A STATISTICAL MODEL FOR RELIABILITY GENERALIZATION FORMULATED AS A MIXTURE MODEL

#3201

Desirée Blázquez-Rincón[1], Juan I. Durán[2], Juan Botella[2], and Manuel Suero[2]
1. Universidad a Distancia de Madrid 2. Universidad Autónoma de Madrid

Financial support:
Project PID2021-122404NB-I00, Ministerio de Ciencia e Innovación, Gobierno de España

INTRODUCTION

Reliability generalization (RG) is a meta-analytic approach that aims to synthesize reliability estimates from independent studies that had administered the same test. Most RG meta-analyses are based on Cronbach's alpha, as it is the most widely reported. Given the skewed nature of its distribution, Hakstian & Whalen (1976) proposed a transformation for $\hat{\alpha}$ leading to the following point estimate and its sampling variance:

$$\hat{t}_{HW} = \sqrt[3]{1-\hat{\alpha}}$$

$$V(\hat{t}_{HW}) = \frac{18J(N-1)(1-\hat{\alpha})^{2/3}}{(J-1)(9N-11)^2}$$

The "classic" random-effects model and its limitations

RG meta-analysis primarily provide an estimate of the average reliability and the variability across the reliability parameters, that is, the between-study heterogeneity. Both parameters are routinely estimated based on the "classic" random-effects model, which assumes that the differences found among the effect estimates $\hat{\theta}$ ($\hat{\alpha}$ in RG meta-analysis) are a consequence of two sources of error:

- random sampling error, $e_i \sim N(0, \sigma_i)$
- variability across θ, $u_i \sim N(0, \tau)$

Assuming u_i and e_i are independent, the variability of $\hat{\theta}$ is decomposed as (Hedges, 1983):

$$\sigma_{\theta_i} = \sigma_{u_i} + \sigma_{e_i} = \tau + \sigma_i$$

Among other limitations (Suero et al., 2023; see poster #4201), those effect sizes indices where the conditional within-study variance (σ_i) includes the value of the effect size itself (θ_i) leads to a dependence between e_i and $\hat{\theta}_i$, which makes the expression σ_{θ_i} incorrect. This is what happens in the case of RG meta-analysis when the *Hakstian & Whalen* transformation is used.

An alternative from mixture models

An alternative random-effects model can be developed as a mixture model (Suero et al., 2023; see poster #4201), avoiding the limitations of the "classic" model. From a mixture models perspective, the marginal distribution of $\hat{t}_{HW}$ can be defined as:

$$h_{\hat{T}_{HW}}(\hat{t}_{HW}; \mu_{T_{HW}}, \tau_{T_{HW}}, N, j) =$$

$$\int_{-\infty}^{\infty} f_{\hat{T}_{HW}}(\hat{t}_{HW}/t_{HW}; N, j) \cdot f_{T_{HW}}(t_{HW}; \mu_{T_{HW}}, \tau_{T_{HW}}) dt_{HW}$$

Where the theoretical mean and variance of the distribution of $\hat{t}_{HW}$ are respectively expressed as:

$$\mu_{\hat{T}_{HW}} = \mu_{T_{HW}}$$

$$\sigma^2_{\hat{T}_{HW}} = \sum_{i=1}^{k} \frac{1}{k}\left[A_i(\tau^2_{T_{HW}} + \mu^2_{T_{HW}})\right] - \mu^2_{\hat{T}_{HW}}, \text{ where } A_i = \frac{18j(N_i-1)}{(j-1)(9N_i-11)^2} + 1.$$

EMPIRICAL TEST OF ACCURACY

A reliability score was generated in the form of *Hakstian & Whalen* values for a test made up of $j = 20$ items administered to a sample of $N = 100$ subjects in the "fixed sample sizes" condition and to a sample of N = 40, 60, 80, 100, 120, 140, 160, or 180 subjects in the "variable sample size" condition along 10,000,000 replicas/studies.

Data generation

1. Parametric Hakstian & Whalen values (t_{HW}, were i is 1,..., 10,000,000) were extracted from $N(\mu_{T_{HW}} = 0.4642$ – given a $\mu_\alpha = 0.90$ -, $\tau_{T_{HW}} = 0.07)$
2. t_{HW} values were transformed to parametric Cronbach's alpha, α, and their sampling variances, $\sigma^2_{T_{HW}}$, were obtained.
3. Lastly, observed effect sizes for t_{HW}, $\hat{t}_{HW}$, were obtained from $N(\mu = t_{HW}, \sigma^2 = \sigma^2_{T_{HW}})$

Results

Theoretical mean and variance of $\hat{t}_{HW}$ under the mixture model are consistent with the corresponding empirical values of $\hat{t}_{HW}$. The magnitude of the discrepancies are within the expected margin of error with 95% of confidence for the number of replicas.

	Fixed sample sizes			*Variable sample sizes*		
	Empirical	Theoretical	Theoretical - empirical	Empirical	Theoretical	Theoretical - empirical
$\mu_{T_{HW}}$	0.46420	0.46420	0.0000022	0.46419	0.46419	-0.0000021
$\tau_{T_{HW}}$	0.00543	0.00542	-0.0000021	0.00550	0.00550	-0.0000001

REFERENCES

- Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, *41*, 219–231. doi:10.1007/BF02291840
- Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, *93*, 388-395. https://doi.org/10.1037/0033-2909.93.2.388
- Suero, M., Botella, J., Durán, J. I., & Blázquez-Rincón, D. (2023, September 12). Reformulating the meta-analytical random effects model as a mixture model. https://doi.org/10.17605/OSF.IO/V2FDE

PERFORMANCE OF THE HETEROGENEITY ESTIMATOR

A point estimator of the between-study variance/heterogeneity parameter can be derived from the random-effects model expressed as a mixture model:

$$\hat{\tau}^2_{MM} = \frac{\sum_{i=1}^{k} \hat{t}^2_{HW_i}}{\sum_{i=1}^{k} A_i} - \frac{\sum_{i=1}^{k} \frac{1}{k} A_i}{(k-1)}\left[\frac{k\left(\overline{\hat{t}_{HW}}\right)^2}{\sum_{i=1}^{k} \frac{1}{k} A_i} - \frac{\sum_{i=1}^{k} \hat{t}^2_{HW_i}}{\sum_{i=1}^{k} A_i}\right]$$

A Monte Carlo simulation study was carried out to examine the performance of $\hat{\tau}^2_{MM}$ in comparison with the most recommended - restricted maximum-likelihood $\hat{\tau}^2_{REML}$ - and frequently used - DerSimonian-Laird $\hat{\tau}^2_{DL}$ - procedures. Neither of them were truncated to zero.

For a meta-analysis made up of $k = 10, 30, 50, 70, \text{or } 90$ primary studies, a reliability score was generated in the form of *Hakstian & Whalen* values for a test made up of $j = 20$ items administered to an average sample size of $N = 100$ subjects. The number of replicas per condition was 10,000 meta-analysis, resulting in 50,000 generated meta-analysis.

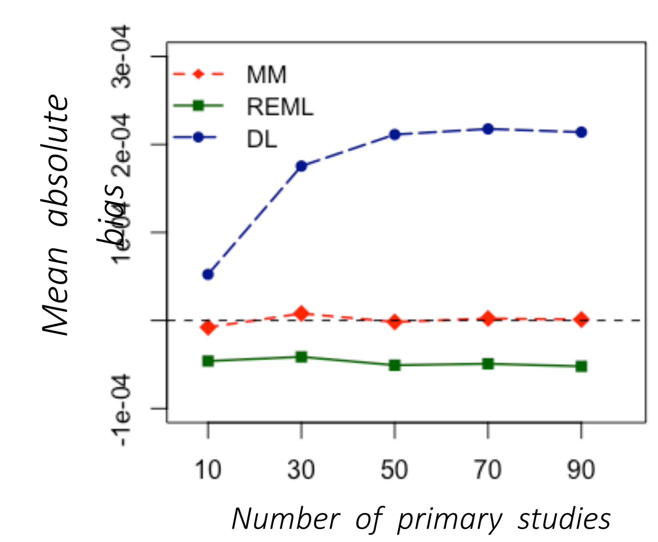
Data generation

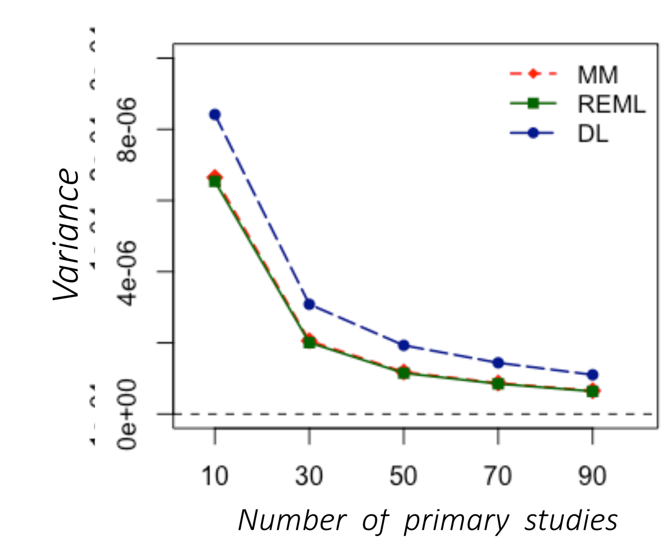
1-3. In order to generate the k observed $\hat{t}_{HW}$ values for each meta-analysis, steps 1 to 4 described in the previous simulation were followed.

4. For each meta-analysis, $\hat{\tau}^2_{MM}$, $\hat{\tau}^2_{DL}$, and $\hat{\tau}^2_{REML}$ were computed.

Results

Compared to the other methods examined, $\hat{\tau}^2_{MM}$ showed the best performance regarding absence of bias without a significant loss in efficiency.





CONCLUSIONS

- ❑ Mixture models are an appropriate framework for conducting RG meta-analysis using the *Hakstian & Whalen* transformation for Cronbach's alpha coefficients.
- ❑ Bias when estimating heterogeneity from empirical data can be reduced by using $\hat{\tau}^2_{MM}$.