

Supplementary Material for Bayesian model-based outlier detection in network meta-analysis

Silvia Metelli*, Dimitris Mavridis, Perrine Créquit, Anna Chaimani
*email: silvia.metelli@u-paris.fr

1 Complete simulation results

In this section, we report additional results for the extensive simulation study described in Section 5 of the main document. The number of studies per comparison was set to 10 for all comparisons in an ideally balanced design and ranged from 1 to 9 to reflect values more often encountered in practice in three different unbalanced designs. The amount of heterogeneity and number of induced outliers also varied across scenarios. In total, we explored 32 different scenarios. Table 1 summarized all the scenarios explored. We report additional results from our proposed Bayesian diagnostic tools for the remaining scenarios analyzed in the simulation study. Specifically, Table 2, Table 3 and Table 4 respectively report mean Bayes factors and posterior predictive p -value results for the case of a balanced fairly connected network, an unbalanced but well-connected network and finally, an unbalanced and poorly connected network. Then, in Table 5 we report some representative results based on our initial choice of discrepancy measure - which turned out to show poor performance - namely the well-known Gelman's discrepancy measure given by

$$D_i^G = \frac{|y_{i,k} - E(Y_{i,k}|\theta, \tau^2)|^2}{Var(Y_{i,k}|\theta, \tau^2)}.$$

Finally, we report the results obtained from our comparison with the forward search (FS) algorithm (Petropoulou et al., 2021) and the conditional predictive ordinate (CPO) estimated via Integrated Nested Laplace Approximation (INLA) (Gelfand, 1995; Rue et al., 2009) respectively in Figure 1 and Figure 2, where we focus on the exemplificatory case of simulated scenarios with three outliers and fairly connected network (scenarios 21-24). As briefly described in the main paper, outlier detection in network meta-analysis via FS algorithm operates by choosing a small starting subset of studies considered to be outlier-free, called the basic set and sequentially adding the remaining studies according to some measure of how close the study is to the evolving set, under some posited model. Abrupt changes in one or more of the monitored measures may indicate outlyingness. Here, we monitored mean Cook's distance as diagnostic measure: values falling well above the threshold of 1 correspond to all or some of the artificially generated outliers. FS search for NMA is available via the 'NMAoutlier' R package. The CPO expresses the posterior probability of observing the value (or set of values) of y_i when the model is fitted to all data except y_i , with a larger value indicating a better fit of the model to y_i and very low CPO values suggest that y_i is an outlier and an influential point. The CPO is connected with the frequentist studentized residual test for outlier detection, as it is known that data points with large studentized residuals have small CPOs and will be

detected as outliers. For interpretation, it is usual to plot the Inverse-CPOs (ICPOs), where a value larger than 40 can be considered as possible outlier, and higher than 70 as extreme values (Ntzoufras, 2009). For study i , the general formula for CPO can be expressed as

$$CPO_i = P(y_i | \mathbf{y}_{-i}) = \int P(y_i | \mathbf{y}_{-i}, \boldsymbol{\theta}) P(\boldsymbol{\theta} | \mathbf{y}_{-i}) d\boldsymbol{\theta} = \left(\int \frac{1}{P(y_i | \mathbf{y}_{-i}, \boldsymbol{\theta})} P(\boldsymbol{\theta} | \mathbf{y}_{-i}) d\boldsymbol{\theta} \right)^{-1}$$

where \mathbf{y}_{-i} is the vector of observed data after having removed the i -th observation, and $\boldsymbol{\theta}$ represents the parameter collection. Thus, the CPO can be estimated by taking the inverse of the posterior mean of the inverse density function of y_i . In words, CPO estimates the probability of observing y_i in the future, after having already observed \mathbf{y}_{-i} . Estimation of the CPO is available via Integrated Nested Laplace Approximation (INLA), using ‘nmaINLA’ R package.

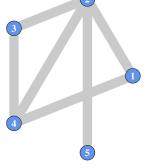
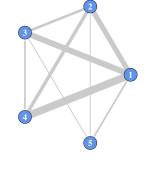
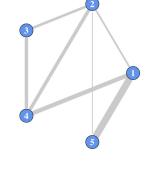
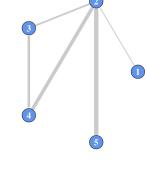
Network geometry	Design	Scenario	Heterogeneity	Number of outliers
 Balanced 10 studies per comparison	1	0	1	
	2	0.032	1	
	3	0.096	1	
	4	0.287	1	
	5	0	3	
	6	0.032	3	
	7	0.096	3	
	8	0.287	3	
 Unbalanced well-connected network	9	0	1	
	10	0.032	1	
	11	0.096	1	
	12	0.287	1	
	13	0	3	
	14	0.032	3	
	15	0.096	3	
	16	0.287	3	
 Unbalanced fairly-connected network	17	0	1	
	18	0.032	1	
	19	0.096	1	
	20	0.287	1	
	21	0	3	
	22	0.032	3	
	23	0.096	3	
	24	0.287	3	
 Unbalanced poorly-connected network	25	0	1	
	26	0.032	1	
	27	0.096	1	
	28	0.287	1	
	29	0	3	
	30	0.032	3	
	31	0.096	3	
	32	0.287	3	

Table 1: Summary of the different scenarios analyzed in the simulation study.

τ^2	N outliers	BF	p^{DL}	p^{DSDO}
0	1	872.6	0.001	0.01
		1135.8	0.001	0.001
	3	128.2	0.001	0.01
		582.5	<0.001	0.01
	1	1102.1	<0.0001	0.005
		458.1	0.001	0.001
0.032	3	840.1	<0.001	0.01
		44.7	0.01	0.01
	1	18.1	0.05	0.05
		11.6	0.01	0.01
	3	7.2	0.01	0.01
		3.6	0.10	0.07
0.096	1	7.3	0.15	0.15
		3.1	0.34	0.12
	3	1.7	0.07	0.11
		0.99	0.12	0.20
	1	580.0	0.001	0.01
		835.4	0.01	0.001
0.032	3	118.2	0.03	0.01
		284.1	0.001	0.01
	1	982.2	0.001	0.001
		777.1	0.01	0.001
	3	540.1	0.02	0.03
		540.1	0.01	0.01
0.096	1	32.1	0.01	0.005
		9.2	0.01	0.01
	3	12.7	0.04	0.05
		2.2	0.10	0.07
	1	2.5	0.01	0.01
		2.5	0.34	0.12
0.287	3	1.3	0.07	0.15
		1.2	0.13	0.20

Table 2: Mean Bayes factors and mean posterior predictive p-values for the induced outliers out of 1000 replicates of simulations for the balanced scenario and fairly connected network, comparing all methods proposed (scenarios 1-8).

τ^2	N outliers	BF	p^{DL}	p^{DSDO}
0	1	580.0	0.001	0.01
		835.4	0.01	0.001
	3	118.2	0.03	0.01
		284.1	0.001	0.01
	1	982.2	0.001	0.001
		777.1	0.01	0.001
0.032	3	540.1	0.02	0.03
		540.1	0.01	0.01
	1	32.1	0.01	0.005
		9.2	0.01	0.01
	3	12.7	0.04	0.05
		2.2	0.10	0.07
0.096	1	2.5	0.01	0.01
		2.5	0.34	0.12
	3	1.3	0.07	0.15
		1.2	0.13	0.20

Table 3: Mean Bayes factors and mean posterior predictive p-values for the induced outliers out of 1000 replicates of simulations for the unbalanced scenario with well-connected network, comparing all methods proposed (scenarios 9-16).

τ^2	N outliers	BF	p^{D^L}	$p^{D^{SDO}}$
0	1	111.0	0.01	0.01
		153.6	0.01	0.01
	3	118.2	0.03	0.03
		14.6	0.05	0.03
	1	28.1	0.10	0.05
		188.2	0.05	0.07
0.032	3	234.1	0.05	0.04
		10.2	0.05	0.03
	1	8.6	0.10	0.10
		9.1	0.13	0.10
	3	2.7	0.08	0.05
		2.2	0.20	0.1
0.096	1	2.5	0.11	0.12
		2.1	0.21	0.22
	3	1.8	0.13	0.10
		1.2	0.17	0.25

Table 4: Mean Bayes factors and mean posterior predictive p-values for the induced outliers out of 1000 replicates of simulations for the unbalanced scenario with poorly connected network, comparing all methods proposed (scenarios 25-32).

	Posterior predictive p-value (D^G)			
	$\tau^2 = 0$	$\tau^2 = 0.032$	$\tau^2 = 0.096$	$\tau^2 = 0.287$
Outlying study 1	0.05	0.06	0.10	0.10
Outlying study 2	0.01	0.01	0.22	0.21
Outlying study 1	0.01	0.01	0.06	0.20
No outliers	None	None	$2 \cdot 2p^{D^G} \sim 0.01$	$3p^{D^G} \sim 0.01$

Table 5: Results from simulated network (fairly-connected) using Gelman's discrepancy with three outliers (scenarios 21-24).

$$\tau^2 = 0 \quad \tau^2 = 0.032 \quad \tau^2 = 0.096 \quad \tau^2 = 0.287$$

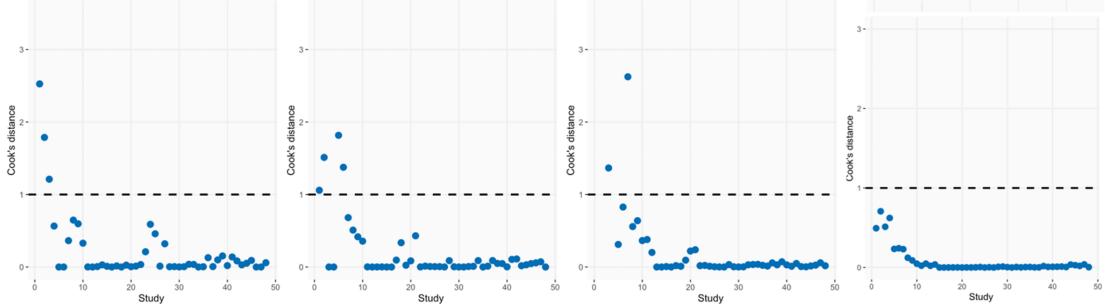


Figure 1: Forward search algorithm outlier detection based on Cook's distance. Points above the dashed line correspond to potential outliers.

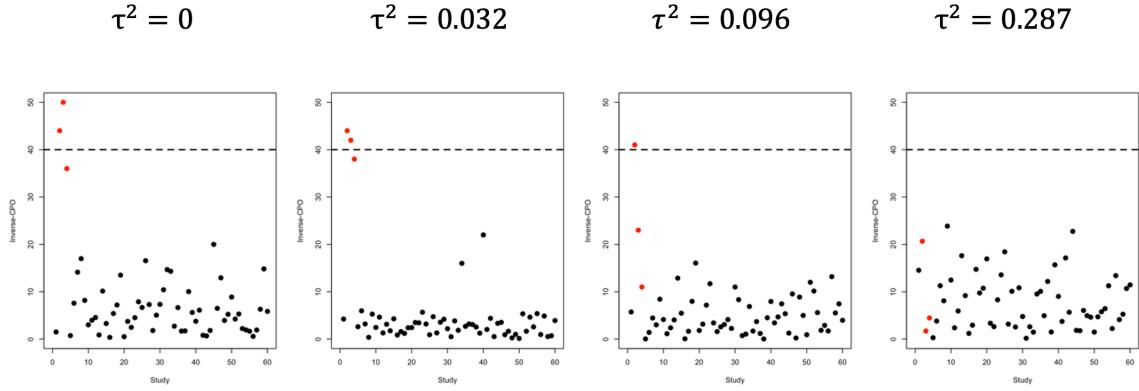


Figure 2: Conditional predictive ordinate outlier detection. Points in red above the dashed line correspond to potential outliers.

2 Additional details and results for the down-weighting scheme

In this section we provide additional details which can constitute useful insight for appropriately choosing the hyperparameters of the beta distribution for the down-weighting scheme presented in Section 4 of the main paper and we report additional results for the relative bias of network meta-analysis estimates with and without down-weighting for the remaining simulated scenarios with three outliers. More specifically, in Figure 3 we show a number of different Beta distributions which correspond to applying less and more severe down-grading while in Figure 4, Figure 5 and Figure 6 we show the relative bias of the estimates under the balanced scenario, unbalanced scenario with well-connected network and unbalanced scenario with poorly-connected network.

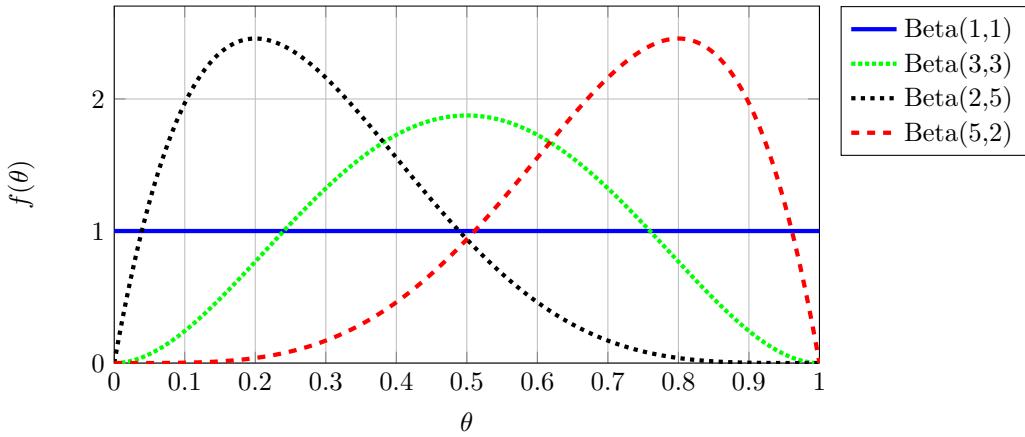


Figure 3: Examples of possible Beta distribution to be used in the down-weighting scheme.

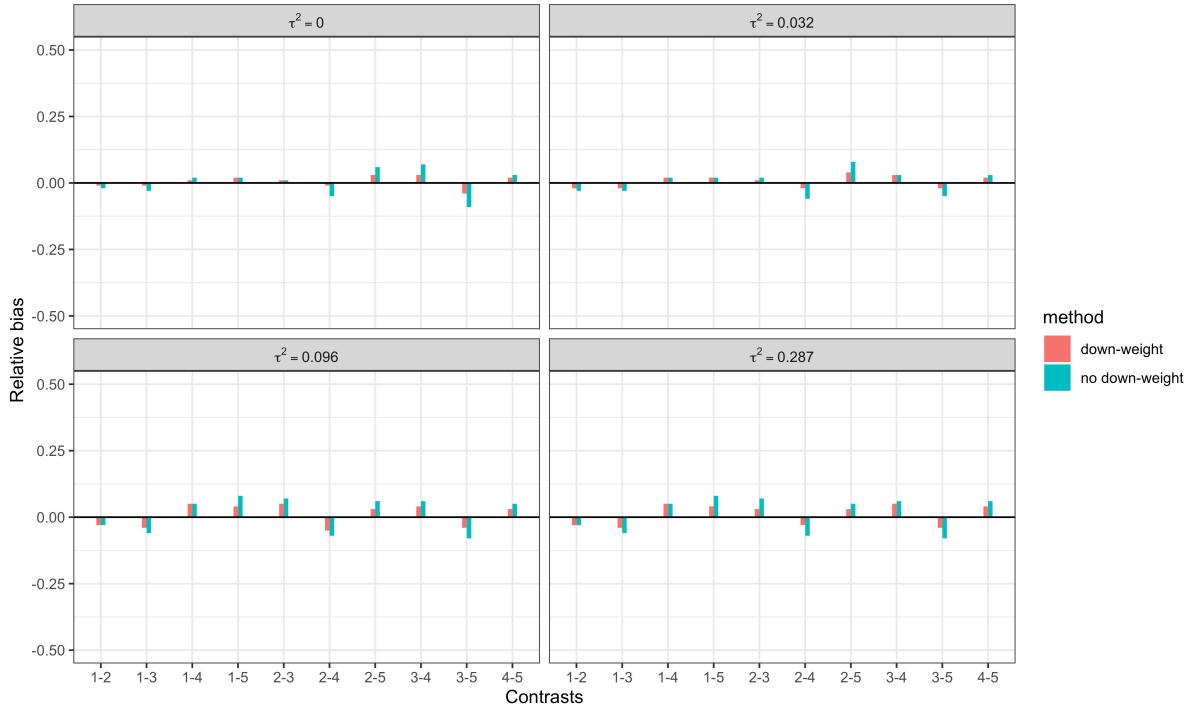


Figure 4: Relative bias plot at varying heterogeneity for the balanced design.

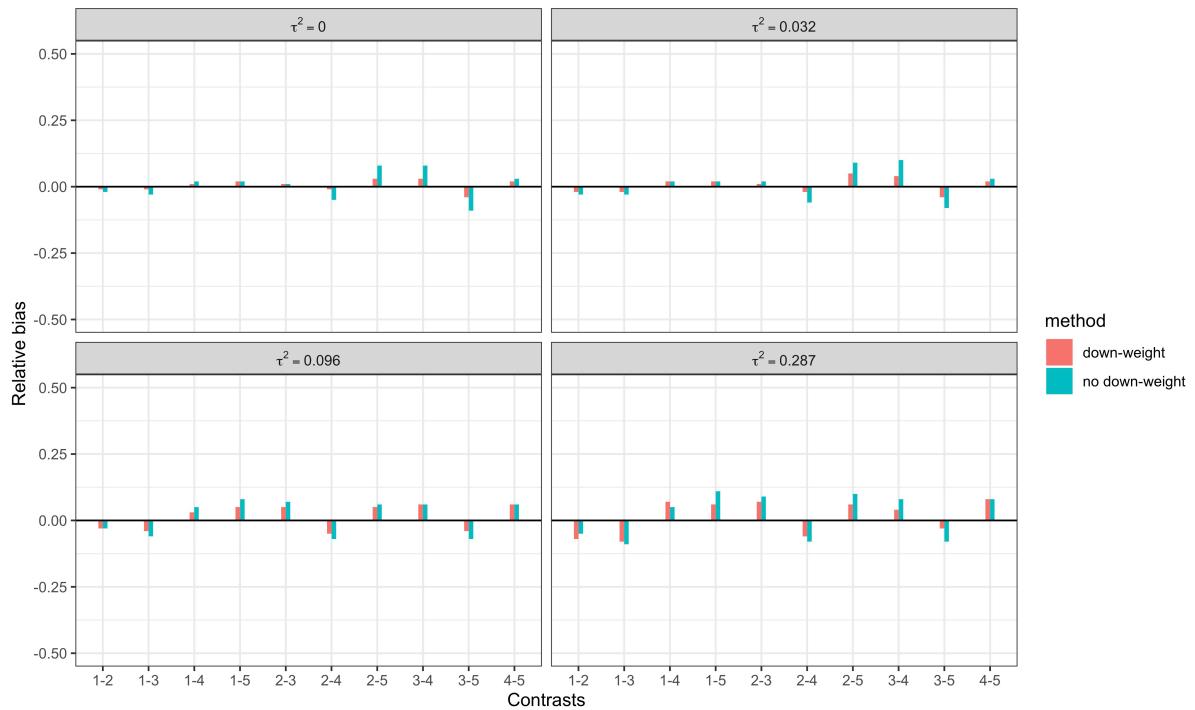


Figure 5: Relative bias plot at varying heterogeneity for the unbalanced, well-connected design.

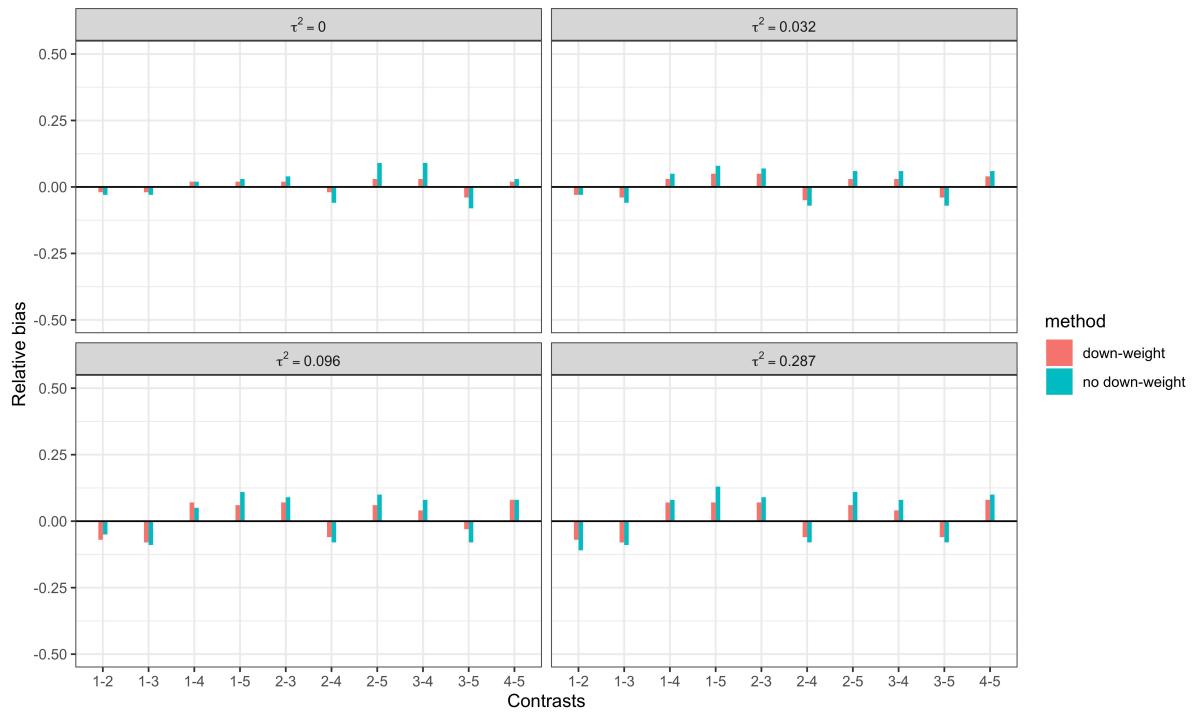


Figure 6: Relative bias plot at varying heterogeneity for the unbalanced, fairly-connected design.

3 Complete real data results

The network geometries for both NSCLC data and smoking cessation data are depicted in Figure 7, while Figure 8 reports additional results from posterior predictive assessments. In Figure 9 we report funnel plots for both networks, i.e. we plot each study's effect estimates centered at the comparison-specific effect against their reversed standard error to further investigate the distribution of the effect sizes (e.g. an asymmetry in the plot might suggest that larger effects tend to be systematically found in smaller studies) and can in some cases raise additional flags of outlyingness. From both plots it is evident that some of the detected outliers lie far from the concentration of data points and create some asymmetry, while both Study 42 and 44 (Monochemotherapy vs Immunotherapy) - which were associated with highest BFs and predictive p -values - are not identified as suspicious in the funnel plot. Finally, the contribution matrix for the lung cancer data is reported in Table 6.

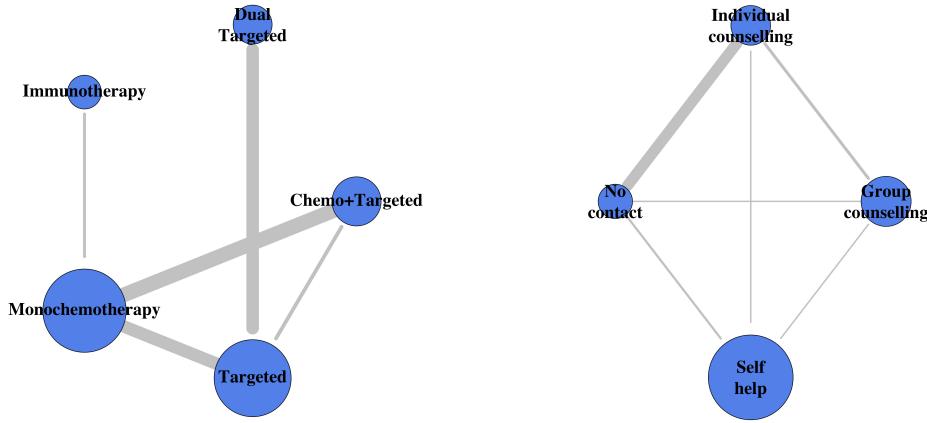


Figure 7: Two real networks of interventions: (left panel) non-small cell lung cancer full network, (right panel) smoking cessation network. The size of the node is proportional to the number of patients randomised and the size of the edge is proportional to the degree of the node

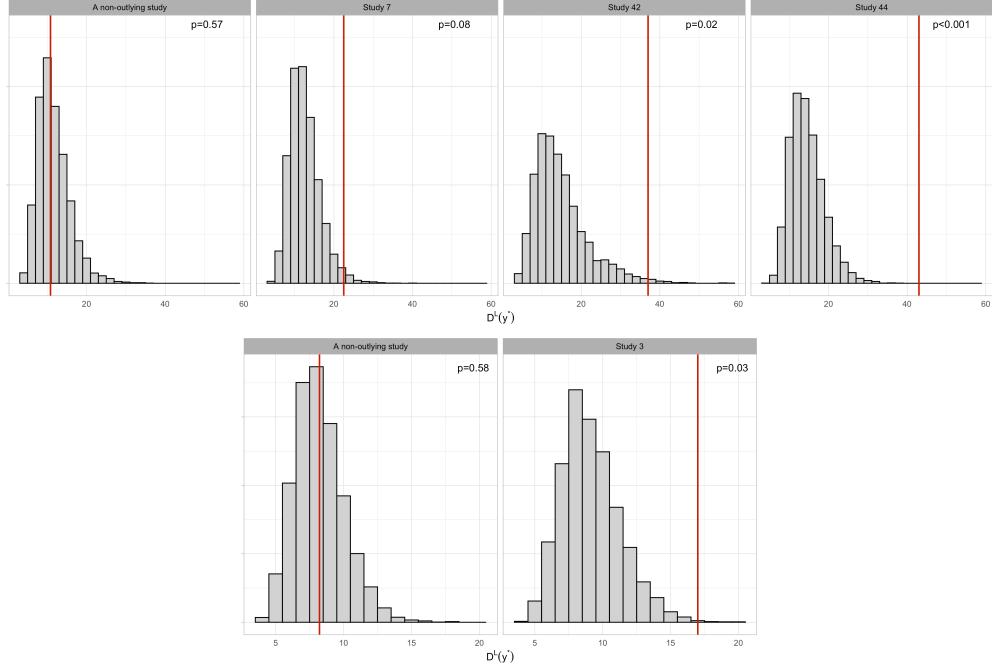


Figure 8: Histograms of draws from the posterior predictive distribution for the replicated vs. realized likelihood-based discrepancy (vertical line) with a randomly chosen non-outlying study used as comparison (annotated posterior predictive p -values) for lung cancer data (top) and smoking cessation data (bottom).

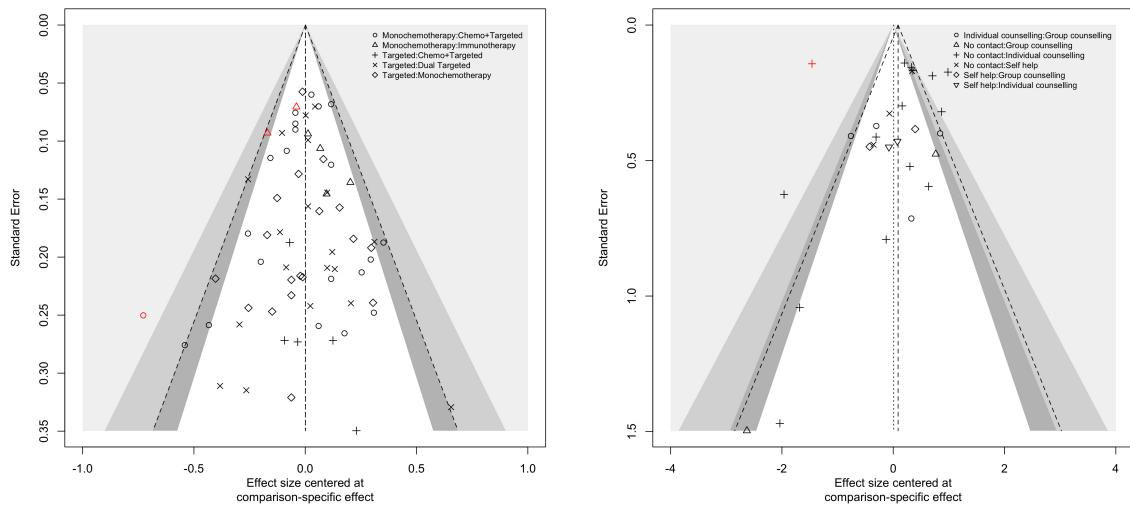


Figure 9: Comparison-adjusted funnel plots centered at comparison-specific effect with pseudo confidence intervals at 90%, 95% and 99% approximate confidence levels for the lung cancer data (left panel) and smoking cessation data (right panel). Treatments ordered from oldest to newest in both networks. Studies in red correspond to the potential outliers detected.

	Chemo+Dual vs Dual	Chemo+Dual vs Immuno	Dual vs Immuno	Dual vs Monochemo	Immuno vs Targeted
Study 1	1.1881	0	1.073	1.5956	0
Study 2	2.04130	1.8437	2.7415	0	
Study 3	1.6037	0.1072	1.7111	2.5667	2.5667
Study 4	0.4828	0	0.436	0.6484	0
Study 5	3.7462	0	3.3834	5.031	0
Study 6	0.4387	0.0293	0.4681	0.7022	0.7022
Study 7	0.3918	0.6588	0.0355	0.0473	0.0473
Study 8	0.7692	0.0514	0.8208	1.2312	1.2312
Study 9	0 5.1246	3.3932	0	5.0456	
Study 10	0	9.1686	6.0709	0	9.0272
Study 11	1.3957	0.3313	0.4489	0.5985	0.5985
Study 12	1.16	0.0776	1.2377	1.8566	1.8566
Study 13	1.0392	0	0.9386	1.3956	0
Study 14	1.8226	0.1219	1.9447	2.9171	2.9171
Study 15	0.6708	1.1278	0.0608	0.081	0.081
Study 16	0.4425	0	0.3996	0.5942	0
Study 17	1.0522	0	0.9503	1.413	0
Study 18	0.3661	0.6155	0.0332	0.0442	0.0442
Study 19	2.3525	0.1573	2.5101	3.7651	3.7651
Study 20	0.5048	0.8487	0.0457	0.061	0.061
Study 21	6.9758	0.4665	7.4432	11.1647	11.1647
Study 22	1.2917	0.0864	1.3782	2.0673	2.0673
Study 23	0.3495	0.5876	0.0317	0.0422	0.0422
Study 24	1.4085	0.3344	0.453	0.604	0.604
Study 25	0.7256	0.0485	0.7743	1.1614	1.1614
Study 26	0.7432	0.0497	0.793	1.1895	1.1895
Study 27	1.6599	0.111	1.7711	2.6566	2.6566
Study 28	0	7.5961	5.0297	0	7.479
Study 29	2.4362	4.0962	0.2207	0.2943	0.2943
Study 30	1.5671	2.635	0.142	0.1893	0.1893
Study 31	0.9227	0.0617	0.9845	1.4767	1.4767
Study 32	0.5293	0.89	0.048	0.0639	0.0639
Study 33	0.9124	0.061	0.9736	1.4604	1.4604
Study 34	0.368	0.6187	0.0333	0.0445	0.0445
Study 35	0.9047	0.0605	0.9653	1.448	1.448
Study 36	0.3987	0.6704	0.0361	0.0482	0.0482
Study 37	0.9314	0.0623	0.9938	1.4908	1.4908
Study 38	3.7002	6.2214	0.3352	0.447	0.447
Study 39	2.2471	3.7783	0.2036	0.2715	0.2715
Study 40	0.8664	0.2057	0.2786	0.3715	0.3715
Study 41	5.2328	0	4.7261	7.0276	0
Study 42	2.3506	0	2.123	3.1568	0
Study 43	0.4943	0	0.4464	0.6638	0
Study 44	0	13.3458	8.8368	0	13.1399
Study 45	0 4.5442	3.0089	0	4.4741	
Study 46	1.2925	0	1.1674	1.7358	0
Study 47	2.8403	4.7756	0.2573	0.3431	0.3431
Study 48	3.2163	5.4078	0.2914	0.3885	0.3885
Study 49	5.9404	0	5.3652	7.9779	0
Study 50	4.1172 0	3.7186	5.5294	0	
Study 51	0.3254	0.5471	0.0295	0.0393	0.0393
Study 52	1.7079	2.8716	0.1547	0.2063	0.2063
Study 53	0.8102	0.0542	0.8645	1.2967	1.2967
Study 54	1.0484	0	0.9469	1.408	0
Study 55	1.4037	0	1.2678	1.8851	0
Study 56	0.8126	0	0.7339	1.0913	0
Study 57	1.2512	0.0837	1.3351	2.0026	2.0026
Study 58	1.4085	0.3344	0.453	0.604	0.604
Study 59	0.7981	0	0.7208	1.0718	0
Study 60	0.7246	1.2184	0.0656	0.0875	0.0875
Study 62	0.5845	0.9828	0.053	0.0706	0.0706
Study 63	2.8261	0.6709	0.909	1.2119	1.2119
Study 64	1.4457	2.4308	0.131	0.1746	0.1746
Study 65	3.1145	5.2366	0.2822	0.3762	0.3762
Study 66	0.5745	0.9659	0.052	0.0694	0.0694
Study 67	0	9.2824	6.1463	0	9.1393
Study 68	0.7068	0	0.6384	0.9493	0
Study 69	1.7788	0	1.6065	2.3888	0

Table 6: Per-study contribution matrix in the lung cancer data set.

References

- Petropoulou, M., Salanti, G., Rücker, G., Schwarzer, G., Moustaki I., and Mavridis, D. (2021). A forward search algorithm for detection of extreme study effects in network meta-analysis. *Statistics in Medicine* 1, 1-25.
- Gelfand, A. E. (1995). Model Determination Using Sampling-Based Methods, In: Gilks W, Richardson S and Spiegelhalter D (eds) *Markov Chain Monte Carlo In Practice*, London, Chapman Hall.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B* **71**, 319–392.
- Ntzoufras I. (2009). *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, West Sussex, England.