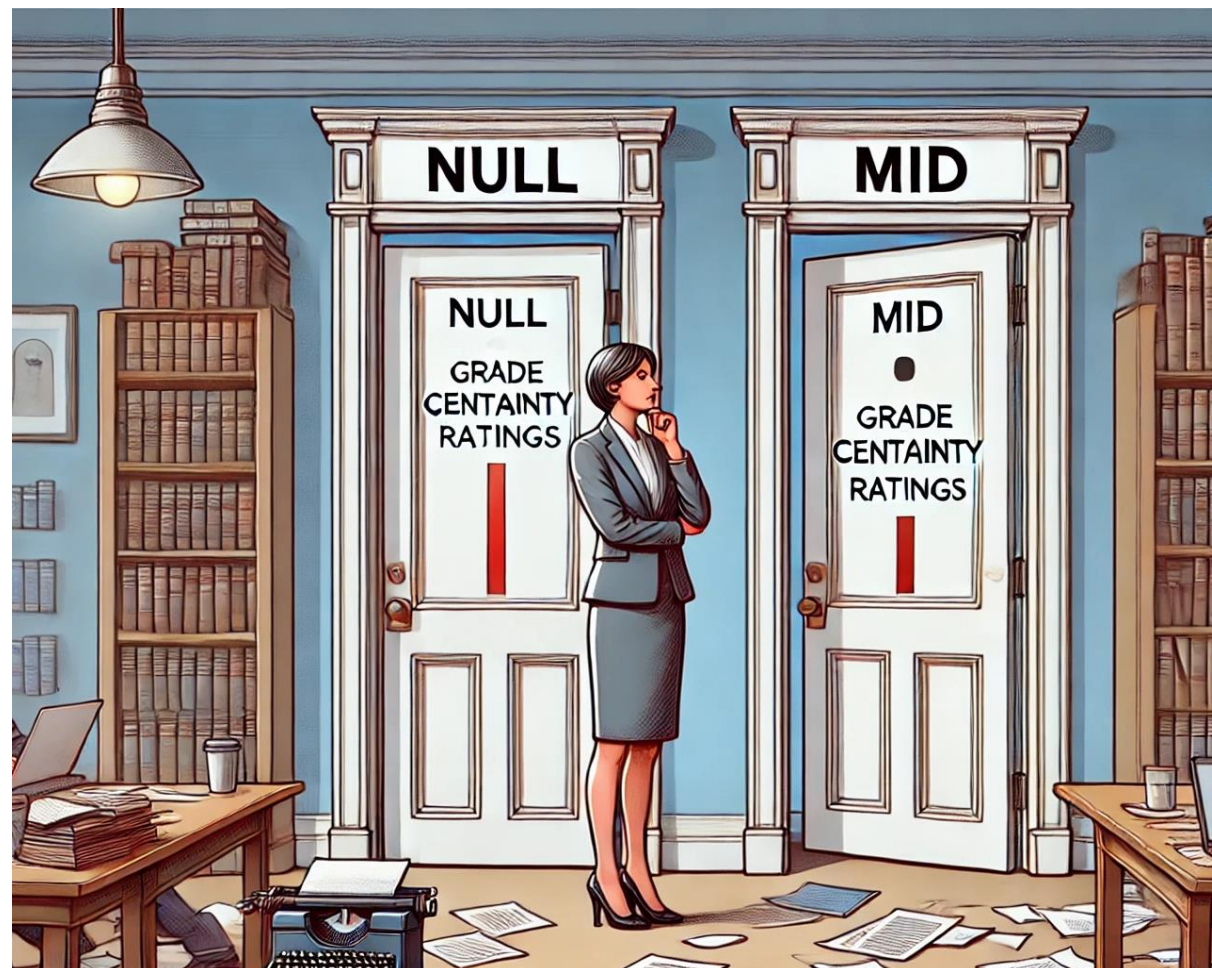


# What to consider when choosing thresholds for GRADE certainty ratings?

Monica Hultcrantz  
2024-11-08



# Disclosure

- I have no actual or potential conflict of interest in relation to this presentation
- Active in the GRADE Working group: member of the GRADE guidance group and lead of the certainty in evidence project group
- Part of the co-ordinating team for the Nordic GRADE network





Journal of Clinical Epidemiology 64 (2011) 401–406

**Journal of  
Clinical  
Epidemiology**

## GRADE guidelines: 3. Rating the quality of evidence

Howard Balshem<sup>a,\*</sup>, Mark Helfand<sup>a,b</sup>, Holger J. Schünemann<sup>c</sup>, Andrew D. Oxman<sup>d</sup>,  
Regina Kunz<sup>e</sup>, Jan Brozek<sup>c</sup>, Gunn E. Vist<sup>d</sup>, Yngve Falck-Ytter<sup>f</sup>, Joerg Meerpohl<sup>g,h</sup>,  
Susan Norris<sup>i</sup>, Gordon H. Guyatt<sup>c</sup>

proaches on all patient-important outcomes [1]. In the context of a systematic review, the ratings of the quality of evidence reflect the extent of our confidence that the estimates of the effect are correct. In the context of making recommendations, the quality ratings reflect the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.



Journal of Clinical Epidemiology ■ (2017) ■

**Journal of  
Clinical  
Epidemiology**

## ORIGINAL ARTICLE

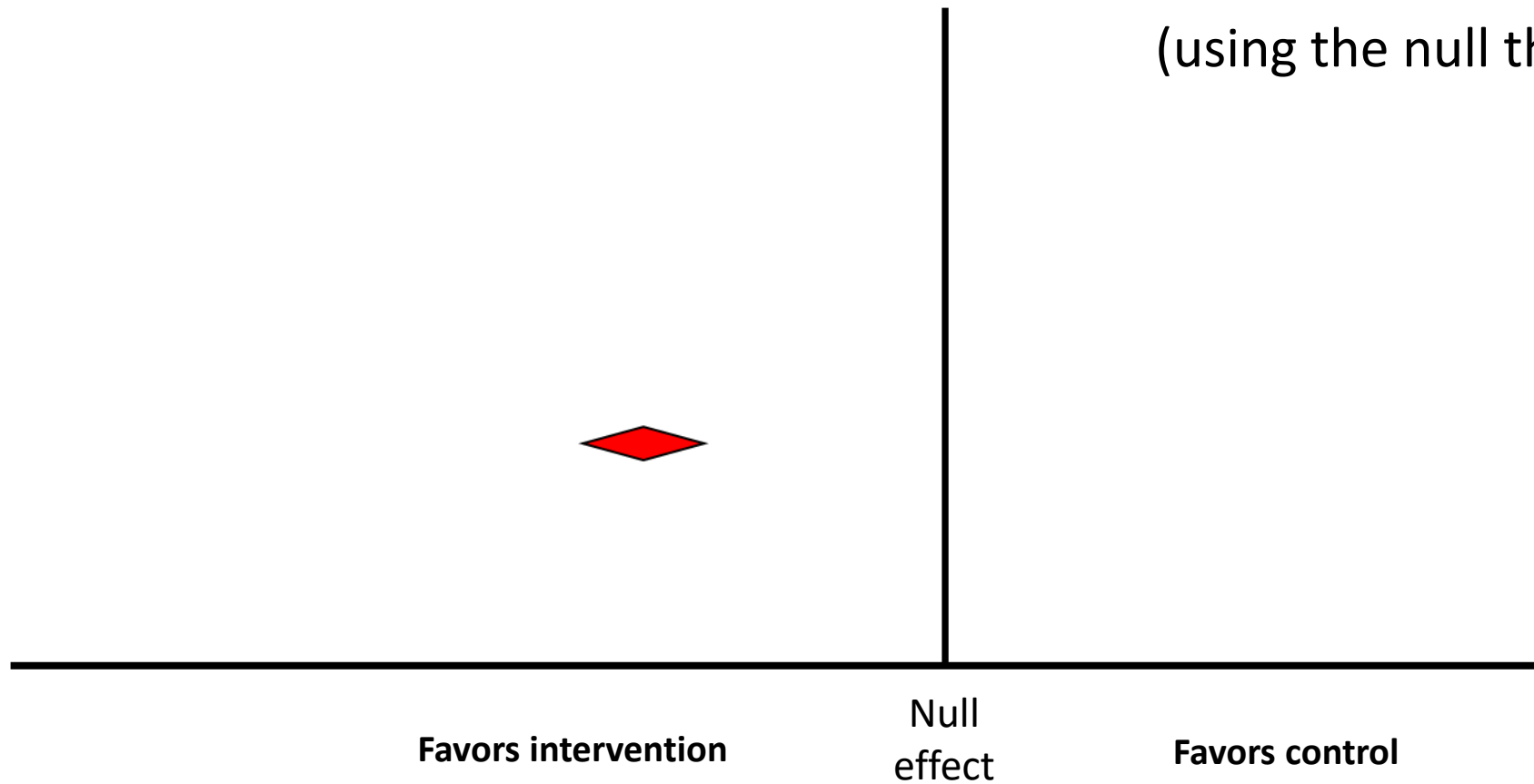
# The GRADE Working Group clarifies the construct of certainty of evidence

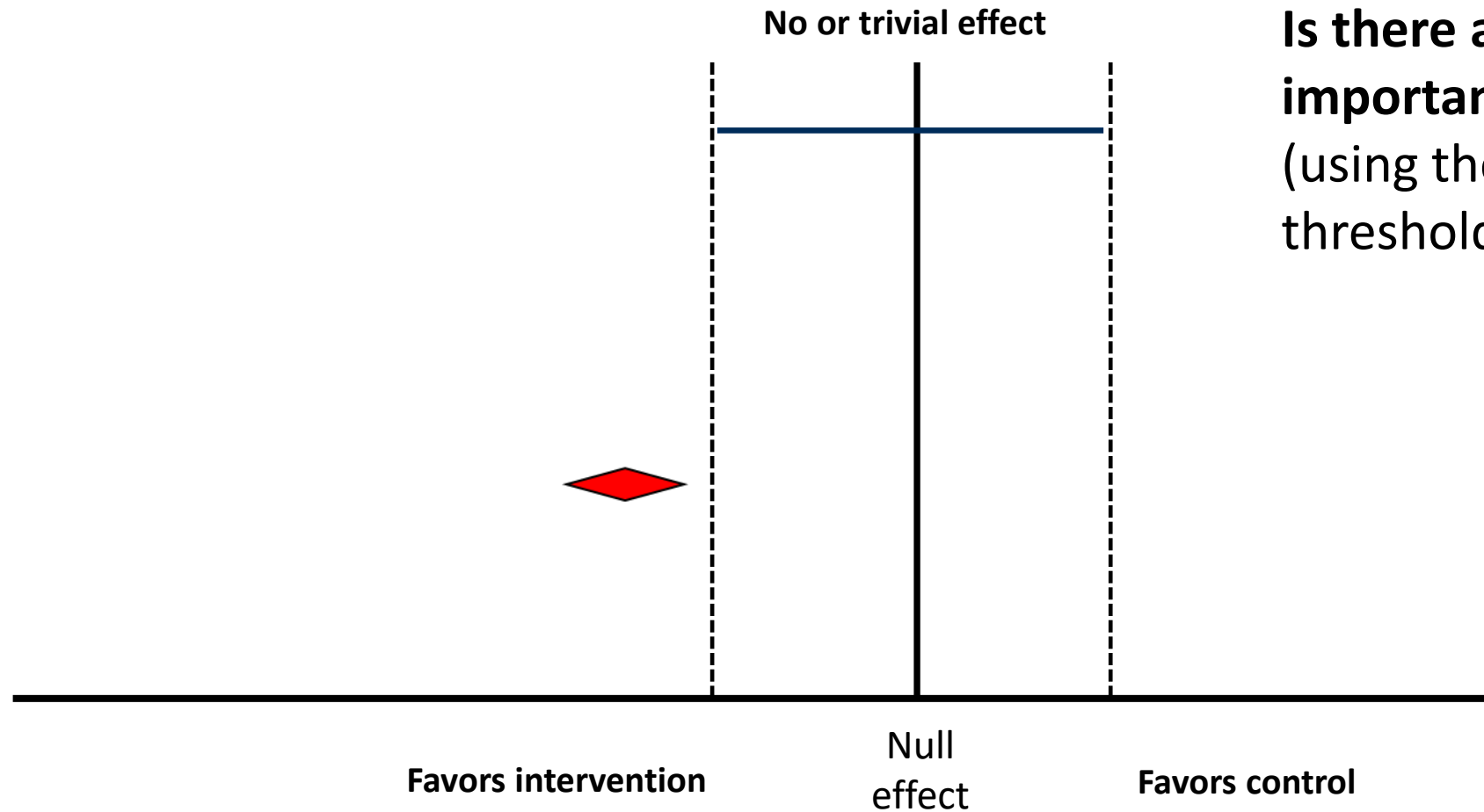
Monica Hultcrantz<sup>a,b,\*</sup>, David Rind<sup>c,d</sup>, Elie A. Akl<sup>e,f</sup>, Shaun Treweek<sup>g</sup>, Reem A. Mustafa<sup>e,h</sup>,  
Alfonso Iorio<sup>e,i</sup>, Brian S. Alper<sup>j,k</sup>, Joerg J. Meerpohl<sup>l,m</sup>, M Hassan Murad<sup>n</sup>,  
Mohammed T. Ansari<sup>o</sup>, Srinivasa Vittal Katikireddi<sup>p</sup>, Pernilla Östlund<sup>a,q</sup>, Sofia Tranæus<sup>a,q,r</sup>,  
Robin Christensen<sup>s</sup>, Gerald Gartlehner<sup>t,u</sup>, Jan Brozek<sup>e,i</sup>, Ariel Izcovich<sup>v</sup>, Holger Schünemann<sup>e,i</sup>,  
Gordon Guyatt<sup>e,i</sup>

## Key findings

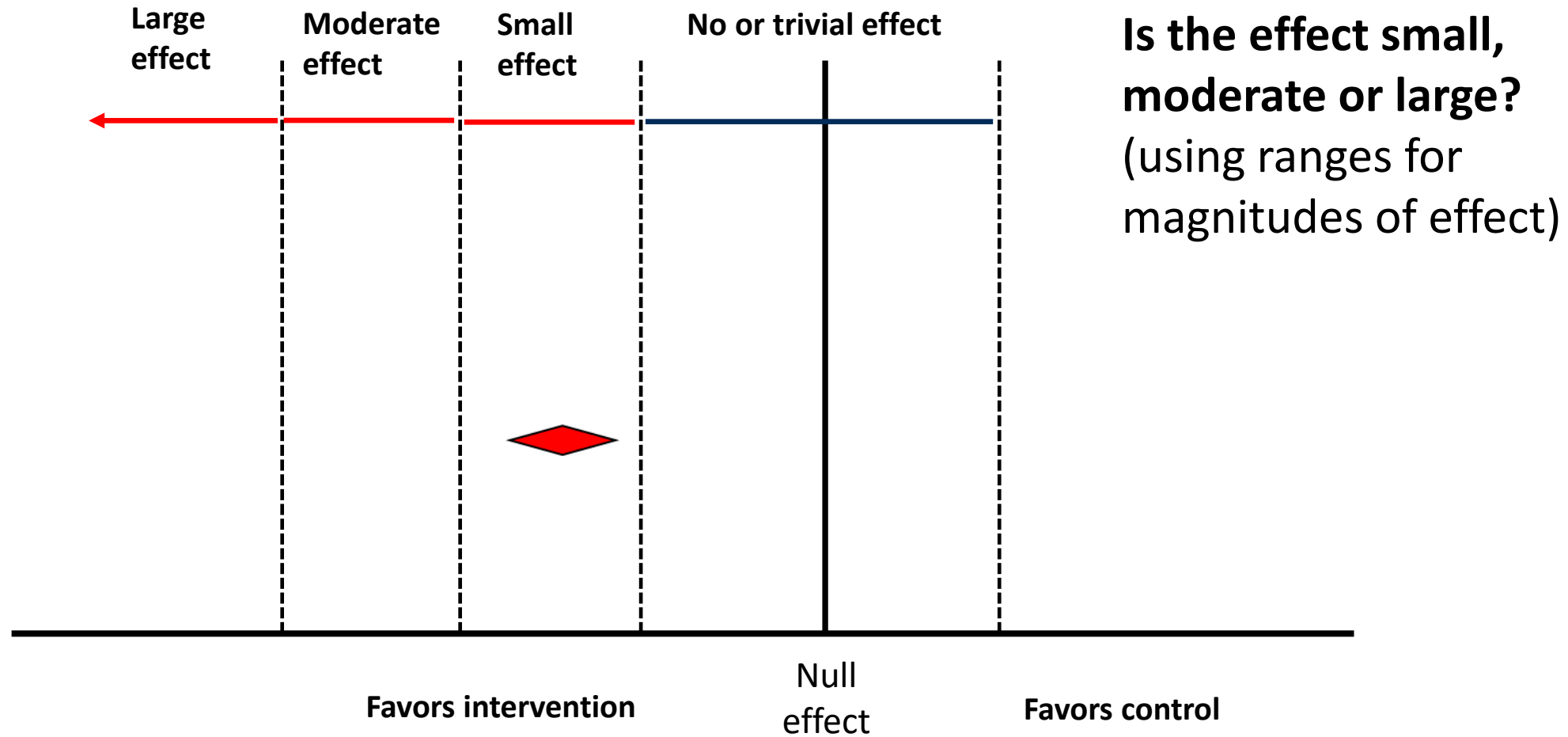
- The grading of recommendations assessment, development and evaluation (GRADE) Working Group clarifies that when rating certainty of the evidence for an individual outcome, we are rating how certain we are that the true effect lies within a particular range or on one side of a threshold.

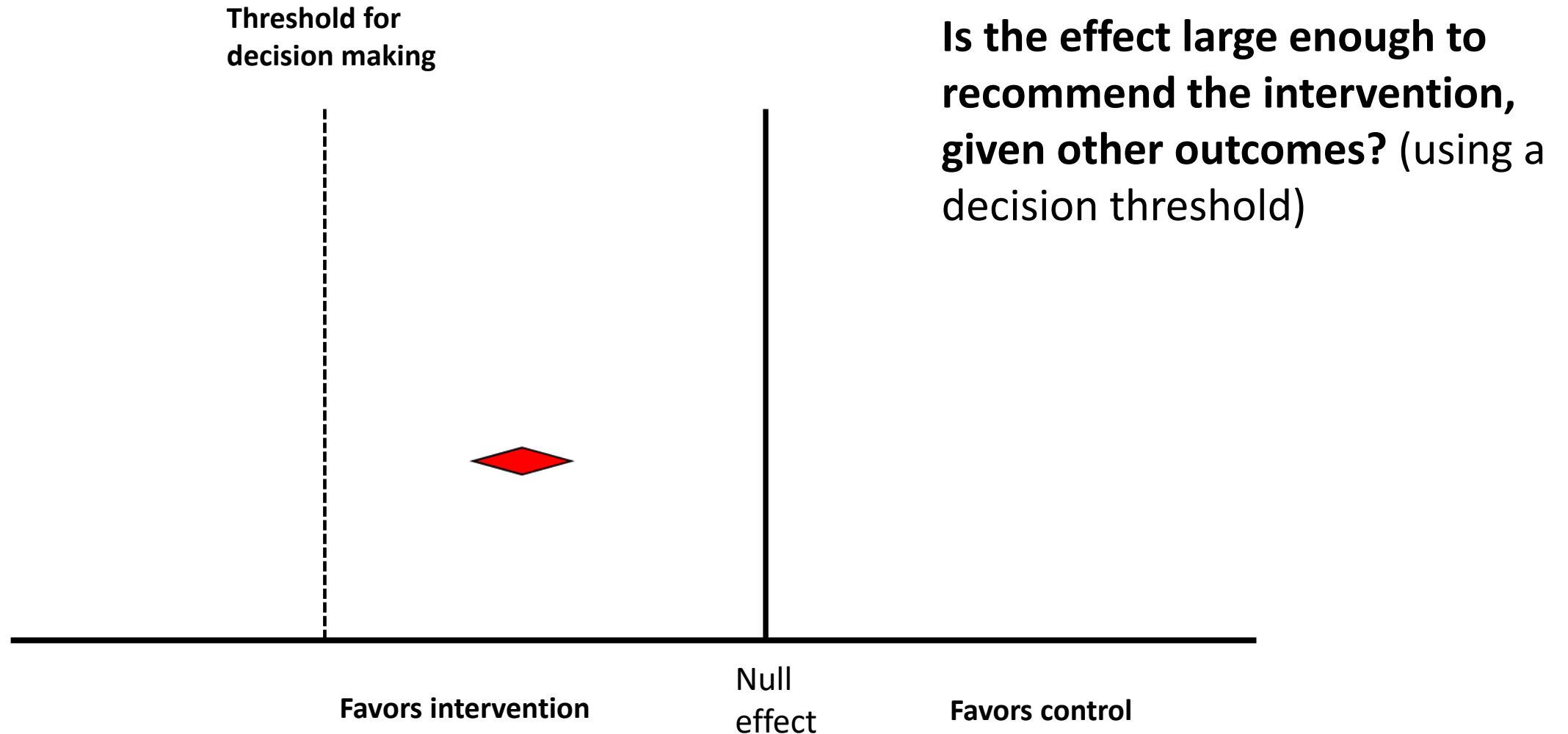
**Is there a true underlying effect?**  
(using the null threshold)





**Is there an effect that is important to patients?**  
(using the MIDs as threshold/range)







## Rating certainty of evidence

- GRADE users rate how certain they are that the true effect lies within a particular range or on one side of a threshold.
- It is important that authors of systematic reviews, health technology assessments, and guidelines specify the thresholds or ranges they are using.
- The thresholds/ranges were initially categorized into different degrees of “contextualization” – non- , partially-, and fully contextualized

## GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework

Romina Brignardello-Petersen,<sup>1</sup> Ivan D Florez,<sup>1,2</sup> Ariel Izcovich,<sup>3</sup> Nancy Santesso,<sup>1</sup> Glen Hazlewood,<sup>4</sup> Waleed Alhazanni,<sup>1</sup> Juan José Yepes-Nuñez,<sup>5</sup> George Tomlinson,<sup>6,7</sup> Holger J Schünemann,<sup>1</sup> Gordon H Guyatt,<sup>1</sup> on behalf of the GRADE working group

Cite this as: *BMJ* 2020;**371**:m3900  
<http://dx.doi.org/10.1136/bmj.m3900>

## GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework

Romina Brignardello-Petersen,<sup>1</sup> Ariel Izcovich,<sup>2</sup> Bram Rochwerf,<sup>1</sup> Ivan D Florez,<sup>1,3</sup> Glen Hazlewood,<sup>4</sup> Waleed Alhazanni,<sup>1</sup> Juan Yepes-Nuñez,<sup>5</sup> Nancy Santesso,<sup>1</sup> Gordon H Guyatt, Holger J Schünemann,<sup>1,6</sup> on behalf of the GRADE working group

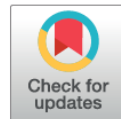
A minimally contextualised framework minimises value judgments regarding the magnitude of intervention effects. A partially contextualised approach will involve making such judgments.<sup>9</sup> This article focuses on the minimally contextualised approach.

# Problems with categories of contextualization

- Difficult concept, non-intuitive
- Papers with different terminologies
- Not only an issue within GRADE for treatment interventions



ELSEVIER



Journal of Clinical Epidemiology 165 (2024) 111189

**Journal of  
Clinical  
Epidemiology**

### GRADE GUIDANCE SERIES

## GRADE guidance 37: rating imprecision in a body of evidence on test accuracy

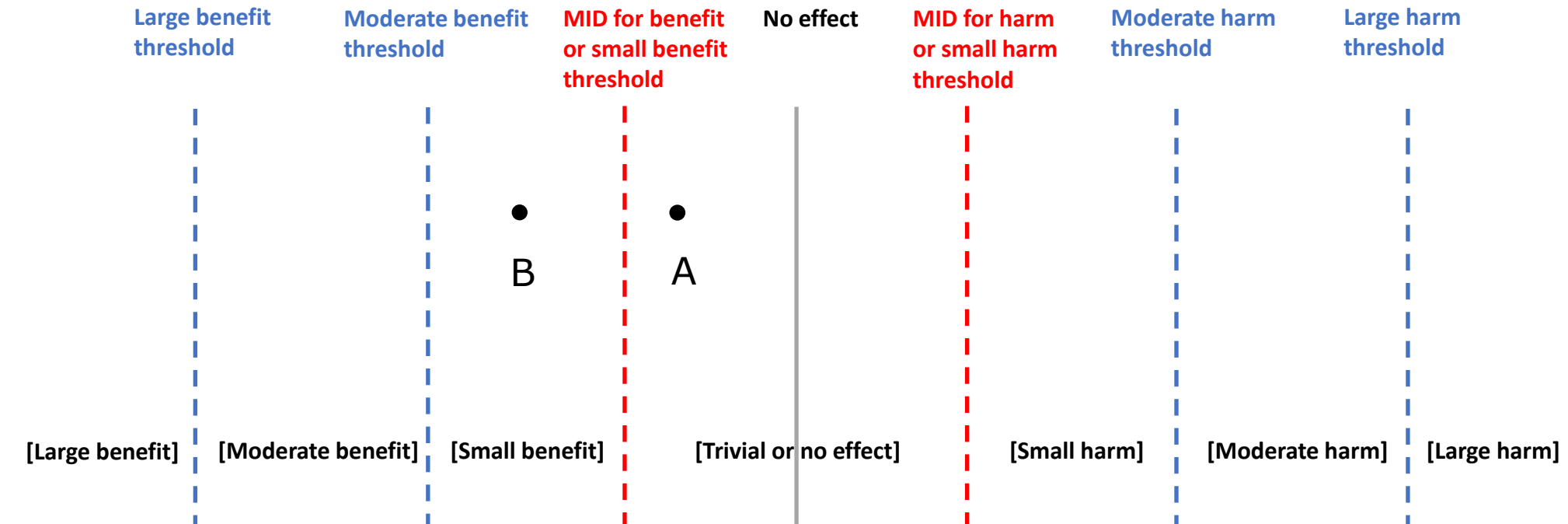
Reem A. Mustafa<sup>a,b,\*</sup>, Ibrahim K. El Mikati<sup>c</sup>, M. Hassan Murad<sup>d</sup>, Monica Hultcrantz<sup>e</sup>,  
Karen R. Steingart<sup>f</sup>, Bada Yang<sup>g,h</sup>, Mariska M.G. Leeflang<sup>i</sup>, Elie A. Akl<sup>b,j</sup>, Philipp Dahm<sup>k</sup>,  
Holger J. Schünemann<sup>b,l,m,n</sup>

### **Box 2 Rationale for using judgment threshold(s) instead of discussing the level of contextualization**

...[for multiple reasons], when addressing the question of imprecision in test accuracy studies, and because contextualized decisions are required, **instead of discussing the concept as minimally or partially contextualized approaches, the focus should be on establishing judgment threshold(s).** This may be one or more judgment thresholds, depending on the setting.

# Solution

- Drop contextualization categories
- Refer only to thresholds and targets
- Note: no change in underlying concepts



Additional threshold: Net benefit

# What determines the choice of threshold/s?

- If MID or magnitude of effect thresholds are available and reliable – first choice
- If not, GRADE user must consider their remit and target audience

# When might choose the null?

- Minimize value and preference judgments
- Perception: establishing thresholds challenging, time-consuming
- Beyond remit, may not seem themselves as right group to collect evidence or interpret
- As a first step in a complex review
- Early in the process going from evidence to decision



# Minimal important difference (MID)

- Smallest difference in an outcome people consider important
  - People in general, or particular population
- Value and preference judgment
  - May be influenced by e.g. age, gender and prior experience

# Ranges of magnitude of effect

- Useful for a fully structured Evidence to Decision framework (EtD)
  - Require large/moderate/small/trivial
  - All important benefits and harms
- Ratings of individual outcomes informs judgments across all benefits and harms

# If selecting thresholds for magnitude of effect

- consider new GRADE Guidance

Focus on 3 approaches, categorized as:

## Research-based

- Approach 1: using empirically derived generic coefficients as DTs
- Approach 2: calculating utility adjusted risk difference DTs or considering range of outcome-specific thresholds, matching to the given outcome for decision-making

## Expert (evidence)-based

- Approach 3: DTs obtained from surveying decision-makers (e.g. GDG) to directly estimate thresholds, prior guidelines

## Mix of approaches (triangulation)

- Using approach 3 with one of the two approaches

# Net benefit

- Threshold between overall benefit vs harm
- GRADE concept paper describes approach to arriving at threshold
- Implicit judgments of certainty
  - Strong recommendation high certainty
  - Conditional/weak low certainty
- Useful concept, practical application for certainty judgments ongoing work

Open access

Communication

**BMJ Open** Defining certainty of net benefit: a GRADE concept paper

Brian S Alper,<sup>1,2</sup> Peter Oettgen,<sup>1,3</sup> Ilkka Kunnamo,<sup>4,5</sup> Alfonso Iorio,<sup>6</sup> Mohammed Toseef Ansari,<sup>7</sup> M Hassan Murad,<sup>8</sup> Joerg J Meerpohl,<sup>9,10</sup> Amir Qaseem,<sup>11</sup> Monica Hultcrantz,<sup>12,13</sup> Holger J Schünemann,<sup>14</sup> Gordon Guyatt,<sup>14</sup> on behalf of The GRADE Working Group

## In summary

- Certainty ratings represent how certain we are that the true effect lies within a particular range or on one side of a threshold
- Important to specify the threshold/s used and target of certainty rating
- GRADE will stop referring to levels of contextualization – an important change in terminology, but not in essential concepts
- In choosing threshold/s, consider whether reliable thresholds exist as well as remit and target audience

**Thank you!**

[Monica.Hultcrantz@regionstockholm.se](mailto:Monica.Hultcrantz@regionstockholm.se)

[www.nordicgradenetwork.org](http://www.nordicgradenetwork.org)

[GRADE home](#)

**GRADE** working group

Nordic  
**GRADE**  
Network



# References

1. Hultcrantz M , Rind D , Akl EA , et al. *The GRADE Working Group clarifies the construct of certainty of evidence*. J Clin Epidemiol 2017;87:4–13.
2. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. *GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings*. J Clin Epidemiol 2021;137:163-175
3. Wiercioch W, Morgano GP, Piggott T, et al. GRADE Guidance: Using thresholds for judgments on health benefits and harms in decision-making. Submitted for publication.
4. Alper BS, Oettgen P, Kunnamo I, et al. Defining certainty of net benefit: a GRADE concept paper. BMJ Open. 2019;9(6)