SOLUTION: Diagnostic Test Accuracy Meta-Analysis Worksheet

(Suzanne Freeman, Nicola Cooper and Alex Sutton)

Diagnostic tests generally comprise of a measure which splits individuals into healthy or diseased. To assess accuracy, a diagnostic test is compared to the "gold standard" test which is assumed to provide the true diagnosis of individuals. There are two parameters which are often used to assess the accuracy of diagnostic tests. Sensitivity is the proportion of patients with the disease correctly diagnosed by the test. Specificity is the proportion of patients without the disease correctly diagnosed by the test. For those new to this topic, an interactive primer on the evaluation of diagnostic test accuracy can be found here: https://crsu.shinyapps.io/diagprimer/.

A meta-analysis of diagnostic test accuracy (DTA) studies synthesises both sensitivity and specificity from multiple studies to evaluate the performance of a diagnostic test. The results are often presented either around a mean point or as a summary receiver operating curve (ROC). *MetaDTA* is a web-based App for conducting meta-analysis of diagnostic test accuracy studies. *MetaDTA* produces summary ROC plots, and pooled estimates for sensitivity and specificity together with uncertainty in their estimation and can also be used to aid sensitivity analyses by excluding studies.

After completing this worksheet, you should be able to:

- Use *MetaDTA* to perform a meta-analysis of diagnostic test accuracy studies, obtain pooled estimates of sensitivity and specificity and interpret the results
- Use *MetaDTA* to produce and download a summary ROC plot
- Use *MetaDTA* to examine the influence of studies when they are included and excluded in a sensitivity analysis

MetaDTA can be accessed from: https://crsu.shinyapps.io/dta_ma/. From here please follow the link to the Beta version of MetaDTA. A copy of the user guide can be downloaded from the home page.

MetaDTA has four inbuilt datasets which can be used to familiarise yourself with the features of MetaDTA. The four datasets are also available to download in csv format and can be used to help ensure that your own data is in the correct format for upload to MetaDTA. These datasets can be accessed from the grey box on the Load Data page. The example datasets come from a systematic review investigating the accuracy of an informant-based questionnaire, for detection of all cause dementia in adults. The datasets consist of thirteen studies assessing the use of the IQCODE (Informant Questionnaire on Cognitive Decline in the Elderly) tool for identifying adults with dementia within a secondary care setting.

The IQCODE tool contains a number of questions which are scored on a five point scale. The IQCODE tool has a number of different variants, depending on how many questions are asked. The questions are based on the performance of everyday tasks related to cognitive function. These are then rated on a scale of 1-5. The final score is an average score for each question. The threshold used in each study is included as a covariate (see below). The IQCODE tool is only a screening tool and does not offer a definitive diagnosis of dementia. A high value on the IQCODE tool is taken to indicate that a patient is at risk of dementia and further clinical investigations are needed for diagnosis.

Start by selecting the 'With Quality Assessment and Covariates' dataset from the grey box on the left hand side of the 'Load Data' page and then use this dataset to complete the following questions:

1. Click on the 'Load Data' page and select the 'Data for Analysis' tab. Check you understand the data that is loaded and ready for analysis. (Optional: Calculate sensitivity by hand for Flicker and check your result using the 'Study-level Outcomes' tab on the 'Meta-Analysis' page).

The dataset consists of thirteen trials. In the third column TP represents the number of patients with a true positive test result. In the fourth column FN represents the number of patients with a false negative test result. In the fifth column FP represents the number of patients with a false positive test result. In the sixth column TN represents the number of patients with a true negative test result. The dataset also contains quality assessment results and three covariates. Risk of bias and applicability concerns were assessed in each of the studies using the QUADAS-2 tool with the results of the seven domains reported in columns 7-13. The values 1, 2 and 3 represent low, high or unclear risk of bias/applicability concerns respectively. The final three columns contain covariates: the threshold above which patients would be identified as possibly having dementia, the country the study was conducted in and IQCODE variant (there are three variants 16-item, 26-item and 32-item).

Optional: Sensitivity = Number of true positives / Total number of patients with the disease
Flicker: Sensitivity = 188 / (188+28) = 0.87

2. a) Use the 'Study-level Outcomes' tab on the 'Meta-Analysis' page to complete the following table:

Author	**Year**	**Sensitivity**	**Specificity**
Flicker	1997	0.870	0.578
Garcia	2002	0.922	0.826
Gonclaves	2011	0.717	0.673
Hancock	2009	0.859	0.390
Harwood	1997	1.000	0.784
Jorm	1991	0.708	0.800

b) Across all studies:

i) Which studies have the largest and smallest values for sensitivity?

Largest sensitivity = Mulligan and Harwood
Smallest sensitivity = Jorm

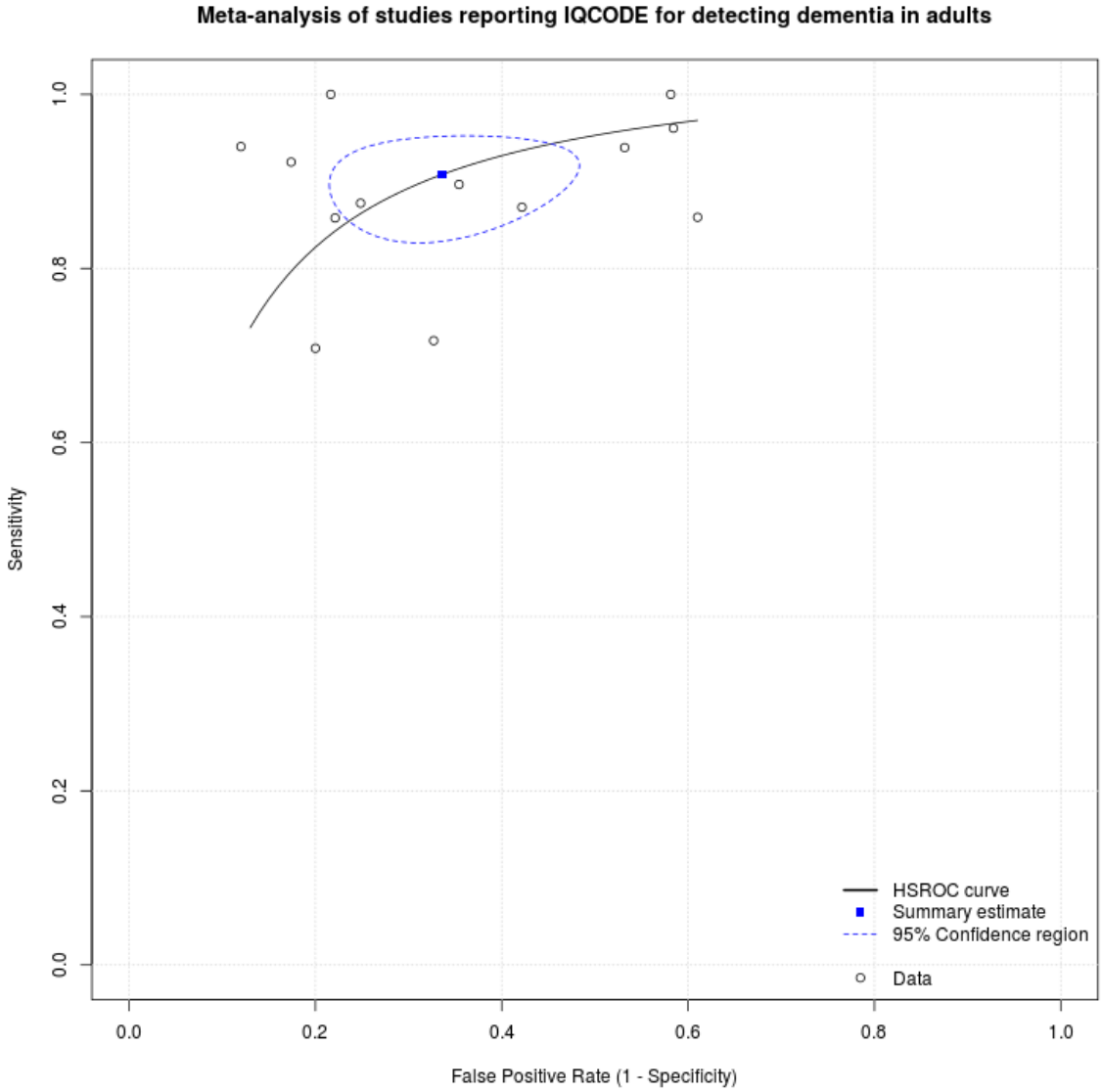
ii) Which studies have the largest and smallest values for specificity?

Largest specificity = Siri
Smallest specificity = Hancock

The largest and smallest values of sensitivity and specificity can be obtained in two ways:

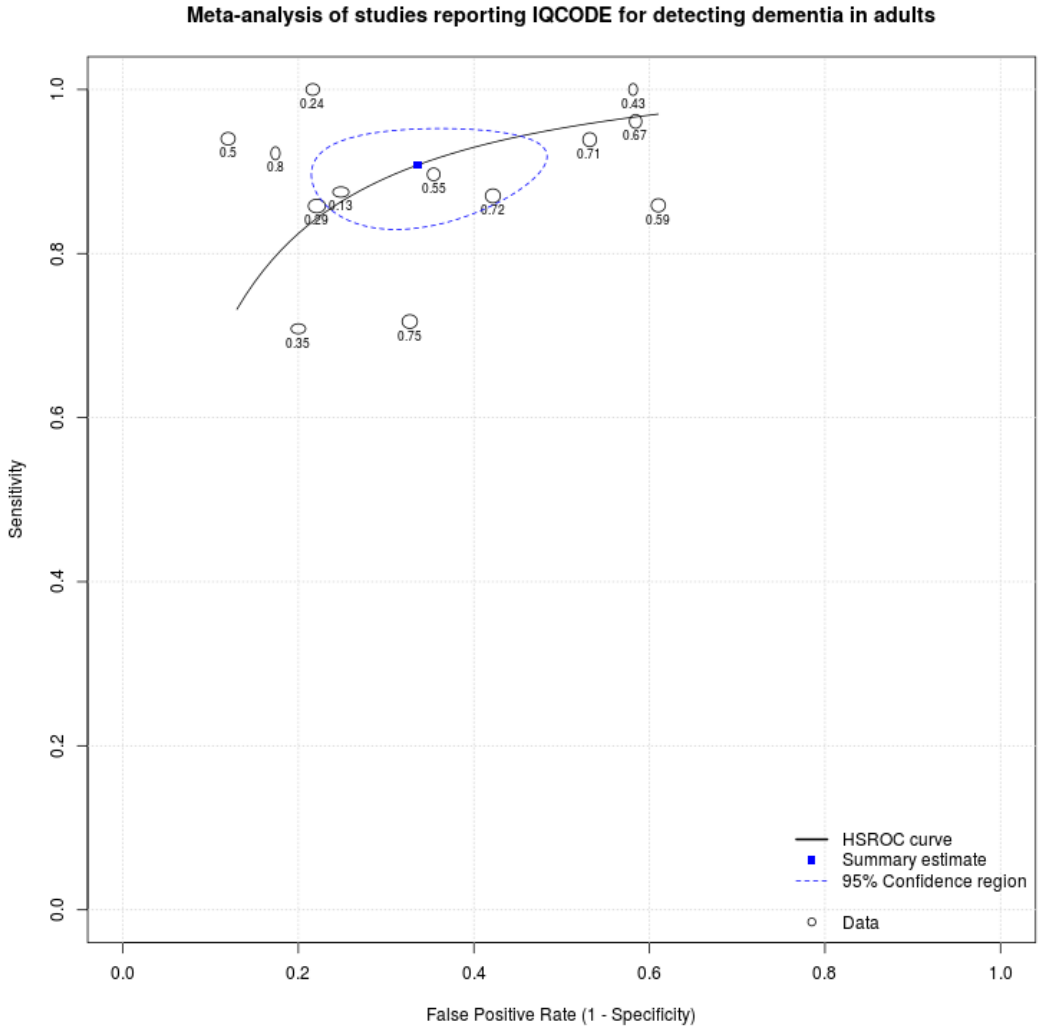
- *On the 'study-level outcomes' tab the triangles to the right of the column headings can be used to sort the data in the table by ascending or descending order.*
- *Using the 'ROC curve' tab on the 'sensitivity analysis' page. When you hover over a data point on the ROC plot the sensitivity and specificity corresponding to that study are displayed below the plot.*

3. a) Use the 'ROC Curve' tab on the 'Meta-Analysis' page to produce a plot which shows the HSROC curve (this stands for hierarchical summary receiver operatic characteristics curve and is a pooled curve taking all the data points into account and including random effects to allow for between study heterogeneity), summary estimate, 95% confidence region and individual study estimates. Change the title of the plot. (Optional: Download the plot and place into a Word document).



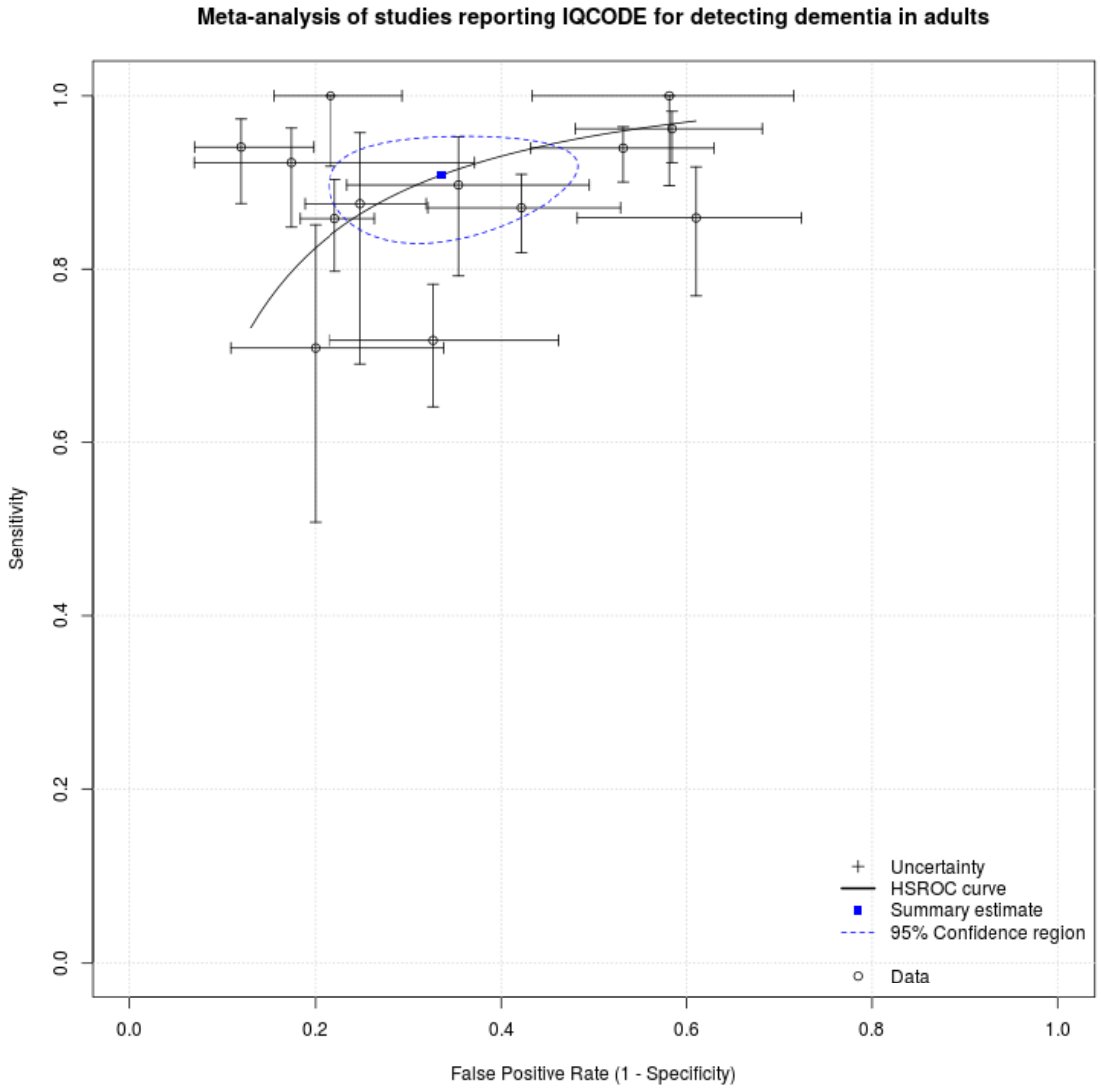
b) Display the disease prevalence and percentage study weights for each study

Disease prevalence and percentage study weights can be displayed by selecting the options from the grey box on the left hand side of the page. Disease prevalence is displayed numerically below each data point. The shape of the circles representing the study estimates are changed to reflect the percentage study weights for sensitivity and specificity.



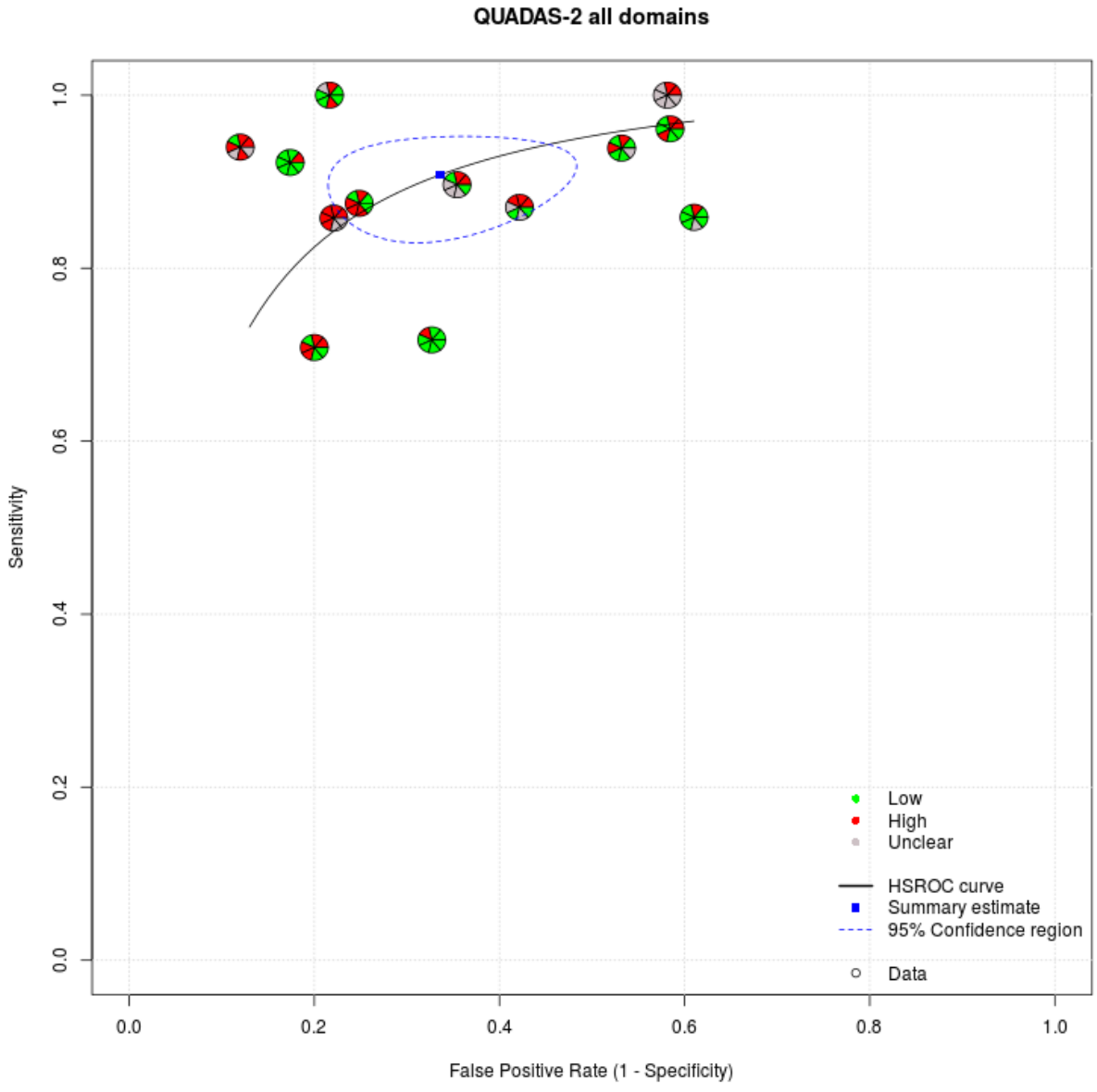
c) Display the 95% study-level confidence levels for sensitivity and specificity

95% study-level confidence intervals for sensitivity and specificity can be added to the HSROC plot by selecting the options from the grey box on the left hand side of the page.



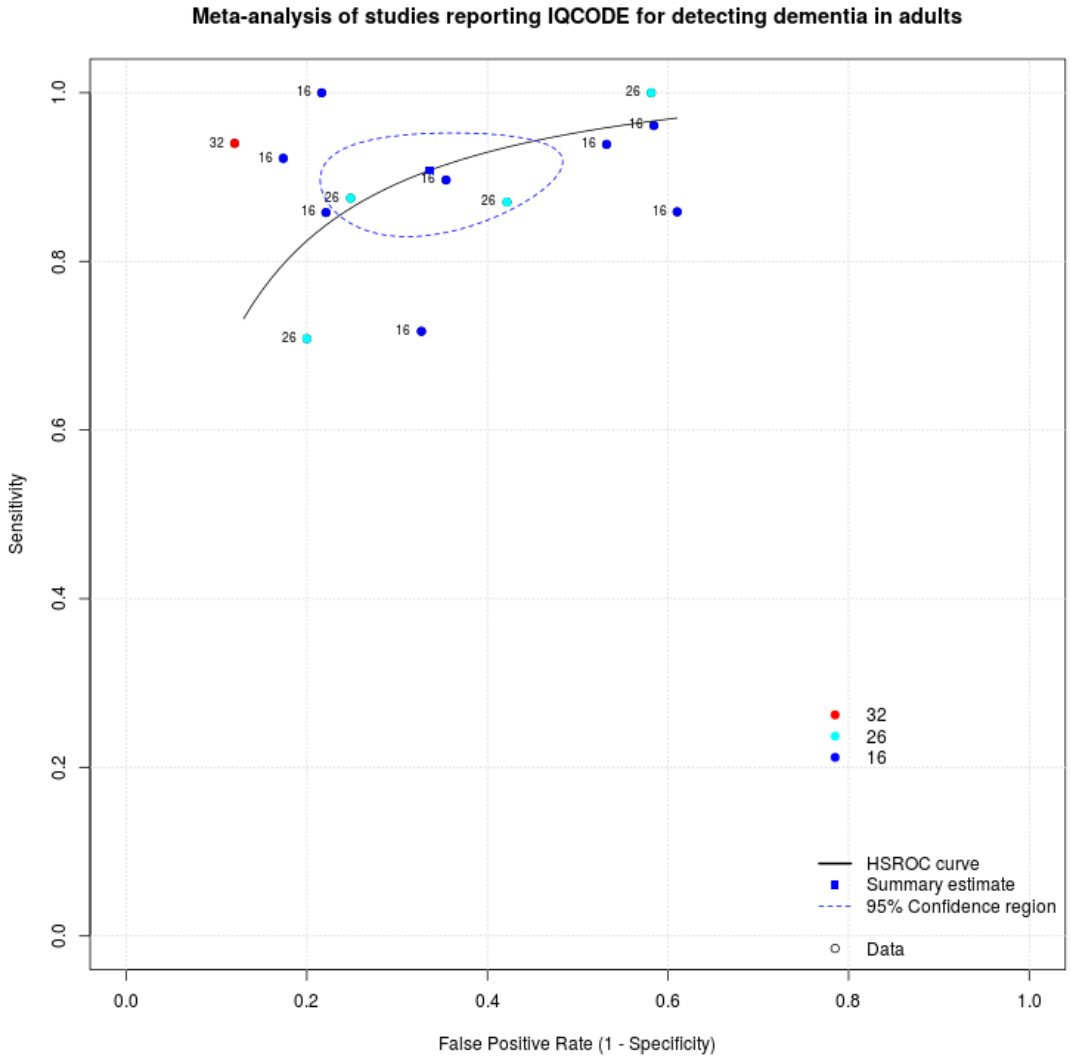
d) From the 'Display quality assessment scores' drop down menu in the grey box select the option 'Both risk of bias and applicability concerns'. Identify any studies where all seven QUADAS-2 domains are high or unclear risk of bias.

The data points have been replaced with glyphs with seven segments. Each segment represents one of the seven QUADAS-2 domains. Clicking on the centre of the data points produces a larger version of the glyph below the HSROC plot alongside the study name and study estimates of sensitivity and specificity. The two studies which have high or unclear risk of bias for all seven QUADAS-2 domains are Mulligan and Narasimhalu.



e) From the 'Display covariates' drop down menu in the grey box select the covariate IQCODE. Identify the one study which used the 32-item IQCODE variant.

Once a covariate is selected from the drop down menu another option will appear allowing you to display the covariate as text, coloured points or both. The only study which used the 32-item IQCODE variant was Siri.



f) Optional: Play around with the graphical options and see which options can be used in combination

4. a) Use the 'Statistics' tab on the 'Meta-Analysis' page to complete the following table:

Parameter	**Estimate**	**2.5%**	**97.5%**
Sensitivity	0.908	0.858	0.942
Specificity	0.664	0.563	0.752
False Positive Rate	0.336	0.248	0.437
Diagnostic Odds Ratio	19.501	8.025	30.977

b) Interpret the sensitivity, specificity and false-positive rate

Sensitivity of 0.908 (95% confidence interval (CI): 0.858, 0.942) indicates that 91% of people with a high IQCODE score would be correctly identified as being at risk of dementia by the screening test. Specificity of 0.664 (95% CI: 0.563, 0.752) indicates that 66% of patients with a low IQCODE score would be correctly identified as not being at risk of dementia by the screening test. The false positive rate is equal to 1-specificity. Therefore the false positive rate indicates that 34% of people with a high IQCODE score would be incorrectly identified as being at risk of dementia.

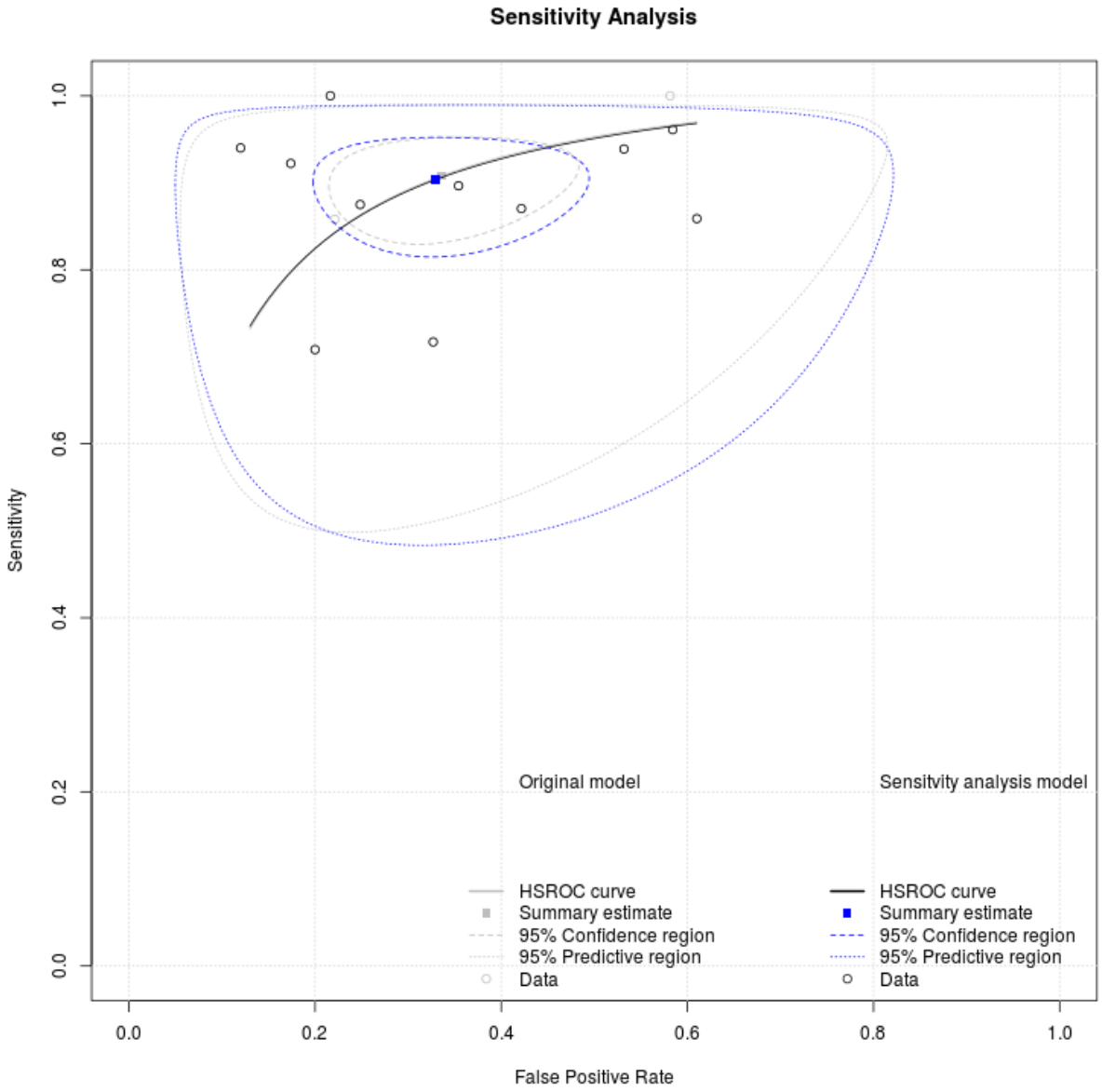
c) Interpret the diagnostic odds ratio

The diagnostic odds ratio represents how many times more likely a patient is to have the disease given that they have a positive test result compared to a patient with a negative test result. The diagnostic odds ratio of 19.5 (95% CI: 8.03, 31.0) indicates that patients with a high IQCODE score are almost 20 times more likely to be at risk of dementia than patients with a low IQCODE score.

5. Use the 'Sensitivity Analysis' page to exclude the two trials with high or unclear risk of bias for all seven QUADAS-2 domains:

   Exclude Mulligan and Narasimhalu

   a) Use the 'ROC Curve' tab to produce a plot which shows the data points, summary estimate, 95% confidence region and 95% predictive region. Change the title to "Sensitivity Analysis". (Optional: Download the plot and place in a Word document)



b) How do the estimates of sensitivity and specificity change compared to the analysis with all studies?

Excluding Mulligan and Narasimhalu results in little difference between the two estimates of sensitivity and specificity. Sensitivity is reduced from 0.908 (95% CI: 0.858, 0.942) to 0.904 (95% CI: 0.849, 0.940) and specificity is increased from 0.664 (95% CI: 0.563, 0.752) to 0.671 (95% CI: 0.562, 0.764).

6. The Prevalence page predicts how many patients in practice you would expect to have true positive, false positive, true negative and false negative results for a given disease prevalence based on the meta-analysis results and helps to give the results some clinical context. Use the 'Meta-analysis' tab on the 'Prevalence' page to identify the expected number of true positive, false positive, true negative and false negative results for a prevalence of 50% and how this would change if prevalence was increased to 70%.

50% prevalence: True positives = 454 (95% CI: 429, 471), False positives = 168 (95% CI: 124, 218), True negatives = 332 (95% CI: 282, 376), False negatives = 46 (95% CI: 29, 71)

70% prevalence: True positives = 636 (95% CI: 601, 659), False positives = 101 (95% CI: 74, 131), True negatives = 199 (95% CI: 169, 226), False negatives = 64 (95% CI: 41, 99)