Current Application Status of Different Transformation Methods for Proportions in Meta-Analysis of Single Proportions

Authors: Zeng Zhiyu

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

Abstract

Objective: To examine the actual usage of data transformation methods for rates in current literature on meta-analysis of single proportions. Methods: A search was conducted in PubMed for literature on meta-analysis of single proportions published in 2017, and 145 articles were selected for analysis from 481 records. Results: Among the 123 articles with full text, only 33 (26.8%) reported the use of rate transformation methods, including 20 using the double arcsine method, 8 using logit transformation, 3 using square root arcsine transformation, 1 using logarithmic transformation, and 1 using raw rates directly. In these 33 articles, the use of rate transformation method was not associated with the magnitude of the pooled rate (P=0.217). Conclusion: The transformation method for rates is an important factor in meta-analysis of single proportions, but the relative merits of various transformation methods remain inconclusive; published literature should provide more detailed descriptions of methods such as data transformation for rates.

Full Text

Current Status of Data Transformation Methods for Proportions in Meta-Analysis of Single Rates

Boston Scientific Corporation BSC Int'l Medical Trading (Shanghai) Co., Ltd. Beijing, 100020, China

Abstract

Objective: To investigate the current status of data transformation methods in meta-analyses of single proportions. **Methods:** A literature search in PubMed was performed to retrieve meta-analyses of single proportions published in 2017,

yielding 481 records, from which 145 articles were included in the analysis. **Results:** Among 123 articles with full text, only 33 (26.8%) described the use of proportion transformation methods. Specifically, double arcsine transformation was used in 20 articles, logit transformation in 8, squared arcsine transformation in 3, log transformation in 1, and raw proportions were used directly in 1. In these 33 articles, the choice of transformation method was not associated with the magnitude of pooled proportions (P=0.217). **Conclusion:** Data transformation is an important consideration in meta-analysis of single proportions, yet there is no consensus on the optimal method. Published studies should provide clearer descriptions of data transformation methods for proportions.

Keywords: Single proportion; Meta-analysis; Data transformation

Meta-analysis is a statistical method for quantitatively synthesizing multiple studies on the same research question [1]. Its fundamental concepts and methods can be traced back to 1904 when Pearson began combining results from multiple studies, followed by Fisher's method for combining P-values. However, the term "Meta-analysis" was not formally coined until 1976 by Glass GV [2, 3].

Meta-analysis encompasses various types, including a special category known as meta-analysis of single proportions. This approach primarily deals with uncontrolled dichotomous data from cross-sectional studies, commonly used to investigate prevalence, detection rates, awareness rates, case-fatality rates, and infection rates. Such data are characterized by having only single-group event counts and total observations without a control group. This study focuses on meta-analysis of this type of data, excluding incidence rates expressed as time-to-event data and diagnostic test metrics such as sensitivity and specificity.

In clinical practice, the normal approximation method is commonly used for proportion data due to its convenience. However, substantial bias can arise when the proportion or sample size is small. Additionally, when using the inverse variance method for pooling in single-proportion meta-analysis, smaller proportions may receive disproportionately large weights. In such cases, alternative transformation methods for proportions should be considered to meet statistical assumptions. Currently, several transformation options exist for effect size pooling in single-proportion meta-analysis, including no transformation (using raw proportions directly), log transformation, logit transformation, squared arcsine transformation, and double arcsine transformation. Nevertheless, guidance on when to apply transformations and which method to prefer remains unclear. This study examines the actual usage of data transformation methods for proportions in current meta-analysis literature.

2. Methods

2.1 Literature Search

We searched PubMed for articles published in 2017 using terms including: Meta analysis, prevalence, incidence, rate, percentage, frequency,

proportion, mortality, and complication. Free-text searching was employed with appropriate logical combinations of search terms. After iterative refinement, the final search strategy was: `((((meta[title] AND analysis[title]) AND (proportion[Title] OR prevalence[Title] OR percentage[Title] OR frequency[Title] OR incidence[Title] OR rate[Title] OR complication[title] OR mortality[title])) AND ("2017/01/01"[PDat] : "2017/12/31"[PDat]))) AND meta-analysis[Publication Type]`, which retrieved 481 records on June 17, 2018.

2.2 Literature Screening

The 481 records were imported into NoteExpress reference management software. All titles and abstracts were reviewed, with inclusion criteria requiring that abstracts reported pooled rates expressed as percentages in meta-analyses. Full texts were then obtained. The literature screening flowchart is shown in [Figure 2: see original paper]. Ultimately, 145 articles were included, with full text available for 123 articles.

2.3 Data Extraction

[Figure 1: see original paper] Literature screening flowchart. We extracted parameters including type of proportion, number of studies, minimum sample size, meta-analysis software, transformation method, pooled proportion, and I^2 value (see Error: Reference source not found). When a single article contained multiple single-proportion meta-analyses, we selected either the primary analysis or the first one reported.

2.4 Statistical Analysis

The analysis was primarily descriptive. Categorical data were expressed as counts and percentages, while continuous data were presented as mean ± standard deviation along with medians. No formal hypothesis testing was conducted. Simple linear regression was performed to examine the relationship between pooled proportion magnitude and transformation method. Statistical analyses were conducted using R software (version 3.5.2).

Results

Among the 145 included articles, the majority originated from mainland China (37 articles, 25.5%), followed by the United States (17, 11.7%), Canada (15, 10.3%), and the United Kingdom (13, 9.0%). The term “prevalence” was used in most articles (122, 84.1%), with other terms including incidence (11, 7.6%), rate (5, 3.4%), and proportion (3, 2.1%). Notably, some articles using “prevalence” actually described incidence concepts but without person-time denominators, while others used “incidence” to refer to prevalence. Since the statistical methods were identical, all were included in the analysis.

The number of studies included in each meta-analysis ranged from 2 to 184 (mean 38.2 ± 26.2, median 28 [IQR: 16–49]). Minimum sample sizes ranged from 2 to 39,211 (mean 628.3 ± 4,202.3, median 49 [IQR: 20–134]). Pooled proportions ranged from 0.0003 to 0.984 (mean 0.245 ± 0.227, median 0.179 [IQR: 0.070–0.351]).

Among the 123 articles with full text, STATA was the most frequently used software (49 articles, 39.8%), followed by Comprehensive Meta-Analysis (CMA, 26, 21.1%) and R (21, 17.1%). Other software included MetaXL (4, 3.3%), Open Meta-Analysis (3, 2.4%), MedCalc (3, 2.4%), and StatsDirect (3, 2.4%).

Only 33 of the 123 full-text articles (26.8%) described the use of proportion transformation methods. Among these, double arcsine transformation was used in 20 articles, logit transformation in 8, squared arcsine transformation in 3, log transformation in 1, and raw proportions were used directly in 1. Only one article (reference 111) [4] compared the normality assumptions across five transformation methods before selecting the squared arcsine method. In these 33 articles, the choice of transformation method was not associated with the magnitude of pooled proportions (simple linear regression, P=0.217).

Only five articles specified the weighting method for pooling proportions, with four using inverse variance and one using sample size. Just two articles mentioned adding 0.5 as a correction for zero-event studies. The vast majority reported I^2 values, which indicated substantial heterogeneity across studies, and most employed random-effects models.

Discussion

This study provides a detailed description of the current application of proportion transformation methods in single-proportion meta-analysis. The double arcsine method was most commonly used, followed by logit and squared arcsine transformations. However, overall reporting of transformation methods was inadequate, with insufficient description of weighting schemes, continuity corrections, and other methodological details.

Our search of Chinese literature identified several introductory articles on software implementation for single-proportion meta-analysis [5-8]. For example, Wang et al. introduced methods for uncontrolled dichotomous data and STATA implementation [5]; Luo et al. described performing single-proportion meta-analysis in R [6]; Xiao et al. presented Matlab implementation [7]; and Chen et al. explained the process in RevMan [8]. However, these articles focused primarily on software usage rather than comparing different transformation methods.

International literature on this topic is also limited [9-11]. Trikalinos et al. [9] conducted simulation studies comparing methods for meta-analysis of proportions and rates, creating over 700 meta-analysis scenarios based on distribution, mean, variance, number of studies, and sample sizes, with 1,000 random samples per scenario under both fixed- and random-effects models. They examined three

transformation methods (no transformation, logit, and arcsine) and found arcsine transformation performed best. Barendregt et al. [10] simulated prevalence meta-analyses with nine studies (sample sizes 20–180), assuming a prevalence of 0.05 with normally distributed rates across studies (mean 0.05, SD 0.005) under random-effects models. After 1,000 random samples comparing three methods (no transformation, logit, and double arcsine), double arcsine performed best. Unfortunately, both studies compared only three methods each without direct comparison between arcsine and double arcsine transformations. More concerningly, a recent article argued that the double arcsine transformation has serious flaws in single-proportion meta-analysis and recommended generalized linear mixed models instead [11].

Methodological research on meta-analysis is crucial, particularly for small proportions. For instance, a 2007 New England Journal of Medicine meta-analysis on rosiglitazone safety [12] concluded that the drug increased myocardial infarction risk and showed a trend toward increased cardiovascular mortality. This publication generated extensive controversy and methodological debate, particularly regarding handling of small proportions, with different meta-analyses yielding contradictory results [13-17]. Nevertheless, this meta-analysis became pivotal evidence for drug safety concerns, contributing to rosiglitazone's withdrawal in Europe in 2010 and strict usage restrictions by the US FDA in 2011. After more clinical evidence became available, the FDA removed these restrictions in 2013 and reaffirmed the drug's safety in 2015 [18].

While meta-analysis of rare events with control groups has been extensively discussed [19-23], research on transformation method applicability for single-proportion meta-analysis remains insufficient [9, 10]. Clinical demand for summarizing single proportions is increasing, including analyses of adverse events for drugs and devices. In this context, meta-analyses should provide detailed methodological descriptions to ensure reproducibility.

This study has several limitations: (1) The literature search was limited to one year in PubMed, and some full texts were unavailable, though the results likely reflect the general underreporting of methodological details in single-proportion meta-analyses. (2) When multiple single-proportion meta-analyses were present in one article, only the primary or first analysis was selected. (3) While focusing on transformation methods, we only briefly mentioned other important parameters such as weighting and correction methods, particularly the construction of 95% confidence intervals. Nevertheless, our findings demonstrate that these parameters are seldom reported in published literature.

In conclusion, data transformation is an important consideration in single-proportion meta-analysis, but the relative merits of different methods remain undetermined. Beyond further methodological research, published studies should provide clearer descriptions of transformation methods and related parameters.

5. References

[1] Li Y. Evidence-Based Medicine[M]. Beijing: People's Medical Publishing House, 2014.
[2] Gurevitch J, Koricheva J, Nakagawa S, et al. Meta-analysis and the science of research synthesis[J]. Nature,2018,555(7695):175-182.
[3] Glass G V. Primary, secondary, and meta-analysis of research[J]. Educ. Res.,1976,5(10):3-8.
[4] Wang T, Liu Y, Li Z, et al. Prevalence of intimate partner violence (IPV) during pregnancy in China: A systematic review and meta-analysis[J]. PLoS One,2017,12(10):e175108.
[5] Wang P, Li H, Liu J. Meta-Analysis Methods for Uncontrolled Dichotomous Data and Implementation in Stata[J]. Evidence-Based Medicine, 2012, 12(01): 52-55.
[6] Luo M, Tan H, Zhou Q, et al. Implementing Meta-Analysis of Single Rates in R Software[J]. Evidence-Based Medicine, 2013, 13(03): 181-184.
[7] Xiao J, Wu G, Gao Y, et al. Implementation of Single-Rate Meta-Analysis Using Matlab Software[J]. Chinese Journal of Health Statistics, 2014, 31(04): 715-717.
[8] Chen Y, Du L, Geng X, 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.
[9] 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.
[10] 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.
[11] Schwarzer G, Chemaitelly H, Abu-Raddad L J, et al. Seriously misleading results using inverse of freeman-tukey double arcsine transformation in meta-analysis of single proportions[J]. Res Synth Methods,2019 doi: 10.1002/jrsm.1348. [Epub ahead of print].
[12] Nissen S E, Wolski K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes[J]. New England Journal of Medicine,2007,356(24):2457-2471.
[13] Lu Y, Ma D, Yu X. Cardiovascular Safety of Rosiglitazone and Pioglitazone[J]. Adverse Drug Reactions Journal, 2015(2).
[14] Chen Y, Tang X, Zeng L. Re-evaluation of Cardiovascular Safety of Rosiglitazone[J]. Chinese Journal of Endocrinology and Metabolism, 2014, 30(6).
[15] Dahabreh I J. Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of rosiglitazone[J]. Clinical Trials: Journal of the Society for Clinical Trials,2008,5(2):116-120.
[16] Cai T, Parast L, Ryan L. Meta-analysis for rare events[J]. Statistics in Medicine,2010,29(20):2078-2091.
[17] Nissen S E, Wolski K. Rosiglitazone Revisited[J]. Archives of Internal Medicine,2010,170(14):1191-1201.

[18] FDA Drug Safety Communication: FDA eliminates the Risk Evaluation and Mitigation Strategy (REMS) for rosiglitazone-containing diabetes medicines[Z]. 2015:2018.
[19] Bhaumik D K, Amatya A, Normand S T, et al. Meta-Analysis of Rare Binary Adverse Event Data[J]. Journal of the American Statistical Association,2012,107(498):555-567.
[20] Ma Y, Chu H, Mazumdar M. Meta-analysis of Proportions of Rare Events–A Comparison of Exact Likelihood Methods with Robust Variance Estimation[J]. Communications in Statistics - Simulation and Computation,2015,45(8):3036-3052.
[21] Bakbergenuly I, Kulinskaya E, Morgenthaler S. Inference for binomial probability based on dependent Bernoulli random variables with applications to meta-analysis and group level studies[J]. Biometrical Journal,2016,58(4):896-914.
[22] Bai O, Chen M, Wang X. Bayesian Estimation and Testing in Random Effects Meta-Analysis of Rare Binary Adverse Events[J]. Statistics in Biopharmaceutical Research,2016,8(1):49-59.
[23] Jackson D, Law M, Stijnen T, et al. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio[J]. Stat Med,2018,37(7):1059-1085.

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