Meta-epidemiologic analysis indicates that MEDLINE searches are sufficient for diagnostic test accuracy systematic reviews

Wynanda A. van Enst[a,b,*,1], Rob J.P.M. Scholten[a,b,1], Penny Whiting[c,d], Aeilko H. Zwinderman[b], Lotty Hooft[a,b,1]

[a]Dutch Cochrane Center, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Meibergdreef 9, 1100 DD, Amsterdam, The Netherlands
[b]Biostatistics and Bioinformatics, Academic Medical Center, Meibergdreef 9, 1100 DD, Amsterdam, The Netherlands
[c]Kleijnen Systematic Reviews, Unit 6, Escrick Business Park, YO19 6FD, UK
[d]School of Social and Community Medicine, University of Bristol, 39 Whatley Road, Bristol BS8 2PSB, UK

Abstract

Objectives: To investigate how the summary estimates in diagnostic test accuracy (DTA) systematic reviews are affected when searches are limited to MEDLINE.

Study Design and Setting: A systematic search was performed to identify DTA reviews that had conducted exhaustive searches and included a meta-analysis. Primary studies included in selected reviews were assessed to determine whether they were indexed on MEDLINE. The effect of omitting non-MEDLINE studies from meta-analyses was investigated by calculating the summary relative diagnostic odds ratio (RDORs) = $DOR_{MEDLINE\ only}/DOR_{all\ studies}$. We also calculated the summary difference in sensitivity and specificity between all studies and only MEDLINE-indexed studies.

Results: Ten reviews contributing 15 meta-analyses met inclusion criteria for quantitative analysis. The RDOR comparing MEDLINE-only studies with all studies was 1.04 (95% confidence interval [CI], 0.95, 1.15). Summary estimates of sensitivity and specificity remained almost unchanged (difference in sensitivity: −0.08%; 95% CI −1% to 1%; difference in specificity: −0.1%; 95% CI −0.8% to 1%).

Conclusion: Restricting to studies indexed on MEDLINE did not influence the summary estimates of the meta-analyses in our sample. In certain circumstances, for instance, when resources are limited, it may be appropriate to restrict searches to MEDLINE. However, the impact on individual reviews cannot be predicted. © 2014 Elsevier Inc. All rights reserved.

Keywords: Diagnostic test accuracy; Meta-analyses; Dissemination bias; Comprehensive searches; MEDLINE; Systematic review

1. Introduction

Systematic reviews of diagnostic test accuracy (DTA) studies are important to inform evidence-based use of diagnostic tests in clinical practice. A comprehensive search across multiple databases combined with screening the search results to identify studies for inclusion in the review is a key part of any systematic review [1,2]. This process can be time consuming and costly, especially for DTA reviews that often involve screening several thousand references. Methods for efficient searching are therefore needed without introducing bias by missing relevant studies.

There are many electronic bibliographic databases that can be used to identify biomedical studies [3]. Most reviewers only search a small subset of the available databases, even in a comprehensive search. The best-known databases include MEDLINE and EMBASE. As from January 2010, MEDLINE records are included in EMBASE, whereas some EMBASE records are not covered by MEDLINE. EMBASE covers other journals especially drug therapy journals, more European journals, and more non-English journals compared with MEDLINE [4]. Regional databases like PASCAL and LILACS or specialized databases like PsychINFO may include studies additional to EMBASE and MEDLINE. Thus, if one of these databases is not searched when conducting a systematic review, there is a risk that some relevant studies will be missed.

What is new?

Key findings
- Less than half of the diagnostic test accuracy (DTA) systematic reviews (43%) included studies that are not indexed in MEDLINE.

- Omitting non-MEDLINE studies from the meta-analysis did not significantly hamper the diagnostic odds ratio, sensitivity, or specificity.

What this adds to what was known?
- This is the first metaepidemiologic evidence on the impact of search strategies for DTA systematic reviews.

What is the implication and what should change now?
- Empirical evidence indicates that searching in databases beyond MEDLINE for a DTA systematic review may no longer be regarded an absolute necessity to produce valid outcomes.

When time or financial resources are limited, simplifying the searches can be a practical solution. However, this may compromise the quality of the review by missing relevant studies. Much research has been done to develop search filters to enhance the precision of the search, defined as the number of relevant records identified by a search divided by the number of records identified. Therefore the number needed to read (NNR), defined as the number of records needed to read to find one relevant additional paper, can be reduced [5,6]. However, empirical evidence has found that even the most sensitive methodological filters for searching for DTA studies miss relevant studies [7,8].

Reducing the number of databases to be searched could reduce the amount of work involved in searching and also the NNR for screening search results and so be time and cost effective. In particular, costs will be reduced if only MEDLINE is searched, as this database is freely accessible through the PubMed interface. Empirical research has shown that excluding EMBASE when searching for randomized controlled trials (RCTs) will affect the results of intervention reviews. This is the consequence of a systematic difference between the two databases for RCTs. Trials that are indexed on MEDLINE on average find larger effects and have more significant results compared with studies indexed on EMBASE. Searching exclusively in MEDLINE may lead to an overestimation of the magnitude of treatment effects, which could affect patient management [9]. Although the publication process of trials is often dependent on identification of a significant effect, there is no such effect in diagnostic studies as the main outcomes are accuracy measures such as the diagnostic odds ratio (DOR), sensitivity, and specificity. Because of the

nature of these outcomes, it is not obvious to specify a hypothesis and test for it. Other factors may influence the publication process, but it is not clear whether these factors are associated with particular databases.

A previous review has shown that failure to search multiple databases to identify studies for inclusion in DTA reviews misses relevant studies [2]. However, this review did not investigate the impact of these missing studies on the results of the review. Restricting a review to studies indexed on a single database, for example MEDLINE, is only problematic if this leads to biased results. We would assume that reviews based exclusively on studies indexed on MEDLINE could have biased results if the results of those studies differ systematically from relevant studies indexed on other databases. We therefore aimed to assess whether restriction of databases influences the estimation of measures of accuracy in DTA reviews.

2. Methods

2.1. Identification of reviews

MEDLINE was searched through the PubMed interface to identify DTA reviews published between January 2006 and January 2011. The methodological filter of Devillé [10] was applied to identify reviews covering diagnostic test accuracy combined with the review filter that is available in PubMed to identify systematic reviews. Search results were limited to 622 journals that had an impact factor ≥4 in 2010 [11] and were accessible through the medical library of the University of Amsterdam. The complete search strategy can be found in Appendix A at www.jclinepi.com. In addition, the Cochrane Database of Systematic Reviews was searched in March 2011 for all DTA reviews. The literature search and the presentation of the review was structured according to the PRISMA guidelines [12].

2.2. Inclusion criteria

We included reviews in which the authors evaluated the diagnostic accuracy of one or more tests against a reference standard and reported measures of accuracy: the DOR, sensitivity, and specificity. We only included DTA reviews that had conducted a meta-analysis and had searched MEDLINE and at least one other biomedical database. We excluded narrative reviews, genomic reviews, animal reviews, reviews that had applied a language or quality restriction, reviews that had assessed the analytical validity of tests, and reviews that only evaluated other measures of diagnostic performance such as reproducibility and reliability. Two reviewers independently assessed titles and abstracts of the references identified by the electronic search for relevance. Inclusion screening of full-text articles was conducted independently by two reviewers. A third reviewer was consulted in case of disagreement. Only meta-analyses that included both studies indexed on MEDLINE and studies not indexed on MEDLINE were included for further analysis.

2.3. Data extraction

Descriptive characteristics (author, publication year, test under evaluation, and purpose of the test) and full references for each included primary study were extracted from each review. Data to populate two-by-two tables (the number of true positives, false positives, false negatives, and true negatives) were extracted for all individual studies of all included meta-analyses. We contacted the authors of reviews when two-by-two tables were not reported in the review. In case of no response, we sent two reminders requesting the missing data. When no reply was received from the authors, we extracted data from the primary studies ourselves.

2.4. Comprehensiveness of the searches

We aimed to assess whether the comprehensiveness of the search was associated with finding studies not indexed on MEDLINE [2]. We assessed the comprehensiveness of the search according to the AMSTAR checklist [13]. AMSTAR, a measurement tool to assess the methodological quality of a systematic review, has several items that determine the comprehensiveness of a search for a systematic review: (1) at least two electronic sources should be searched, (2) the years and names of databases should be reported, (3) key words and/or MeSH or EMTREE index terms should be provided, and (4) extra effort should be made to identify extra studies. Each of these items was scored individually, although in AMSTAR, all are scored as a single item (item 3). We also added two extra items that we believe contribute to the comprehensiveness of a search: (5) whether the search was performed without using a search filter for DTA studies as we know that can lead to missing studies [8,14] and (6) whether the full search strategy was available. Each item could be scored with 1 when the item was fulfilled or with 0 when the item was not fulfilled or unclear. Altogether, each search could thus score between 0 and 6.

2.5. Data analysis

We assessed which of the primary studies included in each review were indexed on MEDLINE by means of a known-item search. A known-item search implies a user who is looking for one particular study; in our case, the study included in the review for which we had the full reference extracted from the review [9]. References that could be identified in MEDLINE were labeled as "in MEDLINE" (iM) and those that were not as "Not in MEDLINE" (NiM).

We selected the primary meta-analysis from each systematic review which we defined as the (sub)set of clinically relevant studies on which the conclusions of the review were based. The selection process was double checked by a second reviewer for a subset of reviews (20%) and those considered most complicated to classify

(W.v.E., P.W.). For reviews that assessed more than one index test, we selected all meta-analyses that contributed to the conclusion.

First, the impact of not including NiM studies was measured quantitatively by redoing the meta-analyses and calculating the DOR, sensitivity, and specificity: once with all studies and once without NiM studies. Analyses were performed using the bivariate random-effects model [15] in Stata version 10.0 [16]. The DOR, sensitivity, and specificity were calculated with corresponding 95% confidence intervals (CIs). Per meta-analysis, we quantified the effect of excluding NiM studies by estimating the summary relative diagnostic odds ratio (RDOR) = DOR_{iM}/DOR_{iM+NiM}. We also calculated the asymptotic variance of the RDOR. The log(RDORs) were then pooled by the use of a random-effects generic inverse-variance model to estimate the average effect of restricting analyses to NiM studies. Similarly to the RDOR, the summary of the difference in sensitivity and specificity between iM and NiM was estimated. The statistical methods are explained further in Appendix B at www.jclinepi.com.

We used a sensitivity analysis to assess the impact of leaving out NiM studies for the meta-analyses with the largest proportion of NiM studies [NiM/(iM + NiM) to maximize the likelihood of finding an effect of missing studies by a search limited to MEDLINE].

We assessed if there was an association between the comprehensiveness of the search and finding all studies in MEDLINE. The comprehensiveness of the search was summarized by calculating a score for the search characteristics (1 point for each item fulfilled). This score (range, 0—6) was used as a continuous independent variable in a logistic regression.

3. Results

The searches identified 615 hits of which 116 were considered potentially relevant and were assessed for inclusion based on full-text articles. We identified 42 reviews but in 24 (57%) of these, including two Cochrane reviews, all included primary studies were indexed on MEDLINE and so these were not investigated further. The 18 reviews with contrast presented 52 meta-analyses; 39 meta-analyses were considered as primary meta-analyses as they contributed to the conclusions of the reviews. Eight reviews including 18 primary meta-analyses were excluded because of the characteristics of the primary meta-analyses (all studies included in the primary meta-analyses were indexed on MEDLINE, lack of specification of which primary studies were included in the meta-analyses, or lack of information to populate the two-by-two tables). The final 10 reviews included 18 meta-analyses of which three did not have contrast, leaving 15 meta-analyses for inclusion. The selection process is presented graphically in Fig. 1, and
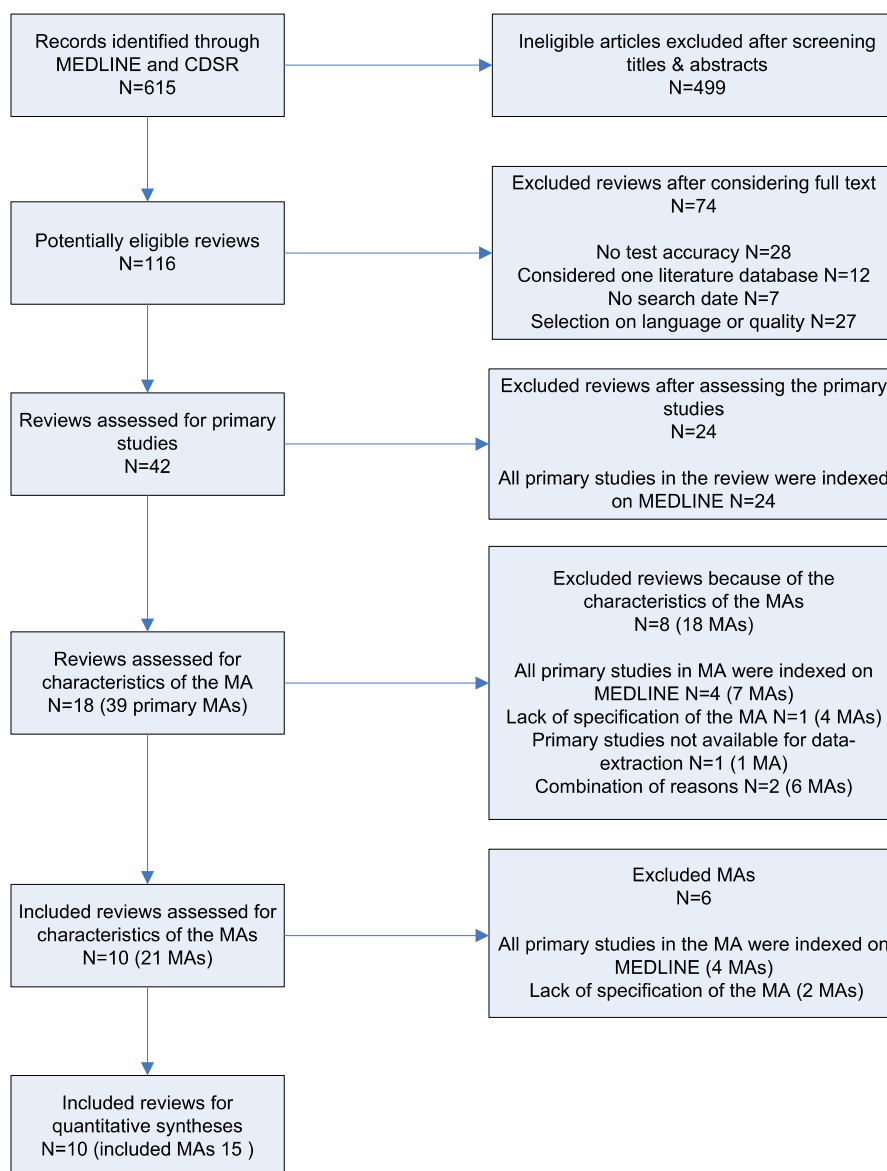
Fig. 1. Flowchart of selection process of systematic reviews and their primary meta-analyses. MA, meta-analysis.

the basic characteristics of the included reviews are presented in Table 1.

The mean percentage of NiM studies in the included meta-analyses was 16% (range, 1.0–33%). The RDORs comparing the DOR for iM studies with the DOR of all studies ranged from 0.77 to 1.23 with a pooled RDOR of 1.04 (95% CI, 0.94, 1.15) suggesting that restricting searches to MEDLINE may slightly overestimate the results compared with searching a broader range of databases (Fig. 2). However, the point estimate is very close to 1 and not significant. None of the individual meta-analyses were significantly affected by leaving out NiM studies. There were no statistically significant differences between the sensitivity and specificity of iM studies and all studies (difference in sensitivity: −0.08%; 95% CI, −1% to 1% range between −5.8% and 4.8% and difference in specificity:

−0.1%; 95% CI, −0.8% to 1% range between − 2.2% and 2.3%; Figs. 3 and 4). Heterogeneity was minimal for all meta-analyses.

The meta-analyses selected as the primary meta-analyses were also those with the largest contrasts in NiM/(iM + NiM) with the exception of one review. For this one review, a greater number of studies would have been missed by a different meta-analysis (NiM 7% primary meta-analyses vs. NiM 15% for other meta-analysis). A sensitivity analysis showed that including this meta-analyses would not have led to different conclusions (RDOR of 1.07; 95% CI, 0.98, 1.17).

There was a nonsignificant association between the comprehensiveness of the search and finding all studies in MEDLINE (odds ratio, 0.82 [95% CI, 0.53, 1.26]). A score below 1 indicates that a higher search score is

Table 1. Characteristics of included systematic reviews ($n = 10$) and meta-analyses ($n = 15$)

Review author (publication year)	Diagnostic test	Target condition	Search score (1–5)	Studies in the review, N (%NiM)	Included MA, N	Studies in MA, N (%NiM)
Cnossen (2008) [25]	All uterine artery Doppler indices	Pre-ecplamsia and intrauterine growth restriction	5	132 (10)	1	17 (18)
Diel (2010) [26]	Interferon- γ release assays	Active tuberculosis	1	124 (0.8)	1	19 (11)[a]
Haase (2009) [27]	Neutrophil gelatinase-associated lipocalin	Acute kidney injury	2	19 (21)	1	19 (21)
Leeflang (2009) [28]	Serum galactomannan ELISA	Invasive aspergillosis	5	42 (7)	1	18 (11)
Medeiros (2008) [29]	Dynamic contrast-enhanced breast magnetic resonance	Breast lesion and cancer	3	69 (3)	1	69 (1)
Mitchell (2008) [30]	One or two simple verbal questions	Depression in cancer settings	3	10 (10)	3	9 (11) 5 (20) 3 (33)
Musso (2010) [31]	ELISA (cytokeratin-18), NAFLD fibrosis score and fibroscan	NAFLD	1	32 (25)	3	11 (27) 9 (22) 6 (17)
Stein (2009) [32]	Pelvic ultrasonography	Ectopic pregnancy	2	10 (10)	1	10 (10)
Wang (2011) [33]	Myocardial perfusion scintigraphy and dobutamine stress echocardiography	Coronary artery disease	5	17 (6)	1	7 (14)
Whiting (2010) [34]	Anti—citrullinated peptide antibodies	Rheumatoid arthritis	3	151 (17)	2	15 (7) 138 (15)[b]
Total	—	—	—	482 (11)	15	355 (16)

Abbreviations: ELISA, enzyme-linked immunosorbent assay; MA, meta-analysis; NAFLD, nonalcoholic fatty liver disease; NiM, not in MEDLINE.
[a] Data limited to diseased participants.
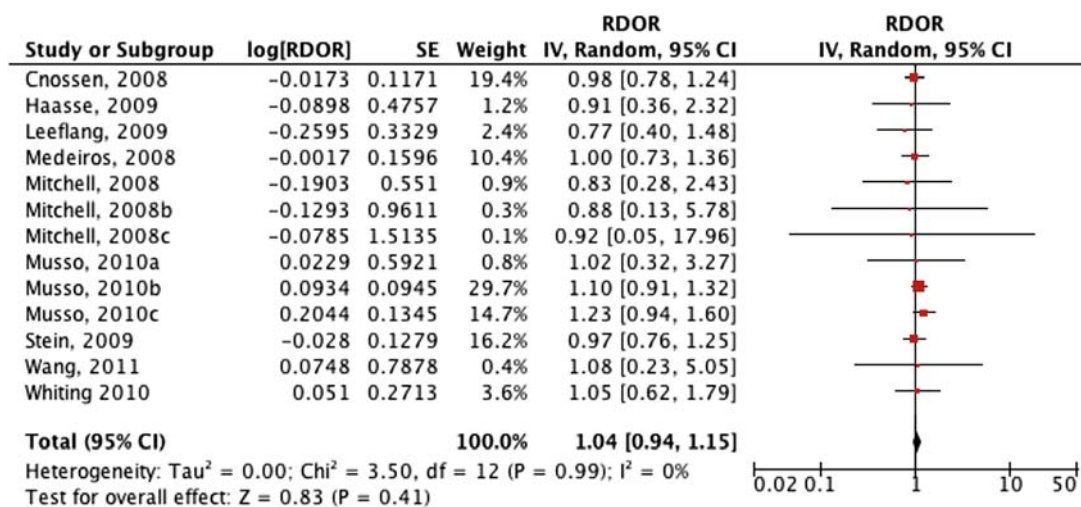[b] Meta-analysis only included to study the robustness of the results in a sensitivity analysis.

associated with reducing the odds of finding all studies in MEDLINE.
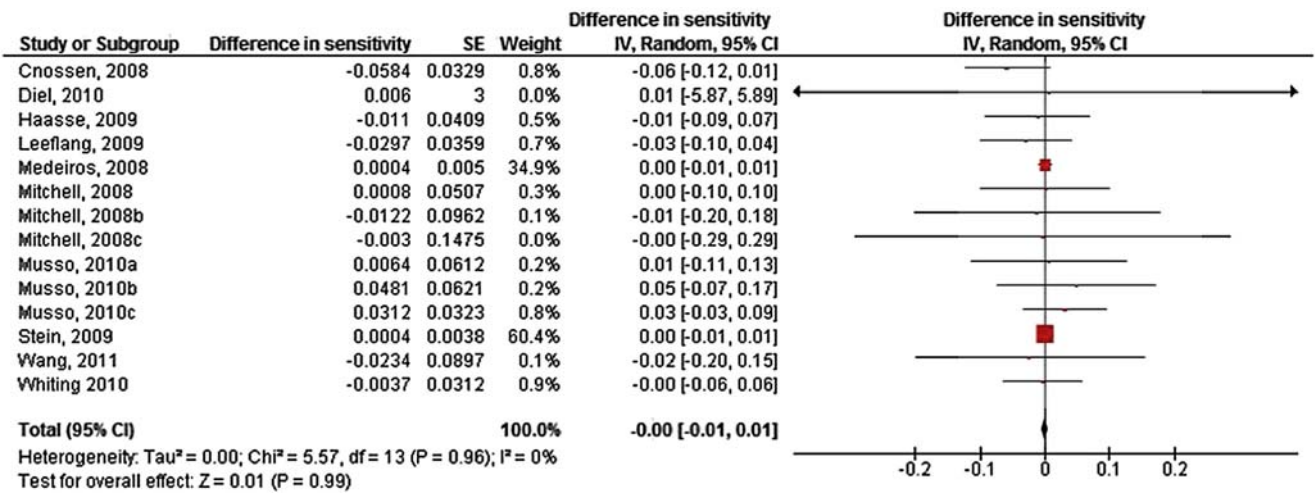
4. Discussion

Our results suggest that restricting searches to MEDLINE would only have had a negligible nonsignificant impact on the summary estimates of the meta-analyses included in our study. In addition, for most reviews (57%) assessed in our metaepidemiologic study, all the included studies were indexed on MEDLINE, although the review authors had searched at least one other database. These reviews would not have been impacted at all had the search been restricted to MEDLINE.

Our results differ from evidence found for intervention reviews. For intervention reviews, it has been shown that studies that are uniquely indexed on EMBASE on average yield 27% lower effect estimates than studies from other databases [9]. These, and similar results from other studies [17,18], are the founding to score reviews that have searched in only one electronic database as low quality [13]. Our study indicates that the summary estimates of the meta-analyses in DTA reviews that searched only MEDLINE

Study or Subgroup	log[RDOR]	SE	Weight	RDOR IV, Random, 95% CI
Cnossen, 2008	-0.0173	0.1171	19.4%	0.98 [0.78, 1.24]
Haasse, 2009	-0.0898	0.4757	1.2%	0.91 [0.36, 2.32]
Leeflang, 2009	-0.2595	0.3329	2.4%	0.77 [0.40, 1.48]
Medeiros, 2008	-0.0017	0.1596	10.4%	1.00 [0.73, 1.36]
Mitchell, 2008	-0.1903	0.551	0.9%	0.83 [0.28, 2.43]
Mitchell, 2008b	-0.1293	0.9611	0.3%	0.88 [0.13, 5.78]
Mitchell, 2008c	-0.0785	1.5135	0.1%	0.92 [0.05, 17.96]
Musso, 2010a	0.0229	0.5921	0.8%	1.02 [0.32, 3.27]
Musso, 2010b	0.0934	0.0945	29.7%	1.10 [0.91, 1.32]
Musso, 2010c	0.2044	0.1345	14.7%	1.23 [0.94, 1.60]
Stein, 2009	-0.028	0.1279	16.2%	0.97 [0.76, 1.25]
Wang, 2011	0.0748	0.7878	0.4%	1.08 [0.23, 5.05]
Whiting 2010	0.051	0.2713	3.6%	1.05 [0.62, 1.79]
Total (95% CI)			100.0%	1.04 [0.94, 1.15]

Heterogeneity: Tau² = 0.00; Chi² = 3.50, df = 12 (P = 0.99); I² = 0%
Test for overall effect: Z = 0.83 (P = 0.41)

Fig. 2. Forest plot for the relative diagnostic odds ratio indicating the difference between including only MEDLINE studies in the primary meta-analysis vs. including all studies. (For some reviews, multiple meta-analyses were included as they considered different tests).

Study or Subgroup	Difference in sensitivity	SE	Weight	Difference in sensitivity IV, Random, 95% CI
Cnossen, 2008	-0.0584	0.0329	0.8%	-0.06 [-0.12, 0.01]
Diel, 2010	0.006	3	0.0%	0.01 [-5.87, 5.89]
Haasse, 2009	-0.011	0.0409	0.5%	-0.01 [-0.09, 0.07]
Leeflang, 2009	-0.0297	0.0359	0.7%	-0.03 [-0.10, 0.04]
Medeiros, 2008	0.0004	0.005	34.9%	0.00 [-0.01, 0.01]
Mitchell, 2008	0.0008	0.0507	0.3%	0.00 [-0.10, 0.10]
Mitchell, 2008b	-0.0122	0.0962	0.1%	-0.01 [-0.20, 0.18]
Mitchell, 2008c	-0.003	0.1475	0.0%	-0.00 [-0.29, 0.29]
Musso, 2010a	0.0064	0.0612	0.2%	0.01 [-0.11, 0.13]
Musso, 2010b	0.0481	0.0621	0.2%	0.05 [-0.07, 0.17]
Musso, 2010c	0.0312	0.0323	0.8%	0.03 [-0.03, 0.09]
Stein, 2009	0.0004	0.0038	60.4%	0.00 [-0.01, 0.01]
Wang, 2011	-0.0234	0.0897	0.1%	-0.02 [-0.20, 0.15]
Whiting 2010	-0.0037	0.0312	0.9%	-0.00 [-0.06, 0.06]
Total (95% CI)			**100.0%**	**-0.00 [-0.01, 0.01]**

Heterogeneity: Tau² = 0.00; Chi² = 5.57, df = 13 (P = 0.96); I² = 0%
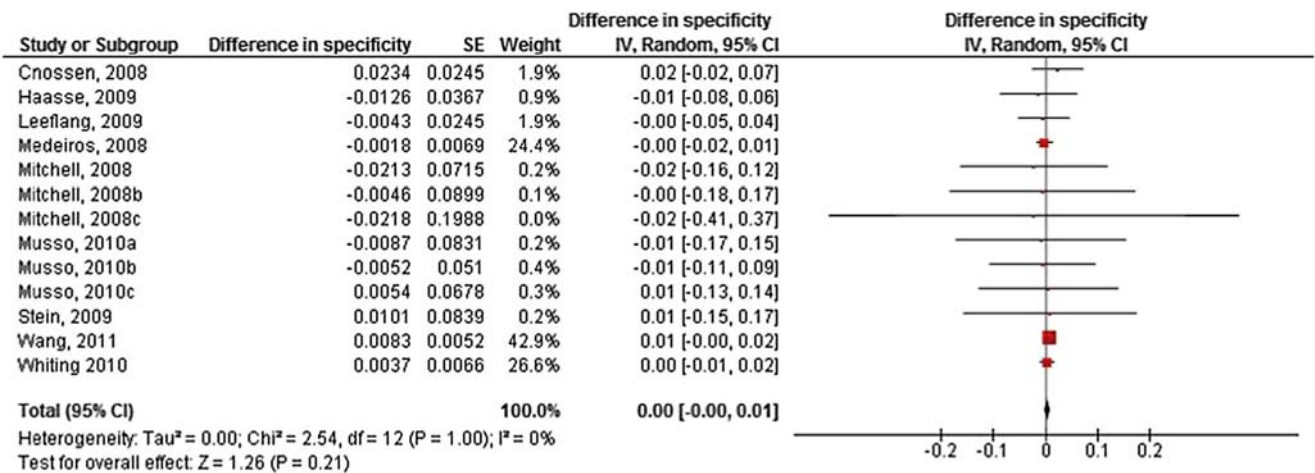Test for overall effect: Z = 0.01 (P = 0.99)

Fig. 3. Forest plot for the relative sensitivity indicating the difference between including only MEDLINE studies in the primary meta-analysis vs. including all studies. (For some reviews, multiple meta-analyses were included as they considered different tests).

are not significantly affected. Consequently, DTA reviews may therefore not automatically be dismissed as "low quality" if they have searched only one electronic database.

Although our results indicate that limiting the search to MEDLINE studies did not alter the summary estimates of the reviews, there are other consequences of a more restrictive search that should be considered. Missing relevant studies that could potentially have been included in the review will decrease the power of the meta-analysis and make the CIs around summary measures of accuracy wider. This will potentially lower the confidence in the results of the meta-analysis. Heterogeneity is a feature of almost all DTA reviews [19]. Investigation of heterogeneity is important to determine the most reliable estimate of accuracy. Heterogeneity can be investigated using metaregression or subgroup analysis, but this requires sufficient power. Missing studies will therefore also decrease the potential to investigate heterogeneity. Second, different databases

use different indexing systems and almost no search strategy has perfect sensitivity; even strategies designed to be very sensitive miss relevant studies [14]. Therefore, searching in multiple databases may increase the likelihood of identifying studies also available in MEDLINE.

This study has some limitations. The number of meta-analyses that could be included in our analysis was small, also because most meta-analyses had no NiM studies and could therefore not add information to our analysis. A second reason was poor reporting of studies, proving too few details to include the study for analysis. Poor reporting of DTA studies is a common problem in DTA studies [20]. To prevent from this, we selected reviews with an impact factor of ≥4 that have been shown to have higher quality of reporting. Still, this seemed to be insufficient. Because of the small number of included meta-analyses, we had low power for assessing the effect of limiting the meta-analysis to iM studies. Moreover, the fact that we did not

Study or Subgroup	Difference in specificity	SE	Weight	Difference in specificity IV, Random, 95% CI
Cnossen, 2008	0.0234	0.0245	1.9%	0.02 [-0.02, 0.07]
Haasse, 2009	-0.0126	0.0367	0.9%	-0.01 [-0.08, 0.06]
Leeflang, 2009	-0.0043	0.0245	1.9%	-0.00 [-0.05, 0.04]
Medeiros, 2008	-0.0018	0.0069	24.4%	-0.00 [-0.02, 0.01]
Mitchell, 2008	-0.0213	0.0715	0.2%	-0.02 [-0.16, 0.12]
Mitchell, 2008b	-0.0046	0.0899	0.1%	-0.00 [-0.18, 0.17]
Mitchell, 2008c	-0.0218	0.1988	0.0%	-0.02 [-0.41, 0.37]
Musso, 2010a	-0.0087	0.0831	0.2%	-0.01 [-0.17, 0.15]
Musso, 2010b	-0.0052	0.051	0.4%	-0.01 [-0.11, 0.09]
Musso, 2010c	0.0054	0.0678	0.3%	0.01 [-0.13, 0.14]
Stein, 2009	0.0101	0.0839	0.2%	0.01 [-0.15, 0.17]
Wang, 2011	0.0083	0.0052	42.9%	0.01 [-0.00, 0.02]
Whiting 2010	0.0037	0.0066	26.6%	0.00 [-0.01, 0.02]
Total (95% CI)			**100.0%**	**0.00 [-0.00, 0.01]**

Heterogeneity: Tau² = 0.00; Chi² = 2.54, df = 12 (P = 1.00); I² = 0%
Test for overall effect: Z = 1.26 (P = 0.21)

Fig. 4. Forest plot for the relative specificity indicating the difference between including only MEDLINE studies in the primary meta-analysis vs. including all studies. (For some reviews, multiple meta-analyses were included as they considered different tests).

identify a significant difference for the summary estimates does not exclude that no difference does exist. In addition, if we had been able to estimate the difference between the various diagnostic parameters assessed in the NiM and iM separately, the contrast between these two subgroups might have been bigger, but unfortunately the number of NiM studies was too low to perform a separate meta-analysis upon. Furthermore, the low power of our analysis and poor reporting of reviews also limited our ability to investigate the relationship between the comprehensiveness of the search and finding studies outside of MEDLINE. It would have been more interesting to assess the specific features of the searches individually to provide insight into which strategies are most likely to identify studies beyond MEDLINE, but the analysis did not have sufficient power to do so. It would been preferable to have used a validated checklist for this purpose like the Canadian Agency for Drugs and Technologies in Health peer-review checklist which is used by information specialists to peer review search strategies [21]. However, these tools require access to the full strategy, but these are very rarely reported in a review.

A known-item search was used to assess whether studies were indexed on MEDLINE. This type of search distinguishes itself from searches used in systematic reviews. As the name indicates, the study that is searched for is known in a known-item search, whereas for a systematic review, explorative searches are used. Explorative searches depend on the indexing system of the databases. Therefore it is possible that studies indexed on MEDLINE had not been found by the review authors when using an explorative search, although we have identified it using the known-item search [2]. Explorative searches in MEDLINE are likely to miss studies because of inadequate assignment of MeSH terms [17]. Performing the known-item search for this study therefore may have resulted in an overestimation of the number of iM studies. Access to the full search strategies would have enabled us to replicate the original searches and to determine which studies would be identified by searching MEDLINE, but these full search strategies are rarely reported. Because of this limitation, we might have overestimated the number of studies a reviewer will identify on MEDLINE but did not affect our result that accuracy results of iM studies are not significantly different than accuracy results in studies uniquely indexed on other electronic databases.

Despite the fact that the results of our study are based on a small sample of meta-analyses, our confidence in the results is strengthen by the fact that no individual-included review had significant differences between the DOR based on all studies compared with the DOR based on the studies that were iM studies only. In addition, when the meta-analyses with the largest proportion of NiM studies were included, that is, the analyses with the greatest number of missing studies, still no significant difference was found. Another strength of our study is that the inclusion of reviews was restricted to those that had no inclusion limitations on language or quality. A selective inclusion of

primary studies for the reviews could otherwise have interfered with our results. Additionally, we investigated the association between the search characteristics of the included reviews and finding all studies in MEDLINE; no association was found. A further strength is that the analyses to estimate the impact of leaving out NiM studies were controlled for the dependency that exists for every meta-analysis with all studies included and the meta-analysis including the subset of MEDLINE studies.

Although this study has been undertaken to answer challenges with respect to searching for primary studies, it also gives insights into the underlying topic of selective reporting. It is well known that publication and dissemination bias can influence the results of a meta-analysis [22]. However, evidence for its existence and manifestation in the field of diagnostic test accuracy is scarce [23,24]. Our results did not identity a systematic difference in accuracy measures between studies indexed on MEDLINE and other database–indexed studies, but this does not exclude that there is a systematic difference between MEDLINE and other databases. It would be worthwhile to investigate if our results also apply for specific fields of medical test accuracy research as the results of this study may differ between fields.

5. Conclusion

Restricting the search to MEDLINE did not have a significant impact on the summary estimates of the reviews included in our study. Missing relevant studies, however, may lower the precision of the summary measures of accuracy and the power of the analyses to investigate heterogeneity. When financial resources are low, restricting searches to MEDLINE does not seem to affect the summary estimates of the review. However, the impact for individual reviews is still unpredictable.

statistical analysis, contributed to the data analysis, and wrote the statistical methods for the manuscript. L.H. contributed to the development of the protocol and data extraction and commented on the manuscript.

Appendix

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jclinepi.2014.05.008.

References

[1] de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: searching for studies. In: Cochrane handbook for systematic reviews of diagnostic test Accuracy. Version 0.4 [updated September 2008] ed. The Cochrane Collaboration; 2008.

[2] Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol 2008;61:357−64.

[3] Indiana University. University Library; Ulrichsweb Global Serials Directory. 2013. Available at http://www.ulib.iupui.edu/. Accessed October 22, 2013.

[4] Lefebvre C, Manheimer E, Glanville J. Chapter 6. Searching for studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. 5.1.0 ed. Wiley; 2008.

[5] Bachmann LM, Coray R, Estermann P, Ter RG. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc 2002;9:653−8.

[6] Vincent S, Greenley S, Beaven O. Clinical evidence diagnosis: developing a sensitive search strategy to retrieve diagnostic studies on deep vein thrombosis: a pragmatic approach. Health Info Libr J 2003;20:150−9.

[7] Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane Database Syst Rev 2013;(9):MR000022.

[8] Leeflang MMG, Scholten RJPM, Rutjes AWS, Reitsma JB, Bossuyt PMM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 2006;59:234−40.

[9] Sampson M, Platt R, StJohn PD, Moher D, Klassen TP, Pham B, et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 2003;56:943−55.

[10] Deville WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. J Clin Epidemiol 2000;53:65−9.

[11] Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Reitsma JB, Bossuyt PM, et al. Quality of reporting of diagnostic accuracy studies. Radiology 2005;235:347−53.

[12] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.

[13] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;15:10.

[14] Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. J Clin Epidemiol 2011;64:602−7.

[15] Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982−90.

[16] Stata Statistical Software: Release 10 [computer program]. College Station, TX, USA. StataCorp LP; 2007.

[17] Suarez-Almazor ME, Belseck EF, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Control Clin Trials 2000;21:476−87.

[18] Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. Eval Health Prof 2002;25(1):38−64.

[19] Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess 2005;9:1−113. iii.

[20] Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med 2014;19:47−54.

[21] Canadian Agency for Drugs and Technologies in Health. CADTH peer review checklist for search strategies. Ottawa, Canada: The Agency; 2013. Available at http://www.cadth.ca/en/resources/finding-evidence-is/peer-review-search-strat. Accessed October 18, 2013.

[22] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010;14:iii. ix-iii,193.

[23] DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. JAMA 2004;292:1363−4.

[24] Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002;31:88−95.

[25] Cnossen JS, Morris RK, ter RG, Mol BW, van der Post JA, Coomarasamy A, et al. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariate meta-analysis. CMAJ 2008;178(6):701−11.

[26] Diel R, Loddenkemper R, Nienhaus A. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. Chest 2010;137:952−68.

[27] Haase M, Bellomo R, Devarajan P, Schlattmann P, Haase-Fielitz A. Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis 2009;54(6):1012−24.

[28] Leeflang MM, Debets-Ossenkopp YJ, Visser CE, Scholten RJ, Hooft L, Bijlmer HA, et al. Galactomannan detection for invasive aspergillosis in immunocompromised patients. Cochrane Database Syst Rev 2008;(4):CD007394.

[29] Medeiros LR, Duarte CS, Rosa DD, Edelweiss MI, Edelweiss M, Silva FR, et al. Accuracy of magnetic resonance in suspicious breast lesions: a systematic quantitative review and meta-analysis. Breast Cancer Res Treat 2011;126:273−85.

[30] Mitchell AJ. Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. Br J Cancer 2008;98:1934−43.

[31] Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. Ann Med 2011;43:617−49.

[32] Stein SC, Fabbri A, Servadei F, Glick HA. A critical comparison of clinical decision instruments for computed tomographic scanning in mild closed traumatic brain injury in adolescents and adults. Ann Emerg Med 2009;53(2):180−8.

[33] Wang LW, Fahim MA, Hayen A, Mitchell RL, Lord SW, Baines LA, et al. Cardiac testing for coronary artery disease in potential kidney transplant recipients: a systematic review of test accuracy studies. Am J Kidney Dis 2011;57(3):476−87.

[34] Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M, et al. Systematic review: accuracy of anti-citrullinated peptide antibodies for diagnosing rheumatoid arthritis. Ann Intern Med 2010;152:456−64.