

ORIGINAL RESEARCH

Evidence for differences in patterns of temporal trends in meta-analyses of diagnostic accuracy studies in the Cochrane database of systematic reviews

Jacqueline Murphy*, Thomas R. Fanshawe

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, Oxfordshire, UK

Accepted 15 July 2024; Published online 22 July 2024

Abstract

Objectives: Temporal trends in comparative meta-analyses of interventions are well-recognized in the medical literature. For studies of diagnostic test accuracy (DTA), evidence of temporal trends is growing and the importance of assessing and reporting them has been highlighted in recent guidelines on postmarket surveillance in several jurisdictions. In this study, we evaluate the prevalence and patterns of time trends using a larger and more up-to-date set of DTA systematic reviews than has previously been examined, from the Cochrane Database of Systematic Reviews.

Study Design and Setting: Cumulative meta-analysis was conducted on bivariate random effects meta-analysis estimates of sensitivity and specificity, after ranking studies by publication date. Trends for all studies were assessed graphically using plots of summary estimates by study rank, and using receiver operating characteristic plots of sensitivity vs specificity. Linear trends were also described using weighted linear regression with autocorrelated errors of summary estimates against study rank. Various patterns of nonlinear trends were characterized descriptively.

Results: The analysis included 46 reviews (92 meta-analyses) conducted between 2017 and 2022. The total number of studies within all reviews was 1486, with a median (IQR) 7134 (2782–16,406) participants per review. Reviews had a median (IQR) time span of 19 (15–25) publication years. Time trends in at least 1 DTA measure were observed in 40 (87%) reviews, and statistically significant linear trends in 32 (70%) reviews. Nonlinear time trends were observed in 14 (30%) reviews. There was no evidence for a trend in either DTA measure in 6 (13%) reviews.

Conclusion: The study contributes evidence on the variety in patterns of linear and nonlinear temporal DTA trends which has not previously been described. We recommended researchers check statistical assumptions of trend analysis methods, eg, using graphical methods. Further research into potential reasons for time trends could contribute to the robustness of future meta-analyses. © 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Diagnostic accuracy; Meta-analysis; Cumulative meta-analysis; Time trend analysis; Temporal trends; Generalised least squares

1. Introduction

Time trends in estimates of effects from meta-analyses are well-recognized in the scientific literature [1–3]. It has also been found that time trends occur in diagnostic test accuracy

(DTA) studies; however, evidence is currently limited to single study settings [4], single meta-analyses [5], or groups of systematic reviews from a limited range of publication years [6,7]. Recent guidance on postmarket surveillances in several jurisdictions, including the European Union [8], United States [9], and United Kingdom [10], incorporate recommendations to monitor changes in the performance of diagnostic devices over time postapproval. Therefore, there is an urgent need to evaluate the extent to which time trends are present in meta-analyses of diagnostic studies across a range of settings, and to identify the methods that are the most suitable for detecting these trends. In this study, we aim to add to the limited evidence base using a larger and more up-to-date source of DTA systematic reviews than has previously been examined.

This study was funded by the National Institute for Health and Care Research (NIHR) Community Healthcare MedTech and In Vitro Diagnostics Co-operative at Oxford Health NHS Foundation Trust (MIC-2016-018). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

* Corresponding author. Nuffield Department of Primary Care Health Sciences, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG, UK.

E-mail address: jacqueline.murphy@phc.ox.ac.uk (J. Murphy).

<https://doi.org/10.1016/j.jclinepi.2024.111472>

0895-4356/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

What is new?

Key findings

- We found evidence for temporal trends in the majority of primary meta-analyses of diagnostic accuracy studies from recent systematic reviews in the Cochrane Library. We also identified nonlinear trends in a substantial proportion of diagnostic accuracy meta-analyses and characterized the different patterns of trend in these studies.

What this adds to what is known

- The findings provide up-to-date evidence on the existence of time trends in a large sample of diagnostic accuracy studies and illustrate cases where existing methods of trend analysis, which assume linearity, may not be appropriate.

What is the implication and what should change now?

- We recommend that researchers performing diagnostic accuracy meta-analysis should examine temporal effects when assessing reasons for between-study heterogeneity, and should allow for possible nonlinear trends.

The Cochrane Database of Systematic Reviews (CDSR) holds records of systematic reviews conducted by Cochrane Review Groups in a range of clinical areas [11]. Cochrane Systematic Reviews are conducted and reported according to a single set of guidelines [12] and are updated periodically as new evidence becomes available [13]. This database therefore provides a consistent resource of high methodological quality to inform an evaluation of the performance of diagnostic tests over time.

Building on previous evidence for time trends in comparative meta-analyses of interventions in the CDSR [14], the objectives of this study were to evaluate the prevalence of time trends in meta-analyses of studies reported in DTA systematic reviews registered with the CDSR, and to assess the extent to which the assumptions of existing methods are appropriate for evaluation of such trends.

2. Materials and methods

2.1. Identification and selection of reviews

Data for the analysis were taken from the most recent update of all DTA systematic reviews published in the CDSR between August 20, 2016, and August 20, 2023. We considered the primary meta-analysis from each review, which was identified as either the primary meta-analysis as specified by the authors in the original articles, or (where

none was indicated by the authors) the largest subset of studies for any meta-analysis reported in the original review by index diagnostic test, threshold, or subgroup analysis. For index tests consisting of a continuous biomarker measurement, we only considered meta-analyses performed at each single threshold value of this biomarker. Screening of reviews was performed through full-text review and inspection of the accompanying data files from the CDSR website. Reviews were excluded based on the following criteria relating to the primary meta-analysis.

1. No meta-analysis performed by the original review authors (due to, e.g., too many differing test thresholds and/or substantial heterogeneity between studies);
2. Too few studies included the review (at least 5 studies were required to perform the meta-analyses);
3. Fewer than 5 publication years after pooling the first 5 studies of the primary meta-analysis (required for the regression analysis);

Within each review, we included only studies for which both sensitivity and specificity could be estimated. Further details about identification and screening of reviews, preparation of data files, and details of included and excluded reviews are provided in the Supplementary Information.

2.2. Statistical methods

2.2.1. Cumulative meta-analysis of diagnostic accuracy

Within each review, bivariate random effects meta-analysis was conducted on sensitivity and specificity estimates taking into account correlation between the DTA measures using generalized linear mixed models, with binomial likelihood for incorporating within-study variability [15–19]. Cumulative meta-analysis was conducted after ranking studies by publication date. The first meta-analysis for each review was performed on all studies published within 5 years of the first included study to ensure enough data for the bivariate meta-analysis. Subsequent meta-analyses were performed with the addition of each further study in chronological order. The year of publication for each of the included studies within each primary meta-analyses was available from the CDSR data files, but as full dates of publication were not provided, when conducting the cumulative meta-analyses we randomized the order of studies within each publication year. In 1 review, obtaining a specificity of less than 100% was not possible by definition, and for this review we analyzed sensitivity using univariate random effects models as in the original publication (CD013021 [15,17,20]).

2.2.2. Characterizing time trends in diagnostic accuracy estimates

Based on previously published methods [6,21], we estimated time trends among cumulative bivariate meta-analyses of sensitivity (specificity) within each review by

fitting a linear regression to the logit of sensitivity (specificity) against the number of studies in each meta-analysis in chronological order (ie, study rank), weighted by the inverse variance of each meta-analysis, with autocorrelated errors of order 1. The method takes into account correlation between sequential estimates of DTA arising from the overlap in the studies included in each sequential meta-analysis, and therefore nonindependence of the summary estimates [21].

The appropriateness of the linearity assumption was assessed by visual inspection of the cumulative meta-analysis plots with fitted regression lines. For meta-analyses where the linearity assumption appeared appropriate (ie, a positive or negative linear trend, or no trend) the slope of the trend was quantified using the odds ratio of logit sensitivity (specificity), with confidence intervals estimated using the standard error of the slope coefficient and the normal approximation. Statistical significance in the regression slope was interpreted as indicating the presence of a linear trend in these studies. Time trends where the linearity assumption was not appropriate based on visual inspection were instead plotted graphically without fitted regression lines. As we had not prespecified the shape that a nonlinear trend might follow, for these studies the shape of the nonlinear trend was characterized descriptively and the estimate of the linear slope is not reported. The number and proportion of studies with a linear trend (positive or negative), nonlinear trend, or no time trend were summarized.

As an alternative method of illustrating trends over time in both sensitivity and specificity, the DTA summary estimates from each review were also presented in the receiver operating characteristic (ROC) space. Sensitivity vs specificity was plotted for each sequential meta-analysis, and a line was added to connect neighboring points to illustrate the change in both DTA estimates as the number of included studies increased.

2.2.3. Sensitivity analysis

We conducted a sensitivity analysis to test whether randomizing the order of studies within each year affected the results. We reanalyzed the data by conducting the cumulative meta-analysis for each review on studies pooled by publication year to remove the randomization of the ordering, and we assessed the impact on the study conclusions.

Statistical analysis was conducted using R software [22] and the “lme4” package for bivariate meta-analysis [23].

3. Results

3.1. Overview of included reviews and primary meta-analyses

A total of 125 reviews were identified from the online search of the CDSR (Fig 1). After 49 exclusions based on

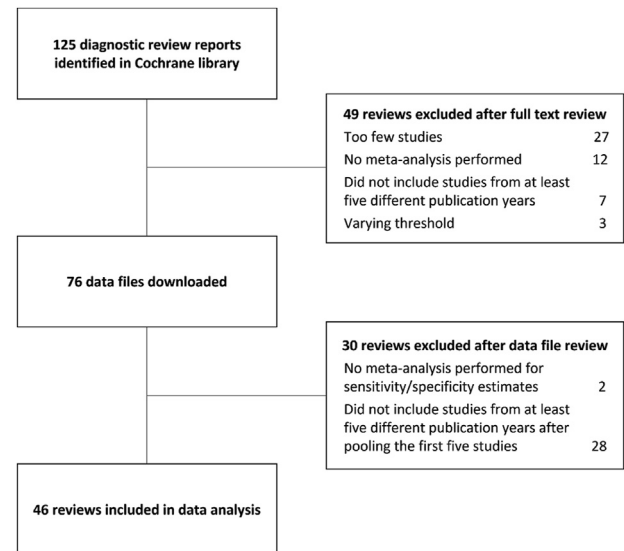


Figure 1. Flowchart for selection of systematic review studies.

full-text review and 30 further exclusions following inspection of the accompanying data files (due to insufficient data for meta-analysis, or too few years to perform analysis of trend), 46 reviews were included in the analysis. Included reviews were published between 2017 and 2022 by 18 different Cochrane Review Groups (Table 1).

Across all primary meta-analyses in the 46 included reviews there were a total of 1486 studies with 3,158,814 study participants (Table 1). The median (IQR) number of studies included in the primary meta-analysis of each review was 24 (17 – 34) from a median (IQR) time span of 19 (15 – 25) publication years from the first included study to the final included study. The median (IQR) total number of participants in the primary meta-analysis of each review (including all studies) was 7134 (2782–16,406). Further details about the included reviews and the selected primary meta-analysis of each review, along with a full list of citations are given in the Supplementary Information.

3.2. Characterizing time trends in diagnostic accuracy estimates

3.2.1. Time trends in diagnostic accuracy

Time trends in meta-analyses for at least one DTA measure (linear or nonlinear trends in sensitivity and/or specificity) were observed in 40 (87%) of 46 reviews (Table 2). By DTA measure, 28 (61%) meta-analyses of sensitivity and 31 (67%) meta-analyses of specificity demonstrated time trends.

3.2.2. Meta-analyses with linear trends in diagnostic accuracy

Statistically significant linear trends in at least 1 DTA measure were observed in 32 (70%) reviews (Table 2). Among meta-analyses of sensitivity, statistically significant

Table 1. Characteristics of included reviews and the primary meta-analysis for each review

Publication year	Number of reviews
2017	8
2018	8
2019	7
2020	4
2021	12
2022	7
Cochrane review group	Number of reviews
Anesthesia	1
Colorectal	3
Dementia and cognitive improvement	3
Developmental, psychosocial, and learning problems	1
Eyes and vision	1
Gut	1
Gynecological, neuro-oncology, and orphan cancer	5
Gynecology and fertility	1
Hepato-biliary	7
Infectious diseases	4
Injuries	1
Kidney and transplant	1
Neonatal	3
Oral health	4
Pregnancy and childbirth	4
Sexually transmitted infections	1
Skin	4
Vascular	1
Characteristics of included primary meta-analyses	n or median (IQR)
Total number of studies in included reviews	1486
Number of studies per review	24 (17-34)
Total number of participants in included reviews	3,158,814
Number of participants per review	7134 (2782-16,406)
Time range of studies per review (years)	19 (15-25)

Data are for studies within each review after exclusion of studies with incalculable sensitivity and specificity.

positive linear trends were observed in 4 (9%) meta-analyses, and negative linear trends in 15 (33%) meta-analyses. For specificity, 9 (20%) meta-analyses had positive linear trends, and 15 (33%) had negative linear trends. 3 (7%) reviews had negative linear trends for both sensitivity and specificity, and none had positive linear trends for both sensitivity and specificity. Illustrative examples

of studies with linear time trends in at least 1 DTA measure are shown in [Figure 2](#) (Reviews CD010276 [24], CD011902 [25], and CD011926 [26]).

3.2.3. Meta-analyses with nonlinear trends in diagnostic accuracy

Nonlinear time trends were observed in 16 (17% of 92) meta-analyses of either DTA measure (27% of the 59 meta-analyses demonstrating a time trend) from 14 (30%) reviews; illustrative examples are given in [Figure 3](#). The first pattern of nonlinearity was characterized by an observable time trend (either positive or negative) in earlier summary estimates, which then reached a plateau with approximately zero slope of change in the summary estimate as the number of studies increased. This pattern was observed in 3 meta-analyses of sensitivity (CD011964 [27], CD013346 [28] shown in [Figure 3](#), CD014546 [29]) and 2 of specificity (CD009551 [30], CD014546 [29]). The second pattern, in 4 meta-analyses of specificity (CD008874 [31], CD008892 [32] shown in [Figure 3](#), CD013186 [33], CD013346 [28]) arose when earlier summary estimates were inconsistent with the observed trends (or absence of trend) in subsequent meta-analyses. Finally, 6 meta-analyses of sensitivity (CD009977 [34], CD012245 [35], CD013021 [20] shown in [Figure 3](#), CD013194 [36] shown in [Figure 3](#), CD013855 [37], CD014545 [38]) and 1 of specificity (CD012080 [39]) demonstrated a third pattern of nonlinearity, where apparent time trends existed but with nonconstant slope. All studies with nonlinear trends are indicated in [Supplementary Table 4](#), and plots are presented in [Supplementary Figure 1](#).

3.2.4. Meta-analyses with no evidence for time trends in diagnostic accuracy

In 27 (59%) of 46 reviews there was no evidence for a trend in at least one DTA measure as the number of studies in the meta-analysis increased ([Table 2](#)). There was no evidence for a trend in either DTA measure in 6 (13%) reviews. An illustrative example of a review with no observed time trends for either sensitivity or specificity is shown in [Figure 2](#) (CD013208 [40]).

3.2.5. Trends in diagnostic accuracy represented in the ROC space

Various trends in sensitivity and specificity were also evident from the ROC plots ([Supplementary Figure 2](#)). Illustrative examples are provided in [Figure 4](#) of reviews with no time trend in either DTA measure (CD011767 [41]), a trend in sensitivity only (CD012245 [35]), and a trend in both sensitivity and specificity (CD011902 [25]). Some studies demonstrated large changes in DTA as the number of studies increased (e.g., CD012245 [35] or CD013855 [37]). However for many reviews the changes in DTA observed in the ROC space were relatively small

Table 2. Summary of observed time trends in diagnostic accuracy cumulative meta-analyses

	<i>n</i> (% of 46 reviews)	Specificity				Total
		Nonlinear trend	No trend	Significant negative linear trend	Significant positive linear trend	
Sensitivity	Nonlinear trend	2 (4)	4 (9)	1 (2)	2 (4)	9 (20)
	No trend	2 (4)	6 (13)	8 (17)	2 (4)	18 (39)
	Significant negative linear trend	3 (7)	4 (9)	3 (7)	5 (11)	15 (33)
	Significant positive linear trend	0 (0)	1 (2)	3 (7)	0 (0)	4 (9)
	Total	7 (15)	15 (33)	15 (33)	9 (20)	46 (100)

despite statistically significant linear trends (e.g., CD008587 [42]). The plots also showed nonlinear trends for several reviews involving a change in direction of either DTA measure over time (e.g., CD008874 [31] or CD012546 [43]).

3.2.6. Sensitivity analysis

Following reanalysis of the cumulative meta-analyses with studies pooled by publication year, 3 additional meta-analyses (2 of sensitivity and 1 of specificity) exhibited statistically significant linear trends where none had been observed in the main analysis (Supplementary Figure 3). Linear trends observed in the unpooled analysis were no longer statistically significant in 2 meta-analyses of sensitivity, and 5 meta-analyses of specificity. The overall

conclusions of the study were largely unchanged; linear or nonlinear time trends were observed in at least 1 DTA measure in 55 (60% of 92) meta-analyses from 36 (78% of 46) reviews.

4. Discussion

We evaluated time trends in cumulative bivariate meta-analyses of sensitivity and specificity in 46 DTA systematic reviews published between 2017 and 2022. We found evidence for time trends in at least 1 DTA measure in the majority (87%) of primary meta-analysis of included reviews, and in almost two-thirds of sensitivity (61%) or specificity (67%) meta-analyses when considered separately.

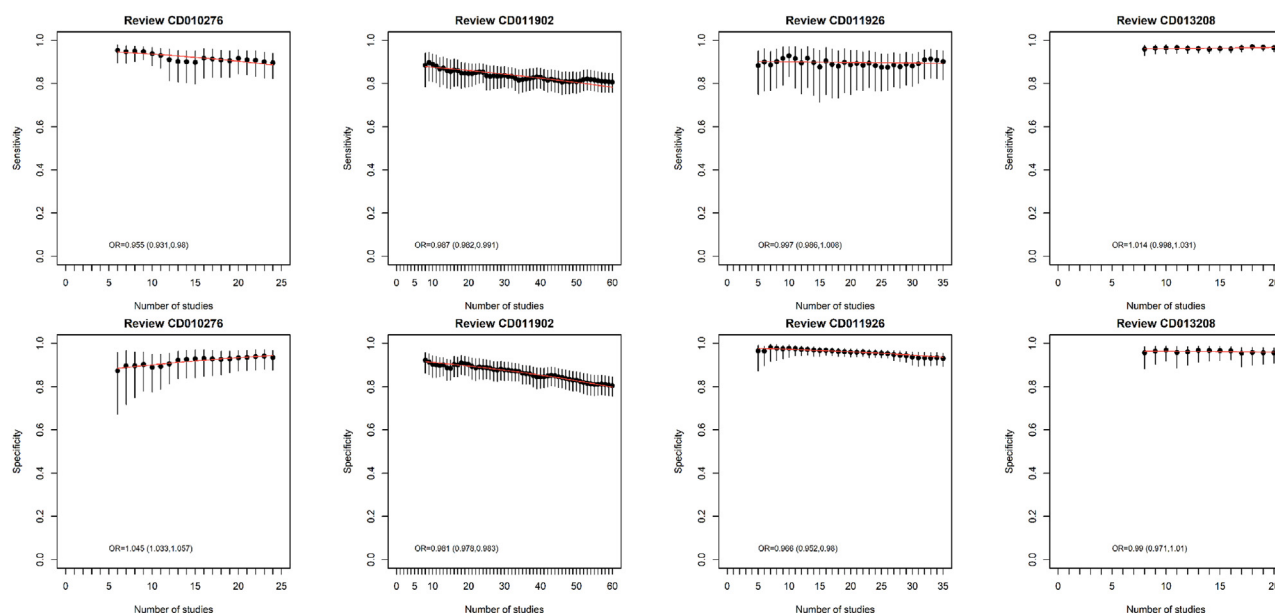


Figure 2. Patterns of linear trends in cumulative meta-analyses of diagnostic accuracy (selected examples). Plots show summary diagnostic accuracy estimates from cumulative meta-analyses estimated using bivariate random effects meta-analysis [15–17] against number of included studies ranked by publication date. Slope of time trend was estimated using linear regression of the logit of sensitivity (or specificity) on the number of studies included in each meta-analyses, weighted by the inverse variance of each meta-analysis, assuming autocorrelated errors of order 1 [6, 21]. ORs reflect the increase in logit sensitivity (or specificity) per 1 additional study. Illustrative results are shown from four selected reviews: CD010276 [24]: statistically significant opposing linear trends in sensitivity and specificity; CD011902 [25]: statistically significant negative linear trends in sensitivity and specificity; CD011926 [26]: statistically significant (negative) linear trend in specificity only; CD013208 [40]: no statistically significant linear trend in either sensitivity or specificity.

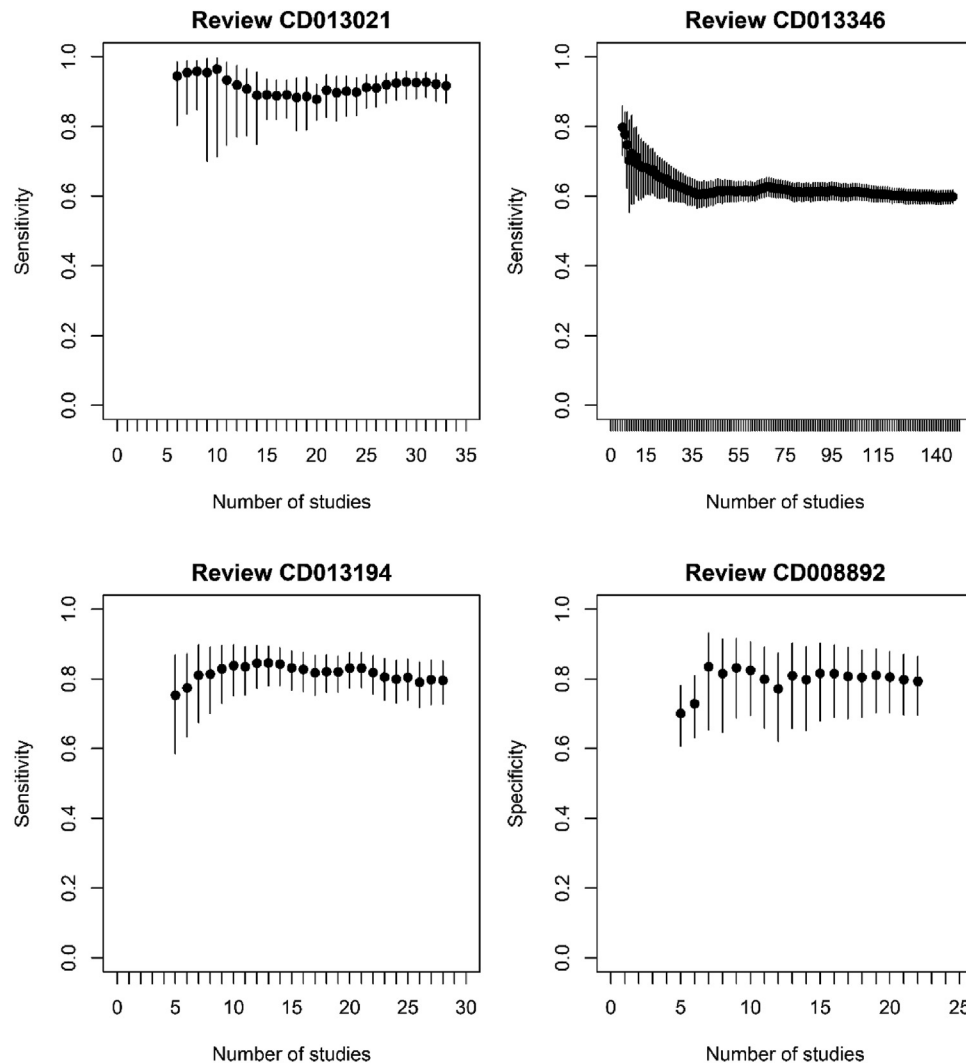


Figure 3. Trends in cumulative meta-analyses of diagnostic accuracy poorly characterized by linearity assumption (selected examples). Plots show summary diagnostic accuracy estimates from cumulative meta-analyses estimated using bivariate random effects meta-analysis [15–17] against number of included studies ranked by publication date. Illustrative results are shown from 4 selected reviews: CD013021 [20] and CD013194 [36]: nonlinear trend in sensitivity with nonconstant slope; CD013346 [28]: nonlinear trend in sensitivity with plateau; CD008892 [32]: nonlinear trend in specificity with early estimates inconsistent with later trend.

This study builds on a previous analysis of DTA time trends in 48 studies published between 2011 and 2012 identified via MEDLINE [6], by providing up-to-date evidence from more recent systematic reviews published between 2017 and 2022. The methods used in previous studies involved separate meta-analyses of sensitivity and specificity [6]. In the present study we incorporated bivariate meta-analysis to take into account correlation between the DTA measures [15–17], which more closely replicated the methods used in the original systematic reviews. We found a greater proportion of reviews exhibited time trends than previously identified [6]. This difference could be partly attributable to the use of a different database (MEDLINE vs CDSR), in addition to the longer time span between the first and last included studies for the meta-analyses in our analysis (median 19 years vs 12.3 years [6]).

Previous research evaluating time trends in DTA [6] employed similar metaregression trend analysis methods which rely on an assumption of linearity [21]. This study provides further evidence for the existence of time trends in DTA meta-analyses, while also identifying that in a substantial proportion of meta-analyses these cannot be appropriately described using a linearity assumption. Other studies have modeled nonlinear time trends, such as in Baker and Jackson [44] where an assumption of effects tending toward the null (“fading” effects) was incorporated into metaregression models using an exponential decaying term. In the present analysis we identified further patterns not consistent with fading effects. In some cases, the trends in DTA reached a plateau after which the summary estimates were unchanged with the addition of further studies into the meta-analysis. Further research into methods of

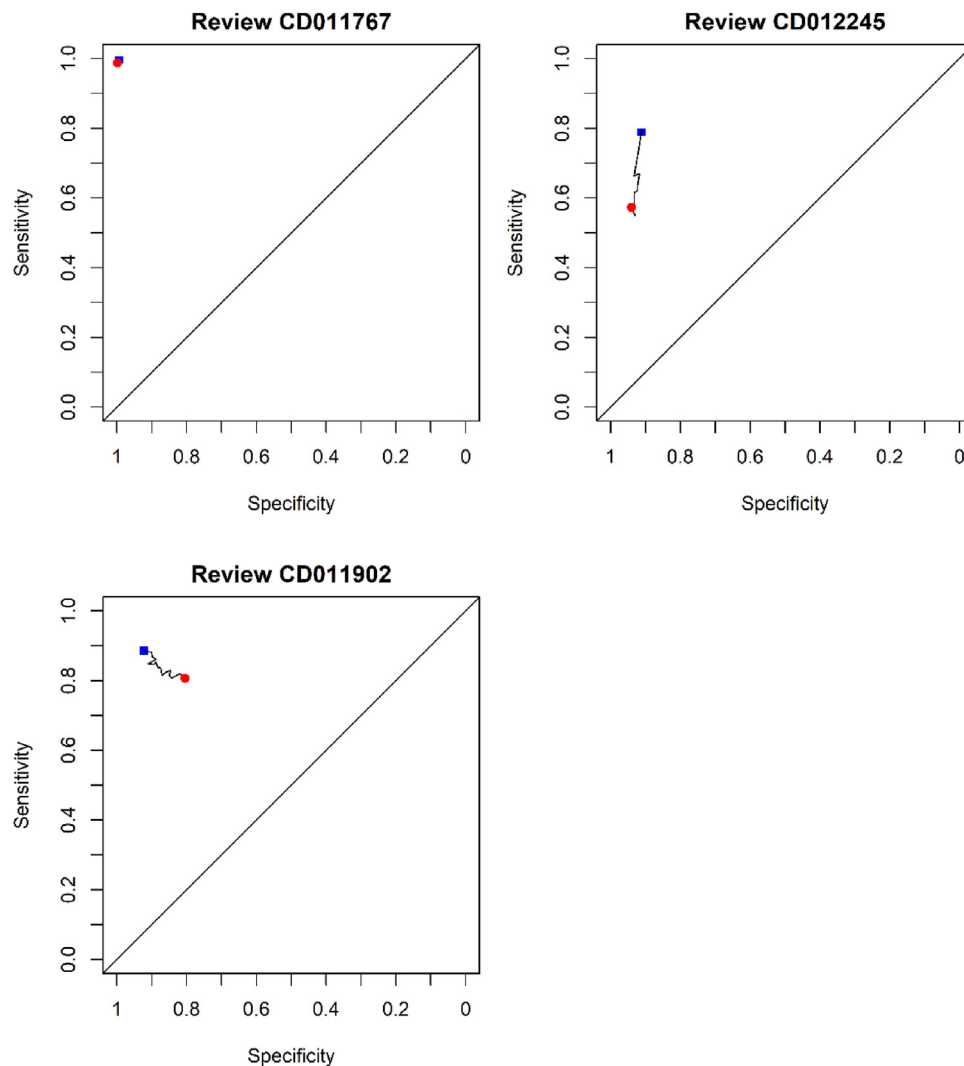


Figure 4. Trends in cumulative meta-analyses of diagnostic accuracy measures in Cochrane systematic reviews, represented in the receiver operating characteristic space (selected examples). Plots show trajectory of summary diagnostic accuracy estimates from cumulative meta-analyses estimated using bivariate random effects meta-analysis [15–17], as number of included studies ranked by publication date increases. Black point indicates first summary estimate (including the first 5 studies); red point indicates the final summary estimate (including all studies). Illustrative results are shown from three selected reviews: CD011767 [41], CD012245 [35], CD011902 [25].

determining when additional studies may contribute little to conclusions of a DTA meta-analysis could be beneficial; eg, as a component of a living systematic review [45].

In large meta-analyses, there is increased power to detect even small temporal effects, and many of the effect sizes observed in this study corresponded to changes in sensitivity or specificity of only a few percentage points. Therefore, in addition to analyzing presence and statistical significance of temporal trends, consideration should be given to the clinical relevance of the magnitude, direction, or pattern of any detected trends in diagnostic accuracy specific to each particular review setting.

There are several possible reasons why an analysis such as this may suggest the existence of a temporal trend in a DTA systematic review. First, there may be a genuine change in performance as the diagnostic test becomes used more widely. Second, an apparent temporal trend may act

as a proxy for another source of heterogeneity; eg, if a test is taken up in different settings (including ‘diagnostic downshift’ [46]) or in patient groups with varying underlying risk. Third, there is the possibility of the “Proteus effect”, where the earliest studies of a diagnostic test or intervention may have substantially different estimates compared to later studies [47], which may be related to an association between earlier time to publication and more optimistic estimates of diagnostic accuracy [48,49]. Finally, it may be the result of statistical artifact relating to the numbers of included participants in early vs late meta-analysis [6,7] or regression to the mean [14].

Our study is based on reports published in the CDSR to ensure we used a source of reviews conducted using a single set of reporting guidelines with high methodological standards [12]. Although this enhances the coherence of the evidence and there is no reason to suspect findings

would differ in non-Cochrane reviews, we cannot guarantee that this would be the case. Our study has some other limitations. First, although we conducted bivariate meta-analysis of the DTA studies, we used existing methods for estimating coefficients of linear trend which analyze sensitivity and specificity separately. To our knowledge, bivariate regression with autocorrelated errors has not yet been implemented for analyzing time trends in diagnostics studies. There is scope for further research to assess the suitability of such methods, particularly when analyzing small numbers of studies such as in early cumulative meta-analysis estimates. Second, it is possible that reasons for time trends relate to differences in the design and conduct of earlier vs later studies within a review. Analysis of study design, quality assessment (such as the bias and applicability ratings [50] conducted as part of Cochrane reviews), or formal analysis of study heterogeneity [51,52] could provide insights into how such factors influence DTA estimates over time. Third, included reviews spanned 18 different Cochrane Review Groups, and therefore the analysis included evidence from a wide range of clinical areas. However, the small number of included reviews from most Cochrane Review Groups and analysis of only the primary meta-analysis from each review resulted in insufficient data for identification of any “typical” patterns of time trends by clinical area.

5. Conclusion

In summary, this study adds to the growing evidence for the existence for the existence of time trends in many DTA systematic reviews by identifying various patterns of linear and nonlinear trends in the majority of the 46 included reviews from the CDSR. We recommend that the authors of systematic reviews consider possible temporal effects when summarizing their findings, and at a minimum explore these graphically, as a means of exploring sources of between-study heterogeneity. When fitting regression models to test for temporal effects, the appropriateness of statistical assumptions such as linearity should be checked, eg, via visual inspection of the cumulative meta-analysis plots.

Future research could include further exploration of the reasons for time trends and patterns of trend, and identification of methods for detecting temporal trends prospectively as part of a living DTA systematic review.

Ethics approval and consent to participate

Not applicable.

CRedit authorship contribution statement

Jacqueline Murphy: Writing — review & editing, Writing — original draft, Visualization, Formal analysis,

Data curation, Methodology. **Thomas R. Fanshawe:** Writing — review & editing, Visualization, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Data availability

The original data used for the analysis are available to download from www.cochranelibrary.com. R code for the analysis is available from <https://github.com/murphyjfe>.

Declaration of competing interest

T.R.F received funding from the National Institute for Health and Care Research Applied Research Collaboration Oxford and Thames Valley at Oxford Health NHS Foundation Trust. There are no competing interests for any other author.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2024.111472>.

References

- [1] Brisco E, Kulinskaya E, Koricheva J. Assessment of temporal instability in the applied ecology and conservation evidence base. *Res Synth Methods* 2024;15:398–412.
- [2] Gehr BT, Weiss C, Porzsolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Med Res Methodol* 2006;6:25.
- [3] Lambert AM, Parretti HM, Pearce E, Price MJ, Riley M, Ryan R, et al. Temporal trends in associations between severe mental illness and risk of cardiovascular disease: a systematic review and meta-analysis. *PLoS Med* 2022;19:e1003960.
- [4] Jansen BHE, Oudshoorn FHK, Tijans AM, Yska MJ, Lont AP, Collette ERP, et al. Local staging with multiparametric MRI in daily clinical practice: diagnostic accuracy and evaluation of a radiologic learning curve. *World J Urol* 2018;36:1409–15.
- [5] Hsiao V, Massoud E, Jensen C, Zhang Y, Hanlon BM, Hitchcock M, et al. Diagnostic accuracy of fine-needle biopsy in the detection of thyroid malignancy: a systematic review and meta-analysis. *JAMA Surg* 2022;157:1105–13.
- [6] Cohen JF, Korevaar DA, Wang J, Leeflang MM, Bossuyt PM. Meta-epidemiologic study showed frequent time trends in summary estimates from meta-analyses of diagnostic accuracy studies. *J Clin Epidemiol* 2016;77:60–7.
- [7] van Enst WA, Naaktgeboren CA, Ochodo EA, de Groot JA, Leeflang MM, Reitsma JB, et al. Small-study effects and time trends in diagnostic test accuracy meta-analyses: a meta-epidemiological study. *Syst Rev* 2015;4:66.
- [8] European Parliament. In vitro Diagnostic Medical Devices Regulation (IVDR). Regulation (EU) 2017/746 of the European parliament and of the council of 5 April 2017 on in vitro diagnostic medical devices and repealing directive 98/79/EC and commission decision 2010/227/EU. Annex XIII: performance evaluation, performance studies and post-market performance follow-up. PART B: post-

- market performance follow-up. 2023. Document 32017R0746. (English). Updated 20/3/2023. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02017R0746-20230320>. Accessed June 12, 2024.
- [9] US Food & Drug Administration. Postmarket surveillance under section 522 of the federal food, drug, and cosmetic act: guidance for industry and food and drug administration staff. 2022. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarket-surveillance-under-section-522-federal-food-drug-and-cosmetic-act>. Accessed June 12, 2024.
- [10] Medicines & Healthcare products Regulatory Agency. Chapter 8 - post-market surveillance, vigilance, market surveillance. 2022. Updated 26 June 2022. Available at: <https://www.gov.uk/government/consultations/consultation-on-the-future-regulation-of-medical-devices-in-the-united-kingdom/outcome/chapter-8-post-market-surveillance-vigilance-market-surveillance>. Accessed March 19, 2024.
- [11] Cochrane library. Review Groups. 2024. Available at: <https://www.cochrane.org/about-us/our-global-community/review-groups>. Accessed February 27, 2024.
- [12] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4. Cochrane; 2023: (updated August 2023). Available at: www.training.cochrane.org/handbook. Accessed September 4, 2024.
- [13] Cochrane Library. About the Cochrane Database of Systematic Reviews. Publication; 2024. Available at: <https://www.cochranelibrary.com/cdsr/about-cdsr>. Accessed February 27, 2024.
- [14] Fanshawe TR, Shaw LF, Spence GT. A large-scale assessment of temporal trends in meta-analyses using systematic review reports from the Cochrane Library. *Res Synth Methods* 2017;8:404–15.
- [15] Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017;26:1896–911.
- [16] Takwoingi YDN, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Supplementary material 1 to Chapter 10: code for undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0. Cochrane; 2023. (updated July 2023) Available at: <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>. Accessed May 8, 2024.
- [17] Takwoingi YDN, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0. Cochrane; 2023. (updated July 2023) Available at: <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>. Accessed May 8, 2024.
- [18] Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331–2. author reply 2-3.
- [19] Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Making* 2010;30:499–508.
- [20] Nagar H, Wietek N, Goodall RJ, Hughes W, Schmidt-Hansen M, Morrison J. Sentinel node biopsy for diagnosis of lymph node involvement in endometrial cancer. *Cochrane Database Syst Rev* 2021;6:CD013021.
- [21] Bagos PG, Nikolopoulos GK. Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology. *J Clin Epidemiol* 2009;62:1037–44.
- [22] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023. Available at: <https://www.R-project.org/>. Accessed September 4, 2024.
- [23] Bates DM, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using {lme4}. *J Stat Softw* 2015;67:1–48.
- [24] Walsh T, Macey R, Kerr AR, Lingen MW, Ogden GR, Warnakulasuriya S. Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. *Cochrane Database Syst Rev* 2021;7:CD010276.
- [25] Dinnes J, Deeks JJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR, et al. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. *Cochrane Database Syst Rev* 2018;12:CD011902.
- [26] Pammi M, Flores A, Versalovic J, Leeflang MM. Molecular assays for the diagnosis of sepsis in neonates. *Cochrane Database Syst Rev* 2017;2:CD011926.
- [27] Davenport C, Rai N, Sharma P, Deeks JJ, Berhane S, Mallett S, et al. Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women. *Cochrane Database Syst Rev* 2022;7:CD011964.
- [28] Colli A, Nadarevic T, Miletic D, Giljaca V, Fraquelli M, Stimac D, et al. Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. *Cochrane Database Syst Rev* 2021;4:CD013346.
- [29] Macey R, Walsh T, Riley P, Glenny AM, Worthington HV, O'Malley L, et al. Visual or visual-tactile examination to detect and inform the diagnosis of enamel caries. *Cochrane Database Syst Rev* 2021;6:CD014546.
- [30] Cruciani M, Mengoli C, Barnes R, Donnelly JP, Loeffler J, Jones BL, et al. Polymerase chain reaction blood tests for the diagnosis of invasive aspergillosis in immunocompromised people. *Cochrane Database Syst Rev* 2019;9:CD009551.
- [31] Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018;5:CD008874.
- [32] Wijedoru L, Mallett S, Parry CM. Rapid diagnostic tests for typhoid and paratyphoid (enteric) fever. *Cochrane Database Syst Rev* 2017;5:CD008892.
- [33] Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, et al. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Database Syst Rev* 2018;12:CD013186.
- [34] Rud B, Vejborg TS, Rapoport ED, Reitsma JB, Wille-Jørgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev* 2019;2019:CD009977.
- [35] Heazell AE, Hayes DJ, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *Cochrane Database Syst Rev* 2019;5:CD012245.
- [36] Dinnes J, Deeks JJ, Grainge MJ, Chuchu N, Ferrante di Ruffano L, Matin RN, et al. Visual inspection for diagnosing cutaneous melanoma in adults. *Cochrane Database Syst Rev* 2018;12:CD013194.
- [37] Macey R, Walsh T, Riley P, Hogan R, Glenny AM, Worthington HV, et al. Transillumination and optical coherence tomography for the detection and diagnosis of enamel caries. *Cochrane Database Syst Rev* 2021;1:CD013855.
- [38] Walsh T, Macey R, Riley P, Glenny AM, Schwendicke F, Worthington HV, et al. Imaging modalities to inform the detection and diagnosis of early caries. *Cochrane Database Syst Rev* 2021;3:CD014545.
- [39] Best LM, Takwoingi Y, Siddique S, Selladurai A, Gandhi A, Low B, et al. Non-invasive diagnostic tests for *Helicobacter pylori* infection. *Cochrane Database Syst Rev* 2018;3:CD012080.
- [40] Ochodo EA, Olwanda EE, Deeks JJ, Mallett S. Point-of-care viral load tests to detect high HIV viral load in people living with HIV/AIDS attending health facilities. *Cochrane Database Syst Rev* 2022;3:CD013208.
- [41] Badeau M, Lindsay C, Blais J, Nshimyumukiza L, Takwoingi Y, Langlois S, et al. Genomics-based non-invasive prenatal testing for detection of fetal chromosomal aneuploidy in pregnant women. *Cochrane Database Syst Rev* 2017;11:CD011767.
- [42] Koliopoulos G, Nyaga VN, Santesso N, Bryant A, Martin-Hirsch PP, Mustafa RA, et al. Cytology versus HPV testing for cervical cancer

- screening in the general population. *Cochrane Database Syst Rev* 2017;8:CD008587.
- [43] Manzotti C, Casazza G, Stimac T, Nikolova D, Gluud C. Total serum bile acids or serum bile acid profile, or both, for the diagnosis of intrahepatic cholestasis of pregnancy. *Cochrane Database Syst Rev* 2019;7:CD012546.
- [44] Baker R, Jackson D. Inference for meta-analysis with a suspected temporal trend. *Biom J* 2010;52:538–51.
- [45] Simmonds M, Elliott JH, Synnot A, Turner T. Living systematic reviews. *Methods Mol Biol* 2022;2345:121–34.
- [46] Sajid IM, Frost K, Paul AK. 'Diagnostic downshift': clinical and system consequences of extrapolating secondary care testing tactics to primary care. *BMJ Evid Based Med* 2022;27:141–8.
- [47] Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; 58:543–9.
- [48] Korevaar DA, van Es N, Zwinderman AH, Cohen JF, Bossuyt PM. Time to publication among completed diagnostic accuracy studies: associated with reported accuracy estimates. *BMC Med Res Methodol* 2016;16:68.
- [49] Treanor LM, Frank RA, Atyani A, Sharifabadi AD, Hallgrimson Z, Fabiano N, et al. Reporting bias in imaging diagnostic test accuracy studies: are studies with positive conclusions or titles submitted and published faster? *AJR Am J Roentgenol* 2021;216:225–32.
- [50] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155: 529–36.
- [51] Leeflang MMG, Kraaijpoel N. Systematic reviews of diagnostic test accuracy in CMI. *Clin Microbiol Infect* 2018;24:1115–6.
- [52] Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002; 21:1525–37.