Strengthening methods of diagnostic accuracy studies
Ochodo, Eleanor

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Survey revealed a lack of clarity about recommended methods for meta-analyses of diagnostic accuracy data

Eleanor A. Ochodo, Johannes B. Reitsma, Patrick M. Bossuyt, Mariska M.G. Leeflang

Abstract

Objectives: To collect reasons for selecting the methods for meta-analysis of diagnostic accuracy from authors of systematic reviews and improve guidance on recommended methods.

Study design and Setting: Online survey in authors of recently published meta-analyses of diagnostic accuracy.

Results: We identified 100 eligible reviews, of which 40 had used more advanced methods of meta-analysis (hierarchical random-effects approach), 52 more traditional methods (summary receiver operating characteristic curve based on linear regression or a univariate approach), and 8 combined both. Fifty-nine authors responded to the survey; 29 (49%) authors had used advanced methods, 25 (42%) authors traditional methods, and 5 (9%) authors combined traditional and advanced methods. Most authors who had used advanced methods reported to do so because they believed that these methods are currently recommended (n = 27; 93%). Most authors who had used traditional methods also reported to do so because they believed that these methods are currently recommended (n = 18; 75%) or easy to understand (n = 18; 75%).

Conclusion: Although more advanced methods for meta-analysis are recommended by The Cochrane Collaboration, both authors using these methods and those using more traditional methods responded that the methods they used were currently recommended. Clearer and more widespread dissemination of guidelines on recommended methods for meta-analysis of test accuracy data is needed.
5.1 Introduction
The last few years have witnessed a large increase in the need to make evidence-based decisions about the use and interpretation of medical tests. [1,2] One way of making valid statements about the accuracy of tests is by systematically analyzing results in previously published and unpublished primary studies. Accuracy is defined as the ability of a test to discriminate between patients with and without the disease of interest. Within a systematic review, accuracy results of prior studies can be included to generate a single and more precise summary estimate, and to analyze sources of heterogeneity, a process referred to as meta-analysis. Meta-analyses, if rigorously prepared, can objectively summarize results of prior studies, help identify the risk of bias in primary studies, and improve the reliability and accuracy of conclusions and recommendations. [3-5]

The challenge in diagnostic accuracy studies is that there are usually two outcome measures of interest: sensitivity and specificity, for example, or positive and negative predictive values. These two measures of accuracy can be negatively correlated, in particular when studies applied different thresholds to define a positive test result. [4, 5, 6] Additionally, primary diagnostic studies tend to have small sample sizes, are carried out in diverse settings and, as a consequence, can display substantial variability in study results. [4-9]

Different methods for meta-analyzing diagnostic accuracy data have been proposed in the last 20 years. [10] The earlier introduced and more traditional methods include independent pooling of sensitivity and specificity [11], pooling of diagnostic odds ratios, [11] pooling of likelihood ratios [11, 12] and generating a summary receiving operating characteristic (SROC) curve based on linear regression. [13, 14] More advanced methods, proposed in the last decade are hierarchical methods, such as the hierarchical SROC, [15] the bivariate random effects models, [16] and trivariate analysis of sensitivity, specificity and prevalence. [17]

Univariate or independent pooling of accuracy measures do not account for correlations between sensitivity and specificity. Ignoring this correlation can
underestimate accuracy and may provide misleading results. [4, 6, 18] Generating a SROC curve based on linear regression (Moses-Littenberg model) neither fully accounts for imprecision of study estimates nor between-study heterogeneity. [4, 6] The advanced methods have been shown to be more statistically sound and flexible than the traditional methods. [4, 6, 19] Unlike traditional methods, advanced methods typically take into account both the within and between study variability and estimate the correlation between sensitivity and specificity.

The Cochrane collaboration is an international organization that helps people make well-informed decisions about health care by promoting the preparation and use of systematic reviews. [20] A few years ago, the collaboration also started including systematic reviews of test accuracy studies in the Cochrane library. The Cochrane Methods group currently recommends the use of two hierarchical methods of meta-analysis: the hierarchical SROC method and bivariate random effects methods. [6]

Despite advanced methods being available and recommended, previous reports have shown that the uptake of these methods of meta-analysis is slow. A majority of authors still use more traditional methods of meta-analysis. [2, 21] To understand why, and to improve guidance on the recommended methods to use, we asked authors of diagnostic accuracy reviews about selecting the type of methods for meta-analyzing the data in their publication.

5.2 Methods

Since diffusion of novel methods takes time, and dissemination typically progresses slowly, we wanted to focus on recently published reviews. To collect a sample of recently published quantitative diagnostic accuracy reviews, we searched MEDLINE for articles published between September 2011 and January 2012. This search was done in February 2012 by one author (E.O) using the following search strategy: (systematic[sb] AND ("diagnostic test accuracy" OR DTA[tiab]
OR "SENSITIVITY AND SPECIFICITY"[MH] OR SPECIFICIT*[TW] OR "FALSE NEGATIVE"[TW] OR ACCURACY[TW])}.

To be eligible for this study, articles should be reviews of published diagnostic accuracy studies, published in English, with a meta-analysis. We excluded meta-analyses of individual patient data as the methodology of such studies differ from those of meta-analysis of published data. [22]

Two authors (E.O and M.L) extracted the method of meta-analysis used in the eligible articles by reading the full text of the articles for the methods employed and also by examining the references cited. Disagreements were resolved through discussion and consensus. We then classified the method of meta-analysis used into three groups: The traditional methods group, the advanced methods group, and the combined traditional and advanced methods group (for those that used both methods). The traditional methods of meta-analysis included independent pooling of sensitivity and specificity, the summary receiving operating characteristic method based on linear regression (Moses-Littenberg model), pooling of diagnostic odds ratios, and independent pooling of likelihood ratios. The advanced methods are the hierarchical models: bivariate logitnormal random effects meta-analysis, hierarchical summary receiving operating characteristic model (HSROC) and trivariate analysis of sensitivity, specificity and prevalence.

We then designed an online questionnaire using the software SurveyMonkey and pre-tested it on four authors of previously published diagnostic test accuracy reviews. We asked authors both general questions relating to the methods used for analysis and specific questions on reasons they selected the method of meta-analysis in their publications. The responses to the reasons of using the meta-analysis methods were collected using a 5 point Likert scale (Strongly Agree, Agree, Neither Agree or Disagree, Disagree, Strongly Disagree). The questionnaire can be found in Appendix 1.
We released this survey to email addresses of the corresponding authors of the included diagnostic test accuracy reviews. The email addresses of these authors were extracted from the publications. Of fourteen publications, the corresponding author was also a corresponding author for another publication. In this situation, we sent the survey to a co-author, whose email address was provided in the publication. If the email addresses of other co-authors were not provided, only the most recent publication of the corresponding author was included in the survey.

We sent two reminders to authors who did not respond to the survey. The first and second reminders were sent 10 days and 30 days respectively after the release of the initial survey. This survey was run for five weeks (28th March to 28th April 2012).

5.3 Analysis
We analysed the level of inter-rater agreement in scoring the methods for meta-analysis. The survey results were downloaded from SurveyMonkey into an Excel sheet. In reporting, we collapsed the Likert scale results into 3 categories (Agree, Neither Agree or disagree, Disagree) and calculated the proportion of responses to each Likert item. We analysed these reasons per the type of meta-analysis used in their publication (Traditional and Advanced method).

5.4 RESULTS
5.4.1 Search results
The initial search identified 1,335 articles. After screening titles and abstracts of these articles, 1,183 articles were deemed ineligible. After reading the full texts of the remaining 152 potentially eligible abstracts, 48 were excluded. Four articles had corresponding authors identical to authors already included and were, therefore, excluded. The reasons for exclusion are shown in Appendix 2. The absolute level of agreement (between E.O & M.L) for scoring the method of meta-analysis used was 87% (kappa 0.77 [95% CI: 0.61 to 0.93]) before discussion and 100% after discussion.
The survey was sent to 100 authors. Of these, 40 authors (40%) had used more advanced methods of meta-analysis, 52 authors (52%) more traditional methods and 8 (8%) combined both advanced and traditional methods. A majority of corresponding authors were first authors (n=49, 49%) followed by last authors (n=38, 38%).

Fifty-nine authors responded to the survey, giving a response rate of 59%. We excluded one respondent who had performed an individual patient data (IPD) meta-analysis. This respondent had obtained separate estimates of sensitivity and specificity, with justification. Of the 58 respondents included in our analysis, 29 (50%) had used advanced methods of meta-analysis, 24 (41%) had used traditional methods of meta-analysis, and 5 (9%) combined traditional and advanced methods. A majority of the 58 respondents were first authors (n=34, 59%) followed by last authors (n=17, 29%).

5.4.2 Survey results

5.4.2.1 General results

In response to the question “Which method of meta-analysis are you familiar with?” most authors responded that they were familiar with separate/independent pooling of estimates of sensitivity and specificity (n=44, 76%) followed by the bivariate random effects meta-analysis (n=40, 69%). (Figure 1)
Two-thirds of authors had a statistician involved in conducting the meta-analysis (n=37, 64%), with more authors in advanced methods group (n=21/29, 72%) using a statistician compared to the traditional methods group (n=11/24, 46%). All authors who combined traditional and advanced methods had consulted a statistician (n=5, 100%).

Overall, the most frequently used software packages to perform meta-analyses were Stata using metandi (n=23, 40%), Review Manager (n=21, 36%) and MetaDiSc (n=20, 35%). Less often used software packages were; Stata using midas (n=12, 21%), SAS (n=14, 24%) and R (n=6, 10%). Some authors had used more than one software package. A majority of authors in the traditional methods group (N=14) used Meta-DiSc (n=16, 67%) followed by Review Manager (n=7, 29%), whereas most authors in the advanced methods group (N=29) used Stata metandi (n= 16, 55%) followed by Review Manager (n=13, 45%). Those who combined traditional and advanced methods (N=5) mostly used Meta-DiSc, Stata metandi, SAS and R, in equal proportions (n=2, 40%). (Figure 2)

Forty-four authors responded to an open-ended question about the challenges encountered when conducting a meta-analysis. Responses ranged from “no
problems” to “problems in interpretation” (See appendix 3). Most authors cited challenges in performing the analysis (n=20/44), particularly in analyzing heterogeneity (n= 9/20).

5.4.2.2 Reasons for selecting methods for meta-analysis

Most authors in the traditional group (n=24) had selected that method for meta-analysis because they felt that that method was easy to understand (n=18, 75%) and was currently recommended (n=18, 75%). Seventeen authors felt that the method yielded precise estimates (71%). About two-thirds of the authors in this group disagreed that the method was the only one they knew (n=15, 62%) (See figure 3).

For the advanced methods group (n=29), most authors had used that method of meta-analysis because they felt the method was currently recommended (n=27, 93%). Seventy-five percent (n=21/28) of authors felt that this method yielded precise estimates. Like the traditional methods group, a majority in this group disagreed that the method was the only one they knew (n=21/28, 75%) (See figure 3).

In analysing the reasons for selecting methods of meta-analysis, we excluded the group that combined traditional and advanced methods of meta-analyses in their
It was unclear to which group they referred to when answering our questions.

We also asked all authors if they had considered using the hierarchical methods. A majority (n=46, 79%) responded that they did. For those who did not, we asked them about their reasoning. Six indicated that they had never heard of these methods (n=6/11, 55%) and did not know enough about these methods (n=8/11, 72%). In addition, authors were undecided about the following items ‘the methods are time consuming’ (n=8/9, 89%), and ‘I didn’t think using them will significantly change the estimates of my study’ (n=6/11, 60%) (See figure 4).

5.5 Discussion

In our sample of 100 systematic reviews with meta-analysis of test accuracy data, published between September 2011 and January 2012, we found that more than half of the articles relied on more traditional methods for meta-analysis of accuracy studies. Our survey among the study authors shows that most authors who had used traditional methods of meta-analysis felt that the methods were currently recommended and were easy to understand, while those who had used the more advanced methods also did so because they felt that these methods were currently recommended. For those who had not considered using the
advanced methods, most responded that they did not know enough about these methods. Further, most were undecided as to whether the methods were time consuming, or would significantly change study estimates.

Our study has limitations. The search of articles to be included in the review was done by one person, and despite the sensitive filter some reviews could have been missed. Nevertheless we were able to include one hundred reviews, with high agreement in identification of methods and an almost sixty percent response rate in review authors. A potential limitation is that the corresponding authors may not have been well versed in statistical methods to correctly answer our questions. One author responded “Sorry I could not give you more information to inform why the decision to use the bivariate method was made. My answers “neither disagree nor agree” actually mean unknown, as the decision was made by the statistician.”

We also did not consider the specific research question or objectives of the reviews in our study. The objective of a review may influence the choice of meta-analysis used, and the benefits of the hierarchical methods may also depend on the review question. [6]

Our study confirms findings from previous, comparable analyses, which also showed that traditional methods remain most widely used. [2, 21.] A comprehensive review by Dahabreh and colleagues of 760 diagnostic reviews in MEDLINE published between 1966 and 2009 revealed that univariate meta-analysis and SROC by Moses-Littenberg were used in 87% and 86% of the reviews respectively. Only 11% of the reviews had used advanced models such as the bivariate method and HSROC. [2] Willis and Quigley examined 236 diagnostic accuracy reviews from eight databases published before 31st December 2008. They found that independent pooling of sensitivity and specificity and using the SROC based on linear regression were used in 70% of the studies whereas the bivariate random effects and HSROC models were only used in 22% and 5% of studies respectively. These two previous studies also
showed that the uptake of the advanced methods was increasing over time. This could partly explain an increased proportion of those who used advanced methods (47%) in the diagnostic reviews included in our sample.

The analysis entailed in meta-analyses of accuracy tests can be complex and preferably a statistician needs to be consulted during the design and conduct of the analysis. [6] In our survey two thirds of authors had involved a statistician in the review. Notably, the proportion of those using traditional methods who involved a statistician was comparable to those using advanced methods. This may indicate that some statisticians are not aware of newer methods, or that they decided not to use them, for various reasons. This can also explain why those who used traditional methods believed that the method was currently recommended.

Some authors in the advanced methods group responded that they used Meta-DiSc or Review Manager to perform their meta-analysis. Meta-DiSc and Review Manager are software packages that cannot be used to perform advanced meta-analyses. [23, 24] This response may also reflect lack of knowledge of the statistical methods used in the papers by the corresponding authors. It could be that they used Meta-DiSc or Review Manager only to generate forest plots and figures, but this is unclear.

Responses from our survey generally reflect a lack of clarity or understanding of the appropriate methods to meta-analyse diagnostic accuracy data. Peer reviewers and editors of journals may also be unaware of these methods. This sentiment was echoed by one of the respondents as follows: “The issue with meta-analyses is lack of agreement on the best methods and use of the best approach... Further, a related issue is lack of understanding and interpretation of the methods by peer reviewers and editors....”

Poor uptake or lack of knowledge of advanced methods may be explained in part by the lack of clarity in communicating the benefits of advanced models over traditional methods for clinical practice. [21] Studies have shown divergent
views on the differences of these methods on summary estimates. Some studies have shown that these methods produce comparable summary estimates [25, 26] while others have shown that the advanced models produce significantly different estimates. [27] Most of these studies have been based on empirical data.

In practice, physicians or policy makers may just require a summary measure to distinguish between a useful and useless test. However, bearing in mind the variability of results of test accuracy studies, it is essential that meta-analyses account for and, if possible, explain this variability. This explanation will help physicians or policy makers make an objective assessment about the suitability of the tests to their settings and hopefully diminish the uptake of tests based on inadequate evidence. Advanced models of meta-analyses of test accuracy in this case have an upper hand. However, circumstances such as statistical models failing to converge or a few studies included in a review may hinder the use of advanced models of meta-analyses. In such cases, traditional methods can be used.[6] Nonetheless, it is important for authors to clearly state in their papers if they encountered this challenge to enable an objective assessment of their results. In our study, only two authors who used traditional methods cited the reasons above in their publications.

Our study findings may guide further implementation plans of the Cochrane Collaboration. In order to guide authors, the diagnostic test accuracy working group of the Cochrane collaboration provides annual courses on preparing reviews of test accuracy. However, these courses given are carried out at selected places. This teaching method may limit the number and scope of people who can be guided on appropriate methods of meta-analysis. The diagnostic test working group also organizes workshops during their annual colloquium on various aspects of test accuracy including meta-analysis. However, these workshops only benefit people working for or in collaboration with the organization. A handbook on performing diagnostic reviews is also available on the website (http://srdta.cochrane.org/handbook-dta-reviews).
Proposals for improvements include publishing simple and practical tutorials in medical journals and providing simple to use online tutorials with sample data sets. These online tutorials can be availed freely or at a reduced fee on the Cochrane website. These tutorials need to explain which methods of meta-analysis are currently appropriate for test accuracy data, the advantages and disadvantages of proposed methods, and need to include a simple step wise approach in conducting a meta-analysis of test accuracy. The need for simple guidance is reflected by comments of one of the respondents as follows: “There is a need for some reviews on method selection that are tailored to a less mathematically inclined audience. It doesn't need to be non-mathematical but something to help review authors work effectively with statisticians”. Another possible method of improvement is writing letters to journal editors in response to published reviews that used inappropriate methods to meta-analyse test accuracy data. These letters will, however, need to take into account the research question or the objectives of the review in order to provide an objective assessment. If accepted, these letters can increase awareness among journal editors and peer-reviewers on recommended methods of meta-analysis of test accuracy data.

As systematic reviews and meta-analyses form a fundamental part of evidence-based practice, more research and guidance is needed to convince authors which methods of meta-analyses are most appropriate and benefit clinical practice. We hope that clearer guidance to review authors will encourage more authors to conduct and publish robust diagnostic reviews that will effectively guide policy makers or clinicians.

References


Appendices

Appendices can be accessed at http://www.jclinepi.com/article/S0895-4356(13)00203-5