Bimodality of the likelihood function of the random-effects model in meta-analysis

Society for Research Synthesis Methodology
14th Annual Meeting
Chicago, IL
July 22-24, 2019

Wolfgang Viechtbauer
Maastricht University

2019-07-23

Random-Effects Model

- let y_i be an estimate of some underlying parameter θ_i with (approximately) known sampling variance v_i
- assume $y_i|\theta_i \sim \mathrm{N}(\theta_i, v_i)$ (at least approximately)
- have $i, \ldots, k$ studies, each providing such an estimate
- assume $\theta_i \sim \mathrm{N}(\mu, \tau^2)$
- want to estimate μ and τ^2

Random-Effects Model

- random-effects model:

$$y_i = \mu + u_i + e_i$$

where $u_i \sim \mathrm{N}(0, \tau^2)$ and $e_i \sim \mathrm{N}(0, v_i)$

- assume $\mathrm{Cov}[u_i, u_{i'}] = \mathrm{Cov}[e_i, e_{i'}] = \mathrm{Cov}[u_i, e_i] = 0$

$$y_i \sim \mathrm{N}(\mu, \tau^2 + v_i)$$

Example: BCG Vaccine

- meta-analysis by Colditz et al. (1994) on the effectiveness of the Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis
- each included study provides the proportion of TB positive cases in the vaccinated and non-vaccinated group

Results from One Trial

	Positive	Negative	Total
Vaccinated	4	119	123
Not Vaccinated	11	128	139

$$p_T = 4/123 = .0325$$

$$p_C = 11/139 = .0791$$

$$RR = \frac{4/123}{11/139} = .41$$

$$y = \ln\left[\frac{4/123}{11/139}\right] = -.89$$

$$v = \frac{1}{4} - \frac{1}{123} + \frac{1}{11} - \frac{1}{139} = .326$$

Results from 13 Trials

trial	author	year	tpos	tneg	cpos	cneg	yi	vi
1	Aronson	1948	4	119	11	128	-0.89	0.326
2	Ferguson & Simes	1949	6	300	29	274	-1.59	0.195
3	Rosenthal et al	1960	3	228	11	209	-1.35	0.415
4	Hart & Sutherland	1977	62	13536	248	12619	-1.44	0.020
5	Frimodt-Moller et al	1973	33	5036	47	5761	-0.22	0.051
6	Stein & Aronson	1953	180	1361	372	1079	-0.79	0.007
7	Vandiviere et al	1973	8	2537	10	619	-1.62	0.223
8	TPT Madras	1980	505	87886	499	87892	0.01	0.004
9	Coetzee & Berjak	1968	29	7470	45	7232	-0.47	0.056
10	Rosenthal et al	1961	17	1699	65	1600	-1.37	0.073
11	Comstock et al	1974	186	50448	141	27197	-0.34	0.012
12	Comstock & Webster	1969	5	2493	3	2338	0.45	0.533
13	Comstock et al	1976	27	16886	29	17825	-0.02	0.071

Maximum Likelihood Estimation

- log likelihood function (Hardy & Thompson, 1996):

$$ll(\mu, \tau^2) = -\frac{1}{2}\sum_{i=1}^{k} \ln(\tau^2 + v_i) - \frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - \mu)^2}{\tau^2 + v_i}$$

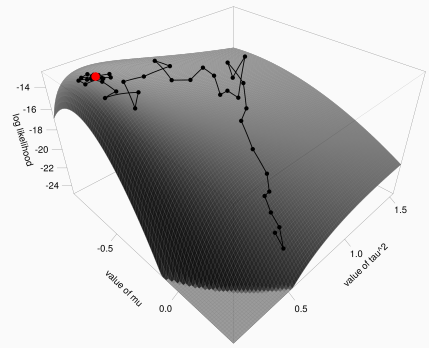
- MLEs of μ and τ^2:

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i}$$

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2 \left((y_i - \mu)^2 - v_i\right)}{\sum_{i=1}^{k} w_i^2}$$

  where $w_i = 1/(\tau^2 + v_i)$
- can use standard optimization techniques to obtain the MLEs (e.g., quasi-Newton methods, Nelder-Mead, ...)

Log Likelihood Surface with Optimization Trace



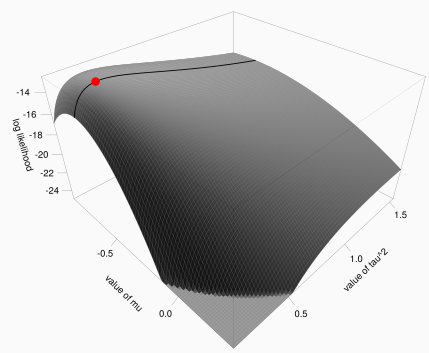
Profiled Log Likelihood

- given τ^2, the MLE of μ is $\hat{\mu}$ as defined above
- can reduce this to a one-dimension optimization problem
- profiled log likelihood function:

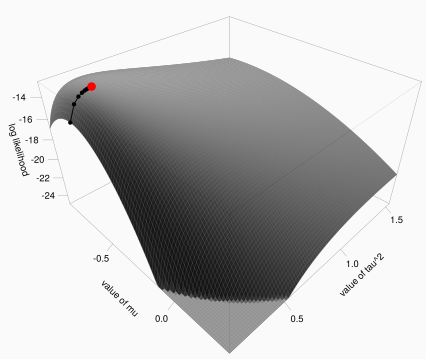
$$ll_P(\tau^2) = -\frac{1}{2}\sum_{i=1}^{k} \ln(\tau^2 + v_i) - \frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - \hat{\mu})^2}{\tau^2 + v_i}$$

- now only need to optimize over τ^2

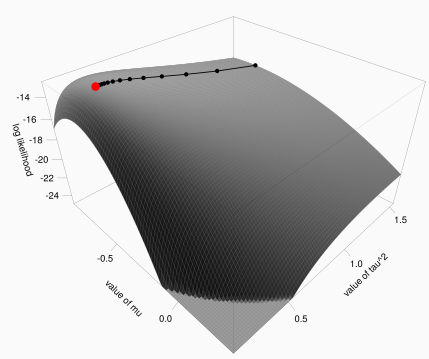
Profiled Log Likelihood



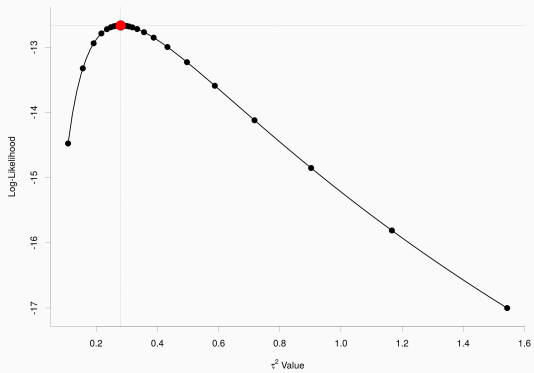
Profiled Log Likelihood with Optimization Trace



Profiled Log Likelihood with Optimization Trace



Profile of the Log Likelihood Function



Restricted Maximum Likelihood Estimation

- restricted log likelihood function (Viechtbauer, 2005):

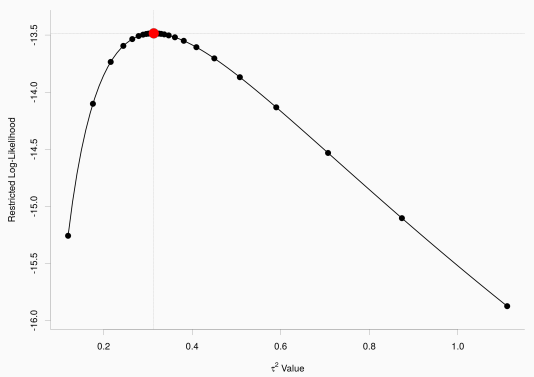
$$ll_R(\tau^2) = -\frac{1}{2}\sum_{i=1}^{k}\ln(\tau^2 + v_i) - \frac{1}{2}\ln\sum_{i=1}^{k}\frac{1}{\tau^2 + v_i} - \frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - \hat{\mu})^2}{\tau^2 + v_i}$$

- REML estimator of τ^2:

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2\left((y_i - \hat{\mu})^2 - v_i\right)}{\sum_{i=1}^{k} w_i^2} + \frac{1}{\sum_{i=1}^{k} w_i}$$

  where $w_i = 1/(\tau^2 + v_i)$ and $\hat{\mu}$ as defined previously
- one-dimensional optimization problem

Profile of the Restricted Log Likelihood Function



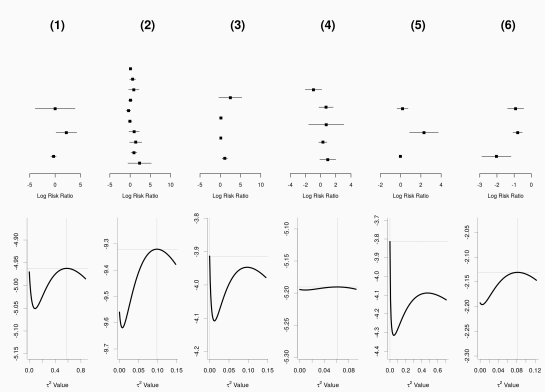
Can There Be Local Optima?

- checked for non-unimodality by profiling the (restricted) log likelihood function and searching for multiple peaks in 1991 meta-analyses from Cochrane Database of Systematic Reviews
- analyses conducted using odds/risk ratios and risk differences
- restricted to meta-analyses with at least three studies
- will focus here on results when using risk ratios as the outcome measure and when using ML estimation

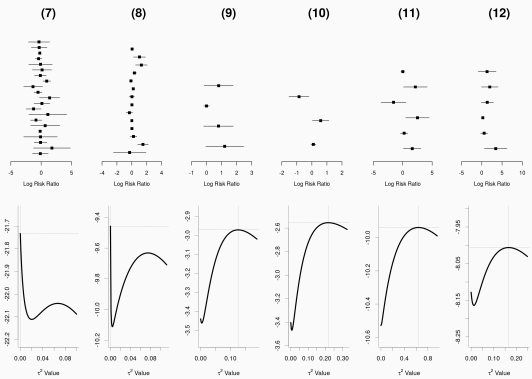
Results

- found 12 cases where the log likelihood function was bimodal
- one peak at $\hat{\tau}^2 = 0$ and another at $\hat{\tau}^2 > 0$
- either one of the two peaks could be the ML estimate of τ^2

Results



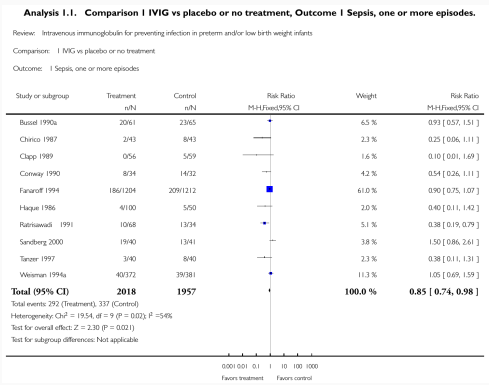
Results



Illustrative Case

- review on intravenous immunoglobulin for preventing infection in preterm and/or low birth weight infants (Ohlsson & Lacy, 2013)
- ten studies provided data on the occurrence of sepsis
- authors used the Mantel-Haenszel method for the meta-analysis

Illustrative Case: Cochrane Review Results



Analysis 1.1. Comparison 1 IVIG vs placebo or no treatment, Outcome 1 Sepsis, one or more episodes.

Review: Intravenous immunoglobulin for preventing infection in preterm and/or low birth weight infants

Comparison: 1 IVIG vs placebo or no treatment

Outcome: 1 Sepsis, one or more episodes

Study or subgroup	Treatment n/N	Control n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Bussel 1990a	20/61	23/65		6.5 %	0.93 [0.57, 1.51]
Chirico 1987	2/43	8/43		2.3 %	0.25 [0.06, 1.11]
Clapp 1989	0/56	5/59		1.6 %	0.10 [0.01, 1.69]
Conway 1990	8/34	14/32		4.2 %	0.54 [0.26, 1.11]
Fanaroff 1994	186/1204	209/1212		61.0 %	0.90 [0.75, 1.07]
Haque 1986	4/100	5/50		2.0 %	0.40 [0.11, 1.42]
Ratrisawadi 1991	10/68	13/34		5.1 %	0.38 [0.19, 0.79]
Sandberg 2000	19/40	13/41		3.8 %	1.50 [0.86, 2.61]
Tanzer 1997	3/40	8/40		2.3 %	0.38 [0.11, 1.31]
Weisman 1994a	40/372	39/381		11.3 %	1.05 [0.69, 1.59]
Total (95% CI)	**2018**	**1957**		**100.0 %**	**0.85 [0.74, 0.98]**

Total events: 292 (Treatment), 337 (Control)
Heterogeneity: $Chi^2 = 19.54$, df = 9 (P = 0.02); $I^2 = 54\%$
Test for overall effect: Z = 2.30 (P = 0.021)
Test for subgroup differences: Not applicable

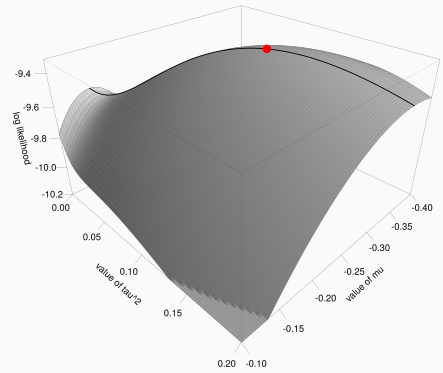
Illustrative Case: Mantel-Haenszel Method

```
Fixed-Effects Model (k = 10)

I^2 (total heterogeneity / total variability):  53.95%
H^2 (total variability / sampling variability): 2.17

Test for Heterogeneity:
Q(df = 9) = 19.54, p-val = 0.02

Model Results (log scale):

estimate    se   zval  pval  ci.lb  ci.ub
   -0.17  0.07  -2.30  0.02  -0.31  -0.02

Model Results (RR scale):

estimate  ci.lb  ci.ub
    0.85   0.74   0.98
```

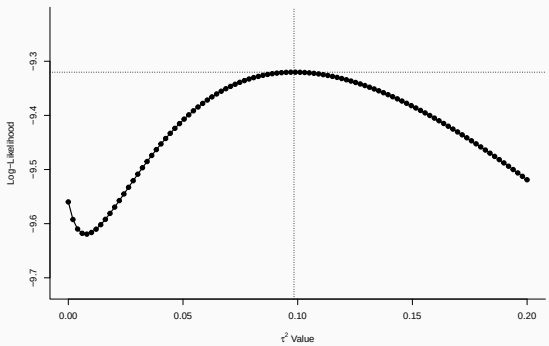

Illustrative Case: Random-Effects Model ($\hat{\tau}^2 = 0$)

```
Random-Effects Model (k = 10; tau^2 estimator: ML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.01)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00

Test for Heterogeneity:
Q(df = 9) = 19.47, p-val = 0.02

Model Results:

estimate    se   zval  pval  ci.lb  ci.ub
   -0.15  0.07  -2.02  0.04  -0.29  -0.00


 pred ci.lb ci.ub cr.lb cr.ub
 0.86  0.75  1.00  0.75  1.00
```


Illustrative Case: Log Likelihood Surface



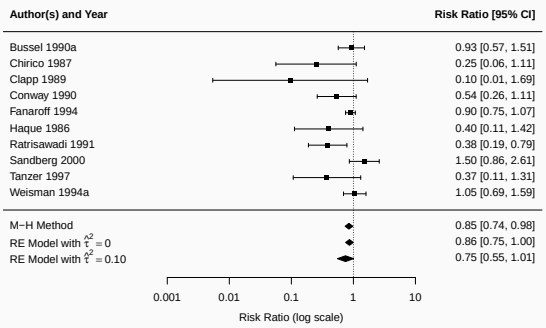
Illustrative Case: Profiled Log Likelihood



Illustrative Case: Random-Effects Model ($\hat{\tau}^2 = 0.10$)

```
Random-Effects Model (k = 10; tau^2 estimator: ML)

tau^2 (estimated amount of total heterogeneity): 0.10 (SE = 0.09)
tau (square root of estimated tau^2 value):      0.31
I^2 (total heterogeneity / total variability):   55.08%
H^2 (total variability / sampling variability):  2.23

Test for Heterogeneity:
Q(df = 9) = 19.47, p-val = 0.02

Model Results:

estimate    se   zval  pval  ci.lb  ci.ub
   -0.29  0.15  -1.91  0.06  -0.59   0.01


 pred ci.lb ci.ub cr.lb cr.ub
 0.75  0.55  1.01  0.38  1.48
```


Illustrative Case: Forest Plot



Conclusions

- local maxima in likelihood function can occur for various types of models (e.g., Henn & Hodges, 2014; Hoeschele, 1989)
- this can also happen in the meta-analytic random-effects model (for both ML and REML estimation)
- estimates obtained using iterative algorithms may then not reflect the true ML/REML estimates, which can lead to over- or underestimation of the amount of heterogeneity
- is it time to **panic**? probably not ...
- still good to check for this (construct a profile likelihood plot!)
- or fit the model with a low and a high starting value for $\hat{\tau}^2$
- can this also happen in other meta-analytic models? **yes!**

Acknowledgments

- thanks to **Dan Jackson** and **Rebecca Turner** for their collaboration on this topic and Rebecca for providing the Cochrane dataset

References

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature. *Journal of the American Medical Association*, *271*(9), 698–702.

Hardy, R. J., & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, *15*(6), 619–629.

Henn, L., & Hodges, J. S. (2014). Multiple local maxima in restricted likelihoods and posterior distributions for mixed linear models. *International Statistical Review*, *82*(1), 90–105.

Hoeschele, I. (1989). A note on local maxima in maximum likelihood, restricted maximum likelihood, and Bayesian estimation of variance components. *Journal of Statistical Computation & Simulation*, *33*(3), 149–160.

Ohlsson, A., & Lacy, J. B. (2013). Intravenous immunoglobulin for preventing infection in preterm and/or low birth weight infants. *Cochrane Database of Systematic Reviews*, (*7*), CD000361.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293.