Vignette raremeta

Karl Westphal

2022-11-29

Vignette **raremeta**

Abstract

This article gives an introduction to the R-software package **raremeta**. We aim to motivate and explain the use of the package in the context of conducting meta-analysis of binary data of rare events. Special interest is put into variants of a common way to incorporate studies with no events: *Continuity Corrections*.

Introduction

In the following vignette, we aim to give the reader a starting point to work with the **raremeta** package for R. For a quick dive into the workflow, we begin with some examples interlaced with the corresponding R-code. For anybody familiar with the mathematical modeling and methodological techniques used when conducting meta analysis and their implementation in R this might be sufficient. For anybody new to either the statistical concepts or similar programs, there is a more in-depth description following the examples. To do so, the reader will be given an overview of the main workflow for model fitting using **raremeta**. After this, there will be a dedicated section introducing continuity corrections on a theoretical and practical level. Some more information will be in an appendix.

raremeta in Action

We begin by setting up **raremeta** and examine the dataset `dat.nissen2007`

```
install.packages("raremeta")
```

```
library(raremeta)
dat <- dat.nissen2007
head(dat,3)
```

```
##       study nRosiglitazone miRosiglitazone cvRosiglitazone nControl miControl
## 1 49653/011            357               2               1      176         0
## 2 49653/020            391               2               0      207         1
## 3 49653/024            774               1               0      185         1
##   cvControl methodControl
## 1         0       Placebo
## 2         0     Glyburide
## 3         0       Placebo
```

We see that the dataset `dat` contains event counts for two different events, `cv` (cardiovascular death) and `mi` (myocardial infarction) under the treatment `Rosiglitazone`. Suppose we are interested in analyzing the effect of the treatment on `cv` - cardiovascular death. We use the function `rareDescribe` to prepare the dataset.

```
dat <- rareDescribe(ai=cvRosiglitazone,ci=cvControl,n1i=nRosiglitazone,
                    n2i=nControl,data=dat)
```

Now, `dat` is an object of type `rareData`, which we can use as an argument for the `rareIV` function to model fit. Non specified arguments have a default, which can also be checked in the next section. Let us go through some examples of doing so. We begin by fitting a random effects model specifying the effect size to be the logarithm of the odds ratio, the heterogeneity estimator `"DL"` - *DerSimonian-Laird*, and the continuity correction to add 0.5 to all cells of all studies if there is a single-zero study and to none of the cells if there is no single-zero study (double-zero studies got droped by default).

```
Fit1 <- rareIV(x=dat,measure="logOR",method="DL",cc="constant",ccto="if0all")
summary(Fit1)
```

```
## Random-effects meta-analysis using the inverse variance model (tau^2 estimator: DL):
##
## Number of studies: 23
## Continuity correction: constant, applied to 23 studies
## Double-zero studies were excluded from the analysis.
##
## Heterogeneity:
##
##  Q(df = 22) = 4.7302, p-val 1
##
##  tau^2 (estimated amount of total heterogeneity): 0
##  tau (square root of estimated tau^2 value):      0
##  I^2 (total heterogeneity / total variability):   0  %
##
## Model results:
##
##   logOR     se   zval  pval   ci.lb  ci.ub
##  0.2471 0.2367 1.0441 0.296 -0.2168 0.7111
##
## ---
## Signif. codes:  '***': < .001 '**': < .01 '*': < .05 '.': < .1
```

Next, we fit a fixed effects model specifying the effect size to be the risk difference and the contiuity correction to add 0.1 to all studies, without dropping double-zero studies.

```
Fit2 <- rareIV(x=dat,measure="RD",method="FE",cc="constant",
               ccval=0.1 ,ccto="all",drop00=FALSE)
```

For a exhaustive account of the possible inputs, continue reading or look into the corresponding documentation.

Calculation: An Overview of the raremeta-Workflow

Let us go through the process of calculating effect sizes using the software package **raremeta** in `R`. We will introduce a common workflow: Preparing event-count-data using `rareDescribe` to calculate estimations of effect sizes using `rareIV`.

```
install.packages("raremeta")
```

```
library(raremeta)
```

Suppose `dat` is a dataframe containing k units of event-counts, i.e. for the i'th study

i'th study	event	no-event	
treatment	a_i	b_i	n_{1i}
control	c_i	d_i	n_{2i}

in form of a table, i.e. k rows (one for each study) and columns reporting the event-counts of the individual studies.

dat	ai	bi	ci	di	n1i	n2i
Study 1	a_1	b_1	c_1	d_1	n_{11}	n_{21}
Study 2	a_2	b_2	c_2	d_2	n_{12}	n_{22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Study k	a_k	b_k	c_k	d_k	n_{1k}	n_{2k}

Let us look at `rareDescribe` and its parameters:

```
rareDescribe(ai,bi,ci,di,n1i,n2i,data)
```

The meaning of its parameters becomes clear from the table above. The columns of `dat` are named in the same fashion as the parameters, i.e. `dat$ai` is a vector of length k reporting the number of events in the treatment-group etc. If we then use `rareDescribe` on the dataframe `dat` an object of the class `rareData` is returned.

```
dat <- rareDescribe(ai,bi,ci,di,n1i,n2i,data=dat)
```

Just as before, `dat` includes the event-counts of the k studies, now expanded by information about the number of double-zero studies, means and medians of event-counts and much more. Most importantly, `dat` is now in the right format to put into `rareIV` to calculate effect sizes. Again, let us look at `rareIV` and its parameters:

```
rareIV(x,measure,method,cc,ccval = 0.5,tccval,cccval,ccsum=1,ccto = "only0",drop00=TRUE,
       weighted=TRUE,level = 95,test="z",digits=4,verbose=FALSE,control)
```

The parameter `x` stands for an object of type `rareData`. Through `measure` we decide which effect size shall be estimated. Possible inputs are `logOR, logRR, RD` standing for the logarithm of the Odds-Ratio, the logarithm of the Relative-Risk and the Risk-Difference, respectively. By `method` we indicate the mathematical modeling, deciding between a fixed-effects and a random effects model and also decide the heterogeneity-estimator to be used. The arguments `cc` up to `ccto` specify the type of continuity correction to be applied. There is a dedicated section explaining the meaning and usage. The standard method of adding 0.5 to all cells of all single- and double-zero studies is achieved by setting `cc="constant"`. The logical `drop00` indicates whether studies with no events in both the treatment-group and the control group should be left out from the analysis. The argument `weighted` specifies whether estimations for effect sizes shall come by weighted averages (i.e. normed inverse variances) or unweighted averages (i.e. arithmetical means). The confidence interval is specified through `level` and through `test` we specify how test statistics and confidence intervals for the fixed effects should be computed. By `digits` we specify the number of decimal places to which the printed result should be rounded.

Let us now feed our dataset `dat` (now of type `rareData`) into this function. We choose to stick to the defaults as much as possible, specifying the effect size to be `logOR`, a fixed effects model and constant continuity correction.

```
Fit <- rareIV(x = dat,measure="logOR",method="FE",cc="constant")
```

The returned object `Fit` includes estimations the parameters of interest, for example the effect size, p-value, z-value, confidence intervals and much more. We can display its information by way of the `summary()` function or simply through the command `head()`.

Continuity Corrections

Theoretical Introduction

When confronted with data expressed in terms of a 2×2-table many standard methods of estimating effect sizes may fail.

	event	no-event
treatment	a_i	b_i
control	c_i	d_i

Let us, for example, consider the standard method of estimating the logarithm of the standard estimator of the relative-risk RR_i of the two groups in the i-th study:

$$\log(\hat{RR}_i) := \log \frac{a_i/(a_i + b_i)}{c_i/(c_i + d_i)}.$$

If either a_i, the number of events in the treatment group in study i, or c_i, the number of events in the control group in study i, is equal to zero, one is faced with an undefined term. Continuity corrections present a way of handling this problem. To apply the continuity correction we add specified values to the cells of the specified study. There are various methods to do so. Three of them will be discussed below. As an example consider the following study:

	event	no-event
treatment	0	n_T
control	0	n_C

a *double-zero study* with n_T and n_C being the size of the treatment group and the control group, respectively. The standard method of continuity correction is adding 0.5 to each cell of the study. If this method is applied, we end up with:

	event	no-event
treatment	0.5	$n_T + 0.5$
control	0.5	$n_C + 0.5$

Now, many methods of estimating effect sizes are available again.

Continuity corrections can be applied in various ways. Two main questions arise:

Which studies should the continuity correction be applied to?
Assume we are conducting meta-analysis in the presence of studies where no event is observed in either the treatment or the control group. Besides the two obvious ways of applying the continuity correction to all or none of the studies it is standard to apply the continuity correction to only those studies with no event in either the control or the treatment group.

Which method of continuity correction do I apply to the specified studies?

1. Constant Continuity Correction

We add a constant value to each cell of the specified studies. If we decide to add value x_i in the i-th study, this amounts to

	event	no-event
treatment	$a_i + x_i$	$b_i + x_i$
control	$c_i + x_i$	$d_i + x_i$.

The default is adding 0.5 to each cell of each specified study but it is also possible to add different values to different studies. The important characteristic of this approach is that the added value does not depend on the size or outcome of the studies. The possibility to do so is reflected in the next two approaches.

2. Treatment Arm Continuity Correction

In the specified studies, we add a value dependent on the size of the control group to the cells of the treatment group and vice versa. The value is calculated as the reciprocal of the total size of the other group. For the i-th study this amounts to:

	event	no-event
treatment	$a_i + \frac{1}{c_i+d_i}$	$b_i + \frac{1}{c_i+d_i}$
control	$c_i + \frac{1}{a_i+b_i}$	$d_i + \frac{1}{a_i+b_i}$.

3. Empirical Continuity Correction

In the specified studies, we add a value dependent on the respective group sizes and an estimation of the effect size only considering the non-zero studies. Again, let us consider the i'th study

	event	no-event
treatment	a_i	b_i
control	c_i	d_i.

Suppose we are interested in estimating the Odds Ratio. Let n_T be the size of the treatment group, n_C be the size of the control group of the i'th study. Let $\hat{\Omega}_{OR}$ be the standard estimation of the pooled Odds Ratio calculated only considering the non-zero studies. We choose the continuity corrections k_T for the treatment group and k_C for the control group in such a way that

$$\frac{k_T(n_C + k_C)}{k_C(n_T + k_T)} = \hat{\Omega}_{OR}.$$

This ensures that the estimated Odds Ratio which is obtained for a double-zero study after the continuity correction is applied amounts to $\hat{\Omega}_{OR}$.\ Applying the continuity correction to the i-th study makes us end up with:

	event	no-event
treatment	$a_i + k_T$	$b_i + k_T$
control	$c_i + k_C$	$d_i + k_C$.

To receive unique solutions for k_T and k_C we use the following restriction:

$$k_T + k_C = 1,$$

which, for example, holds true for the standard constant continuity correction of $k_T = k_C = 0.5$. Now, with $R := \frac{n_C}{n_T}$, the group ratio imbalance, and the approximation

$$\frac{k_T(n_C + k_C)}{k_C(n_T + k_T)} \approx \frac{Rk_T}{k_C}$$

which is true for large enough groups, we now see that it holds that:

$$k_C \approx \frac{R}{R + \hat{\Omega}_{OR}} \text{ and } k_T \approx \frac{\hat{\Omega}_{OR}}{R + \hat{\Omega}_{OR}}.$$

Continuity Corrections in raremeta

We have seen different ways of continuity correcting studies. Let us go through the syntax of **raremeta** to learn how to apply them. It all comes down to specifying parameters the parameters from `cc` up to `ccto` in

```
rareIV(x,measure,method,cc,ccval = 0.5,tccval,cccval,ccsum=1,ccto = "only0",drop00=TRUE,
    weighted=TRUE,level = 95,test="z",digits=4,verbose=FALSE,control).
```

The first parameter, `cc`, decides which of the three introduced methods of continuity correction should be applied. Possible inputs are `none`, which stands for the option to not apply any continuity correction and `"constant", "tacc"` and `"empirical"` which stand for the constant continuity correction, the treatment arm continuity correction and the empirical continuity correction, respectively. The argument `"ccto"` specifies the studies, continuity correction shall be applied to. Possible inputs are `"only0", "all", "if0all"`. While `"only0" and "all"` stand for applying continuity correction to those studies with no event in either the treatment- or the control group and all studies (regardless the existence of single-zero or double-zero studies), respectively, `"if0all"` leads to the application of continuity corrections to all studies if there is either a single-zero study or a double-zero study or none of the studies if there are neither single-zero studies nor double-zero studies. If one opts for `cc = "constant"`, it must be specified, which value should be applied to the relevant cells. This happens either through `"ccval"` or the two arguments `"tccval"` and `"cccval"`. While specifying `"ccval"` is used when continuity corrections in the control group and the treatment group shall be the same, the arguments `"tccval"` and `"cccval"` enable the user to input continuity corrections for the treatment group and the control group, respectively. In both cases, if a single value is put in, all corresponding cells of the specified studies will be continuity corrected with this value. If the input comes in form of vectors, whose length is equal to the number of studies, then the corresponding cells of the i'th study will be continuity corrected through the i'th entry of the vector. Obviously, this enables the user simulate her own way of continuity correcting studies. If one opts for `cc = "tacc"` or `cc = "empirical"`, it must be specified, to which value the values k_T and k_C from the theoretical introduction above should add up to. This happens through the argument `ccsum`.

If one, for example, wants to apply the treatment arm continuity correction where $k_T + k_C = 0.1$ to all studies of the `rareData`-object `dat`, one may input

```
rareIV(x = dat,measure="logOR",method="FE",cc="tacc",ccsum=0.1,ccto="all").
```

To apply the constant continuity correction with specified values (potentially vectors of length > 1) t for the treatment group and c for the control group to all single- and double-zero studies, one may input

```
rareIV(x= dat,measure="logOR",method="FE",cc="constant",
       tccval=t,cccval=c,ccto="only0",drop00=FALSE).
```

Appendix

Appendix A: Motivating the Empirical Continuity Correction

In this section we want to shed light on a certain aspect of continuity corrections by answering the following question:

What is the estimated effect size for a double-zero study after the continuity correction is applied?

Suppose we are interested in the Odds Ratio. Let n_T and n_C refer to the size of the treatment group and the control group, respectively.

	event	no-event
treatment	0	n_T
control	0	n_C

Let k_T and k_C be the continuity corrections applied to the treatment group and the control group respectively. After the continuity correction is applied we end up with:

	event	no-event
treatment	k_T	$n_T + k_T$
control	k_C	$n_C + k_C$.

Let $(\neg)E$ stand for (no-)event, T for the treatment group and C for the control group. If we now estimate the Odds Ratio via plug in, using estimations of the risk in the two study arms:

$$\widehat{\mathbb{P}(E|T)} = \frac{k_T}{n_T + 2k_T} \text{ and } \widehat{\mathbb{P}(E|C)} = \frac{k_C}{n_C + 2k_C},$$

we end up with:

$$\begin{aligned}
\hat{OR} &:= \frac{\widehat{\mathbb{P}(E|T)}/(1 - \widehat{\mathbb{P}(E|T)})}{\widehat{\mathbb{P}(E|C)}/(1 - \widehat{\mathbb{P}(E|C)})} \\
&= \frac{\frac{k_T}{n_T+2k_T}/(1 - \frac{k_T}{n_T+2k_T})}{\frac{k_C}{n_C+2k_C}/(1 - \frac{k_C}{n_C+2k_C})} \\
&= \frac{\frac{k_T}{n_T+2k_T}/\frac{n_T+k_T}{n_T+2k_T}}{\frac{k_C}{n_C+2k_C}/\frac{n_C+k_C}{n_C+2k_C}} \\
&= \frac{k_T/(n_T + k_T)}{k_C/(n_C + k_C)} \\
&= \frac{k_T(n_C + k_C)}{k_C(n_T + k_T)}
\end{aligned}$$

Now, with the group ratio imbalance $R := \frac{n_C}{n_T}$, we can easily describe what the three approaches amount to. For the constant continuity correction with $k_T = k_C = \alpha$ sufficiently small (e.g. 0.5) we approximately get

$$\hat{OR} = \frac{\alpha(n_T\ R + \alpha)}{\alpha(n_T + \alpha)} \approx \frac{\alpha\ n_T\ R}{\alpha\ n_T} = R.$$

For the reciprocal continuity correction with $k_T = 1/n_C$ and $k_C = 1/n_T$ get

$$\hat{OR} = \frac{1/n_C(n_C - 1/n_T)}{1/n_T(n_T - 1/n_C)} = 1$$

and, by definition, for the empirical continuity correction for the prior $\hat{\Omega}$ this amounts to

$$\hat{OR} = \hat{\Omega}_{OR}.$$

In summary, the constant continuity correction pulls the estimated Odds Ratio towards the group ratio imbalance, the reciprocal continuity correction towards no effect and the empirical continuity correction towards the estimated pooled Odds Ratio using only the non-zero studies.