

Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores

Cue Hyunkyu Lee^{1,2}, Seungho Cook^{1,3}, Ji Sung Lee^{1,4}, Buhm Han^{1,2*}

¹Asan Institute for Life Sciences, Asan Medical Center, Seoul 05505, Korea,

²Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul 05505, Korea,

³School of Systems Biomedical Science, Soongsil University, Seoul 06978, Korea,

⁴Department of Medicine, University of Ulsan College of Medicine, Seoul 05505, Korea

The meta-analysis has become a widely used tool for many applications in bioinformatics, including genome-wide association studies. A commonly used approach for meta-analysis is the fixed effects model approach, for which there are two popular methods: the inverse variance-weighted average method and weighted sum of z-scores method. Although previous studies have shown that the two methods perform similarly, their characteristics and their relationship have not been thoroughly investigated. In this paper, we investigate the optimal characteristics of the two methods and show the connection between the two methods. We demonstrate that each method is optimized for a unique goal, which gives us insight into the optimal weights for the weighted sum of z-scores method. We examine the connection between the two methods both analytically and empirically and show that their resulting statistics become equivalent under certain assumptions. Finally, we apply both methods to the Wellcome Trust Case Control Consortium data and demonstrate that the two methods can give distinct results in certain study designs.

Keywords: fixed effects model, genome-wide association study, inverse variance-weighted average, meta-analysis, optimality, weighted sum of z-scores

Introduction

The meta-analysis is a tool for pooling information from multiple independent studies [1-4]. In the field of genetics, the meta-analysis has become a popular way of aggregating information from multiple genome-wide association studies (GWASs) in order to increase statistical power while controlling for the rate of false positive findings [5-13]. The meta-analysis has also become a useful tool for many applications of bioinformatics, such as neuroimage processing [14] and expression quantitative trait loci analysis [15].

There exist several approaches for combining information from multiple studies. Statistical methods can differ depending on the scenario: when (1) test statistics are unknown but only p-values are available, (2) test statistics are known but data are not available, or (3) actual data are available. In

this paper, we focus on scenario (2), which is a common situation in genetic studies. We note that for scenario (1), Fisher's method for combining p-values is commonly used [16]. In scenario (3), we can combine actual data, which is rarely doable in retrospective studies or in genetic studies where transferring genotype data is difficult due to privacy issues. For scenario (2), which we focus on, the fixed effects model meta-analysis is the most common approach for synthesizing test statistics from multiple studies [1, 17].

To perform a fixed effects model meta-analysis, there are two popular methods: the inverse variance-weighted average and the weighted sum of z-scores (SZ) [2, 17, 18]. The inverse variance-weighted average method (IVW) summarizes effect sizes from multiple independent studies by calculating the weighted mean of the effect sizes using the inverse variance of the individual studies as weights. The weighted SZ method constructs a new z-score by calculating

Received October 18, 2016; Revised December 3, 2016; Accepted December 3, 2016

*Corresponding author: Tel: +82-2-3010-4176, Fax: +82-2-3010-4182, E-mail: buhm.han@amc.seoul.kr

Copyright © 2016 by the Korea Genome Organization

It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).