Grading the quality of evidence in complex interventions: a guide for evidence-based practitioners

M Hassan Murad, Jehad Almasri, Mouaz Alsawas, Wigdan Farah

Evidence-based Practice Center, Mayo Clinic, Rochester, Minnesota, USA

Correspondence to:
Dr M Hassan Murad, Evidence-based Practice Center, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA; murad.mohammad@mayo.edu

Abstract

Evidence-based practitioners who want to apply evidence from complex interventions to the care of their patients are often challenged by the difficulty of grading the quality of this evidence. Using the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach and an illustrative example, we propose a framework for evaluating the quality of evidence that depends on obtaining feedback from the evidence user (eg, guideline panel) to inform: (1) proper framing of the question, (2) judgements about directness and consistency of evidence and (3) the need for additional contextual and qualitative evidence. Using this framework, different evidence users and based on their needs would consider the same evidence as high, moderate, low or very low.

Complex interventions in medicine are defined as interventions that contain several interacting components within the experimental and control interventions, multiple behaviours required by those delivering or receiving the intervention, multiple groups or organisational levels targeted by the intervention, and permit for variable flexibility and tailoring of the intervention.[1] An example is care management programmes for type 2 diabetes. In one systematic review, programmes varied across studies in terms of team composition (physician, nurse, case manager, diabetes educator), delivery method (face-to-face, telephone, online, outpatient, inpatient), intensity and frequency of the intervention, and other factors.[2] Complex interventions are increasingly used in studies of behavioural change, psychotherapy, education, public health, health services, quality improvement, social policy and many other fields.

The challenge

Just as with any other intervention, evidence-based practitioners who want to apply evidence from complex intervention to the care of their patients need to select the evidence (hopefully through a systematic review process) and appraise it (ie, identify the extent to which this evidence is trustworthy, also called, rating the quality of evidence). The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach is a modern method for rating the quality of evidence with good transparency and reliability.[3] A study of Cochrane systematic reviews showed that the outcomes of complex interventions were more likely to be rated as 'very low' quality of evidence compared with those of simple interventions (37.5% vs 9.1%). None of the outcomes of complex intervention reviews were rated as 'high'.[4] We believe these low ratings are inconsistent with the definition of quality of evidence in GRADE, which is a construct that reflects the trustworthiness of evidence, and we attribute this phenomenon to improper framing of the clinical question for which the quality of evidence is being rated.[5]

Rationale

Complex interventions are inherently heterogeneous; thus, we may be tempted to lower the quality of evidence for heterogeneity (inconsistency across studies). Complex interventions are also likely to have components that are different from the ones we can or want to implement in our setting; thus, we may be tempted to lower the quality of evidence for indirectness (lack of applicability). Therefore, by rating down for heterogeneity and indirectness, most complex intervention studies will turn out to provide low-quality evidence; which is frustrating to evidence users.[4]

Proposed approach

We present a framework for evaluating the quality of evidence using GRADE. The framework hinges on obtaining feedback from the end user of evidence (eg, a guideline panel) during the process of evidence synthesis (eg, conducting a systematic review to support a guideline). This feedback seeks to inform: (1) proper framing of the question, (2) judgements about directness and consistency of evidence (the two domains that are highly relevant to complex interventions) and (3) the need for additional contextual and qualitative evidence to provide information about the circumstances under which the intervention works best (figure 1). This qualitative evidence can explain barriers and facilitators of implementation, as well as the cultural and social factors that can modify the effect of the intervention.

Although engaging patients and evidence users in the process of evidence synthesis is not new[6] and is routinely carried out in some systematic reviews (eg, stakeholder engagement is mandated in reviews conducted by Evidence-based Practice Centers funded by the Agency on Healthcare Quality and Research),[6] the role of this engagement has been to develop the scope and key questions. Here, we advocate for using their feedback to inform rating the quality of evidence of complex interventions. Finally, evidence synthesis methods other than traditional quantitative meta-analysis will likely be needed to address long causal chains in complex interventions, such as model driven and realistic systematic reviews and meta-narrative evidence synthesis.

Example

The framework is applied through an illustrative example that demonstrates how the perspective of the evidence user affects rating of quality of evidence and shows how the same evidence can warrant different levels of trustworthiness for different users (table 1). For

BMJ
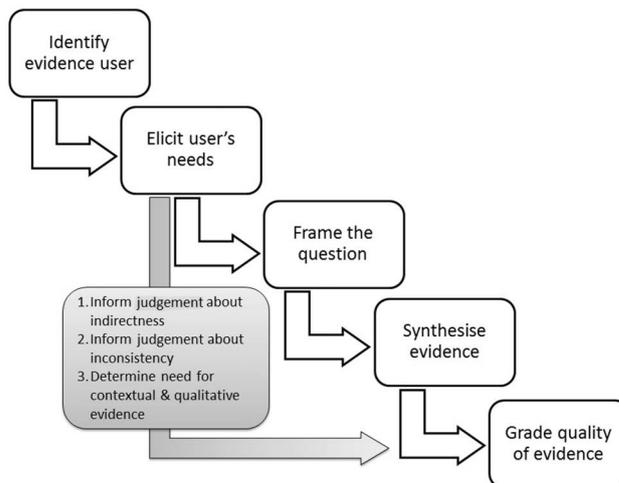
Figure 1 The traditional process of rating the quality of evidence is in white colour. The additional proposed steps that address complex interventions are in grey colour.

Table 1 Care management interventions for type 2 diabetes

Evidence user	Perspective	Framing of the question	Clinical heterogeneity	Statistical heterogeneity	Quality of evidence (certainty in evidence)*
Policymaker/ payer	Should I cover (pay for) this intervention?	To what extent is this bundle of interventions effective?	No concern, we are interested in the effect of the whole bundle.	No concern, we are interested in the effect of the whole bundle.	High
Physician and patient	Which part of the intervention should we provide to this patient?	Which component of the intervention is most effective?	Variability of components is concerning particularly if a component is not well described or not feasible in own practice.	No concern if metaregression explains heterogeneity (ie, demonstrate the effect of each component).	▸ *Moderate* (if statistical heterogeneity was explained) ▸ *Low* (if statistical heterogeneity was not explained)
Diabetes counsellor/ educator	How can I apply the intervention?	Under what circumstances does the intervention work best?	No concern	No concern	The effectiveness evidence is inadequate to judge certainty. Additional contextual evidence from qualitative research is identified leading to high certainty.†

Evidence: meta-analysis demonstrated that the interventions significantly reduced HbA1c (−0.21%, 95% CI −0.40% to −0.03%). Across studies, there was important clinical heterogeneity (varying components of the intervention) and statistical heterogeneity (varying effect size with I^2>75%). Following a traditional approach, the evidence may be graded down twice for indirectness (numerous components make application challenging) and inconsistency (heterogeneity of effect).
*The focus in this example is only on the domains of indirectness and inconsistency. We assumed that there were no concerns in other quality of evidence domains such as precision, risk of bias or concerns about HbA1c being a surrogate outcome.
†Contextual evidence shows that a significant amount of qualitative studies are available to provide context and details on implementation. For example, studies demonstrated that before providing education in diabetes, we should address negative feelings that individuals have (anger about the initial diagnosis, and subsequently negotiate loss of control of a person's own lifestyle and environment).[7] Qualitative studies also provided explanation to lack of adherence to a programme (individual, organisational and content factors).[8] Addressing these factors is essential to programme success.
HbA1c, glycated haemoglobin.

example, policymakers or payers (health insurance carriers) may not be concerned about clinical and statistical heterogeneity because they are interested in *any* management programme that improves diabetes control. They realise that different health systems under their jurisdiction may develop different programmes with variable components based on each system setting and resources (some delivered face-to-face, some delivered remotely, some led by nurse, etc). Therefore, from a policymaker perspective, there may be little interest in knowing more about the individual components of the intervention. This differs substantially from the patient

and clinician perspectives because they may be interested in only choosing the most effective components of diabetes management that can fit in the patient capacity, schedule and daily routine. The third perspective presented in the table is that of a diabetes educator responsible for implementing the programme, which requires knowing much more details about implementation and contextual factors. Adopting this perspective leads to considering the evidence of effectiveness to be insufficient for the question at hand. Two qualitative studies[7][8] were identified and provided some of the needed details. The studies shed light on the psychological factors that should be addressed when delivering diabetes management programmes and provided insight on the factors that affect adherence to such programmes (table 1). With these contextual factors known, the persons providing the programme will have increased confidence in their ability to deliver the programme and they may rate the quality of the evidence as high.

Implications

We believe this approach will hopefully lead to more appropriate and possibly consistent judgements about studies of complex interventions and better application of GRADE. This approach, however, will not solve other challenges of complex interventions. For example, lack of blinding of outcome assessors and inadequate allocation concealment (both are feasible even if the study was unblinded and open) and poor reporting of the details of the intervention (which makes them hard to replicate in practice). These challenges can be solved at the primary research level, and not at the level of evidence synthesis, appraisal or application.

References

1. **Craig P**, Dieppe P, Macintyre S, *et al*. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655.
2. **Egginton JS**, Ridgeway JL, Shah ND, *et al*. Care management for type 2 diabetes in the United States: a systematic review and meta-analysis. *BMC Health Serv Res* 2012;**12**:72.
3. **Mustafa RA**, Santesso N, Brozek J, *et al*. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;**66**:736–42; quiz 42 e1–5.
4. **Movsisyan A**, Melendez-Torres GJ, Montgomery P. Outcomes in systematic reviews of complex interventions never reached "high" GRADE ratings when compared with those of simple interventions. *J Clin Epidemiol* 2016;**78**:22–33.
5. **Murad MH**. Applying GRADE to studies of complex interventions is possible but requires proper framing of the question. *J Clin Epidemiol* 2016;**77**:138–139.
6. **Shippee ND**, Domecq Garces JP, Prutsky Lopez GJ, *et al*. Patient and service user engagement in research: a systematic review and synthesized framework. *Health Expect* 2015;**18**:1151–66.
7. **Boyle E**, Saunders R, Drury V. A qualitative study of patient experiences of type 2 diabetes care delivered comparatively by general practice nurses and medical practitioners. *J Clin Nurs* 2016;**25**:1977–86.
8. **Schwennesen N**, Henriksen JE, Willaing I. Patient explanations for non-attendance at type 2 diabetes self-management education: a qualitative study. *Scand J Caring Sci* 2016;**30**:187–92.