Summary HTA

german agency for HTA of German Institute of Medical Documentation and Information

HTA-Report | Summary

Comparison of tools for assessing the methodological quality of primary and secondary studies in health technology assessment reports in Germany

Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U

German Agency for HTA
at DIMDI (DAHTA)
Waisenhausgasse 36-38a
50676 Köln
Germany

Tel.: +49 221 4724-525
Fax: +49 221 4724-444
dahta@dimdi.de
www.dimdi.de

Health political background

Healthcare policy decisions should be based on the best available scientific evidence. Scientific evidence is based on the synthesis of study results, which are if possible unbiased and thus have a high credibility.

Scientific background

Quality assessments to evaluate the credibility of studies is an inherent component of HTA reports (HTA = Health Technology Assessment) and systematic reviews. There are various quality assessment tools (QAT) that rate the extent of systematic distortion in study results by confounding or bias (internal validity).

There is no gold standard for assessing the study quality, since the true associations of exposures/interventions and outcomes are unknown. The existing tools for assessing study quality can be classified into scales, checklists and component ratings. In a scale, each item receives a numerical rating that will be added to a sum score. Scales are no longer recommended, because they do not reflect the correct extent of validity. A checklist consists of at least two items without a numerical rating system. The component rating includes components like “randomization” and “blinding”, which are also not evaluated numerically, but qualitatively.

In this report methodological quality that is used synonymously with the expression study quality and must be distinguished from the reporting quality, which is not part of this report.

The quality of health economic studies is determined by (a) the validity of study results, (b) the compliance with methodological standards of health economic evaluation and (c) the access to appropriate cost data. The methodological standards of health economic evaluations are described in health economic literature and international guidelines for providing health economic evaluations. Health economic evaluations are based on the theoretical concepts of welfare economics and decision analysis. The standards of economic evaluation have reached a broad consensus regarding the constitutive elements of health economic evaluation and approaches to cost analysis and outcome determination. Nevertheless, some guidelines recommend different approaches to be used. The elements of health economic evaluation contain (1) the justification and the choice of the evaluation type, (2) the identification and the selection of comparators, (3) the perspective, (4) the identification of resource use and costs, (5) the identification of all relevant effects and benefits, (6) the declaration of the time horizon, (7) modelling, (8) discounting, (9) incremental analysis, (10) uncertainty analysis.

All HTA reports are available for free of charge as full texts in the DAHTA database (only in German) and at German Medical Science (GMS).

Within the scope of the

Bundesministerium für Gesundheit

Research questions

What QAT are available to assess the quality of systematic reviews/HTA reports, intervention studies, observational studies, diagnostic studies and health economic studies, how do they differ among each other and what conclusions can be drawn from these results for quality assessments?

Methods

A systematic search of relevant electronic databases from 1988 onwards is done to identify QAT, supplemented by screening of the references of the HTA reports of the German Agency for Health Technology Assessment (DAHTA) and in addition an internet search. Formal characteristics and substantive elements of the tools are extracted. The substantive elements of the QAT are extracted specific to systematic reviews, intervention studies, observational studies, diagnostic studies, and health economic studies. The literature search, the data extraction and the quality assessment are carried out independently by two reviewers. Different ratings of the reviewers are solved by consensus.

The content of tools for the quality assessment of systematic reviews, intervention studies, observational studies, and diagnostic studies is extracted by using modified criteria lists. The elements of the lists are made up of study characteristics, which have either empirically demonstrated evidence of an effect on the level of the study results or its distorting effect on study results is generally accepted. The elements for study characteristics of systematic reviews, intervention studies, and observational studies are summarised in several domains. Out of all elements, those elements with empirical evidence as a potential source of bias or elements being classified on a theoretical basis as essential for internal validity are defined as relevant elements.

In order to provide a basis for the selection of a tool, only generic tools and their elements of internal validity are considered. Furthermore, the presence of sufficient operationalisation is required. The tools are distinguished by the total number of covered elements, covered relevant elements, and covered domains. Tabular summaries of the results are prepared for each study design and the results across the QAT are assessed qualitatively to identify more and less comprehensive tools.

For the data extraction of the basic elements of health economic studies, a form is developed, because there are no systematic reviews that can provide a basis for the data extraction. In the first step of the development process health economic literature and current national and international guidelines for creating health and pharmacoeconomic studies are screened. Literature and guidelines address mainly similar topics (elements of health economic evaluation). In the second step, the key elements are worked out to investigate the relation to study quality (internal validity) of health economic studies. Domains and items are developed based on the elements of health economic evaluations adapted from literature and guidelines. Domains and items are transferred into a form for analysing the quality assessment tools for health economic evaluation studies. This form helps to extract the various tools. In the development of domains and items, effort is made to ensure that items relate primarily to the internal validity.

In the health economic extraction form a gradation for rating the different items is made as such: “appropriate”, “justified”, “reported” and “missing”. If

a quality assessment tool asks for a special item addressed in a study, a rating with “reported” is made (e. g. perspective of analysis, outcome parameter or discount rate). An item is rated with “justified”, the quality assessment tool asks for the rationale for choosing a special specification. The rating “appropriate” is assigned when the quality assessment tool asks for the adequacy of used methodology in an item.
In order to find out about problems in the practical application of tools, a workshop is conducted. Objectives of the workshop are to exchange and discuss user experiences with quality assessment tools for intervention studies, requirements, and content of tools on the quality of intervention studies. These discussions will examine practical issues that are rarely discussed in the literature. A consensus on individual aspects is not pursued. The target audience include authors of the German HTA reports and systematic reviews of the German Institute for Medical Documentation and Information (DIMDI) and the Institute for Quality and Efficiency in Health Care (IQWiG), experts in the field of methodology, researchers (from the disciplines of medicine conducting public health, epidemiology, prevention, health economics), involved in healthcare policy-relevant evaluations, as well as institutes/associations conducting systematic reviews. Topics are introduced by presentations of invited experts followed by moderated discussions. Presentations and discussions are documented by audio recordings and transcriptions.

Results

The extensive literature search yields a total of 147 tools to assess the study quality: 15 for systematic reviews/HTA reports/meta-analysis, 80 for intervention studies, 30 for observational studies, 17 for diagnostic studies and 22 for health economic studies. Among the QAT are 16 tools that can be used both for intervention and observational studies.
An initial screening of HTA reports in the DAHTA database indicates that a quality assessment was reported in 87 % of the identified documents. However, in only half of these reports the chosen QAT was mentioned.
The tools show a wide variation of the formal and content characteristics. Some tools contain not only items of internal validity, but also of reporting quality and external validity. Design-specific generic tools for the assessment of systematic reviews/HTA reports/meta-analysis, intervention studies, observational studies and diagnostic studies are identified, which cover most elements for internal validity, most of the domains with at least one, or 50 % of the contained elements as well as the most relevant elements. More and less comprehensive tools can be distinguished.
The tools that examine the quality of health economic studies also reveal significant differences both in the consideration of various topics, as well as in the assessment of quality. In addition, substantial differences exist in the operationalisation of the items. Across all study designs, none of the included tools meet all elements.
A total of 27 people from HTA and EBM-associated (EBM = evidence-based medicine) institutions take part in the workshop. The following discussion points are suggested by the participants: the external validity as a part of assessment tools, the subjectivity of the assessment process, dealing with low reporting quality, endpoint versus study related quality assessment and

incorporation of the results of the quality assessment. As consensus at the workshop is not intended, individual opinions are presented. External and internal validity should be assessed separately from each other. Items, which leave much room for subjective ratings, lead to a lack of interrater reliability and result in a high need for discussions. This can be avoided by a precise and detailed operationalisation of the items.

Discussion

The quality of studies can be defined in various ways. It is a dominating view that an assessment of study quality can either express the level of internal validity or the possibility of distortion. However, the inventory of the numerous identified tools shows that many of them include the assessment of reporting quality. Mixing the reporting quality and the internal validity can lead to a misinterpretation of the study quality, if the elements of the reporting quality are used as a surrogate for assessing the methodological quality.
Based on the tabular presentation of covered content items, the identified QAT can be compared. However, this approach has limitations, since there is no scientific consensus on the necessary elements of the internal validity, and not all of the generally accepted elements are based on empirical evidence. Therefore, the highest possible number of covered elements is not necessarily an indication of an appropriate tool.
For further differentiation of the QAT, the number of covered relevant elements is presented. While for relevant elements of intervention and diagnostic studies only evidence based elements affecting the internal validity are selected, this is true for only some of the relevant elements of observational studies and systematic reviews. Overall, the performance of relevant elements should be used cautiously to identify tools that are more or less comprehensive. Depending on the topic, it should be examined, whether all items of a chosen tool are relevant, and whether additional quality items should be assessed as a part of the assessment.
Some elements of QAT cannot be clearly assigned to the reporting quality, the internal or external validity. For example, the calculation of the required sample size is only associated with the precision of the results without affecting the size of the effect estimator. However, the precision of the effect estimates may affect the significance of the results.
Not all the tools ever used have been found. However, the possibility of having missed important and frequently applied tools is low, since different data sources including the internet were screened.
In general, the higher the scope for subjective assessments, the lower the agreement between the reviewers is. Therefore, every item of a tool should be operationalised as detailed and precisely as possible. Where necessary, the instructions can be adjusted to ensure that all reviewers are clear on how to score study quality. About 40 % of the included tools provide more detailed guidance for assessment.
The quality assessment of health economic studies is an essential part of creating HTA reports. A total of 22 health economic QAT is identified. There are considerable differences regarding:

- the number of included items of the health economic extraction form (elements of health economic evaluation)
- the assessment quality: appropriate – justified – reported
- the diversity of quality sampling

None of the analysed QAT covers the whole range of relevant themes (elements of health economic evaluation). Only few consider most domains of the extraction form. Only three tools check the adequacy of the methodological procedures. Many tools ask for the methodological adequacy in few items. None of the QAT defines what is meant with adequacy. Most tools demand a justification for the methodological procedures or analyse, which items are reported.
Significant differences also exist in the sophistication of the quality assessment. The question how differentiated an assessment tool discusses the different elements of health economic evaluation can be answered by the number of items in a QAT. Because a tool is based only on few items, questions have to be more generally introduced. Reviewers will have a considerable scope for interpretation. For extensive tools with a great number of items, they can be operationalised to be more specific, so the scope for interpretation will be significantly reduced and more objective assessments are supported.

Conclusions

The quality assessment of studies is a mandatory part of systematic reviews, and has to be documented transparently. There are different, design-specific QAT available that can be selected according to their substantive coverage of the elements of internal validity.
There is consensus that scales should not be used for quality assessments or should be used without quantitative assessment. To minimise the subjectivity of the evaluation, tools with a detailed and precise operationalisation of the items are preferable. If possible, the chosen tool should be tested in a few studies in advance to check if the operationalisation of the items needs to be supplemented or clarified to minimise the subjectivity of the evaluation and to ensure uniform scoring of all reviewers.
Further research is needed to identify study characteristics that influence the internal validity of studies, especially for observational studies. So far, there is no evidence that qualitative overall assessment of study quality is correctly associated with the internal validity.
For assessing the quality of health economic studies, tools should be developed, which (1) cover all relevant elements of health economic evaluation, (2) assess the appropriate use of methodological procedures and (3) differentiate the various topics sufficiently. The adequacy should be based on the standards of health economic evaluation (defined by standard literature and international guidelines). Advice for filling in and operationalisations should be part of the assessment tools and, in addition, adequacy should be accurately described and defined.