4 Pooling effect sizes

Author: Chaofan Chen
Date: 2025-05-31 ~ 06-01

- 在 Meta 分析中有很多统计技术与方法可以被我们应用，用哪种技术或方法更好？一般取决于上下文（效应量的类型）。
- 因此，在开始使用R进行分析之前，我们必须对meta分析的统计假设及其背后的数学原理有一个基本的了解。我们还将讨论 Meta 分析背后的“理念”。在统计学中，这个“想法”转化为一个模型，我们将看到 Meta 分析模型是什么样子的：
- As we will see, the nature of the meta-analysis requires us to make a fundamental decision right away: we have to assume：
  1. ~~either a~~ ~~fixed-effect model（FE）~~
  2. or a random-effects model（RE）
- 了解模型基于的理念可以帮助我们选择**更合适的模型**：

1. The Fixed-Effect and Random-Effects Model

通过构建统计模型,我们试图找到数据背后“真实世界”的近似表征。我们需要一个数学公式来解释观察到的结果,找到研究背后的真正效应量。

在进行元分析时，我们的**最终目标之一**是：

找到一个总体效应量（effect size），用来总结多个研究中的结果，形成一个关于某种干预或关系的总体判断。但现实中我们常会发现，不同研究中的效应量是**不完全一样的**，甚至差异较大。产生这些差异的原因可能包括：

- 研究设计不同（如干预方式、时间、强度等）
- 样本特征不同（如年龄、性别、健康状况等）
- 测量工具或分析方法不同
- 抽样误差（sampling error）

因此，Meta 分析模型的一个**关键任务**就是：

> 解释这些研究之间效应量差异的“原因和程度”，即评估研究之间的“异质性”(heterogeneity)。

即便我们认为所有研究实际上是测量同一个True effect size（只有一个总体效应），这些差异依然必须通过合适的模型来处理（FE）。

The Fixed-Effect Model

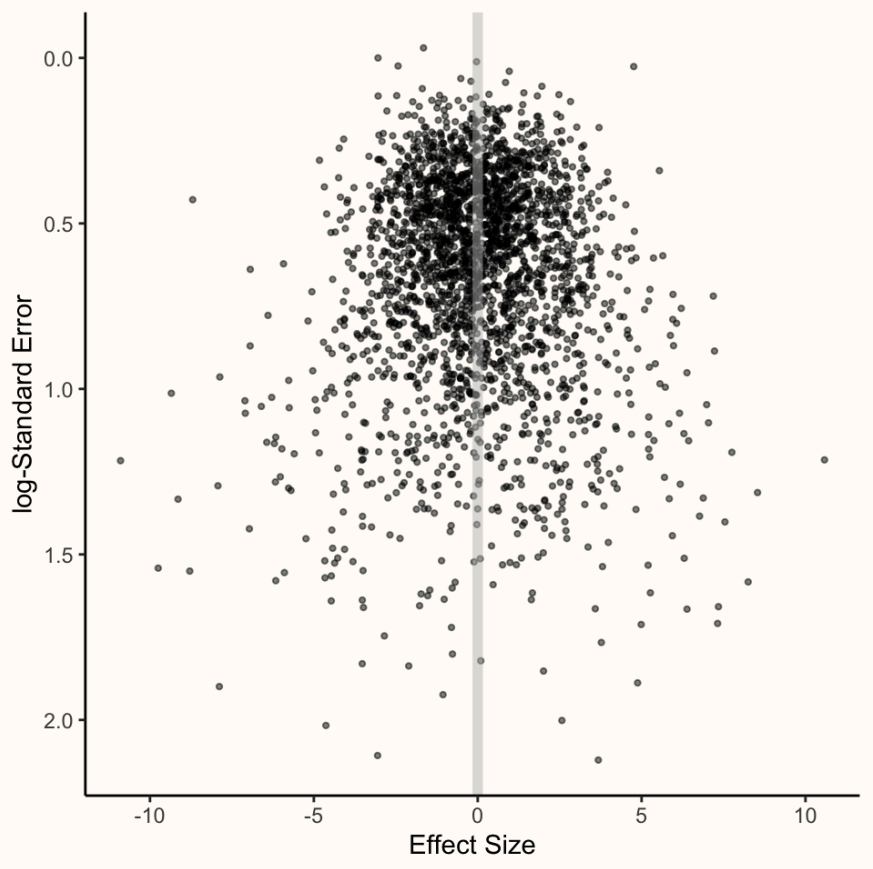
FE假设一个 true effect size 存在，用 θ 表示，一切的差异均来源于 sampling error。

We can describe the relationship like this (Borenstein et al. 2011, chap. 11):

$$\theta_k = \theta + \epsilon_k$$

固定效应模型告诉我们，如果我们找到研究的真实效应值 θ_k，这个效应值不仅和第k个研究的真值一致，还和总体的真值一致，具体来说，一项研究的真实效应量以及总体的综合效应量是一致的（一个 true effect size 存在）。

上述内容被我们弄清楚，我们知道FE假设 observed effect size 和 true effect size 的差异只来源于 sampling error ，前面提到，根据公式，抽样误差越小的研究越接近真实效应量，样本量越大的研究抽样误差越小，抽样误差可以用标准误 stantard error 表示，所以我们模拟一下：



可以看到拥有较小抽样误差（SE）的观测效应量更接近于真实效应量

这种表现可以用固定效应模型的公式来预测：我们知道，标准误较小的研究具有较小的抽样误差，因此它们对总体效应量的估计更有可能接近真实效应量。

所以：当我们在分析中**池化效应**时，我们应该**给予具有更高精度（即较小标准误）的效应量更大的权重**。

因此，如果我们想要计算固定效应模型下的**合并效应量**，我们只需使用所有研究的**加权平均值**。

现在我们要知道**权重如何确定**：

We can use the `standard error` 来计算其方差 s_k^2 ，再取**方差的倒数**作为该研究的权重 W_k ，称为`逆方差加权（Inverse-variance weighting）`：

$$w_k = \frac{1}{s_k^2} = \frac{1}{(SE_k)^2}$$

- 所以方差越小的效应量，倒数越大，合并时的权重越大。

现在我们知道了权重，我们就可以计算加权平均值（对真实池化效应量 θ的估计）：

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \hat{\theta}_k \cdot w_k}{\sum_{k=1}^{K} w_k}$$

$meta$ 包使执行固定效应元分析变得非常容易。然而，在此之前，让我们在R中“手动”尝试逆方差分析：

- 数据集：`SuicidePrevention`
- 包括：未进行预计算的原始效应

所以我们先计算效应：

计算Hedge's g，使用 `esc` 包的 `esc_mean_sd` function：

```
# Load dmetar, esc and tidyverse (for pipe)
library(dmetar)
```

```
library(esc)
library(tidyverse)

# Load data set from dmetar
data(SuicidePrevention)

# Calculate Hedges' g and the Standard Error
# - We save the study names in "study".
# - We use the pmap_dfr function to calculate the effect size
#   for each row.
SP_calc ← pmap_dfr(SuicidePrevention,
                   function(mean.e, sd.e, n.e, mean.c,
                            sd.c, n.c, author, ...){
                     esc_mean_sd(grp1m = mean.e,
                                 grp1sd = sd.e,
                                 grp1n = n.e,
                                 grp2m = mean.c,
                                 grp2sd = sd.c,
                                 grp2n = n.c,
                                 study = author,
                                 es.type = "g") %>%
                       as.data.frame()})

# Let us catch a glimpse of the data
# The data set contains Hedges' g ("es") and standard error ("se")
glimpse(SP_calc)
```

```
## Rows: 9
## Columns: 9
## $ study       <chr> "Berry et al.", "DeVries et ...
## $ es          <dbl> -0.14279447, -0.60770928, -0...
## $ weight      <dbl> 46.09784, 34.77314, 14.97625...
## $ sample.size <dbl> 185, 146, 60, 129, 100, 220,...
## $ se          <dbl> 0.1472854, 0.1695813, 0.2584...
## $ var         <dbl> 0.02169299, 0.02875783, 0.06...
## $ ci.lo       <dbl> -0.4314686, -0.9400826, -0.6...
## $ ci.hi       <dbl> 0.145879624, -0.275335960, 0...
## $ measure     <chr> "g", "g", "g", "g", "g", "g"...
```

现在我们计算了每个研究的独特的 `g` ，应用FE合并一下：

```
# Calculate the inverse variance-weights for each study
SP_calc$w ← 1/SP_calc$se^2

# Then, we use the weights to calculate the pooled effect
pooled_effect ← sum(SP_calc$w*SP_calc$es)/sum(SP_calc$w)
pooled_effect
```

```
## [1] -0.2311121
```

- **别忘记**计算每项研究的权重

我们的计算结果表明，假设固定效应模型，合并效应大小为 $g \approx -0.23$

▼ 4.1.2 The Random-Effects Model

1. 准备好n, mean, sd
2. 使用随机效应模型
3. tau squared 的计算使用REML
4. 使用knha校正，基于t distribution
5. 使用meta包的metacont()
6. 模型更新使用update()

*模型代码在下面路径内

The question is FE确切的反映了真实情况吗?

固定效应模型：由于样本来源是同质的，观测到的效应量间的差异只有一个来源（sampling error），即**一个真实效应量**存在。

但荟萃分析中的研究总是完全同质的是不现实的，各种各样的异质性存在（即使是细微的差别），Meta 分析中纳入的研究往往同时存在多个维度的异质性，这种复合差异通常会导致显著的研究间真实效应量变异。

因此，我们假设单个研究的效应量差异不仅来源于抽样误差，而且还有另一个变异来源（未知的异质性来源）。

RE假设不存在一个 真实性效应量 ，存在的是一个 真实效应量分布 。每个研究之间的真实效应量有差异，同时又由于抽样误差的原因，促成了另一个变异来源。所以RE的目标是估计真实效应量分布的均值。

	固定效应模型（FE）	随机效应模型（RE）
假设	所有研究共享一个真实效应值 θ	每个研究有自己的真实效应 θ_k，来自分布
目标	估计这个**唯一**的真实效应值	估计所有真实效应的**平均值**μ
差异来源	抽样误差（sampling error）	抽样误差 + 异质性 τ^2

让我们看看如何用公式来表示随机效应模型：

1. 首先是和FE一样的假设：

$$\hat{\theta}_k = \theta_k + \epsilon_k$$

$$\theta_k = \mu + \zeta_k$$

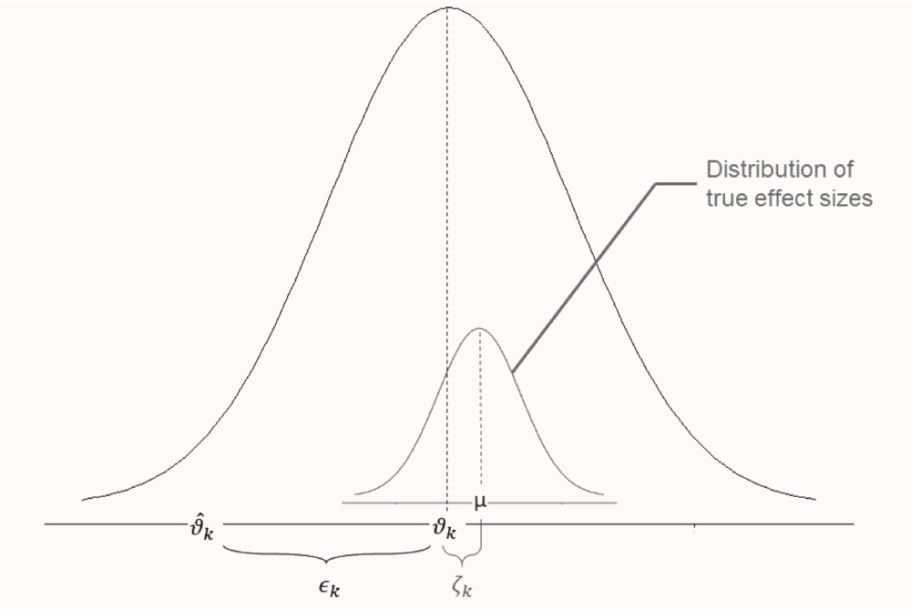
我们期望求得 μ ，但是单个研究的 θ_k 和 μ 之间的差异来源于 ζ_k ，而同时θ_k 本身又受 ϵ_k 的影响，所以：

由于抽样误差，研究中观察到的效应值偏离了它们的真实值。但即使是真正的效应值也只是真实效应的一部分，它们的平均值μ是我们想估计meta分析的综合效应。

我们可以用这个公式来表示随机效应模型（Borenstein et al. 2011，第12章）：

$$\hat{\theta}_k = \mu + \zeta_k + \epsilon_k$$

这个公式清楚地表明，我们观察到的效应量偏离了合并效应量 μ 因为有两个误差项 ζ_k 和 ϵ_k
这种关系如图4.2所示。



ζ_k 是随机波动的，并无先验信息。

epsilonk很好计算，使用标准误进行近似，所以我们需要计算tau Squred

Which Model Should I Use?

4.1.2.1 Estimators of the Between-Study Heterogeneity

In the random-effects model, we therefore calculate an adjusted **random-effects weight** w_k^* for each observation. The formula looks like this: **tau squared is unknown...we need to cauculate it...使用se的平方表示抽样分布的不确定性**

$$w_k^* = \frac{1}{s_k^2 + \tau^2}$$

Using the adjusted random-effects weights, we then calculate the pooled effect size using the inverse variance method, just like we did using the fixed-effect model:

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \hat{\theta}_k w_k^*}{\sum_{k=1}^{K} w_k^*}$$

所以我们现在要计算 τ^2 ，R中的 `meta` 包提供了多种方法：

- The **DerSimonian-Laird** (`"DL"`) estimator (DerSimonian and Laird 1986).
- The **Restricted Maximum Likelihood** (`"REML"`) or **Maximum Likelihood** (`"ML"`) procedures (Viechtbauer 2005).
- The **Paule-Mandel** (`"PM"`) procedure (Paule and Mandel 1982).
- The **Empirical Bayes** (`"EB"`) procedure (Sidik and Jonkman 2019), which is practically identical to the Paule-Mandel method.
- The **Sidik-Jonkman** (`"SJ"`) estimator (Sidik and Jonkman 2005).

> **对于连续性结局推荐使用** `"REML"`

▼ 4.1.2.2 Knapp-Hartung Adjustments

除了对 τ^2 进行估计，我们还必须决定是否要应用所谓的Knapp-Hartung调整 (Knapp and Hartung 2003; Sidik and Jonkman 2002)

- 这个方法影响 `tau^2` 和 `标准误` 的计算方法（进而影响置信区间的计算）而不是影响点估计，一般来说，经过 `"knha"法` 矫正后的置信区间会**变得更宽**。
- 不同于一般假设为正态分布的合并方法，Knha **基于 t 分布**

在方法学中报告你所使用的模型：
强烈建议在荟萃分析报告的方法部分指定您使用的模型类型。下面是一个例子：

> "As we anticipated considerable between-study heterogeneity, a random-effects model was used to pool effect sizes. The **restricted maximum likelihood estimator** (Viechtbauer, 2005) was used to calculate the heterogeneity variance τ2. We used Knapp-Hartung adjustments (Knapp & Hartung, 2003) to calculate the confidence interval around the pooled effect."

- 应用Knapp-Hartung调整通常是明智的。几项研究(IntHout、Ioannidis和Borm 2014；Langan et al. 2019)表明，这可以**减少假阳性**的机会，特别是**在研究数量较少**的情况下。

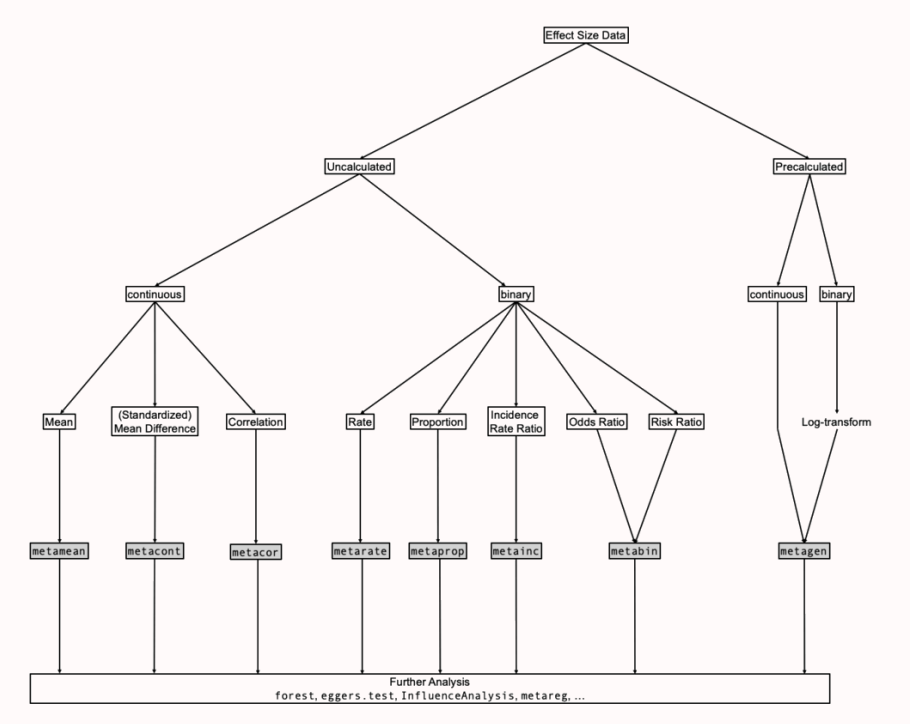
▼ Effect Size Pooling in *R*

D:\Academic research\Meta-analysis\Doing meta analysis with R\Doing meta analysis with R-代码

Time to put what we learned into practice.

使用：R - $meta$

Figure 4.3 provides an overview of **{meta}**'s **structure**.



所以，使用**Raw data**，需要计算**ES**时请使用 `metacont` 函数

There are six core arguments which can be specified in each function:

- `studlab`. This argument associates each effect size with a **study label**. If we have the name or authors of our studies stored in our data set, we simply have to specify the name of the respective column (e.g. `studlab = author`).
- `sm`. This argument controls the **summary measure**, the effect size metric we want to use in our meta-analysis. This option is particularly important for functions using raw effect size data. The **{meta}** package uses codes for

different effect size formats, for example `"SMD"` or `"OR"`. The available summary measures are not the same in each function, and we will discuss the most common options in each case in the following sections.

- `fixed`. We need to provide this argument with a logical (`TRUE` or `FALSE`), indicating if a fixed-effect model meta-analysis should be calculated22.
- `random`. In a similar fashion, this argument controls if a random-effects model should be used. If both `comb.fixed` and `comb.random` are set to `TRUE`, both models are calculated and displayed23.
- `method.tau`. This argument defines the `tau^2` estimator. All functions use the codes for different estimators that we already presented in the previous chapter (e.g. for the DerSimonian-Laird method: `method.tau = "DL"`). REML是迭代法实现的。
- `method.random.ci`. This argument controls how confidence intervals should be calculated when using the random-effects model. When setting `method.random.ci = "HK"`, the Knapp-Hartung adjustment will be applied. In older versions of **{meta}**, this feature was controlled by setting the `hakn` argument to `TRUE` instead. Using the Knapp-Hartung adjustment is advisable in most cases (see Chapter 4.1.2.2).
- `data`. In this argument, we provide **{meta}** with the name of our meta-analysis data set.
- `title` (**not mandatory**). This argument takes a character string with the name of the analysis. While it is not essential to provide input for this argument, it can help us to identify the analysis later on.

▼ 4.2.2 (Standardized) Mean Differences

- Raw effect size data in the form of means and standard deviations of two groups can be pooled using `metacont` function

there are seven function-specific arguments we have to provide:

- `n.e`. The number of observations in the treatment/experimental group.
- `mean.e`. The mean in the treatment/experimental group.
- `sd.e`. The standard deviation in the treatment/experimental group.
- `n.c`. The number of observations in the control group.
- `mean.c`. The mean in the control group.
- `sd.c`. The standard deviation in the control group.
- `method.smd`. This is only relevant when `sm = "SMD"`. `"Hedges"` (default and recommended)

现在使用 `SuicidePrevention` dataset with using RE and Knha method and “REML” method.

```
# Make sure meta and dmetar are already loaded
library(meta)
library(dmetar)
library(meta)

# Load dataset from dmetar (or download and open manually)
data(SuicidePrevention)

# Use metcont to pool results.
m.cont ← metacont(n.e = n.e,
                  mean.e = mean.e,
                  sd.e = sd.e,
                  n.c = n.c,
                  mean.c = mean.c,
                  sd.c = sd.c,
                  studlab = author,
                  data = SuicidePrevention,
                  sm = "SMD",
                  method.smd = "Hedges",
                  fixed = FALSE,
                  random = TRUE,
                  method.tau = "REML",
```

```
            method.random.ci = "HK",
            title = "Suicide Prevention")
```

值得注意的是 meta 包和 esc 包的计算结果保持一致。

- 为了更好的理解计算过程，我们来手动计算 **Berry et al.** 研究的 **Hedges' g**：

组别	样本量 n	均值 $\bar{X}$	标准差 S
实验组 (e)	$n_e = 90$	$\bar{X}_e = 14.98$	$s_e = 3.29$
对照组 (c)	$n_c = 95$	$\bar{X}_c = 15.54$	$s_c = 4.41$

标准化均值差的计算（未校正）：

$$SMD_{\text{between}} = \frac{\bar{x}_1 - \bar{x}_2}{S_{\text{pooled}}}$$

求未知项 S_{pooled} ，公式如下：

$$s_{\text{pooled}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$$

代入计算：

$$s_{\text{pooled}} = \sqrt{\frac{(90-1)(3.29)^2 + (95-1)(4.41)^2}{90+95-2}} = \sqrt{\frac{89 \times 10.8241 + 94 \times 19.4481}{183}} = \sqrt{\frac{963.345 + 1828.11}{183}}$$

求Cohen'd:

$$d = \frac{\bar{X}_e - \bar{X}c}{s\text{pooled}} = \frac{14.98 - 15.54}{3.9053} = \frac{-0.56}{3.9053} \approx -0.1434$$

做一个Hedge's g校正：

修正因子：

$$J = 1 - \frac{3}{4(n_e+n_c)-9} = 1 - \frac{3}{4(185)-9} = 1 - \frac{3}{731} \approx 0.9959$$

进行g校正：

$$g = J \cdot d = 0.9959 \times (-0.1434) \approx \boxed{-0.1428}$$

```
> summary(m.cont)
Review:     Suicide Prevention

                  SMD            95%-CI %W(random)
Berry et al.   -0.1428 [-0.4315;  0.1459]      15.6
```

可以看到，效应量的计算结果（手算和代码）保持一致。

*95%CI的计算为：
CI.lower ← TE - 1.96 * seTE
CI.upper ← TE + 1.96 * seTE

现在看看输出的结果：

Let us see what the results are:

```
summary(m.cont)

## Review:     Suicide Prevention
##
##                    SMD            95%-CI %W(random)
## Berry et al.    -0.1428 [-0.4315;  0.1459]      15.6
## DeVries et al.  -0.6077 [-0.9402; -0.2752]      12.3
## Fleming et al.  -0.1112 [-0.6177;  0.3953]       5.7
## Hunt & Burke    -0.1270 [-0.4725;  0.2185]      11.5
## McCarthy et al. -0.3925 [-0.7884;  0.0034]       9.0
## Meijer et al.   -0.2676 [-0.5331; -0.0021]      17.9
## Rivera et al.    0.0124 [-0.3454;  0.3703]      10.8
## Watkins et al.  -0.2448 [-0.6848;  0.1952]       7.4
## Zaytsev et al.  -0.1265 [-0.5062;  0.2533]       9.7
##
## Number of studies: k = 9
## Number of observations: o = 1147 (o.e = 571, o.c = 576)
##
##                         SMD            95%-CI     t p-value
## Random effects model -0.2304 [-0.3734; -0.0874] -3.71  0.0059
##
## Quantifying heterogeneity:
##  tau^2 = 0.0044 [0.0000; 0.0924]; tau = 0.0661 [0.0000; 0.3040]
##  I^2 = 7.4% [0.0%; 67.4%]; H = 1.04 [1.00; 1.75]
##
## Test of heterogeneity:
##     Q d.f. p-value
##  8.64    8  0.3738
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-Profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model (df = 8)
## - Hedges' g (bias corrected standardised mean difference; using exact formulae)
```

▼ Summary

- 注意区分固定效应模型和随机效应模型，两者基于不同的理论背景进行效应量合并
- 固定效应模型假设存在一个 `真实效应量`，而随机效应模型基于的是存在 `真实效应量的分布`。
- 真实效应量的方差 `tau^2` （也被称为研究间异质性方差），在随机效应模型分析中必须被估计。有几种方法可以做到这一点，哪一种效果最好取决于上下文（不同效应量），对于continuous outcome，推荐使用“最大似然估计（REML）”
- `meta` 包提供了强大的分析函数（例如metagen，metacont等）