Formulas for estimating and pooling Hedges' g parameters in a meta-analysis

Roger B. Newson

May 27, 2016

1 Introduction

The parameter Hedges' g was advocated by Hedges (1981)[1] as a measure of effect of a treatment, to be derived from a 2-sample comparison between treated and untreated subjects, which can be compared, and pooled in a meta-analysis, between studies where different outcomes are measured for the same treatment comparison. It is expected to be useful in areas such as physiotherapy, where there may be no consensus as to how to measure the efficacy of a treatment, and where different symptom scores are used in different studies. An example of an application of Hedges' g appears in Diong *et al.*, 2016[2]. Having done a meta–analysis of Hedges' g parameters on a messy set of journal articles, the current author would like to share some of the formulas used with other meta–analysts for future reference.

2 Formulas

Hedges' g was defined by Hedges (1981)[1] as a general measure of treatment effect, valid when comparing 2 treatment groups regarding a Normally–distributed outcome that is equally variable in the 2 sub–populations from which the treatment groups are sampled. Following Hedges, we assume that there are K studies (or study-outcomes), each featuring 2 treatment groups, an intervention group (with index 1) and a control group (with index 0). For each i from 1 to K, we assume that there are n_{0i} subjects in the control group, and n_{1i} subjects in the intervention group, and that the outcomes measured in Study i are Y_{0ij} for j from 1 to n_{0i} in the control group and Y_{1ij} for j from 1 to n_{1i} in the intervention group. We assume that, for each i, the control values Y_{0ij} are sampled from a common Normal distribution with mean μ_{0i} and standard deviation (SD) σ_{0i}, and that the intervention values Y_{1ij} are sampled from another common Normal distribution with mean μ_{1i} and SD σ_{1i}.

The standard sampling estimators for μ_{0i} and μ_{1i} are the respective sample means,

$$\begin{aligned} \overline{Y}_{0i} = & \ n_{0i}^{-1} \sum_{j=1}^{n_{0i}} Y_{0ij}, \\ \overline{Y}_{1i} = & \ n_{1i}^{-1} \sum_{j=1}^{n_{1i}} Y_{1ij}, \end{aligned} \quad (1)$$

and the standard sample estimators for σ_{0i} and σ_{1i} are the respective sample SDs,

$$\begin{aligned} \overline{SD}_{0i} = & \ \sqrt{(n_{0i}-1)^{-1} \sum_{j=1}^{n_{0i}} \left(Y_{0ij} - \overline{Y}_{0i}\right)^2}, \\ \overline{SD}_{1i} = & \ \sqrt{(n_{1i}-1)^{-1} \sum_{j=1}^{n_{1i}} \left(Y_{1ij} - \overline{Y}_{1i}\right)^2}. \end{aligned} \quad (2)$$

However, if we think that we can assume that the control and intervention SDs are equal in each study ($\sigma_{0i} = \sigma_{1i} = \sigma_i$), then we can estimate each common study–specific SD σ_i by weighting the two squared treatment–specific SDs to give the estimate

$$\overline{SD}_i = \sqrt{\frac{(n_{0i}-1)\overline{SD}_{0i}^2 + (n_{1i}-1)\overline{SD}_{1i}^2}{n_{0i}+n_{1i}-2}}. \quad (3)$$

This estimate enables us to estimate the difference between the intervention and control study mean in each study, expressed in units of the common SD. In the ith study, this difference, in the study population at large, is the ith population Hedges' g, defined as

$$\gamma_i = (\mu_{1i} - \mu_{0i})/\sigma_i, \quad (4)$$

and is estimated using the ith sample Hedges' g, defined as the estimator

$$g_i = \left(\overline{Y}_{1i} - \overline{Y}_{0i}\right) / \overline{SD}_i. \tag{5}$$

For the purposes of making funnel plots or carrying out heterogeneity tests, we may use the standard–error formula for the equal–variance t-test to define an approximate standard error for the individual g_i using the variance formula of Hedges (1981)[1].

In a meta-analysis, we wish to estimate the weighted mean of the γ_i, which is equal to the common value if all the γ_i are equal. Usually, this meta–population parameter is defined as a weighted sum of the γ_i, using weights w_i that we hope are approximately inversely proportional to the sampling variance of the g_i. (These weights are frequently the pooled intervention and control sample numbers $n_i = n_{1i} + n_{0i}$.) It is defined as

$$\overline{\gamma} = \frac{\sum_{i=1}^{K} w_i \gamma_i}{\sum_{i=1}^{K} w_i}. \tag{6}$$

If these weights w_i can be estimated using consistent estimators W_i, then the estimator for the meta–population $\overline{\gamma}$ is the weighted sum

$$\overline{g} = \frac{\sum_{i=1}^{K} W_i \, g_i}{\sum_{i=1}^{K} W_i}. \tag{7}$$

In the case where the w_i are the pooled n_i, the W_i are also the n_i. However, the w_i might be inverse sampling variances for the g_i, and then the W_i may be inverse variance estimates.

2.1 The connection with Somers' D

A possible justification for Hedges' g is as a less robust, but more widely available, substitute for Somers' D (Newson, 2006)[3]. Somers' D can be viewed as a common currency for associations, which can be compared, and meta–analysed, between associations between variables defined on a variety of scales (Newson, 2015)[4]. In the two–sample case considered here, it measures how little overlap there is between the 2 treatment groups (intervention and control).

If the assumption of equal variances between treatment groupss in the same study for the same outcome is true, then the ith Hedges' g is related to the corresponding Somers' D using the formula

$$D_i = 2\Phi\left(\frac{\gamma_i}{\sqrt{2}}\right) - 1, \tag{8}$$

where $\Phi(\cdot)$ is the cumulative standard Normal distribution function, and D_i is the Somers' D, in the ith study sub–population, of outcome with respect to treatment (defined as 0 for control, 1 for intervention).

In practice, of course, we cannot always estimate Somers' D for each study in a meta-analysis, except if we have the original data for each study. Hedges' g has the advantage that journal articles usually provide enough information for the user to estimate the study–treatment means μ_{0i} and μ_{1i} and the study–treatment SDs σ_{0i} and σ_{1i}, allowing us to estimate Hedges' g for each study, using the formula (5).

2.2 Alternative estimates for the study parameters

Usually, when we do a meta-analysis, the report from a component study provides estimates for the study–specific μ_{0i}, μ_{1i}, σ_{0i} and σ_{1i}. In most cases, these are the corresponding $\overline{Y}_{0i}$, $\overline{Y}_{1i}$, $\overline{SD}_{0i}$ and $\overline{SD}_{1i}$. However, sometimes these are not available. In these cases, the meta–analyst must find substitutes.

Sometimes, instead of treatment–group means and SDs, the study reports give treatment–group means with confidence limits. For the ith study, and for h equal to zero for the control sub–sample and 1 for the intervention sub–sample, the lower and upper confidence limits for $\overline{Y}_{hi}$ may be denoted $\overline{Y}_{hi}^{(\text{lower})}$ and $\overline{Y}_{hi}^{(\text{upper})}$, respectively. The estimated SD is then given by the alternative formula

$$\overline{SD}_{hi} = \frac{\sqrt{n_{hi}}\left(\overline{Y}_{hi}^{(\text{upper})} - \overline{Y}_{hi}^{(\text{lower})}\right)}{\text{invt}\left(n_{hi} - 1, 1 - \frac{1}{2}\alpha\right) - \text{invt}\left(n_{hi} - 1, \frac{1}{2}\alpha\right)}. \tag{9}$$

Here, α is defined so that the confidence interval has confidence level $100(1 - \alpha)$, so that $\alpha = 0.05$ for a 95 percent confidence interval, and $\text{invt}(a, b)$ is defined as the inverse cumulative Student's t–distribution function, with a degrees of freedom, of b, as implemented in Stata using the Stata statistical function `invt()` (StataCorp, 2015)statacorp.

Alternatively, a report from a component study may provide confidence limits only for the difference between intervention and control study means, and not for the study means themselves. In these cases, we may denote the lower and upper confidence limits for the ith treated–control difference as $\text{diff}_i^{(\text{lower})}$ and $\text{diff}_i^{(\text{upper})}$, respectively. We may then reconstruct the standard error of the difference from the confidence limits, using the formula

$$\text{SE}_i^{\text{diff}} = \frac{\text{diff}_i^{(\text{upper})} - \text{diff}_i^{(\text{lower})}}{\text{invt}\left(n_{0i} + n_{1i} - 2, 1 - \frac{1}{2}\alpha\right) - \text{invt}\left(n_{0i} + n_{1i} - 2, \frac{1}{2}\alpha\right)}, \tag{10}$$

where α and $\text{invt}(\cdot, \cdot)$ are defined as before. Assuming that the population SDs are equal in the control and treated subpopulations for the ith study, we may then estimate the common SD σ_i using the formula

$$\overline{SD}_i = \frac{\text{SE}_i^{\text{diff}}}{\sqrt{1/n_{0i} + 1/n_{1i}}}. \tag{11}$$

We can then substitute (12) into (5) to estimate the ith Hedges' g.

Alternatively, the study may present no means, standard deviations or standard errors, but only medians and other percentiles, or even only a median and an interpercentile range. In this case, the investigators may have done this because the outcome does not have a perfect Normal distribution. However, the best possible approximation to a Hedges' g may still be to treat the median as the mean, and to estimate the SD from the percentiles, or from the interpercentile range. For each probability q, we will denote by $\xi_{hi}^{(q)}$ the $100q$th population percentile, in treatment group h of study i, and we will denote by $\bar{\xi}_{hi}^{(q)}$ the corresponding estimate of $\xi_{hi}^{(q)}$ from the study report. The treatment–group mean is then estimated as the corresponding median,

$$\overline{Y}_{hi} = \bar{\xi}_{hi}^{(0.5)}. \tag{12}$$

To estimate the SD, we will need 2 percentiles, or at least their difference. For instance, if the percentiles given are the median, the 25th percentile, and the 75th percentile, then we usually estimate the SD using the 25th and 75th percentiles, or even using just their difference, known as the interquartile range, which is sometimes given in a study report without the original percentiles. We will denote by $q^{(\text{lower})}$ and $q^{(\text{upper})}$ the lower and upper proportions corresponding to the reported percentiles, such that the study report has presented percentiles $100q^{(\text{lower})}$ and $100q^{(\text{upper})}$, or just their difference. We will denote by $\xi_{hi}^{(\text{lower})}$ and $\xi_{hi}^{(\text{upper})}$ the lower and upper population percentiles for treatment group h in study i, and denote by $\bar{\xi}_{hi}^{(\text{lower})}$ and $\bar{\xi}_{hi}^{(\text{upper})}$ the corresponding sample percentiles, which were given in the report (or at least their difference was). The population SD for treatment group h of study i is then given by the formula

$$\sigma_{hi} = \frac{\xi_{hi}^{(\text{upper})} - \xi_{hi}^{(\text{lower})}}{\Phi^{-1}\left(q^{(\text{upper})}\right) - \Phi^{-1}\left(q^{(\text{lower})}\right)}, \tag{13}$$

where $\Phi^{-1}(\cdot)$ is the inverse standard Normal cumulative distribution function, computed in Stata using the `invnormal()` function[5]. This SD can therefore be estimated using the formula

$$\overline{SD}_{hi} = \frac{\bar{\xi}_{hi}^{(\text{upper})} - \bar{\xi}_{hi}^{(\text{lower})}}{\Phi^{-1}\left(q^{(\text{upper})}\right) - \Phi^{-1}\left(q^{(\text{lower})}\right)}. \tag{14}$$

Note that this formula depends only on the difference between the upper and lower percentiles. Therefore, if $q^{(\text{lower})} = 0.25$ and $q^{(\text{upper})} = 0.75$, then the formula (15) requires only the interquartile range.

2.3 The positive–beneficial Hedges' g

Sometimes, the meta–analysis is complicated by the fact that some of the study outcomes are higher if the subject is doing well, and other study outcomes are higher if the subject is doing badly. In this case, we need to summarize the results using a revised Hedges' g, known as a positive–beneficial Hedges' g, which will be positive if subjects do better (on average) in the intervention group (as is expected to happen if the treatment is beneficial), and negative if subjects do less well (on average) in the intervention group (as is expected to happen if the treatment is harmful).

To deal with this possibility, we define the study–specific beneficiality sign for the ith study (or study–outcome) as

$$\beta_i = \begin{cases} 1, & \text{if higher } Y_{hij} \text{ values are better,} \\ -1, & \text{if lower } Y_{hij} \text{ values are better.} \end{cases} \tag{15}$$

We can then redefine the ith study–specific population Hedges' g by modifying formula (4) as

$$\gamma_i^* = \beta_i \gamma_i, \tag{16}$$

and redefine the corresponding ith study–specific sample Hedges' g by modifying (5) as

$$g_i^* = \beta_i g_i, \tag{17}$$

for which an approximate standard error, for funnel plots and heterogeneity ttests, is once again given by the variance formula of Hedges (1981)[1].

The parameter we need to estimate, in order to summarize the benefit of intervention, is then

$$\overline{\gamma}^* = \frac{\sum_{i=1}^K w_i \gamma_i^*}{\sum_{i=1}^K w_i}, \tag{18}$$

which we estimate using the weighted mean

$$\overline{g}^* = \frac{\sum_{i=1}^K W_i \, g_i^*}{\sum_{i=1}^K W_i}. \tag{19}$$

Note that, if this method is to work as advertized, then the β_i should be decided *a priori*, before knowing the directions of the g_i.

3 Further extensions

In the simplest case, all the K studies can be assumed mutually independent. However, this is not necessarily the case. In some complicated meta–analyses, we have multiple outcomes per study, each with its own beneficiality sign, and maybe also with its own weight, because some outcomes were measured in more subjects than others.

In those cases, we can make the observational unit a study–outcome instead of a study, and use clustered Huber variances to estimate the standard errors and confidence limits for estimating the parameter (19) using the statistic (20). These clustered and weighted Huber variances use the assumption that we are sampling studies independently from a population of studies, instead of sampling study–outcomes from a population of study–outcomes.

References

[1] Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 1981; **6(2)**: 107–128.

[2] Diong J, Allen N, Sherrington C. Structured exercise improves mobility after hip fracture: a meta–analysis with meta–regression. *British Journal of Sports Medicine* 2016; **50**: 346–355.

[3] Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *The Stata Journal* 2006; **6(3)**: 309–334. Download from *http://www.stata-journal.com/article.html?article=snp15_6*

[4] Newson RB. Somers' D: A common currency for associations. Presented at the 21st UK Stata User Meeting, 10–11 September, 2015. Download from *https://ideas.repec.org/p/boc/usug15/01.html*

[5] StataCorp. *Stata: Release 14. Statistical Software.* College Station, TX: StataCorp LP; 2015.