

应用 R 软件 metaplust 程序包实现 Meta 分析



郭梓鑫¹, 冯雨嘉¹, 王清华¹, 曾宪涛², 刘小平²

1. 武汉大学中南医院/武汉大学第二临床学院(武汉 430071)

2. 武汉大学中南医院循证与转化医学中心(武汉 430071)

【摘要】 R 软件是一款具备强大的数据处理和图形绘制功能的免费软件,其在 Meta 分析中的应用日趋广泛。R 软件中有 meta、metafor、metaplust 等程序包可实现 Meta 分析。本文以实例对 R 软件 metaplust 程序包实现 Meta 分析的过程和结果进行展示。

【关键词】 metaplust 程序包; Meta 分析; R 软件; 离群值检测

How to implement meta-analysis using R package metaplust

GUO Zixin¹, FENG Yujia¹, WANG Qinghua¹, ZENG Xiantao², LIU Xiaoping²

1. Zhongnan Hospital of Wuhan University, the Second Clinical Medical College of Wuhan University, Wuhan, 430071, P.R.China

2. Center for Evidence-Based and Transitional Medicine, Zhongnan Hospital of Wuhan University, Wuhan, 430071, P.R.China

Corresponding author: ZENG Xiantao, Email: zengxiantao1128@163.com; LIU Xiaoping, Email: liuxiaoping2015@hotmail.com

【Abstract】 R software is a free, powerful statistical and graphing software, including metafor, meta as well as metaplust packages. They can be used to conduct meta-analysis. This article introduces detailed operations of the metaplust package for meta-analysis using cases.

【Key words】 Metaplust package; Meta-analysis; R software; Outlier detection

随着循证医学的发展,Meta 分析理论与方法学也随着需求而不断发展。目前 Meta 分析被公认为客观评价和合成针对某一特定问题研究证据的最佳手段,是循证决策的重要依据^[1-3]。R 软件是一款功能强大、免费开放的统计分析及绘图软件^[4],metaplust 程序包是 R 专用于实现 robust meta 分析和 Meta 回归的程序包^[5]。metaplust 程序包由 Beath 等^[6]研发,版本为 Version 0.7~9,该程序包是通过标准 Meta 分析法和基于轮廓似然置信区间的稳健法来实现的。本文以 Lawlor 等^[7]《运动干预在抑郁症治疗中的有效性:基于随机对照试验的系统评价与 Meta 回归分析》中的数据为例,介绍 metaplust 程序包的使用方法。

1 R 软件及 metaplust 程序包的安装及加载

R 软件的简介及安装详见前文^[8]。本文以 R 3.4.1 版为例,待安装完毕后,双击桌面的 R 软件图

标,即可启动 R 的交互式窗口(R-GUI)。于命令提示符“>”后输入命令 install.packages("metaplust"),在弹出的对话框中选择某个镜像安装(CRAN),安装完成后可由 library("metaplust")命令完成加载。至此,R 软件及 metaplust 程序包就已安装及加载完毕。

2 数据加载

在进行数据加载之前须对数据进行处理,整理后的数据排列格式见表 1。该研究数据为运动与不运动比较、运动与认知治疗比较的贝克抑郁量表评分的效应量和加权平均差的标准化均值差,并采用 Meta 分析和 Meta 回归方法来验证运动干预在抑郁症治疗中的有效性。

在 Excel 软件中整理数据并以 csv 格式保存。导入数据语句为 mydata<-read.csv(file.choose()),输入到 R 软件后按回车键即可选择文件夹位置,导入要分析的 csv 数据。数据导入后可使用命令 edit(mydata)对数据进行编辑和修改,还可使用命令 summary(mydata)输出基本的数据描述性信息。

DOI: 10.7507/1672-2531.201708086

通信作者:曾宪涛,Email: zengxiantao1128@163.com;刘小平,Email: liuxiaoping2015@hotmail.com

表 1 整理后的数据排列格式

study	smd	varsmd	sesmd	duration	duration4	duration8	duration12
Mutrie	-2.53	0.160 0	0.40	4	0	-4	-8
McNeil	-1.07	0.168 1	0.41	6	2	-2	-6
Reuter	-2.1	0.160 0	0.40	10	6	2	-2
Doyne	-1.2	0.184 9	0.43	8	4	0	-4
Hess-Homeier	-0.82	0.324 9	0.57	8	4	0	-4
Epstein	-0.84	0.211 6	0.46	8	4	0	-4
Martinsen	-1.16	0.078 4	0.28	9	5	1	-3
Singh	-0.45	0.115 6	0.34	10	6	2	-2
Klein	0.25	0.260 1	0.51	12	8	4	0
Veale	-0.53	0.057 6	0.24	12	8	4	0

study: 纳入研究的作者名; smd (study effect estimate): 研究效应量; varsmd (study effect variance): 研究效应量的方差; sesmd (study effect standard error): 研究效应量的标准误; duration: 干预时间, 单位为周; duration4: 干预时间与 4 周差值; duration8: 干预时间与 8 周差值; duration12: 干预时间与 12 周差值。

3 数据分析

metaplust 程序包总体上可以分为 Meta 分析和 Meta 回归两大功能。其主要执行函数命令有 metaplust 函数、testOutliers 函数和 outlierProbs 函数。其中, metaplust 函数用于模型拟合、选择合适模型及绘制森林图等; testOutliers 及 outlierProbs 函数用于检测数据中的离群值。下面以实例说明数据分析的过程。

3.1 模型的拟合与选择

命令: `mydata<-mydata[order (mydata $ duration),]`

3.1.1 标准正态随机效应模型 命令: `mydata.meta<-metaplust (smd, sesmd, mods=duration, slab=study, random="normal", data=mydata)`

`summary (mydata.meta)`

命令中 smd 为观察的效应量, sesmd 为效应量的标准误, mods 为各研究相对应的协变量数据框架, slab 为与每个研究相对应的字符串向量 (Vector of character strings corresponding to each study), random 为随机效应的类型 (normal 表示正态分布, t-dist 表示 t 分布, mixture 表示混合分布), data 为数据。上述数据的标准正态随机效应模型 Meta 分析结果见表 2 和表 3。

3.1.2 t 分布随机效应模型 metaplust 软件包进行 t 分布随机效应模型 Meta 分析与标准正态随机效应模型类似, 只需要将参数 random 赋值为 "t-dist" 即可, 具体命令如下:

`mydata.meta<-metaplust (smd, sesmd, mods=duration, slab=study, random="t-dist", data=mydata)`

`summary (mydata.meta)`

表 2 标准正态随机效应模型的 metaplust 命令结果 (合并效应量)

条目	Est.	95%ci.lb	95%ci.ub	P value
muhat	-2.899 4	-4.300 6	-1.522 2	0.000 884
tau2	0.117 1	-	-	-
duration	0.207 8	0.058 4	0.363 2	0.011 570

Est.: 合并后的效应量; 95%ci.lb: 95% 可信区间下限; 95%ci.ub: 95% 可信区间上限; P-value: P 值; tau2: 异质性检验的效应值。

表 3 标准正态随机效应模型的 metaplust 命令结果 (拟合效应量)

logLik	AIC	BIC
-8.133 435	22.266 87	23.174 62

logLik: 对数似然函数值; AIC: 赤池信息量准则; BIC: 贝叶斯信息规则。

上述数据的 t 分布随机效应模型 Meta 分析结果见表 4 和表 5。

3.1.3 混合随机效应模型 metaplust 进行混合随机效应模型合并只需要参数 random 赋值为 "mixture" 即可, 具体命令如下:

`mydata.meta<-metaplust (smd, sesmd, mods=duration, slab=study, random="mixture", data=mydata)`

`summary (mydata.meta)`

上述数据的混合随机效应模型 Meta 分析结果见表 6 和表 7。

3.2 离群值检验

对于 t 分布随机效应模型, 离群值的检验需要自由度趋近为无穷大; 对于混合随机效应模型, 离群值的检验需要离群研究概率密度函数的权重趋近为零。由于这两个检验都涉及参数空间边界上的参数, 渐近理论不适用, 所以在零假设下运用参数自举的方法来检验统计量的经验分布。然后将

表 4 *t* 分布随机效应模型的 metapuls 命令结果 (合并效应量)

条目	Est.	95%ci.lb	95%ci.ub	P-value
muhat	-2.906 87	-4.244 82	-1.574 51	0.000 806
tau2	0.038 73			
vinv	0.426 05			
duration	0.211 76	0.064 37	0.357 42	0.010 068

Est.: 合并后的效应量; 95%ci.lb: 95% 可信区间下限; 95%ci.ub: 95% 可信区间上限; P-value: P 值; tau2: 异质性检验的效应值; vinv: *t* 分布自由度逆概率分布函数值, 当 vinv 为 0 时, 相当于自由度为无穷大, 则模型倾向于标准正态随机效应模型。

表 5 *t* 分布随机效应模型的 metapuls 命令结果 (拟合效应量)

logLik	AIC	BIC
-8.009 885	24.019 77	25.230 11

logLik: 对数似然函数值; AIC: 赤池信息量准则; BIC: 贝叶斯信息规则。

表 6 混合随机效应模型的 metapuls 命令结果 (合并效应量)

条目	Est.	95%ci.lb	95%ci.ub	P-value
muhat	-2.884 72	-4.110 82	-1.482 62	0.000 649
tau2	0.000 00			
tau2out	0.593 98			
Outlier prob.	0.251 69			
duration	0.210 86	0.078 07	0.345 86	0.007 052

Est.: 合并后的效应量; 95%ci.lb: 95% 可信区间下限; 95%ci.ub: 95% 可信区间上限; P-value: P 值; tau2: 标准研究的随机效应方差; tau2out: 离群研究的随机效应方差, 当 tau2 和 tau2out 值接近时, 表示该研究没有离群值; Outlier prob: 离群概率。

表 7 混合随机效应模型的 metapuls 命令结果 (拟合效应量)

logLik	AIC	BIC
-7.691 39	25.382 78	26.895 7

logLik: 对数似然函数值; AIC: 赤池信息量准则; BIC: 贝叶斯信息规则。

表 8 离群值检验命令结果

似然比统计量的观测值	P value
0.9	0.082

观察到的统计量似然比与此分布进行比较, 以确定 *P* 值。由于参数自举中使用大量随机生成的数据, *P* 值是不同的^[9]。metapuls 提供了 testOutliers 函数进行离群值检验, 命令如下:

```
mydata.testOutliers<- testOutliers (mydata.mix)
summary (mydata.testOutliers)
离群值检验的结果见表 8。
```

此外, metapuls 还提供 outliersProbs 用于计算离群概率, 对于混合随机效应模型和 *t* 分布随机效应模型可以计算出每个研究的离群概率, 并使用 plot 命令绘制离群概率图像。命令如下:

```
mydata.outlierProbs<- outlierProbs (mydata.mix)
```

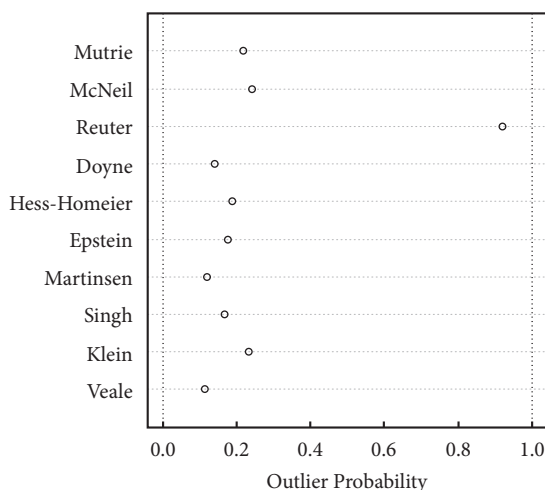


图 1 离群概率图

plot (mydata.outlierProbs)

离群概率图结果见图 1。

3.3 模型选择

对于相关模型的拟合与选择, 总体上依据 AIC (赤池信息量准则) 和 BIC (贝叶斯信息规则) 的数值。AIC 和 BIC 都是对模型的拟合效果进行评价的指标, 若 AIC 和 BIC 值越小, 则模型对数据的拟合越好。当 *t* 分布随机效应模型和混合随机效应模型比标准正态随机效应模型的 AIC 和 BIC 的值低, 那么倾向于选择 *t* 分布随机效应模型和混合随机效应模型。结合 testOutliers 和 outlierProbs 函数, 可以对研究数据中的离群值进行检测, 若数据中存在明显的离群值, 则倾向于选择混合随机效应模型^[9]。因本例中数据存在明显的离群值, 故选择应用混合随机效应模型。

3.4 输出 Meta 分析结果

经过前期的模型拟合与选择, 并完成离群值检验后, metapuls 程序包可以调用 metafor 程序包的绘制森林图功能, 完成森林图的绘制。命令如下:

```
plot (exercise.mix, extrameta=NULL)
森林图结果见图 2。
```

3.5 Meta 回归

目前, metapuls 程序包尚未有预测方法, 计算 4、8 和 12 周效应的替代方法是在这些时间将数据置于中心, 并对每个数据进行 Meta 回归^[10]。命令如下:

```
mydata $ duration4<-mydata $ duration-4
mydata $ duration8<-mydata $ duration-8
mydata $ duration12<-mydata $ duration-12
mydata.nodurn<-metapuls (smd, sesmd, label =
"Random Mixture (No Duration)", slab=study,
```

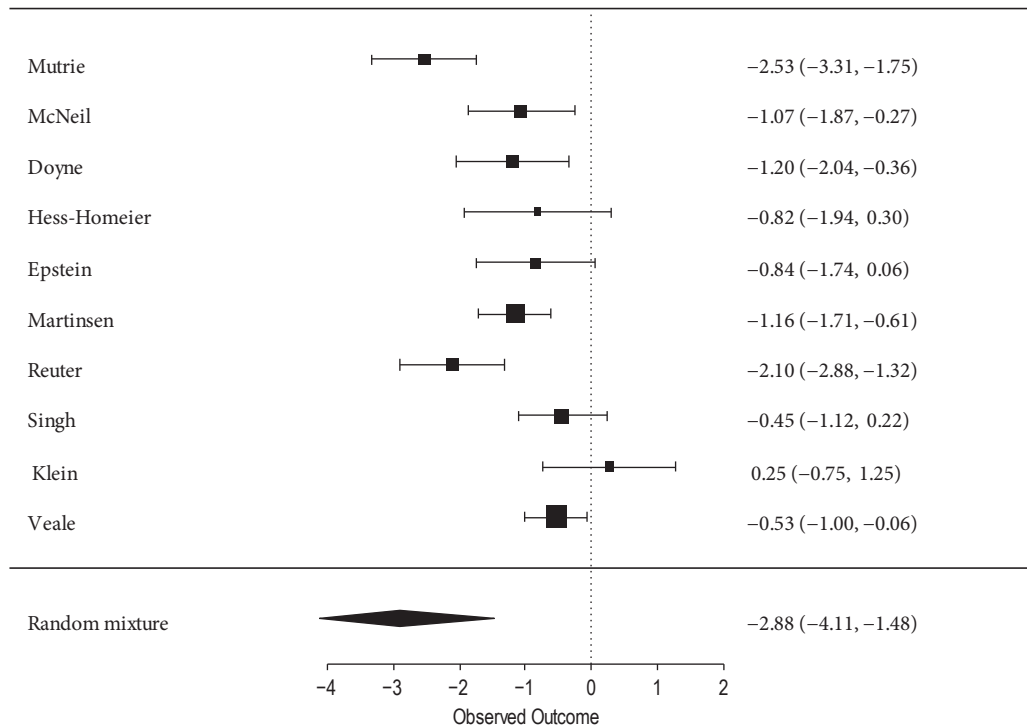


图2 Meta分析森林图结果

random="mixture", data = mydata)

```
mydata.wk4 <- metaplust (smd, sesmd,
  mods=duration4, label = "Random Mixture (Week
  4)", slab = study, random = "mixture", data=mydata)
```

```
mydata.wk8 <- metaplust (smd, sesmd,
  mods=duration8, label="Random Mixture (Week 8)",
  slab = study, random = "mixture", data = mydata)
```

```
mydata.wk12 <- metaplust (smd, sesmd,
  mods=duration8, label="Random Mixture (Week
  12)", slab=study, random="mixture", data=mydata)
```

```
plot (mydata.nodurn, extrameta=list
  (mydata.wk4, mydata.wk8, mydata.wk12),
  xlab="Standardised mean difference")
```

Meta回归森林图见图3, Meta回归的截距可以估计对应的平均效应。图3结果表明, 研究效应在持续时间延长时会下降, 可能与安慰剂效应减或消失有关。

4 结果汇总

在完成上述运算命令后, 即可对结果进行汇总和图形的绘制。执行表9中的命令后, 即可展示以上图表结果。

5 模型的比较

在Meta分析中, 最常见的两种模型是固定效

应模型和随机效应模型。固定效应模型假设所有纳入的研究拥有共同的真实效应量, 而随机效应模型中的真实效应量随研究的不同而改变^[11]。基于不同模型的运算, 所得到的合并后的效应量的均数值也不相同^[12]。尽管不一定适用于实践, 但为了数据处理的方便, 传统上假定随机效应模型的真实效应量呈正态分布^[13]。而t分布可以更灵活地来替代这个假设, 通过马尔科夫链蒙特卡罗方法实现正态分布和t分布的偏态扩张^[14, 15], 比常规正态分布更加灵活。混合随机效应模型可以通过比较具有和不具有离群值的模型来确定是否存在异常值, 并且使用后验概率来确定, 其优点在于Meta分析时考虑了含离群值的研究, 且具有适当的权重^[16]。

6 结语

随着循证医学的发展, 为了适应不同类型Meta分析需求, 多种Meta分析软件被研发^[17]。目前, 国内有多篇介绍Meta分析相关软件^[18]的文章, 包括STATA软件^[19]、RevMan软件^[20]、SAS软件^[21]、R软件^[8]等。R软件作为一款免费的统计分析及绘图软件, 受到广大的欢迎。R软件中有meta、metafor等程序包可以实现Meta分析。本文以实例对R软件metaplust程序包实现Meta分析的过程和结果进行了展示。

metaplust程序包是R软件中实现Meta分析和

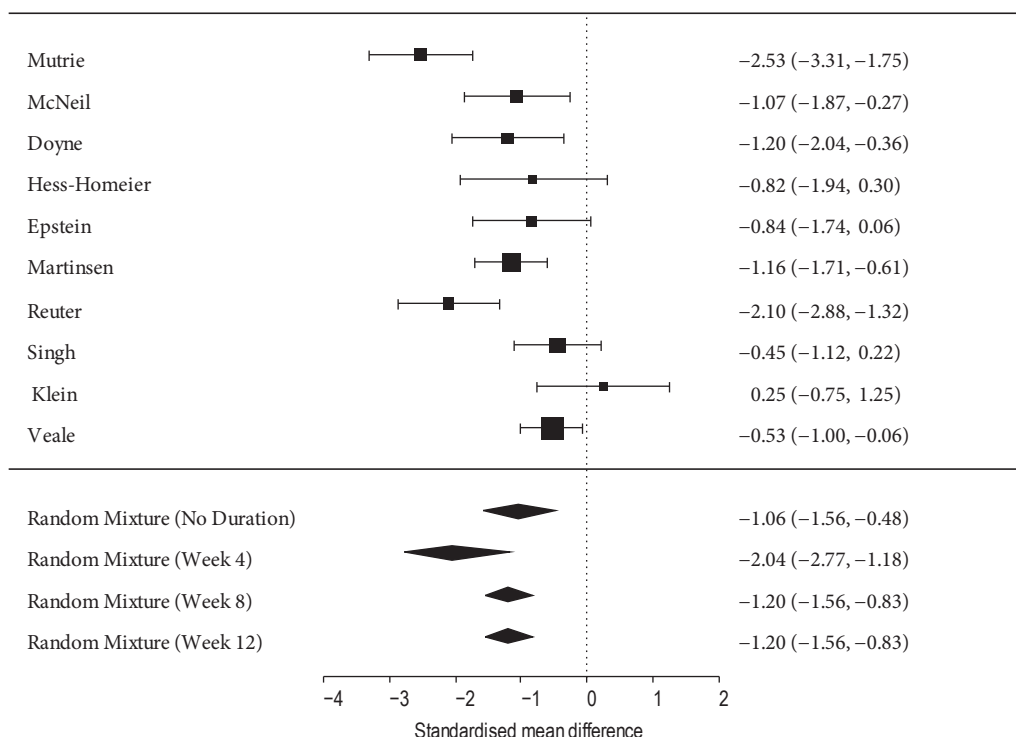


图3 持续时间与研究效应量关系的 Meta 回归森林图

表9 执行结果汇总的命令

数据分析函数	命令	结果
metaplus	summary (mydata.meta)	表 2
testOutliers	summary (mydata.testOutliers)	表 8
outlierProbs	plot (mydata.outlierProbs)	图 1
plot.metaplus	plot.metaplus	图 2

Meta 回归的程序包。metaplus 运用的是标准 Meta 分析法和基于轮廓似然置信区间的 robust 方法。metaplus 程序包的目的是为了取代一般的 Meta 分析程序包, 如 metafor 程序包^[22], 而是为了提供额外的专业分析, 如离群值的检测和离群值的概率计算等。metaplus 程序包可以绘制离群概率图, 并调用 Metafor 程序包绘制森林图等结果图形。然而, 值得注意的是, metaplus 程序包的不足在于进行大量数据分析计算时会延长计算时间, 特别是进行 testOutliers() 命令时。另一个不足之处是 metaplus 程序包只适用于标准正态随机效应、t 分布随机效应和混合正态随机效应 3 种模型, 没有扩展到效应量的其他分布类型^[9]。

总之, metaplus 程序包可以在实现基本 Meta 分析功能的基础上进行离群值的检测和概率计算。虽然 metaplus 程序包有些许不足, 但是 R 软件的开放特性也会使该程序包不断地完善。

参考文献

- 1 曾宪涛, 冷卫东, 李胜, 等. 如何正确理解及使用GRADE系统. 中国循证医学杂志, 2011, 11(9): 985-990.
- 2 Sutton AJ, Abrams KR, Jones DR, et al. Methods for meta-analysis in medical research. 1st edition. John Wiley & Sons, Ltd.: London. 2000: 1-348
- 3 曾宪涛, 田国祥, 张超, 等. Meta分析系列之十五: Meta分析的进展与思考. 中国循证心血管医学杂志, 2013, 5(6): 561-563.
- 4 R Core Team (2015). R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available at: <https://www.R-project.org/>.
- 5 Beath KJ, Bolker B. R Core Team. Metaplus: robust meta-analysis and meta-regression, 2017. Available at: <https://CRAN.R-project.org/package=metaplus.Rpackageversion0.7-9>.
- 6 Beath KJ. The reference manual of metaplus(2016). Available at: <https://cran.r-project.org/web/packages/metaplus/metaplus.pdf>.
- 7 Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. BMJ, 2001, 322(7289): 763-767.
- 8 董圣杰, 曾宪涛, 郭毅. R软件Metafor程序包在Meta分析中的应用. 中国循证医学杂志, 2012, 12(9): 1141-1147.
- 9 Beath KJ. Metaplus: an R package for the analysis of robust meta-analysis and meta-regression. R Journal, 2016, 8(1): 5-16.
- 10 Johnson BT, Huedo-Medina TB. Depicting estimates using the intercept in meta-regression models: the moving constant technique. Res Synth Method, 2011, 2(3): 204-220.
- 11 程里礼, 雷鹏, 陶园, 等. 基于统计学角度: 解读固定效应模型和随机效应模型. 中国循证心血管医学杂志, 2017, 9(3): 261-264.
- 12 Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med, 2001, 20(6): 825-840.

- 13 Goldstein H. Multilevel statistical models. Arnold: London, 2003.
- 14 Baker R, Jackson D. A new approach to outliers in meta-analysis. *Health Care Manag Sci*, 2008, 11(2): 121.
- 15 Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. *Br J Math Stat Psychol*, 2008, 27(3): 418.
- 16 Beath KJ. A finite mixture method for outlier detection and robustness in meta-analysis. *Res Synth Method*, 2014, 5(4): 285.
- 17 曾宪涛, 冷卫东, 郭毅, 等. Meta分析系列之一: Meta分析的类型. *中国循证心血管医学杂志*, 2012, 4(1): 3-5.
- 18 曾宪涛, Kwong Joey SW, 田国祥, 等. Meta分析系列之二: Meta分析的软件. *中国循证心血管医学杂志*, 2012, 4(2): 89-91.
- 19 田金徽, 李伦, 杨克虎. 频率统计方法网状Meta分析在STATA软件中的实现. *中国循证儿科杂志*, 2014, 9(6): 472-474.
- 20 徐世侠, 汤先华, 陈海青. Meta分析及RevMan软件介绍. *中华医学图书情报杂志*, 2009, 18(3): 61-63.
- 21 鄢金柱, 李胜, 翁鸿, 等. 应用SAS软件宏命令实现二分类数据的Meta分析. *中国循证心血管医学杂志*, 2014, 6(2): 121-124.
- 22 Viechtbauer W. Conducting meta-analyses in r with the metafor package. *J of Stat Soft*, 2010, 36(3): 1-48.

收稿日期: 2017-08-22 修回日期: 2018-01-24

本文编辑: 樊斯斯