OPEN ACCESS

University of BRISTOL

Phillippo, D., Dias, S., Welton, N., Caldwell, D., Taske, N., & Ades, T. (2018). *Confidence in recommendations based on Network Meta-Analysis: threshold analysis as an alternative to GRADE*. Abstract from Guidelines International Network Conference 2018, Manchester, United Kingdom.

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). Please refer to any applicable terms of use of the conference organiser.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms

Confidence in recommendations based on Network Meta-Analysis: threshold analysis as an alternative to GRADE

David M. Phillippo[1], Sofia Dias[1], Nicky J. Welton[1], Deborah M. Caldwell[1], Nichole Taske[2], A. E. Ades[1]

[1]School of Social and Community Medicine, University of Bristol, UK

[2]National Institute for Health and Care Excellence (NICE), London, UK

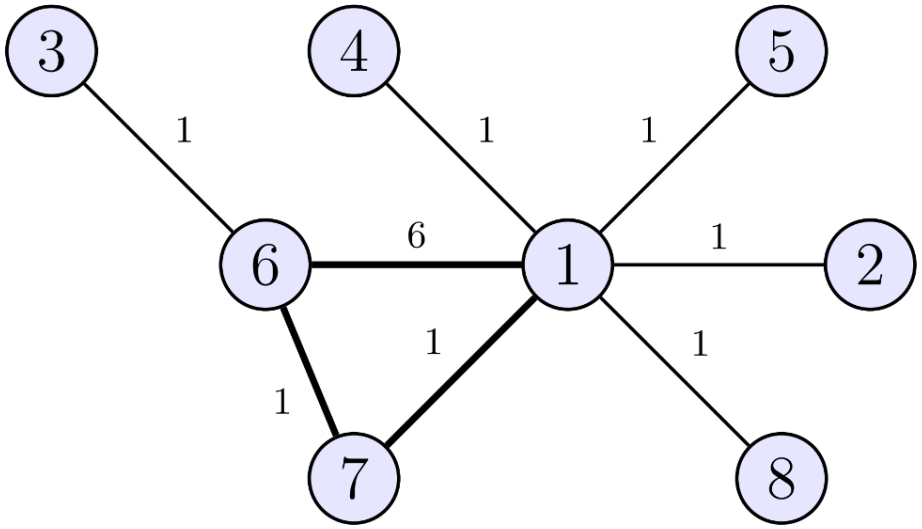
Background

Network Meta-Analysis (NMA)

- Combines evidence on multiple treatments from several studies
- Arranges treatments on a network structure joined by study evidence
- Provides a consistent set of treatment effect estimates
- Is routinely used to inform clinical guideline recommendations, technology appraisals

University of BRISTOL

Example: Headaches clinical guideline



Minimal Clinically Important Difference (MCID) = 0.5 days per month

(NICE CG151.1, 2015)

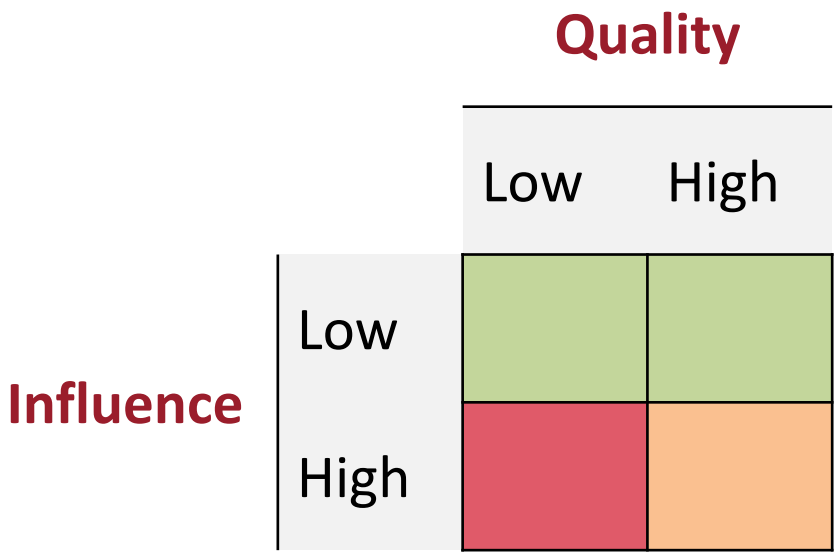
	Treatment	Mean change in headache days per month (95% CrI)
1	Placebo	0
2	Telmisartan	-0.51 (-2.32, 1.27)
3	Amitriptyline	-1.14 (-2.45, 0.16)
4	Divalproex Sodium	0.13 (-0.99, 1.23)
5	Gabapentin	0.00 (-1.60, 1.58)
6	Topiramate	-1.04 (-1.52, -0.58)
7	Propranolol	-1.19 (-2.20, -0.20)
8	Propranolol/Nadolol	-0.60 (-1.65, 0.45)

University of BRISTOL

Motivation

How robust are the recommendations based on NMA?

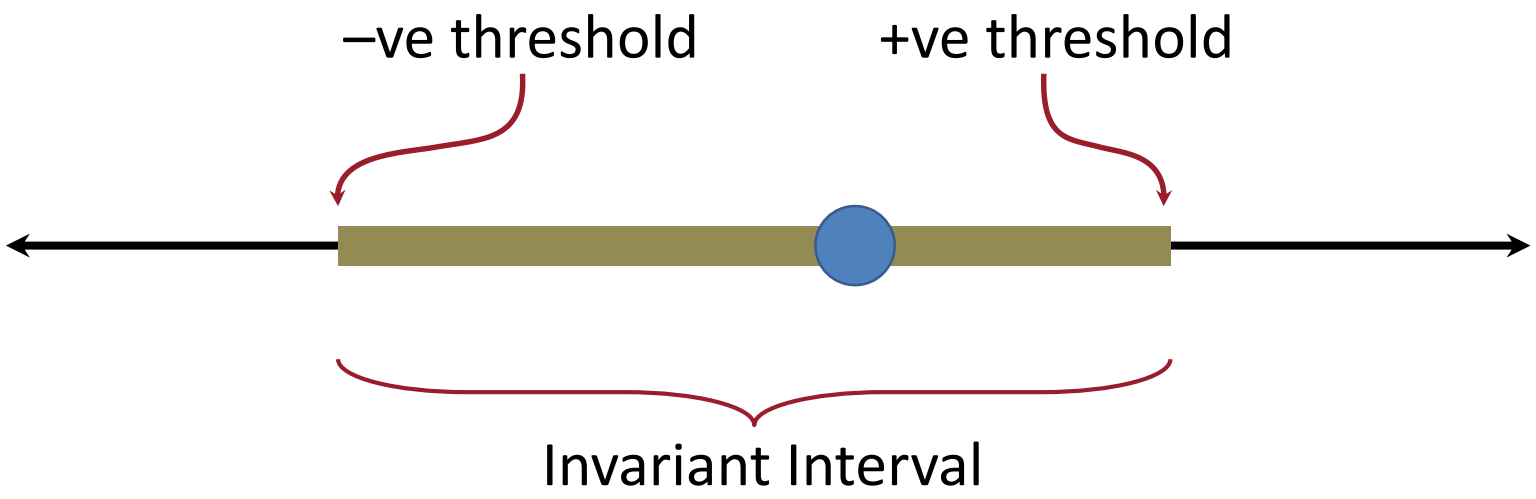
The quality of evidence reflects the extent to which our **confidence in an estimate of the effect** is adequate to **support a particular recommendation**. *(GRADE Handbook, 2013)*



University of BRISTOL

Threshold Analysis

Create an *invariant interval* for a data point:



How much would the evidence have to change before we reach a new recommendation?

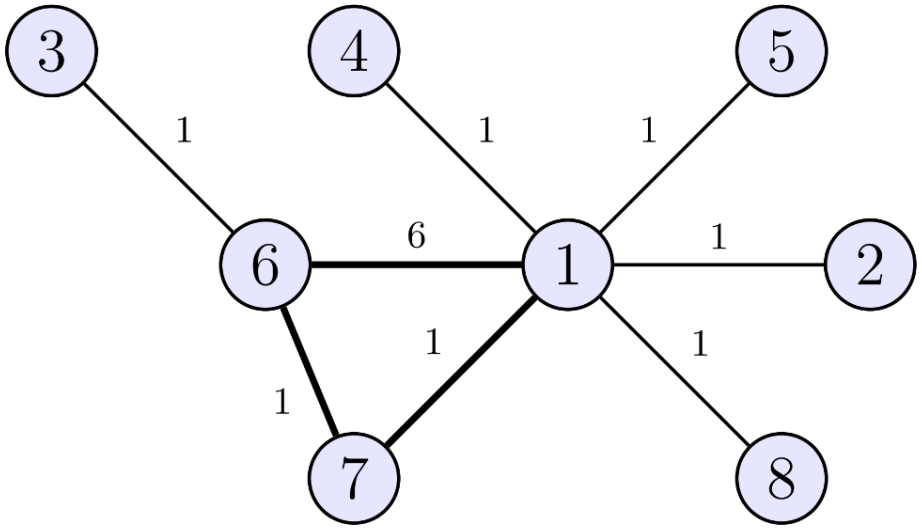
Threshold Analysis

We can do this at two “levels”:

- Study level
  - Thresholds for each individual study estimate
- Contrast level
  - Thresholds for combined body of evidence on a contrast
  - Highly flexible due to approximation step

University of BRISTOL

Example: Headaches clinical guideline



Minimal Clinically Important Difference (MCID) = 0.5 days per month

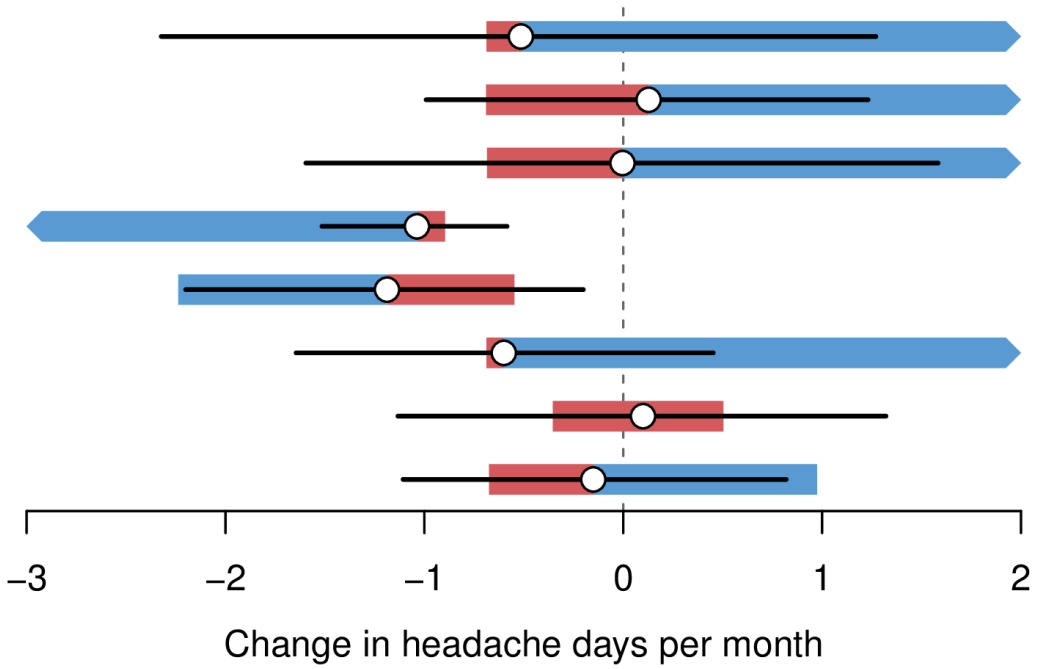
(NICE CG151.1, 2015)

	Treatment	Mean change in headache days per month (95% CrI)
1	Placebo	0
2	Telmisartan	-0.51 (-2.32, 1.27)
3	Amitriptyline	-1.14 (-2.45, 0.16)
4	Divalproex Sodium	0.13 (-0.99, 1.23)
5	Gabapentin	0.00 (-1.60, 1.58)
6	Topiramate	-1.04 (-1.52, -0.58)
7	Propranolol	-1.19 (-2.20, -0.20)
8	Propranolol/Nadolol	-0.60 (-1.65, 0.45)

University of BRISTOL

Example: Headaches clinical guideline – contrast level

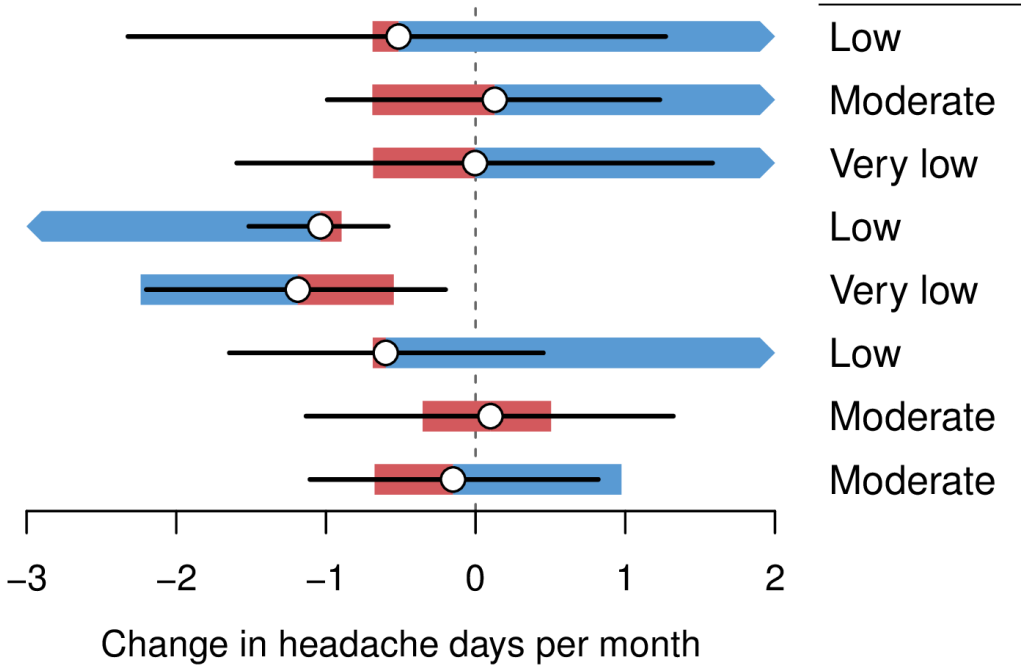
Contrast	Mean	95% CrI	Invariant Interval and $\tilde{k}^*$		
2 vs. 1	–0.51	(–2.32, 1.27)	2, 3, 6, 7	(–0.69, NT)	–
4 vs. 1	0.13	(–0.99, 1.23)	3, 4, 6, 7	(–0.69, 12.02)	3, 6, 7, 8
5 vs. 1	–0.00	(–1.60, 1.58)	3, 5, 6, 7	(–0.69, 24.04)	2, 3, 6, 7
6 vs. 1	–1.04	(–1.51, –0.58)	–	(NT, –0.90)	3, 6, 7, 8
7 vs. 1	–1.19	(–2.20, –0.20)	3, 7	(–2.24, –0.55)	3, 6, 7, 8
8 vs. 1	–0.60	(–1.64, 0.45)	3, 6, 7, 8	(–0.69, NT)	–
6 vs. 3	0.10	(–1.13, 1.32)	6, 7	(–0.35, 0.50)	3, 7
7 vs. 6	–0.15	(–1.11, 0.82)	3, 7	(–0.67, 0.97)	3, 6



Base–case optimal treatment set is 3, 6, 7.

Example: Headaches clinical guideline – contrast level

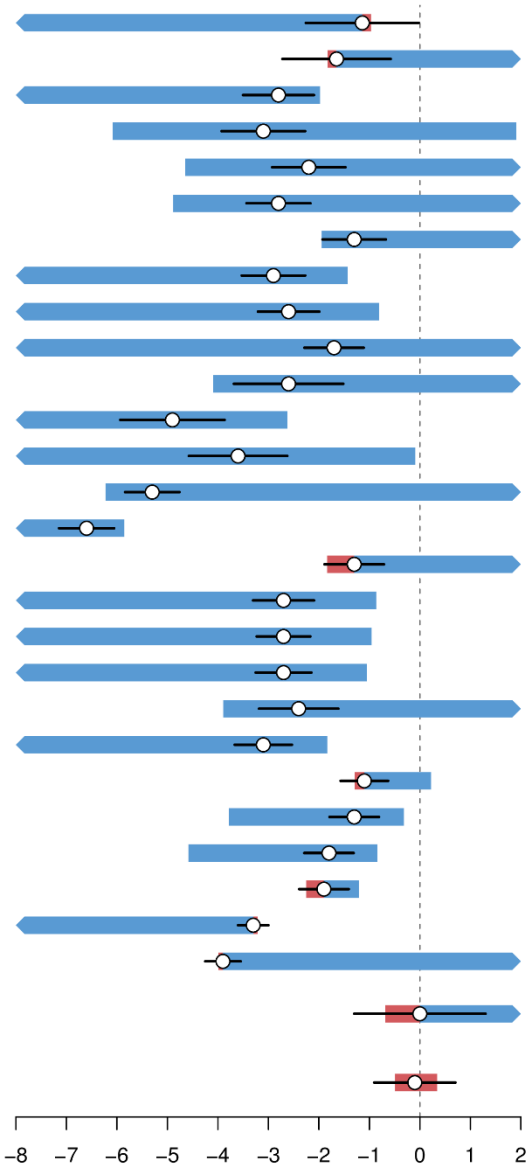
Contrast	Mean	95% CrI	Invariant Interval and $\tilde{k}$*			GRADE
2 vs. 1	–0.51	(–2.32, 1.27)	2, 3, 6, 7	(–0.69, NT)	–	Low
4 vs. 1	0.13	(–0.99, 1.23)	3, 4, 6, 7	(–0.69, 12.02)	3, 6, 7, 8	Moderate
5 vs. 1	–0.00	(–1.60, 1.58)	3, 5, 6, 7	(–0.69, 24.04)	2, 3, 6, 7	Very low
6 vs. 1	–1.04	(–1.51, –0.58)	–	(NT, –0.90)	3, 6, 7, 8	Low
7 vs. 1	–1.19	(–2.20, –0.20)	3, 7	(–2.24, –0.55)	3, 6, 7, 8	Very low
8 vs. 1	–0.60	(–1.64, 0.45)	3, 6, 7, 8	(–0.69, NT)	–	Low
6 vs. 3	0.10	(–1.13, 1.32)	6, 7	(–0.35, 0.50)	3, 7	Moderate
7 vs. 6	–0.15	(–1.11, 0.82)	3, 7	(–0.67, 0.97)	3, 6	Moderate



Base–case optimal treatment set is 3, 6, 7.

Example: Headaches clinical guideline – study level

Study (Trt.)	Mean	95% CI	Invariant Interval and $\tilde{k}^*$		
Diener 2009 (1)	−1.14	(−2.26, −0.02)	–	(NT, −0.97)	2, 3, 6, 7
Diener 2009 (2)	−1.65	(−2.72, −0.58)	2, 3, 6, 7	(−1.82, NT)	–
Apostol 2008 (1)	−2.80	(−3.50, −2.10)	3, 6, 7, 8	(−9.95, −1.98)	3, 4, 6, 7
Apostol 2008 (4)	−3.10	(−3.93, −2.27)	3, 4, 6, 7	(−6.09, 1.91)	3, 6, 7, 8
Apostol 2008 (4)	−2.20	(−2.93, −1.47)	3, 4, 6, 7	(−4.65, NT)	–
Apostol 2008 (4)	−2.80	(−3.43, −2.17)	3, 4, 6, 7	(−4.88, 5.19)	3, 6, 7, 8
Brandes 2004 (1)	−1.30	(−1.93, −0.67)	3, 6, 7, 8	(−1.95, NT)	–
Brandes 2004 (6)	−2.90	(−3.53, −2.27)	–	(NT, −1.43)	3, 6, 7, 8
Brandes 2004 (6)	−2.60	(−3.21, −1.99)	–	(NT, −0.81)	3, 6, 7, 8
Brandes 2004 (6)	−1.70	(−2.29, −1.11)	–	(NT, 15.76)	2, 3, 6, 7
Lewis 2009 (1)	−2.60	(−3.68, −1.52)	3, 6, 7, 8	(−4.10, NT)	–
Lewis 2009 (6)	−4.90	(−5.93, −3.87)	–	(NT, −2.62)	3, 6, 7, 8
Lewis 2009 (6)	−3.60	(−4.57, −2.63)	–	(NT, −0.09)	3, 6, 7, 8
Lipton 2011 (1)	−5.30	(−5.84, −4.76)	3, 6, 7, 8	(−6.22, NT)	–
Lipton 2011 (6)	−6.60	(−7.14, −6.06)	–	(NT, −5.85)	3, 6, 7, 8
Silberstein 2004 (1)	−1.30	(−1.89, −0.71)	3, 6, 7, 8	(−1.84, NT)	–
Silberstein 2004 (6)	−2.70	(−3.30, −2.10)	–	(NT, −0.87)	3, 6, 7, 8
Silberstein 2004 (6)	−2.70	(−3.23, −2.17)	–	(NT, −0.96)	3, 6, 7, 8
Silberstein 2004 (6)	−2.70	(−3.25, −2.15)	–	(NT, −1.05)	3, 6, 7, 8
Winner 2005 (1)	−2.40	(−3.18, −1.62)	3, 6, 7, 8	(−3.89, NT)	–
Winner 2005 (6)	−3.10	(−3.67, −2.53)	–	(NT, −1.83)	3, 6, 7, 8
Diener 2004 (1)	−1.10	(−1.57, −0.63)	3, 6, 7, 8	(−1.29, 0.22)	3, 7
Diener 2004 (6)	−1.30	(−1.79, −0.81)	3, 6	(−3.78, −0.32)	3, 7
Diener 2004 (6)	−1.80	(−2.29, −1.31)	3, 6	(−4.58, −0.85)	3, 7
Diener 2004 (7)	−1.90	(−2.39, −1.41)	3, 7	(−2.25, −1.21)	3, 6
Holroyd 2010 (1)	−3.30	(−3.60, −3.00)	6, 7	(−17.43, −3.21)	3, 6, 7, 8
Holroyd 2010 (8)	−3.90	(−4.25, −3.55)	3, 6, 7, 8	(−3.99, 10.77)	6, 7
Silberstein 2013 (5 vs. 1)	0.00	(−1.30, 1.30)	3, 5, 6, 7	(−0.69, 11.76)	2, 3, 6, 7
Dodick 2009 (3 vs. 6)	−0.10	(−0.90, 0.70)	3, 7	(−0.49, 0.34)	6, 7

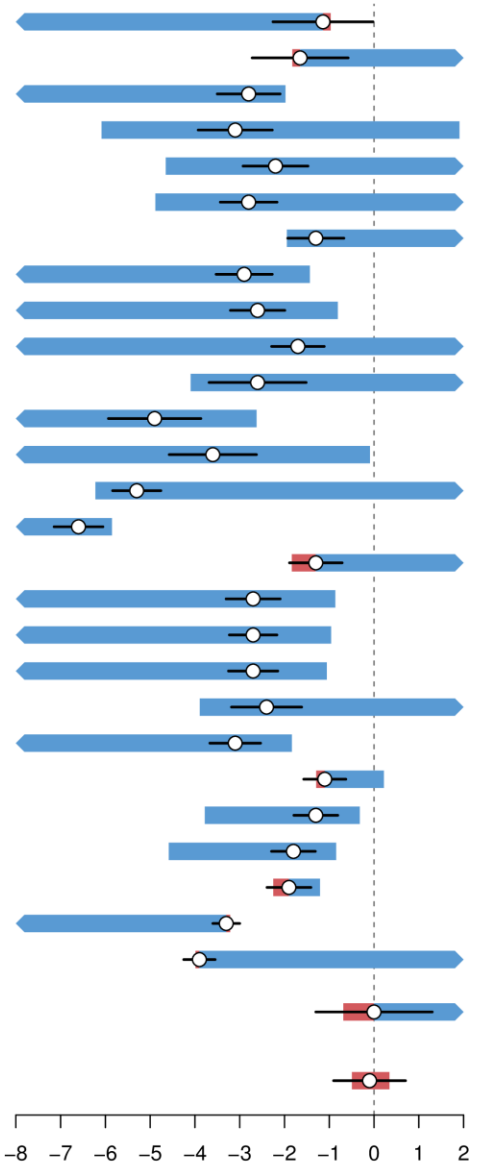


○ Mean — 95% CI ▭ Invariant Interval

Base–case optimal treatment set is 3, 6, 7.

Example: Headaches clinical guideline – study level

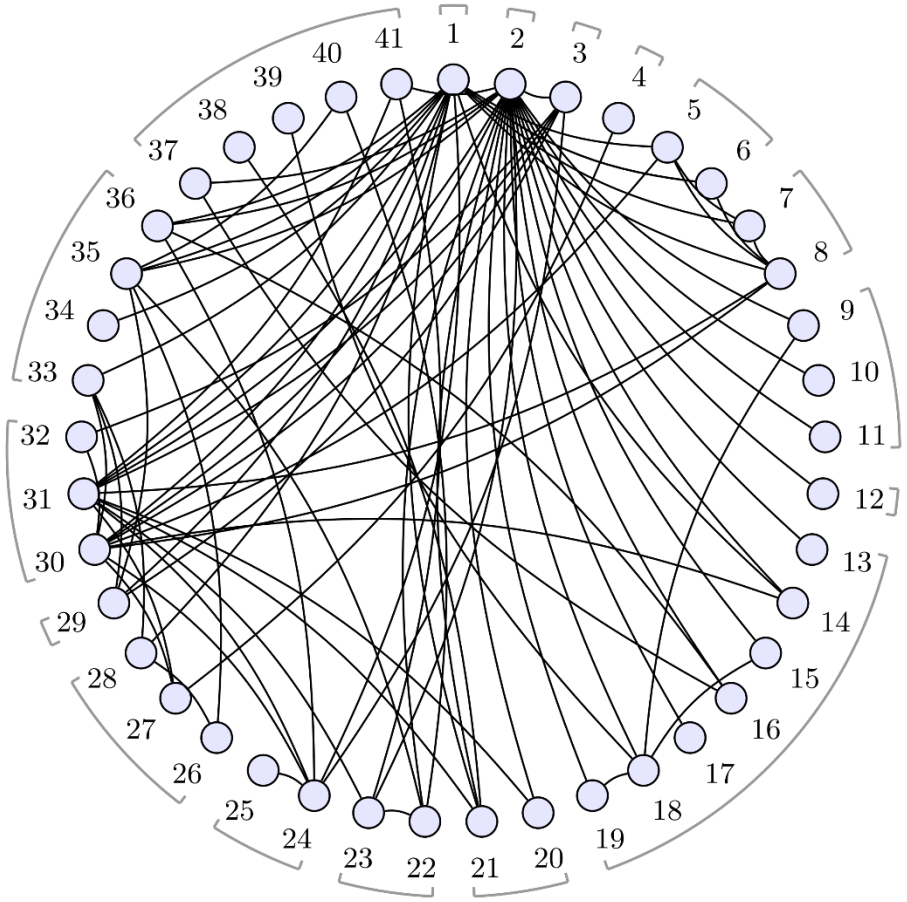
Study (Trt.)	Mean	95% CI	Invariant Interval and $\tilde{k}^*$			Sel	Perf	Det	Att	Rep	Oth
Diener 2009 (1)	−1.14	(−2.26, −0.02)	–	(NT, −0.97)	2, 3, 6, 7	U	L	L	L	U	H
Diener 2009 (2)	−1.65	(−2.72, −0.58)	2, 3, 6, 7	(−1.82, NT)	–						
Apostol 2008 (1)	−2.80	(−3.50, −2.10)	3, 6, 7, 8	(−9.95, −1.98)	3, 4, 6, 7	U	L	L	L	U	L
Apostol 2008 (4)	−3.10	(−3.93, −2.27)	3, 4, 6, 7	(−6.09, 1.91)	3, 6, 7, 8						
Apostol 2008 (4)	−2.20	(−2.93, −1.47)	3, 4, 6, 7	(−4.65, NT)	–						
Apostol 2008 (4)	−2.80	(−3.43, −2.17)	3, 4, 6, 7	(−4.88, 5.19)	3, 6, 7, 8						
Brandes 2004 (1)	−1.30	(−1.93, −0.67)	3, 6, 7, 8	(−1.95, NT)	–	L	L	L	H	U	L
Brandes 2004 (6)	−2.90	(−3.53, −2.27)	–	(NT, −1.43)	3, 6, 7, 8						
Brandes 2004 (6)	−2.60	(−3.21, −1.99)	–	(NT, −0.81)	3, 6, 7, 8						
Brandes 2004 (6)	−1.70	(−2.29, −1.11)	–	(NT, 15.76)	2, 3, 6, 7						
Lewis 2009 (1)	−2.60	(−3.68, −1.52)	3, 6, 7, 8	(−4.10, NT)	–	U	L	L	L	U	L
Lewis 2009 (6)	−4.90	(−5.93, −3.87)	–	(NT, −2.62)	3, 6, 7, 8						
Lewis 2009 (6)	−3.60	(−4.57, −2.63)	–	(NT, −0.09)	3, 6, 7, 8						
Lipton 2011 (1)	−5.30	(−5.84, −4.76)	3, 6, 7, 8	(−6.22, NT)	–	L	L	L	H	U	L
Lipton 2011 (6)	−6.60	(−7.14, −6.06)	–	(NT, −5.85)	3, 6, 7, 8						
Silberstein 2004 (1)	−1.30	(−1.89, −0.71)	3, 6, 7, 8	(−1.84, NT)	–	U	L	L	H	U	L
Silberstein 2004 (6)	−2.70	(−3.30, −2.10)	–	(NT, −0.87)	3, 6, 7, 8						
Silberstein 2004 (6)	−2.70	(−3.23, −2.17)	–	(NT, −0.96)	3, 6, 7, 8						
Silberstein 2004 (6)	−2.70	(−3.25, −2.15)	–	(NT, −1.05)	3, 6, 7, 8						
Winner 2005 (1)	−2.40	(−3.18, −1.62)	3, 6, 7, 8	(−3.89, NT)	–	U	L	L	L	U	L
Winner 2005 (6)	−3.10	(−3.67, −2.53)	–	(NT, −1.83)	3, 6, 7, 8						
Diener 2004 (1)	−1.10	(−1.57, −0.63)	3, 6, 7, 8	(−1.29, 0.22)	3, 7	U	L	L	H	U	U
Diener 2004 (6)	−1.30	(−1.79, −0.81)	3, 6	(−3.78, −0.32)	3, 7						
Diener 2004 (6)	−1.80	(−2.29, −1.31)	3, 6	(−4.58, −0.85)	3, 7						
Diener 2004 (7)	−1.90	(−2.39, −1.41)	3, 7	(−2.25, −1.21)	3, 6						
Holroyd 2010 (1)	−3.30	(−3.60, −3.00)	6, 7	(−17.43, −3.21)	3, 6, 7, 8	L	L	L	L	U	L
Holroyd 2010 (8)	−3.90	(−4.25, −3.55)	3, 6, 7, 8	(−3.99, 10.77)	6, 7						
Silberstein 2013 (5 vs. 1)	0.00	(−1.30, 1.30)	3, 5, 6, 7	(−0.69, 11.76)	2, 3, 6, 7	U	L	L	L	H	L
Dodick 2009 (3 vs. 6)	−0.10	(−0.90, 0.70)	3, 7	(−0.49, 0.34)	6, 7	U	L	L	L	U	L



○ Mean — 95% CI ▬ Invariant Interval

Base−case optimal treatment set is 3, 6, 7.

Example: Social Anxiety



- 41 treatments, 100 studies
- Class effect model with 17 classes
- 84 direct comparisons and many “first order” loops

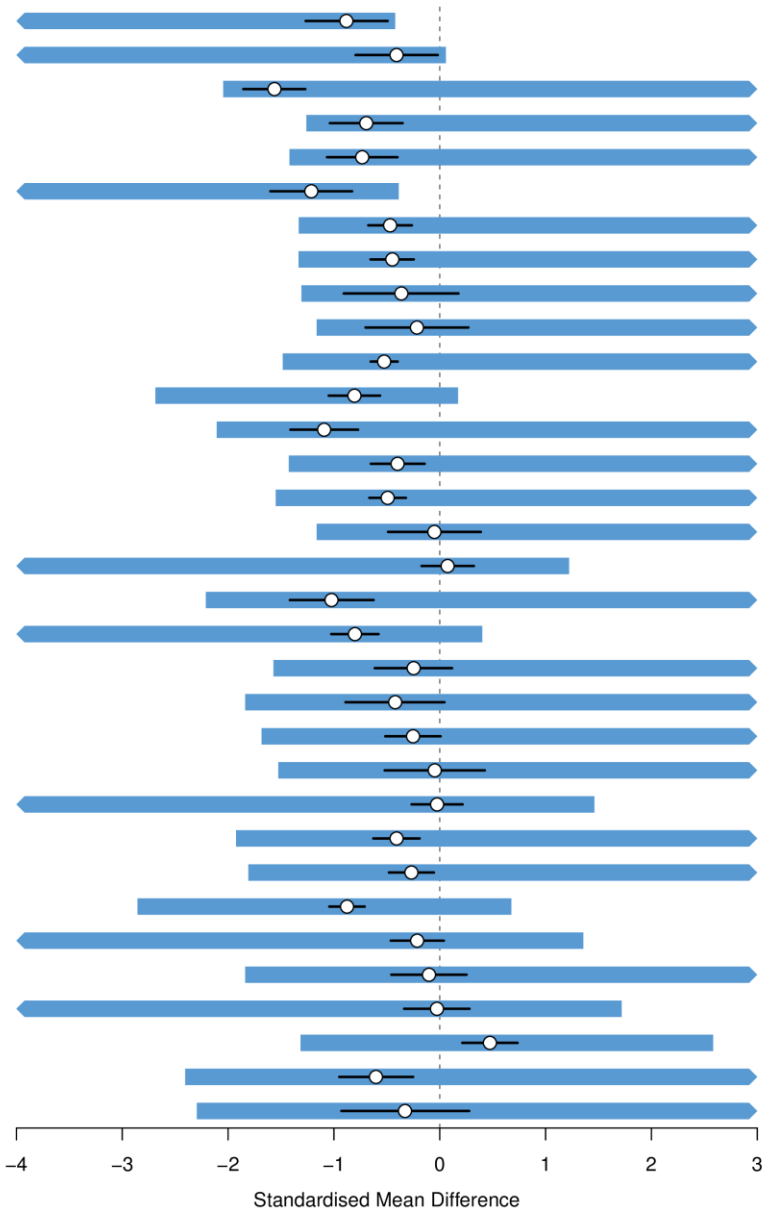
  ⇒ GRADE impractical

Based on efficacy alone recommendation is CBT + Phenelzine (41)

(NCC-MH, 2013)

Example: Social Anxiety – contrast level

Contrast	SMD	95% Credible Interval	Invariant Interval and $\tilde{k}^*$		
41 vs. 31	−0.88	(−1.27, −0.49)	–	(NT, −0.42)	36
41 vs. 23	−0.41	(−0.80, −0.02)	–	(NT, 0.06)	36
36 vs. 1	−1.56	(−1.86, −1.27)	36	(−2.05, 13.61)	7
36 vs. 16	−0.69	(−1.04, −0.35)	36	(−1.26, 4.91)	16
36 vs. 24	−0.73	(−1.06, −0.40)	36	(−1.42, 5.61)	25
41 vs. 2	−1.21	(−1.60, −0.83)	–	(NT, −0.39)	36
17 vs. 2	−0.47	(−0.68, −0.26)	17	(−1.33, 50.71)	12
13 vs. 2	−0.45	(−0.65, −0.25)	13	(−1.33, 139.66)	4
39 vs. 18	−0.36	(−0.91, 0.18)	39	(−1.31, 3.60)	36
38 vs. 21	−0.22	(−0.70, 0.27)	38	(−1.16, 3.71)	36
18 vs. 2	−0.53	(−0.65, −0.40)	39	(−1.48, 8.79)	36
23 vs. 2	−0.81	(−1.05, −0.56)	23	(−2.69, 0.17)	36
36 vs. 2	−1.09	(−1.41, −0.77)	36	(−2.11, 34.05)	18
16 vs. 2	−0.40	(−0.65, −0.14)	36	(−1.43, 17.40)	12
19 vs. 2	−0.49	(−0.67, −0.32)	19	(−1.55, 29.73)	23
25 vs. 24	−0.05	(−0.49, 0.39)	25	(−1.16, 554.74)	39
31 vs. 8	0.07	(−0.17, 0.32)	31	(−13.67, 1.22)	36
34 vs. 1	−1.02	(−1.42, −0.63)	34	(−2.21, 101.29)	5
31 vs. 1	−0.80	(−1.02, −0.58)	31	(−16.38, 0.40)	36
32 vs. 30	−0.25	(−0.61, 0.12)	32	(−1.57, 5.66)	36
11 vs. 2	−0.42	(−0.89, 0.04)	11	(−1.84, 856.05)	34
9 vs. 2	−0.25	(−0.51, 0.01)	9	(−1.68, 46.63)	36
40 vs. 35	−0.05	(−0.52, 0.43)	40	(−1.53, 3.12)	36
30 vs. 24	−0.02	(−0.27, 0.22)	30	(−8.80, 1.46)	36
15 vs. 2	−0.41	(−0.63, −0.19)	15	(−1.92, 49.40)	23
22 vs. 2	−0.27	(−0.48, −0.06)	22	(−1.81, 5.92)	18
8 vs. 1	−0.88	(−1.04, −0.71)	8	(−2.86, 0.68)	36
8 vs. 6	−0.21	(−0.46, 0.04)	8	(−13.40, 1.36)	6
37 vs. 30	−0.10	(−0.46, 0.25)	37	(−1.84, 7.10)	36
8 vs. 7	−0.03	(−0.34, 0.28)	8	(−17.13, 1.72)	7
31 vs. 23	0.47	(0.21, 0.74)	36	(−1.32, 2.58)	23
21 vs. 2	−0.60	(−0.95, −0.25)	21	(−2.41, 10.57)	36
12 vs. 2	−0.33	(−0.93, 0.28)	12	(−2.30, 59.55)	36



○ SMD — 95% Credible Interval ■ Invariant Interval

Base−case recommended treatment is 41.

Thresholds smaller than 0.8 SMD

Thresholds larger than 0.8 SMD

Example: Social Anxiety

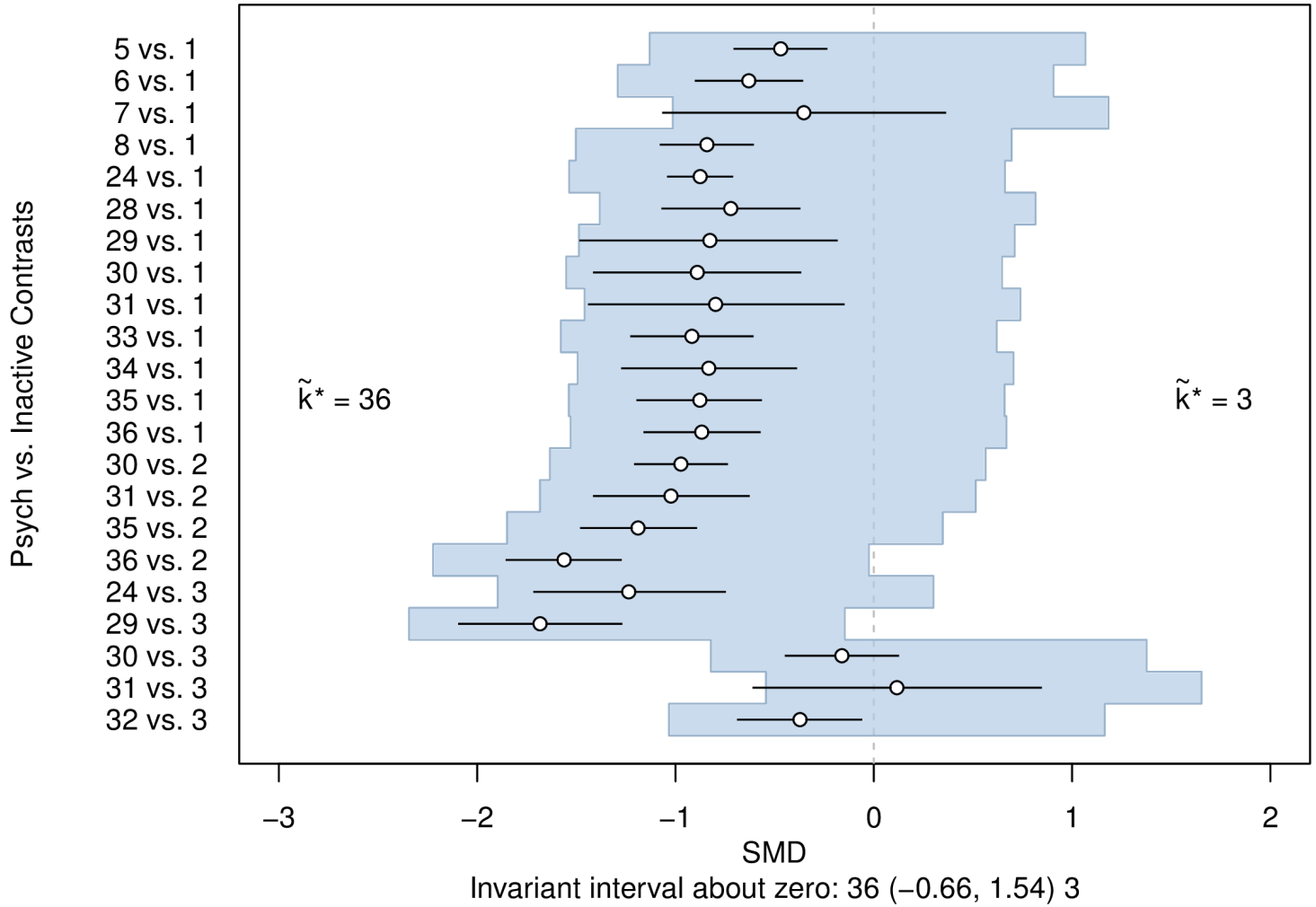
We can also use threshold analysis to investigate more specific concerns

- E.g. groups of treatments or studies with certain characteristics

Psychological treatment bias – could this affect the recommendation?

University of BRISTOL

Example: Social Anxiety – psychological treatment bias



University of BRISTOL

Conclusions

- Evidence quality and risk of bias are not sufficient to assess robustness of decisions
- Threshold analysis provides insight into the effects of changes in the evidence on treatment decisions
  - We can have more confidence in recommendations where thresholds are large
  - We can focus attention on the quality of decision-sensitive trials and contrasts
- More complex analyses can investigate specific concerns in the evidence
- Can be used with a range of decision rules or for decisions based on cost-effectiveness

Phillippo DM, Dias S, Ades AE, Didelez V and Welton NJ (2018). *Sensitivity of treatment recommendations to bias in network meta-analysis*. J. R. Stat. Soc. A, 181: 843-867. doi:10.1111/rssa.12341

R package *nmathresh* package available on CRAN

University of BRISTOL

Thank You

NICE
National Institute for Health and Care Excellence

This work was supported by the Centre for Clinical Practice, NICE, with funding to the Clinical Guidelines Technical Support Unit, University of Bristol.

MRC Medical Research Council

This work was undertaken with the support of the MRC ConDuCT-II Hub (MR/K025643/1), and the MRC grant MR/M005232/1.