Comparison of Different Transformation Methods for Rates in Single-Rate Meta-Analysis

Authors: Zeng Zhiyu, Zeng Zhiyu

Date: 2023-04-19T00:00:00+00:00

Abstract

Objective To compare different transformation methods for rates in meta-analysis of single proportions. Methods Two sets of simulated data were constructed for meta-analysis of single proportions to evaluate the performance under five data transformation methods (no transformation, logarithmic transformation, logit transformation, arcsine square root transformation, and double arcsine transformation), considering both fixed-effect and random-effects models, and incorporating different constant values when event counts were zero. The mean of pooled rates (Mean), bias (Bias), proportion bias (Proportion Bias), mean squared error (Mean Squared Error, MSE), proportion mean squared error (Proportion MSE), and coverage of 95% confidence intervals (Coverage) were calculated. Results For meta-analysis of single proportions based on binomial distribution, the arcsine square root transformation demonstrated the best overall performance. When event counts were zero, adding different constant values substantially affected the results; however, this correction provided no benefit to the no-transformation strategy and was even detrimental, while offering very limited improvement for logarithmic and logit transformations. When the overall rate was <0.05, the pooled rate from meta-analysis of single proportions exhibited substantial bias. Conclusion The arcsine square root transformation performs optimally in meta-analysis of single proportions. Caution is warranted when employing meta-analysis for overall rates <0.05.

Full Text

Comparison of Different Data Transformations for Meta-Analysis of Single Proportions

ZENG Zhi-yu
Boston Scientific International Medical Trading (Shanghai) Co., Ltd., Beijing, 100020, China

Abstract

Objective: To compare different data transformation methods in meta-analysis for single proportions. **Methods:** Two sets of simulation data were constructed for meta-analysis of single proportions, examining five data transformation methods (no transformation, log transformation, logit transformation, arcsine transformation, and double arcsine transformation) under both fixed-effect and random-effects models, with different fixed values added when event counts were zero. The pooled proportion mean, bias, proportion bias, mean squared error (MSE), proportion MSE, and 95% confidence interval coverage were calculated. **Results:** For meta-analysis of single proportions based on binomial distribution, the arcsine transformation demonstrated the best overall performance. When event counts were zero, adding different fixed values substantially affected results, but this correction did not improve the no-transformation strategy and was even detrimental; improvements for log and logit transformations were also very limited. When the population proportion was less than 0.05, the pooled proportion from meta-analysis showed substantial bias. **Conclusion:** The arcsine transformation performs optimally in meta-analysis of single proportions. Caution is advised when using meta-analysis for population proportions below 0.05.

Keywords: Single Proportions; Meta-Analysis; Data Transformation; Arcsine Transformation

Introduction

Meta-analysis encompasses various types, with single-proportion meta-analysis primarily based on uncontrolled dichotomous data from cross-sectional studies. Such data are commonly used to investigate prevalence, detection rates, awareness rates, case fatality rates, and infection rates. These data are characterized by having only a single group with event counts and total observations, without a control group. This study focuses on meta-analysis of such data.

The data objects in single-proportion meta-analysis belong to count data. Descriptive statistics for count data include ratios, proportions, and rates, though these terms are often confused in clinical research practice. Generally, a distinction should be made between proportion (e.g., prevalence, cure rate) and rate (or incidence, e.g., mortality rate, morbidity rate). This study excludes rates expressed as time-to-event data and also excludes effect measures such as sensitivity and specificity from diagnostic tests.

Single-proportion meta-analysis typically employs the inverse variance method for pooling, making appropriate data transformations necessary when sample sizes or proportions are small. Many software packages provide different data transformation methods, including no transformation (using raw proportions), log transformation, logit transformation, arcsine transformation, and double arcsine transformation. However, operational guidelines on when to apply transformations and which method to prefer remain unclear. This study uses simulation analysis to compare these transformation methods for single-proportion meta-analysis.

Methods

2.1 Simulation Data

Two sets of simulation data were constructed to evaluate five proportion transformation methods in meta-analysis of single proportions.

Simulation Data 1 was based on previously published simulation parameters. Nine studies were assumed with sample sizes increasing from 20 to 180 in increments of 20. Event counts in each study were assumed to follow a binomial distribution given the sample size and population proportion (e.g., prevalence). For the fixed-effect model, the true prevalence P0 was set at 0.01, 0.05, and 0.2. For the random-effects model, P was assumed to follow a normal distribution with mean P0 and standard deviations of 0.1P0, 0.2P0, 0.5P0, and P0. This generated a total of 3 + 3$×$3 = 12 scenarios. Each scenario was randomly sampled 1,000 times, with each sample undergoing meta-analysis using five transformation methods. For zero-event cases, the effects of adding values from 0 to 1 in increments of 0.1 were examined, resulting in $12 \times 5 \times 11 \times 1{,}000 = 660{,}000$ meta-analyses.

Simulation Data 2 represented medical complication scenarios. Five studies were assumed with sample sizes of 50, 100, 200, 300, and 1,000 (one small, three medium, and one large). Only the fixed-effect model was considered, with complication rates ranging from 0.01 to 0.2 in increments of 0.01 (20 conditions). Each condition was sampled 1,000 times, with each sample undergoing meta-analysis using five transformation methods. For zero-event cases, either no adjustment or addition of 0.5 was applied, yielding $20 \times 5 \times 2 \times 1{,}000 = 200{,}000$ meta-analyses.

2.2 Meta-Analysis Methods for Single Proportions

Statistical analyses were performed using R (version 3.5.2) with the Metaprop function from the meta package. For each of the 1,000 meta-analyses under each condition, the pooled proportion mean, bias, proportion bias, mean squared error (MSE), proportion MSE, and 95% confidence interval coverage were calculated. Because normal approximation for 95% confidence intervals is clearly disadvantageous for no-transformation scenarios when proportions are small, the Clopper-Pearson method was uniformly adopted for all confidence interval calculations to avoid methodological bias.

Results

3.1 Results from Simulation Data 1

[Figure 1: see original paper] presents results for the fixed-effect model (P0 = 0.05), with the x-axis showing values added when event counts were zero (from 0 to 1 in increments of 0.1). Panel A shows that the pooled proportion means from the arcsine transformation (green) and double arcsine transformation (blue) were closest to 0.05. The raw proportion performed poorly, and while adding values to zero events noticeably affected the pooled proportion, the correction was unsatisfactory. Panel B shows proportion bias, where the arcsine transformation again exhibited the smallest absolute bias, followed by double arcsine, logit, and log transformations. Panels C and D show proportion MSE and 95% confidence interval coverage, respectively, where all four transformation methods performed similarly and substantially better than no transformation. provides specific values for scenarios where zero events were either left unchanged or adjusted by adding 0.5.

Similarly, [Figure 2: see original paper] shows results for the random-effects model (P = 0.05, SD = 0.005). Panel A demonstrates that arcsine and double arcsine transformations produced pooled proportion means closest to 0.05, while raw proportions performed poorly. Adding values to zero events had noticeable but unsatisfactory effects, and log and logit transformations also performed poorly. Panel B shows that arcsine and double arcsine transformations had the smallest absolute proportion bias, outperforming raw proportions, logit, and log transformations. Panel C shows that arcsine and double arcsine transformations were optimal for proportion MSE. Panel D shows that arcsine and double arcsine transformations achieved the best and nearly identical 95% confidence interval coverage. presents specific values for zero-event scenarios with or without addition of 0.5. Additionally, in random-effects models, pooled proportion bias and coverage gradually deteriorated as the standard deviation of the complication rate P increased (results not shown).

3.2 Results from Simulation Data 2

[Figure 3: see original paper] displays results from Simulation Data 2 (hypothetical complication data), with the x-axis showing population complication rates (from 0.01 to 0.20 in increments of 0.01). Panels A and B both show proportion bias: Panel A leaves zero events unchanged, while Panel B adds 0.5. Panel A reveals that all methods exhibited substantial bias when the population proportion was less than 0.05, though bias decreased rapidly as the population proportion increased. Log and logit transformations overestimated complication rates, while no transformation, arcsine, and double arcsine transformations underestimated them. Overall, the arcsine transformation (green) showed the smallest absolute proportion bias. Panel B shows that adding 0.5 did not substantially improve bias and actually worsened bias for raw proportions. Panel C shows proportion MSE, and Panel D shows 95% confidence interval coverage, where arcsine, log, and logit transformations achieved higher coverage rates.

Discussion

This simulation study found that for meta-analysis of single proportions based on binomial distribution, the arcsine transformation demonstrated the best overall performance. When event counts were zero, adding different fixed values substantially affected results, but this correction neither helped nor even harmed the no-transformation strategy, while providing very limited improvement for log and logit transformations. Zero events did not affect results from arcsine and double arcsine transformations. When the population proportion was less than 0.05, meta-analysis produced substantially biased pooled proportions.

Single-proportion meta-analysis using the inverse variance method assigns excessive weight to small proportions, making transformation methods essential. Various transformation methods exist depending on data characteristics and objectives. Five common methods for single-proportion meta-analysis are: (1) no transformation (direct calculation from raw proportions), (2) log transformation, (3) logit transformation, (4) arcsine transformation, and (5) double arcsine transformation (also called Freeman-Tukey transformation). These methods are incorporated in many meta-analysis software packages, yet guidelines on when to use each method remain inconclusive.

Domestic literature includes introductions to software for single-proportion meta-analysis, but similar international studies are scarce. Trikalinos et al. conducted simulation studies for meta-analysis of single proportions and rates, creating over 700 meta-analysis scenarios based on proportion distributions, means, variances, number of studies, and sample sizes, with 1,000 random samples per scenario under both fixed-effect and random-effects models. They examined no transformation, logit transformation, and arcsine transformation, comparing meta-analytic estimates to true values using difference, percentage difference, MSE, percentage MSE, and 95% confidence interval coverage (using Wald method for individual analyses then summarizing across 1,000 simulations). Their results favored arcsine transformation, consistent with our findings. Barendregt et al. simulated meta-analysis of prevalence, assuming nine studies with sample sizes from 20-180, prevalence of 0.05, and random-effects model with normally distributed prevalence (mean = 0.05, SD = 0.005). After 1,000 random samples, they compared no transformation, logit transformation, and double arcsine transformation using bias, MSE, and 95% confidence interval coverage (Wald method). Double arcsine transformation slightly outperformed logit transformation (e.g., in fixed-effect models, proportion bias was -0.022, 0.004, and 0.002 for raw, logit, and double arcsine transformations, respectively). Our Simulation Data 1 yielded similar results: arcsine and double arcsine transformations were comparable in fixed-effect models, while double arcsine outperformed logit in random-effects models. In Simulation Data 2, double arcsine and logit transformations produced similar results, with logit showing better performance in proportion MSE and coverage, though both were inferior to arcsine transformation.

These two studies each compared only three transformation methods without direct comparison between arcsine and double arcsine transformations, and Barendregt et al. used relatively simple scenarios. Our study compared five transformation methods across two simulation datasets, consistently demonstrating arcsine transformation as optimal.

Study limitations include: (1) exclusive use of R software, which may not represent other software, though R is a powerful, flexible, and increasingly popular free software for meta-analysis; and (2) limited simulation scenarios, though our two datasets examining various zero-event adjustments and population proportions from 0.01 to 0.20 should provide representative insights.

In conclusion, arcsine transformation is recommended for single-proportion meta-analysis, while adding a fixed value to zero events does not substantially improve results.

References

[1] Barendregt J J, Doi S A, Lee Y Y, et al. Meta-analysis of prevalence[J]. Journal of Epidemiology and Community Health,2013,67(11):974-978.
[2] Sun Zhenqiu, Xu Yongyong. Medical Statistics[M]. 4th ed. Beijing: People's Medical Publishing House, 2014.
[3] Nyaga V N, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data[J]. Arch Public Health,2014,72(1):39.
[4] Trikalinos T A, Trow P, Schmid C H. Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates[M]. Rockville (MD):Agency for Healthcare Research and Quality (US),2013.
[5] Schwarzer G. General Package for Meta-Analysis[Z]. 2017:2018.
[6] Wang Peixin, Li Hongtian, Liu Jianmeng. Meta-analysis methods for uncontrolled dichotomous data and Stata implementation[J]. Evidence-Based Medicine, 2012, 12(01): 52-55.
[7] Luo Meiling, Tan Hongzhuan, Zhou Quan, et al. Implementing meta-analysis of single rates in R software[J]. Evidence-Based Medicine, 2013, 13(03): 181-184.
[8] Xiao Jing, Wu Guiyun, Gao Yuexia, et al. Implementation of single-rate meta-analysis using Matlab software[J]. Chinese Journal of Health Statistics, 2014, 31(04): 715-717.
[9] Chen Yuehong, Du Liang, Geng Xingyuan, et al. Implementation of meta-analysis for uncontrolled dichotomous data in RevMan software[J]. Chinese Journal of Evidence-Based Medicine, 2014, 14(07): 889-896.

Note: Figure translations are in progress. See original paper for figures.