metaumbrella: An R Package for Conducting Umbrella Reviews

Corentin J. Gosling*
Paris Nanterre University, France

Aleix F. Solanes*
Institut d'Investigacions Biomediques
August Pi I Sunyer, Barcelona, Spain.

Paolo Fusar-Poli
King's College London, UK

Joaquim Radua
Institut d'Investigacions Biomediques
August Pi I Sunyer, Barcelona, Spain.

Abstract

This supplementary materials is designed to complement the paper submitted to *'Evidence-Based Mental Health'* describing the **metaumbrella** package. In Section 1, we present details on the calculations performed by the package to conduct the main analyses (meta-analytic models, assessment of publication bias and excess of statistical significance, and stratification of the evidence). In the Section 2, we present the calculations performed by the package to adapt to various user inputs and to convert effect sizes from one to another.

Keywords: evidence, umbrella review, meta-review, meta-analysis, R.

1. Umbrella calculations

1.1. Effect size measures

The **metaumbrella** package allows to work with different effect size measures. For studies comparing means, users can work with standardized mean differences (SMD), Hedges' g (G), or mean difference (MD). It is worth noting that, based on the published literature, SMD can alternatively be used to describe a Cohen's d or a Hedges' g measure (Higgins, Thomas, Chandler, Cumpston, Li, Page, and Welch 2019). For the sake of clarity, we use SMD to refer only to a Cohen's d. For studies comparing frequencies, users can work with odds ratio or its logarithm (OR), or risk ratio or its logarithm (RR). For studies comparing incidence and hazard rates, users can work with hazard ratio or its logarithm (HR), and incident rate ratio or its logarithm (IRR). For studies exploring associations between dimensional variables, users can work with raw Pearson's correlation coefficients (R) or Fisher's z (Z).

1.2. Meta-analytic models

In the **metaumbrella** package, users can fit either fixed-effect or random-effects meta-analytic models (Hedges and Olkin 1985). The fixed-effect model assumes that the observed differences

in effect sizes between studies arise from sampling error. Therefore, this type of model should mainly be used to pool effect sizes coming from studies with similar methods (such as in the dosage of the intervention or the tool used to measure the outcome), and similar sample characteristics (such as in the age, sex, or severity of the condition). Considering $i = 1, \ldots, k$ independent effect sizes of a true effect size, the fixed-effect model is given by

$$es_i = \theta + \epsilon_i \quad (1)$$

where es_i denotes the observed effect in the i-th study, θ is the shared common true effect and ϵ_i a within-study error in the i-th study.

Contrary to the fixed-effect model, the random-effect model assumes that the observed differences in effect sizes arise not only from sampling error but also because different studies estimate different true effects. Thus, this specification allows combining effect sizes that derive from studies with differences in their methods or in their sample characteristics. Considering $i = 1, \ldots, k$ independent effect sizes of a true effect size, this random-effects model is given by

$$es_i = \mu + \beta_i + \epsilon_i \quad (2)$$

where es_i denotes the observed effect in the i-th study, μ is the average true effect across studies, β_i the between-study error for the i-th study, ϵ_i a within-study error in the i-th study.

By default, the between-study variance is estimated using a restricted likelihood maximum estimator, but four other estimators (DerSimonian-Laird, maximum-likelihood estimator, Paule-Mandel estimator or Hartung-Knapp-Sidik-Jonkman) are available. To fit the meta-analyses, the `umbrella()` function relies on the `metagen()` function from the R **meta** package (Balduzzi, Rucker, and Schwarzer 2019).

Afterwards, the `umbrella()` function extracts several meta-analytic statistics (the overall pooled estimate and its 95% confidence interval and p value, three heterogeneity indicators: tau^2, I^2 and Q statistics), calculates the 95% prediction interval and estimates whether the largest study included in the meta-analysis has a significant effect.

1.3. Non-independence of effect sizes

A core assumption of standard meta-analytic models is that all effect sizes come from independent participants and experiments. However, this assumption is frequently violated as some form of dependence often arises between effect sizes (Jackson, Riley, and White 2011). The **metaumbrella** package distinguishes three forms of dependence and proposes a solution to handle each of them. First, dependence can be observed when effect sizes are nested within a larger factor. For example, this situation occurs when several effect sizes originate from either multiple independent studies reported in the same paper or from multiple independent studies reported in different papers, but conducted by the same research laboratory. We name this type of dependence *hierarchical* dependence hereafter. Second, dependence can be observed when effect sizes are generated from the same participants. For example, this situation occurs when several effect sizes are derived from the same participants who have completed multiple outcomes at a unique time-point or who have completed the same outcome at multiple time-points. We name this type of dependence *multivariate* dependence hereafter. Finally, dependence may be observed when effect sizes are generated from the partly

same participants. This situation occurs when several effect sizes of a meta-analysis originate from studies that compare independent experimental or exposed groups to a unique control or non-exposed group. We name this type of dependence *partial* dependence hereafter.

When *hierarchal* dependence is present in the data, a combined effect size across dependent studies is computed (Borenstein, Hedges, Higgins, and Rothstein 2009). More precisely, all dependent effect sizes nested within a larger factor are resumed to a unique effect size by performing a fixed-effect meta-analysis. The effect size and the variance of this independent effect size are equal to the pooled effect size and its variance in the fixed-effect meta-analysis. The sample size associated with this unique effect size is equal to the sum of the sample size of each independent subgroup.

When *multivariate* dependence is present in the data, a combined effect size across outcomes or time-points derived from the same units is computed. More precisely, all dependent effect sizes derived from the same units are resumed to a unique effect size by estimating the non-weighted mean of all effect sizes (Borenstein *et al.* 2009). The correlation between these effect sizes (that can be specified by the user) is used to calculate the variance of this combined effect size, as derived from standard formula (Borenstein *et al.* 2009). The sample size associated with this unique effect size is equal to the largest sample size that completed an outcome or time-point.

When *partial* dependence is present in the data, the shared group is split into several independent subgroups of smaller sample size, as described in the Cochrane Handbook (Higgins *et al.* 2019). More precisely, the number of participants in each independent subgroup is obtained by dividing the total number of participants in a shared group by the number of non-shared groups. These corrected sample sizes are used to re-estimate the effect sizes and their variance.

1.4. Publication bias

To assess the presence of publication bias, the approach described by Egger, Smith, Schneider, and Minder (1997) and Sterne and Egger (2005) is used. This approach proposes to conduct a weighted linear regression in which the effect sizes of the individual studies are regressed against their precision (their standard error). If an association between the effect sizes and their precision is found, this can be interpreted as an indication of publication bias.

$$es_i \quad = \quad \beta_0 + \beta_1 * SE_i \tag{3}$$

where es_i denotes the observed effect in the i-th study and SE_i denotes the standard error of the i-th study. This regression is weighted by the inverse of the variance of the effect sizes $\frac{1}{SE^2}$

When the effect size is a ratio (OR, RR, HR, or IRR), the logarithm of the effect size is used:

$$\log es_i \quad = \quad \beta_0 + \beta_1 * SE_i \tag{4}$$

No publication bias assessment is conducted if the meta-analysis includes less than three studies.

1.5. Test for excess of significance

This test estimates the probability that the number of statistically significant studies in the meta-analysis is as large or more than the observed number of statistically significant studies. If the probability of observing as many significant studies is very small, we may conclude that there are too many more statistically significant studies than we could expect by chance, indicating the possibility of data tortures or reporting biases (Ioannidis and Trikalinos 2007). This test is conducted automatically when running the `umbrella()` function but users who are interested in assessing the excess of significance bias without performing an umbrella review can use the `esb.test()` function available in the **metaumbrella** package. The test for an excess of significance is a simple binomial (or χ^2) test, in which the expected number of statistically significant studies is the sum of the statistical power of the studies (after assuming that the best approximation of the true effect size is the effect size of the largest study, the pooled effect size, or any other estimate given by the user). The following paragraphs refer to the strategies followed to estimate the statistical power of each included study depending on the effect size measure.

- **SMD.** To estimate the power of studies reporting SMD, the `esb.test()` function starts by dividing the best approximation of the true SMD by the standard error of each study. This allows to estimate, for each study, a t-value (`t = true_d / se`) that is then used to estimate the power according to standard formulas (Cohen 1988). Note that if the returned estimated power is larger than 1, the `esb.test()` function uses 1.

- **OR.** Prior to calculation, if any number of cases/controls in the exposed and non-exposed groups is equal to zero, the `esb.test()` function adds 0.5 to the four groups (Weber, Knapp, Ickstadt, Kundt, and Glass 2020).

  First, the function estimates the odds of exposition in controls as the average of the observed odds in the controls sample and the indirect estimation of the odds from the cases sample according to the best approximation of the true OR, weighted by $n_{controls} * (1 + o_{cases})$ for controls and $n_{cases} * (1 + o_{controls})$ for cases, with

$$o_{cases} = \frac{n_{cases\_exp}}{n_{cases\_nexp}} \tag{5}$$

$$o_{controls} = \frac{\frac{w_{controls} * n_{controls\_exp}}{n_{controls\_nexp}} + \frac{w_{cases} * n_{cases\_exp}}{n_{cases\_nexp}}/OR}{w_{controls} + w_{cases}} \tag{6}$$

  where $w_{controls} = n_{controls\_exp} + n_{controls\_nexp}$, $w_{cases} = n_{cases\_exp} + n_{cases\_nexp}$, and where n_{cases} and $n_{controls}$ are the total number of cases and controls, n_{cases_exp}, n_{cases_nexp}, $n_{controls\_exp}$ and $n_{controls\_nexp}$ are the number of cases and controls in the exposed and non-exposed groups.

  Second, it then estimates the odds of exposition in cases multiplying the odds in controls by the best approximation of the true OR .

  Third, it simulates thousands of studies with these parameters creating random numbers of exposed in cases and controls according to binomial distributions with $\pi_{cases} = o_{cases}/(1 + o_{cases})$ and $\pi_{controls} = o_{controls}/(1 + o_{controls})$. Note that $o_{cases}/(1 + o_{cases})$ is the probability of being exposed in the cases sample, and $o_{controls}/(1 + o_{controls})$ is the probability of being exposed in the controls sample.

  Finally, it estimates the statistical power as the proportion of these studies with statistically significant findings.

- **RR.** First, the `esb.test()` function estimates the incidence of the event in non-exposed as the average of the observed incidence in non-exposed and the indirect estimation of the incidence from the exposed sample according to the best approximation of the true RR, weighted by the sample sizes

$$I_{nexp} = \frac{\frac{w_{nexp} * n_{cases\_nexp}}{n_{nexp}} + \frac{w_{exp} * n_{cases\_exp}}{n_{exp}} / RR}{w_{nexp} + w_{exp}} \quad (7)$$

  where $w_{nexp} = n_{nexp}$, $w_{exp} = n_{exp}$, and n_{exp} and n_{nexp} are the number of participants in the exposed and non-exposed groups.

  Second, it estimates the incidence of the event in exposed multiplying the incidence of the event in non-exposed by the best approximation of the true RR.

  Third, it simulates thousands of studies with these parameters creating random numbers of cases in exposed and non-exposed according to binomial distributions with $\pi_{exp} = I_{exp}$ and $\pi_{nexp} = I_{nexp}$.

  Finally, it estimates the statistical power as the proportion of these studies with statistically significant findings.

- **IRR.** First, the `esb.test()` function estimates the incidence of the event in non-exposed as the average of the observed incidence in non-exposed and the indirect estimation of the incidence from the exposed sample according to the best approximation of the true IRR, weighted by the times

$$IR_{nexp} = \frac{\frac{w_{nexp} * n_{cases\_nexp}}{time_{nexp}} + \frac{w_{exp} * n_{cases\_exp}}{time_{exp}} / IRR}{w_{nexp} + w_{exp}} \quad (8)$$

  where $w_{nexp} = time_{nexp}$, $w_{exp} = time_{exp}$ and where $time_{exp}$ and $time_{nexp}$ are the person-time of disease-free observation in the exposed and non-exposed groups.

  Second, it estimates the incidence of the event in exposed multiplying the incidence of the event in non-exposed by the best approximation of the true IRR.

  Third, it simulates thousands of studies with these parameters creating random numbers of cases in exposed and non-exposed according to Poisson distributions with $\lambda_{exp} = IR_{exp} * time_{exp}$ and $\lambda_{nexp} = IR_{nexp} * time_{nexp}$.

  Finally, it estimates the statistical power as the proportion of these studies with statistically significant findings.

- **HR.** First, the `esb.test()` function estimates the ratio between the numbers of exposed and non-exposed groups. The `esb.test()` function estimates this number empirically to match the statistical power of the study (which could differ depending on factors such as the inclusion of one or other covariate in the study analysis). Specifically, it uses the `optim()` function to find the ratio associated with 50% power to detect the HR reported in the study with the p value reported in the study. Afterwards, it calls the `powerCT.default0()` function from the **powerSurvEpi** package (Qiu, Chavarro, Lazarus, Rosner, and Ma 2021) to estimate the power to detect the best approximation of the true HR with the estimated ratio and p value = 0.05.

1.6. Stratification of evidence

The `add.evidence()` function of the **metaumbrella** package has been created to stratify evidence according to several criteria. Two pre-established criteria are proposed but users can also use some personalized criteria to adapt to the requirements of their umbrella review.

- **Pre-established criteria.**

  1. **Ioannidis.** The first pre-established criteria are those proposed by Prof Ioannidis (Fusar-Poli and Radua 2018). These criteria propose to stratify evidence in five ordinal classes: "Class I", "Class II", "Class III", "Class IV", "Class ns", the "Class I" being the highest class that can be reached. The criteria for each class are the following:
     - Class I: number of cases > 1000, p value of the meta-analysis $< 10^-6$, $I^2 < 0.5$, 95% prediction interval excluding the null, p value of the Egger test $> .05$ and p value of the test for excess significance bias $> .05$.
     - Class II: number of cases > 1000, p value of the meta-analysis $< 10^-6$, largest study with a statistically significant effect and class I criteria not met.
     - Class III: number of cases > 1000, p $< 10^-3$ and class I-II criteria not met.
     - Class IV: p < 0.05 and class I-III criteria not met.
     - Class ns: p >= 0.05.
  2. **GRADE.** The second pre-established criteria are inspired by the GRADE criteria (Guyatt, Oxman, Vist, Kunz, Falck-Ytter, Alonso-Coello *et al.* 2008). Importantly, this algorithmic approach should not be taken as an equivalent to the approach underlying the standard GRADE criteria. However, in line with the standard GRADE approach, the GRADE classification used in the **metaumbrella** package stratifies evidence according to four ordinal classes ("High", "Moderate", "Low", "Very low") and use a downgrading procedure. All factors start with a "High" evidence class that could then be downgraded depending on four indicators:
     - Imprecision: a total number of participants included in the meta-analysis giving a lower power than 80% to detect an SMD = 0.20 leads to a downgrading of 1 class. A number of participants giving a lower power than 80% to detect an SMD = 0.50 leads to a downgrading of 2 classes. Note that for IRR, the number of cases and controls to estimate the power is assumed to be equal to half the number of cases included in the meta-analysis.
     - Limitations: a proportion of participants included in studies at low risk of bias $< 75\%$ leads to a downgrading of 1 class. A proportion $< 50\%$ leads to a downgrading of 2 classes.
     - Publication bias: a p value at an Egger test $< .10$ leads to a downgrading of 1 class.
     - Inconsistency: an I^2 value >= 50% leads to a downgrading of 1 class.

- **Personalized criteria.** Because the criteria used to stratify evidence can vary depending on the aim of the umbrella review, the `add.evidence()` function offers the possibility to select the criteria used to stratify evidence as well as the cut-off values to reach each class. Similarly to the "Ioannidis" criteria, evidence is stratified in 5 ordinal

classes, from "Class I" to "Class V" (the "Class I" being the highest class that can be reached).

A total of 13 criteria can be used to stratify evidence: (1) the number of studies included in the meta-analysis, (2) the total number of participants included in the meta-analysis, (3) the number of cases included in the meta-analysis, (4) the p value of the pooled effect size, (5) the inconsistency between individual studies (I^2 statistics), (6) the imprecision of the pooled effect size (the statistical power of the meta-analysis to detect a given SMD), (7) the percentage of participants included in studies at low risk of bias, (8) the methodological quality of the systematic review (AMSTAR score) (Shea, Reeves, Wells, Thuku, Hamel, Moran *et al.* 2017), (9) the p value at the Egger test for publication bias, (10) the p value of the Ioannidis' for excess of significance bias, (11) the p value of the Jackknife meta-analysis, (12) the inclusion of the null value in the 95% CI prediction interval and (13) the statistical significance of the largest study (i.e., with the smallest variance) included in the meta-analysis.

Users can select any of the 13 criteria (minimum 1 and maximum 13) and must set the threshold scores for each selected criteria to reach the 5 possible classes.

2. Adaptation to various inputs

One of the key advantages of the `umbrella()` function over other statistical software and R packages designed to perform meta-analyses lies in the possibility of offering users automatic fitting of numerous meta-analytic models based on a large variety of inputs data. Therefore, users may extract the data reported in the articles without the necessity of undertaking homogenization work if the available information differs between articles. To adapt to the various inputs, the `umbrella()` function includes many internal functions that convert several input statistics into the effect sizes required to conduct the umbrella review.

2.1. Obtention of the value of the effect size

- **SMD, MD, G.** These three effect size measures are used to quantify the differences between one experimental and one control group on some quantitative, normally distributed dependent variable.

  - SMD is obtained by the following formulas

$$SMD = \frac{mean\_cases - mean\_controls}{pooled\_sd} \tag{9}$$

  where $mean\_cases$ and $mean\_controls$ are equal to the means of the two groups and where $pooled\_sd$ is equal to

$$pooled\_sd = \sqrt{\frac{(n\_cases - 1) * sd\_cases^2 + (n\_controls - 1) * sd\_controls^2}{df}} \tag{10}$$

  where sd_cases and $sd\_controls$ are equal to the standard deviations of the two groups, n_cases and $n\_controls$ are the sample sizes of the two groups and df is equal to $n\_cases + n\_controls - 2$

- G is obtained by adding a correction to the SMD for the positive bias (Hedges and Olkin 1985).

$$G = SMD * J \tag{11}$$

where J is equal to

$$J = exp(\log_{gamma}(df/2) - 0.5 * \log(df/2) - \log_{gamma}((df - 1) / 2)) \tag{12}$$

We implemented this correction using the R functions `exp()` and `lgamma()` instead of `gamma()` to avoid numerical errors when the degrees of freedom are large (Albajes-Eizagirre, Solanes, and Radua 2018)

- For MD, users must directly enter this value.

- **OR and RR.** These two effect size measures are used to quantify the differences between exposed and non-exposed groups on some dichotomous dependent variables. OR and RR are obtained using the following formulas

$$OR = \frac{(n\_cases\_exp \,/\, n\_cases\_nexp)}{(n\_controls\_exp \,/\, n\_controls\_nexp)} \tag{13}$$

$$RR = \frac{(n\_cases\_exp \,/\, n\_exp)}{(n\_cases\_nexp \,/\, n\_nexp)} \tag{14}$$

where n_exp and n_nexp are numbers of participants in the exposed and non-exposed groups, n_cases_exp and $n\_controls\_exp$ to are the numbers of cases and controls in the exposed group and n_cases_nexp and $n\_controls\_nexp$ are the numbers of cases and controls in the non-exposed group.

Note that if any of the n_cases_exp, n_cases_nexp, $n\_controls\_exp$ or $n\_controls\_nexp$ is equal to zero, 0.5 is added to the four values (Weber *et al.* 2020). That said, studies with no participants exposed to the risk factor or with 0 cases are eliminated because they provide no information.

- **IRR.** This effect size measure is used to compare the incidence rates of events occurring at any given point in time between exposed and non-exposed groups. For this measure, we use

$$IRR = \frac{(n\_cases\_exp \,/\, time\_exp)}{(n\_cases\_nexp \,/\, time\_nexp)} \tag{15}$$

where n_cases_exp and n_cases_nexp are the numbers of cases in the exposed and non-exposed groups and time_exp and time_nexp are the person-time rates of the exposed and non-exposed groups.

- **HR.** This effect size measure is used to compare hazard rates of events between exposed and non-exposed groups. Users must directly enter this value.

2.2. Obtention of the variance of the effect size

When information on the variance of the effect size is not directly reported in the dataset, the `umbrella()` function includes several functions to estimate the variance of the effect sizes from raw information.

3.2.1. Using raw information

- **SMD, MD, G.** For SMD and G, their variance is estimated as follows

$$var_{SMD} = \frac{1}{n\_cases} + \frac{1}{n\_controls} \quad (16)$$

$$var_{G} = var_{SMD} + (1 - (df - 2)/(df * J^2)) * G^2 \quad (17)$$

  For MD, users must enter the variance or any information to estimate it (i.e., the standard error or the 95% CI, see formula in Section 2.2.2).

- **OR, RR.** The standard formulas allowing to estimate the variance of the logarithm of OR and RR are

$$var_{\log(OR)} = \frac{1}{n\_cases\_exp} + \frac{1}{n\_cases\_nexp} + \frac{1}{n\_controls\_exp} + \frac{1}{n\_controls\_nexp} \quad (18)$$

$$var_{\log(RR)} = \frac{1}{n\_cases\_exp} - \frac{1}{n\_exp} + \frac{1}{n\_cases\_nexp} - \frac{1}{n\_nexp} \quad (19)$$

  For OR, when the information regarding these sample sizes is not available and that any information to estimate the variance is available (i.e., the 95% CI, the standard error, or the variance), another approach is used. In this case, an estimation of the variance is provided using the value of the OR and the total number of cases and controls. Specifically, a function simulates all combinations of the possible number of exposed/non-exposed participants compatible with both the value of the OR and the total number of cases and controls reported and averages the corresponding variances.

- **IRR.** For this effect size, the variance of the logarithm of IRR is estimated as

$$var_{\log(IRR)} = \frac{1}{n\_cases\_exp} + \frac{1}{n\_cases\_nexp} \quad (20)$$

  When the IRR and its standard error are known, the `umbrella()` function (re)estimates the number of exposed and the time of exposition from IRR. The aim of this action is two-fold. On the one hand, the `umbrella()` function estimates any missing number of exposed or time of exposition. On the other hand, it makes $var_{\log(IRR)}$ coincide with the original analyses even when those included covariates. Otherwise, $var_{\log(IRR)}$ would be unfairly larger in studies that controlled for potential sources of variability. The formulas to conduct the (re)estimation of the number of exposed are based on the above formulas of IRR and $var_{\log(IRR)}$, although the function uses the R function `optim` to avoid squared roots of negative numbers in internal calculations.

- **HR.** Users must enter the variance of the HR or any information allowing to estimate it (i.e., the standard error or the 95% CI).

3.2.2. Using the 95 percent CI

For studies in which users report neither the variance nor the standard error of the effect size, nor the raw information allowing to estimate it, this information is obtained from the 95% CI.

- **SMD, MD, G.** The variance of these effect size measures is converted from the 95% CI using the formula

$$var_{differences} = (\frac{upper\ bound\ 95\%CI - lower\ bound\ 95\%\ CI}{2 * qt(0.975, df)})^2 \quad (21)$$

  where $qt(x, df)$ returns the value of the inverse cumulative density function of the Student t distribution given a variable (x) and degrees of freedom (df).

- **OR, RR, HR, IRR.** The variance of these effect size measures is converted from the 95% CI using the formula

$$var_{ratios} = (\frac{\log(upper\ bound\ 95\%\ CI) - \log(lower\ bound\ 95\%\ CI)}{2 * qnorm(0.975, 0, 1)})^2 \quad (22)$$

  where $qnorm(x, \mu, SD)$ returns the value of the inverse cumulative density function of the normal distribution given a variable (x), a population mean (μ) and population standard deviation (SD).

2.3. Conversions between effect sizes

In three instances, the **metaumbrella** package performs conversions between effect size measures.

1. When the input effect size measure is G and/or MD, they are first automatically converted into an SMD to have all mean comparisons on the same scale. The Ioannidis' test for excess of significance bias is performed on these SMD values. However, SMD values are automatically converted into G before running meta-analytic calculations and Egger's test for publication bias.

   To convert the MD into an SMD, the variance of the MD or any information allowing to estimate it (the standard error or the 95% CI) is required. Then, the variance of the MD is used to obtain the pooled standard deviation of the MD, which then allows to calculate the SMD

$$pooled_{sd} = \frac{\sqrt{var_{MD}}}{\sqrt{1/n_{cases} + 1/n_{controls}}} \quad (23)$$

$$SMD = \frac{MD}{pooled_{sd}} \quad (24)$$

   In line with Formula 11, to convert the G into an SMD, the formula used is

$$SMD = \frac{G}{J} \quad (25)$$

$$var_{SMD} = var_G - (1 - (df - 2)/(df * J^2)) * G^2 \quad (26)$$

2. When the effect size measures vary within the same factor, they are converted in the same metric to allow the realization of the meta-analysis. This situation may occur, for example, when the authors of an umbrella review pool together effect sizes from several meta-analyses or pool together effect sizes found in a systematic review that reported effect sizes in different metrics. The `umbrella()` function performs three types of conversion. It converts OR to SMD (situation 1), RR to OR (situation 2), and HR to OR (situation 3). When SMD is included in a meta-analysis with multiple effect size measures, it is used as the main effect size measure and OR, RR and HR are converted into an SMD (in the case of RR and HR, they are first converted into an OR, and are then converted into an SMD). When SMD is not included in a meta-analysis with multiple effect size measures, OR is used as the main effect size measure and RR and HR are converted into an OR. IRR is not converted by the `umbrella()` function.

   Situation 1: when users report both SMD and OR within the same meta-analysis, the OR is converted into an SMD using the standard formula (Borenstein *et al.* 2009)

$$SMD = \frac{\log(OR) * \sqrt{(3)}}{\pi} \tag{27}$$

   Situation 2: when users report both OR and RR within the same factor, the RR is converted into an OR. Two distinct approaches can be used to perform this conversion. First, when users indicate the number of cases and controls in the exposed and non-exposed groups, this information is used to estimate an OR using the standard formula to estimate this effect size (Section 13). In contrast, if users provide only the value of the RR plus any information regarding the variance (i.e., either the variance, standard error or 95% CI) plus the total number of cases and controls, the number of cases and controls in both the exposed and non-exposed groups are estimated using approach described in Section 2.4 and then the OR is calculated using the standard formula to estimate this effect size.

   Situation 3: when users report both OR and HR within the same factor, the HR is assumed to be equal to an OR.

3. Last, the `umbrella()` function reports all pooled effect sizes of meta-analyses in their original metric but also in equivalent odds ratio (eOR) and equivalent Hedges' G (eG) to facilitate the comparison of effect sizes between meta-analyses.

   Pooled effect sizes expressed as eG are converted into an eOR using the formula described in (Borenstein *et al.* 2009).

$$eOR = exp(\frac{eG * \pi}{\sqrt{3}}) \tag{28}$$

   Pooled effect sizes expressed as OR, RR, HR and IRR are assumed to be equal to an eOR.

$$eOR = OR = RR = HR = IRR \tag{29}$$

   Pooled effect sizes expressed as OR, RR, HR and IRR are converted into an eG using the formula described in (Borenstein *et al.* 2009).

$$eG = \frac{\log(eOR) * \sqrt{3}}{\pi} \tag{30}$$

   where eOR can be any of OR, RR, HR or IRR.

2.4. Obtention of missing variables

The `umbrella()` function derives missing variables from existing variables, either from obvious relationships (e.g., the number of cases in the non-exposed sample is the total number of cases minus the number of cases in the exposed sample), or from standard formulas (e.g., from the formula of the variance).

If there is no formula to obtain the missing figure, in some cases the missing variable can have any value as far as some relationships are kept. In such cases, the simplest value is used. For instance, in studies reporting IRR but with missing follow-up times in the exposed and non-exposed groups, the function sets the overall time to 1 and then splits it among the exposed and non-exposed groups to match the reported IRR.

Otherwise, the function finds the value that best matches the reported data. For instance, when working with OR or RR, users do not necessarily have the information on the number of cases and controls in the exposed and non-exposed groups and may report only the effect size value and the overall number of cases and controls. In this case, the `umbrella()` function simulates all combinations of the possible number of cases and controls in the exposed and non-exposed groups compatible with the actual value of OR or RR. Then, it selects the combination whose variance coincides with the variance reported. Similarly, when the number of cases in the exposed and non-exposed groups is not reported when working with IRR, the `umbrella()` function uses the `optim()` to find the number of cases in the exposed and non-exposed groups that results in a variance that coincides with the reported variance. Afterwards, it recalculates the times so that the resulting IRR coincides with the reported IRR.

2.5. Unrounding of extracted effect size estimates

To replicate the meta-analyses in an umbrella review, it is necessary to rely on information reported in the articles when the raw data are not shared publicly. Among the different pieces of information that permit replication of the meta-analyses, the effect sizes of the individual studies along with their 95% CI are often available in forest plots. However, authors must round off the information reported which leads to a decrease in the precision when using this information to replicate the meta-analysis. Therefore, the `umbrella()` function unrounds this information when the input information to replicate a meta-analysis is the effect size along with the 95% confidence interval. To do so, we used the function `optim()` to find the mean and standard error resulting in a confidence interval that, once rounded, is identical to the one reported in the paper. For instance, imagine the authors find an OR = 3.140 and the standard error of the log(OR) is 0.170, resulting in a 95% CI = 2.250-4.382. If authors rounded these values to one decimal figure, they would report OR = 3.1 with 95% CI = 2.3-4.4. However, the `umbrella()` function unrounds these figures to OR = 3.136344 with 95% CI = 2.248973-4.373843, closer to the true statistics.

References

Albajes-Eizagirre A, Solanes A, Radua J (2018). "Meta-Analysis of Non-Statistically Significant Unreported Effects." *Statistical Methods in Medical Research*, **28**, 3741–3754.

Balduzzi S, Rucker G, Schwarzer G (2019). "How to Perform a Meta-Analysis With R: A Practical Tutorial." *Evidence Based Mental Health*, (22), 153–160.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009). *Overview*. John Wiley & Sons, Ltd. doi:doi.org/10.1002/9780470743386.ch22.

Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Publishers.

Egger M, Smith GD, Schneider M, Minder C (1997). "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *BMJ*, **315**(7109), 629–634. doi:10.1136/bmj.315.7109.629.

Fusar-Poli P, Radua J (2018). "Ten Simple Rules for Conducting Umbrella Reviews." *Evidence based mental health*, **21**, 95–100. doi:10.1136/ebmental-2018-300014.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, *et al.* (2008). "GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations." *International Journal of Epidemiology*, **49**(3), 885–893. doi:10.1093/ije/dyz241.

Hedges L, Olkin I (1985). *Statistical Methods for Meta-Analysis*. CA: Academic Press.

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd. doi:10.1002/9781119536604.fmatter.

Ioannidis JP, Trikalinos TA (2007). "An Exploratory Test for an Excess of Significant Findings." *Clinical Trials*, **4**(3), 245–253. doi:10.1177/1740774507079441.

Jackson D, Riley R, White IR (2011). "Multivariate Meta-Analysis: Potential and Promise." *Statistics in Medicine*, **30**(20), 2481–2498. doi:10.1002/sim.4172.

Qiu W, Chavarro J, Lazarus R, Rosner B, Ma J (2021). *powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies*. R package version 0.1.3, URL https://CRAN.R-project.org/package=powerSurvEpi.

Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, *et al.* (2017). "AMSTAR 2: A Critical Appraisal Tool for Systematic Reviews That Include Randomised or Non-randomised Studies of Healthcare Interventions, or Both." *BMJ*, **358**. doi:10.1136/bmj.j4008. URL https://www.bmj.com/content/358/bmj.j4008.

Sterne JA, Egger M (2005). *Regression Methods to Detect Publication and Other Bias in Meta-Analysis*, chapter 6, pp. 99–110. John Wiley & Sons, Ltd. doi:https://doi.org/10.1002/0470870168.ch6.

Weber F, Knapp G, Ickstadt K, Kundt G, Glass A (2020). "Zero-Cell Corrections in Random-effects Meta-analyses." *Research Synthesis Methods*, **11**(6), 913–919. doi:https://doi.org/10.1002/jrsm.1460.

Affiliation:

Corentin J. Gosling*
DysCo lab, Paris Nanterre University
200 Avenue de la Republique, F-92000 Nanterre, France
and
Universite de Paris, Laboratoire de Psychopathologie et des Processus de Sante, F-92100 Boulogne-Billancourt, France
E-mail: cgosling@parisnanterre.fr

Aleix Solanes*
Institut d'Investigacions Biomediques,
August Pi I Sunyer, Barcelona, Spain
and
Department of Psychiatry and Legal Medicine,
Autonomous University of Barcelona, Barcelona, Spain
E-mail: al.solanes@gmail.com
*Joint first authorship

Paolo Fusar-Poli
Department of Psychosis Studies, Institute of Psychiatry, Psychology, and Neuroscience
King's College London, London, UK
and
University of Pavia, Italy
E-mail: Paolo.fusar-poli@kcl.ac.uk

Joaquim Radua
Institut d'Investigacions Biomediques,
August Pi I Sunyer, Barcelona, Spain
and
Centro de Investigacion Biomedica en Red de Salud Mental (CIBERSAM),
Barcelona, Spain
and
Department of Psychosis Studies, Institute of Psychiatry, Psychology, and Neuroscience
King's College London, London, UK
and
Centre for Psychiatric Research and Education, Department of Clinical Neuroscience,
Karolinska Institutet, Stockholm, Sweden
E-mail: radua@clinic.cat