ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 1198−1207

Journal of Clinical Epidemiology

Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program

David Atkins[a], Stephanie M. Chang[b,*], Gerald Gartlehner[c], David I. Buckley[d], Evelyn P. Whitlock[e], Elise Berliner[b], David Matchar[f,g]

[a]*Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development, Washington, DC, USA*
[b]*Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, Rockville, MD, USA*
[c]*Department for Evidence-Based Medicine and Clinical Epidemiology, Danube University, Krems, Austria*
[d]*Oregon Evidence-Based Practice Center, Oregon Health and Science University, Portland, OR, USA*
[e]*Oregon Evidence-Based Practice Center, Center for Health Research, Kaiser Permanente Northwest, 3800 N Interstate Avenue, Portland, OR, USA*
[f]*Duke Center for Clinical Health Policy Research, Durham, NC, USA*
[g]*Duke-NUS Medical School, Singapore*

Accepted 5 November 2010; Published online 3 April 2011

Abstract

Objective: To describe a systematic approach for identifying, reporting, and synthesizing information to allow consistent and transparent consideration of the applicability of the evidence in a systematic review according to the Population, Intervention, Comparator, Outcome, Setting domains.

Study Design and Setting: Comparative effectiveness reviews need to consider whether available evidence is applicable to specific clinical or policy questions to be useful to decision makers. Authors reviewed the literature and developed guidance for the Effective Health Care program.

Results: Because applicability depends on the specific questions and needs of the users, it is difficult to devise a valid uniform scale for rating the overall applicability of individual studies or body of evidence. We recommend consulting stakeholders to identify the factors most relevant to applicability for their decisions. Applicability should be considered separately for benefits and harms. Observational studies can help determine whether trial populations and interventions are representative of ''real world'' practice. Reviewers should describe differences between available evidence and the ideally applicable evidence for the question being asked and offer a qualitative judgment about the importance and potential effect of those differences.

Conclusion: Careful consideration of applicability may improve the usefulness of systematic reviews in informing practice and policy. Published by Elsevier Inc.

Keywords: Applicability; Generalizability; External validity; Heterogeneity of treatment effect; Comparative effectiveness; Systematic review

1. Introduction

A defining characteristic of comparative effectiveness research is that it includes ''the conduct and synthesis of research comparing the benefits and harms of different interventions… in 'real world' settings'' with the purpose of determining ''which interventions are most effective for which patients under specific circumstances'' [1]. A comparative effectiveness review must therefore make judgments about whether the available research evidence reflects ''real world'' practice and should make clear for which patients and which circumstances the review's conclusions can be used to make clinical or policy decisions. Existing guidance on conducting systematic reviews has

Funding support from the Agency for Healthcare Research and Quality's Effective Health Care Program.

Disclaimer: The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of Agency for Healthcare Research and Quality, the Veterans Health Administration, or the Health Services Research and Development Service. Therefore, no statement in this report should be construed as an official position of these entities, the US Department of Health and Human Services, or the US Department of Veterans Affairs.

* Corresponding author. Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, USA. Tel.: +301-427-1490; fax: +301-427-1520.

E-mail address: stephanie.chang@ahrq.hhs.gov (S.M. Chang).

0895-4356/$ - see front matter Published by Elsevier Inc.
doi: 10.1016/j.jclinepi.2010.11.021

What is new?

Key points

- The Patient, Intervention, Comparator, Outcome, Setting framework is a useful way of organizing the review and presentation of factors that affect applicability.
- Input from clinical experts and stakeholders can help identify specific study elements that should be routinely abstracted to examine applicability.
- Population-based surveys, pharmacoepidemiologic studies, and large case series or registries of devices or surgical procedures can be used to determine whether the populations, interventions, and comparisons in existing studies are representative of current practice.
- Reviewers should assess whether benefits or harms vary along with differences in patient or intervention characteristics (i.e., effect modification) or with differences in underlying risk.
- Reports should clearly highlight important issues relevant to applicability of individual studies in a ''Comments'' or ''Limitations'' section of evidence tables and in text.
- Metaregression, subgroup analysis, and/or separate applicability summary tables may help reviewers, and those using the reports see how well the body of evidence applies to the question at hand.
- Judgments about applicability of the evidence should consider the entire body of studies.
- Important limitations of the applicability of the evidence should be described within each summary conclusion.

focused on the risk of bias in individual studies and judging whether conclusions of the review are internally valid, rather than this equally important aspect of the review process [2].

A variety of terms have been used to describe this aspect—*applicability, external validity, generalizability, directness, and relevance*. Shadish et al. [3] defined *external validity* as ''inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments, and outcomes.'' The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group has used the term *directness* to cover applicability and other distinct aspects of the relationship between the evidence and making recommendations [4]. We prefer *applicability*, which we define as the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under ''real-world'' conditions. This better reflects the perspective of reviews conducted by the Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program and by many other groups (e.g., guideline developers) in which systematic review aims to answer specific clinical or policy questions involving particular populations and then must make judgments about whether the available evidence is *applicable* to the questions at hand.

Relatively few clinical trials are designed with applicability in mind, and, furthermore, clinical studies typically report that only a few of the factors needed to fully assess applicability. In contrast to the accumulating body of empirical data on factors affecting the risk of bias, or internal validity, there have been much less empirical data to determine which factors affect applicability. For these reasons, to date there has not been any detailed guidance for assessing applicability of evidence in producing systematic reviews.

This article outlines specific steps to ensure that systematic reviews describe and characterize the evidence so that users of a review can apply it appropriately in their decisions. The first step, identifying factors that may affect applicability, should be considered at the very earliest stages of a review, when defining key questions and the populations, interventions, comparators, and outcomes of interest. Defining inclusion and exclusion criteria inevitably takes into account factors that may affect the applicability of studies—for example, reviews meant to inform decision makers in developed countries exclude studies in developing countries because they may not be applicable to the patients and health care settings in Western countries. This article focuses on subsequent steps in a review to describe a systematic but practical approach for considering applicability in the process of reviewing, reporting, and synthesizing evidence from eligible studies.

To develop this guidance, we searched the literature using the terms *applicability* and *external validity* and reviewed our own experience working with users of reviews produced by the Evidence-based Practice Center (EPC) program. We extracted specific study characteristics that were proposed as relevant to external validity or applicability in the literature; the article of Rothwell [5] provided an extensive list to which we added from other literature, prioritized based on the experience of our program, and organized under the Patient, Intervention, Comparator, Outcome, Setting (PICOS) framework. We presented draft guidance at in-person meetings of the EPC program and circulated multiple drafts for review by EPC investigators. Parts of an earlier draft were posted for public comment. The final guidance document has incorporated peer and public review comments.