Introduction to meta-analysis

Steve Simon

11/14/2018

Overview

This presentation will show how to use base R functions to reproduce the elements of a forest plot. It will also offer some suggestions on how you should interpret patterns in this plot.

BCG description

"Results from 13 studies examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis."

The BCG data comes with the metafor package in R, but is also available at https://rdrr.io/cran/metafor /man/dat.colditz1994.html (https://rdrr.io/cran/metafor/man/dat.colditz1994.html)

BCG data, first three columns

```
##    trial                author year
## 1      1               Aronson 1948
## 2      2       Ferguson & Simes 1949
## 3      3        Rosenthal et al 1960
## 4      4      Hart & Sutherland 1977
## 5      5 Frimodt-Moller et al 1973
## 6      6        Stein & Aronson 1953
## 7      7       Vandiviere et al 1973
## 8      8             TPT Madras 1980
## 9      9        Coetzee & Berjak 1968
## 10    10        Rosenthal et al 1961
## 11    11         Comstock et al 1974
## 12    12       Comstock & Webster 1969
## 13    13         Comstock et al 1976
```

BCG data, last six columns

```
##     tpos   tneg cpos   cneg ablat       alloc
## 1      4    119   11    128    44      random
## 2      6    300   29    274    55      random
## 3      3    228   11    209    42      random
## 4     62  13536  248  12619    52      random
## 5     33   5036   47   5761    13   alternate
## 6    180   1361  372   1079    44   alternate
## 7      8   2537   10    619    19      random
## 8    505  87886  499  87892    13      random
## 9     29   7470   45   7232    27      random
## 10    17   1699   65   1600    42  systematic
## 11   186  50448  141  27197    18  systematic
## 12     5   2493    3   2338    33  systematic
## 13    27  16886   29  17825    33  systematic
```

```
Fixed-Effects Model (k = 13)


Test for Heterogeneity:
Q(df = 12) = 163.1649, p-val < .0001


Model Results:


estimate       se      zval     pval    ci.lb     ci.ub
 -0.4361   0.0423   -10.3190   <.0001   -0.5190   -0.3533   ***


---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
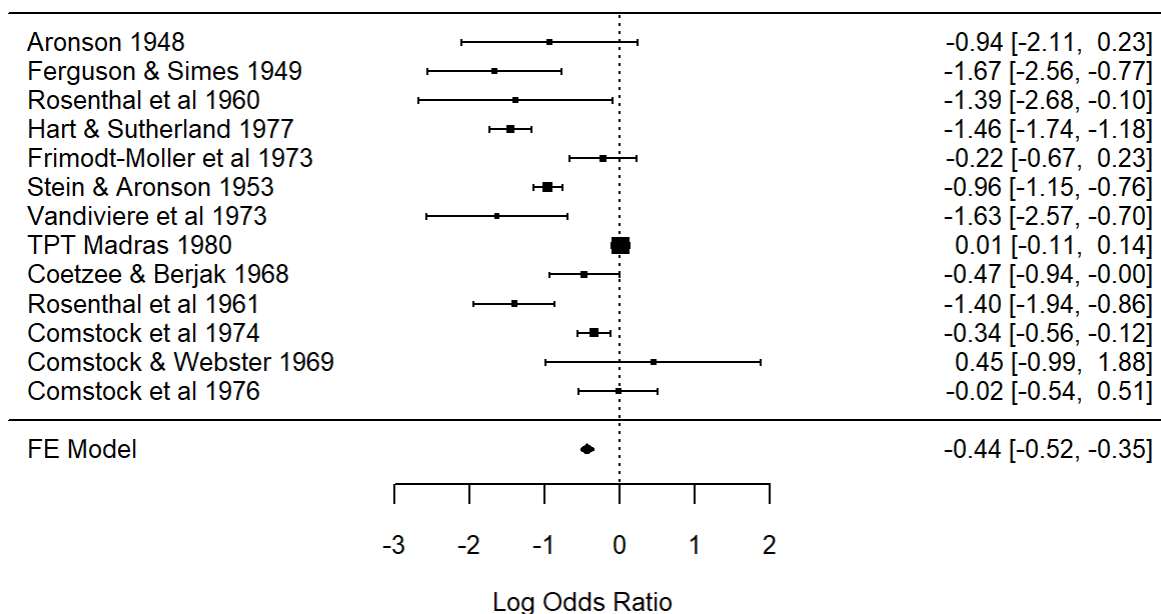
Study		Log Odds Ratio [95% CI]
Aronson 1948		-0.94 [-2.11, 0.23]
Ferguson & Simes 1949		-1.67 [-2.56, -0.77]
Rosenthal et al 1960		-1.39 [-2.68, -0.10]
Hart & Sutherland 1977		-1.46 [-1.74, -1.18]
Frimodt-Moller et al 1973		-0.22 [-0.67, 0.23]
Stein & Aronson 1953		-0.96 [-1.15, -0.76]
Vandiviere et al 1973		-1.63 [-2.57, -0.70]
TPT Madras 1980		0.01 [-0.11, 0.14]
Coetzee & Berjak 1968		-0.47 [-0.94, -0.00]
Rosenthal et al 1961		-1.40 [-1.94, -0.86]
Comstock et al 1974		-0.34 [-0.56, -0.12]
Comstock & Webster 1969		0.45 [-0.99, 1.88]
Comstock et al 1976		-0.02 [-0.54, 0.51]
FE Model		-0.44 [-0.52, -0.35]

Log Odds Ratio

-3   -2   -1   0   1   2

Here are the results for the first study, listed in a two by two table. A positive result is bad, indicating that the vaccine did not work.

```
##     Pos Neg
## T 4    119
## C 11   128
```

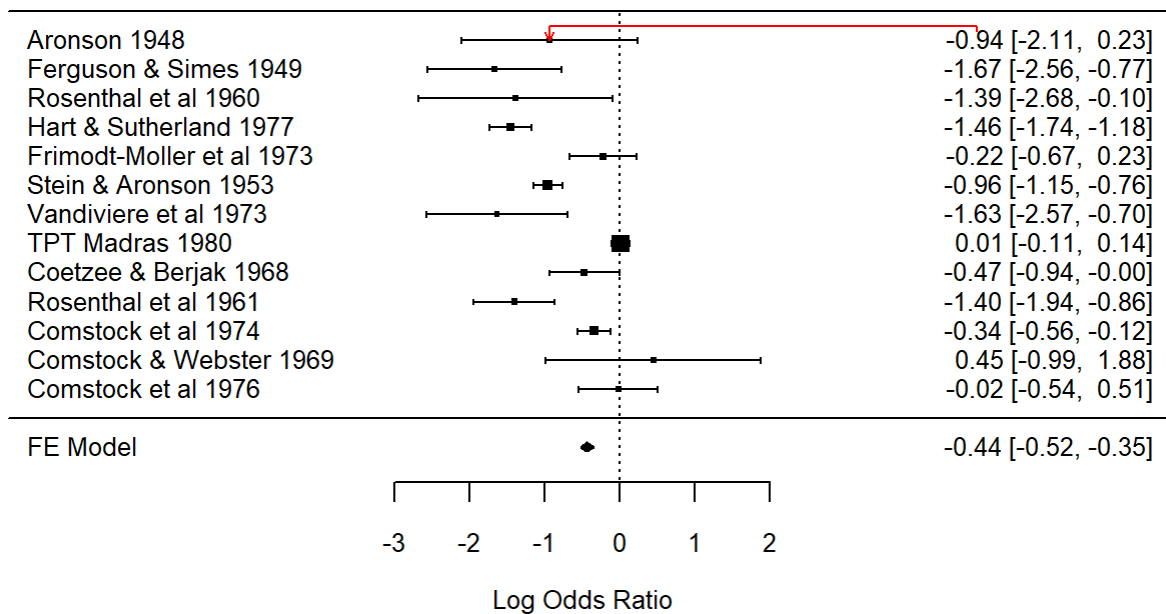Calculate the odds for each row.

```
## 4 / 119 = 0.034
```

```
## 11 / 128 = 0.086
```

The ratio of these two odds is the odds ratio.

```
## 0.034 / 0.086 = 0.391
```

The underlying distribution of the odds ratio is almost always skewed right and bounded below by zero, so you should consider a log transformation.

```
## log(0.391)=-0.939
```

Study	Log Odds Ratio [95% CI]
Aronson 1948	-0.94 [-2.11, 0.23]
Ferguson & Simes 1949	-1.67 [-2.56, -0.77]
Rosenthal et al 1960	-1.39 [-2.68, -0.10]
Hart & Sutherland 1977	-1.46 [-1.74, -1.18]
Frimodt-Moller et al 1973	-0.22 [-0.67, 0.23]
Stein & Aronson 1953	-0.96 [-1.15, -0.76]
Vandiviere et al 1973	-1.63 [-2.57, -0.70]
TPT Madras 1980	0.01 [-0.11, 0.14]
Coetzee & Berjak 1968	-0.47 [-0.94, -0.00]
Rosenthal et al 1961	-1.40 [-1.94, -0.86]
Comstock et al 1974	-0.34 [-0.56, -0.12]
Comstock & Webster 1969	0.45 [-0.99, 1.88]
Comstock et al 1976	-0.02 [-0.54, 0.51]
FE Model	-0.44 [-0.52, -0.35]

Log Odds Ratio (x-axis: -3, -2, -1, 0, 1, 2)

The variance of the log odds ratio is approximately

$$\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$$

where n11, n12, etc. are the count of each cell in the two by two table. For the first study, you get
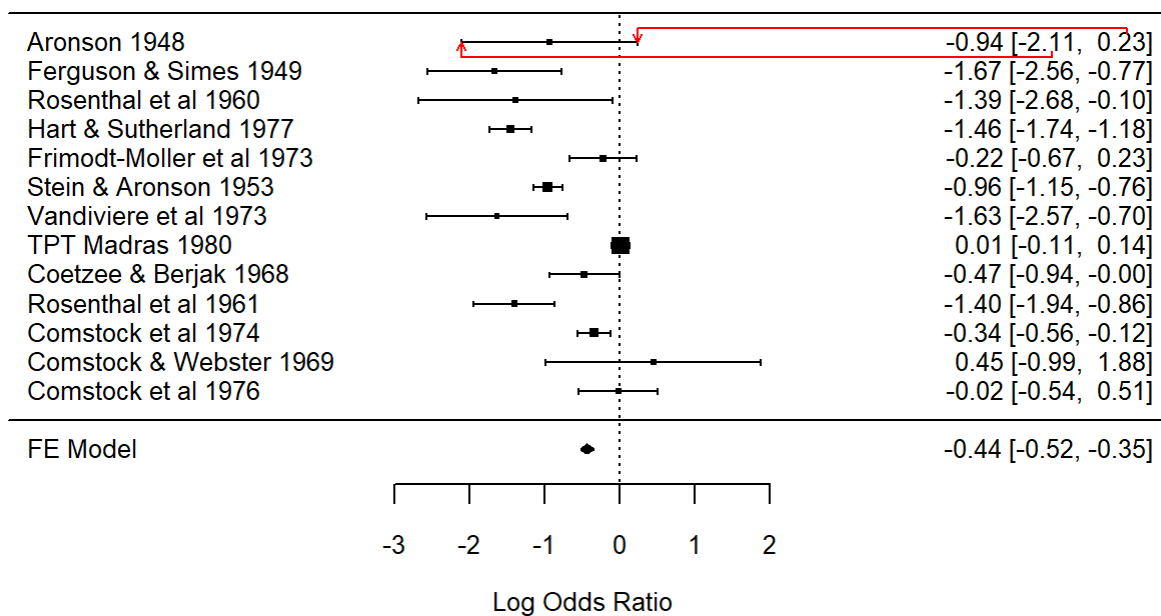
```
## 1/4 + 1/119 + 1/11 + 1/128 = 0.357
```

The approximate confidence interval for the log odds is

```
## -0.939-1.96*sqrt(0.357)=-2.11
```

```
## -0.939+1.96*sqrt(0.357)=0.232
```

These limits include the value of zero, so this particular study failed to reject the null hypothesis.

Aronson 1948		-0.94 [-2.11, 0.23]
Ferguson & Simes 1949		-1.67 [-2.56, -0.77]
Rosenthal et al 1960		-1.39 [-2.68, -0.10]
Hart & Sutherland 1977		-1.46 [-1.74, -1.18]
Frimodt-Moller et al 1973		-0.22 [-0.67, 0.23]
Stein & Aronson 1953		-0.96 [-1.15, -0.76]
Vandiviere et al 1973		-1.63 [-2.57, -0.70]
TPT Madras 1980		0.01 [-0.11, 0.14]
Coetzee & Berjak 1968		-0.47 [-0.94, -0.00]
Rosenthal et al 1961		-1.40 [-1.94, -0.86]
Comstock et al 1974		-0.34 [-0.56, -0.12]
Comstock & Webster 1969		0.45 [-0.99, 1.88]
Comstock et al 1976		-0.02 [-0.54, 0.51]
FE Model		-0.44 [-0.52, -0.35]

-3   -2   -1   0   1   2

Log Odds Ratio
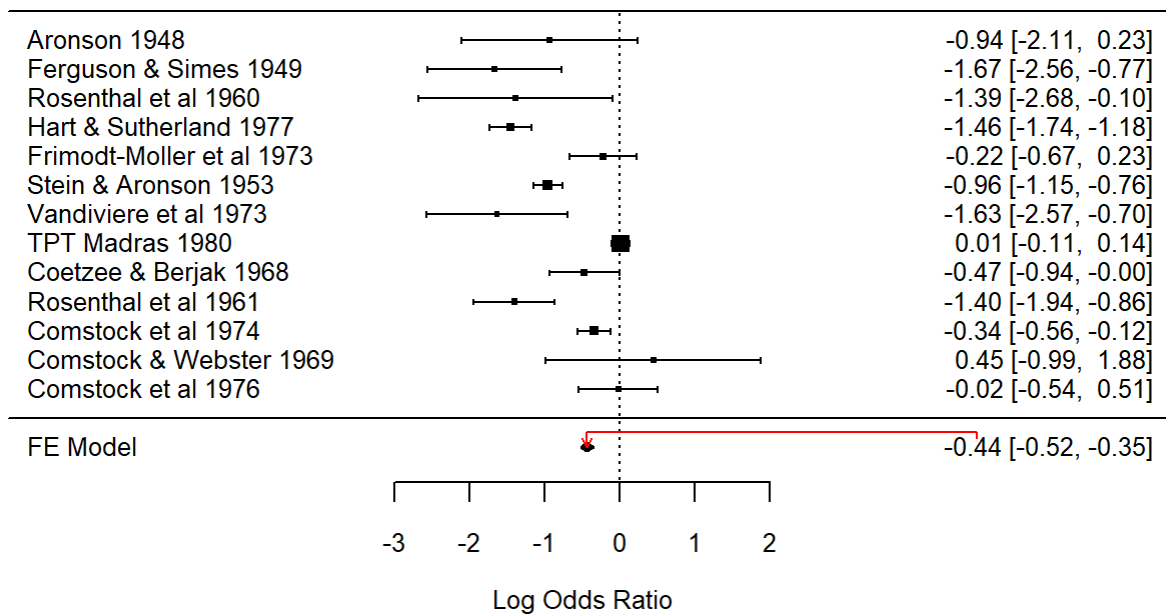
The log odds ratios and variances for the full data set are listed below.

```
##        y      w
## 1   -0.94    2.8
## 2   -1.67    4.8
## 3   -1.39    2.3
## 4   -1.46   49.2
## 5   -0.22   19.2
## 6   -0.96  101.0
## 7   -1.63    4.4
## 8    0.01  249.6
## 9   -0.47   17.6
## 10  -1.40   13.3
## 11  -0.34   79.8
## 12   0.45    1.9
## 13  -0.02   14.0
```

The overall estimate is a weighted average of the individual study estimates.

```
## (-0.94)(2.8)+(-1.67)(4.8)+...+(-0.02)(14)=-0.44
```

	Log Odds Ratio	
Aronson 1948		-0.94 [-2.11, 0.23]
Ferguson & Simes 1949		-1.67 [-2.56, -0.77]
Rosenthal et al 1960		-1.39 [-2.68, -0.10]
Hart & Sutherland 1977		-1.46 [-1.74, -1.18]
Frimodt-Moller et al 1973		-0.22 [-0.67, 0.23]
Stein & Aronson 1953		-0.96 [-1.15, -0.76]
Vandiviere et al 1973		-1.63 [-2.57, -0.70]
TPT Madras 1980		0.01 [-0.11, 0.14]
Coetzee & Berjak 1968		-0.47 [-0.94, -0.00]
Rosenthal et al 1961		-1.40 [-1.94, -0.86]
Comstock et al 1974		-0.34 [-0.56, -0.12]
Comstock & Webster 1969		0.45 [-0.99, 1.88]
Comstock et al 1976		-0.02 [-0.54, 0.51]
FE Model		-0.44 [-0.52, -0.35]

The variance of the overall estimate is simply the reciprocal of the sum of the weights.

```
## 1/(2.8+4.8+ ... +14)=0.00179
```

The confidence limits for the overall estimate are
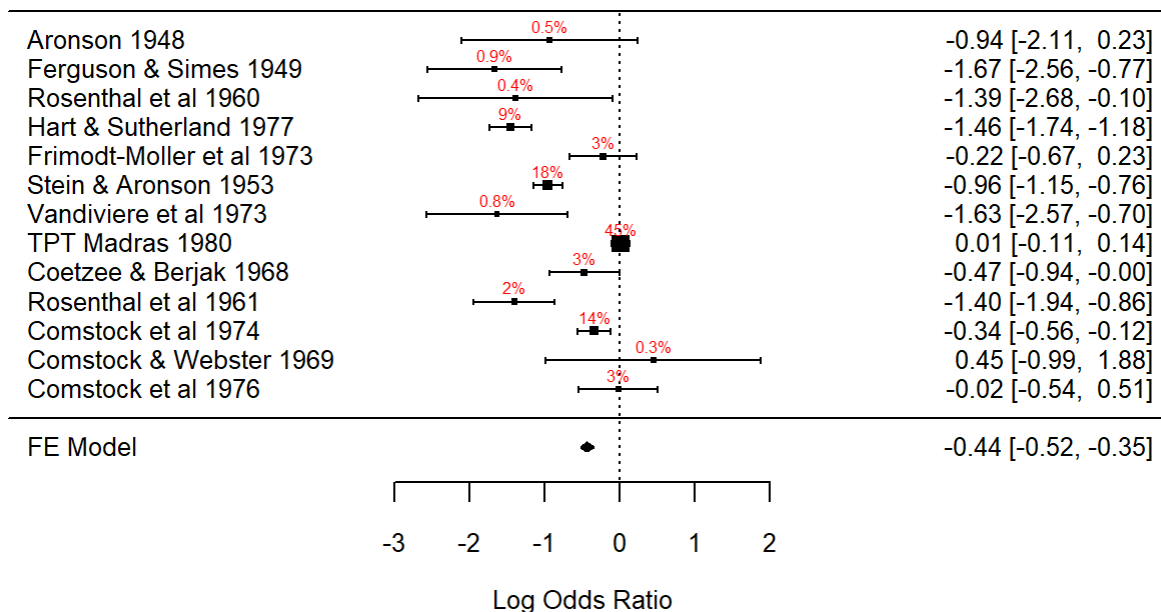
```
## -0.44-1.96 sqrt(0.00179)=-0.52
```

```
## -0.44+1.96 sqrt)0.00179)=-0.36
```
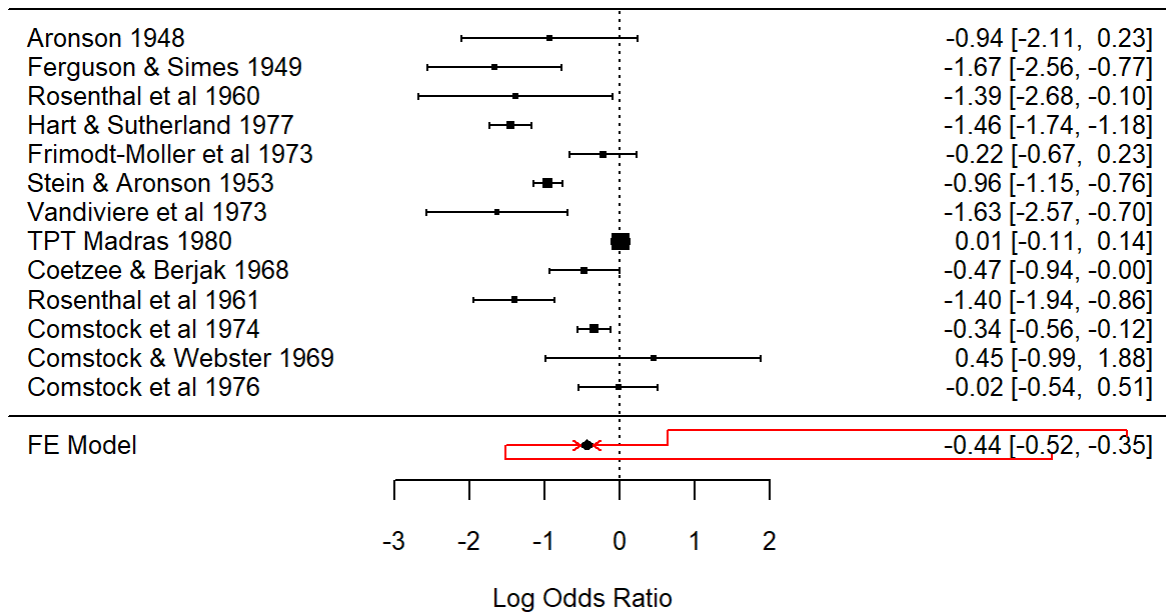
The relative weights given to each study are listed below.

```
##     study     wt rel_wt
## 1       1    2.8   0.5%
## 2       2    4.8   0.9%
## 3       3    2.3   0.4%
## 4       4   49.2     9%
## 5       5   19.2     3%
## 6       6  101.0    18%
## 7       7    4.4   0.8%
## 8       8  249.6    45%
## 9       9   17.6     3%
## 10     10   13.3     2%
## 11     11   79.8    14%
## 12     12    1.9   0.3%
## 13     13   14.0     3%
## 14  Total  559.9  99.9%
```

The relative weights determine the size of the squares in the forest plot.

Aronson 1948	-0.94 [-2.11, 0.23]
Ferguson & Simes 1949	-1.67 [-2.56, -0.77]
Rosenthal et al 1960	-1.39 [-2.68, -0.10]
Hart & Sutherland 1977	-1.46 [-1.74, -1.18]
Frimodt-Moller et al 1973	-0.22 [-0.67, 0.23]
Stein & Aronson 1953	-0.96 [-1.15, -0.76]
Vandiviere et al 1973	-1.63 [-2.57, -0.70]
TPT Madras 1980	0.01 [-0.11, 0.14]
Coetzee & Berjak 1968	-0.47 [-0.94, -0.00]
Rosenthal et al 1961	-1.40 [-1.94, -0.86]
Comstock et al 1974	-0.34 [-0.56, -0.12]
Comstock & Webster 1969	0.45 [-0.99, 1.88]
Comstock et al 1976	-0.02 [-0.54, 0.51]
FE Model	-0.44 [-0.52, -0.35]

Finally, notice a dotted vertical line at the log odds ratio=0. This is the null value.

What to look for in a forest plot

The forest plot provides information about individual studies, and how they relate to the overall estimate. Here are some things to look for.

1. Look at the overall effect first. Does the overall confidence interval diamond intersect the vertical dotted line of no effect?

2. Count the number of individual studies which fail to achieve statistical significance. These are individual confidence intervals that intersect the vertical dotted line. The count of non-significant studies is not a primary measure of evidence, because you can easily have a meta-analysis where every single individual study is not statistically significant, but the overall result is strongly significant. Still, it's an interesting measure, and gives you an initial impression of what the existing publications have said so far.

3. Count the number of studies where the point estimate is on the "wrong" side of the dotted line. Again, this is not a primary measure of evidence. Informally, a meta-analysis strengthens the results of each individual study if all the individual studies are pointing in the same direction. A few studies pointing in the opposite direction will tend to dilute the strength of a meta-analytic finding (as they should).

4. Look at the relative weights given to each study. Does one study tend to dominate the findings?

5. Look for evidence of heterogeneity. This can occur several ways, but the easiest source of heterogeneity is a bimodal pattern in the studies. Most of the studies might show little or no effect, for example, but a smaller group of studies might cluster in a region of a strong effect. If you are familiar with the individual studies, ask yourself if there is a feature that is unique to the outlying cluster of studies.