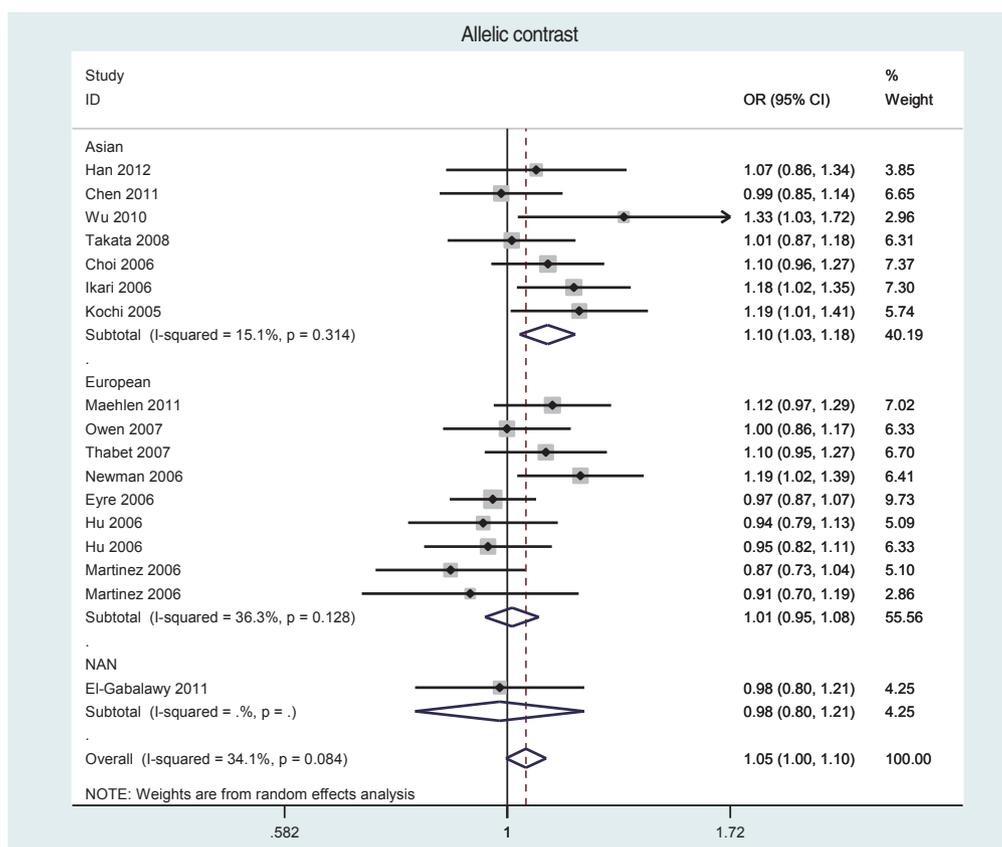
Hence, Song et al. [18] performed subgroup analyses by dividing the subjects into three races: Asians, Europeans, and Native North Americans. In addition, differences in allele frequencies may also induce heterogeneity among studies [32].

If heterogeneity is determined to persist, a random effect model may be applied [33,34]. However, a meta-regression may be applied to identify the cause of the heterogeneity [29,35]. Meta-regression is recommended only for analysis of ten or more articles, and its STATA command is <metareg> [30].

## CONCLUSION AND SUGGESTIONS

Two features that distinguish GWMA from the intervention meta-analyses are that GWMA uses HWE to verify the validity of a study and performs meta-analyses according to the five possible types of genetic models.

If individual patient data, as opposed to the findings of the selected literature, are used, the STATA <metagen> command



**Figure 2.** A forest plot of an allelic contrast model, using the STATA ‘metan’ command of Song et al. [18]. OR, odds ratio; NAN, North American Natives; CI, confidence interval.

may be used [36]. Furthermore, there may be a hypothesis in which the outcome variables are continuous and not dichotomous. A case in point is the investigation of differences in bone density according to vitamin D receptor polymorphisms [17]. We plan to describe the process of GWMA involving continuous outcome variables in a future article. In addition, we shall introduce genome search meta-analysis (GSMA), which was developed for meta-analysis for ordinal outcome variables [37], at another time.

Currently, genome-wide epidemiology is evolving into system epidemiology using multi-omics, including proteomics, metabolomics, and epigenomics, in pursuit of precision medicine [19,38,39]. Amid this trend, GWMA is vital in that it can reinterpret existing studies and suggest future research directions. We hope this article provides inspiration for further studies.

**ACKNOWLEDGEMENTS**

This work is the product of research activities of Meta-analysis Study Group in Korea (President: In-Soo Shin).

**CONFLICT OF INTEREST**

The authors have no conflicts of interest to declare for this study.

**SUPPLEMENTARY MATERIAL**

Supplementary material (Korean version) is available at <http://www.e-epih.org/>.

**ORCID**

- Sungryul Shim <http://orcid.org/0000-0003-4143-7383>
- Jiyoung Kim <http://orcid.org/0000-0002-0843-2241>
- Wonguen Jung <http://orcid.org/0000-0001-5595-7510>
- In-Soo Shin <http://orcid.org/0000-0001-7535-3511>
- Jong-Myon Bae <http://orcid.org/0000-0003-3080-7852>

## REFERENCES

1. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. *N Engl J Med* 2015;372:2229-2234.
2. McCarthy JJ, McLeod HL, Ginsburg GS. Genomic medicine: a decade of successes, challenges, and opportunities. *Sci Transl Med* 2013; 5:189sr4.
3. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: A: background concepts. *JAMA* 2009;301:74-81.
4. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: B: are the results of the study valid? *JAMA* 2009;301:191-197.
5. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: C: what are the results and will they help me in caring for my patients? *JAMA* 2009; 301:304-308.
6. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; 33:177-182.
7. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361: 865-872.
8. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91-99.
9. Ioannidis JP. Genetic associations: false or true? *Trends Mol Med* 2003;9:135-138.
10. Lee YH. Meta-analysis of genetic association studies. *Ann Lab Med* 2015;35:283-287.
11. Gwinn M, Ioannidis JP, Little J, Khoury MJ. Editorial: updated guidance on human genome epidemiology (HuGE) reviews and meta-analyses of genetic associations. *Am J Epidemiol* 2014;180:559-561.
12. Sagoo GS, Little J, Higgins JP. Systematic reviews of genetic association studies. *Human Genome Epidemiology Network. PLoS Med* 2009;6:e28.
13. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 2009;10:191-201.
14. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008;17:R122-R128.
15. Thompson JR, Attia J, Minelli C. The meta-analysis of genome-wide association studies. *Brief Bioinform* 2011;12:259-269.
16. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* 2013;14:379-389.
17. Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J. A method for meta-analysis of molecular association studies. *Stat Med* 2005; 24:1291-1306.
18. Song GG, Bae SC, Kim JH, Kim YH, Choi SJ, Ji JD, et al. Association between functional Fc receptor-like 3 (FCRL3) -169 C/T polymorphism and susceptibility to seropositive rheumatoid arthritis in Asians: a meta-analysis. *Hum Immunol* 2013;74:1206-1213.
19. Casado-Vela J, Cebrián A, Gómez del Pulgar MT, Lacal JC. Approaches for the study of cancer: towards the integration of genomics, proteomics and metabolomics. *Clin Transl Oncol* 2011;13:617-628.
20. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* 2008;5:e184.
21. Wallace BC, Small K, Brodley CE, Lau J, Schmid CH, Bertram L, et al. Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. *Genet Med* 2012;14:663-669.
22. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003;56:297-303.
23. Khoury MJ, Bertram L, Boffetta P, Butterworth AS, Chanock SJ, Dolan SM, et al. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. *Am J Epidemiol* 2009;170:269-279.
24. Thakkinstian A, McEvoy M, Minelli C, Gibson P, Hancox B, Duffy D, et al. Systematic review and meta-analysis of the association between {beta}2-adrenoceptor polymorphisms and asthma: a HuGE review. *Am J Epidemiol* 2005;162:201-211.
25. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Eur J Epidemiol* 2009; 24:37-55.
26. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:967-986.
27. Han SW, Sa KH, Kim SI, Lee SI, Park YW, Lee SS, et al. FCRL3 gene polymorphisms contribute to the radiographic severity rather than susceptibility of rheumatoid arthritis. *Hum Immunol* 2012;73: 537-542.
28. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol* 2005;34:1319-1328.
29. Zintzaras E, Lau J. Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. *J Clin Epidemiol* 2008;61:634-645.
30. Shim SR, Shin IS, Bae JM. Intervention meta-analysis using STATA software. *J Health Inform Stat* 2016;41:123-134 (Korean).
31. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;7:781-791.
32. Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A* 2008;105:617-622.
33. Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* 2008;123:1-14.
34. Munafò MR, Clark TG, Flint J. Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. *Psychiatry Res* 2004;129:39-44.
35. Shim SR, Shin IS, Yoon BH, Bae JM. Dose-response meta-analysis using STATA software. *J Health Inform Stat* 2016;41:351-358 (Korean).
36. Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res* 2012;40:3777-3784.
37. Wise LH, Lanchbury JS, Lewis CM. Meta-analysis of genome searches. *Ann Hum Genet* 1999;63:263-272.
38. Gonzalez de Castro D, Clarke PA, Al-Lazikani B, Workman P. Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. *Clin Pharmacol Ther* 2013;93:252-259.
39. Dammann O, Gray P, Gressens P, Wolkenhauer O, Leviton A. Systems epidemiology: what's in a name? *Online J Public Health Inform* 2014;6:e198.