RESEARCH IN BRIEF

Double arcsine transform not appropriate for meta-analysis

Christian Röver [iD] | Tim Friede [iD]

Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

Correspondence
Christian Röver, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany.
Email: christian.roever@med.uni-goettingen.de

Abstract
The variance-stabilizing Freeman–Tukey double arcsine transform was originally proposed for inference on single proportions. Subsequently, its use has been suggested in the context of meta-analysis of proportions. While some erratic behavior has been observed previously, here we point out and illustrate general issues of monotonicity and invertibility that make this transform unsuitable for meta-analysis purposes.

KEYWORDS
Freeman–Tukey, prevalences, proportions, variance stabilizing transform

Highlights

What Is Already Known
- Meta-analyses commonly involve transformations for endpoints or effect measures.
- Use of the variance-stabilizing double-arcsine transforms has been proposed in the context of meta-analyses of proportions.
- Some erratic behavior of the double-arcsine transform has been observed previously.

What Is New
- We demonstrate why a double-arcsine transformation is not suitable for meta-analysis.

Potential Impact for Readers Outside the Authors' Field
- While meta-analyses of proportions are common in many application areas, transformations of effect measures should in general be carefully selected to avoid paradoxical behavior, in particular the double-arcsine transformation should not be used. Alternatives are available including the arcsine transformation or one-stage approaches.

1 | INTRODUCTION

Freeman and Tukey[1] proposed the *double arcsine transform* in order to facilitate confidence interval construction or testing of *single* proportions based on a normal approximation on the transformed scale. Within a binomial model, for a given number a of events among a total of N trials (i.e., an

observed *proportion* of $p = a/N$), the transformation is given by

$$\theta = \frac{1}{2}\left(\arcsin\left(\sqrt{\frac{a}{N+1}}\right) + \arcsin\left(\sqrt{\frac{a+1}{N+1}}\right)\right). \quad (1)$$

The double arcsine transform is a *variance stabilizing transformation*, that is, the variance of transformed deviates is (at least approximately) independent of the underlying binomial probability.[2] Differing conventions are common regarding the overall scaling; above, a factor 1/2 was applied, but other definitions are also used.[1–3] It is worth also pointing out the closely related *arcsine transform* here, which is defined as

$$\vartheta = \arcsin\left(\frac{a}{\sqrt{N}}\right) = \arcsin\left(\sqrt{p}\right), \quad (2)$$

and which constitutes the limiting case for large sample sizes N.[4] While the (single) arcsine transform had been introduced in the 1930s,[5,6] Freeman and Tukey proposed the double arcsine variant suggesting improved variance stabilizing properties[1,7,8]; The transforms' differing variances are also illustrated in this note's supplementary material.

In the context of meta-analysis, Miller[9] suggested the use of the double arcsine transform, and also worked out the inverse of the transform in order to back-transform combined estimates to the probability scale, which is essential for interpretation of the results. While the forward transformation is unambiguous, back-transformation requires specification of a corresponding "sample size" N. Miller (1978) suggested the harmonic mean of the original sample sizes here (without providing a rationale for this choice), and alternative conventions have been proposed in the meantime.[10] Some general properties of arcsine transformation methods and related alternatives have been discussed by Lin and Xu.[10]

Schwarzer et al.[3] have previously pointed out certain pathologies arising when performing meta-analyses of prevalences on the double arcsine transformed scale and subsequently back-transforming to the probability scale. In particular, they demonstrated that one may end up with a zero estimate along with a zero-width confidence interval despite having observed positive proportions in the original data. In the present note, we will more closely investigate the double arcsine transform's properties and the origins of pathologies arising from its use in the meta-analysis context.
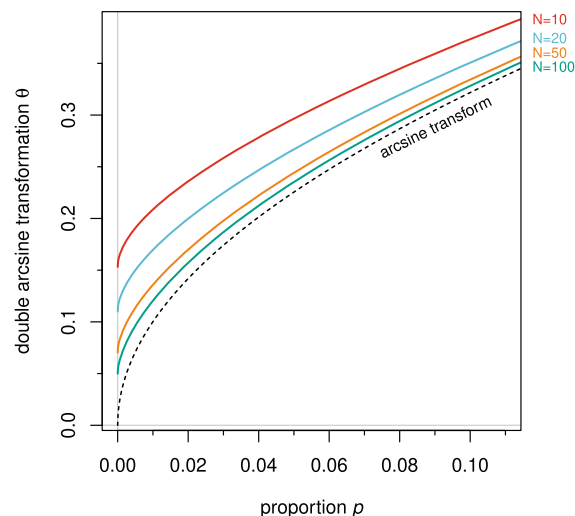
FIGURE 1 Illustration of the double arcsine transform for varying sample sizes N. Effectively, proportions (or proportion *estimates*) corresponding to different sample sizes are not mapped to a common scale; even the ranges of possible θ values are only partly overlapping. The dashed line shows the limiting ("single") arcsine transform [Colour figure can be viewed at wileyonlinelibrary.com]

2 | PROPERTIES OF THE DOUBLE ARCSINE TRANSFORM

In Figure 1, the double arcsine transform is sketched for selected values of N, and in addition the limiting case (for large N) of the arcsine transform is also shown.

For *fixed N*, the transform is bijective and monotonic, but once differing N values are considered, it becomes clear that proportions p are effectively mapped to differing scales. Even the resulting *images*, that is, the ranges of θ values that the proportions are mapped to, are only partly overlapping. These issues do not arise for the limiting ("single") arcsine transform.

When considering differing sample sizes simultaneously, one of the consequences is that the transformation from p to θ in general is not monotonic; an example is given in Table 1. These are results of two studies ("study 10" and "study 13") from the example data quoted by Schwarzer et al.[3] While *study* 10 observed a greater proportion p of events than *study* 13, the order is reversed on the transformed scale (θ).

Figure 2 illustrates the two differing transformations that are effectively applied for the two proportions, resulting in the reverse order on the transformed scale. In such a setup, the combined estimate may eventually also not end up *between* the observed proportions after back-transformation; for the example data from Table 1, this happens, for example, when using the inverse-variance convention[11] to determine the N value used for

TABLE 1 Paradoxical behavior of the double arcsine transform exemplified in the example data reported by Schwarzer et al.[3]; the transformed values (θ) are not proportional to the original proportions (p) here (see also Figure 2)

	Cases (a)	Total (N)	Proportion (p)	Transformation (θ)
Study 10	32	16,557	0.00193	0.0443
Study 13	1	676	0.00148	0.0464
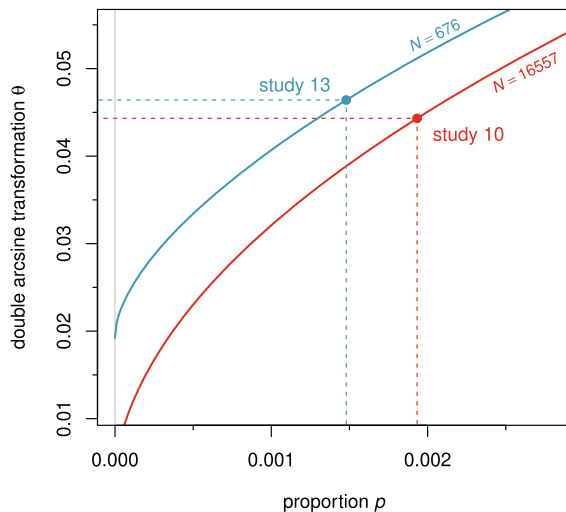
FIGURE 2 Illustration of the double arcsine transformation for the example data from Table 1 [Colour figure can be viewed at wileyonlinelibrary.com]

back-transformation (here: $N = 17,234$), resulting in an estimated proportion of $\hat{p} = 0.00194$, which is greater than both observed proportions. Such effects may easily lead to nonsensical inferences; for example, a meta-analysis of four proportions ($a/N = 1/10$, $10/100$, $100/1000$, $1000/10000$) may yield a confidence interval excluding the common proportion of 10% (when using the harmonic mean sample size for back-transformation; here: $N = 36$). The above calculations are demonstrated in the supplementary material, R code.

As previously pointed out by Schwarzer et al.,[3] another consequence is that for certain (small or large) θ values no corresponding proportion p exists. For example, in Figure 2 a value of $\theta = 0.015$ cannot be mapped to a corresponding proportion p for $N = 676$ (blue line); sometimes such cases are then pragmatically mapped to $p = 0$.

3 | DISCUSSION AND CONCLUSIONS

All the above issues do not pose a problem in the context in which the transform was originally proposed, namely, the investigation of a *single* binomial trial,[1] but it means that proportions from differing sample sizes simply are not transformed to a common scale. While the double

arcsine transformation may have good variance-stabilizing properties, it clearly has also some properties that make it unsuitable for meta-analysis. Use of the transform in a meta-analysis does not only lead to occasional erratic behavior, but also to generally questionable inference. Note that the problems arise once the (forward) transformation is done, and it is not a matter of improving upon the inverse transformation (or finding a suitable N value), as has previously been suggested[12]; any back-transformation (seemingly) avoiding pathologies may in fact only be obscuring the problems. While one might argue that such issues may be less important when all sample sizes are similar, or when all sample sizes are large, it is also important to note that these issues do not arise for the obvious alternative of the closely related plain ("single") arcsine transform.

A number of more appropriate alternative approaches of differing technical complexity are available, as has been discussed, for example, by Schwarzer et al.[3] and Lin and Chu.[13] A very obvious one among these may be the closely related, and in many ways simpler ("single") arcsine transform, which also constitutes the limiting case of the double arcsine transformation for large sample sizes. Recent empirical and numerical comparisons of meta-analysis methods for proportions did not indicate any advantages of the double arcsine transform over the plain ("single") arcsine transform.[13,14] If a two-stage approach, that is, the combination of proportion estimates on a transformed scale, is desired, then alternative transformations should be considered, an obvious candidate being the ("single") arcsine transform. In case there are still concerns regarding the quality of the approximations involved, one-stage models avoiding the intermediate transformation step should be used.[3,13]

In summary, while the double arcsine transform is well-suited for the originally intended purpose of providing a variance stabilizing transformation for a (single) proportion (and in this regard is superior to the arcsine transformation), it is clearly not appropriate for meta-analysis of *several* binomial probability estimates.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplemental material of this article.

ORCID

Christian Röver https://orcid.org/0000-0002-6911-698X
Tim Friede https://orcid.org/0000-0001-5347-7441

REFERENCES

1. Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Stat*. 1950;21(4):607-611. doi:10.1214/aoms/1177729756
2. Guan Y. Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. *Stat Probab Lett*. 2009;79(14):1621-1629. doi:10.1016/j.spl.2009.04.010
3. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods*. 2019;10(3):476-483. doi:10.1002/jrsm.1348
4. Anscombe FJ. The transformation of Poisson, binomial and negative binomial data. *Biometrika*. 1948;35(3–4):246-254. doi:10.1093/biomet/35.3-4.246
5. Bartlett MS. The square root transformation in analysis of variance. *Suppl J R Stat Soc*. 1936;3(1):68-87. doi:10.2307/2983678
6. Eisenhart C. Inverse sine transformation of proportions. In: Eisenhart C, Hastay MW, Wallis WA, eds. *Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production Management Engineering*. McGraw-Hill; 1947:395-416.
7. Freeman MF, Tukey JW. *Transformation Related to the Angular and the Square Root. Memorandum Report 24, Statistical Research Group*. Princeton University; 1949.
8. Blom G. Transformations of the binomial, negative binomial, Poisson and χ^2 distributions. *Biometrika*. 1954;41(3/4):302-316. doi:10.1093/biomet/43.1-2.235
9. Miller JT. The inverse of the Freeman-Tukey double-arcsine transformation. *Am Stat*. 1978;32(4):1238. doi:10.1080/00031305.1978.10479283
10. Lin L, Xu C. Arcsine-based transformations for meta-analysis of proportions: pros, cons, and alternatives. *Health Sci Rep*. 2020;3(3):e178. doi:10.1002/hsr2.178
11. Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. *J Epidemiol Community Health*. 2013; 67(11):974-978. doi:10.1136/jech-2013-203104
12. Doi SA, Xu C. The Freeman-Tukey double arcsine transformation for the meta-analysis of proportions: recent criticisms were seriously misleading. *Journal of. Evid Based Med*. 2021;14(4):259-261. doi:10.1111/jebm.12445
13. Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. *Epidemiology*. 2020;31(5):713-717. doi:10.1097/EDE.0000000000001232
14. Lin L, Xu C, Chu H. Empirical comparisons of 12 meta-analysis methods for synthesizing proportions of binary outcomes. *J Gen Intern Med*. 2022;37:308-317. doi:10.1007/s11606-021-07098-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.