Quality assessment tools for observational studies: lack of consensus

It is a good development that systematic reviews nowadays take a broader perspective about including different study types than 10 years ago, when many reviews of interventions focused on randomised trials only. Considering different types of studies within one review allows questions to be answered, which might have been neglected previously, for example about long-term or important but rare adverse effects of interventions, which cannot be captured in randomised trials. Also, for many interventions it is actually very difficult to carry out randomised trials, for example if they are used in rare (orphan) diseases.

Observational studies may be the only feasible alternative. It remains a dilemma that no gold standard exists with which to evaluate the quality (external and internal validity) of observational studies. A large variety of tools and checklists are available but the large numbers of these probably reflect the lack of consensus between researchers as to which one is the best.

The reviews presented in this issue contain various observational studies. The lack of a good standard means the validity of these studies is difficult to judge. Shamliyan et al. highlighted this issue in their recent systematic review of 46 scales and 51 checklists to assess the quality of observational studies. Overall they found that the checklists and scales varied in their content, validity and applicability to different study designs. Of particular note was that the essential criteria of quality (allocation concealment, intention to treat, sample size) were infrequently reported. There was no consensus around the individual criteria of validity and ranking overall quality. It was concluded that numerical scores were meaningless when examining the quality of studies in systematic reviews, in part due to their lack of transparency. However none of the available tools could discriminate poor reporting from the quality of the studies and did not give separate conclusions about external and internal validity.

Shamliyan and colleagues were unable to provide recommendations as to what should be used for carrying out quality assessments of observational studies. Subjective judgements in the evaluation process should be avoided. A previous Health Technology Assessment quality assessment tools were potentially useful for systematic reviews of non-randomised studies, but all omitted key quality domains and therefore each would require refining.

Alternative checklists are available via the equator network website (http://www.equator-network.org) but caution should be used here as these lists provide reporting checklists rather than quality checklists. For example, both the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (http://www.strobe-statement.org) and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) checklist for meta-analyses of observational studies are both reporting checklists and were not developed to consider quality.

More efforts should focus on developing quality assessment tools for non-randomised studies, possibly by refining existing tools. Future collaboration is essential to determine consensus on criteria and to develop checklists for transparent quality assessment of observational research.

References


© 2010 The Authors
International Journal of Evidence-Based Healthcare © 2010 The Joanna Briggs Institute