COMMENTARY

AHRQ Series Commentary 2: Informative, timely, and valuable: an outsider view of the Comparative Effectiveness Review articles

David Tovey*


Accepted 11 January 2010

The development of the Methods Guide for Comparative Effectiveness Reviews (CERs) by the Agency for Healthcare Research and Quality (AHRQ) and Effective Health Care (EHC) Program is an important and timely development [1]. The articles published within the Journal of Clinical Epidemiology (JCE) represent a sample from the work completed so far and are informative in describing current strategic thinking within the program. They will also have relevance for many others who are responsible for conducting systematic reviews both within and beyond the United States. “The Guide” builds on the experience and expertise gained since the introduction of the Evidence-based Practice Center (EPC) Programs in 1997 and has also been influenced by similar programs internationally, including the National Institute for Health and Clinical Excellence in the United Kingdom, the Canadian Agency for Drugs and Technology in Health, and also The Cochrane Collaboration.

No health system is immune from the pressure to contain costs, but the situation is particularly critical in the United States and, as we have seen over the past 12 months, also somewhat highly charged. Recent controversies in relation to the United States Preventive Services Task Force recommendations on mammography and others indicate some of the challenges faced by system providers and policy makers and the fragile nature of the trust between vocal elements of the public and media, some politicians, and those who are advocating a more evidence-based and realistic approach to health care decision making.

The purpose of the JCE articles and the Methods Guide for CERs is described as to “improve consistency within and transparency of the EHC programme” [2]. The thread of transparency runs throughout the program. Interested members of the public are encouraged to propose and help to focus topics for CERs, and more recently, CERs in development have been posted publicly to encourage comment [2].

Inevitably, those who are working outside the program will be interested to compare the CER methodological approaches with those of others conducting systematic reviews. The articles raise important issues both at the strategic level and also within the details of individual items, such as the reporting of harms.

One of the historic challenges for people advocating evidence-based health care has been the gap between systematic reviews and clinical practice. Traditionally, systematic reviews were explicitly limited in their scope simply to present the evidence, with a tentative nod toward its “implications.” This led to some complaints from clinicians that systematic reviews were not sufficient to guide clinical actions in practice (not, presumably, the same ones who moaned the perceived threat to their clinical autonomy and the dangers of “cook-book medicine”). The articles published in the JCE demonstrate how the EHC Program has sought to address such concerns. At the outset, the priority of the program is to provide “understandable and actionable information for patients, clinicians and policy makers” [2].

As such they can be seen to represent one part of a more widespread strategic shift; from seeing the systematic review as being simply the task of identifying the presence or otherwise of high-quality evidence on the effects (predominantly the benefits) of an intervention in a given clinical situation toward the accumulation and synthesis of the best current evidence to guide practice and policy in response to a given health care question. In Mark Helfand’s editorial, [3] he suggests that CERs are “better described as ‘complex evidence reports’ than as ‘systematic reviews’ in recognition of this broader, and more outcome focused task.” Apart from the potential of this approach of creating difficulties of understanding what people mean when they use the abbreviation CER, I think this is slightly problematic, because the purpose of conducting systematic reviews should always be to ensure that they provide the best and most actionable evidence to guide practice and policy. This necessitates, as the CER articles make clear,
that systematic reviews are not simply concerned with accuracy (validity) but also and crucially concerned with applicability to end users, relevance, and timeliness. There are already a number of terms that are used, each with differences that may appear somewhat subtle to the end users, including systematic reviews, meta-analysis, health technology assessment, guidelines, and so on. If reviews such as those produced by the EHC program need to be re-labeled, it might be preferable to retain some elements of the familiar ones and refer to them as “extended” or “applied” systematic reviews, to reflect that they are based in a systematic approach, but that this has been augmented for a specified purpose, such as to make them more relevant or applicable to real-life decision making in the U.S. context.

Inevitably, the articles published in the JCE and the Methods Guide for CERs will be compared with the Cochrane Handbook for Systematic Reviews of Interventions [4]. On the basis of the CER articles published to date, there is considerable common ground between the Cochrane methodological approaches and those described for CERs. This is not surprising, because both the Cochrane Collaboration and AHRQ share a recognition of the importance of reviewers adopting the highest possible standards of research practice and reporting. The Cochrane Handbook and the Methods Guide for CERs appear on the basis of the content published so far to be highly complementary and would be useful reference points for reviewers outside either organization. The Cochrane Handbook describes what is perhaps a more prescriptive and fine detailed approach, whereas the Methods Guide for CERs articles appear to be directed more at describing the underlying principles of the intended approach. To an extent, this reflects a difference in the anticipated readership. Teams undertaking CERs will be expected to be experienced systematic reviewers, and the review questions will be chosen specifically because of the high impact on health care delivery and policy. The Cochrane Collaboration has traditionally encouraged a mix of less-experienced volunteers teaming up with experienced reviewers to undertake reviews in their shared area of interest. The Cochrane Handbook, therefore, seeks to ensure that there is consistency in the rigorousness of methodological approach and to provide detailed guidance.

The chapters that describe the preferred approaches to assessing harms and those describing the quality of evidence provide an insight into the CER approach at a more tangible level [5,6]. Both are highly instructive and important. In relation to identifying harms, it has lately become a point of principle for many systematic reviewers not to limit the search to the narrow confines of randomized studies. There is far less agreement on what this means in practice and how this can be optimally informative to the end user without becoming unfeasible from the point of view of the reviewer and the use of resources. The article by Chou et al., describing harms, makes the EHC approach both explicit and clear: “Observational studies are almost always necessary to assess harms adequately” [5]. There follows a useful description of the sources of information that can provide valid material to inform the review, including unpublished supplemental trials data, cohort and case—control studies, studies based on patient registries, for example, and their various limitations and risk of bias. Inevitably, in an article of its length, there are few examples to inform decisions about limiting or extending the search, but it would be useful for these to be developed further. I would also have liked to see some more explicit connection between the articles on harms and grading the quality of evidence, because they seem to overlap naturally. The only mention that I could find was within the “Discussion” section of the article by Owens et al. [6]. In this article, the authors proposed that CER reviewers might choose to upgrade the initial assessment of observational studies from low to moderate and downgrade randomized studies accordingly from high to moderate or low. In itself this represents a substantial and potentially controversial departure from the approach described by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group (www.gradeworkinggroup.org) [7].

The general approach to grade the strength of a body of evidence is closely aligned to GRADE, and two leading members of GRADE Working Group are acknowledged at the end of the article. Perhaps, most importantly, the proposed approach maintains several core aspects of GRADE. These include grading the quality of evidence by outcome rather than study and allowing the overall rating to be based on more than study type, to include at a minimum, risk of bias assessment, consistency, directness, and precision. There are, however, specific and interesting differences beyond those mentioned in the earlier paragraph. First, the CER approach seems to be much more relaxed about taking nonclinical or surrogate outcomes than are usually recommended by the GRADE Group. Second, although the four-level grading is maintained, the bottom level, which the GRADE Group describes as “very low,” has been amended to “insufficient.” Having allowed the use of an “insufficient evidence” label on the BMJ’s Clinical Evidence product when I was its Editor, I am somewhat persuaded by this experience and my minor involvement in the GRADE Working Group that this is both less “actionable” to the end user and also becomes potentially an easy way of “sitting on the fence.” The final difference in the use of the GRADE approach is the splitting out of applicability within CERs as a separate section. The stated and understandable purpose of this is to enable users from different backgrounds and in different environments to tailor their action or recommendation to their own circumstances, taking into account the “directness” score within the GRADE tool and other factors. It will be interesting to record how useful this approach is in practice, because it is often these elements of external validity that are the most
challenging decisions to make in determining whether and how to convert evidence into practice.

Although the articles published in the *JCE* are enormously valuable, there are obvious gaps, many of which are evidently work in progress. There will be considerable interest in the proposed articles covering, for example, risk of bias assessment, searching for studies, quantitative synthesis, and comparing diagnostic technologies. One element that seems almost absent from the current articles is a consideration of the importance of updating. This will be addressed in a separate article, but it is a cross-cutting issue that is central to the perceived credibility of CER reports among user communities, in particular, clinicians.

Another issue noticeable for its absence is consideration of the role of collaborative work with other agencies and groups. It is, therefore, gratifying that there is a paper in preparation that addresses the issue of using existing systematic reviews to replace de novo processes in CERs. A major benefit of an important agency, such as AHRQ, describing its methods in this way, is to promote convergence in approach between bodies responsible for conducting systematic reviews. These articles and the *Methods Guide for CERs* are an important step forward in this respect. The challenge remains to influence evidence “sceptics” in the United States and elsewhere, and these articles will be influential in addressing many of the criticisms made from such quarters. What is also needed is a greater sharing of expertise and ideas within the Evidence Based Health Care community. Perhaps, organizations, such as AHRQ, the Cochrane Collaboration, and many others, could and should do more to compare their processes and explore the differences in pursuit not only of refining their own methods but also of encouraging shared approaches where appropriate between organizations and reducing duplication of effort. As an example, there are currently 2,000 active Cochrane reviewers in the United States, and it would be possible to increase the overall capacity substantially if there was more exchange and interaction between such organizations sharing core beliefs. These articles are a great step forward, but as recent experience in the United States shows, there are many challenges ahead.

References


