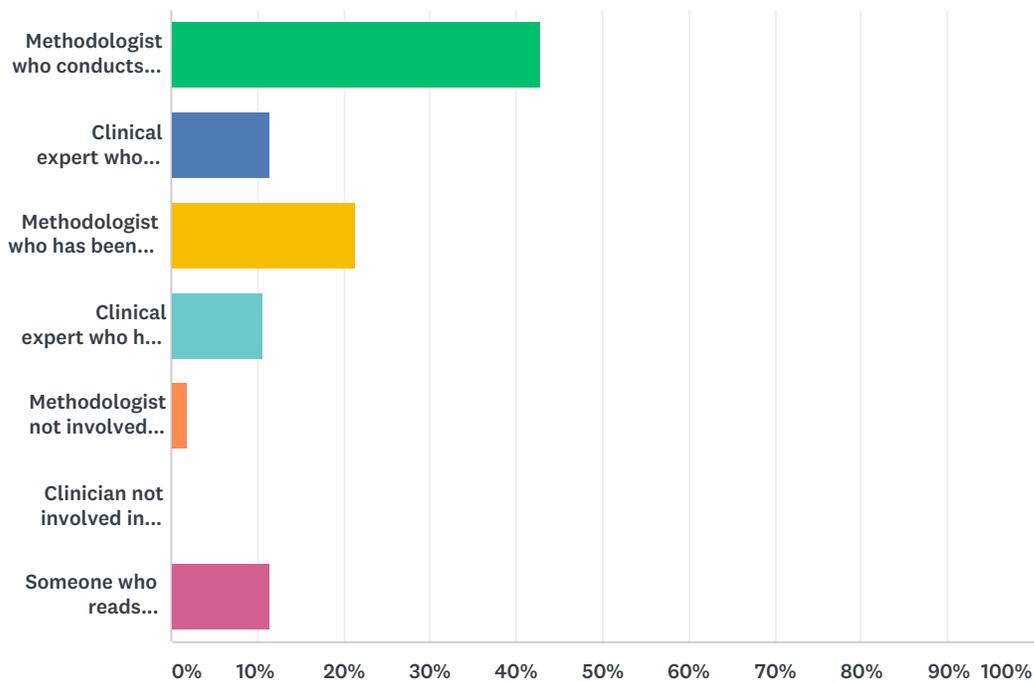


APPENDIX 2: Results of the survey of participants (not including GRADE project group members)

Q1 What is your primary role related to systematic reviews and guidelines? (Select one that best represents your role)

Answered: 112 Skipped: 0



| ANSWER CHOICES | RESPONSES | |
|--|-----------|------------|
| Methodologist who conducts systematic reviews | 42.86% | 48 |
| Clinical expert who conducts systematic reviews | 11.61% | 13 |
| Methodologist who has been involved in guideline development | 21.43% | 24 |
| Clinical expert who has been involved in guideline development | 10.71% | 12 |
| Methodologist not involved in systematic reviews or guidelines | 1.79% | 2 |
| Clinician not involved in systematic reviews or guidelines | 0.00% | 0 |
| Someone who reads systematic reviews | 11.61% | 13 |
| TOTAL | | 112 |

Q2 What is your education in epidemiology?

Answered: 87 Skipped: 25

| # | RESPONSES | DATE |
|----|--|--------------------|
| 1 | MSc epidemiology | 4/12/2018 12:45 PM |
| 2 | Masters | 4/9/2018 5:02 PM |
| 3 | Minimal; a class or two. | 4/9/2018 9:39 AM |
| 4 | little, basic courses | 4/9/2018 4:36 AM |
| 5 | MSc Doctor of Medicine [UK] | 4/8/2018 8:02 PM |
| 6 | undergraduate level formal teaching as well as experience and knowledge gained in the workplace. | 4/8/2018 3:13 AM |
| 7 | MSc | 4/6/2018 12:10 PM |
| 8 | no formal education | 4/6/2018 5:11 AM |
| 9 | Masters | 4/5/2018 3:26 PM |
| 10 | PhD | 4/5/2018 12:19 PM |
| 11 | MSc Epidemiology | 4/4/2018 8:08 AM |
| 12 | One post-graduate level paper | 4/4/2018 4:43 AM |
| 13 | Post graduate master epidemiology. | 4/4/2018 4:41 AM |
| 14 | PhD in statistical modelling | 4/4/2018 3:31 AM |
| 15 | One year training in clinical epidemiology. Autodidact learning and teaching. Participating in guideline development and implementation and in knowledge transfer. | 4/3/2018 9:51 PM |
| 16 | phd | 4/3/2018 9:11 AM |
| 17 | Master in Public Health | 4/3/2018 8:40 AM |
| 18 | One semester Phd course. | 4/3/2018 6:35 AM |
| 19 | None | 4/3/2018 3:26 AM |
| 20 | As part of MPH degree | 4/3/2018 3:25 AM |
| 21 | PhD | 4/2/2018 11:28 PM |
| 22 | PhD | 4/2/2018 2:50 PM |
| 23 | PhD epidemiology and biostats | 4/2/2018 2:11 PM |
| 24 | MSc in Clinical Epidemiology | 4/2/2018 7:35 AM |
| 25 | MSc | 4/2/2018 2:41 AM |
| 26 | none | 4/1/2018 7:14 AM |
| 27 | Diploma in evidence based healthcare | 3/31/2018 10:30 PM |
| 28 | basic | 3/31/2018 3:28 PM |
| 29 | MSc Clinical Epidemiology and Biostatistics | 3/30/2018 5:34 PM |
| 30 | MSc Clinical Epidemiology | 3/30/2018 11:55 AM |
| 31 | Masters in Public Health | 3/29/2018 6:32 PM |
| 32 | Master of Public Health from Columbia University | 3/29/2018 4:47 PM |
| 33 | Not sure how to answer this. Do you mean how many courses we had as master's and doctoral students we have had? I have had training in classical epidemiology and clinical epidemiology. | 3/29/2018 4:05 PM |

How to communicate the results of systematic reviews using standard statements

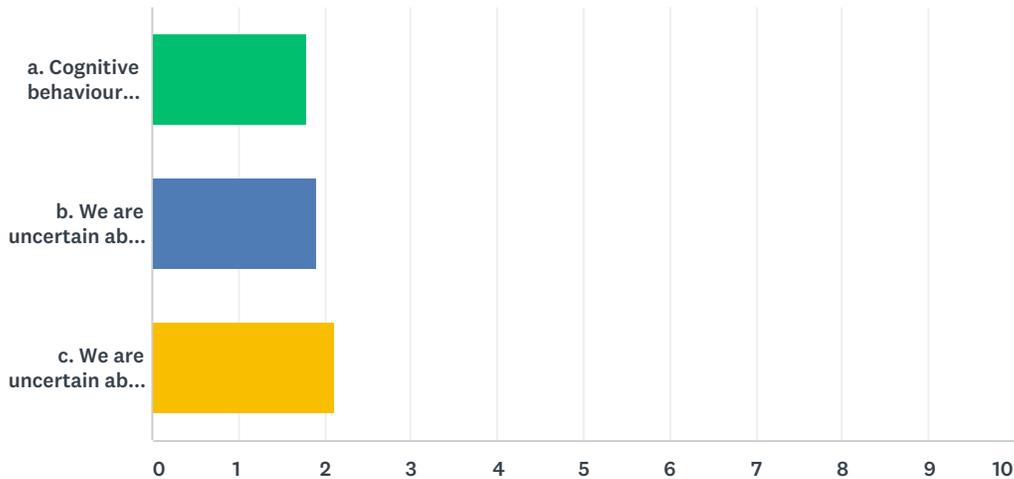
| | | |
|----|---|--------------------|
| 34 | None. Masters in psychology. | 3/29/2018 2:34 PM |
| 35 | 15 hrs online course | 3/29/2018 1:42 PM |
| 36 | No formal | 3/29/2018 12:35 PM |
| 37 | PhD in nursing | 3/29/2018 12:15 PM |
| 38 | Master in Public Health | 3/29/2018 9:18 AM |
| 39 | Several courses as part of education. | 3/29/2018 8:22 AM |
| 40 | Phd | 3/29/2018 8:08 AM |
| 41 | I have a PhD in epidemiology and teach epidemiology for medical students. | 3/29/2018 7:34 AM |
| 42 | PhD | 3/29/2018 6:19 AM |
| 43 | during my PhD I became MSc in the epidemiology within a two year programma | 3/29/2018 6:01 AM |
| 44 | Undergraduate level. I now work in communication of data. | 3/29/2018 5:40 AM |
| 45 | None. | 3/29/2018 2:04 AM |
| 46 | PhD, with over 10 years experience following | 3/28/2018 9:59 PM |
| 47 | None formally. 20 years of experience with Cochrane | 3/28/2018 7:03 PM |
| 48 | Master in clinical and translational science | 3/28/2018 6:51 PM |
| 49 | MSc | 3/28/2018 4:31 PM |
| 50 | msc | 3/28/2018 4:12 PM |
| 51 | Training in Clinical epidemiology 450 h | 3/28/2018 4:05 PM |
| 52 | Masters in Health Science specialising in critical appraisal of clinical practice guidelines | 3/28/2018 1:29 PM |
| 53 | MSc Health Science. Post-grad training in population health. | 3/28/2018 1:22 PM |
| 54 | PhD in public health | 3/28/2018 11:44 AM |
| 55 | Master in Public Health | 3/28/2018 8:53 AM |
| 56 | PhD | 3/28/2018 8:26 AM |
| 57 | MSc and PhD in public health | 3/28/2018 8:24 AM |
| 58 | none | 3/28/2018 7:55 AM |
| 59 | PhD training; methodologist who conducts SR and participates in guidelines development, and has clinical training | 3/28/2018 7:47 AM |
| 60 | minimal, I am a physician specialist | 3/28/2018 7:03 AM |
| 61 | No fomal education, but have undertaken postgraduate short courses in epidemiology, various analysis techniques, systematic review and meta-analysis courses and have been working in an epidemiological department for 20 years. | 3/28/2018 6:57 AM |
| 62 | Masters and PhD in medical statistics | 3/28/2018 6:23 AM |
| 63 | registration epidemiologist B | 3/28/2018 4:06 AM |
| 64 | No degrees, just self-readings | 3/27/2018 9:41 PM |
| 65 | I'm an M.D. | 3/27/2018 8:58 PM |
| 66 | Masters degree | 3/27/2018 4:48 PM |
| 67 | I have studied epidemiology in medical school | 3/27/2018 4:47 PM |
| 68 | None | 3/27/2018 4:30 PM |
| 69 | No education in epidemiology. My education is a BScN, and a MA in Psychology. | 3/27/2018 2:49 PM |
| 70 | No formal education | 3/27/2018 2:45 PM |
| 71 | It was a course in my MPH degree | 3/27/2018 2:05 PM |
| 72 | MPH in Epidemiology | 3/27/2018 1:48 PM |

How to communicate the results of systematic reviews using standard statements

| | | |
|----|--|--------------------|
| 73 | masters in health studies | 3/27/2018 12:47 PM |
| 74 | MSc | 3/27/2018 11:26 AM |
| 75 | MD and MPH | 3/27/2018 10:06 AM |
| 76 | No formal education. | 3/27/2018 9:29 AM |
| 77 | Masters degree in public health | 3/27/2018 8:20 AM |
| 78 | Masters level | 3/27/2018 7:16 AM |
| 79 | MSc | 3/27/2018 6:27 AM |
| 80 | Master in health economics | 3/27/2018 3:42 AM |
| 81 | Master level | 3/27/2018 3:10 AM |
| 82 | Medical school, and later personal education in relevant literature | 3/26/2018 3:35 PM |
| 83 | None | 3/26/2018 12:16 PM |
| 84 | None formal - work experience assisting with population research | 3/26/2018 11:02 AM |
| 85 | medical sociologist, undergraduate psychology, PHD in Survey design, 12 years IQWiG. | 3/26/2018 10:17 AM |
| 86 | no specific education | 3/26/2018 10:10 AM |
| 87 | Limited | 3/26/2018 9:51 AM |

Q3 A systematic review compared the effects of cognitive behavioural therapy versus a waiting list for military suffering from post-traumatic stress disorder on depression. It found that cognitive behaviour therapy reduced depression by 8 points more on a scale from 1-100 (95% confidence interval from 21 point reduction to 12 point increase). This reduction is small but important. The evidence came from a meta-analysis with very few people (91) and very serious concern that the studies were at high risk of bias because of unclear randomisation and large loss to follow-up. The conclusion about the effect of cognitive behaviour therapy could be worded in the following three ways. Please indicate the acceptability of each statement.

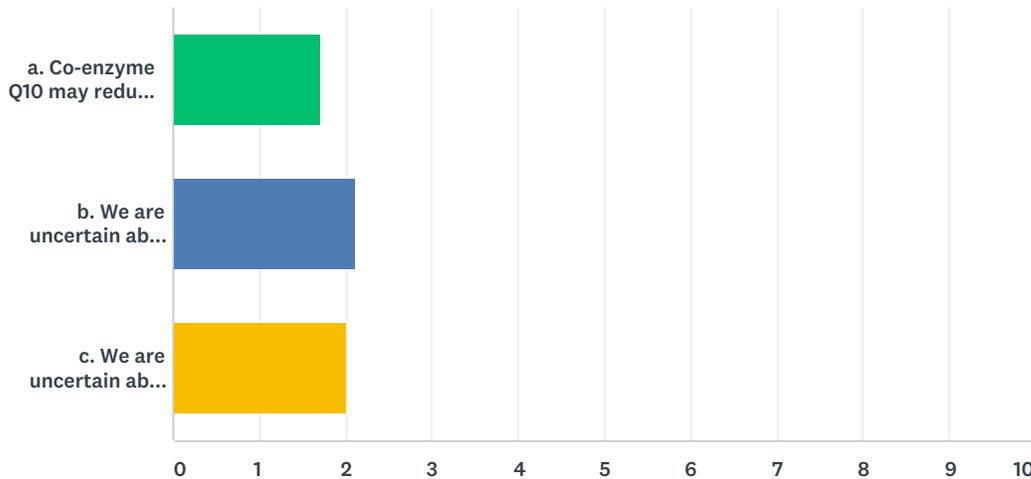
Answered: 91 Skipped: 21



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|--------------|--------------|-------|------------------|
| a. Cognitive behaviour therapy may reduce depression slightly more than no therapy but we are uncertain. | 36.67% 33 | 47.78% 43 | 15.56% 14 | 90 | 1.79 |
| b. We are uncertain about the effect of cognitive behaviour therapy on depression. | 28.57% 26 | 51.65% 47 | 19.78% 18 | 91 | 1.91 |
| c. We are uncertain about whether cognitive behaviour therapy reduces depression more than no therapy. | 14.29% 13 | 60.44% 55 | 25.27% 23 | 91 | 2.11 |

Q4 The conclusion about the effect of co-enzyme Q10 on blood pressure could be worded in the following three ways. Please indicate the acceptability of each statement.

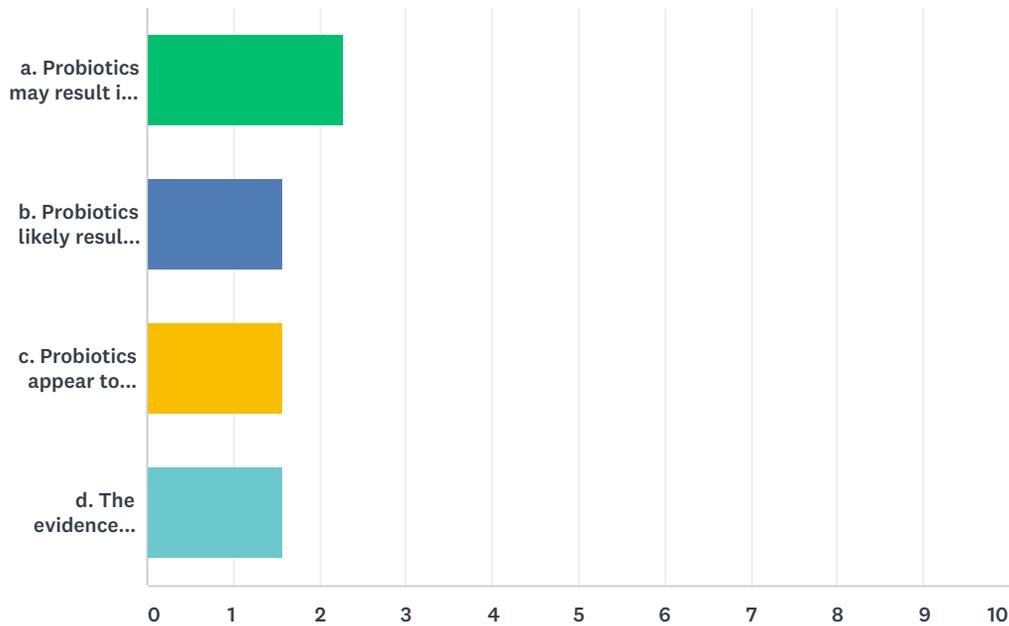
Answered: 90 Skipped: 22



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|---|--------------|--------------|--------------|-------|------------------|
| a. Co-enzyme Q10 may reduce blood pressure slightly but we are uncertain. | 46.07% 41 | 35.96% 32 | 17.98% 16 | 89 | 1.72 |
| b. We are uncertain about the effect of co-enzyme Q10 on blood pressure | 15.56% 14 | 56.67% 51 | 27.78% 25 | 90 | 2.12 |
| c. We are uncertain about whether co-enzyme Q10 reduces blood pressure. | 16.67% 15 | 65.56% 59 | 17.78% 16 | 90 | 2.01 |

Q5 The authors of the review considered that the cut-off for a large effect is RR 0.60. Please indicate the acceptability of the statements below.

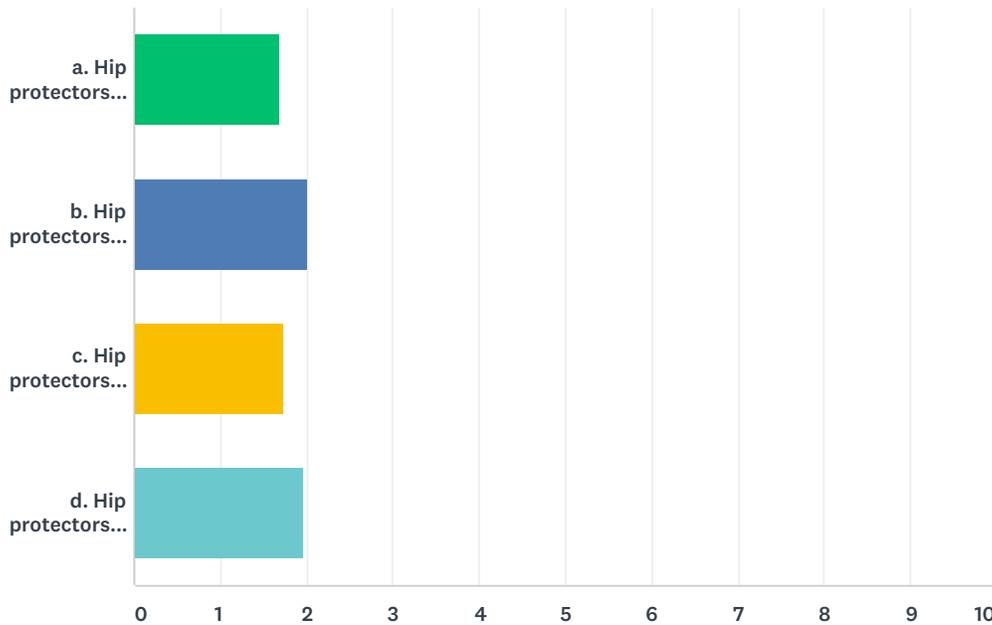
Answered: 91 Skipped: 21



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|--------------|--------------|-------|------------------|
| a. Probiotics may result in a large reduction in the incidence of diarrhea. | 10.11% 9 | 52.81% 47 | 37.08% 33 | 89 | 2.27 |
| b. Probiotics likely result in a large reduction in the incidence of diarrhea. | 51.69% 46 | 39.33% 35 | 8.99% 8 | 89 | 1.57 |
| c. Probiotics appear to result in a large reduction in the incidence of diarrhea. | 50.00% 45 | 43.33% 39 | 6.67% 6 | 90 | 1.57 |
| d. The evidence suggests that probiotics result in a large reduction in the incidence of diarrhea. | 56.67% 51 | 31.11% 28 | 12.22% 11 | 90 | 1.56 |

Q6 The authors indicate that the effect found was less than their cut-off for an effect. Please indicate the acceptability of the statements to communicate the effects of hip protectors compared to no hip protectors on the number of hip fractures.

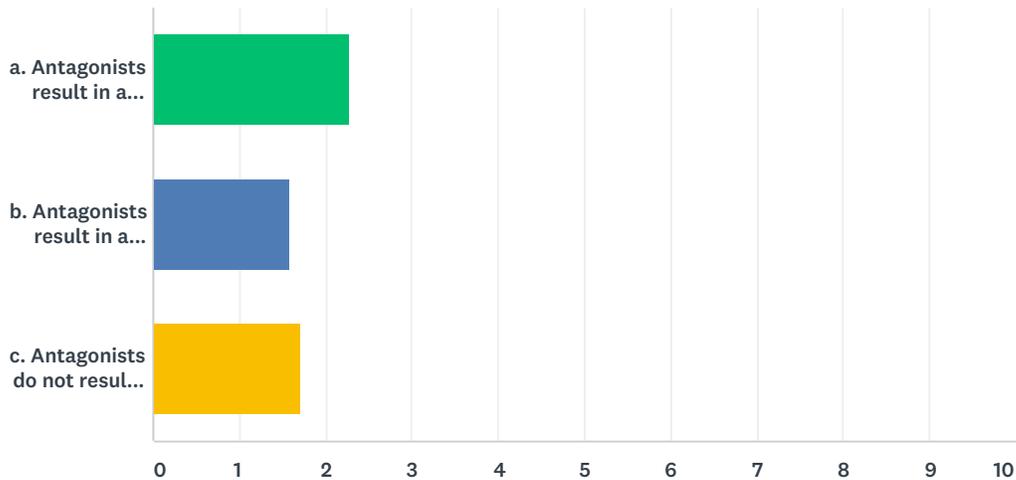
Answered: 90 Skipped: 22



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|--------------|--------------|-------|------------------|
| a. Hip protectors likely do not reduce hip fractures. | 44.94% 40 | 40.45% 36 | 14.61% 13 | 89 | 1.70 |
| b. Hip protectors likely result in little to no difference in hip fractures. | 23.33% 21 | 52.22% 47 | 24.44% 22 | 90 | 2.01 |
| c. Hip protectors probably do not reduce hip fractures. | 39.33% 35 | 47.19% 42 | 13.48% 12 | 89 | 1.74 |
| d. Hip protectors probably result in little to no difference in hip fractures. | 21.35% 19 | 61.80% 55 | 16.85% 15 | 89 | 1.96 |

Q7 Please indicate the acceptability of the statements below to communicate the effect of antagonists compared to placebo on daytime nasal symptoms. Note: the authors considered the SMD and confidence interval a small but not important effect.

Answered: 92 Skipped: 20



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|---|--------------|--------------|--------------|-------|------------------|
| a. Antagonists result in a small effect that may not be an important reduction in daytime symptoms. | 15.22% 14 | 42.39% 39 | 42.39% 39 | 92 | 2.27 |
| b. Antagonists result in a small possible unimportant reduction in daytime symptoms. | 49.45% 45 | 41.76% 38 | 8.79% 8 | 91 | 1.59 |
| c. Antagonists do not result in an important reduction in daytime symptoms. | 45.05% 41 | 38.46% 35 | 16.48% 15 | 91 | 1.71 |

Q8 Please take 5 minutes to review this list of options for statements to communicate results at various levels of evidence and size of effect. If you'd like, please provide any general comments about the statements below.

Answered: 39 Skipped: 73

| # | RESPONSES | DATE |
|---|---|-------------------|
| 1 | Suggestion not important effect: small clinically not relevant reduction For example: High/not important: results in a small clinically not relevant reduction | 4/12/2018 3:05 PM |
| 2 | Question 5b in this survey was worded slightly different than below in a way that affected my answer: "small possible unimportant" instead of "small possibly unimportant". I think the addition of "possibly unimportant" is very helpful, but found it confusing the way it was worded in Question 5b. I would add commas: "small, possibly unimportant,..." so the grouping of these words is clear. No effect/high certainty: you can't say that something doesn't have effect. This is just wrong. Small effect important/moderate certainty: I find the word slightly to be a bit confusing. Your mind has to keep track of probably, and direction (reduces) and then amount. I would use "to some degree" instead of "slightly". "probably" works much better for me than "likely", just because it is more familiar and uses less brain power to combine the meaning of it with the other terms. | 4/6/2018 5:32 AM |
| 3 | 1. Regarding 'no effect', I think we need to differentiate clearly between two different situations which seem to currently be treated as one and described as 'no effect': 1. where there is a tight confidence interval around the null/line of no effect and the CI is completely within threshold limits - in this case, we can say that there is little to no difference in the outcome. 2. where there is a large confidence interval that crosses the null/line of no effect AND one or both important threshold limits - in this case, I think it is misleading to state that there is "little or no difference" as we cannot confirm or exclude an important difference. In this case, should we call this an 'uncertain' effect? Or that the effect is compatible with both little or no effect and an important effect? It is more difficult to describe this in plain language, but I think it is important to convey the uncertainty of the effect that exists in the second case. 2. Re the wording, "does not reduce/may not reduce" etc, I think that given that there is always some uncertainty around the confidence interval, that the word "not" is too strong. Prefer to state 'little or no difference'. 3. The rest of the statements outlined below are reasonable. This is a very useful project which will help ensure correct interpretation in the results of Cochrane Reviews. Thank you! | 4/5/2018 1:08 PM |
| 4 | None of the statements specify a comparator. I think it would be useful to ensure that the comparator is always stated. For example, "Compared to Y, X results in a large reduction in mortality" or "X results in a large reduction in mortality compared to Y". Some of the statements include the word "evidence", while most do not. I think it would be useful to ensure that the word evidence appears in these statements to make clear they are based on evidence rather than opinion. For example, "The evidence shows that X results in a large reduction in mortality compared to Y". It is not clear to me why statements may be phrased using "probably" or "likely" and if there is a meaningful difference. I suggest picking one and discouraging use of the other. Some of the statements allude to the concept of certainty of evidence (using words like "may" and "appears"), while others make certainty explicit. I would favour making certainty explicit in all phrases. For example "The evidence shows we can be reasonably certain that X results in a large reduction in mortality compared to Y" or "The evidence shows we cannot be certain whether X reduces mortality compared to Y". The statements do not consider the fact that, while a given intervention may have little or no effect on average, it may have a large and important effect on a small number of people. They also do not consider the potential harms of an intervention. There may be interventions that are inexpensive and very safe, and which have little or no effect in most people, but do have large and important effects in some people. | 4/4/2018 4:31 AM |
| 5 | I like nuanced statements that provide precisions on which criteria are assessed (e.g. effect size, importance) and allow for uncertainty to express (e.g. likely results in little or no difference)... | 4/3/2018 10:30 PM |

How to communicate the results of systematic reviews using standard statements

| | | |
|----|--|--------------------|
| 6 | When you think about importance of an effect size, do you consider whether a small effect could be important if it applied across a whole population? I am asking because of a recent systematic review where we and our clinician experts judged the size of the effect to be "little to no effect", but others who read our review thought the size could be important if applied at a population level | 4/3/2018 9:21 AM |
| 7 | HIGH - Instead consider "X slightly reduces mortality" (small effect (important)) - For small effect (not important) -- "X does not result in an important reduction in mortality". The direction of the observed effect is missing, and this seems important as a small effect may be meaningful to some but not all. I prefer the statements in this category that include reference to the direction of effect. MODERATE - I don't feel that "likely" and "probably" are good synonyms for MODERATE evidence. Somehow I feel that "likely" is stronger and well suited here, but "probably" is less strong and therefore not as well suited. Perhaps "likely results in" and "seems to result in" could work if you want options? LOW - I feel that "appears to" belongs in the MODERATE category and not the LOW category. I like "may" and "the evidence suggests" though. - for the "no effect" category, "X may not reduce mortality" sounds like "X may or may not reduce mortality". I find it confusing and may introduce ambiguity about being uncertain about the direction of the effect VERY LOW - I feel that some qualifier as to where the uncertainty arises (i.e. low quality evidence) is needed. What do you think about "X may reduce mortality, but the quality of evidence (or studies) is very low"? or "X may reduce mortality, but we are very uncertain due to the low quality of the evidence (or studies)" | 4/2/2018 2:43 PM |
| 8 | Wordings are fine at HIGH or MODERATE certainty of the evidence, but problems arise at LOW level of evidence. For example 'X may result in a large reduction in mortality' raises the question whether the certainty of the evidence might be higher if 'large reduction' is replaced by 'reduction' (clinically relevant reduction). In my view the conclusion should relate to what is considered a clinically relevant effect, a clinically relevant reduction (or more general: the clinical decision threshold) i.e. 'large' should not be part of the statement. The level of evidence relates to the clinical decision threshold, usually the MID, and not to a 'larger' (or smaller) effect size. In principle this is also true at HIGH or MODERATE certainty of the evidence, but because of the higher certainty of the evidence, there is less of a problem. Similarly, I think that we should be careful about conclusions on 'small (unimportant)' effects: all of our judgements should concern clinically relevant effects, statements about 'unimportant' effects are confusing and may be misleading | 4/2/2018 8:45 AM |
| 9 | Regarding the statements of 'No effect' (HIGH, MODERATE and LOW certainty): well, I have learnt that you cannot keep the null hypothesis of no effect even if you cannot reject it. So you should not say 'does not reduce mortality' or 'little or NO difference'. When the effect crosses the line of 'no effect' - the results are uncertain and inconclusive. Very low certainty: I am not sure that 'may reduce mortality but we are uncertain' is a good idea. What about saying: "The results show a reduction in mortality, but the certainty of the evidence for this is/ is assessed as very low." Also, I think in general that the expression 'we are uncertain' sounds strange. Why are 'we' uncertain? It is either the results (crossing line of no effect) or very low certainty of evidence in that case that make us uncertain - so I think we should stress that it is the results/certainty of the evidence as assessed that are uncertain - not we. We are not uncertain of that the results/evidence are uncertain: "It is uncertain whether...", "the results are uncertain because of the possibility of either a reduction or an increase in ..." | 4/1/2018 1:59 PM |
| 10 | I think it is important to say x results in a large reduction...COMPARED to no intervention/placebo | 3/31/2018 7:21 AM |
| 11 | Overall fewer words are preferred to communicate essential findings of quality and effect size | 3/30/2018 12:23 PM |
| 12 | Prefer likely over probably. Wording "small possibly unimportant" is very confusing to lay readers and is cognitively taxing Use of "but we are uncertain" negates "may reduce [...]"; therefore no need for "may reduce" | 3/29/2018 9:25 AM |
| 13 | I am worried that the differences in wording between moderate/small important/small unimportant, with differences in the structure of sentences, use or non-use of adjective quantifiers (eg small vs slightly) and use or non-use of "important", may lead to heterogenous interpretation between these categories. Also, as always, there are issues with translating many of these quantifier terms. | 3/29/2018 8:26 AM |
| 14 | As non-native speaker the difference between likely and probably is hard to grasp. In the previous questions I felt that 'likely' is stronger than 'probably', but that might be personal. If you look these words up in a Dutch-English dictionary the translation is largely the same. In the table I see they're both used for Moderate CiE. Same for the difference between 'appears' and 'may', these words have a different feel for me. 'Appears' is: it could be but we might be completely wrong'. 'May' sounds more certain. This is really about language, and there may be differences between countries/languages. High CiE and small not-important effect: by adding 'may not be' and 'possibly' it seems that you have doubt about the importance of the effect, but as this is High CiE that is not the case, no? | 3/29/2018 7:07 AM |

How to communicate the results of systematic reviews using standard statements

| | | |
|----|---|--------------------|
| 15 | using the word 'appears' by LOW evidence feels too strong for me, I prefer 'may'. | 3/29/2018 6:15 AM |
| 16 | We are currently researching the communication of uncertainty in science, and use GRADE as an example of good practice because of the way it separates the different sources of uncertainty clearly - the calculated uncertainty around measurement is different from the quality of the evidence. It is also vital that the magnitude of the effect and the quality of the data are not confused (as they currently are in scales such as the IARC classifications of cancer risk, leading to statements like 'bacon in same cancer risk bracket as cigarettes' and the genomic evidence scales of 'pathogenic to benign'). I think it is important to give people as clear a view as possible of what data is currently available. However, I appreciate that for a simple statement of conclusions, such as in the preparation of guidelines, then it is important to give people a single sentence. I feel, though, that that sentence should overtly retain the distinction between quality of evidence and size of effect. So, in your examples below, I think they should be of the structure "There is strong/weak evidence that X results in a large/small effect". I don't think that there is anything to be gained in converting the phrases 'moderate evidence' to 'likely' - work on people's interpretation of words used in common parlance suggest that they lead to much more variable interpretation than keeping more precise language. | 3/29/2018 5:58 AM |
| 17 | Effects should not be communicated solely in relative terms with words like „large“, „small“ etc. Absolute effects should be added. In the probiotics example, the absolute effect is about 13 / 100. Is that large? | 3/29/2018 2:17 AM |
| 18 | The suggestion look good. I'm critical though with the choice of “slightly” for small but important effect. In your graduation of statements, it sounds too close to small unimportant instead than in between moderate an unimportant. My preference will be “small but important” difference. I have my preference on the choices provided, but I can live with all but this one. | 3/28/2018 7:24 PM |
| 19 | consistency is appreciated examples will be essential to develop consistency if these statements are for a range of audiences this should be considered | 3/28/2018 4:43 PM |
| 20 | Use appropriate comparisons for context (compared to Y); use plain language | 3/28/2018 2:06 PM |
| 21 | A very helpful tool! | 3/28/2018 1:44 PM |
| 22 | 'Appears to' and 'evidence suggests' seem too strong for low quality evidence. With those statements alone I would assume that there was certainty about the effects. Regarding the inclusion of 'important' in the interpretation - will end-users know what is meant by important? Could they use more elaboration about what is meant by an important difference? It is preferable in my opinion to include the proposed/theoretical direction of effect even if there is no effect. e.g., 'We are uncertain whether it reduces mortality' is preferable to 'we are uncertain about the effect on mortality'. May just be personal preference. | 3/28/2018 1:42 PM |
| 23 | word "appears" seems like a magical statement and not preferred in my opinion. it may be helpful to provide illustrative examples with number to contextualize some of these wording options. | 3/28/2018 12:11 PM |
| 24 | This makes sense and I agree in principle to the standardized wording. As long as you have a critical appraisal checklist whereby you can grade the certainty of the findings of the study according to very low, low, moderate and high then this system works; however not all checklists grade the quality eg CASP | 3/28/2018 9:47 AM |
| 25 | In general, the shortest text that fits the data is best (example: size of effect moderate: reduces mortality, NOT results in a reduction in mortality) if an effect is certain, small, but definitely not important, I would say: does not result in an important reduction in mortality, because clinical relevance is more important than "methodological significance". I prefer "may" over likely or probably; "suggest" is an acceptable alternative | 3/28/2018 8:16 AM |
| 26 | I like the use of certain words to communicate certain levels of certainty; however, I am unsure whether an end-user would distinguish between those words without having previously reviewed a chart like the one below -- for example, I am uncertain whether most clinicians would immediately pick up on the difference between may and likely. Also, for some reason, the word "appears" is not preferable (it may be that it isn't as commonly used in science-writing in our field; appears conjures up "magic" vs. some biologically plausible relationship) Lastly, there are some principals of plain language writing that could be integrated into some of the sample statements that would make them longer, but would also help with easier interpretation of the nuances. For example, I would prefer a statement line: X may reduce mortality; however, it is likely that the reduction is not clinically important. So, state the direction of the relationship first, and then follow with the disclaimer about clinical significance. | 3/28/2018 7:59 AM |

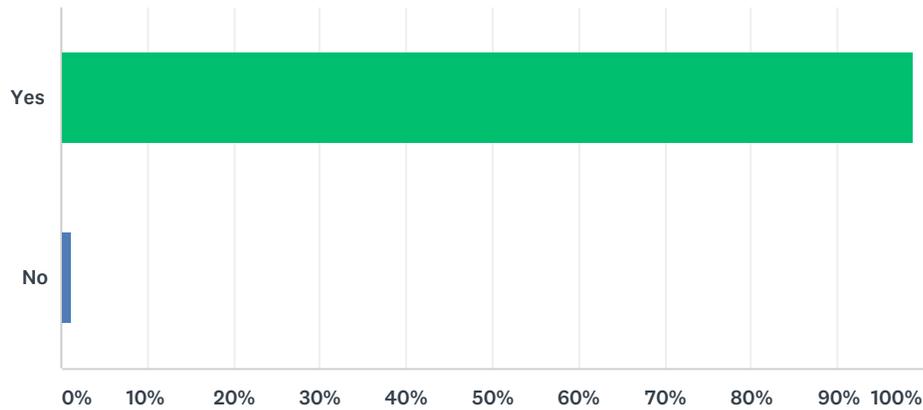
How to communicate the results of systematic reviews using standard statements

| | | |
|----|--|--------------------|
| 27 | Having submitted reviewsz with this wording..most editorial boards have no knowledge of GRADE consider this wording vague (comments received like "is it significant or not as probably and likely or possibly is not wording we are used to see in results, please be more specific...." There is still a long way before clinicians and peer reviewers, editors, guidelin emakers are familiar with GRADE and its terminology | 3/28/2018 7:16 AM |
| 28 | Generally I support the options aligned with the certainty levels (e.g. does, results in, will for HIGH). It's not clear why some constructs are not included. For example there is no "X reduces" for HIGH-Large, and no "X results in" for HIGH-Small. It seems that you could have options such as "X reduces mortality substantially" and "X results in a slight reduction in mortality" for these. I find some phrases too difficult to read and understand. Particularly "X results in a small possibly unimportant effect". There are too many words next to each other! | 3/28/2018 6:52 AM |
| 29 | - I'm not a native speaker, but 'appears' seems to suggest a little more than 'low' | 3/28/2018 4:20 AM |
| 30 | I like all the statements for high and moderate certainty. I know we need to be cautious, but the statements for low certainty are so vague that we run the risk that readers have no idea what we mean. I've had feedback to this extent on recent Cochrane reviews. | 3/27/2018 5:28 PM |
| 31 | I think it is a great idea to use such tool to standardize the way SR results are interpreted. | 3/27/2018 4:52 PM |
| 32 | The double negative of some statements can be confusing. For example: X likely results in a small effect that may not be an important reduction in mortality. I find for someone who is not well versed in research, that can cause confusion. However, many of the other statements are clear. | 3/27/2018 2:58 PM |
| 33 | I prefer the use of "probably" versus "likely" | 3/27/2018 2:11 PM |
| 34 | Language preference is for: HIGH: "small possible unimportant effect in mortality" rather than "does not results in an important reduction in mortality" and "results in little to no difference in mortality" rather than "does not reduce mortality". MODERATE: "likely" rather than "probably". LOW: "may result" rather than "appears to" | 3/27/2018 9:51 AM |
| 35 | Statements need to be simple. "X results in a small effect that may not be an important reduction in mortality" is too complicated a statement, for example - not to mention the equivalent statements for lower QoE. General comments: - For a critical binary outcome, especially mortality, is there a such a thing as an unimportant effect? How small would that be? - Why distinguish between an unimportant effect and no effect? If this is to do with the problem of averages and individual variance for continuous outcome measures (e.g. there is no important effect on average but some patients may have a substantial benefit/Response), this should be made explicit. | 3/27/2018 8:55 AM |
| 36 | I like to work with the standard phrasing as much as possible.It is great for abstracts and is as good as any way to integrate the size and certainty of effect. There are two situations I find troubling to work with: 1. Moderate quality evidence of moderate effect when you have downgraded for imprecision. I always want to describe this in terms of greater uncertainty than 'probably' or even 'may'. I actually end up working something like: 'We found moderate quality evidence that [intervention] reduces mortality by about 4% when compared with control, although we do not have enough data from the studies to rule out there being little or no effect/potential increase in risk of death (18% versus 14%; RR 0.73 (95% CI 0.34 to 1.2)' 2. Very low quality evidence and large/moderate effect when you have a large amount of data from the analyses: SMD 0.5 (0.2 to 0.8; 29 studies, 2300 participants). I don't always think that 'There is insufficient evidence to determine the effects of' is really right here so end up with something like: 'Our confidence in the moderate reduction in symptoms with intervention is very low due to bias and variation between the results of the studies'. | 3/27/2018 7:17 AM |
| 37 | 1)X reduces mortality slightly - why not: reduction in mortality was small 2)"small possibly unimportant" - could be confusing to people. It could be unclear to people what does it mean "possibly unimportant" 3) When talking about "important reduction", some context is necessary. What is important reduction? In the context of this information, it becomes apparent what is important or unimportant reduction 4) "appears to result" has potential to be misleading. The results should be described with more clarity. 5) There is not much difference in appears/may - the wording is not very clear | 3/26/2018 3:48 PM |
| 38 | The table is too complex to retain for mostly SR readers (i.e. non experts about SR production, like primary care physicians...). | 3/26/2018 12:42 PM |

39 Does "importance" always have to do with what is already known about the disease? Should every reported outcome have a known level of importance? At a certain threshold, do all effect sizes become "important" or can a moderate effect be "not important"? What exactly determines the difference between small important/small not important? It would be helpful to see this clarified with effect sizes. 3/26/2018 12:00 PM

Q10 One last question: Do you agree in principle that conclusions should be based on the concepts of the importance/size of the effect and the certainty of the evidence?

Answered: 85 Skipped: 27



| ANSWER CHOICES | RESPONSES | |
|----------------|-----------|----|
| Yes | 98.82% | 84 |
| No | 1.18% | 1 |
| TOTAL | | 85 |

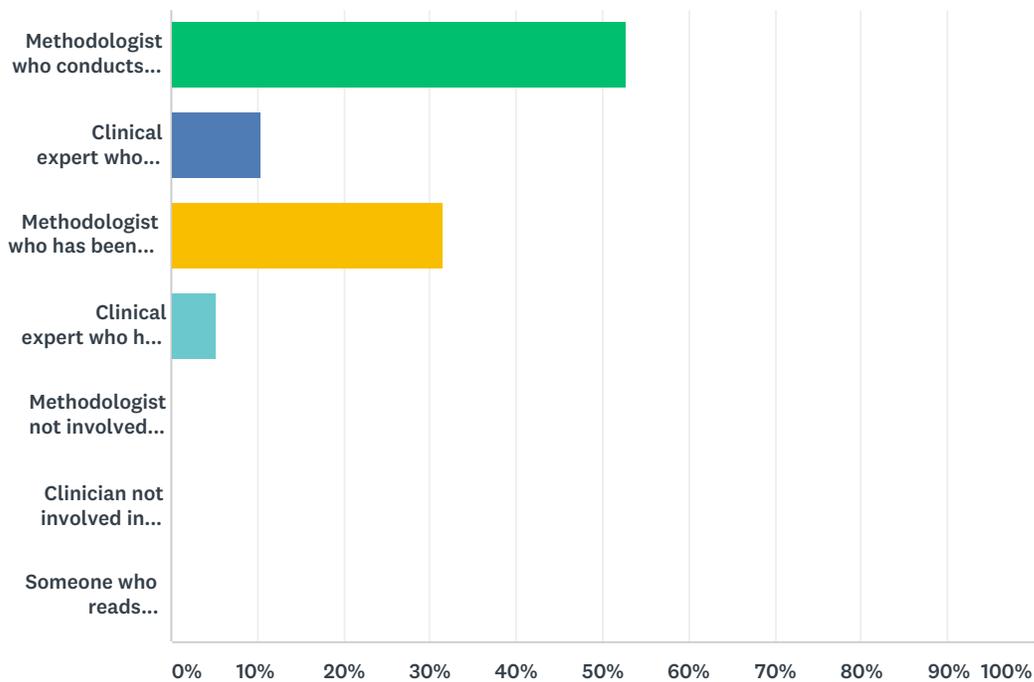
| # | PLEASE PROVIDE COMMENTS IF YOU'D LIKE. | DATE |
|---|---|------------------|
| 1 | good idea! | 4/6/2018 5:32 AM |
| 2 | I do agree with this statement, but I think there is a risk that importance and effect size may be conflated into a single concept. To me, importance includes aspects such as quality of life, possible adverse events, and patient preferences, while effect size is a necessarily specific quantitative measure. I reiterate what I said above about the problem with excluding information about potential harms (perhaps subsumed into the concept of importance), the fact that an intervention that has little or no effect on average could actually have a large and important effect in some people. I think it is also unwise to ignore practical issues such as the cost of an intervention (or cost relative to an alternative) or how difficult it would be to implement the intervention in the target population. | 4/4/2018 4:31 AM |
| 3 | Totally agree with both concepts needing to be included in the conclusion -- I only feel that we may need to explain WHY we are uncertain (i.e. low quality of evidence) as opposed to just saying we're uncertain (for the VERY LOW quality category). Left unqualified "uncertainty" can stem from many situations, so I think clearer to say because of low quality evidence. | 4/2/2018 2:43 PM |
| 4 | Yes and NO: see my earlier comments, I think we should not be making statements on 'unimportant' (i.e. not clinically relevant/ not patient relevant) effects. Also there is a fundamental issue: the level of evidence relates to a clinically relevant effect (or similar clinical decision threshold) and not to a 'larger' (or smaller) effect. For example with a grading HIGH we might be very certain that a particular intervention is superior i.e. has a clinically relevant advantage as compared to the control, but how certain would we be that the advantage is 'large'? In theory one could define 'large' and use this as a threshold to judge e.g. imprecision, and come up with a grading for a 'large effect' (in my example the level of evidence could be HIGH for a clinically relevant effect, and MODERATE for a 'large effect'), but this would make the GRADE system much too complicated and confusing! | 4/2/2018 8:45 AM |
| 5 | However, I must also say that as a methodologist and not a practitioner, it is often difficult to assess the size/importance of an effect. We have been told by our reviewers to leave that judgement to the practitioners. | 4/1/2018 1:59 PM |

How to communicate the results of systematic reviews using standard statements

| | | |
|----|---|--------------------|
| 6 | I think certainty of evidence is an improvement over the concept of quality of evidence. When it comes to small effects, I am not sure I understand how we will decide between important and not important. | 3/30/2018 5:55 PM |
| 7 | results of this survey will be helpful to guideline developers. | 3/30/2018 12:23 PM |
| 8 | What I find difficult is that the wording is about the importance/effect size and CiE, but what I (and many colleagues) automatically do when interpreting the results is to take the CI into account (wide CI make you less certain). I know this is not GRADE guidance, and I see that it makes sense conceptually, but to me it's still counter-intuitive. | 3/29/2018 7:07 AM |
| 9 | Yes, it's vital that guidelines and conclusions reflect both (and indeed the clinical importance rather than the statistical significance of results). It does not mean that we have to wait for strong evidence to be available before doing anything, but the quality of evidence should be taken into account when making decisions. | 3/29/2018 5:58 AM |
| 10 | There are other concepts that should be incorporated | 3/28/2018 2:06 PM |
| 11 | A- methodologists' opinion should not matter when we are trying to standardise the readability for end user, so I think surveying the likes of us is not going to get useful answers. B- A large chunk of SR work and its use is happening in English as a second language people and countries (Think Europe even). There is very little difference between the statements that I just saw from that perspective (I have dealt directly with some decision makers and stakeholders in non English countries). Standardising on our preferences and terms 'likely/ probably/ possibly' won't improve readability for the end user if they are not native English speakers. C- I longed for an open ended answer option for each question throughout, because I don't word a result statement without referring to the whole PICO - personal preference. | 3/28/2018 8:44 AM |
| 12 | Separating size of effect and certainty would be a helpful advance - similar to the separation of QoE and recommendations - especially if judgements/decisions were to be made transparent. | 3/27/2018 8:55 AM |
| 13 | But the context is very important, to explain to people why something is important or not important. | 3/26/2018 3:48 PM |
| 14 | Sure. "(significance + grade of evidence) + (size + direction) of the effect " | 3/26/2018 12:42 PM |
| 15 | Unless patients are not regularly involved in choosing / weighing endpoints for systematic reviews the concept of importance may differ between clinicians / authors and patients | 3/26/2018 10:20 AM |

Q1 What is your primary role related to systematic reviews and guidelines? (Select one that best represents your role)

Answered: 19 Skipped: 0



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----------|
| Methodologist who conducts systematic reviews | 52.63% | 10 |
| Clinical expert who conducts systematic reviews | 10.53% | 2 |
| Methodologist who has been involved in guideline development | 31.58% | 6 |
| Clinical expert who has been involved in guideline development | 5.26% | 1 |
| Methodologist not involved in systematic reviews or guidelines | 0.00% | 0 |
| Clinician not involved in systematic reviews or guidelines | 0.00% | 0 |
| Someone who reads systematic reviews | 0.00% | 0 |
| TOTAL | | 19 |

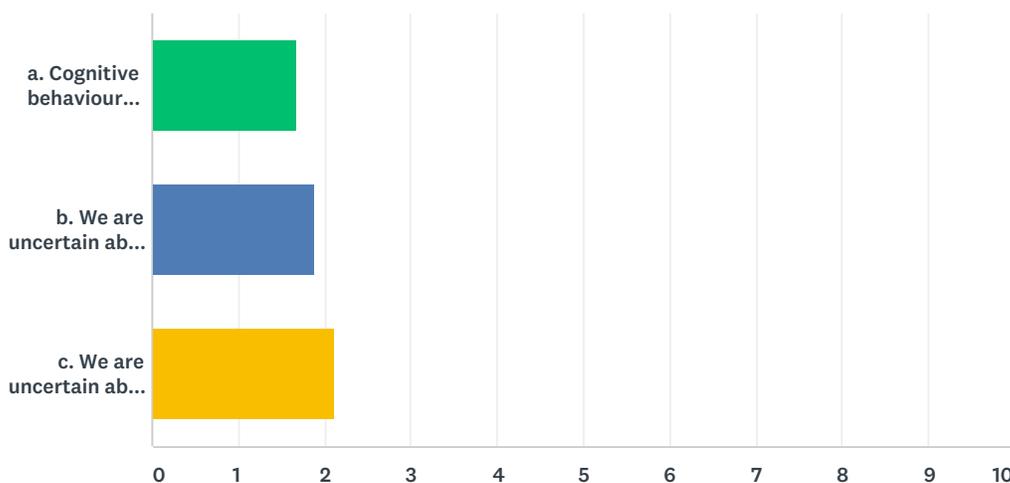
Q2 What is your education in epidemiology?

Answered: 17 Skipped: 2

| # | RESPONSES | DATE |
|----|---|--------------------|
| 1 | PhD | 4/3/2018 3:15 AM |
| 2 | Part of my undergraduate training. 'On-the-job' training undertaking systematic reviews and teaching people how to undertake systematic reviews | 4/2/2018 8:07 PM |
| 3 | Masters degree | 3/30/2018 5:25 AM |
| 4 | None. | 3/28/2018 12:56 AM |
| 5 | PhD in health research methods and MPH | 3/27/2018 3:42 PM |
| 6 | Epidemiologist responsible for postgraduate teaching and supervision of epidemiology | 3/27/2018 10:09 AM |
| 7 | nutritional epidemiology | 3/27/2018 9:38 AM |
| 8 | equivalent to a masters | 3/27/2018 8:55 AM |
| 9 | PhD(c) | 3/26/2018 8:36 PM |
| 10 | I attended workshops | 3/26/2018 3:33 PM |
| 11 | PhD Degree | 3/26/2018 1:32 PM |
| 12 | PhD | 3/26/2018 11:34 AM |
| 13 | No formal education in epidemiology. PhD in experimental medicine. | 3/26/2018 10:08 AM |
| 14 | Master degree in Epidemiology | 3/26/2018 9:45 AM |
| 15 | Mph- masters, epidemiology | 3/26/2018 7:32 AM |
| 16 | Quantitative and qualitative meta-analysis methods. NMA training in WinBugs. | 3/26/2018 6:58 AM |
| 17 | PhD | 3/26/2018 6:37 AM |

Q4 A systematic review compared the effects of cognitive behavioural therapy versus a waiting list for military suffering from post-traumatic stress disorder on depression. It found that cognitive behaviour therapy reduced depression by 8 points more on a scale from 1-100 (95% confidence interval from 21 point reduction to 12 point increase). This reduction is small but important. The evidence came from a meta-analysis with very few people (91) and very serious concern that the studies were at high risk of bias because of unclear randomisation and large loss to follow-up. The conclusion about the effect of cognitive behaviour therapy could be worded in the following three ways. Please indicate the acceptability of each statement.

Answered: 19 Skipped: 0



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|-------------|-------------|-------|------------------|
| a. Cognitive behaviour therapy may reduce depression slightly more than no therapy but we are uncertain. | 47.37% 9 | 36.84% 7 | 15.79% 3 | 19 | 1.68 |
| b. We are uncertain about the effect of cognitive behaviour therapy on depression. | 36.84% 7 | 36.84% 7 | 26.32% 5 | 19 | 1.89 |
| c. We are uncertain about whether cognitive behaviour therapy reduces depression more than no therapy. | 26.32% 5 | 36.84% 7 | 36.84% 7 | 19 | 2.11 |

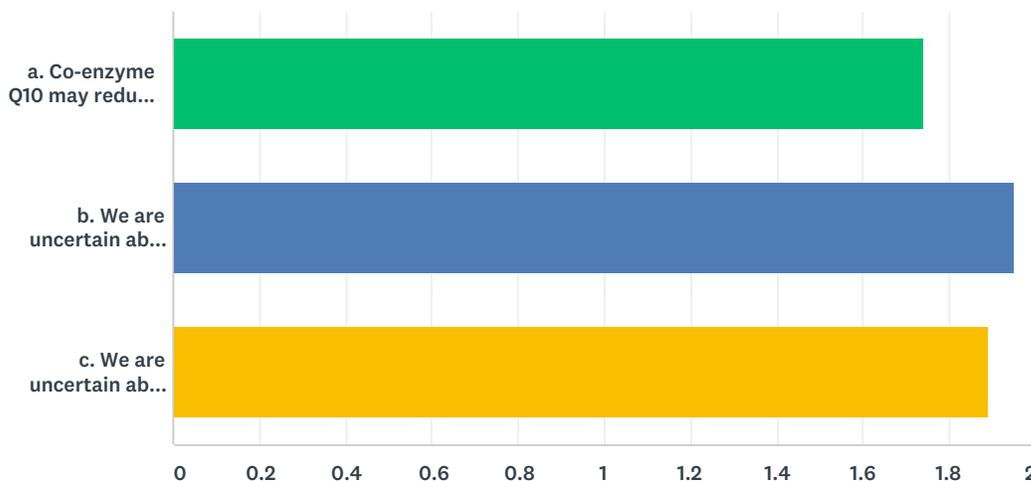
| # | PLEASE PROVIDE COMMENTS IF YOU HAVE ANY. | DATE |
|---|--|--------------------|
| 1 | Difference between b and c depends if it is in the context of an SoF-table. If yes, then I prefer b. if not, I prefer c. | 4/3/2018 3:19 AM |
| 2 | So this is a low certainty of effect situation which is what makes the example hard. I prefer to give a hint to what the effect estimate is even with low certainty evidence. | 3/27/2018 3:45 PM |
| 3 | I assume that the certainty in this example is very low (downgraded twice for risk of bias and once for imprecision). While I appreciate the logic of not indicating any effect direction within the conclusion, I think this is problematic for many/most(?) reviews. In my areas of research, most assessments come out as very low with the occasional low. | 3/27/2018 10:27 AM |

GRADE project group: Communicating the results of systematic reviews using standard statements

| | | |
|---|---|--------------------|
| 4 | Even in case of very low certainty of the evidence I think you should state an effect was observed. | 3/27/2018 9:44 AM |
| 5 | I would find the following acceptable: Cognitive behaviour therapy may reduce depression slightly more than no therapy but we are very uncertain. | 3/26/2018 8:38 PM |
| 6 | a. The uncertainty is reflected in the "may". The "but we are uncertain" part is redundant and may seem contradictory and result in confusion. It does depend, however, on how much uncertainty you believe is expressed in the word "may" c. Acceptable but too many words. Readers may spend a lot of time trying to make sure they understand the statement correctly, and when they finally do they realize that there is no answer to the question | 3/26/2018 1:38 PM |
| 7 | I would have rated this as very low certainty and would rather say "it is unclear whether cognitive..." | 3/26/2018 10:20 AM |

Q5 The conclusion about the effect of co-enzyme Q10 on blood pressure could be worded in the following three ways. Please indicate the acceptability of each statement.

Answered: 19 Skipped: 0

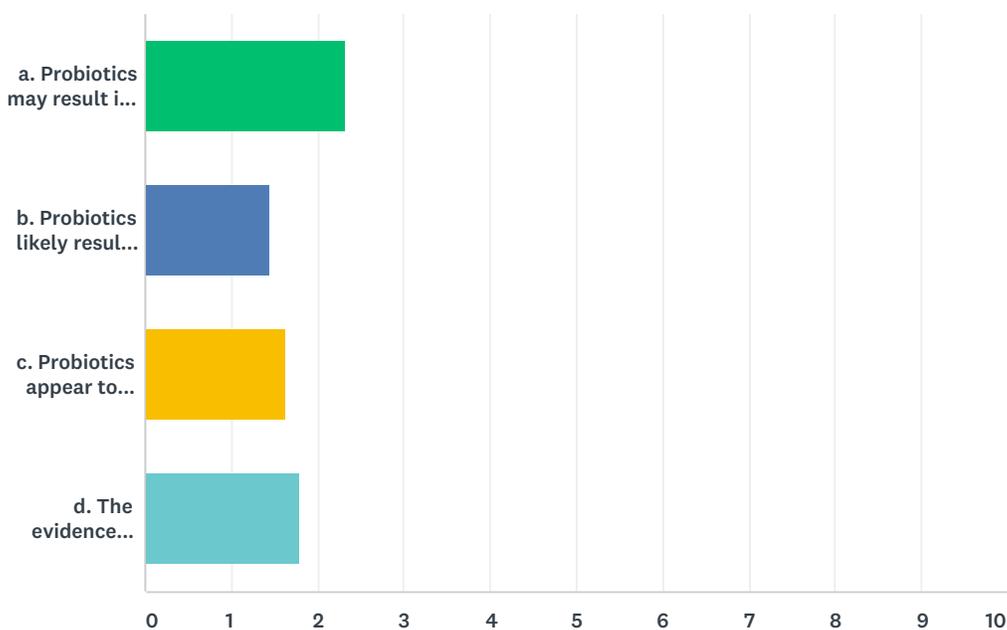


| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|---|--------------|--------------|-------------|-------|------------------|
| a. Co-enzyme Q10 may reduce blood pressure slightly but we are uncertain. | 52.63% 10 | 21.05% 4 | 26.32% 5 | 19 | 1.74 |
| b. We are uncertain about the effect of co-enzyme Q10 on blood pressure | 36.84% 7 | 31.58% 6 | 31.58% 6 | 19 | 1.95 |
| c. We are uncertain about whether co-enzyme Q10 reduces blood pressure. | 26.32% 5 | 57.89% 11 | 15.79% 3 | 19 | 1.89 |

| # | PLEASE PROVIDE COMMENTS IF YOU HAVE ANY. | DATE |
|---|--|--------------------|
| 1 | This example also adds the issue of small effect and the implication from a population point of view. so you may want to say "reduces individual's BP" | 3/27/2018 3:48 PM |
| 2 | the rationale is the same as I stated before | 3/27/2018 9:45 AM |
| 3 | I would find the following acceptable: Co-enzyme Q10 may reduce blood pressure slightly but we are very uncertain. | 3/26/2018 8:39 PM |
| 4 | is it 1.6 the difference between both arms or the absolute effect? I assume the difference is meant, but very misleading in the SOF! Lower by xxx more is quite confusing... In all the statements compared to placebo is missing | 3/26/2018 3:45 PM |
| 5 | a. Same comment as previous question For b vs c, it's a matter of what you are certain or uncertain of. I would say it makes little difference when the certainty is very low, but it may be more important when you have moderate certainty due to imprecision, and you have to communicate whether you have moderate certainty of an effect or of a lack of effect | 3/26/2018 1:46 PM |
| 6 | I would use may if low certainty. Again would rather use unclear, and in that case c) | 3/26/2018 10:22 AM |

Q6 The authors of the review considered that the cut-off for a large effect is RR 0.60. Please indicate the acceptability of the statements below.

Answered: 19 Skipped: 0



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|--------------|-------------|-------|------------------|
| a. Probiotics may result in a large reduction in the incidence of diarrhea. | 15.79% 3 | 36.84% 7 | 47.37% 9 | 19 | 2.32 |
| b. Probiotics likely result in a large reduction in the incidence of diarrhea. | 66.67% 12 | 22.22% 4 | 11.11% 2 | 18 | 1.44 |
| c. Probiotics appear to result in a large reduction in the incidence of diarrhea. | 42.11% 8 | 52.63% 10 | 5.26% 1 | 19 | 1.63 |
| d. The evidence suggests that probiotics result in a large reduction in the incidence of diarrhea. | 42.11% 8 | 36.84% 7 | 21.05% 4 | 19 | 1.79 |

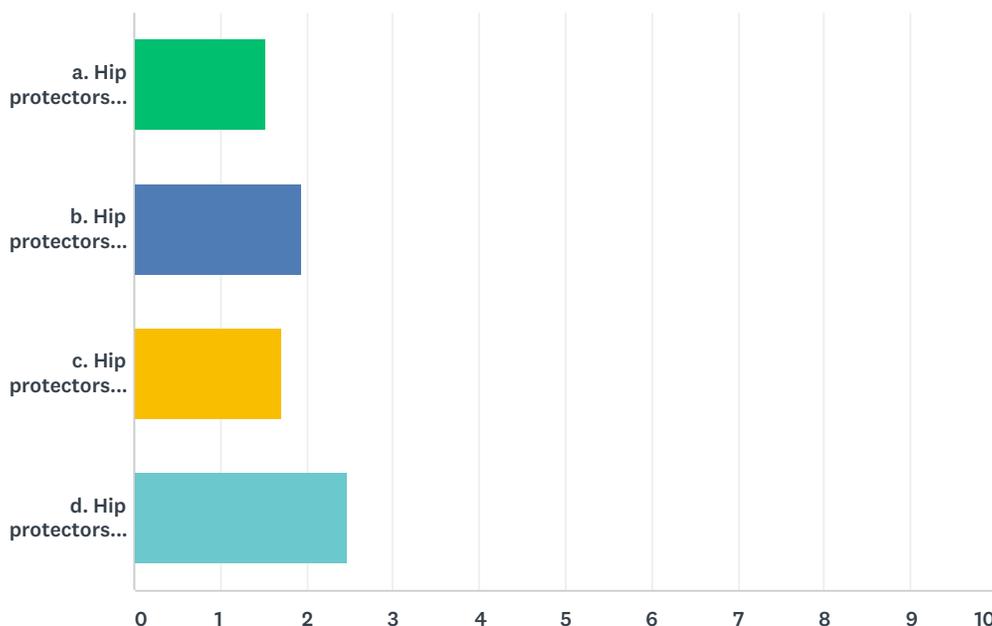
| # | PLEASE PROVIDE COMMENTS IF YOU HAVE ANY. | DATE |
|---|--|--------------------|
| 1 | b&c: related to moderate QoE to me. d: I wouldn't move into this direction, the word 'suggest' might be confused with terminology for weak recs. | 4/3/2018 3:30 AM |
| 2 | It will be very concerning if we ignore the certainty in the statements. I prefer to explicitly state that we are not certain in addition to the "may" or "appear to" | 3/27/2018 3:51 PM |
| 3 | Given there is low certainty, that should ideally be communicated within the conclusion. The word 'suggests' seems to communicate that best but all seem acceptable. | 3/27/2018 10:36 AM |
| 4 | b is unacceptable: "likely" is in contradiction with low QoE. | 3/27/2018 9:48 AM |
| 5 | The wording needs to include some reflection that the evidence is of low certainty. | 3/26/2018 8:46 PM |
| 6 | as an adjunct to antibiotics compared to placebo... | 3/26/2018 3:49 PM |
| 7 | Maybe it's for simplicity for this survey but I wonder if the guidance should be that the narrative statements regarding magnitude should be made based on the absolute effects. May, appear, and suggest all communicate considerable uncertainty to me, which matches the low certainty. Likely sounds too strong. | 3/26/2018 1:51 PM |

GRADE project group: Communicating the results of systematic reviews using standard statements

| | | |
|---|--|-------------------|
| 8 | Option d could be confused with a weak recommendation | 3/26/2018 1:48 PM |
| 9 | I think the word large should be omitted from all statements, since 0.6 is outside the 95% ci. | 3/26/2018 7:39 AM |

Q7 The authors indicate that the effect found was less than their cut-off for an effect. Please indicate the acceptability of the statements to communicate the effects of hip protectors compared to no hip protectors on the number of hip fractures.

Answered: 17 Skipped: 2



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|--|--------------|-------------|--------------|-------|------------------|
| a. Hip protectors likely do not reduce hip fractures. | 58.82% 10 | 29.41% 5 | 11.76% 2 | 17 | 1.53 |
| b. Hip protectors likely result in little to no difference in hip fractures. | 25.00% 4 | 56.25% 9 | 18.75% 3 | 16 | 1.94 |
| c. Hip protectors probably do not reduce hip fractures. | 47.06% 8 | 35.29% 6 | 17.65% 3 | 17 | 1.71 |
| d. Hip protectors probably result in little to no difference in hip fractures. | 11.76% 2 | 29.41% 5 | 58.82% 10 | 17 | 2.47 |

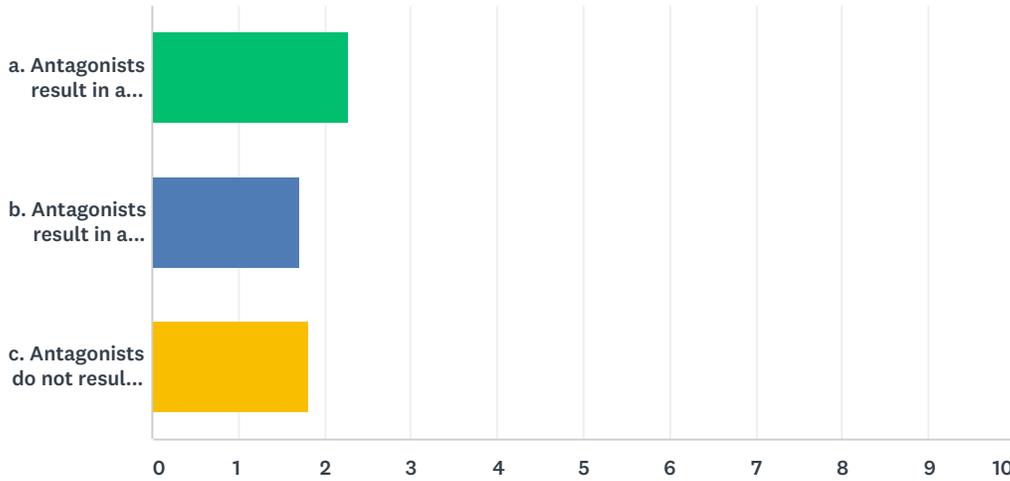
| # | PLEASE PROVIDE COMMENTS IF YOU HAVE ANY. | DATE |
|---|---|--------------------|
| 1 | b&d: ideal (if you feel probably and likely are interchangeable, I do) a&c: I don't like 'do not', 'little to no difference' solves a lot of problems. | 4/3/2018 3:38 AM |
| 2 | The use of likely is important because it reflects the certainty | 3/27/2018 3:52 PM |
| 3 | The clinical question is about reducing hip fracture risk so communicating the take-home message in relation to that makes sense. The 'little to no difference' is obviously technically correct but less intuitive and understandable. | 3/27/2018 10:44 AM |
| 4 | b/d reflects there was an effect although (very) small. Note: In Dutch language it is hard to know what is the difference between likely and probably. | 3/27/2018 9:55 AM |

GRADE project group: Communicating the results of systematic reviews using standard statements

- | | | |
|---|---|-------------------|
| 5 | "The authors indicate that the effect found was less than their cut-off for an effect." - do you mean less than cut off for an important effect? Assuming that the CI interval excludes an important difference, then b or d is acceptable. I feel strongly that we should never ever rate our certainty that there is 'no effect' -- there will always be infinitely too much imprecision. | 3/26/2018 8:52 PM |
| 6 | This one is hard. Some things to consider - The RR seems to be for an increase in hip fractures (there are more fractures in the HP group), so why are all the statements about HP reducing the risk? To me this is about providing guidance with regards to how to frame the statement: based on what we expected to see (in this case, HP and their likelihood of reducing fractures), or what we saw (in this case, HP increased fractures). - If the former, should the statement about the presence of an effect versus the presence of no effect be based on/ modified by what we expected to see? - I said this before, but it may be important to be explicit about whether the statement is based on the point estimate, the CI, or when is which. In this example it you use the CI (no effect vs little to no effect) while in the others you used the point estimate+ CI (effect vs no effect), even if you had less certainty. | 3/26/2018 2:04 PM |
-

Q8 Please indicate the acceptability of the statements below to communicate the effect of antagonists compared to placebo on daytime nasal symptoms. Note: the authors considered the SMD and confidence interval a small but not important effect.

Answered: 17 Skipped: 2



| | UNACCEPTABLE | ACCEPTABLE | IDEAL | TOTAL | WEIGHTED AVERAGE |
|---|--------------|--------------|-------------|-------|------------------|
| a. Antagonists result in a small effect that may not be an important reduction in daytime symptoms. | 11.76% 2 | 47.06% 8 | 41.18% 7 | 17 | 2.29 |
| b. Antagonists result in a small possible unimportant reduction in daytime symptoms. | 35.29% 6 | 58.82% 10 | 5.88% 1 | 17 | 1.71 |
| c. Antagonists do not result in an important reduction in daytime symptoms. | 47.06% 8 | 23.53% 4 | 29.41% 5 | 17 | 1.82 |

| # | PLEASE PROVIDE COMMENTS IF YOU HAVE ANY. | DATE |
|---|--|--------------------|
| 1 | b: I would not try to incorporate any uncertainty about the clinical threshold within this standard statement, I feel cognitively that really is too much, and too confusing. We set a treshold (and that is of course always uncertain and highly depending on a lot of factors, but nevertheless it complicates these statements if we try to incorporate this). I would reserve the word 'possibly' as one of the option for expressing uncertainty in case of QoE LOW. | 4/3/2018 3:51 AM |
| 2 | If the reduction is not of any clinical importance, mentioning it within the conclusion is misleading and muddies the water. The wording "may not be an important reduction" and "possible unimportant reduction" are a bit strange. Surely, there is little uncertainty in the decision about whether these are or are not important (unless it is genuinely on the borderline of clinical importance)? | 3/27/2018 10:57 AM |
| 3 | b is more concise than a. Thatis why I prefer this one. c lacks nuance. | 3/27/2018 9:57 AM |

GRADE project group: Communicating the results of systematic reviews using standard statements

| | | |
|---|---|--------------------|
| 4 | <p>To say with any certainty that the effect is unimportant, then we need to also know the distribution of the effect. Is it normal? For example, are there some people in whom antagonists confer a large important effect and in others no effect? Have a look at how we took into consideration the distribution in this guideline: https://www.bmj.com/content/357/bmj.j1982 Ideal would be: Antagonists result in a small reduction in daytime symptoms that most people would probably not consider important. We also need to have a high degree of certainty in the MID to say anything about importance, don't we? We also need to be certain (or at least make the assumption) that most people have similar MIDs, or at least that most people would not consider it important. I'm not sure that's the case here... So I think really what we are rating is that there is a difference. Probably better to say that there is a difference and leave it at that, unless you can provide more information as above.</p> | 3/26/2018 9:03 PM |
| 5 | <p>a. OK but too wordy and confusing b. OK but reads weird to me. Maybe "small but unimportant", "small but not important" c. Does not present the full picture in which you have high certainty. It could mean that there is a small and unimportant effect or that there is no evidence of effect (not statistically significant)</p> | 3/26/2018 2:07 PM |
| 6 | <p>I prefer c) but important to clearly define the chosen threshold for an important effect</p> | 3/26/2018 10:32 AM |

Q9 Please take 5 minutes to review this list of options for statements to communicate results at various levels of evidence and size of effect. Please add any general comments about the statements below.

Answered: 12 Skipped: 7

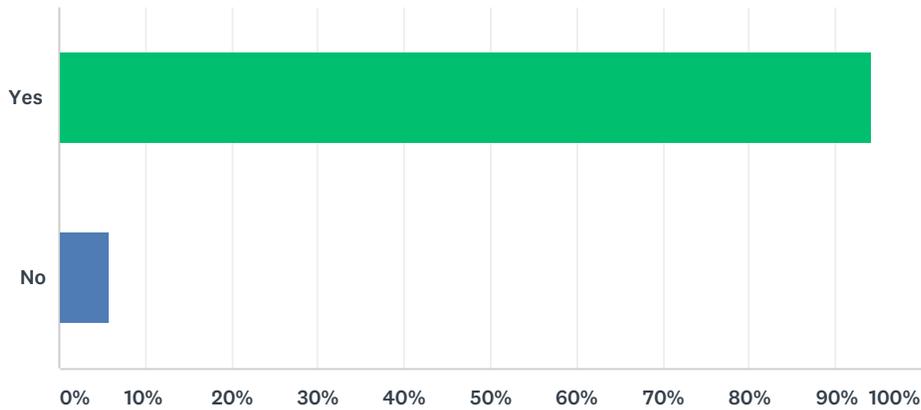
| # | RESPONSES | DATE |
|---|---|--------------------|
| 1 | Some of the statements are very clumsy with poor grammar, hard to understand what they mean without reading them several times | 4/3/2018 9:41 AM |
| 2 | In a lot of situations 'little to no difference' solves a lot of issues because panelist in general seem to have problems when we say there is no effect when we the point estimate suggest there is (but our judgements say there is a small effect that is rated as not important to patients). Major issue is using standardized statements in the context of a SoF. Within the context of a SoF: my preference is to keep them as brief as possible (but giving just enough information readers are able to understand (by for example making the assumption that the comparison is clear form the table: so no need to repeat in the statement), and to be able to present them in user friendly format.) As soon as people decide to go outside the SoF readability usually really decreases, because they often feel the urge to incorporate all the info that is in a SoF:) Last point: I feel this colum in SoF is really crucial, because - if we like it or not - most end-users (naive or not) need help interpreting what the SoF says per outcome. People also use the word 'possibly' indicating LOW quality evidence. Might be another option. I would leave out the option 'X probably may result in a small possibly unimportant effect'. Very simply put: I currently think of these statments as expressing 1) the uncertainty (is/probably is/possibly is/uncertain) and 2) judgement about the magnitude (call it clinically relevance or importance to patient) preferably based on cutt-of value. I used the options below quiet a lot (and we translated them into Dutch), and in almost all instances they work for me. | 4/3/2018 4:19 AM |
| 3 | Think the 'small possibly unimportant' statement is confusing and prefer the 'small effect that may not be an important reduction'. The 'does not result in an important reduction' seems too blunt. The 'moderate' category doesn't have an adjective to it whereas the large and small categories do. Prefer 'likely' as opposed to 'probably' Why in the Low certainty of evidence section is there an option of 'The evidence suggests' when this is not in the other sections? Very low section - prefer the 2nd and 3rd statements as opposed to the 1st | 4/2/2018 9:14 PM |
| 4 | I tend to prefer the terms probably for moderate certainty and may for low certainty. Likely and appears also seem to be acceptable synonyms. I would tend not to use may for very low certainty evidence. | 3/30/2018 5:49 AM |
| 5 | Did you consider adding the following to the list We are certain, x results in large reduction We are mod certain, x results in large effect We are uncertain, x results in large effect | 3/27/2018 4:01 PM |
| 6 | For small not important effect, I prefer the bottom of the three suggested options. I also generally prefer the top option for no effect. For moderate certainty and a small effect (not important), I think it would be better to replace "may not be an important reduction" with "is not an important reduction". The wording of "small possibly unimportant effect" could be replaced by "small but unimportant effect". | 3/27/2018 11:12 AM |
| 7 | Wordings are generally quite acceptable. But in case of very low certainty I would like to suggest to add "very" to uncertain in the last two sentences. | 3/27/2018 10:04 AM |
| 8 | Making mention of a magnitude of effect requires a values judgement. One that can be empirically measured in a population. If the authors use their own judgement (always extremely dubious), then they need to be explicit. I don't think that we can ever have high or moderate certainty that an effect is large or small without also having certainty that all or almost all people would weigh the magnitude in effect similarly. There is probably a huge amount a variability between people in how much value they place on different effect sizes. At least that is my experience. So, if we include a magnitude of effect in the certainty ratings, we need to incorporate our certainty in the typical and distribution of values and preferences of those who the evidence applies to. | 3/26/2018 9:11 PM |

GRADE project group: Communicating the results of systematic reviews using standard statements

| | | |
|----|--|--------------------|
| 9 | High certainty: small effect (not important): x results in a small possibly unimportant sounds very confusing for the reader, I suggest to remove this option same is true for moderate certainty, even more confusing: probably small possibly unimportant. What does this mean?? As I am not a native speaker: Is probably and likely exactly the same in English? low certainty: is may and appear exactly the same? For me, may sounds more convincing...I prefer slightly over may and appears for low certainty. For no effect, low certainty I prefer the sentences with the "little to no difference" very low certainty: I prefer the sentences starting with: we are uncertain | 3/26/2018 4:02 PM |
| 10 | It is not clear to me whether "The evidence suggests...." is a necessary option in case of low evidence. It is not used in other scenarios at all. I could imagine that this makes a difference in terms of communicating results and/or conclusions. The statement "the evidence suggests..." is quite unpersonal. This might also imply that evidence suggests XYZ, but we (the authors) think that ABC. All other statements are more straight forward in this sense, in my opinion. | 3/26/2018 10:00 AM |
| 11 | This is a good start that needs some discussion going forward. | 3/26/2018 7:46 AM |
| 12 | Although stronger statements are reasonable when certainty is higher, the imprecision is already factored in. So the effect size needs to be considered in terms of absolute effects. | 3/26/2018 7:07 AM |

Q10 One last question: Do you agree in principle that conclusions should be based on the concepts of the importance/size of the effect and the certainty of the evidence?

Answered: 17 Skipped: 2



| ANSWER CHOICES | RESPONSES | |
|----------------|-----------|----|
| Yes | 94.12% | 16 |
| No | 5.88% | 1 |
| TOTAL | | 17 |

| # | PLEASE PROVIDE COMMENTS IF YOU'D LIKE. | DATE |
|---|--|--------------------|
| 1 | Also have to consider impact of adverse effects and safety | 4/3/2018 9:41 AM |
| 2 | I fully agree! And this is really a great help formulating conclusion /evidence statements. Thanks for this great work. | 4/3/2018 4:19 AM |
| 3 | This is extremely important to highlight . The certainty of the evidence is what usually gets lost | 3/27/2018 4:01 PM |
| 4 | The importance/size of the effect is crucial and often not given adequate consideration. | 3/27/2018 11:12 AM |
| 5 | The effect size is very important. It will help you in balancing benefits and harms (in the EtD framework) | 3/27/2018 10:04 AM |
| 6 | I don't think we're there yet. | 3/26/2018 9:11 PM |
| 7 | Yes, but these should always be in relation to each other, ie the certainty rating should be made in relation to the defined range of importance/size of effect. | 3/26/2018 10:36 AM |