

Eerste Analyse

Jan van Rongen

2015-02-28 22:50; 2015-03-02 18:30

Introductie en samenvatting

Dit is mijn eerste analyse gebaseerd op de spreadsheet. De gegevens uit de eerste tabel van deze spreadsheet zijn gekopieerd naar de csv-file "cases.csv". De kolomnamen in deze file zijn door mij toegevoegd.

Deze analyse volgt de lijnen van de tutorial voor het metafor pakket bij R (*Conducting Meta-Analyses in R with the metafor Package*, Wolfgang Viechtbauer, Maastricht University. *Journal of Statistical Software*; August 2010, Volume 36, Issue 3.).

De meta-analyse is uitgevoerd op de 39 studies die in de genoemde tabel voldoende gegevens bevatten om een model te kunnen maken. Dit is dus inclusief studies waarvoor de spreadsheet 0 positieven rapporteert. Met behulp van het Random Effect Sizes model vinden we een significant effect (0.71, CI=[0.59, 0.86]) voor het Relatieve Risico.

De funnel-plot en de trim-fill test geven geen indicatie voor publication bias.

De heterogeniteit in de studies is zodanig groot, dat een verdere analyse met een Mixed Effects Model nuttig zou kunnen zijn.

```
library(metafor)
mydata<-read.csv("cases.csv", sep=";", header=TRUE)
# delete incomplete cases
mydata<-mydata[complete.cases(mydata),]
print(mydata[1:3,], row.names=FALSE)
```

```
##          author year tpos tttotal cpos cttotal
##      [2] Selinger 2013   14    117   16    112
##      [7] Hickson  2007    0     69    0     66
##      [8] Beausoleil 2007   21     44   20     45
```

Ik heb een aanname gemaakt over de betekenis van de getallen in de spreadsheet, nl. dat zij resp. de positieve gevallen en de omvang van de totale groep weergeven. Metafor werkt standaard met het aantal positieve en negatieve gevallen. De berekende Relatieve Risico's komen overeen met die in de spreadsheet, dus deze aanname zal wel kloppen.

Random Effect Sizes Model

```
dat <- escalc(measure = "RR", ai = tpos, bi = tttotal-tpos, ci = cpos, di = cttotal-cpos
              , data = mydata, append = TRUE)
res <- rma(yi, vi, data = dat) # make the model
ci <- confint(res) # confidence intervals of important measures
print(dat[1:3,], row.names=FALSE); print(res); print(ci)
```

```
##          author year tpos tttotal cpos cttotal      yi      vi
##      [2] Selinger 2013   14    117   16    112 -0.1772 0.1165
##      [7] Hickson  2007    0     69    0     66 -0.0438 3.9708
##      [8] Beausoleil 2007   21     44   20     45  0.0713 0.0527
```

```
##
## Random-Effects Model (k = 39; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.1645 (SE = 0.0669)
## tau (square root of estimated tau^2 value):      0.4056
## I^2 (total heterogeneity / total variability):    72.35%
## H^2 (total variability / sampling variability):   3.62
##
## Test for Heterogeneity:
## Q(df = 38) = 97.3684, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## -0.3368      0.0967   -3.4838   0.0005   -0.5263   -0.1473      ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
##      estimate      ci.lb      ci.ub
## tau^2      0.1645   0.0429   0.2742
## tau        0.4056   0.2071   0.5236
## I^2(%)     72.3516  40.5540  81.3494
## H^2        3.6168   1.6822   5.3618
```

Met `escalc` berekenen we de gewenste effect groottes. Metafor maakte voor “RR” een log Relative Risk. Dat heeft voordelen, maar in de plot hierna moeten we de resulatten wel terug vertalen naar de Relative Risk zelf.

Het resultaat

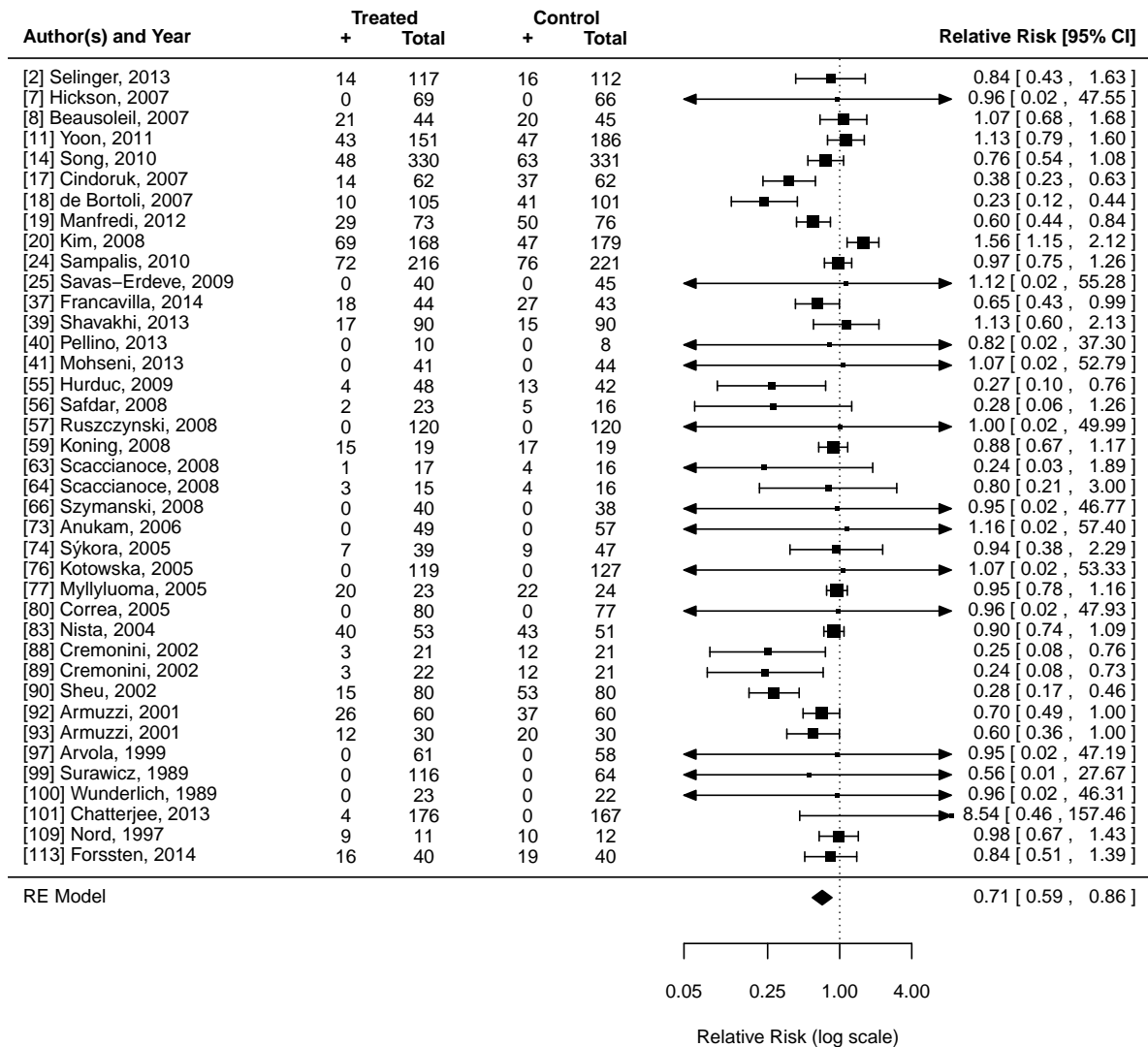
Het resultaat van deze eerste berekeningen geeft een effect van 71%: het risico van de pro-biotica-groep ten opzichte van de controle groep. Het resultaat is robuust met $p < 0.001$

De heterogeniteit staat in de τ^2 statistic. De I^2 statistic (Higgins & Thompson 2002) schat het percentage van de totale variabiliteit dat kan worden toegerekend aan de variabiliteit van de onderliggende testen. Deze maat, tezamen met de Q-test voor heterogeniteit suggereert duidelijk een heterogeniteit onder de werkelijke effecten, die een verder onderzoek *in een mixed effects model* wenselijk maakt.

De forest plot

```
# forest plot
N=dim(mydata)[1]+2
forest(res, slab = paste(dat$author, dat$year, sep = ", ")
      , xlim = c(-16, 6), at = log(c(0.05, 0.25, 1, 4)), atransf = exp
      , ilab = cbind(dat$tpos, dat$tttotal, dat$cpes, dat$cttotal)
      , ilab.xpos = c(-9.5, -8, -6, -4.5), cex = 0.75)
# legends on this plot
op <- par(cex = 0.75, font = 2)
text(c(-9.5, -8, -6, -4.5), N, c("+", "Total", "+", "Total"))
text(c(-8.75, -5.25), N+1, c("Treated", "Control"))
```

```
text(-16, N, "Author(s) and Year", pos = 4)
text(6, N, "Relative Risk [95% CI]", pos = 2)
```

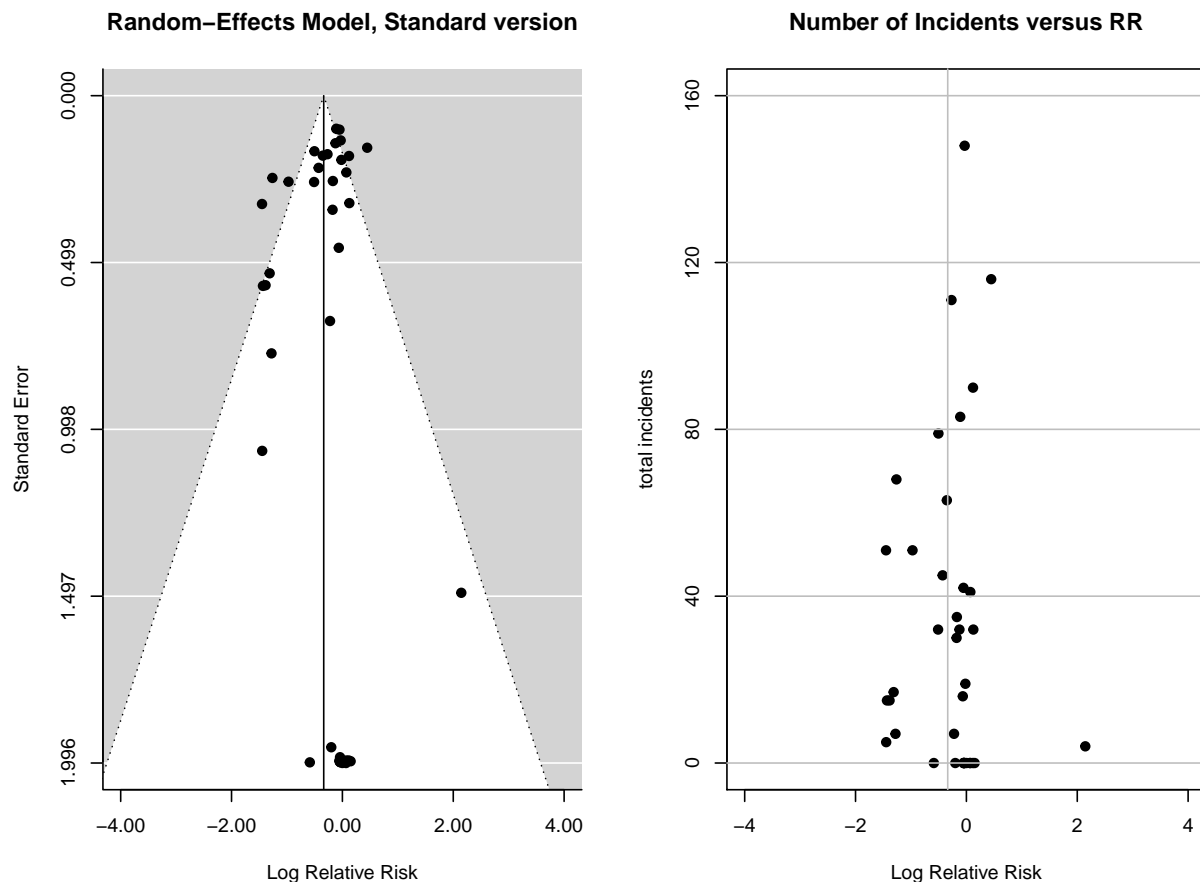


De funnel plot

De standaard funnel plot links, de gevraagde (totaal positieve gevallen versus RR) rechts. De laatste is niet als zodanig aanwezig in Metafor, maar is zonder de achtergrond van een trechter zo gemaakt.

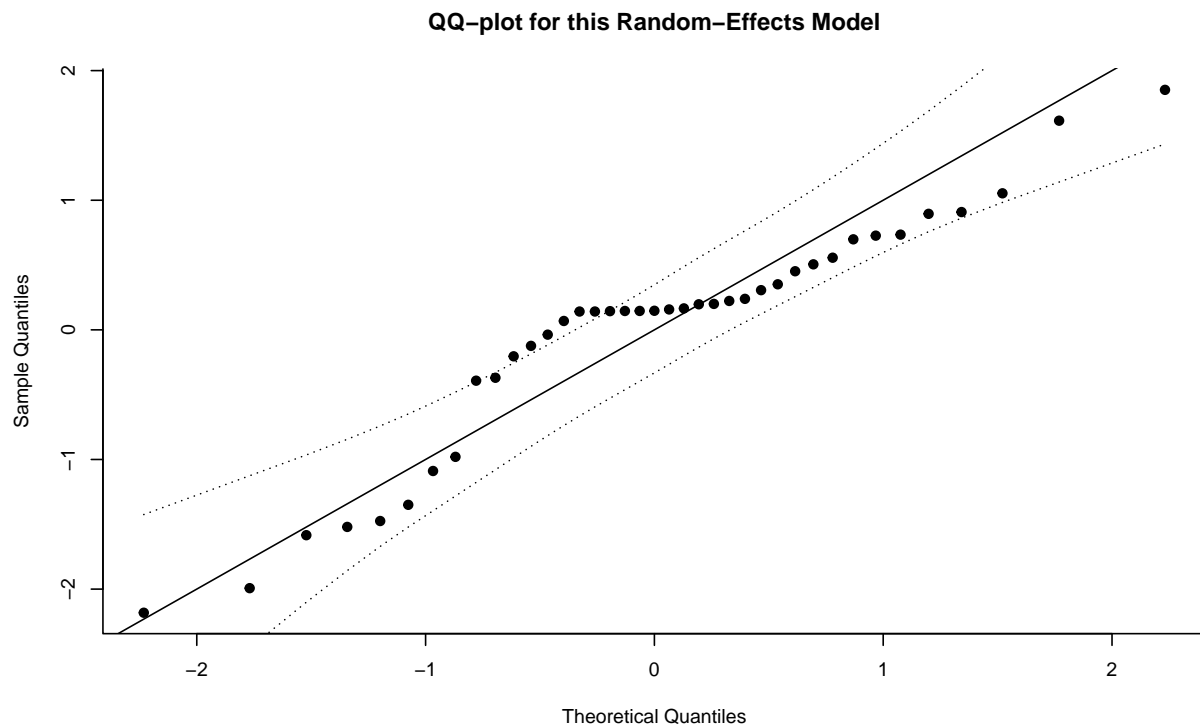
```
par(op)
op=par(mfrow=c(1,2), cex = 0.75, font = 2)
# res <- rma(yi, vi, data = dat) -- already done above
funnel(res, main = "Random-Effects Model, Standard version", xlim=c(-4,4))
```

```
n<-dat$tpos+dat$cpos
plot(pch=19, n~dat$yi, xlim=c(-4,4), ylim=c(0,160), main="Number of Incidents versus RR",
     ylab="total incidents", xlab="Log Relative Risk", axes = FALSE, frame.plot = TRUE)
axis(1, at=seq(-4,4, by=2), labels= formatC(seq(-4,4, by=2)))
axis(2, at=c(0,40,80,120, 160), labels= formatC(c(0,40,80,120, 160)))
abline(h=c(0,40,80,120, 160), col="grey")
abline(v=res$b, col="grey")
```



Dit vraagt een verder uitwerking. Bij Publication Bias is vooral de links-rechts verdeling scheef, maar dat is niet het geval. Hier lijkt eerder de variantie niet in orde, want die is niet trechtervormig. Ook hier is een aanwijzing dat indien mogelijk een mixed effect model wordt onderzocht.

```
par(op)
op=par(cex = 0.75, font = 2)
qqnorm(res, main = "QQ-plot for this Random-Effects Model")
```



```
regtest(res, model = "lm")
```

```
##
## Regression Test for Funnel Plot Asymmetry
##
## model:      weighted regression with multiplicative dispersion
## predictor: standard error
##
## test for funnel plot asymmetry: t = -1.6235, df = 37, p = 0.1130
```

```
trimfill(res)
```

```
##
## Estimated number of missing studies on the right side: 0 (SE = 2.8897)
##
## Random-Effects Model (k = 39; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.1645 (SE = 0.0669)
## tau (square root of estimated tau^2 value):      0.4056
## I^2 (total heterogeneity / total variability):    72.35%
## H^2 (total variability / sampling variability):    3.62
##
## Test for Heterogeneity:
## Q(df = 38) = 97.3684, p-val < .0001
##
## Model Results:
##
```

```
## estimate      se      zval      pval      ci.lb      ci.ub
## -0.3368      0.0967     -3.4838     0.0005     -0.5263     -0.1473      ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De QQ plot is dan ook niet erg vlak en de regressietest geeft wel aanleiding tot enig nadenken. Echter, de trim-fill test geeft een schatting van 0 ontbrekende studies (op de 39), dus een eerste indicatie is dat er *weinig aanwijzingen voor publication bias zijn*.

Vervolg

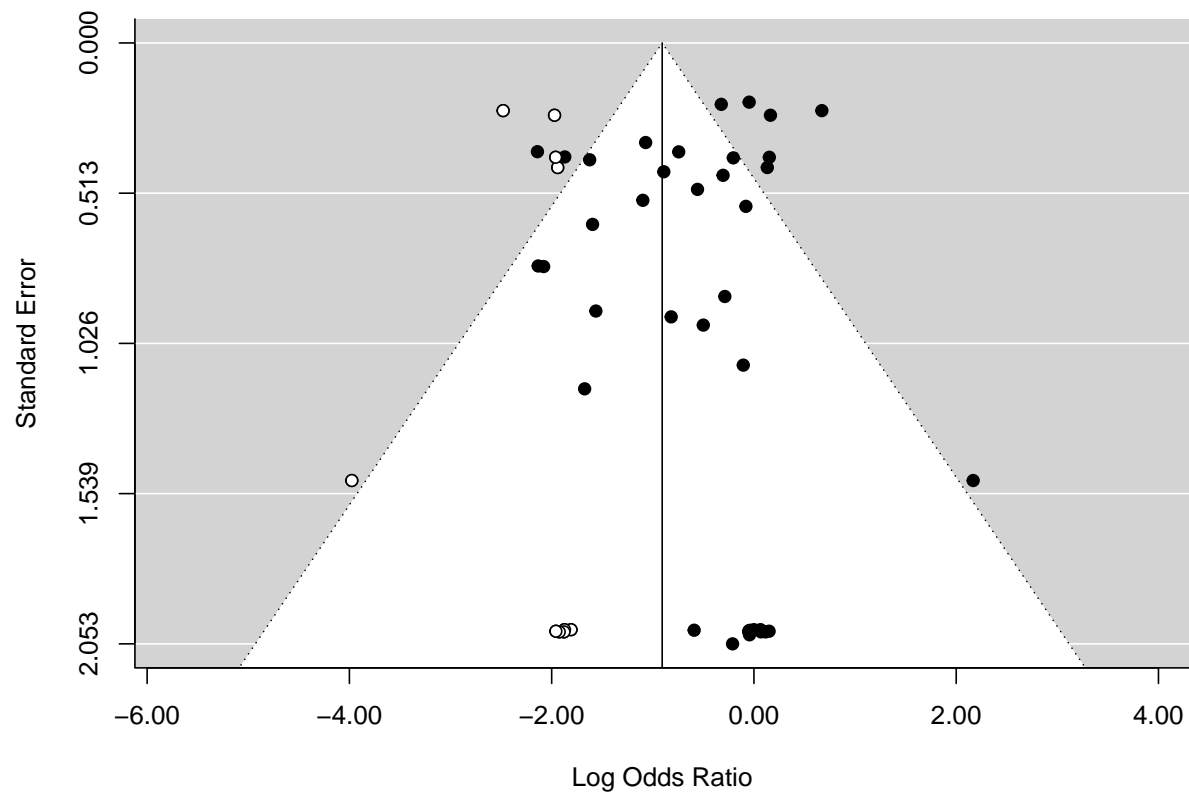
Belangrijkste lijkt me nu de berekeningen te herhalen met een of meer aanvullende gegevens, zodat een mixed effects model kan worden gebruikt. Ondertussen wil ik kijken of een nieuwe methode van Simonsohn e.a. bruikbaar kan zijn, de p-curve analyse (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2256237).

Er is nog iets eigenaardigs: met de Odds ratio (ook een log) vinden we wel een publication bias. Kan ik nog niet verklaren.

```
dat2 <- escalc(measure = "OR", ai = tpos, bi = ttotal-tpos, ci = cpos, di = cttotal-cpos
               , data = mydata, append = TRUE)
res2 <- rma(yi, vi, data = dat2)
print(trimfill(res2))
```

```
##
## Estimated number of missing studies on the left side: 10 (SE = 4.1152)
##
## Random-Effects Model (k = 49; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.6550 (SE = 0.2304)
## tau (square root of estimated tau^2 value):      0.8093
## I^2 (total heterogeneity / total variability):    74.05%
## H^2 (total variability / sampling variability):    3.85
##
## Test for Heterogeneity:
## Q(df = 48) = 229.3906, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## -0.9071      0.1644     -5.5174     <.0001     -1.2294     -0.5849      ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
funnel(trimfill(res2))
```



```
par(op)
```

Jan