



Mariska M. G. Leeflang, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, The Netherlands

Individual study quality in meta analyses of diagnostic test accuracy

Health care professionals, who are looking for the best information about the accuracy of a diagnostic test, turn to systematic reviews more and more. These systematic reviews often contain one or more meta-analyses, in which the results of individual studies are summarized as one overall estimate. This summary estimate, however, can be biased when the individual studies are of poor methodological quality. Then the question arises: if we know how these studies are biased, can we then adjust for these methodological flaws and get a better summary estimate?

INTRODUCTION

Diagnostic test accuracy refers to the ability of a test to discriminate between those who have and those who do not have the target condition. Accuracy is assessed by comparing the results of the index test, the test under evaluation, with the results of the reference standard, which aims to classify patients as having or not having the target condition. Test accuracy is most often expressed as the test's sensitivity (the proportion of those with the target condition who have a positive index test result) and specificity (the proportion of those without the target condition who have a negative index test result). These sensitivity and specificity are not the same as the analytical sensitivity (detection limit, ability to measure an analyte) and specificity (freedom from interference by compounds other than the analyte).

In meta-analyses of diagnostic test accuracy, the sensitivity and specificity from individual diagnostic studies are often combined into

one overall estimate of diagnostic test accuracy: the Diagnostic Odds Ratio (DOR), an overall measure for diagnostic accuracy that combines both sensitivity and specificity. The methodological quality of the original studies is important for the results of a meta-analysis: if the raw material is flawed, one should question the summary estimate. Poor quality studies are more likely to have errors in estimating the outcomes of interest, and to include these estimates in a meta-analysis may result in a flawed overall estimate. Therefore, authors should take into account the methodological quality of the studies they include in their meta-analysis.

The first step in this process, is to assess the methodological quality of the individual studies that will be included in the systematic review or meta-analysis. Many checklists have been developed for this task and the one we used is the QUADAS checklist, developed by Whiting et al (BMC Methodology, 2003). The





results of this quality inventory can be presented in a summary graph, representing the percentage of included studies that fulfilled each checklist-item (see Figure 1 for an example). It can also be presented in a summary table, with the results of each checklist item listed for each study individually (by using symbols as plus, minus, question mark). Such a presentation tells the reader something about the overall quality of the studies in the review, but it will not tell the reader how the flawed studies affect the results of the meta-analysis. The effects that one quality aspect or checklist item may have on the summary estimate, will be different from the effect of the combination of the total off quality aspects, 'the' methodological quality of a certain study.

To investigate the effect of individual study quality on the DOR in meta-analyses, we studied three methods to incorporate the quality assessment results. Because biased studies tend to overestimate diagnostic accuracy and introduce more heterogeneity in the results, we hypothesized that adjustment since quality produces less optimistic results and narrower confidence intervals.

METHODS

We used the DOR as a measure of diagnostic accuracy, and recalculated a summary diagnostic odds ratio of 31 existing meta-analyses, following different strategies to incorporate quality results. We tested one strategy and

named it the 'restrict' strategy. Following this strategy, we restricted our meta-analysis to 'high-quality' studies. To be classified as high quality, a study had to fulfil certain quality items, if these items were not fulfilled, then the study was not of high quality. Items that we included in this definition were verification of the test results and blinding of both the test under evaluation and the reference standard.

Two other strategies were variations of a multivariable regression model. All three strategies were compared to the baseline strategy, ignoring study quality and inclusion of all studies that were also originally included in the meta-analysis. The DOR of the 'ignore' strategy was compared to the DOR that we found with each one of the other strategies. Because we were also interested in heterogeneity, we also compared the confidence intervals of the DOR from the 'ignore' strategy with the confidence intervals of the DOR from each one of the other strategies.

RESULTS

We found no significant differences between the estimated DOR of the 'ignore' strategy and of the other strategies. Neither did we find that there was less heterogeneity among the strategies that adjusted for quality. What we did find was that many studies did not report whether the interpretation of the tests was blinded or not. Another finding was that only 15% of the 487 studies that were included in

the 31 meta-analyses together, could be classified as 'high quality study'. This means that only 15% of these studies fulfilled all items included in our high quality definition (verification of the test results and blinding of both tests under evaluation and the reference standard). As a result, we could only re-analyze 11 reviews of the 31 using the 'restrict' strategy. ▶▶



DISCUSSION

Meta-analysts and authors of systematic reviews can use different methods to tell the reader something about the effect of study quality on their summary estimates. This can be done by assessing the effect of individual quality items, but then the overall effect may still be a completely different one.

One of the strategies that authors can use to give at least a credible summary estimate is to restrict the meta-analysis to studies that have a good methodological quality. By excluding poor quality studies, these authors throw away a lot of (still very useful) information and may end up with only one or two studies in their meta-analysis. Another risk, especially when the authors decide ad-hoc that they will only report the results of the 'good' studies, is that the definition of high quality is chosen in such a way that there will be enough studies left for the meta-analysis.

But what is the definition of good quality worth?

Another strategy can be to include all relevant quality items as covariate into the meta-regression model. 'Relevant' can mean that these items are first tested in an invariable model and that only the items with a significant effect on the DOR will be included into the final regression model. But 'relevant' can also mean that the authors had reasons to decide beforehand that certain items are relevant and others

not, for example based on earlier studies. This method has also disadvantages, for example that significance does not necessarily mean that the item is indeed relevant. More in general, the effect of individual (relevant) quality features may not be always be in the same and predictable direction. Especially not if different features have opposite effects on the summary estimate.

A strategy that we did not investigate, is the calculation of summary quality scores for studies and labelling any study exceeding a certain threshold score as high quality. Such summary quality scores have been extensively studied - and criticized - in systematic reviews of intervention studies. Different shortcomings in study design may cause different forms of bias, making it almost impossible to determine the weight that should be given to each quality item in calculating such quality scores. The same counts for sequential analysis of the studies based on their quality ranking, which would lead to a quality-adjusted cumulative meta-analysis. This strategy also requires a hierarchical approach of study quality by assuming that some criteria are more important than others.

Despite all these disadvantages of incorporating study quality into meta-analyses, study quality should be included in a systematic review in one way or the other. After all, poor quality will at least affect the trustworthiness of the conclusions of that review. We feel that

the results of quality assessment could be summarized in a table or in a figure. When results for all of the included studies are plotted in ROC-space, colour or symbol coding of individual studies can be helpful in helping the readers to recognize characteristics of individual studies. Another solution may be to adjust for study quality, but to report also the unadjusted results (and test for significant differences between the two methods).

Methodological quality is a multidimensional concept in which the importance of individual items will vary from one research project to another. To gain more insight in this concept, we will need studies that fully report the information about their methods and study designs. Until then, adjusting for quality differences in meta-analysis will remain tricky and should be read with cautiousness. ■



The Molecular Diagnostic conference is taking place on 29th January 2008 at the Middle East Laboratory Management and Medicine Congress, Dubai International Convention and Exhibition Centre

