A random effects variance shift model for detecting and accommodating outliers in meta-analysis

Freedom Gumedze[1] Dan Jackson[2]

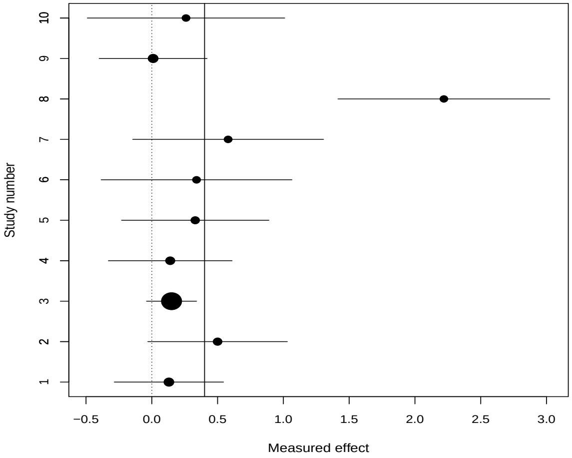
[1]Department of Statistical Sciences
University of Cape Town
Freedom.Gumedze@uct.ac.za

[2]MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

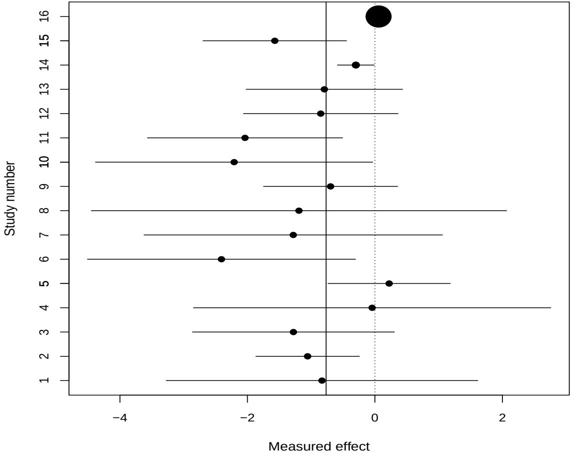
CANNeCTIN - OTN Seminar, PHRI, Canada
14 October 2011

Outline

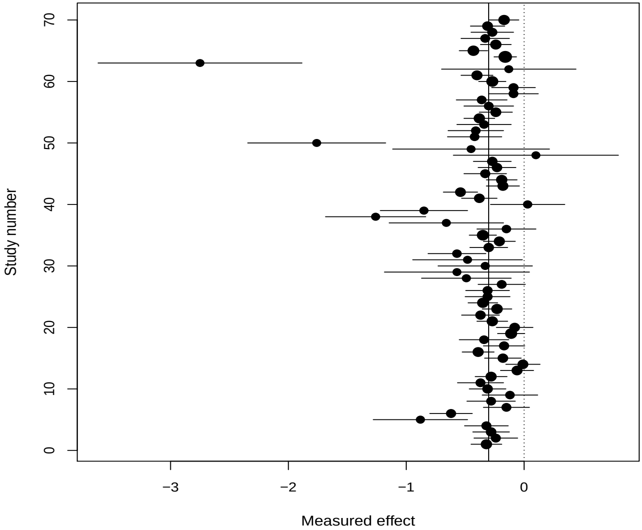
The CDP-choline data



Intravenous magnesium in acute myocardial infarction data



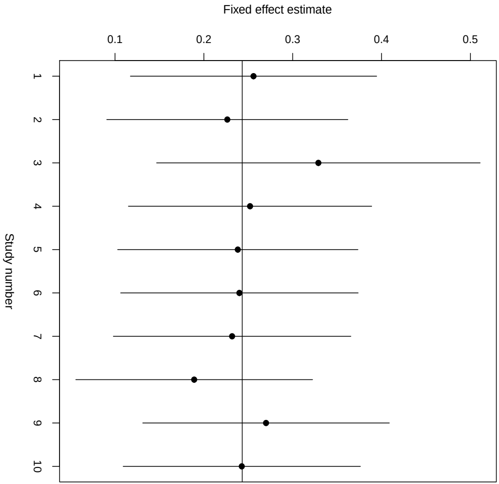
Fluoride toothpaste for preventing dental caries data



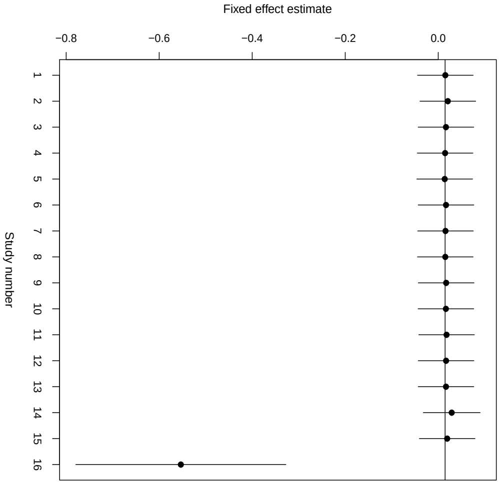
Related work

- Tobias (Stata Tech. Bull., 1999).
- Lee and Thompson (Stat. Med, 2008).
- Baker and Jackson (Health Care Manag. Science, 2008).
- Viechtbauer and Cheung (Research Synth. Meth., 2010).

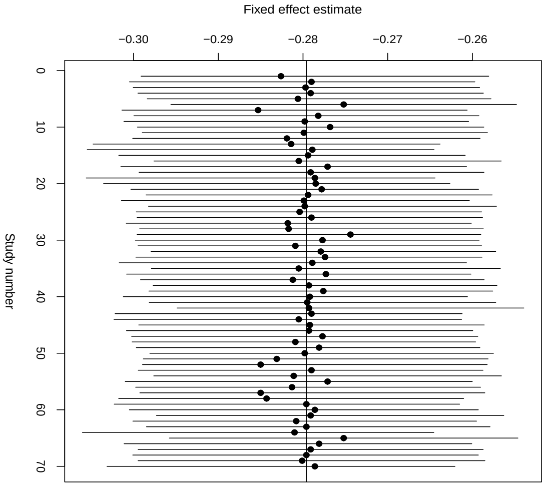
The CDP-choline data



Intravenous magnesium in acute myocardial infarction data



Fluoride toothpaste for preventing dental caries data



Random effects model for meta-analytic data

$$\begin{aligned} \boldsymbol{y} &= \mu \mathbf{1}_n + \boldsymbol{u} + \boldsymbol{e}\,, \\ &\sim N\left(\mu \mathbf{1}_n, \sigma_s^2 \boldsymbol{I}_n + \boldsymbol{R}\right), \end{aligned} \tag{1}$$

where: $\boldsymbol{R} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$.

A RVSOM for the *i*th observation in the random effects model for meta-analytic data

$$\begin{aligned} \boldsymbol{y} &= \mu \mathbf{1}_n + \delta_i \boldsymbol{d}_i + \boldsymbol{u} + \boldsymbol{e} \\ &\sim N\left(\mu \mathbf{1}_n, \omega_i^2 \boldsymbol{d}_i \boldsymbol{d}_i' + \sigma_s^2 \boldsymbol{I}_n + \boldsymbol{R}\right), \end{aligned} \tag{2}$$

where: $\boldsymbol{d}_i$ is a unit vector of length

$\delta_i \sim N\left(0, \omega_i^2\right)$.

Random effects model for meta-analytic data

$$\begin{aligned} \boldsymbol{y} &= \mu \mathbf{1}_n + \boldsymbol{u} + \boldsymbol{e} \, , \qquad (1) \\ &\sim N\left(\mu \mathbf{1}_n, \sigma_s^2 \boldsymbol{I}_n + \boldsymbol{R}\right) , \end{aligned}$$

where: $\boldsymbol{R} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$.

A RVSOM for the *i*th observation in the random effects model for meta-analytic data

$$\begin{aligned} \boldsymbol{y} &= \mu \mathbf{1}_n + \delta_i \boldsymbol{d}_i + \boldsymbol{u} + \boldsymbol{e} \qquad (2) \\ &\sim N\left(\mu \mathbf{1}_n, \omega_i^2 \boldsymbol{d}_i \boldsymbol{d}_i' + \sigma_s^2 \boldsymbol{I}_n + \boldsymbol{R}\right) , \end{aligned}$$

where: $\boldsymbol{d}_i$ is a unit vector of length

$\delta_i \sim N\left(0, \omega_i^2\right)$.

Variance parameter estimation and inference in a under the RVSOM

- Estimation: Variance parameters ω_i^2, σ^2 are estimated iteratively using REML.

- Inference: $H_0 : \omega_i^2 = 0 \quad vs \quad H_A : \omega_i^2 > 0$
  - Likelihood ratio test statistic (LRT):

$$LRT_i = \begin{cases} -2\left\{RL(\hat{\sigma}_s^2; \mathbf{y}) - RL_{(i)}(\hat{\omega}_i^2, \hat{\sigma}_s^2; \mathbf{y})\right\} & \hat{\omega}_i^2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

  - Sampling distribution for LRT_i is obtained using a parametric bootstrap procedure (which also handles the multiple testing problem).

Variance parameter estimation and inference in a under the RVSOM

- Estimation: Variance parameters ω_i^2, σ^2 are estimated iteratively using REML.
- Inference: $H_0 : \omega_i^2 = 0 \quad vs \quad H_A : \omega_i^2 > 0$
  - Likelihood ratio test statistic (LRT):

$$LRT_i = \begin{cases} -2\left\{RL(\hat{\sigma}_s^2; \mathbf{y}) - RL_{(i)}(\hat{\omega}_i^2, \hat{\sigma}_s^2; \mathbf{y})\right\} & \hat{\omega}_i^2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

  - Sampling distribution for LRT_i is obtained using a parametric bootstrap procedure (which also handles the multiple testing problem).

Variance parameter estimation and inference in a under the RVSOM

- Estimation: Variance parameters ω_i^2, σ^2 are estimated iteratively using REML.

- Inference: $H_0 : \omega_i^2 = 0 \quad vs \quad H_A : \omega_i^2 > 0$
    - Likelihood ratio test statistic (LRT):

$$LRT_i = \begin{cases} -2\left\{RL(\hat{\sigma}_s^2; \boldsymbol{y}) - RL_{(i)}(\hat{\omega}_i^2, \hat{\sigma}_s^2; \boldsymbol{y})\right\} & \hat{\omega}_i^2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

    - Sampling distribution for LRT_i is obtained using a parametric bootstrap procedure (which also handles the multiple testing problem).

Variance parameter estimation and inference in a under the RVSOM

- Estimation: Variance parameters ω_i^2, σ^2 are estimated iteratively using REML.
- Inference: $H_0 : \omega_i^2 = 0 \quad vs \quad H_A : \omega_i^2 > 0$
  - Likelihood ratio test statistic (LRT):

$$LRT_i = \begin{cases} -2\left\{RL(\hat{\sigma}_s^2; \mathbf{y}) - RL_{(i)}(\hat{\omega}_i^2, \hat{\sigma}_s^2; \mathbf{y})\right\} & \hat{\omega}_i^2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

  - Sampling distribution for LRT_i is obtained using a parametric bootstrap procedure (which also handles the multiple testing problem).

Sampling distributions of LRTs and multiple testing

Step 1: Fit the null model defined by (1) to the data to obtain estimates $\hat{\mu}$ and $\hat{\sigma}_s^2$.

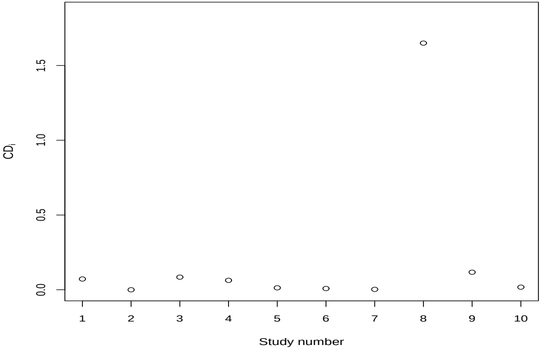
Step 2a: Generate a new data vector: $\boldsymbol{y}^* = \hat{\mu}\mathbf{1}_n + \boldsymbol{u}^* + \boldsymbol{e}^*$.

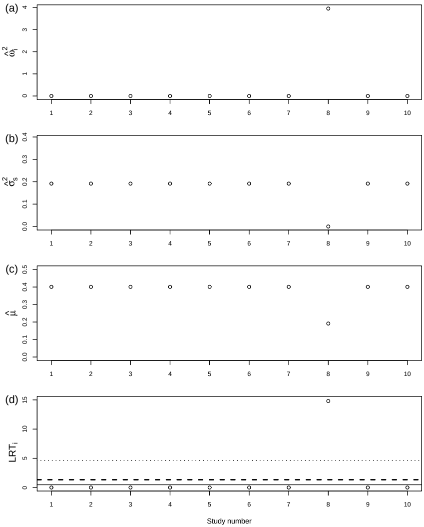
Step 2b: Compute the likelihood ratio test statistics $LRT_i, i = 1, \ldots, n$, by fitting the RVSOM to the simulated data $\boldsymbol{y}^*$ for each observation in turn and compute and save the order statistics of the set $\{LRT_i; i = 1 \ldots n\}$.

Step 3: Repeat steps 2a and 2b R times, for R reasonably large, for example $R = 5000$. This generates an empirical distribution of size R for each order statistic.

Step 4: Calculate the $100(1-\alpha)$th percentile for each order statistic for the required significance level α.

The CDP-choline data




(a)
$\hat{\omega}_i^2$
(b)
$\hat{\sigma}_s^2$
(c)
$\hat{\mu}$
(d)
LRT_i
Study number

The CDP-choline data

	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	0.401	[0.08;0.72]	0.191	[0.058;0.324]
σ_s^2	0.192	-	6.4×10^{-8}	-
ω_8^2	-	-	3.951	-

Under case-deletion: $\hat{\mu} = 0.189$; 95% CI: $[0.056; 0.322]$.

$\sigma_s^2 = 6.4 \times 10^{-8}$

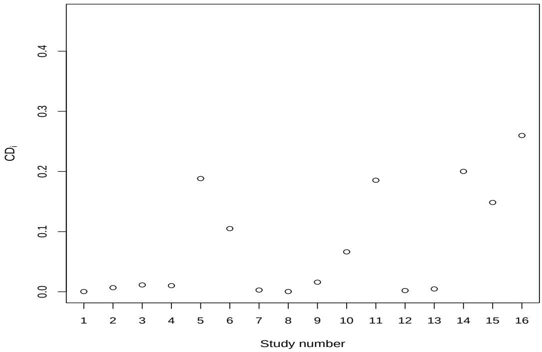
The CDP-choline data

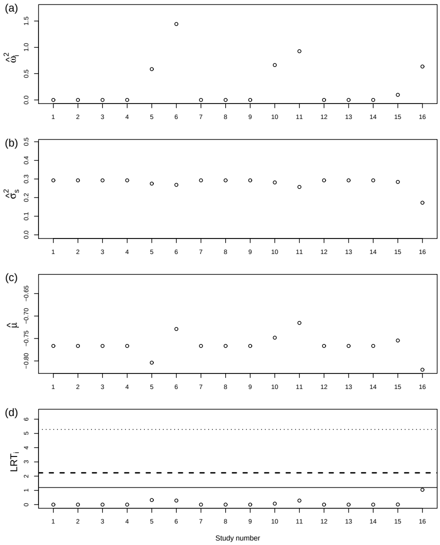
	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	0.401	[0.08;0.72]	0.191	[0.058;0.324]
σ_s^2	0.192	-	6.4×10^{-8}	-
ω_8^2	-	-	3.951	-

Under case-deletion: $\hat{\mu} = 0.189$; 95% CI: $[0.056; 0.322]$.

$$\sigma_s^2 = 6.4 \times 10^{-8}$$

Intravenous magnesium in acute myocardial infarction data




(a)
(b)
(c)
(d)
LRT_i
Study number

Intravenous magnesium in acute myocardial infarction data

	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	-0.766	[-1.18;-0.35]	-0.820	[-1.21;-0.43]
σ_s^2	0.293	-	0.172	-
ω_{16}^2	-	-	0.695	-

Under case-deletion: $\hat{\mu} = -0.875$; 95% CI: $[-1.28; -0.47]$.

$\sigma_s^2 = 0.191$

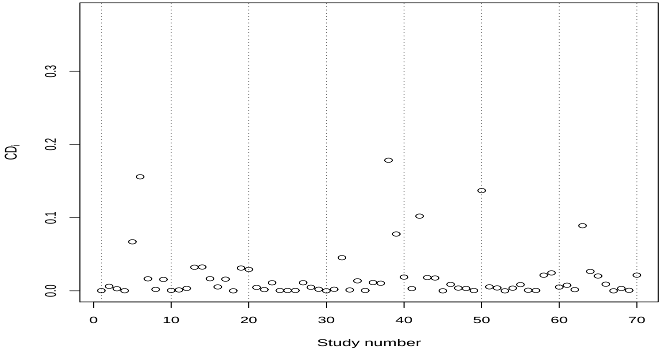
Intravenous magnesium in acute myocardial infarction data

	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	-0.766	[-1.18;-0.35]	-0.820	[-1.21;-0.43]
σ_s^2	0.293	-	0.172	-
ω_{16}^2	-	-	0.695	-

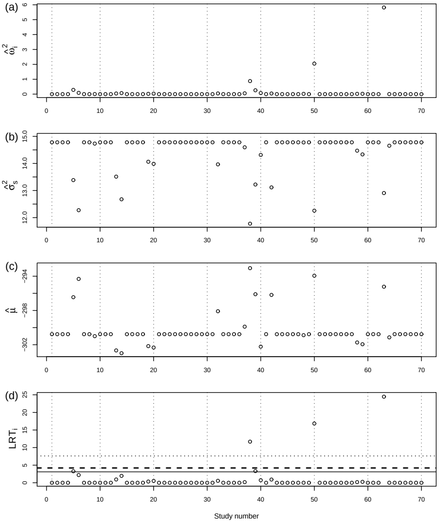
Under case-deletion: $\hat{\mu} = -0.875$; 95% CI: $[-1.28; -0.47]$.

$$\sigma_s^2 = 0.191$$

Fluoride toothpaste for preventing dental caries data



Fluoride toothpaste for preventing dental caries data



Fluoride toothpaste for preventing dental caries data

	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	-0.3008	[-0.33;-0.27]	-0.284	[-0.32;-0.25]
σ_s^2	0.015	-	0.009	-
ω_{38}^2	-	-	0.897	-
ω_{50}^2	-	-	2.082	-
ω_{63}^2	-	-	5.879	-

Under case-deletion: $\hat{\mu} = -0.283$; 95% CI: $[-0.31; -0.25]$.

$\sigma_s^2 = 0.009$

Fluoride toothpaste for preventing dental caries data

	M_0		M_1	
Para.	Est.	95% CI	Est.	95% CI
μ	-0.3008	[-0.33;-0.27]	-0.284	[-0.32;-0.25]
σ_s^2	0.015	-	0.009	-
ω_{38}^2	-	-	0.897	-
ω_{50}^2	-	-	2.082	-
ω_{63}^2	-	-	5.879	-

Under case-deletion: $\hat{\mu} = -0.283$; 95% CI: $[-0.31; -0.25]$.

$$\sigma_s^2 = 0.009$$

Summary and conclusions

- A RVSOM downweights outliers but does not eliminate them from the analysis.
- The LRT gives an objective measure for detecting outliers in meta-analytic data.
- Variance shift outlier model under fixed effects model in meta-analysis, i.e. VSOM version for meta-analytic model with no heterogeneity component.
- RVSOM may not be appropriate when there are *many* outliers are detected.
  - Use heavy-tailed distributions for the random effect (Baker and Jackson, 2008).
- Computation of thresholds for LRTs.

Summary and conclusions

- A RVSOM downweights outliers but does not eliminate them from the analysis.
- The LRT gives an objective measure for detecting outliers in meta-analytic data.
- Variance shift outlier model under fixed effects model in meta-analysis, i.e. VSOM version for meta-analytic model with no heterogeneity component.
- RVSOM may not be appropriate when there are *many* outliers are detected.
  - Use heavy-tailed distributions for the random effect (Baker and Jackson, 2008).
- Computation of thresholds for LRTs.

Summary and conclusions

- A RVSOM downweights outliers but does not eliminate them from the analysis.
- The LRT gives an objective measure for detecting outliers in meta-analytic data.
- Variance shift outlier model under fixed effects model in meta-analysis, i.e. VSOM version for meta-analytic model with no heterogeneity component.
- RVSOM may not be appropriate when there are *many* outliers are detected.
  - Use heavy-tailed distributions for the random effect (Baker and Jackson, 2008).
- Computation of thresholds for LRTs.

Summary and conclusions

- A RVSOM downweights outliers but does not eliminate them from the analysis.
- The LRT gives an objective measure for detecting outliers in meta-analytic data.
- Variance shift outlier model under fixed effects model in meta-analysis, i.e. VSOM version for meta-analytic model with no heterogeneity component.
- RVSOM may not be appropriate when there are *many* outliers are detected.
    - Use heavy-tailed distributions for the random effect (Baker and Jackson, 2008).
- Computation of thresholds for LRTs.

Summary and conclusions

- A RVSOM downweights outliers but does not eliminate them from the analysis.
- The LRT gives an objective measure for detecting outliers in meta-analytic data.
- Variance shift outlier model under fixed effects model in meta-analysis, i.e. VSOM version for meta-analytic model with no heterogeneity component.
- RVSOM may not be appropriate when there are *many* outliers are detected.
  - Use heavy-tailed distributions for the random effect (Baker and Jackson, 2008).
- Computation of thresholds for LRTs.

Acknowledgements

- National Research Foundation
- University of Cape Town
- MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK
- Rothamsted Research, UK