



LUND UNIVERSITY

Challenge of communicating uncertainty in systematic reviews when applying GRADE ratings

Anttila, Sten; Persson, Johannes; Vareman, Niklas; Sahlin, Nils-Eric

Published in:
Evidence-Based Medicine

DOI:
[10.1136/bmjebm-2018-110894](https://doi.org/10.1136/bmjebm-2018-110894)

2018

Document Version:
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):
Anttila, S., Persson, J., Vareman, N., & Sahlin, N.-E. (2018). Challenge of communicating uncertainty in systematic reviews when applying GRADE ratings. *Evidence-Based Medicine*. <https://doi.org/10.1136/bmjebm-2018-110894>

Total number of authors:
4

Creative Commons License:
CC BY-NC

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



OPEN ACCESS

Challenge of communicating uncertainty in systematic reviews when applying GRADE ratings

Sten Anttila,¹ Johannes Persson,² Niklas Vareman,³ Nils-Eric Sahlin³

10.1136/bmjebm-2018-110894

¹Medical Ethics VBE, Lund Universitet, Lund, Sweden

²Department of Philosophy, Lund Universitet, Lund, Sweden

³Department of Medical Ethics, Lund Universitet, Lund, Sweden

Correspondence to:

Dr Sten Anttila, Medical Ethics VBE, Lund Universitet, Lund, Sweden; anttila@sbu.se

One of the most widely used tools for assessing and communicating scientific uncertainty is Grading of Recommendations Assessment, Development, and Evaluation (GRADE), a system for rating the quality of evidence and grading strength of recommendations in healthcare. More than 100 organisations around the world—WHO included¹—are using GRADE or have endorsed it.

In GRADE, a quantitative assessment of uncertainty is qualitatively communicated, so that a result obtained as a CI relative to a threshold is expressed as a finding in which assessors have low, moderate or high certainty, or certainty described with other such qualifiers. What these correspond to in quantitative terms, and how decision-makers interpret them, is our issue here. We confine our attention to GRADE's decision rules for systematic reviews, and do not comment on the problem of multiple outcomes in guideline recommendations.

In a recent guideline article,² GRADE introduced an idea that appears to undermine sound statistical reasoning in systematic reviews: the idea is that a result that is statistically inconclusive because the null hypothesis cannot be ruled out³ is converted into 'moderate certainty'. We fear that, applied as a principle, this GRADE guideline may jeopardise patient health.

What is a statistically inconclusive result? Suppose the potential harm of a treatment is tested. A threshold is set above which the harm is clinically relevant. A confidence level is chosen that reflects how the consequences of erroneous inferences are weighted. If the harm is serious, the level may be 99%, with 1% error risk. If the

harm is less serious, a 95% or a 90% level might be chosen. Then, if the interval estimate includes the threshold, the possibility of harm cannot be excluded. The result is inconclusive given the research question and given the chosen confidence level. More generally, when a CI includes the clinically relevant threshold, the result is inconclusive³ (p 2596).

GRADE presents as an example a hypothetical case⁴ concerning the reduction of incidents of ischaemic stroke² (p 6). The choice of confidence level adopted by GRADE is 95%. The threshold of minimally relevant reduction is set at 1.0% absolute reduction in strokes to reflect the harm associated with the treatment. The resulting interval estimate is 0.6%–2.0%. This means that the threshold is clearly included in the GRADE example. Notwithstanding this, the conclusion of the Grade Working Group (p 7) is the following:

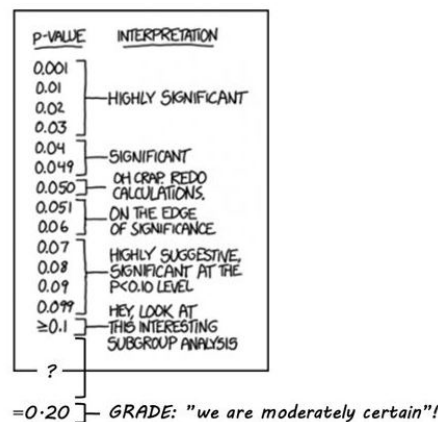
Because the point estimate of 1.3% meets the threshold criterion... the imprecision-generated uncertainty will result in... moderate certainty that the [‘true’] effect is above the threshold [1.0].

In effect, GRADE is downplaying the importance of a prespecified α -level in a protocol by applying the idea that any null hypothesis (threshold) will be rejected to some degree, provided that the point estimate lies on the preferred side of the null hypothesis. This flexibility might be appreciated by guideline developers as well as by stakeholders, but it may also undermine the transparency of the process of the systematic review.

This means that ‘inconclusive’ is converted into ‘moderate certainty’ when GRADE is used. For this specific result to be conclusive, the confidence level must be lowered to less than 80%.ⁱ The corresponding p valueⁱⁱ is 0.20 in a one-sided test and 0.40 in a two-sided test. GRADE's latest stipulation of the meaning of ‘moderate’⁵ is that the ‘true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different’ (p 404). However, in everyday language a common

ⁱGiven a Z distribution, the SE is approximated $0.36 \approx (1.3 - 0.6) / 1.96$.

ⁱⁱThe null hypothesis concerns the threshold $H_0: \leq 1.0$, why $Z = 0.83 \approx (1.3 - 1.0) / 0.36$. A Z value of ± 0.83 divides the probability density function into three areas: 0.20 and 0.60 and 0.20. In a one-sided test, the p value is 0.20, and in a double-sided test it is $0.20 + 0.20 = 0.40$. A Z value of 1.96 divides the PDF into the following familiar areas: 2.5%, 95% and 2.5%.



understanding of 'moderate' is 'within reasonable limits'. If the idea of converting statistically inconclusive results into 'moderate certainty' is understood as a principle, some systematic reviews using GRADE may unintentionally mislead, since it cannot be assumed that users will interpret 'moderate' in accordance with GRADE's stipulation ([Figure 1](#)).

Contributors All authors contributed to the planning. SA wrote the first draft. All other authors contributed equally.

Funding This work is supported by a grant from the Swedish Foundation for Humanities and Social Sciences, grant number M14-0138:1.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non

Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

References

1. WHO handbook for guideline development. 2nd ed. www.who.int/en/
2. Hulterantz 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.
3. Piaggio G, Elbourne DR, Pocock SJ, *et al.* Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:11.
4. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol* 2011;64:1283-93.
5. Balshem H, Helfand M, Schünemann HJ, *et al.* GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-6.

Challenge of communicating uncertainty in systematic reviews when applying GRADE ratings

Sten Anttila, Johannes Persson, Niklas Vareman and Nils-Eric Sahlin

BMJ EBM published online April 12, 2018

Updated information and services can be found at:
<http://ebm.bmj.com/content/early/2018/04/11/bmjebm-2018-110894>

These include:

References

This article cites 3 articles, 0 of which you can access for free at:
<http://ebm.bmj.com/content/early/2018/04/11/bmjebm-2018-110894#ref-list-1>

Open Access

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections

[Open access](#) (28)

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>