A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio

Dan Jackson, Martin Law, Theo Stijnen, Wolfgang Viechtbauer, Ian White

Ian White <ian.white@ucl.ac.uk>

MRC Clinical Trials Unit at UCL

Cochrane Colloquium, Statistical Methods Group

17 Sep 2018

Background

- Setting:
  - meta-analysis (MA) of binary outcome presented as one 2x2 table for each study
  - use random-effects model (allow for heterogeneity)
- Two-stage inverse-variance estimation is easy
  - estimate log odds ratio y_i & its SE s_i in each study
  - estimate heterogeneity variance τ^2
    - DerSimonian & Laird (DL) is common
    - REML is better
  - average the log odds ratios, weighted by their inverse variance = $1/\left(s_i^2 + \tau^2\right)$

Background

- What's wrong with two-stage inverse-variance estimation (DL or REML)? Model 1
  - poor approximation with sparse data (i.e. small counts in the 2x2 tables)

vs.

- Common-effect MA: Mantel-Haenszel or logistic regression are easy one-stage approaches & better than two-stage inverse variance method
- Random-effects MA: random-effects logistic regression is the obvious one-stage approach
  - generalised linear mixed model

Models 2-7

Models 2-7 (random-effects logistic regression)

Model for log odds in arm j of study i: all with study-specific treatment effect $\theta_i \sim N(\theta, \tau^2)$

Generalised linear mixed models

2. Fixed effects of study = control log odds — $\gamma_i + j\theta_i$
3. Random effects of study = control log odds — M2 & $\gamma_i \sim N(\gamma, \sigma^2)$
4. Fixed effects of study = average log odds — $\gamma_i + (j - 1/2)\theta_i$
5. Random effects of study = average log odds — M4 & $\gamma_i \sim N(\gamma, \sigma^2)$

   All above: effects of study uncorrelated with random treatment effects

6. Random effects of study correlated with random treatment effects — M3/5 & γ_i correlated with θ_i
7. Hypergeometric model for one count given table margins

Issues

- Fixed (M2,4) vs. random (M3,5,6) effect of study:
  - fixed effects uses “too many parameters” & may underestimate variances
  - random effect allows use of between-study information
- Coding treatment as $0/1$ (M2,3) or ± 0.5 (M4,5)
  - seems trivial
- Treatment effect associated with effect of study (M6)
- Use of computationally intensive hypergeometric likelihood (M7)

All models fitted in R using Wolfgang Viechtbauer’s metafor

Simulation study: data generating mechanisms

- Base case:
  - 10 studies
  - moderate heterogeneity ($\tau^2 = 0.024$ giving $I^2 = 0.3$)
  - event fractions around 0.2
  - no treatment effect ($\theta = 0$)
- Key variants:
  - no/large heterogeneity
    - settings 2, 3, 15
  - sparser data (fewer events)
    - settings 7-9
  - allocation ratio related to control event fraction
    - setting 12

Simulation study: key finding

- Model 2 underestimates heterogeneity variance
  - e.g. base case: true value 0.024 but mean estimate 0.006
  - with consequent loss of coverage (88% cf 95% nominal and ~93% other methods)
- But model 4 doesn't
  - recall: M2 codes treatment as $0/1$, M4 as ± 0.5
- This really surprised us
- Fixed effects of study means "too many parameters" which could be expected to lead to variance underestimation *in both models*

Simulation study: other findings

- Estimation failure: very rare in models 1-6
- Quite common in model 7. After fine-tuning estimation methods, we got it down to
  - 1.4% of datasets not estimating overall mean θ
  - 2.8% of datasets not estimating its SE
  - a further 0.5% of datasets giving implausible SEs

Simulation study: other findings

Bias in estimating θ: absent except for

- Sparse data (setting 8)
  - M1/DL gave bias 0.02
  - other models didn't remove bias
- Between-study information (setting 12)
  - M3,5,6 (random effects of study) gave bias 0.02

Bias in estimating τ^2:

- Always present when $\tau = 0$
- M1/DL negatively biased when τ is large
  - probably because SE~estimate
- M3,4,5,7 slightly negatively biased when τ is moderate
- M6 slightly positively biased when τ is moderate

Simulation study: other findings

- Precision:
  - all models have similar precision for θ
  - M3-7 are more precise than M1 for τ^2
- Coverage of M1,3-7: mostly around 93%, some variations e.g. higher in sparser/less heterogeneous data

Conclusions

1. One-stage methods (M2-7) are not uniformly superior to two-stage methods (M1)
2. Model 2 (trt=0/1) should not be used
3. Among the models with random study effects, Model 6 (correlated effects) is probably better than 3 and 5
4. Model 7 can be computationally challenging / unstable
5. Models 1, 4, 6 and 7 are all good candidates
   – avoid model 1 with sparse data; 6 with varying allocation ratios; 7 with unsparse data
6. Different models should be compared in sensitivity analysis

For discussion:

- Could we develop criteria for moving from M1 to M4/6/7 & from M4/6/7 to a still better model [Bayes?]

Extra slides

TABLE 5 Simulation study design. All 15 settings were performed with $\theta = 0$ (results shown in tables in the main paper) and $\theta = \log(2)$ (results shown in tables in the Supporting Information). One thousand simulated datasets were produced in each setting and true effect, so that in total, 30 000 datasets were produced. A different random seed was used for each setting and value of θ. All models were applied to the same datasets. This table provides an outline of the simulation study design; see Section 6 for full details. Departures from the defaults are shown in bold

Setting	k	τ^2	Treatment	Control	Baseline probability	Correct models
1	10	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
2	10	**0**	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
3	10	**0.168**	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
4	**3**	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
5	**5**	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
6	**20**	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
7	10	0.024	$N \sim U(\mathbf{10, 100})$	N	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
8	10	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(\mathbf{0.05}), 0.3^2)$	2, 3, 6
9	10	0.024	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(\mathbf{0.01}), 0.3^2)$	2, 3, 6
10	10	0.024	$N \sim U(50, 500)$	**N/2**	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
11	10	0.024	$N \sim U(50, 500)$	**N/2 and N**	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	2, 3, 6
12	10	0.024	$N \sim U(50, 500)$	**N/2 and N (NR)**	$LO_c \sim N(\text{logit}(0.2), 0.3^2)$	None
13	10	0.024	$N \sim U(50, 500)$	N	$\mathbf{P_c \sim U(0.1, 0.3)}$	2
14	10	0.024	$N \sim U(50, 500)$	N	$\mathbf{LO_a} \sim N(\text{logit}(0.2), 0.3^2)$	4, 5, 6
15	10	**2**	$N \sim U(50, 500)$	N	$LO_c \sim N(\text{logit}(\mathbf{0.5}), 0.3^2)$	2, 3, 6

TABLE 6 Simulation study results. The top half of the table shows the mean estimate of the summary log odds ratio θ; empirical standard errors of the estimates are shown in parentheses. Monte Carlo standard errors of the mean estimates can be obtained as the empirical standard errors divided by the square root of 1000. The bottom half of the table shows the mean estimate of τ^2. The true value $\theta = 0$; results for $\theta = \log(2)$ are shown in the Supporting Information. Model 7* indicates that inferences for model 7 have been supplemented with results from the Peto approximation described in Section 3.7.3, as explained in Section 6.3.3

Setting	Model 1 (D & L)	Model 1 (REML)	Model 2 (ML)	Model 3 (ML)	Model 4 (ML)	Model 5 (ML)	Model 6 (ML)	Model 7* (ML)
1 ($\theta = 0$)	−0.001(0.088)	−0.002(0.088)	−0.002(0.089)	−0.006(0.088)	−0.001(0.088)	−0.001(0.088)	−0.006(0.088)	−0.002(0.088)
2 ($\theta = 0$)	−0.002(0.070)	−0.002(0.070)	−0.002(0.071)	−0.004(0.071)	−0.002(0.071)	−0.002(0.071)	−0.002(0.072)	−0.002(0.071)
3 ($\theta = 0$)	0.010(0.153)	0.009(0.154)	0.007(0.156)	−0.002(0.154)	0.010(0.154)	0.013(0.154)	−0.002(0.155)	0.009(0.155)
4 ($\theta = 0$)	0.002(0.159)	0.001(0.159)	0.003(0.157)	−0.003(0.159)	0.003(0.157)	0.003(0.157)	0.000(0.162)	0.003(0.157)
5 ($\theta = 0$)	−0.002(0.119)	−0.002(0.119)	−0.001(0.119)	−0.005(0.119)	−0.001(0.119)	−0.001(0.119)	−0.004(0.121)	−0.001(0.119)
6 ($\theta = 0$)	0.007(0.060)	0.007(0.060)	0.008(0.060)	0.002(0.060)	0.007(0.060)	0.007(0.060)	0.003(0.060)	0.007(0.060)
7 ($\theta = 0$)	0.010(0.163)	0.011(0.163)	0.011(0.168)	0.001(0.169)	0.009(0.169)	0.010(0.166)	0.004(0.171)	0.009(0.167)
8 ($\theta = 0$)	0.020(0.137)	0.020(0.137)	0.021(0.141)	0.010(0.142)	0.021(0.141)	0.021(0.141)	0.013(0.144)	0.021(0.141)
9 ($\theta = 0$)	0.003(0.241)	0.002(0.241)	−0.004(0.292)	−0.035(0.305)	−0.004(0.289)	−0.001(0.285)	−0.015(0.324)	−0.006(0.299)
10 ($\theta = 0$)	−0.006(0.099)	−0.007(0.099)	0.000(0.100)	−0.006(0.099)	−0.001(0.100)	−0.001(0.099)	−0.005(0.101)	0.000(0.100)
11 ($\theta = 0$)	0.001(0.092)	0.001(0.093)	0.004(0.093)	−0.001(0.092)	0.003(0.093)	0.004(0.092)	0.000(0.093)	0.004(0.093)
12 ($\theta = 0$)	0.002(0.092)	0.002(0.092)	0.006(0.093)	0.017(0.093)	0.005(0.093)	0.017(0.093)	0.017(0.094)	0.005(0.093)
13 ($\theta = 0$)	0.005(0.087)	0.005(0.087)	0.006(0.087)	0.000(0.087)	0.005(0.087)	0.005(0.087)	0.001(0.089)	0.005(0.087)
14 ($\theta = 0$)	−0.002(0.090)	−0.002(0.090)	−0.003(0.091)	−0.006(0.091)	−0.002(0.091)	−0.002(0.090)	−0.002(0.092)	−0.002(0.091)
15 ($\theta = 0$)	0.017(0.439)	0.018(0.443)	0.017(0.451)	0.017(0.451)	0.017(0.446)	0.017(0.442)	0.016(0.451)	0.016(0.448)
1 ($\tau^2 = 0.024$)	0.026(0.029)	0.026(0.031)	0.006(0.016)	0.020(0.024)	0.020(0.026)	0.020(0.026)	0.023(0.026)	0.020(0.026)
2 ($\tau^2 = 0$)	0.008(0.016)	0.007(0.015)	0.001(0.005)	0.005(0.013)	0.005(0.012)	0.005(0.012)	0.009(0.014)	0.005(0.012)
3 ($\tau^2 = 0.168$)	0.160(0.107)	0.165(0.110)	0.115(0.099)	0.147(0.097)	0.144(0.099)	0.143(0.098)	0.148(0.100)	0.145(0.100)
4 ($\tau^2 = 0.024$)	0.041(0.071)	0.043(0.082)	0.009(0.033)	0.021(0.049)	0.019(0.044)	0.018(0.042)	0.033(0.056)	0.018(0.044)
5 ($\tau^2 = 0.024$)	0.030(0.046)	0.031(0.051)	0.008(0.027)	0.021(0.039)	0.019(0.037)	0.018(0.036)	0.028(0.041)	0.019(0.038)
6 ($\tau^2 = 0.024$)	0.026(0.022)	0.025(0.022)	0.005(0.012)	0.023(0.020)	0.023(0.021)	0.023(0.021)	0.024(0.021)	0.023(0.021)
7 ($\tau^2 = 0.024$)	0.048(0.076)	0.045(0.075)	0.004(0.022)	0.036(0.065)	0.039(0.070)	0.038(0.068)	0.062(0.076)	0.039(0.070)
8 ($\tau^2 = 0.024$)	0.036(0.057)	0.034(0.058)	0.003(0.019)	0.028(0.047)	0.029(0.052)	0.030(0.055)	0.047(0.058)	0.031(0.056)
9 ($\tau^2 = 0.024$)	0.024(0.084)	0.026(0.090)	0.013(0.211)	0.077(0.210)	0.077(0.182)	0.094(0.246)	0.148(0.274)	0.130(0.780)
10 ($\tau^2 = 0.024$)	0.028(0.035)	0.027(0.036)	0.001(0.007)	0.021(0.030)	0.017(0.028)	0.022(0.032)	0.027(0.033)	0.021(0.031)
11 ($\tau^2 = 0.024$)	0.030(0.035)	0.029(0.036)	0.004(0.013)	0.022(0.029)	0.021(0.030)	0.023(0.031)	0.027(0.031)	0.023(0.031)
12 ($\tau^2 = 0.024$)	0.029(0.035)	0.029(0.037)	0.004(0.017)	0.023(0.032)	0.021(0.031)	0.023(0.032)	0.027(0.032)	0.023(0.032)
13 ($\tau^2 = 0.024$)	0.029(0.031)	0.028(0.032)	0.007(0.017)	0.022(0.028)	0.022(0.028)	0.022(0.028)	0.026(0.028)	0.022(0.028)
14 ($\tau^2 = 0.024$)	0.029(0.032)	0.028(0.033)	0.007(0.019)	0.018(0.024)	0.022(0.028)	0.022(0.028)	0.025(0.028)	0.022(0.028)

TABLE 7 Simulation study results. Actual coverage probability of 95% confidence intervals for θ. The average model based standard errors, as a percentage of the corresponding empirical standard errors, are shown in parentheses. Model 7* indicates that inferences for model 7 have been supplemented with results from the 'Peto approximation' described in section 3.7.3, as explained in Section 6.3

Setting	Model 1 (D & L)	Model 1 (REML)	Model 2 (ML)	Model 3 (ML)	Model 4 (ML)	Model 5 (ML)	Model 6 (ML)	Model 7* (ML)
1	0.933 (98)	0.931 (98)	0.881 (81)	0.923 (93)	0.925 (94)	0.925 (93)	0.926 (97)	0.920 (94)
2	0.957 (105)	0.957 (104)	0.945 (97)	0.952 (102)	0.953 (102)	0.953 (102)	0.956 (105)	0.953 (103)
3	0.910 (94)	0.913 (95)	0.839 (80)	0.899 (90)	0.892 (90)	0.895 (90)	0.894 (90)	0.895 (91)
4	0.935 (106)	0.937 (106)	0.907 (88)	0.921 (95)	0.921 (95)	0.922 (95)	0.947 (103)	0.922 (95)
5	0.939 (104)	0.937 (104)	0.913 (89)	0.933 (99)	0.932 (97)	0.932 (97)	0.949 (103)	0.929 (98)
6	0.943 (102)	0.947 (102)	0.894 (84)	0.936 (100)	0.943 (99)	0.943 (99)	0.942 (100)	0.941 (103)
7	0.955 (105)	0.957 (105)	0.930 (92)	0.944 (98)	0.943 (100)	0.942 (100)	0.956 (104)	0.944 (100)
8	0.952 (102)	0.950 (102)	0.911 (89)	0.937 (96)	0.937 (97)	0.936 (96)	0.958 (102)	0.937 (97)
9	0.986 (121)	0.986 (121)	0.948 (95)	0.958 (99)	0.955 (102)	0.957 (103)	0.968 (102)	0.961 (101)
10	0.939 (101)	0.944 (100)	0.896 (85)	0.934 (97)	0.923 (94)	0.936 (97)	0.952 (99)	0.935 (97)
11	0.936 (101)	0.933 (101)	0.898 (84)	0.930 (97)	0.923 (96)	0.927 (98)	0.939 (100)	0.925 (98)
12	0.931 (101)	0.933 (101)	0.881 (83)	0.927 (97)	0.922 (95)	0.923 (97)	0.931 (98)	0.926 (97)
13	0.946 (101)	0.943 (101)	0.906 (85)	0.941 (97)	0.935 (97)	0.936 (97)	0.945 (99)	0.936 (98)
14	0.920 (97)	0.917 (96)	0.880 (80)	0.906 (89)	0.911 (92)	0.911 (92)	0.916 (93)	0.906 (92)
15	0.876 (84)	0.915 (98)	0.899 (92)	0.898 (93)	0.900 (92)	0.900 (92)	0.899 (92)	0.901 (94)

Abbreviation: REML, restricted maximum likelihood.