

Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data

Patel, Amit; Cooper, Nicola; Freeman, Suzanne; Sutton, Alex

DOI:
[10.1002/jrsm.1439](https://doi.org/10.1002/jrsm.1439)

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Patel, A, Cooper, N, Freeman, S & Sutton, A 2021, 'Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data', *Research Synthesis Methods*, vol. 12, no. 1, pp. 34-44. <https://doi.org/10.1002/jrsm.1439>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data

Amit Patel^{1,2} | Nicola Cooper^{2,3}  | Suzanne Freeman^{2,3} | Alex Sutton^{2,3}

¹Cancer Research UK Clinical Trials Unit, Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

²Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK

³NIHR Complex Reviews Support Unit, Universities of Leicester and Glasgow, UK

Correspondence

Alex Sutton, George Davies Centre,
Biostatistics Research Group, Department
of Health Sciences, University of
Leicester, Leicester UK.
Email: ajs22@le.ac.uk

Funding information

Health Technology Assessment
Programme, Grant/Award Number: PDF-
2018-11-ST2-007; National Institute for
Health Research, Grant/Award Number:
14/178/29; Research Trainees
Coordinating Centre, Grant/Award
Number: NIHR-RM-SR-2017-08-023

Diagnostic test accuracy (DTA) systematic reviews are conducted to summarize evidence on the accuracy of a diagnostic test including a critical evaluation of the primary studies. Where appropriate, the evidence is meta-analyzed to obtain pooled estimates of effectiveness.

In this study, we reviewed and critiqued three DTA guidance documents with respect to the graphical presentation of DTA meta-analysis results. All three documents recommended the use of two forms of graphical presentation: (a) forest plots displaying meta-analysis results for sensitivity (ie, the true positive rate) and specificity (ie, true negative rate) separately, and (b) Summary Receiver Operating Characteristic (SROC) curve to provide a global summary of test performance. Two primary shortcomings were identified: (a) lack of incorporation of quality assessment results into the main analysis and; (b) ambiguity with which the contribution of individual studies is represented on SROC curves. In response, two alternative graphical approaches were developed:

1. *A quality assessment enhanced SROC plot* which displays the results from individual studies in the meta-analysis with multiple indicators of quality assessed using QUADAS-2; and
2. *A percentage study weights enhanced SROC plot* which accurately portrays the percentage contribution each study makes to the meta-analysis.

The proposed enhanced SROC curves facilitate the exploration of DTA data, leading to a deeper understanding of the primary studies included in a DTA meta-analysis including identifying reasons for between study heterogeneity and why specific study results may be divergent. Both plots can easily be produced in the free online interactive application, MetaDTA (https://crsu.shinyapps.io/dta_ma/).

KEYWORDS

diagnostic test accuracy, meta-analysis, receiver operating characteristic curves

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

1 | BACKGROUND

Systematic reviews of diagnostic test accuracy (DTA) aim to evaluate evidence from multiple primary studies. Where appropriate, DTA data, from the studies included in the review, are synthesized using meta-analysis and overall measures of sensitivity and specificity reported as a summary of test performance. Ideally, DTA systematic reviews should also investigate the heterogeneity and any risk of bias in the included studies, and put the synthesized evidence into a clinical context to inform readers and help influence healthcare decisions.¹ Meta-analysis of DTA studies is recommended using either bivariate or hierarchical summary receiver operating characteristic (HSROC) models, as they both take into account the correlation between sensitivity and specificity as well as variability in effects between studies.²⁻⁴ The results can be presented as a mean accuracy point or a summary receiver operating characteristic (SROC) curve which plots the true positive rate (ie, sensitivity) against the false positive rate (ie, 1-specificity) and shows how sensitivity and specificity vary for different thresholds of a test.⁵ Plots can also express the uncertainty in the estimation of the mean point or curve via 95% confidence intervals, and heterogeneity between accuracy estimates using 95% prediction intervals. Additionally, “cross-hairs” plots can display the sensitivity and specificity for individual studies in ROC space with associated confidence intervals.⁶

Quality assessment of included studies in a systematic review is an important component that should be considered in relation to the interpretation of results since individual studies may be at risk of producing bias results or have applicability concerns with respect to the review question. This may go some way to explaining sources of heterogeneity in the individual DTA studies,⁷ although in our experience individual study quality data and individual test accuracy estimates are rarely visualized in the same figure. Similarly, study level covariates may indicate important study differences that contribute to the between study heterogeneity in a meta-analysis.

In Section 2 of this paper, the current guidelines for DTA systematic reviews are critiqued to establish the current best practice recommendations for visualization of results and quality assessment. Section 3, based on the findings from the review of guidelines, describes two new plot enhancements to the SROC space incorporating study quality and percentage study weights. It highlights how improved presentation of current DTA meta-analyses can lead to a deeper understanding and insight into the often heterogeneous results that make up a DTA meta-analysis. Section 4 discusses the implementation of the enhanced SROC plots using the MetaDTA App which allows direct user interactivity to enable exploration and facilitate

Highlights

What is already known?

Three prominent guidelines for the conduct of diagnostic test accuracy (DTA) systematic reviews and meta-analyses (The Cochrane Handbook for DTA studies, PRISMA for DTA and Methods Guide for Medical Test Reviews) suggest the use of two plots to display results:

1. Forest plots to display individual study estimates and pooled meta-analysis results for sensitivity and specificity separately.
2. Summary receiver operating characteristic (SROC) curve plots to display a summary of the performance of a diagnostic test based on data from a meta-analysis.

All three guidance documents acknowledge the importance of quality assessment but do not consider how to incorporate quality assessment alongside forest or SROC plots.

What is new?

Two novel SROC plot enhancements:

1. Quality assessment enhanced SROC plot that *simultaneously* visualizes all dimensions of study quality based on the QUADAS-2 tool.
2. Percentage study weighted SROC plot that visualizes the relative contribution of each study to the pooled sensitivity and specificity results using percentage weights.

Both of these plots can easily be produced with the interactive web application MetaDTA along with a variety of other features.

Potential impact for RSM readers outside the authors' field?

Our enhancements allow for the exploration and investigation of heterogeneity and bias based on the quality of included studies and the identification of influential studies based on the percentage study weights. It is hoped that the enhancements and software implementation detailed here will be used by those conducting DTA reviews to facilitate the discovery of important insights in their data and improve the visualization and reporting of results.

understanding of DTA meta-analysis results. Section 5, the discussion, concludes the paper.

2 | DTA SYSTEMATIC REVIEWS GUIDANCE ON VISUALIZATION OF RESULTS

Three prominent guidelines for conducting DTA systematic reviews and meta-analyses were identified and reviewed to establish current recommendations for graphical presentation of DTA Systematic review results:

1. The Cochrane Handbook for DTA studies,⁴ in which chapters have been published individually since 2008 (but incomplete as of January 2020);
2. PRISMA for DTA,³ an extension of the PRISMA initiative for the transparent reporting of systematic reviews and meta-analyses of interventions. Includes a 27-item reporting check-list underpinned by a recent systematic review of guidance documents for DTA systematic reviews; and
3. Methods Guide for Medical Test Reviews published by the US Agency for Healthcare Research and Quality (AHRQ) in 2012.⁸

All three guidance documents recommend the use of two main forms of graphical display for reporting of DTA meta-analyses: (a) forest plots to display the individual study estimates and meta-analysis pooled result for sensitivity and specificity separately, and (b) SROC plots to display a global summary of the performance of a diagnostic test based on data from a meta-analysis. However, neither PRISMA-DTA³ nor AHRQ handbook⁸ provide specific guidance on what could/should be included in the ROC space; although the AHRQ handbook does provide illustrative analysis examples that include plots of study-level covariate values using numbers and different symbols in the ROC space as well as bounded areas distinguishing likelihood ratio threshold values.

In contrast, the Cochrane Handbook for DTA studies⁴ provides an overview of graphical presentations recommended for use in Cochrane DTA reviews all of which are possible to create within RevMan.⁹ In particular, the display of the results of individual studies in ROC space can be specified in RevMan using: (a) Different plotting symbols or colors to indicate covariate values including markers of quality to enable the exploration of heterogeneity; (b) Different size sensitivity-specificity points to depict the precision of the estimate or sample size of the study; and/or (c) Adding cross-hairs⁶ to each study point to display the confidence limits for sensitivity and specificity. Finally, the presentation of the summary results of the meta-analysis can be displayed as an SROC curve and a summary sensitivity and specificity point with associated confidence and/or prediction regions.

With respect to scaling the study estimates in ROC space relative to their sample size in order to reflect precision of the study estimates, this has been highlighted as misleading as they may be wrongly interpreted as percentage study weights. Percentage study weights quantify the relative contribution of each study to the pooled meta-analysis result¹⁰ and PRISMA states their preferred inclusion on forest plots for meta-analyses of healthcare interventions.¹¹

In terms of quality assessment, all three guidance documents suggest that individual studies should be investigated in terms of risk of bias and applicability to the review question with the methods and results being clearly explained. While the PRISMA-DTA checklist and AHRQ Handbook do not prescribe explicit details, the Cochrane handbook for DTA reviews recommends the use of the QUADAS-2 tool, which assesses four domains: (a) patient selection, (b) index test, (c) reference standard and (d) the flow and timing in terms of risk of bias and additionally the first three with respect to applicability.⁷ Each domain is scored as having “high,” “low,” or “unclear” risk of bias or applicability concern for the relevant domains. It is recommended that Cochrane DTA reviews display results from QUADAS-2 using two different visual presentations: (a) a methodological quality summary plot which presents the seven quality assessment outcomes for each study; and (b) a methodological quality graph which presents a stacked bar chart showing the percentage of studies scoring high, low and unclear for each of the 7 quality outcomes. Both plots use a traffic light system of green, amber and red to display high, low or unclear quality respectively (Figure 1).

Despite all the guidelines acknowledging the importance of a risk of bias/applicability assessment, no explicit consideration of a plot to *simultaneously* present more than one dimension of a studies quality/relevance and its results to allow a visual assessment of whether there is a relationship between the two was present in any of the guidelines. Further, the simultaneous representation of a studies percentage contribution to pooled sensitivity and specificity in the context of ROC space appeared to have been overlooked. A critical exploration and implementation of these omissions, with the aim of improving reporting and facilitating interpretation of DTA meta-analyses is described in the remainder of this paper.

3 | GRAPHICAL ENHANCEMENTS TO SROC PLOTS

3.1 | Quality Assessment enhanced SROC plot

The Quality Assessment enhanced SROC plot incorporates the quality assessment information from all

7 dimensions of QUADAS-2 in the ROC space simultaneously. The QUADAS-2 tool assesses risk of bias across 4 domains and applicability across a further 3 domains. While individual markers of study quality could be represented on SROC plots using different plotting symbols or colors (eg, representing having “high,” “low,” or “unclear” risk of bias), representing each distinct domain in this way would be unwieldy as multiple plots would be required to present all the information.

Our proposed solution is to allow all 4 risk of bias outcomes, all 3 applicability outcomes, or all 7 outcomes across both domains to be displayed simultaneously through the use of glyphs. To enable all 4 risk of bias outcomes to be displayed, we divide a circular symbol (the glyph) into 4 equal segments, each representing one of the four items being assessed, and color each of the segments either green, red or grey to represent “high,” “low,” and “unclear” risk of bias respectively (Figure 2A). Similarly, for applicability concerns the glyph is split into three sections (Figure 2B). To display all of the quality assessment outcomes the glyph can also be split into seven sections (Figure 2C). These glyphs are then used as plotting symbols for individual studies in the ROC space.

To illustrate the use of the quality assessment enhanced SROC plot, we use data from a published systematic review investigating the use of the IQCODE (Informant Questionnaire on Cognitive Decline in the Elderly) tool to screen adults for dementia within a secondary care setting. The review consists of 13 studies and 2745 adults.¹⁴

Figure 3A shows an SROC plot which incorporates quality assessment data from one dimension of the QUADAS-2 tool—risk of bias in patient selection. The study estimates are highlighted either red, green, or grey to indicate high, low or unclear risk of bias for this dimension. For this example it can be seen that the majority of the studies have a high risk of bias with regards to patient selection—8 out of 13 studies. While 6 similar plots could be constructed for each of the other domains, this would require a lot of space with information being strewn across the 7 plots, making it difficult to ascertain overall trends across all domains.

In Figure 3B all 4 risk of bias assessment outcomes and 3 applicability concerns outcomes are presented using the glyph system. Much like the methodological quality summary plot (Figure 1) we can identify which studies are at high risk of bias or have high applicability concerns but now we have the added benefit of being able to see these in relation to the individual study and the meta-analysis results. As well as providing a quick overview of study quality without the need for a separate plot, the plot allows exploration of whether lower (or uncertain) quality studies' results differ systematically

from the higher quality studies, or whether there is more variability across lower quality studies suggesting variable quality may be inducing heterogeneity. Further, in the illustrative plot, since all studies have bias concerns in one or more domains this reminds the reader (and the analyst!) not to over-interpret the pooled result which is an easy mistake to make when quality is reported in a detached manner, as it is typically done, via separate tables or figures (such as Figure 1).

3.2 | Percentage study weight enhanced SROC plot

This plot enhances the SROC plot by incorporating percentage study weights to show the relative contribution of each study to the pooled meta-analysis result. As noted above, including study sample size on a SROC plot can be misleading as it does not represent a study's contribution to the analysis. An alternative is to plot the percentage study weight; however, compared to meta-analyses of interventions, this is much harder to compute when using the recommended bivariate model for DTA data. Our enhanced SROC plot displays the percentage study weight using the method proposed by Burke et al¹⁰ based on a decomposition of Fisher information matrix.¹⁵ It is important to note that as both sensitivity and specificity are being estimated, it is usual for each study to contribute different weightings to each of these outcomes. In order to allow for this differential contribution, instead of representing studies with circles with different radii to represent precision as is typically done on meta-regression plots,¹⁶ ellipses with axes aligned with those of the SROC plot are used. The height and width of the ellipses varies to be proportional to the percentage study weights from the bivariate meta-analysis model.

To illustrate the use of the plot, we use a meta-analysis of 23 studies (containing 4100 children) estimating the accuracy of infrared ear thermometers for diagnosing fever. Most studies defined a cut-off of 38°C for diagnosing fever and used rectal temperature as a reference standard. Further details can be found in the original review.¹³

Figure 4A shows the SROC plot presented in the original meta-analysis.¹³ The solid red line represents the SROC curve, the dashed green lines show the 95% confidence limits and the solid triangles indicate the summary estimates (the upper triangle relates to a fixed effect analysis and the lower relates to a random effect analysis). Study estimates are shown as blue hollow circles with the size of the circle representing the proportion of children in each study. In contrast, Figure 4B displays the study estimates as black hollow circles with the size and shape

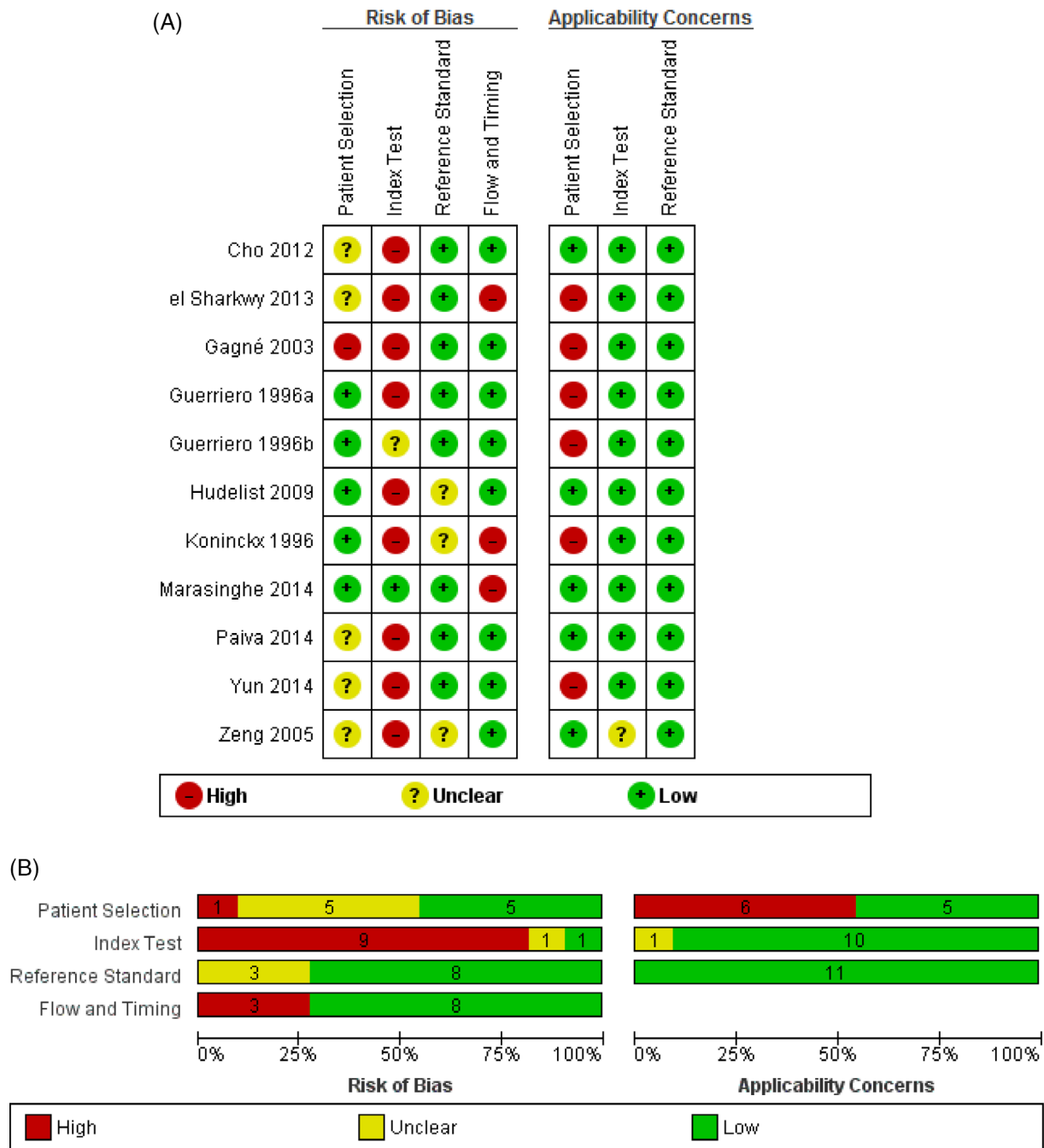


FIGURE 1 A, Example of a methodological quality summary plot used in the following Cochrane DTA systematic review: Nisenblat V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev*, no. 7, 2016.¹² (Figure 3: Risk of bias and applicability concerns summary, reproduced with permission). B, Example of a methodological quality summary table used in the following Cochrane DTA systematic review: Nisenblat V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev*, no. 7, 2016.¹² (Figure 4: Risk of bias and applicability concerns graphs, reproduced with permission). DTA, diagnostic test accuracy [Colour figure can be viewed at wileyonlinelibrary.com]

of the circles representing the percentage study weights for sensitivity and specificity. In this example, it can be observed that all of the points are fairly similar in size indicating an equal weighting of the studies within the meta-analysis; this is typical in the presence of

heterogeneity. It can also be ascertained that the studies with the most extreme values for either sensitivity or specificity (toward the top left-hand-side of the plot) are generally less precise and contribute less weight. In the original meta-analysis displayed in Figure 4A, the

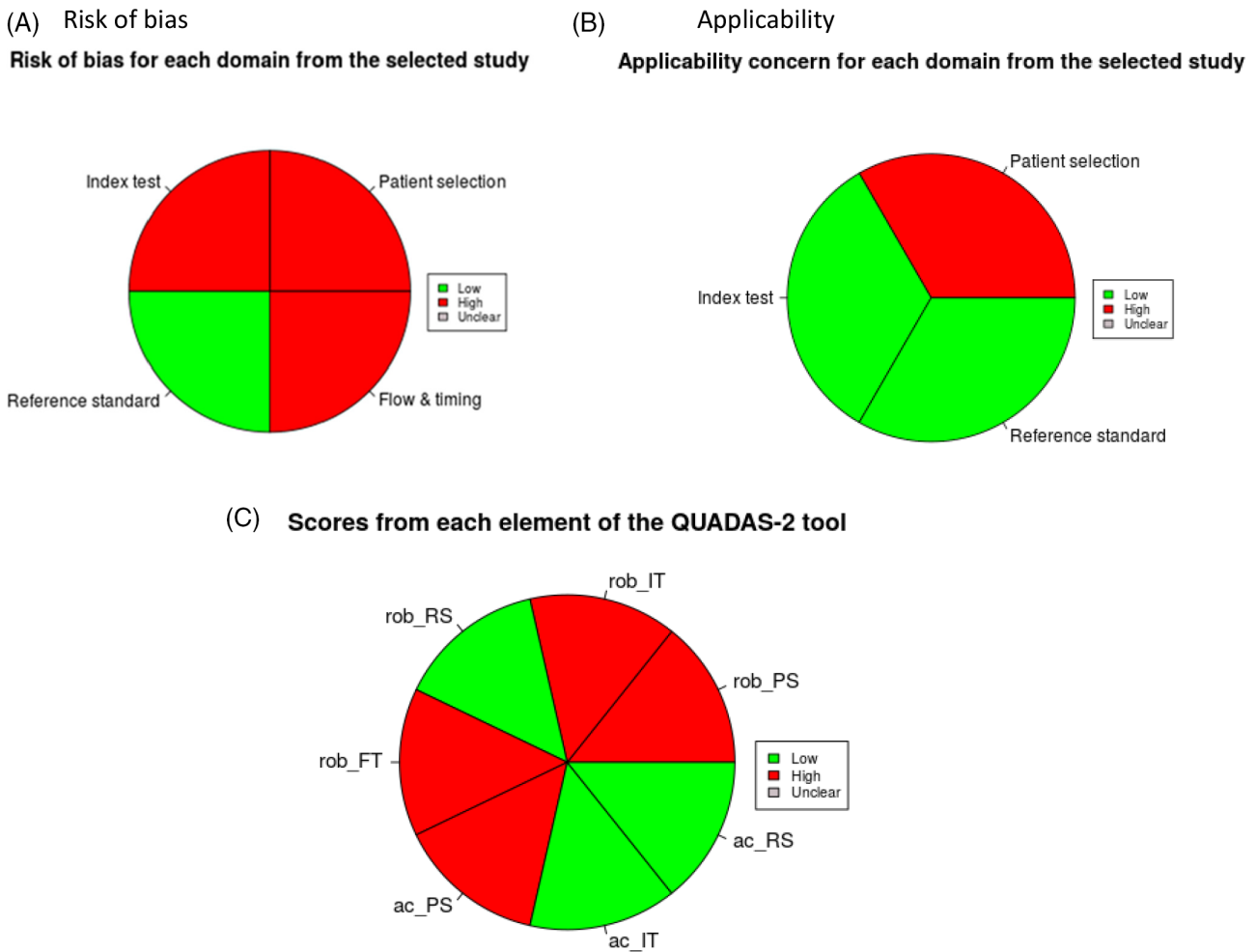


FIGURE 2 Glyphs for presenting QUADAS-2 results. A, Risk of bias. B, Applicability. C, All quality assessment outcomes (risk of bias and applicability). ac_IT, applicability concerns-index test; ac_PS, applicability concerns-patient selection; ac_RS, applicability concerns-reference standard; rob_FT, risk of bias-flow and timing; rob_IT, risk of bias in terms of index test; rob_PS, risk of bias in terms of patient selection; rob_RS, risk of bias-reference standard [Colour figure can be viewed at wileyonlinelibrary.com]

NyPaver study is identified as having the largest sample of children as indicated by the largest blue circle on the ROC plot. However, when the ROC plot is re-created to display percentage study weights rather than sample size (Figure 4B), the size of the ellipse for the Nypaver study is much more comparable to those for other studies in the review, indicating the relative down-weighting of studies in a random effects analysis when heterogeneity is present. Thus, Figure 4B reveals that even though this study is the largest in terms of sample size, its contribution to the meta-analysis is very similar to other studies included in the review.

Percentage study weights are rarely derived in DTA reviews, and we know of no examples where they have been plotted. Thus, this plot allows readers, for the first time, to gauge the relative contribution of each study using an easy visual summary displayed on a SROC plot. If study weighting is more variable than in the example

presented here, the plot will also be useful for identifying influential studies. The robustness of the results to the omission of influential studies can be assessed via sensitivity analysis to ascertain if a review's conclusions are dependent on the inclusion of one or more particular studies. Since symbol size proportionality to indicate study precision/contribution has become a regular feature of forest and meta-regression plots, their use on SROC plots should be intuitive for users to understand.

4 | INTERACTIVITY

The two enhancements to SROC plots presented above were specifically designed to facilitate the conduct, reporting and interpretation of DTA meta-analysis. MetaDTA⁵ (https://crsu.shinyapps.io/dta_ma/) is a free web-based App with a “point and click” interface therefore

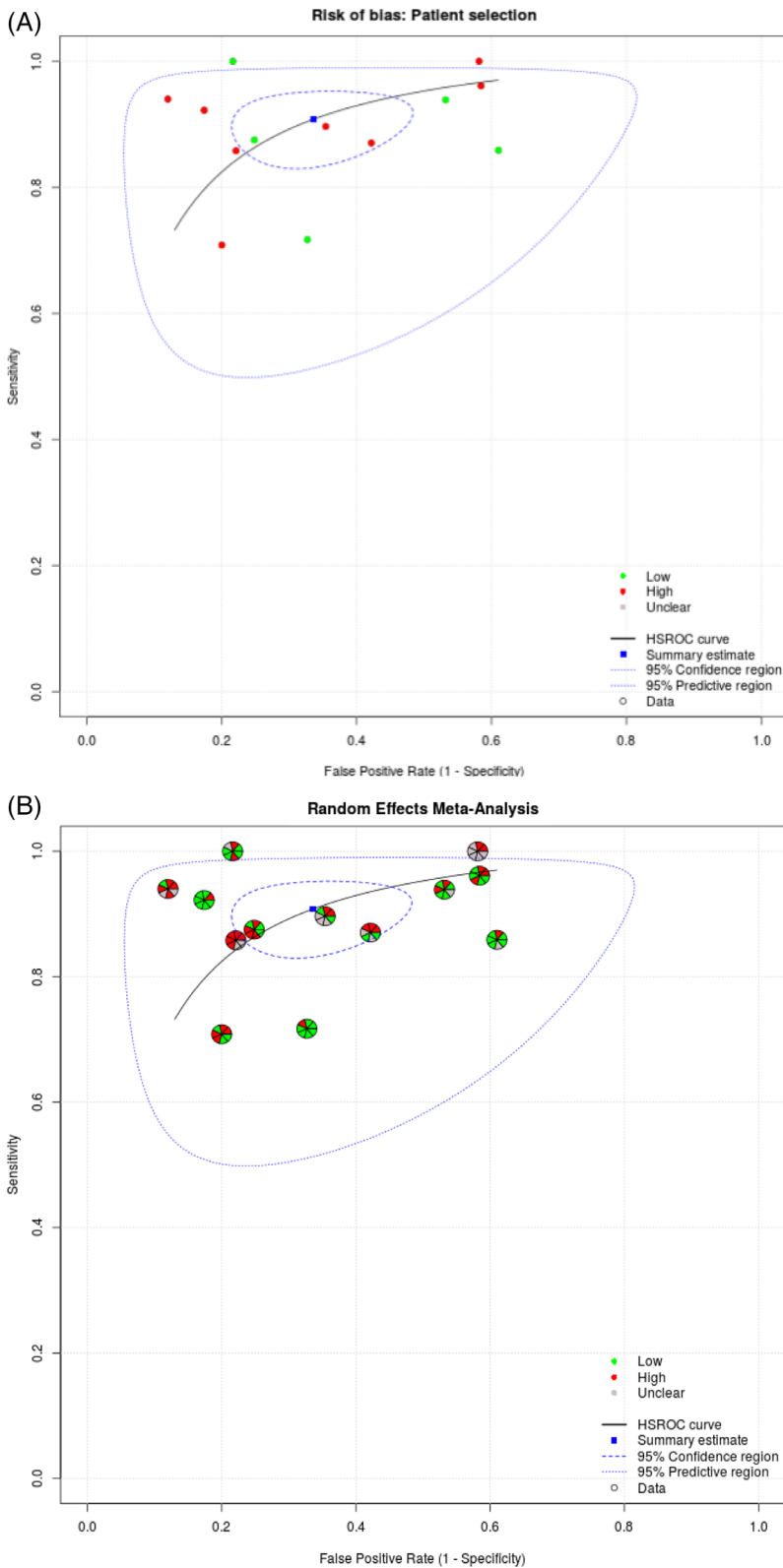
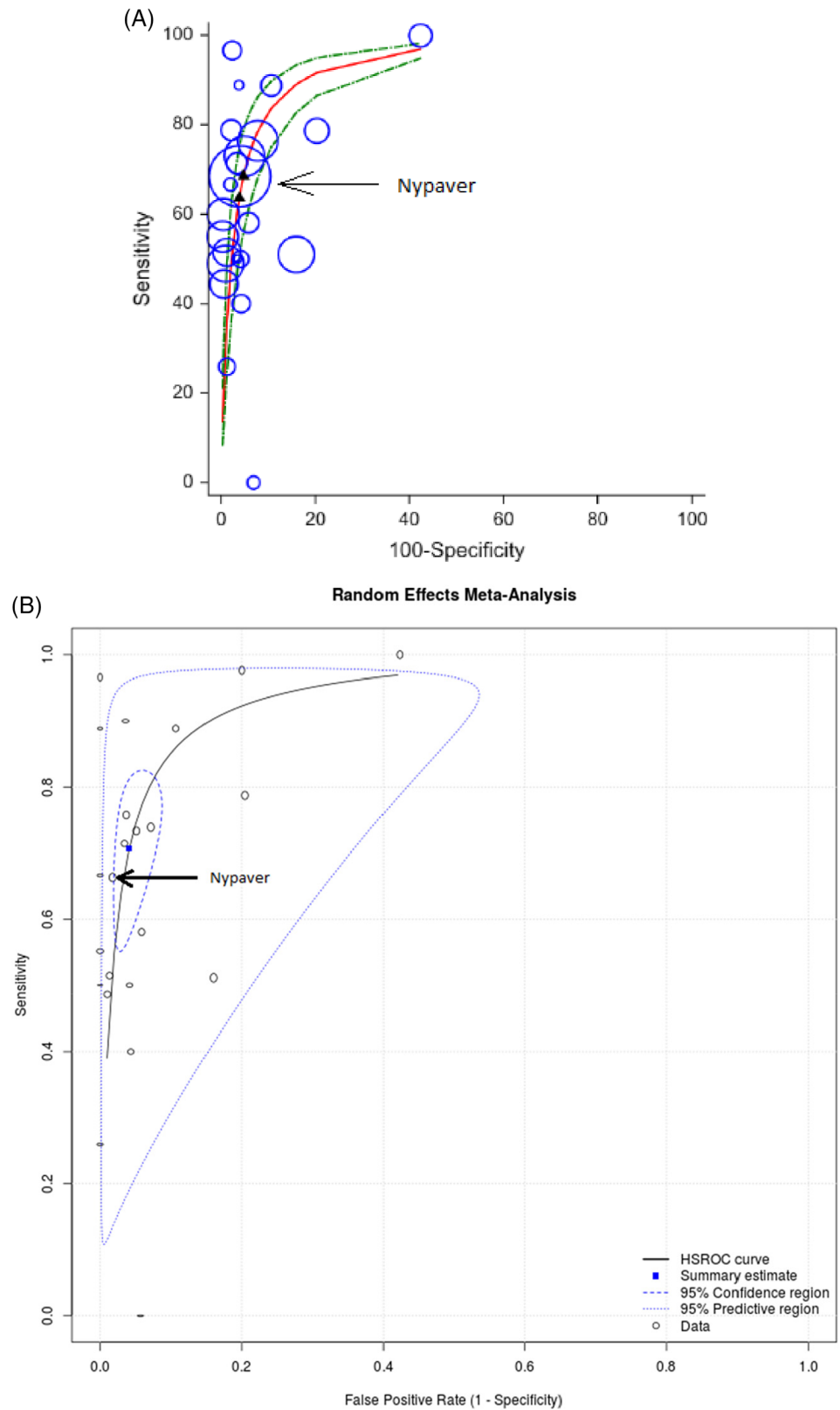


FIGURE 3 Quality assessment plot. A, ROC plot highlighting the quality assessment scores of each study in terms of risk of bias in patient selection. B, ROC plot highlighting the quality assessment scores in terms of risk of bias and applicability concerns. Each quadrant represents one of the seven domains. Going clockwise, from the top, the domains are risk of bias-index test, risk of bias-patient selection, applicability-reference standard, applicability-index test, applicability-patient selection, risk of bias-flow and timing, risk of bias-reference standard. ROC, receiver operating characteristic [Colour figure can be viewed at wileyonlinelibrary.com]

knowledge of statistical software is not required making it accessible to novices and experts alike. The App is fully interactive and can be used as an exploratory tool as well as for analysis of a dataset. In addition to implementing the plot enhancements described above, it implements

pre-existing enhancements suggested in guidelines (see Section 2) and allows for multiple features to be added to the SROC plot simultaneously. For example, Figure 5A shows a plot for the IQCODE test (described in Section 3.1) displaying the risk of bias for patient selection (color coded

FIGURE 4 Percentage study weight plot. A, Original ROC plot used by Dodd et al¹³ to present the results from the meta-analysis. The blue circles represent the study estimates and the size of the circles represent the study sample size. (Figure 1: Summary ROC curve [95% CI] estimated using an inverse weighted random effects model assuming constant DOR, reproduced with permission). B, A re-creation of the ROC plot by Dodd et al. The black circles represent the study estimates, and the size and the shape of the circles represent the percentage study weights. ROC, receiver operating characteristic [Colour figure can be viewed at wileyonlinelibrary.com]



study-specific estimates, the percentage study weights (shape and size of the points) and the covariate test threshold (Note: it is not possible to change the size and shape of the multiple quality domain glyphs). This type of plot could be used to investigate the quality of studies

contributing the most to the meta-analysis. In addition, adding the covariate threshold to the plot allows the relationship between the quality of the studies and threshold to be investigated. Figure 5B shows how crosshairs can be added to the same plot to display the uncertainty in the

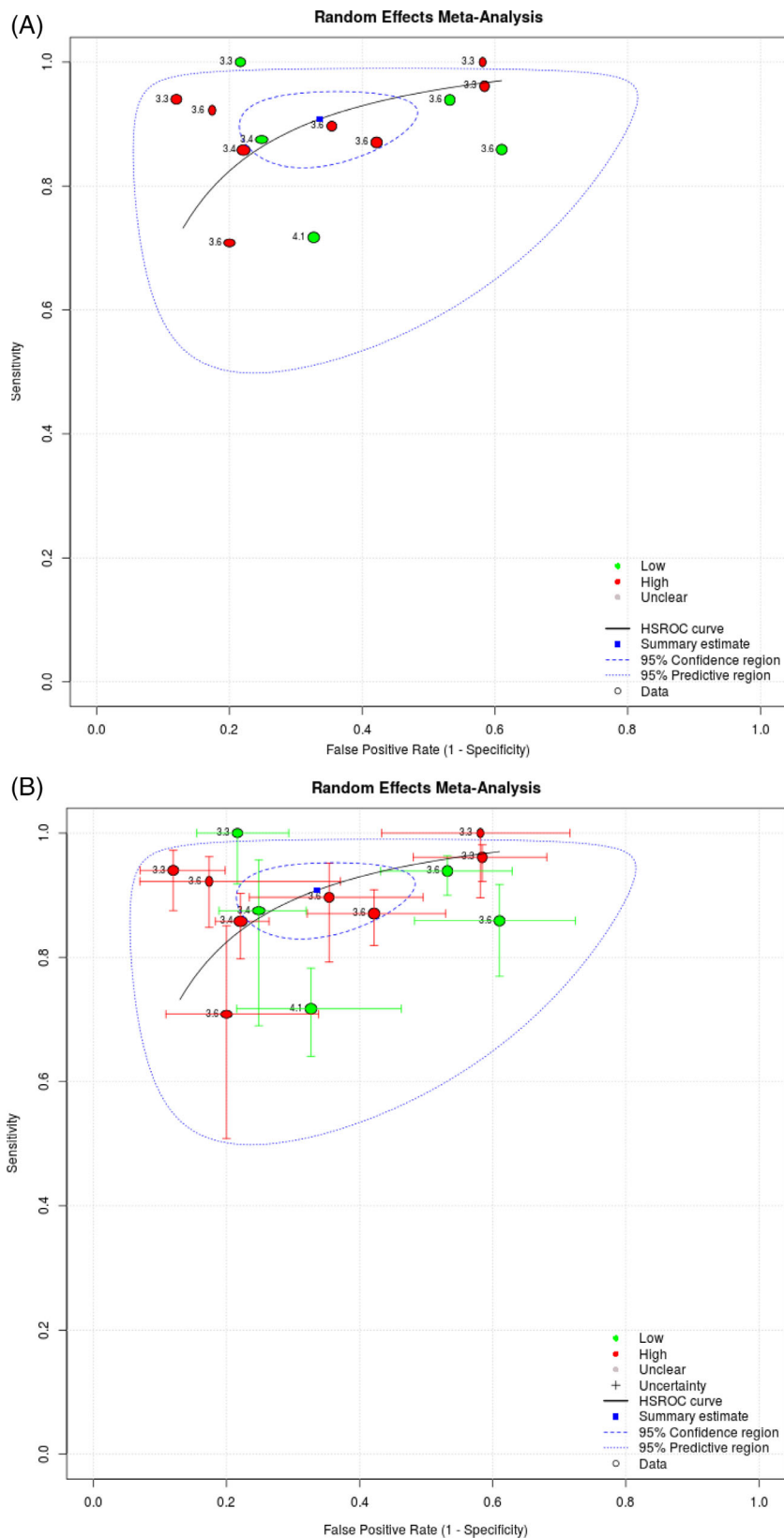


FIGURE 5 An ROC plot using a variety of features. Study estimates are highlighted corresponding to their quality assessment score in terms of risk of bias toward patient selection. The shape of study estimates have also been changed to indicate percentage study weights. Cross-hairs are added to display the uncertainty in sensitivity and specificity for each study. The estimates have also been labeled to indicate test threshold. A, Adding percentage study weight, risk of bias toward patient selection and covariate threshold to the SROC curve. B, As above with cross-hairs added. ROC, receiver operating characteristic; SROC, summary receiver operating characteristic [Colour figure can be viewed at wileyonlinelibrary.com]

sensitivity and specificity for each study. Furthermore, this exploration of the data can inform the sensitivity analysis; for example, studies of higher risk of bias may easily be excluded from the plots and the analysis to assess their

impact on the meta-analysis results. Clicking on study-specific points on the SROC plot displays the following information about the study below the plot—author and year, sensitivity and specificity. In addition, study

weights for sensitivity and specificity, and covariate value may also be displayed below the plot if these options are selected. Also, when using the multiple dimensional study quality glyphs (as described in Section 3.1), by clicking on a glyph it will produce a larger version of the glyph with the different dimensions clearly labeled as shown in Figure 2.

5 | DISCUSSION

In this article, we have reviewed current guidelines for the presentation of results from DTA systematic reviews, and identified shortcomings with respect to visual displays of study quality and study weighting information within the ROC space. In response to these findings, we have detailed two novel enhancements to SROC plots: (a) the quality assessment enhanced SROC plot that simultaneously incorporates multiple individual markers of study quality based on QUADAS-2 results; and (b) percentage study weighted plotting points that show the relative contribution of each study to the pooled sensitivity and specificity results using weights calculated via the method derived by Burke et al.¹⁰ The MetaDTA App enables these plots to be constructed easily and its interactive platform allows further exploration by combining multiple features within a single plot including study quality, percentage study weights and covariates such as test threshold; thus facilitating the investigation of relationships between variables (eg, do the studies with low risk of bias have the largest percentage weight in the meta-analysis, etc).

The main limitation of the proposed enhanced SROC plots is, like all SROC plots, if there are a lot of studies included in the meta-analysis the study-specific estimates may overlap one another and the plots become cumbersome and difficult to read. However, when plotted in MetaDTA, the interactive platform allows the user to click on specific estimates and obtain additional details about the study including author, sensitivity and specificity estimates, and, if selected, an enlarged version of the glyphs displaying the quality assessment results by domain. The interactive platform also allows features to be toggled off if the plot becomes too overcrowded with information. Where multiple features are of interest but cannot be displayed effectively on a single plot, multiple plots can be produced that still aid exploration of the relationships of variables across plots if presented next to one another or quickly toggling between alternative information via the App.

A valuable extension to the implementation of the proposed plots within the App would be the possibility of developing plots for inclusion in digital report formats

(eg, web-based papers or The Cochrane Library) that were interactive. This would allow the interested reader of a meta-analysis similar freedoms to those afforded by the analyst using the App and allow them to explore and even re-analyze the data and not have to rely only on the views of the data included in the report.

Throughout this paper we have focused on the evaluation of the diagnostic accuracy of a single reference test. However, in a decision making context, the interest is often in comparing the performance of multiple diagnostic tests. Methods, analogous to network meta-analysis for comparing the effectiveness of multiple treatment interventions, are less established for diagnostic accuracy although some methods have been proposed.¹⁷⁻²² Such analyses produce a vast quantity of results and therefore presenting them in an informative and interactive way will be a challenge for future research.

In conclusion, it is hoped that the SROC plot enhancements presented here will be used by those conducting DTA reviews and will assist with unpicking what are often heterogeneous meta-analyses. Even when this is not possible, since a meta-analysis is only as good as the studies going into it, presenting quality assessment information in the same plot as the main analysis will go a long way to ensuring that this is not forgotten or ignored when interpreting the results (which is all too easy to do when the quality information is only presented in separate plots within separate sections of a report). In a similar vein, the inclusion of percentage study weights rather than sample size or precision will make it easier to ascertain the influence of individual studies within the analysis.

ACKNOWLEDGMENTS

Amit Patel is funded by a National Institute for Health Research Systematic Review Fellowship (NIHR-RM-SR-2017-08-023) and Suzanne Freeman is funded by a National Institute for Health Research Post-Doctoral Fellowship (PDF-2018-11-ST2-007). Suzanne Freeman, Nicola Cooper and Alex Sutton are funded by the National Institute for Health Research Complex Reviews Support Unit (14/178/29). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the NIHR, NHS or the Department of Health. The NIHR had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

CONFLICT OF INTEREST

The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT

The enhanced SROC plots presented in this paper are available in the MetaDTA App (https://crsu.shinyapps.io/dta_ma/).

ORCID

Nicola Cooper  <https://orcid.org/0000-0002-4486-2791>

REFERENCES

1. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Syst Rev.* 2013;2(1):82.
2. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics.* 2006;8(2):239-251.
3. M. D. McInnes, D. Moher, B. D. Thombs, T. A. McGrath, P. M. Bossuyt, T. Clifford, J. F. Cohen, J. J. Deeks, C. Gatsonis, L. Hooft and others, "Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement," *JAMA*, vol. 319, no. 4, pp. 388–396, 2018.
4. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 0.9.0.* London: The Cochrane Collaboration; 2010.
5. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. *BMC Med Res Methodol.* 2019;19(1):81.
6. Phillips B, Stewart LA, Sutton AJ. 'Cross hairs' plots for diagnostic meta-analysis. *Res Synth Methods.* 2010;1(3–4):308-315.
7. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.
8. Chang SM, Matchar DB, Smetana GW, et al. *Methods guide for medical test reviews.* Rockville: Agency for Healthcare Research and Quality; 2012.
9. The Nordic Cochrane Centre. *Review Manager (RevMan) Version 5.3.* Copenhagen: The Cochrane Collaboration; 2014.
10. Burke DL, Ensor J, Snell KI, van der Windt D, Riley RD. Guidance for deriving and presenting percentage study weights in meta-analysis of test accuracy studies. *Res Synth Methods.* 2018;9(2):163-178.
11. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.
12. Nisenblat V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev.* 2016;7:1-119.
13. Dodd SR, Lancaster GA, Craig JV, Smyth RL, Williamson PR. In a systematic review, infrared ear thermometry for fever diagnosis in children finds poor sensitivity. *J Clin Epidemiol.* 2006;59(4):354-357.
14. Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant questionnaire on cognitive decline in the elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database Syst Rev.* 2015;3:1-78.
15. Riley RD, Ensor J, Jackson D, Burke DL. Deriving percentage study weights in multi-parameter meta-analysis models: with application to meta-regression, network meta-analysis and one-stage individual participant data models. *Stat Methods Med Res.* 2018;27(10):2885-2905.
16. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med.* 2002;21(11):1559-1573.
17. Trikalinos T, Hoaglin D, Small K, Terrin N, Schmid C. Methods for the joint meta-analysis of multiple tests. *Res Synth Methods.* 2014;5(4):294-312.
18. Menten J, Lesaffre B. A general framework for comparative Bayesian meta-analysis of diagnostic studies. *BMC Med Res Methodol.* 2015;15:70.
19. Hoyer A, Kuss O. Meta-analysis for the comparison of two diagnostic tests to a common gold standard: a generalized linear mixed model approach. *Stat Methods Med Res.* 2018;27:1410-1421.
20. Nyaga V, Arbyn M, Aerts M. Beta-binomial analysis of variance model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res.* 2018;27(8):2554-2566.
21. Dimou D, Adam M, Bagos P. A multivariate method for meta-analysis and comparison of diagnostic tests. *Stat Med.* 2016;35:3509-3523.
22. Owen R, Cooper N, Quinn T, Lees R, Sutton A. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds. *J Clin Epidemiol.* 2018;99:64-74.

How to cite this article: Patel A, Cooper N, Freeman S, Sutton A. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. *Res Syn Meth.* 2021; 12:34–44. <https://doi.org/10.1002/jrsm.1439>