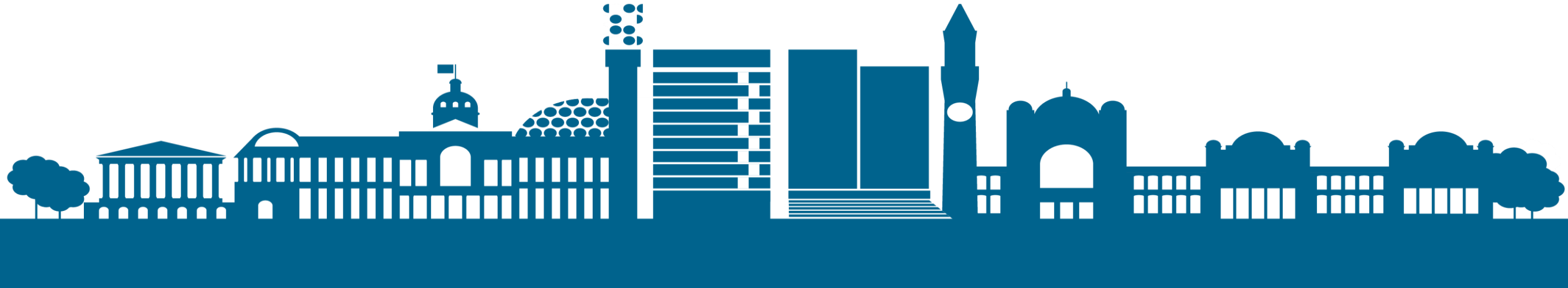
UNIVERSITY OF BIRMINGHAM | INSTITUTE OF APPLIED HEALTH RESEARCH

Comparing test accuracy: from pairwise to network meta-analysis of tests

Yemisi Takwoingi

Professor of Test Evaluation and Evidence Synthesis

Co-convenor Cochrane Screening & Diagnostic Tests Methods Group




TESTS

Outline

- Introduction/background
  - Comparative accuracy study designs
- Current practice for comparing diagnostic test accuracy (DTA)
  - Test comparison strategy
  - Meta-analysis methods
- Extensions to network meta-analysis of diagnostic test accuracy (DTA-NMA)

Scope of a DTA review

- Multiple objectives are possible
- 3 main types of analyses based on review question and objectives
    1) **What is the diagnostic accuracy of a test?**
    2) **How does the accuracy of two or more tests compare?**
    3) How does test accuracy vary with clinical and methodological characteristics?

(1) & (2) are typically primary objectives of a DTA review

Is the sensitivity and (or) specificity of the new test better than that of existing test(s)?

Index and comparator tests

- **Index test**: “new” test or test strategy we wish to evaluate
- **Comparator test**: existing test or diagnostic management strategy which may be standard practice
- We compare the accuracy of the index and the comparator tests
- The term “comparator test” can be confusing so simply put, we compare the accuracy of index tests
- **Reference standard**: the best available way to verify the presence or absence of the target condition.
  - May be a single test or a combination of tests and clinical information not routinely available in practice.

Study designs for comparing test accuracy

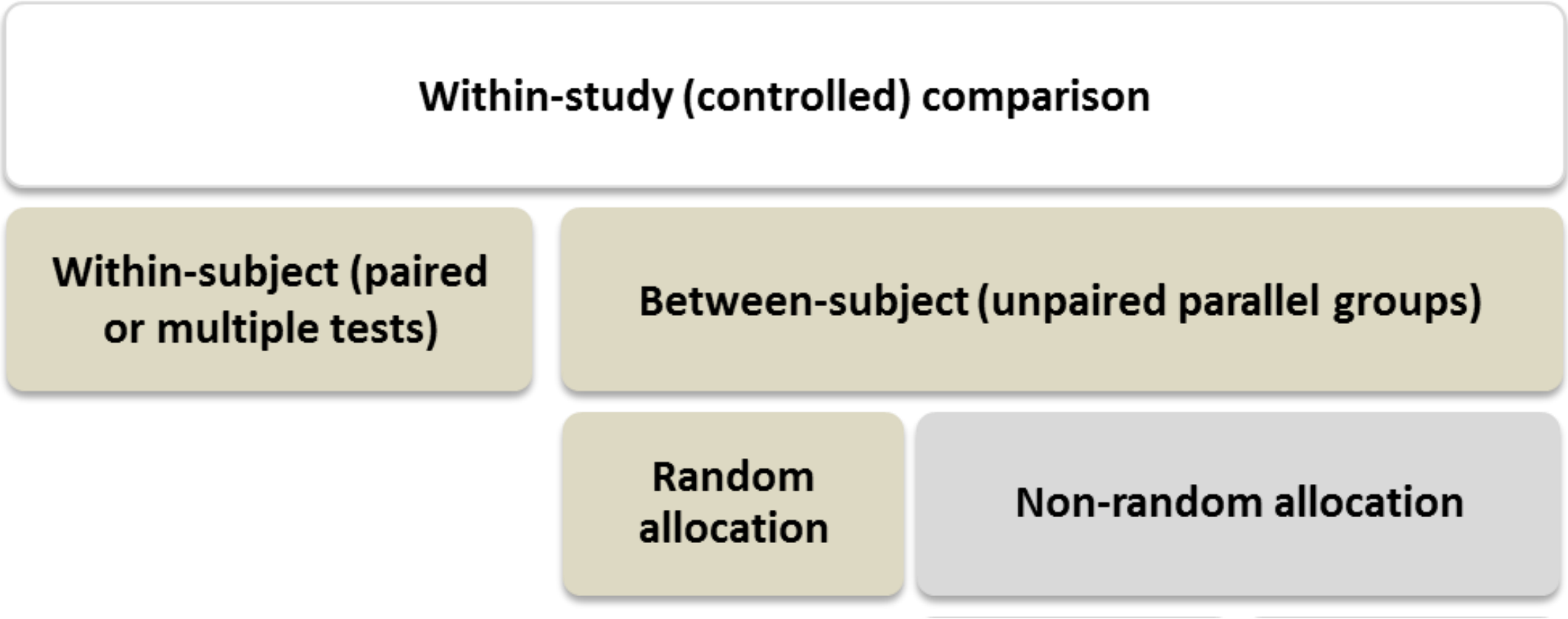
Within-study (controlled) comparison

Within-subject (paired or multiple tests)

Between-subject (unpaired parallel groups)

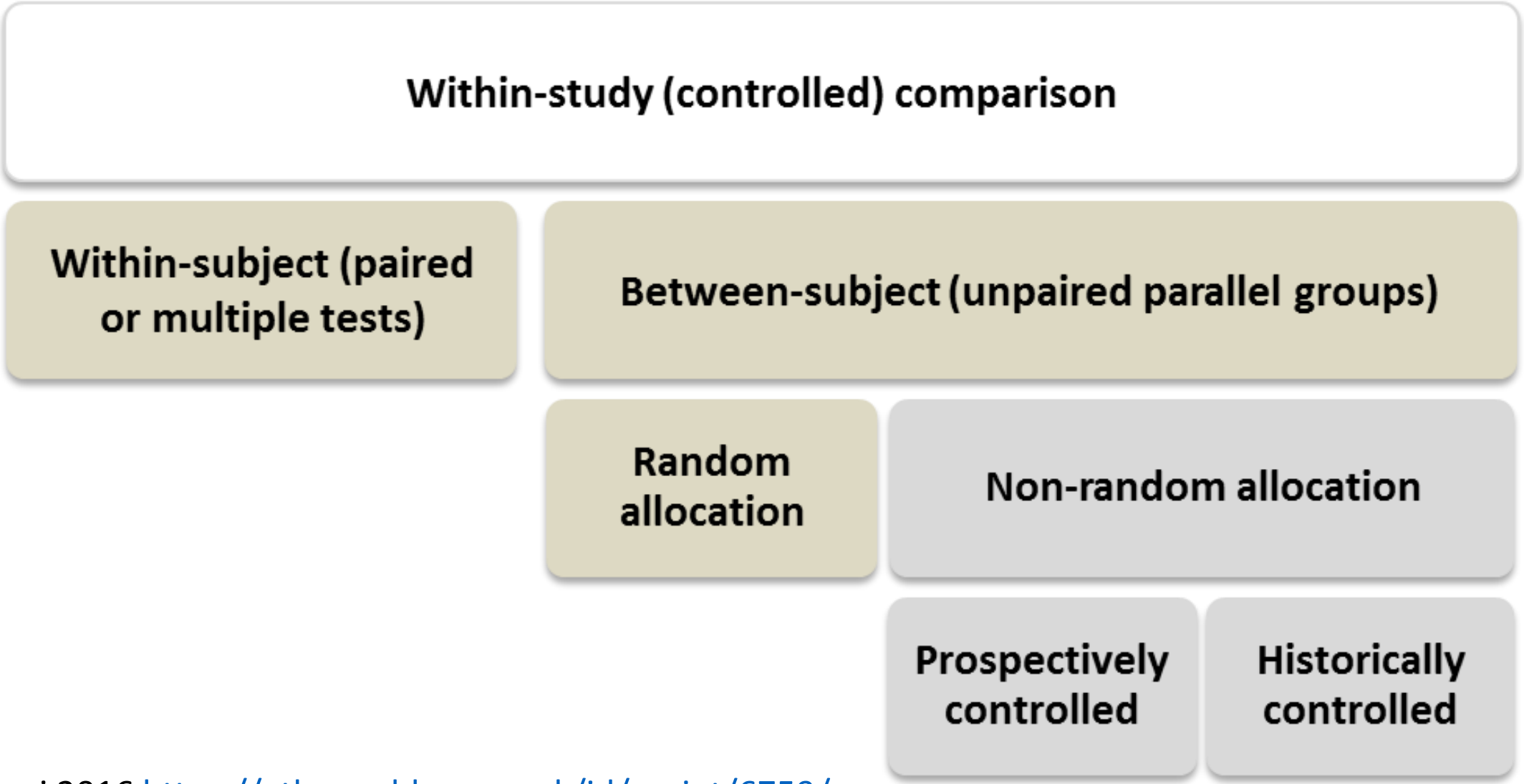
Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

Study designs for comparing test accuracy



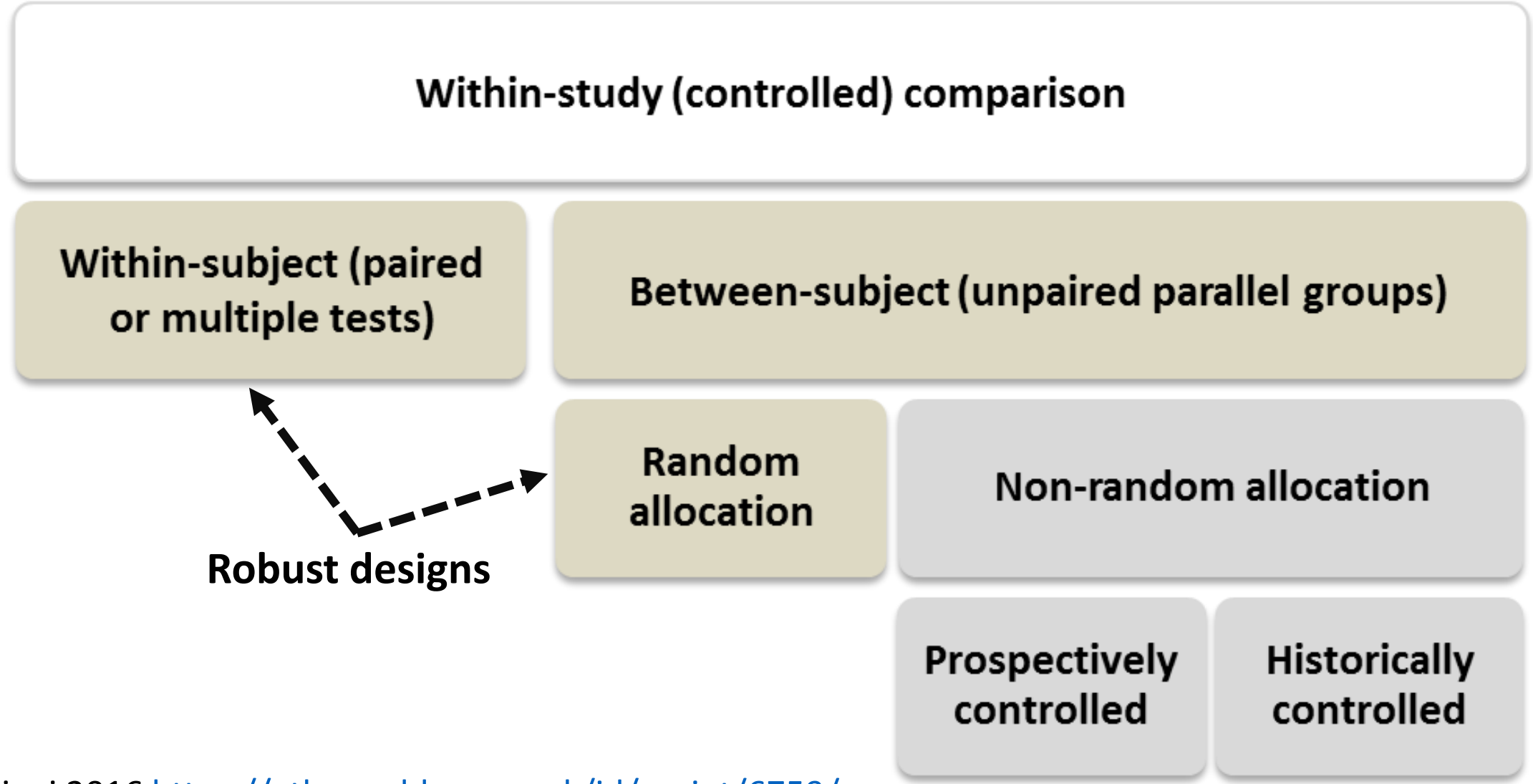
Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

Study designs for comparing test accuracy



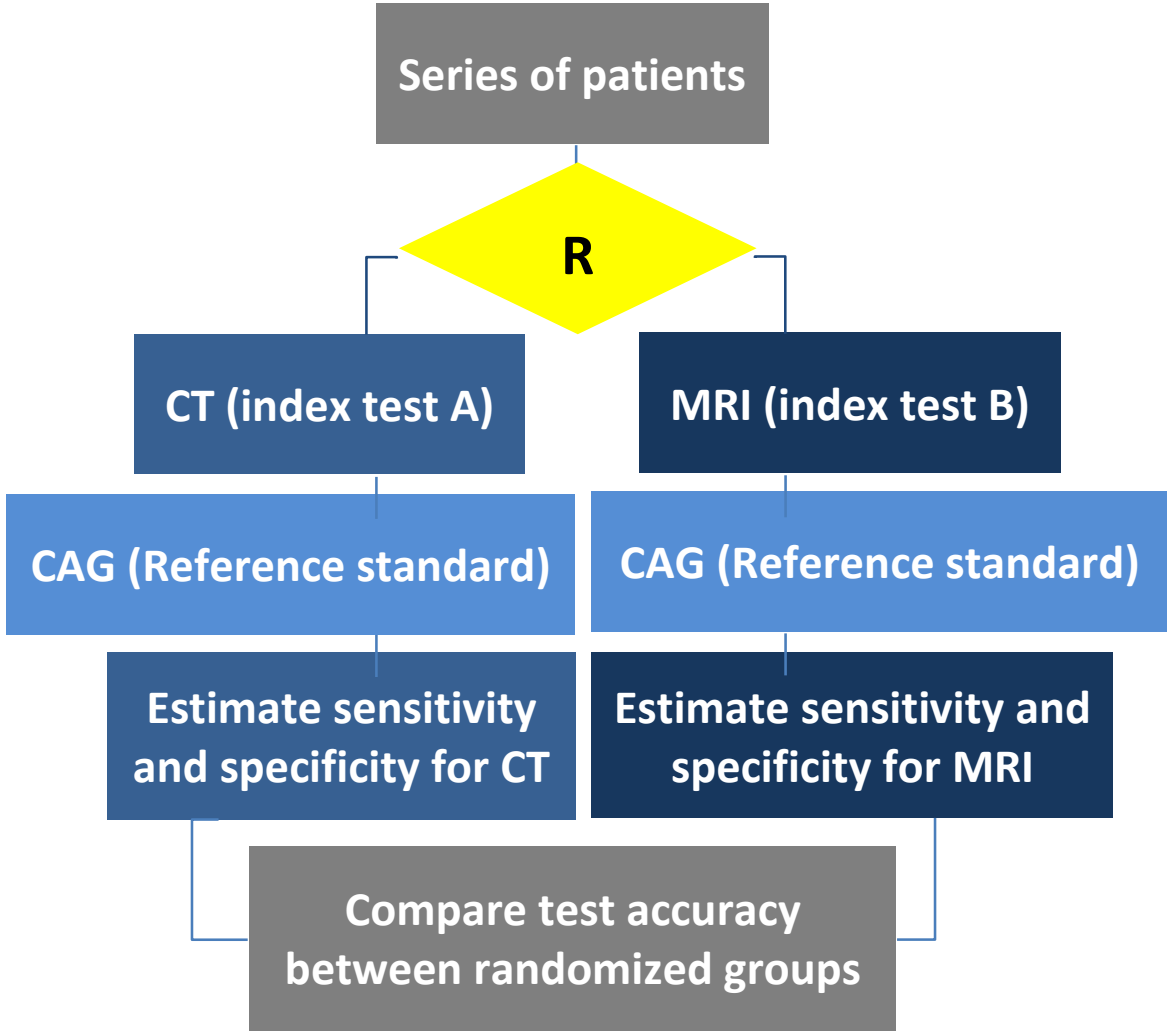
Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

Study designs for comparing test accuracy



Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

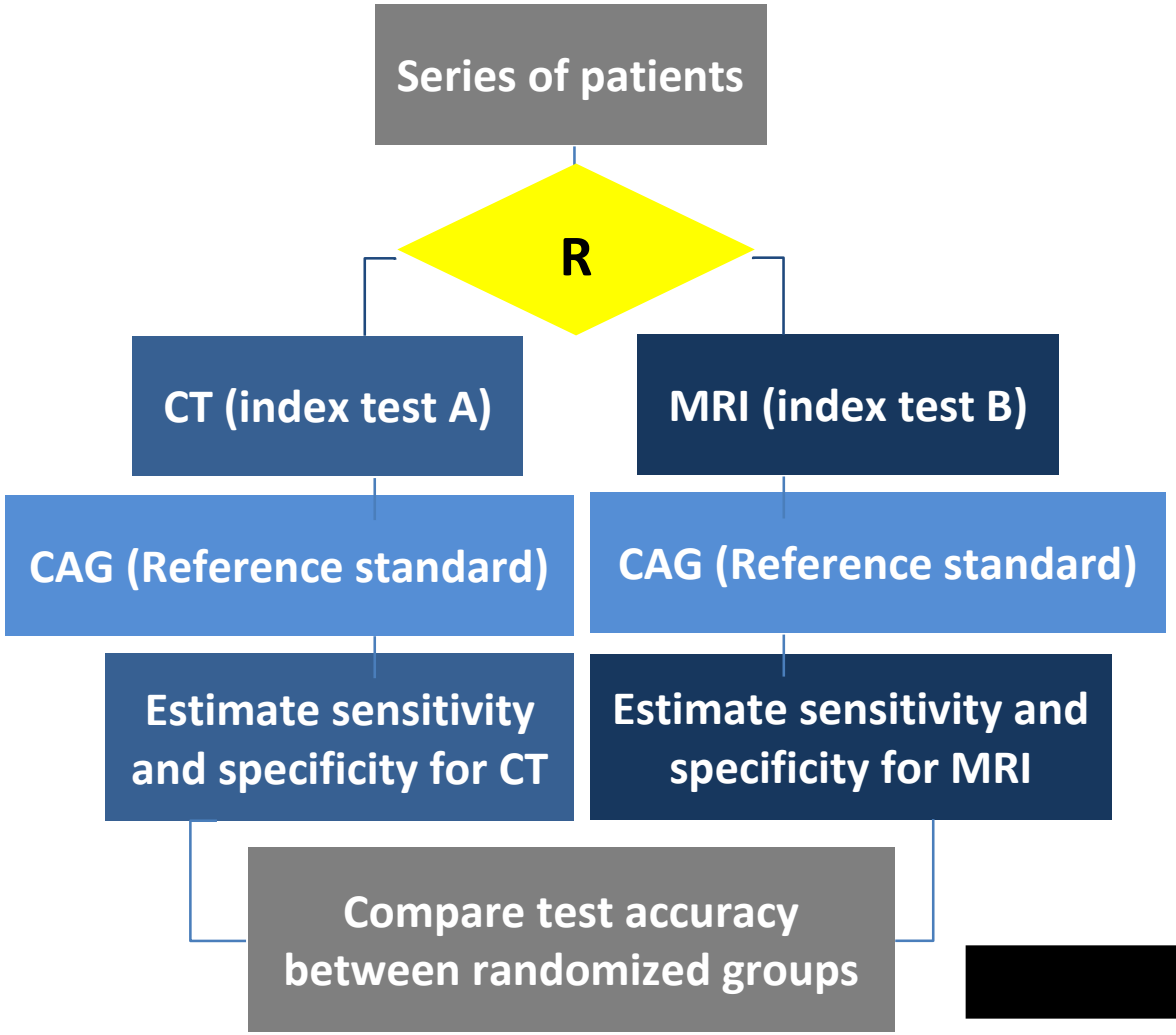
Robust test comparison designs



Unpaired (between-subject randomized) design

CAD = coronary artery disease
CAG = coronary angiography

Robust test comparison designs



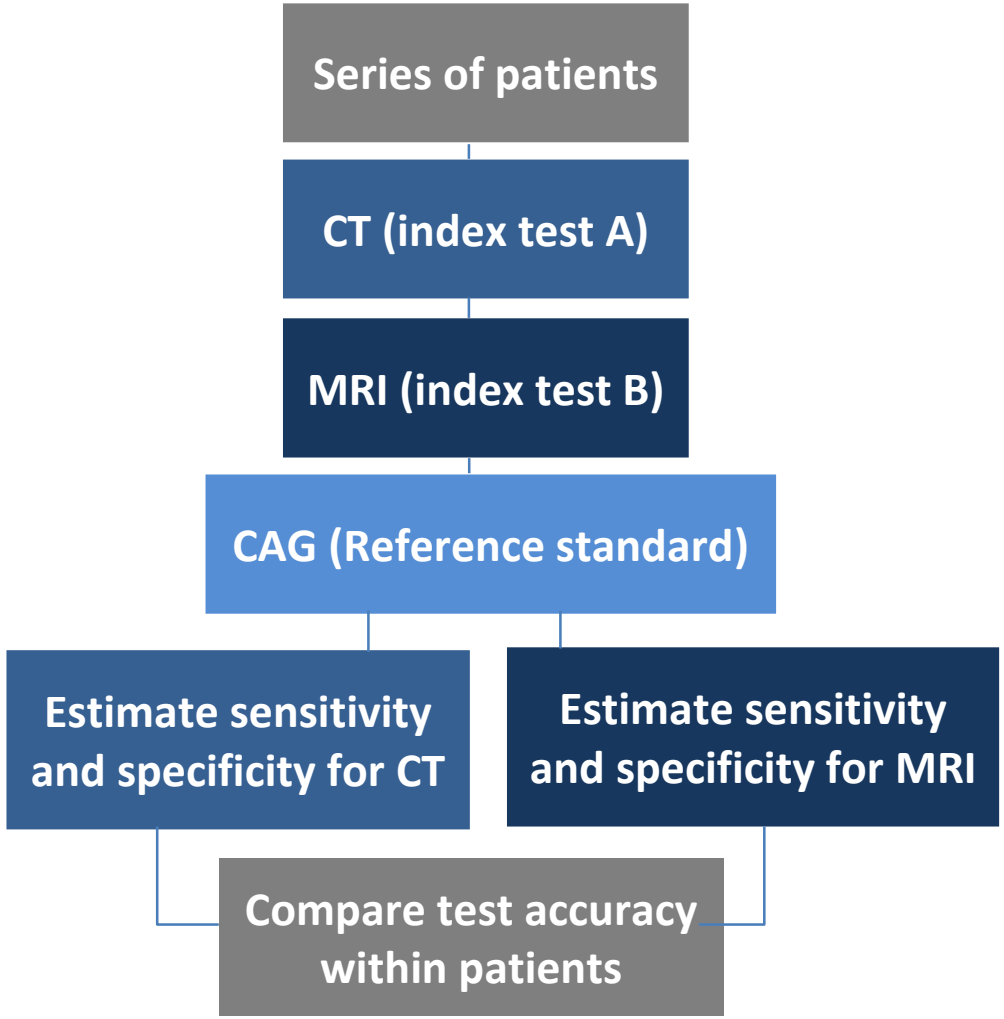
	CAD	No CAD
CT+	TP	FP
CT–	FN	TN

	CAD	No CAD
MRI+	TP	FP
MRI–	FN	TN

Unpaired (between-subject randomized) design

CAD = coronary artery disease
CAG = coronary angiography

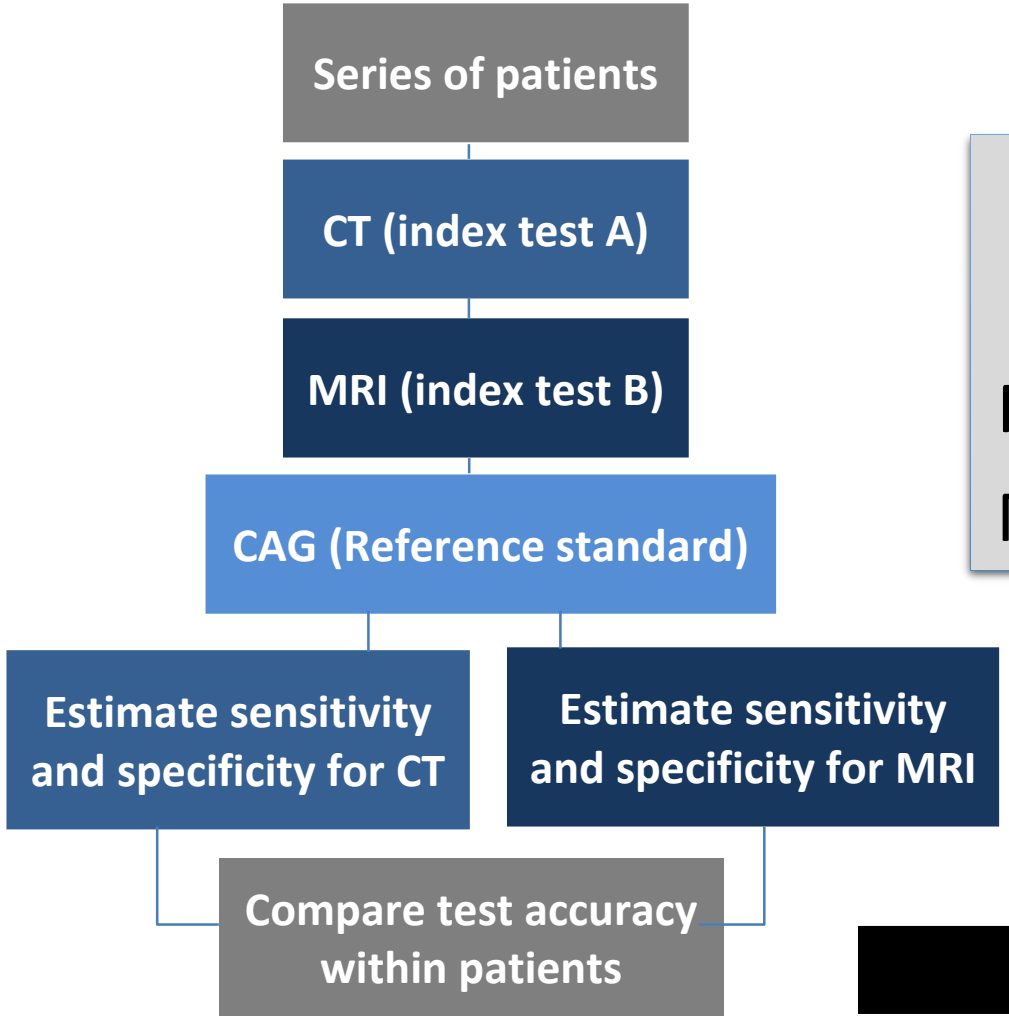
Robust test comparison designs



Paired (within-subject) design

CAD = coronary artery disease
CAG = coronary angiography

Robust test comparison designs



	CAD		No CAD	
	CT+	CT–	CT+	CT–
MRI+	a	b	e	f
MRI–	c	d	g	h

Paired (within-subject) design

CAD = coronary artery disease
CAG = coronary angiography

Joint classification table: an example

Objectives: To compare the diagnostic accuracy and cost-effectiveness of T-SPOT.*TB*® (Oxford Immunotec, Abingdon, UK) and QuantiFERON® TB GOLD In-Tube (Cellestis, Carnegie, VIC, Australia) for diagnosis of suspected active TB and to estimate the diagnostic accuracy of second-generation IGRAs.

Design: Prospective within-patient comparative diagnostic accuracy study.

TABLE 13 Cross-tabulation of T-SPOT.*TB* and QFT-GIT against final diagnosis[27]

		T-SPOT.*TB*, *n*											
		Active TB positive (categories 1 and 2)						Active TB negative (category 4)					
		Positive	Negative	Borderline	Indeterminate	Missing	Total	Positive	Negative	Borderline	Indeterminate	Missing	Total
QFT-GIT	Positive	187	13	6	9	5	220	37	30	3	3	1	74
	Negative	49	41	8	7	2	107	12	250	12	26	4	304
	Indeterminate	16	4	3	1	2	26	2	36	1	8	0	47
	Missing	1	0	0	0	9	10	0	3	0	0	11	14
	Total	253	58	17	17	18	363	51	319	16	37	16	439

Takwoingi Y, Whitworth H, Rees-Roberts M, Badhan A, Partlett C, Green N, et al. Interferon gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation. Health Technol Assess 2019;23(23).

Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy

Yemisi Takwoingi, DVM; Mariska M.G. Leeflang, PhD; and Jonathan J. Deeks, PhD

Background: Systematic reviews that "compare" the accuracy of 2 or more tests often include different sets of studies for each test.

Purpose: To investigate the availability of direct comparative studies of test accuracy and to assess whether summary estimates of accuracy differ between meta-analyses of noncomparative and comparative studies.

Data Sources: Systematic reviews in any language from the Database of Abstracts of Reviews of Effects and the Cochrane Database of Systematic Reviews from 1994 to October 2012.

Study Selection: 1 of 2 assessors selected reviews that evaluated at least 2 tests and identified meta-analyses that included both noncomparative studies and comparative studies.

Data Extraction: 1 of 3 assessors extracted data about review and study characteristics and test performance.

Data Synthesis: 248 reviews compared test accuracy; of the 6915 studies, 2113 (31%) were comparative. Thirty-six reviews (with 52 meta-analyses) had adequate studies to compare results of noncomparative and comparative studies by using a hierarchical summary receiver-operating characteristic meta-regression model for each test comparison. In 10 meta-analyses, noncomparative studies ranked tests in the opposite order of comparative studies. A total of 25 meta-analyses showed more than a 2-fold discrepancy in the relative diagnostic odds ratio between noncomparative and comparative studies. Differences in accuracy estimates between noncomparative and comparative studies were greater than expected by chance ($P < 0.001$).

Limitation: A paucity of comparative studies limited exploration of direction in bias.

Conclusion: Evidence derived from noncomparative studies often differs from that derived from comparative studies. Robustly designed studies in which all patients receive all tests or are randomly assigned to receive one or other of the tests should be more routinely undertaken and are preferred for evidence to guide test selection.

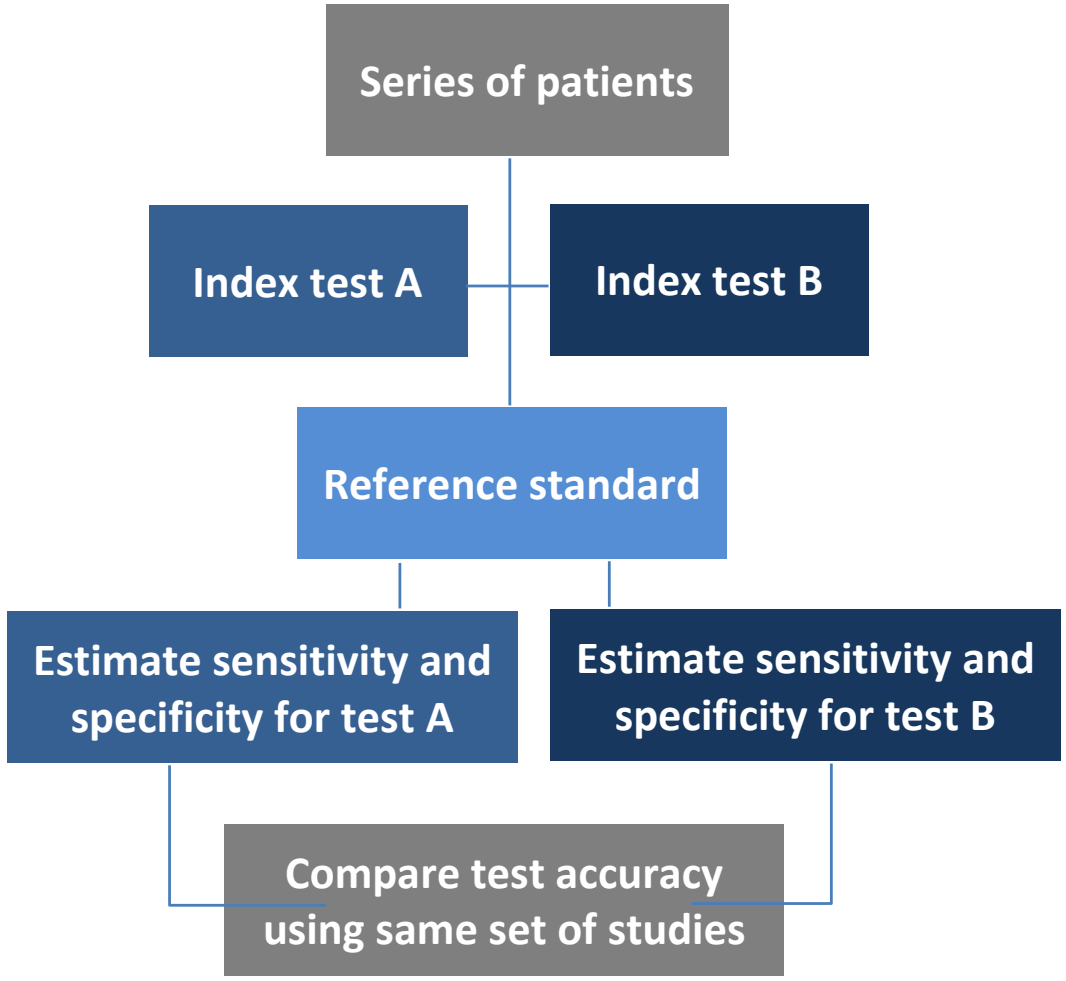
Primary Funding Source: National Institute for Health Research (United Kingdom).

Ann Intern Med. 2013;158:544-554.
For author affiliations, see end of text.

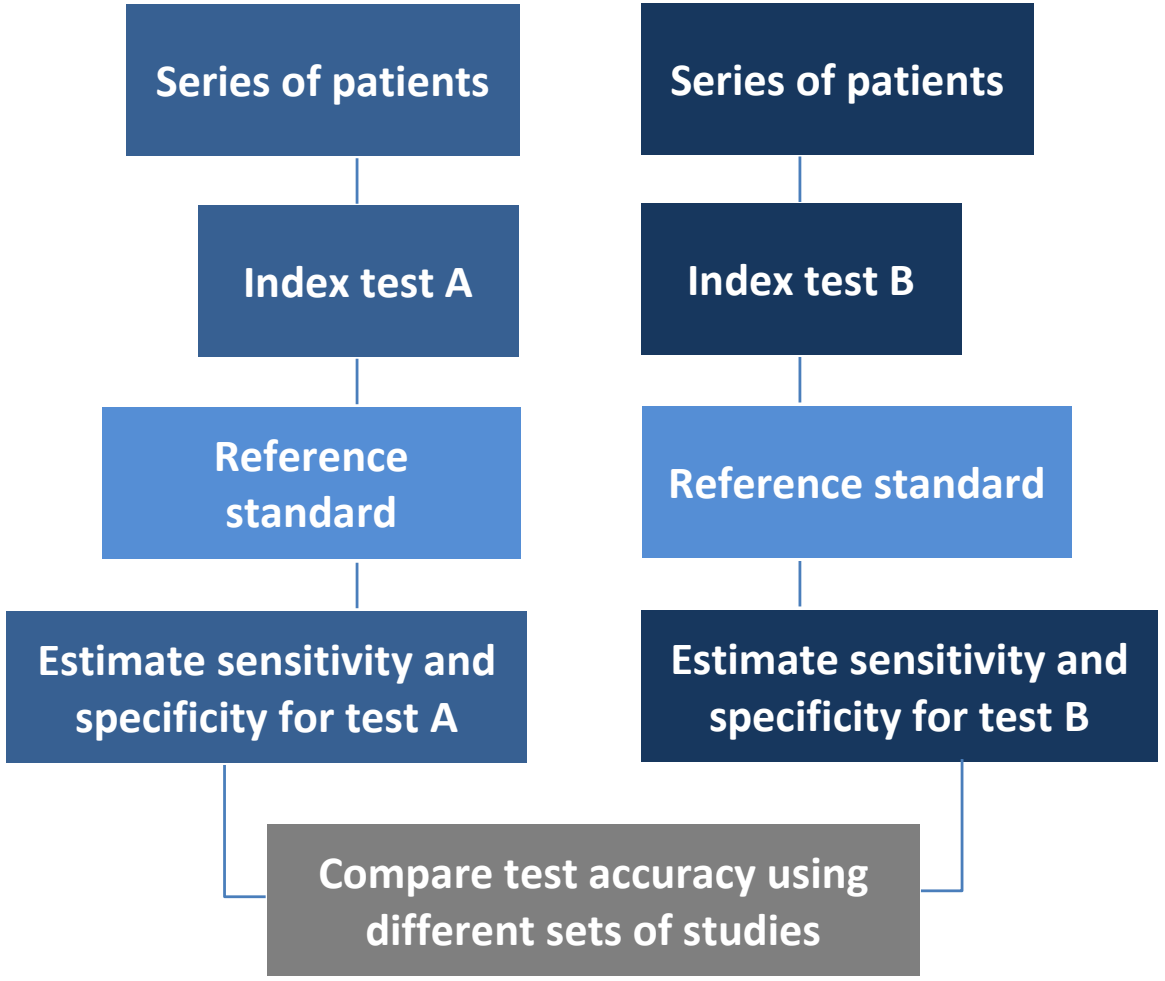
What is the test comparison strategy in the comparative DTA review?

Test comparison strategy

Direct (head-to-head) comparison



Indirect (between-study) comparison



Cochrane DTA Review examples

Cochrane Library
Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

[Diagnostic Test Accuracy Review]

Xpert MTB/RIF and Xpert MTB/RIF Ultra assays for active tuberculosis and rifampicin resistance in children

Alexander W Kay[1], Lucia González Fernández[2], Yemisi Takwoingi[3], Michael Eisenhut[4], Anne K Detjen[5], Karen R Steingart[6a], Anna M Mandalakas[1b]

Rapid diagnostic tests for diagnosing uncomplicated *P. falciparum* malaria in endemic countries

Katharine Abba[1], Jonathan J Deeks[2], Piero L Olliaro[3], Cho-Min Naing[4], Sally M Jackson[1], Yemisi Takwoingi[2], Sarah Donegan[1], Paul Garner[1]

First trimester serum tests for Down's syndrome screening

S Kate Alldred[1], Yemisi Takwoingi[2], Boliang Guo[3], Mary Pennant[4], Jonathan J Deeks[2], James P Neilson[1], Zarko Alfirevic[1]

Example 1: Xpert MTB/RIF and Xpert MTB/RIF Ultra assays for active tuberculosis and rifampicin resistance in children

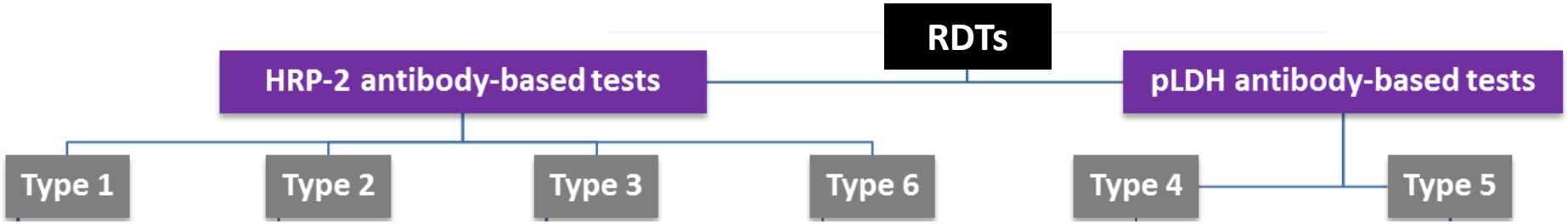
Direct comparison of Xpert MTB/RIF and Xpert Ultra (3 studies)

Kay AW et al. 2020. CD013359

Example 2: Rapid diagnostic tests for *P. falciparum* malaria

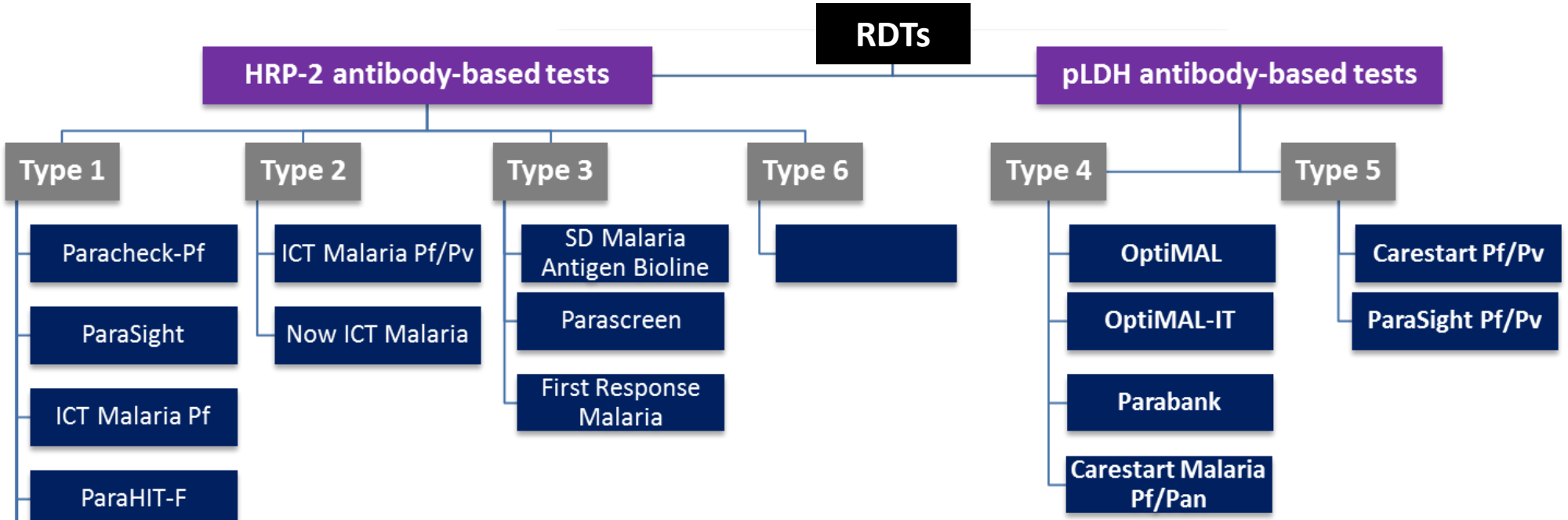
Abba K et al. 2011. CD008122

Example 2: Rapid diagnostic tests for *P. falciparum* malaria



Abba K et al. 2011. CD008122

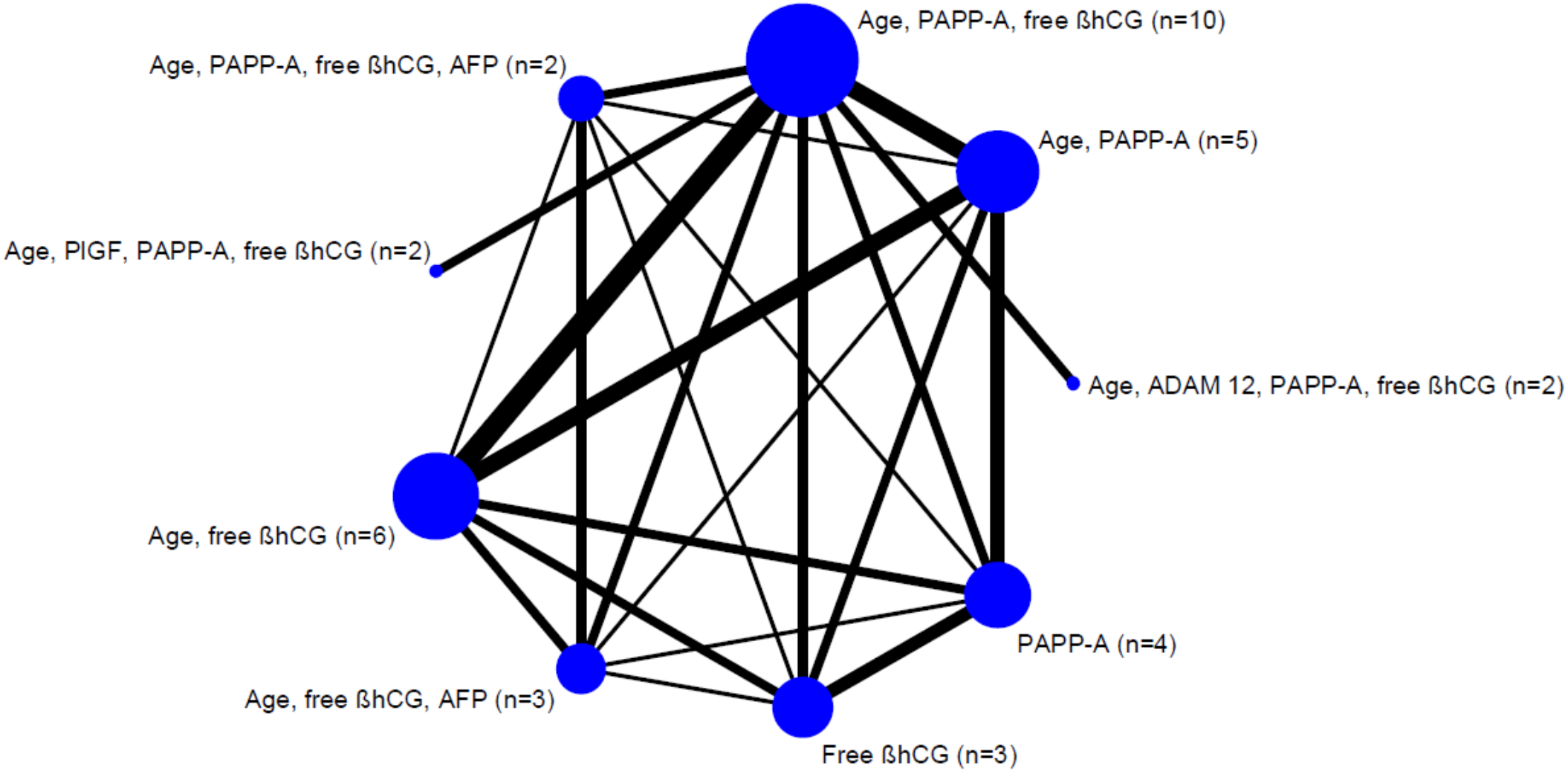
Example 2: Rapid diagnostic tests for *P. falciparum* malaria



- 6 RDT types within 2 groups of antibody-based tests
- Type 1: 10 brands
- Type 2: 2 brands
- Type 3: 3 brands
- Type 4: 4 brands
- Type 5: 2 brands
- Type 6: none
- Total of 21 RDT brands included in 74 studies

Abba K et al. 2011. CD008122

Example 3: First trimester serum test strategies for Down's syndrome screening



Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

Alldred SK et al. 2015. CD011975

ELSEVIER

Journal of Clinical Epidemiology 121 (2020) 1−14

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Methods and reporting of systematic reviews of comparative accuracy were deficient: a methodological survey and proposed guidance

Yemisi Takwoingi[a,b,*], Christopher Partlett[c], Richard D. Riley[d], Chris Hyde[e], Jonathan J. Deeks[a,b]

[a]*Test Evaluation Research Group, University of Birmingham, Birmingham, UK*
[b]*NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK*
[c]*Nottingham Clinical Trials Unit, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, UK*
[d]*Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK*
[e]*Exeter Test Group, College of Medicine and Health, University of Exeter, Exeter, UK*

Accepted 11 December 2019; Published online 14 December 2019

Abstract

Objective: The objective of this study was to examine methodological and reporting characteristics of systematic reviews and meta-analyses which compare diagnostic test accuracy (DTA) of multiple index tests, identify good practice, and develop guidance for better reporting.

POLL

How can I statistically combine the studies to compare test accuracy?

Key challenges for DTA meta-analysis

- Two summary statistics for each study
  - sensitivity and specificity and each have different implications
- Threshold effects induce correlations between sensitivity and specificity and often seem to be present
  - thresholds can vary between studies
  - same threshold can imply different sensitivities and specificities in different groups
- Heterogeneity is the norm
  - substantial variation in sensitivity and specificity are observed in most reviews

Additional key challenges for comparative DTA meta-analysis

- Many DTA studies are not comparative
- Different study designs
  - Correlated data
  - Availability of fully cross-classified data

Meta-analysis methods for comparing test accuracy

(up to July 2014)

	Reference	Method	Test accuracy measure
1	Moses et al 1993; Littenberg and Moses 1993	Comparison of Q*	Q*
2	Hasselblad and Hedges 1995	Standardized distance between the means of two populations	Effectiveness measure (*d*) proportional to log DOR
3	**Rutter and Gatsonis 2001**	**HSROC meta-regression**	Diagnostic odds ratio (DOR)
4	Kowalski et al 2001	Generalized estimating equation	Sensitivity and specificity
5	Lijmer et al 2002	Moses SROC meta-regression	DOR
6	Worster et al 2002	General linear mixed model	Likelihood ratios
7	Suzuki et al 2004	Conditional relative odds ratio	DOR
8	Siadaty and Shu 2004	Proportional odds ratio	DOR
9	Siadaty et al 2004	Repeated measures modelling	DOR
10	**Reitsma et al 2005; Hamza et al 2009**	**Bivariate meta-regression**	Sensitivity and specificity
11	**Cheng et al 2013‡**	**Network meta-analysis**	Sensitivity and specificity
12	**Verde 2013‡**	**Bivariate meta-analysis of paired data**	Sensitivity and specificity
13	**Trikalinos et al 2014**	**Bivariate meta-analysis of paired data**	Sensitivity and specificity

‡Conference presentation

Takwoingi 2016 https://etheses.bham.ac.uk/id/eprint/6759/

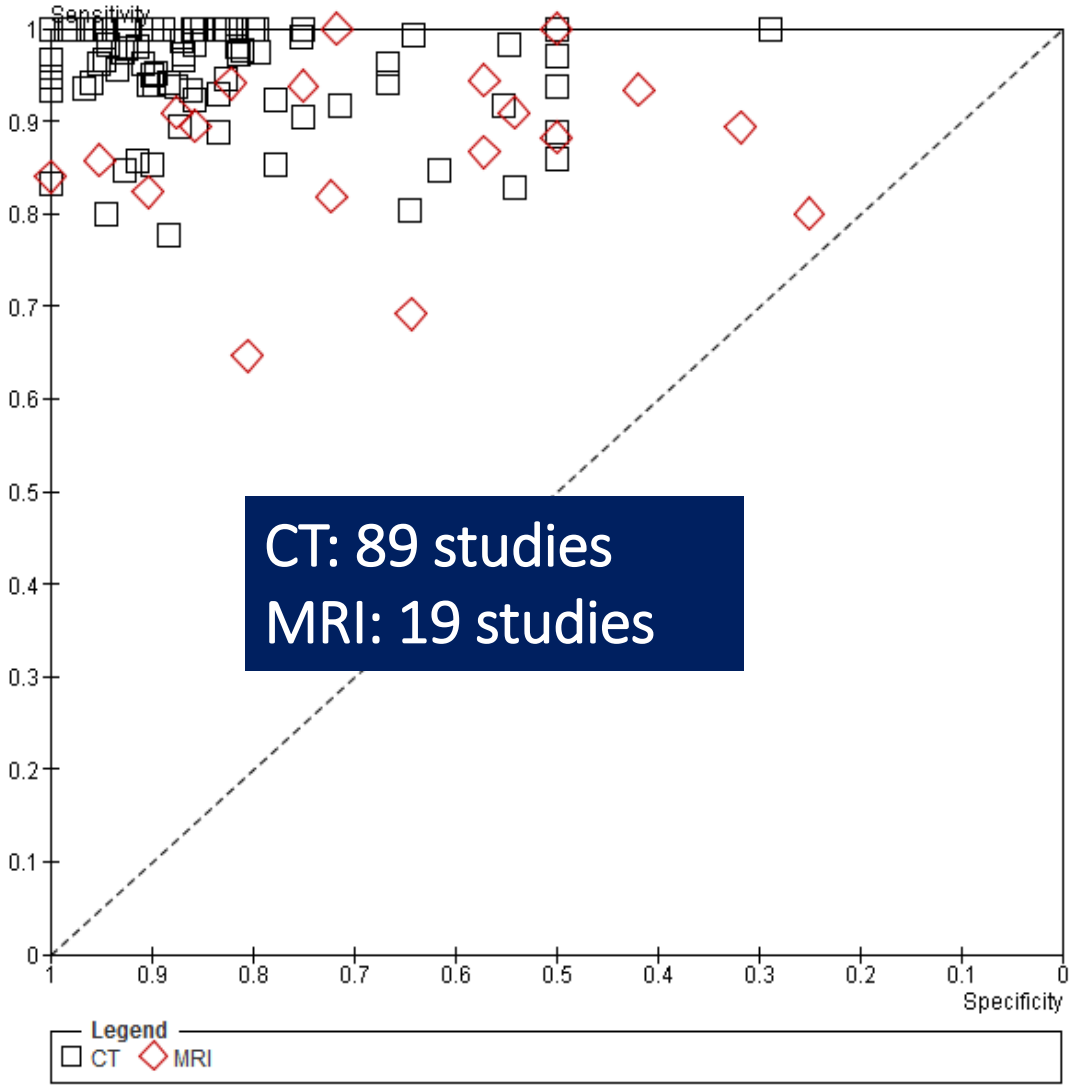
Hierarchical meta-regression

- Hierarchical models can incorporate a study-level covariate to compare test accuracy
- Different questions can be addressed
  - Bivariate model
    - differences in summary points of sensitivity and/or specificity
  - HSROC model
    - differences in overall accuracy
    - differences in threshold
    - differences in shape of SROC curve

Macaskill P et al. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration; 2010. https://methods.cochrane.org/sdt/handbook-dta-reviews

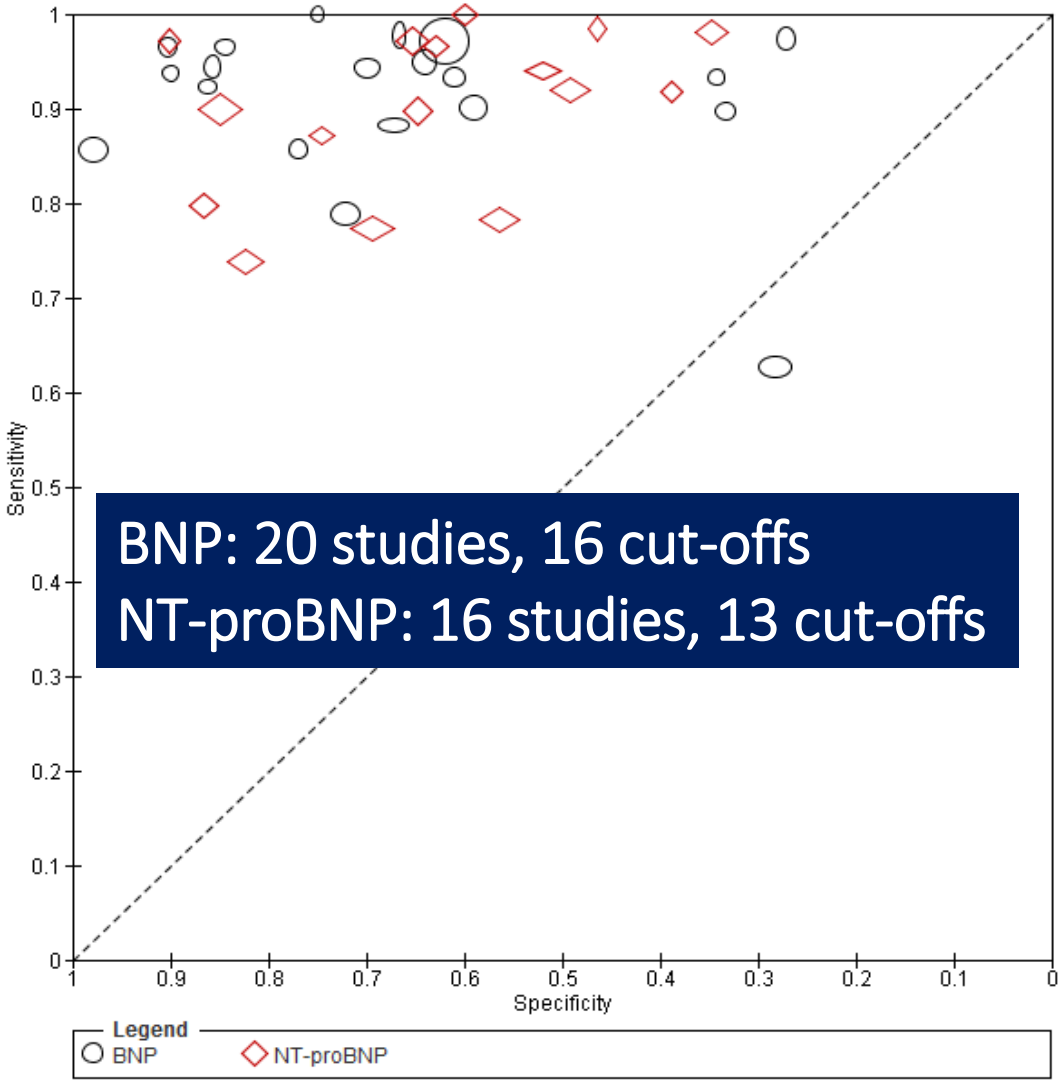
Comparing test accuracy

CT vs MRI for CAD



CT: 89 studies
MRI: 19 studies

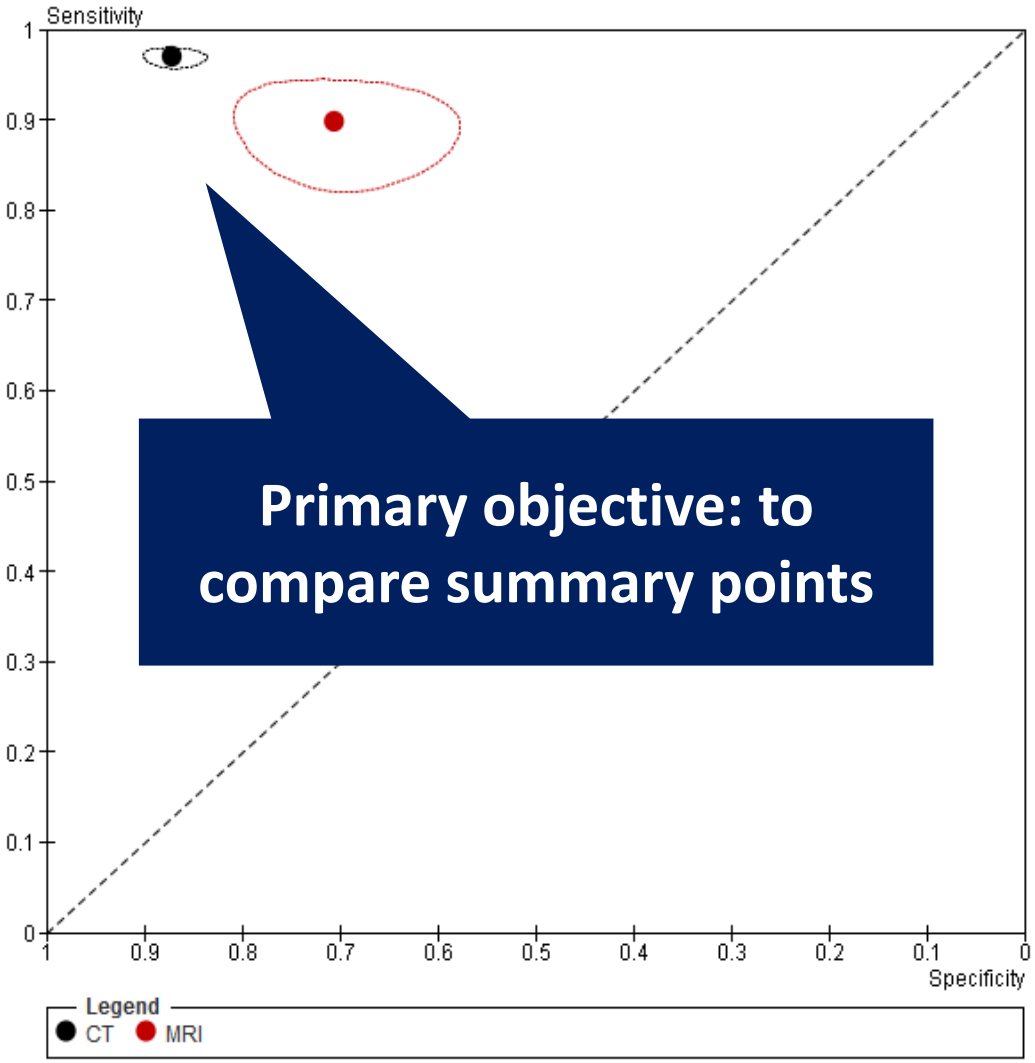
BNP vs NT-proBNP for heart failure



BNP: 20 studies, 16 cut-offs
NT-proBNP: 16 studies, 13 cut-offs

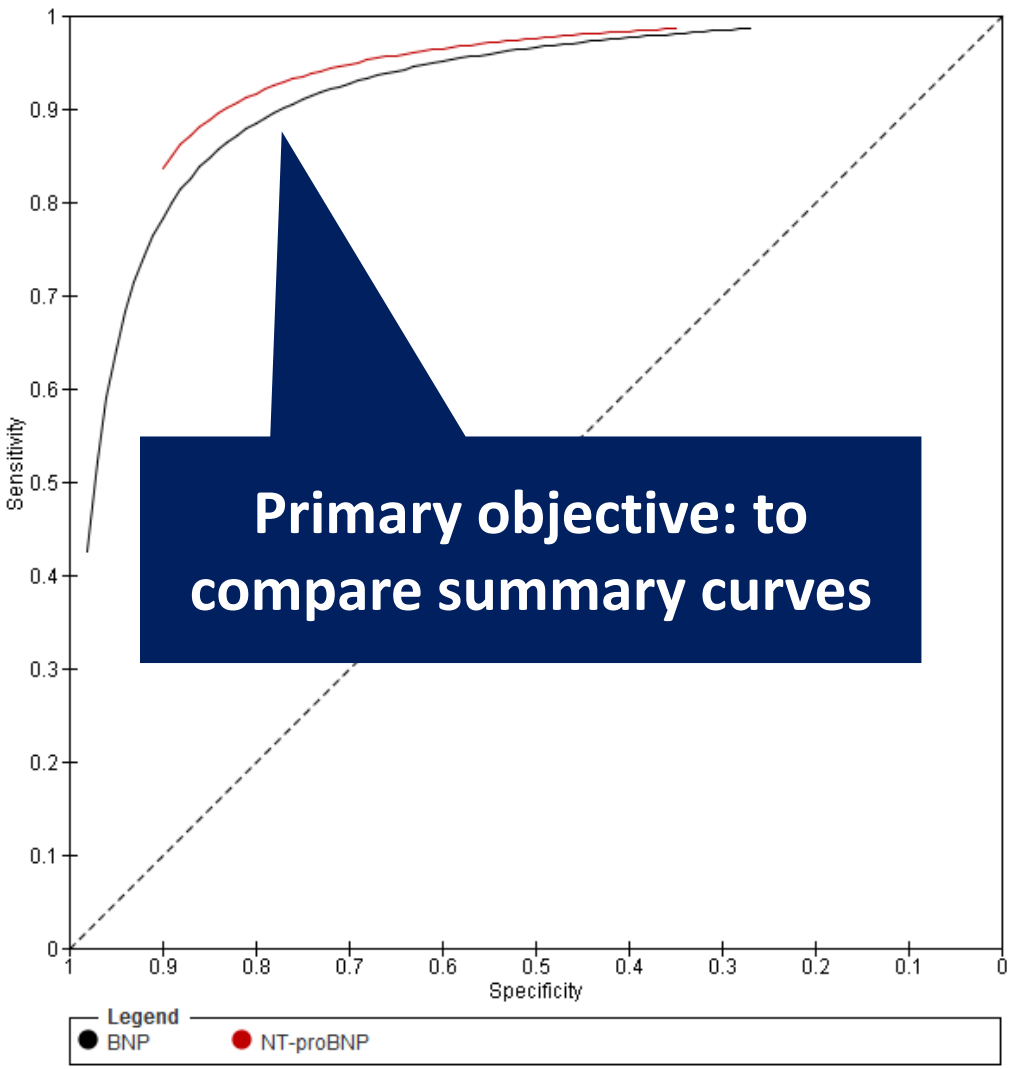
Hierarchical meta-regression models

Bivariate model



Primary objective: to compare summary points

HSROC model



Primary objective: to compare summary curves

Let’s get technical…

Bivariate model specification

Models the proportion in each study (*i*) that have correct test results in diseased and non-diseased groups

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \sum\right) \text{ with } \sum = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

μ_A is the mean logit sensitivity

μ_B is the mean logit specificity

σ_A^2 is the variance of the logit sensitivity

σ_B^2 is the variance of the logit specificity

σ_{AB} is the covariance of logit sensitivity and logit specificity

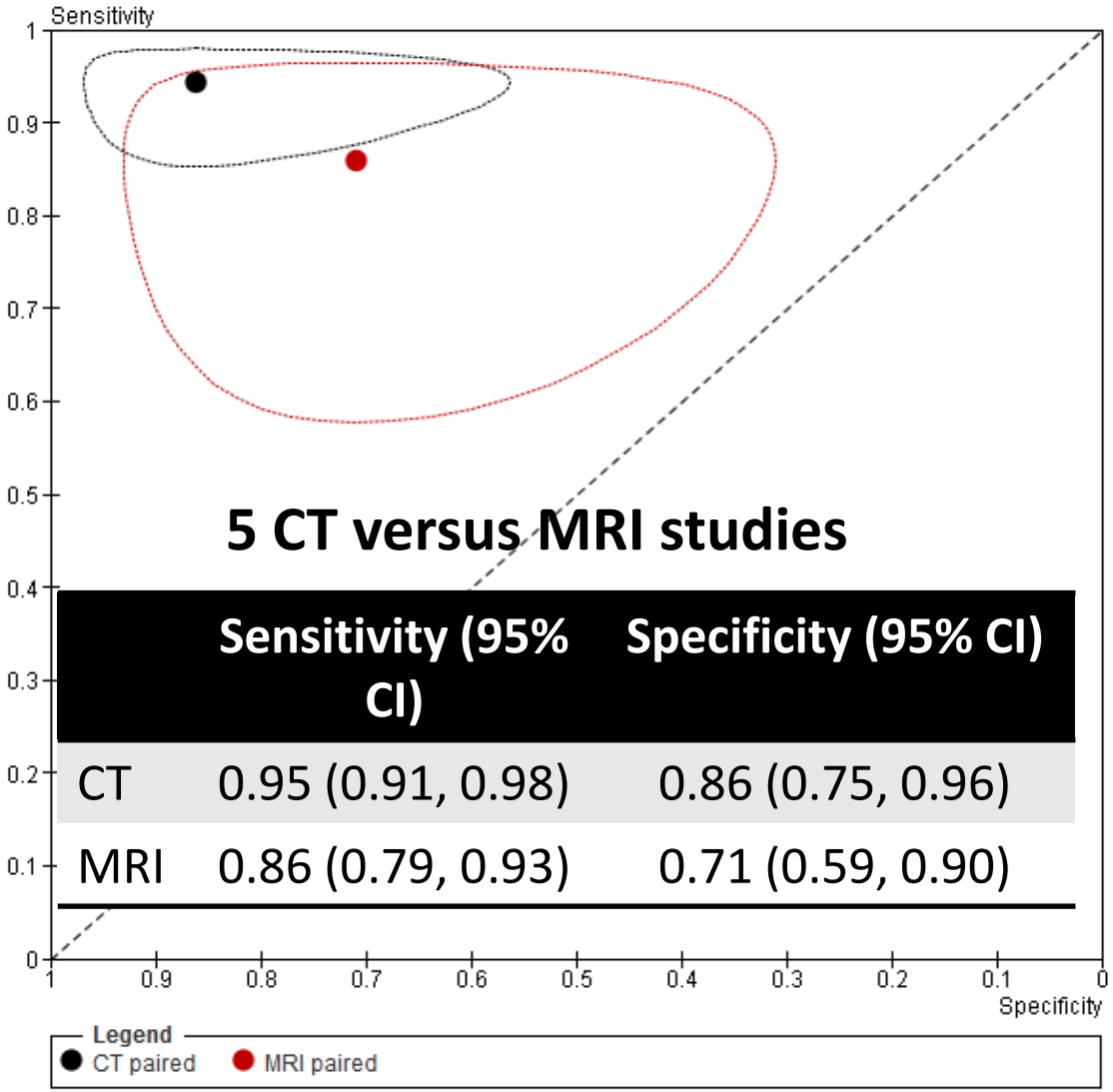
Bivariate model with a covariate

Assuming a test type covariate *t* that may affect both sensitivity and specificity, the model can be extended as follows:

$$\begin{pmatrix} \mu_{Aik} \\ \mu_{Bik} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A + \nu_A t_k \\ \mu_B + \nu_B t_k \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \right)$$

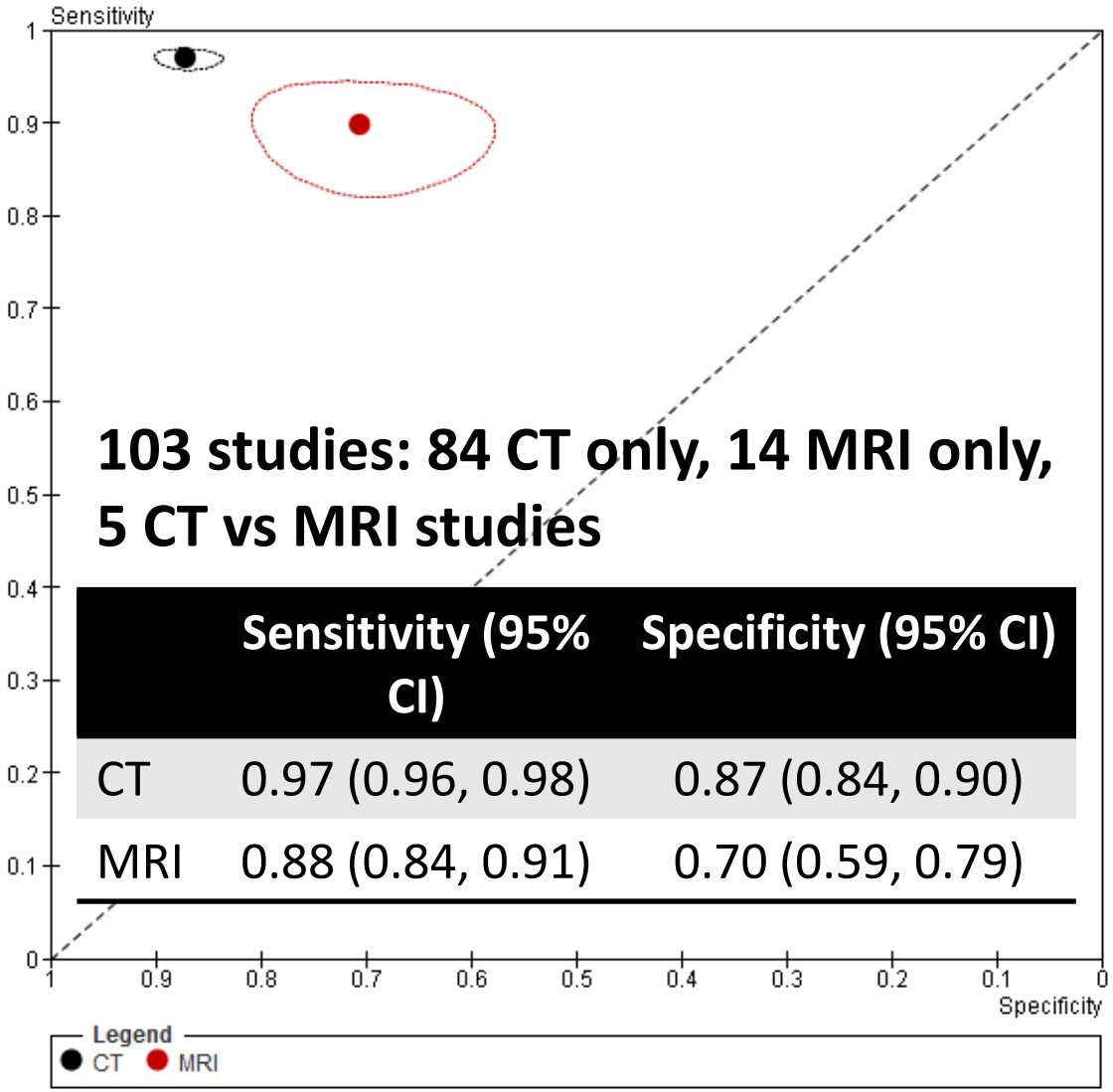
Effect of test type on variance parameters can also be investigated

CT versus MRI for CAD example: direct and indirect comparisons



5 CT versus MRI studies

	Sensitivity (95% CI)	Specificity (95% CI)
CT	0.95 (0.91, 0.98)	0.86 (0.75, 0.96)
MRI	0.86 (0.79, 0.93)	0.71 (0.59, 0.90)

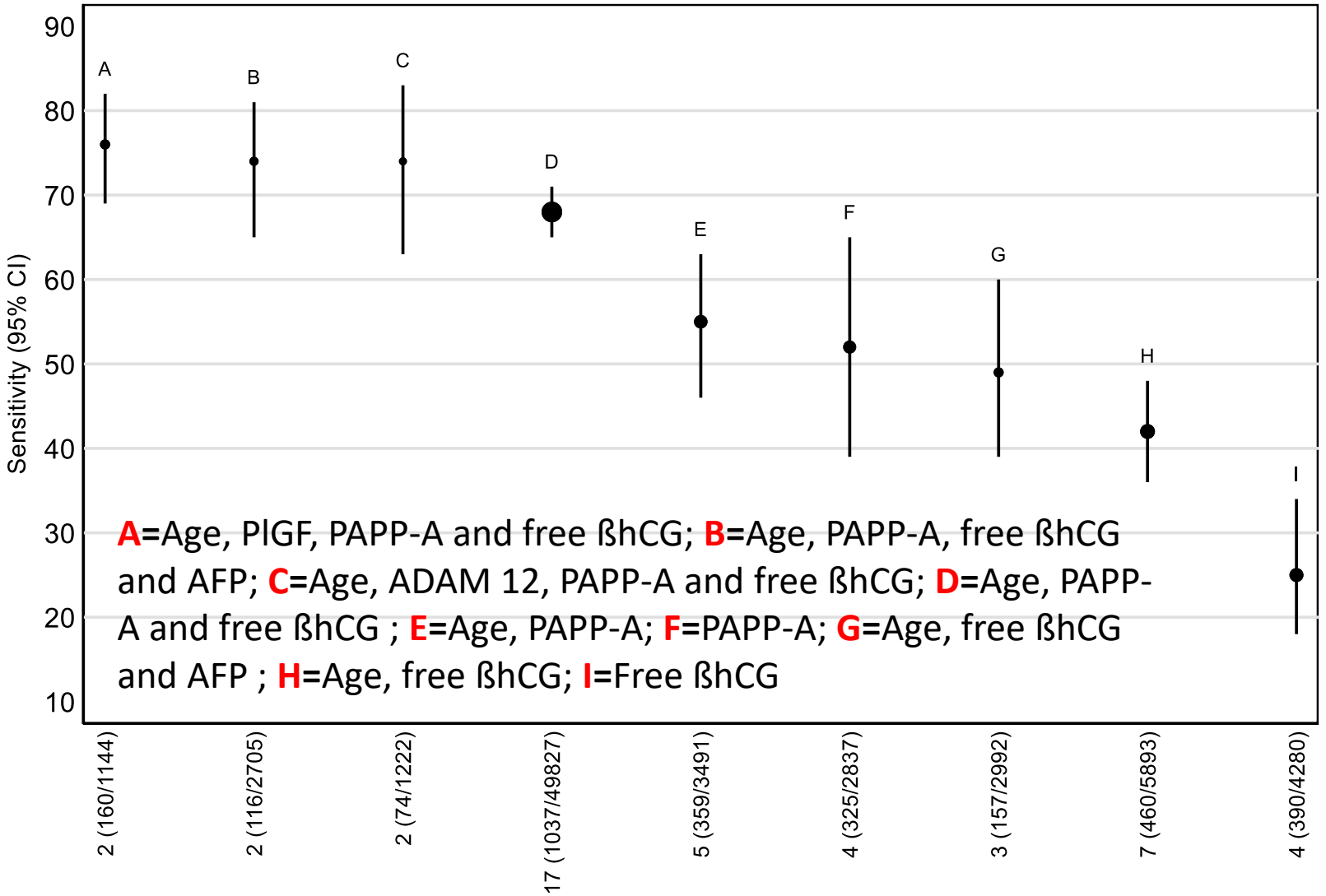


103 studies: 84 CT only, 14 MRI only, 5 CT vs MRI studies

	Sensitivity (95% CI)	Specificity (95% CI)
CT	0.97 (0.96, 0.98)	0.87 (0.84, 0.90)
MRI	0.88 (0.84, 0.91)	0.70 (0.59, 0.79)

Meta-regression not limited to pairwise comparisons

Sensitivity at a 5% false positive rate for 9 first trimester serum test strategies for Down's syndrome screening



Each circle represents the summary sensitivity for a test strategy and the size of each circle is proportional to the number of Down's cases.

The test strategies are ordered according to decreasing sensitivity. The number of studies, cases and women included for each test strategy are shown on the horizontal axis.

Alldred SK et al. 2015. CD011975

HSROC model specification

The model takes the form

$$\text{logit}(\pi_{ij}) = \left(\theta_i + \alpha_i \, dis_{ij}\right)\exp\left(-\beta \, dis_{ij}\right)$$

threshold

i.e. proportion test positive

(random effect)

accuracy

(random effect)

dependence of accuracy on threshold

i.e. shape of the summary curve

(fixed effect)

HSROC model with a covariate

- Assuming a test type covariate *Z* that may affect accuracy, threshold and shape, the model can be extended as:

$$\text{logit}(\pi_{ij}) = \left((\theta_i + \gamma Z_i) + (\alpha_i + \lambda Z_i) dis_{ij}\right) \exp\left(-(\beta + \delta Z_i)\, dis_{ij}\right)$$

- Shape parameter is estimated as β for one test and β+δ for the other test

- If δ = 0 is assumed and covariate terms are removed for shape, SROC curves for the tests will have the same shape (β)

$$\text{logit}(\pi_{ii}) = \left((\theta_i + \gamma Z_i) + (\alpha_i + \lambda Z_i) dis_{ii}\right) \exp\left(-\beta dis_{ii}\right)$$

  - Relative diagnostic accuracy of the two curves can be summarized using the relative DOR = exp(λ)

A ‘non-technical’ summary of the methods

Meta-analysis of diagnostic accuracy studies in mental health

Yemisi Takwoingi,[1] Richard D Riley,[2] Jonathan J Deeks[1]

[1]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK; [2]Research Institute for Primary Care and Health Sciences, Keele University, Staffordshire, UK
Correspondence to Dr Yemisi Takwoingi, y.takwoingi@bham.ac.uk

ABSTRACT
Objectives To explain methods for data synthesis of evidence from diagnostic test accuracy (DTA) studies, and to illustrate different types of analyses that may be performed in a DTA systematic review.
Methods We described properties of meta-analytic methods for quantitative synthesis of evidence. We used a DTA review comparing the accuracy of three screening questionnaires for bipolar disorder to illustrate application of the methods for each type of analysis.
Results The discriminatory ability of a test is commonly expressed in terms of sensitivity (proportion of those with the condition who test positive) and specificity (proportion of those without the condition who test negative). There is a trade-off between sensitivity and specificity, as an increasing threshold for defining test positivity will decrease sensitivity and increase specificity. Methods recommended for meta-analysis of DTA studies –such as the bivariate or hierarchical summary receiver operating characteristic (HSROC) model –jointly summarise sensitivity and specificity while taking into account this threshold effect, as well as allowing for between study differences in test performance beyond what would be expected by chance. The bivariate model focuses on estimation of a summary sensitivity and specificity at a common threshold while the HSROC model focuses on the estimation of a summary curve from studies that have used different thresholds.
Conclusions Meta-analyses of diagnostic accuracy studies can provide answers to important clinical questions. We hope this article will provide clinicians with sufficient understanding of the terminology and methods to aid interpretation of systematic reviews and facilitate better patient care.

Evidence-Based Mental Health Online First, published on October 7, 2015 as 10.1136/eb-2015-102228

Software for meta-analysis of DTA studies

Resources for authors

- DTA Handbook
- Software for meta-analysis

Researchers have prepared macros or modules for statistical models for meta-analysis of data from diagnostic test accuracy studies for several statistical analysis software programs. As these become available we will add them to this page. Currently, there is a macro avaibale for SAS and a package for STATA.

SAS

MetaDAS: A SAS macro for meta-analysis of diagnostic accuracy studies, contains both the bivariate and the HSROC model. Please find the required documents hereunder:

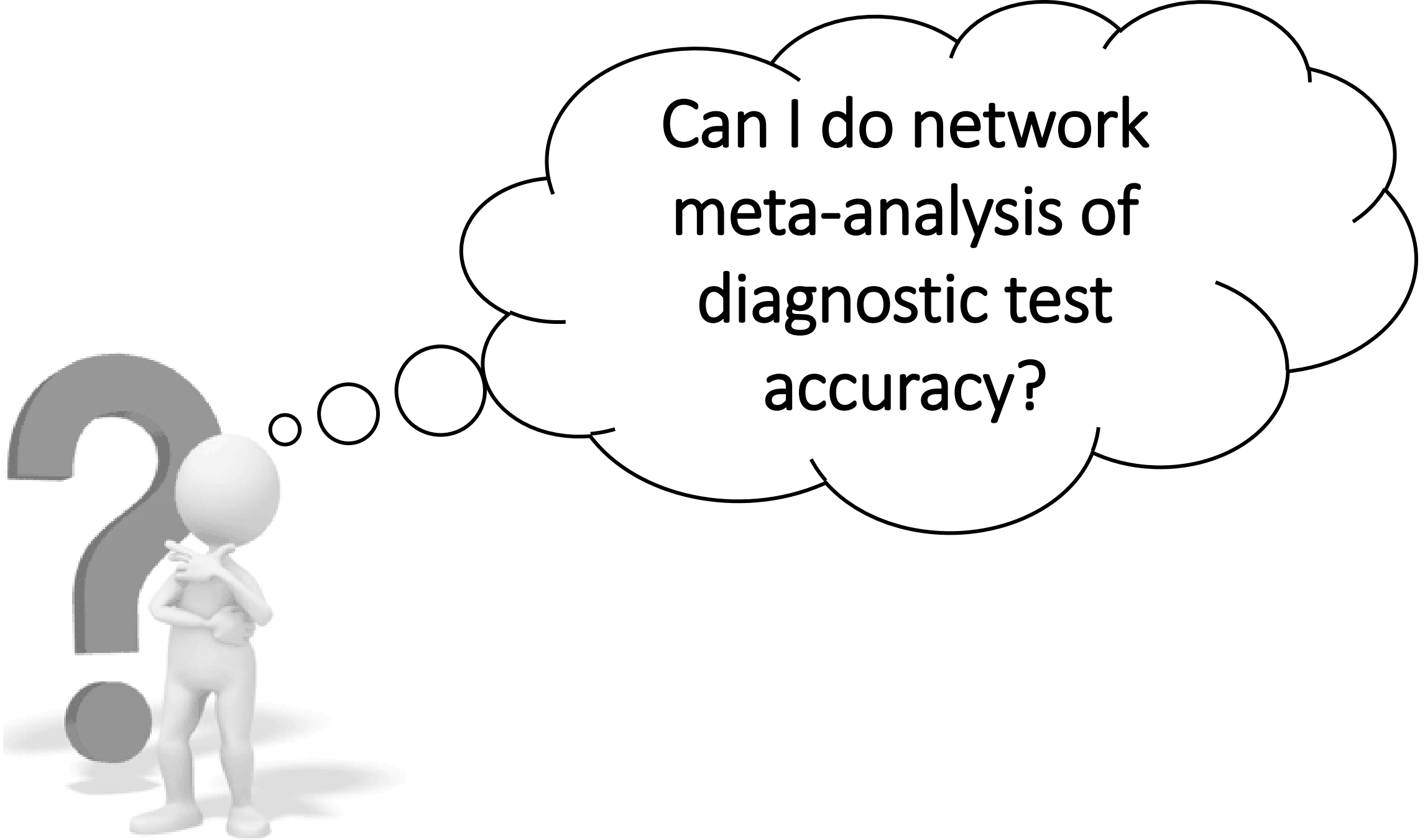
- **User guide** version 1.3 (2012). (PDF 2.7MB, opens in new window)
- **Quick reference** and worked example (2012). (PDF 2.6MB, opens in new window)
- The SAS macro itself: **METADAS v1.3**. This is provided as a text-file and opens in a new window.

R

There are several user-written packages for conducting meta-analysis of diagnostic test accuracy (DTA) studies in R. This tutorial summarises and illustrates some of the packages. Step-by-step instructions are also provided for carrying out the bivariate binomial method by fitting a generalized linear mixed model (GLMM) using the glmer function in the R package lme4. A .R file, “Bivariate binomial meta-analysis of diagnostic test accuracy studies.R” and example dataset based on a review by Schuetz et al. 2010, are included with the **tutorial** in the zipped folder.

STATA

METANDI: A Stata user-written package for meta-analysis of diagnostic accuracy studies (Harbord and Whiting 2009;


Can I do network meta-analysis of diagnostic test accuracy?

What are the DTA-NMA methods and which one should I use?

Network meta-analysis for DTA (DTA-NMA)

EVIDENCE SYNTHESIS METHODS
STATISTICS TEAM


Sofia


Gerta


Argie

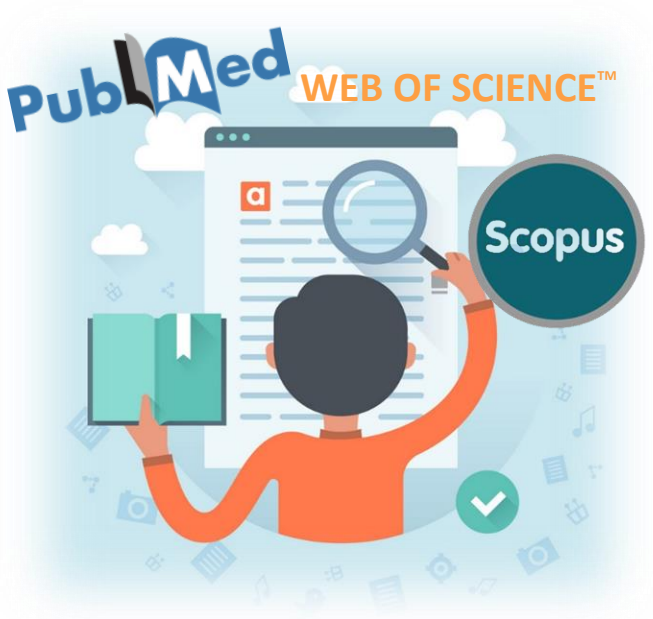

Dimitris


Ridhi

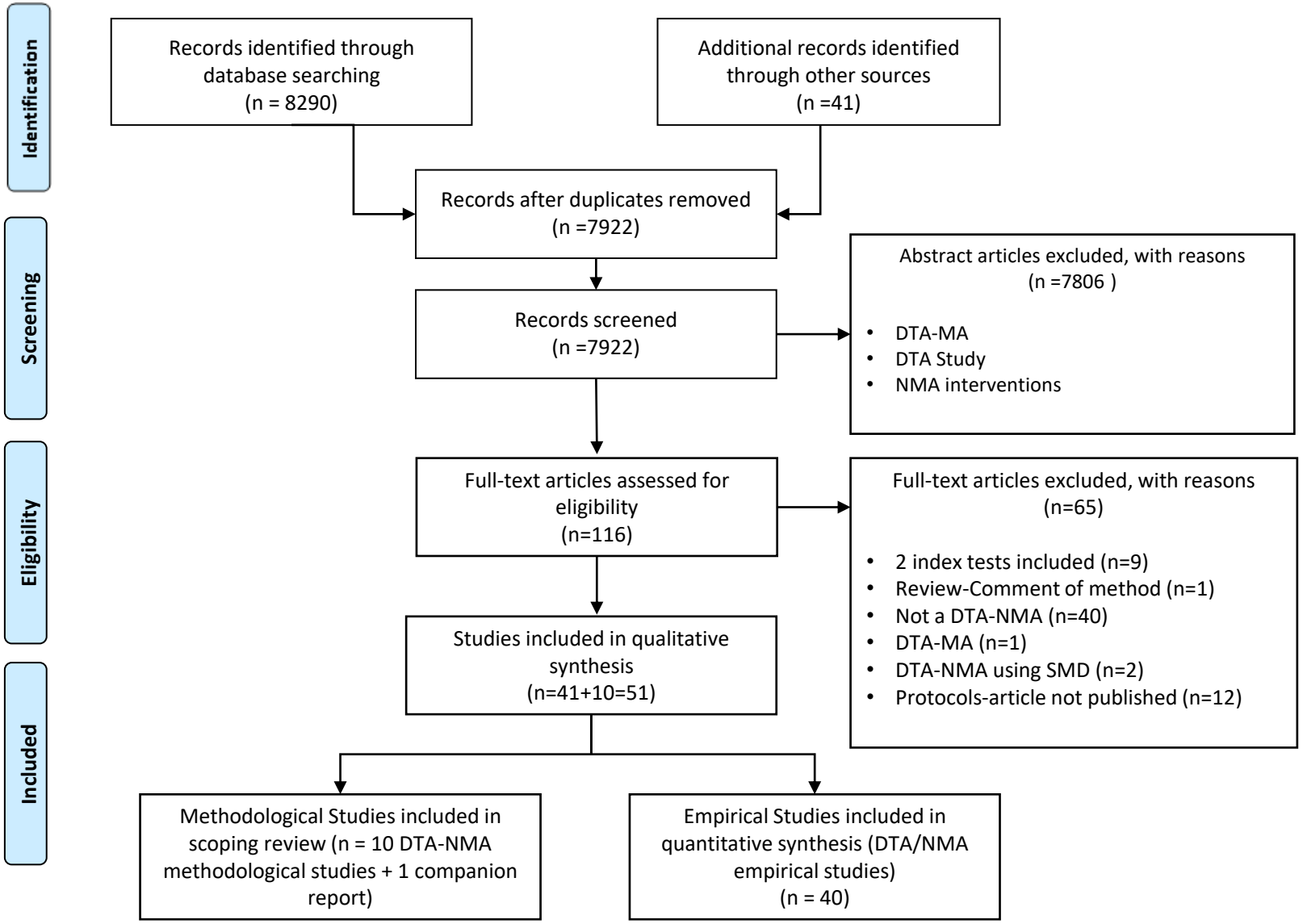

Yemisi

Identification of NMA-DTA methods

10 methodological studies and 40 empirical studies



From inception to end of July 2019



DTA-NMA methods

10 methodological studies of 9 different DTA-NMA methods

Model	Arm-based	Bayesian setting	Imperfect reference standard	Multiple thresholds	Joint classification tables	2x2 tables/ index test
Trikalinos et al. 2014	X	X			X	
Ma 2015	X	X	X		X	
Menten & Lesaffre 2015 (Model A)		X				X
Menten & Lesaffre 2015 (Model B)		X	X	X		X
Menten & Lesaffre 2015 (Model C)		X	X			X
Dimou et al. 2016	X				X	
Cheng 2016 (Model A)	X	X			X	
Cheng 2016 (Model B)	X	X		X	X	
Cheng 2016 (Model C)	X	X			X	
Nyaga et al. 2018a	X	X				X
Nyaga et al. 2018b	X	X				X
Owen et al. 2018	X	X		X		X
Lian et al. 2019	X	X	X	X	X	

HJOG
An Obstetrics and Gynecology International Journal

Review

HJOG 2021, 20 (1), 11-24

Evaluating multiple diagnostic tests: An application to cervical cancer

Areti Angeliki Veroniki[1,2,3], Sofia Tsokani[1], Evangelos Paraskevaidis[4], Dimitris Mavridis[1,5]

Hierarchical meta-regression and DTA-NMA methods

Bivariate meta-regression model

Reitsma et al. (2005)

- A covariate for test type is used to explore sensitivity and specificity between tests
- Assumes that participants undergoing different tests are independent subgroups within each study
- Does not account for the within-study correlation between tests

Normal-binomial model

Nyaga et al. (2018a)

- Hierarchical model using the logit transformation of sensitivity and specificity
- Allows for correlation between tests

Hierarchical latent class model

Menten and Lesaffre (2015)

- Based on differences (contrasts) between the different tests in the network
- Allows for different reference standards
- Correlations between tests from the same study are ignored

Beta-binomial model

Nyaga et al. (2018b)

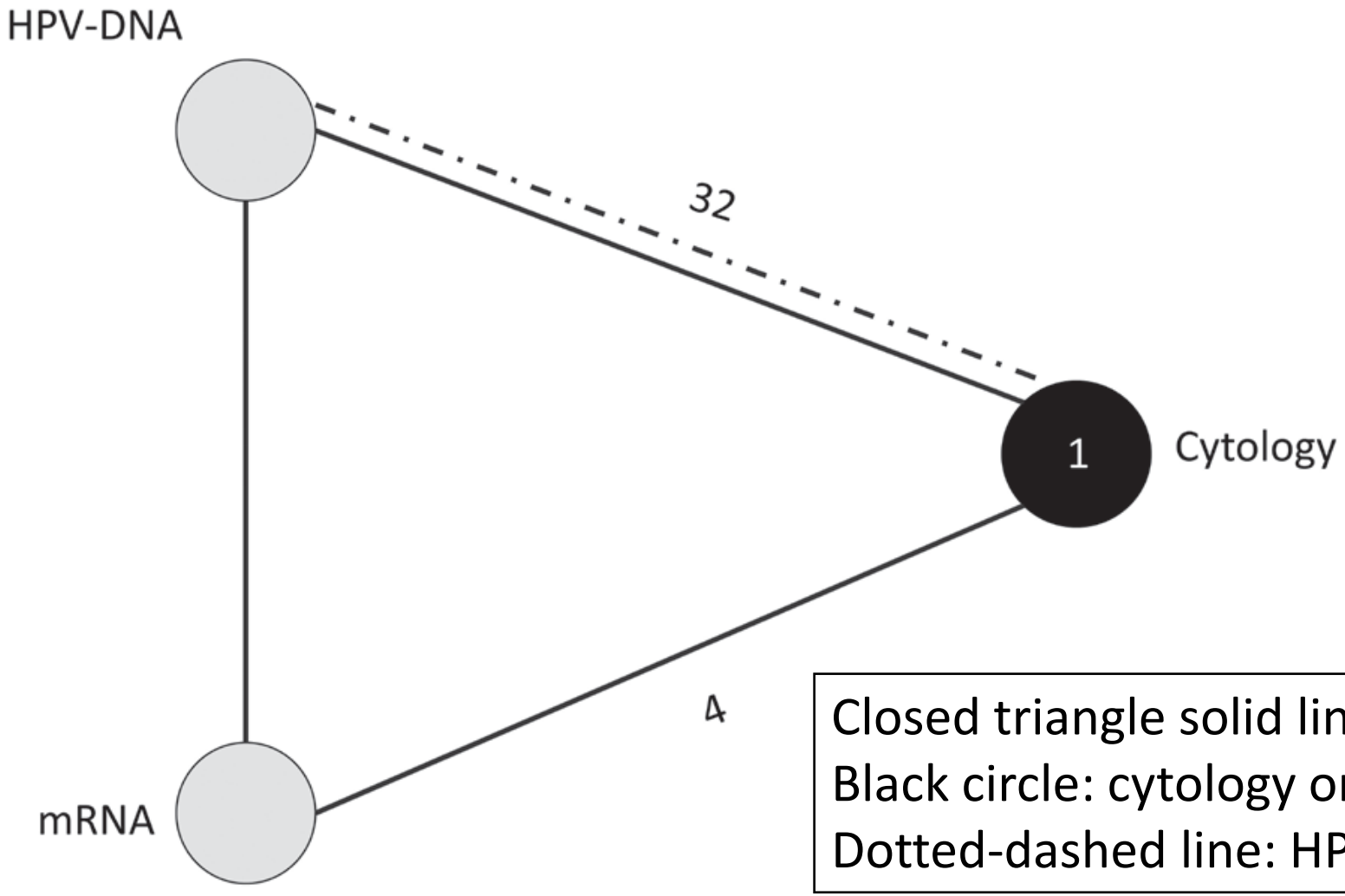
- Sensitivity & specificity are directly modelled using a beta-binomial defined in [0,1]
- Allows for correlation between tests

Variance component model

Owen et al. (2018)

- Extension to the normal-binomial model
- Allows for multiple thresholds
- Incorporates constraints on threshold effects

Network plot of cytology, HPV DNA, and mRNA tests for CIN2+



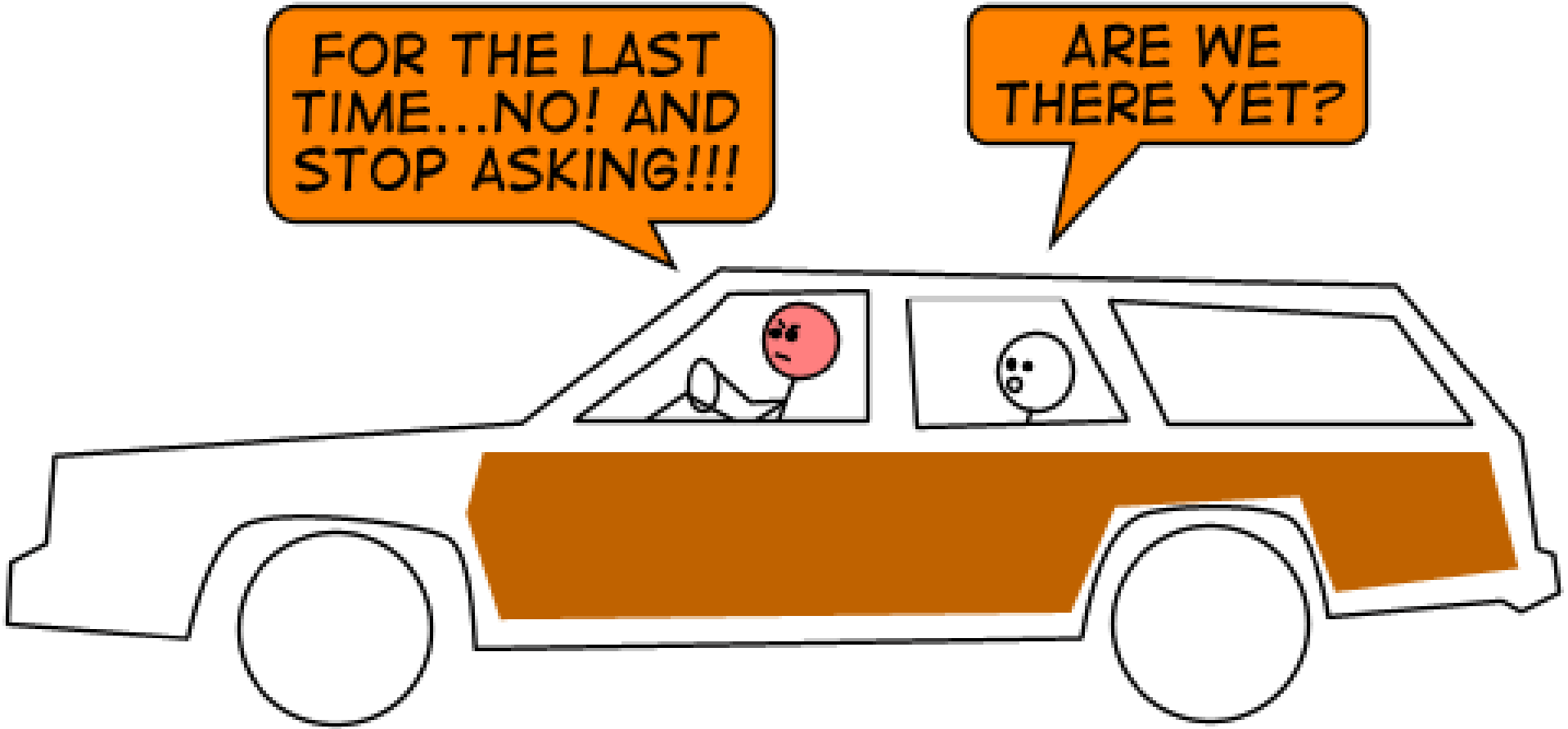
Closed triangle solid line: triple-test studies (n = 4)
Black circle: cytology only (n = 1)
Dotted-dashed line: HPV DNA vs cytology (n = 32)

Veroniki AA, Tsokani S, Praskevaidis E, Mavridis D. Evaluating multiple diagnostic tests: An application to cervical cancer. HJOG. 2021;20 (1): 11-24.

Summary of application to cervical cancer

- Different DTA-NMA methods may lead to different results
  - Differences in point estimates and their uncertainty
- Differences in results across models may be due to differences in how the models deal with
  - Heterogeneity
  - Sensitivity and specificity (logits or proportions)
- Choice of a DTA-NMA method depends on the available data

Are we there yet with DTA-NMA?



Limitations of DTA-NMA

- Comprehensive evaluation is needed to assess the performance of the models
- Complexity: as number of tests increase, number of additional parameters to estimate increase, and so does risk of convergence issues
- Data availability
- Lack of easy to use programs in popular statistical software

Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy

Chapter 10 updated 2021 (online soon)

"Meta-analytic models that account for pairing of test results within an individual within each study have been developed as an extension of the bivariate model. The method proposed by Trikalinos (2014) … The approach of Dimou (2016) … These methods **require further evaluation before they are recommended for routine use**. However, as suggested by Trikalinos (2014) **they may be useful as a sensitivity analysis**.

Network meta-analysis models have also been developed that utilise data from both direct and indirect comparisons of multiple tests... However, **further evaluation of these methods for dealing with complex correlational structures is required before they are implemented in Cochrane reviews**. "

Take home message

- Be clear about the test comparison strategy and strength of the evidence
  - All studies (comparative and non-comparative studies)
  - Restricted to comparative studies that have directly compared the tests
  - Analyses using relevant comparative studies are desirable but may not be feasible

- Hierarchical meta-regression models for comparison of points (bivariate model) or curves (HSROC model) are the norm.

- More complex methods are being published but evaluations are required before they can be adopted in Cochrane DTA reviews.

- A rapidly developing field so watch this space.

Acknowledgements

- Thanks to Argie Veroniki and Sofia Tsokani for slides on the scoping review and the case study.

FUNDED BY

NIHR | National Institute for Health Research

- The views expressed are mine and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Cochrane Methods
Screening and Diagnostic Tests

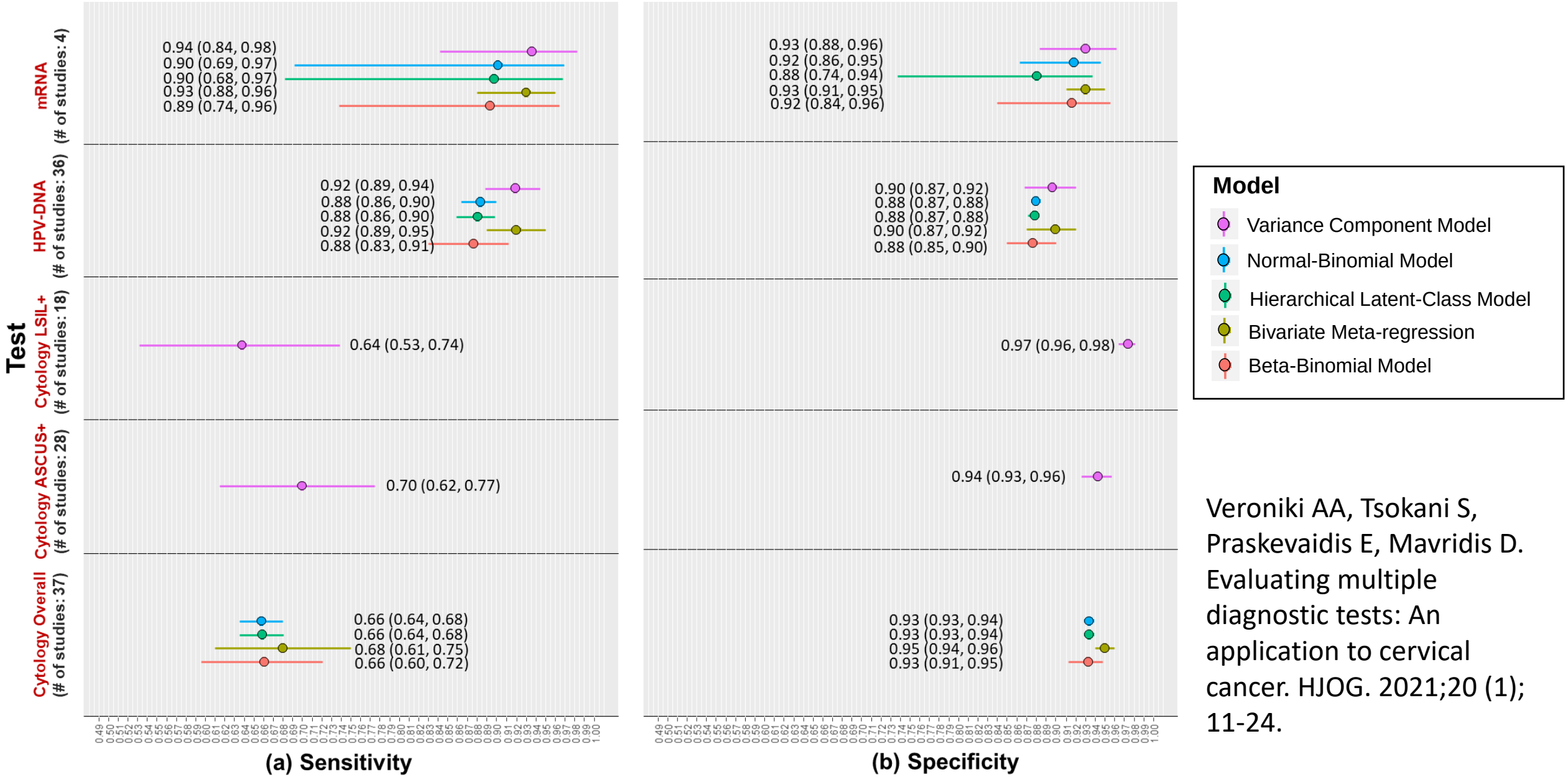
References

- Chu H, Cole SR. Bivariate meta-analysis for sensitivity and specificity with sparse data: a generalized linear mixed model approach (letter to the Editor). J Clin Epidemiol. 2006;59(12):1331-2.
- Dimou NL, Adam M, Bagos PG. A multivariate method for meta-analysis and comparison of diagnostic tests. Statistics in Medicine 2016; 35: 3509-3523.
- Harbord RM et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 2007;8(2):239-51.
- Lian Q, Hodges JS, Chu H. A Bayesian Hierarchical Summary Receiver Operating Characteristic Model for Network Meta-analysis of Diagnostic Tests. J Am Stat Assoc 2019; 114: 949-961.
- Ma X, Lian Q, Chu H, Ibrahim JG, Chen Y. A Bayesian hierarchical model for network meta-analysis of multiple diagnostic tests. Biostatistics 2018;19(1):87-102.
- Macaskill P et al. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration; 2010. Available at https://methods.cochrane.org/sdt/welcome
- Menten J, Lesaffre E. A general framework for comparative Bayesian meta-analysis of diagnostic studies. BMC Med Res Methodol 2015;15:70.

References

- Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. Stat Methods Med Res 2018[a];27(6):1766-1784.
- Nyaga VN, Arbyn M and Aerts M, Beta-binomial analysis of variance model for network meta-analysis of diagnostic test accuracy data. Stat Methods Med Res, 2018[b]. 27(8):2554-66.
- Owen RK, Cooper NJ, Quinn TJ, Lees R, Sutton AJ. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. J Clin Epidemiol 2018;99:64-74.
- Reitsma JB et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982-90.
- Rücker G. Network Meta-Analysis of Diagnostic Test Accuracy Studies. In: Biondi-Zoccai G, editor. Diagnostic Meta-Analysis: A Useful Tool for Clinical Decision-Making: Springer; 2018.
- Rutter CM, Gatsonis C. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20(19)2865-84.
- Trikalinos TA, Hoaglin DC, Small KM, Terrin N, Schmid CH. Methods for the joint meta-analysis of multiple tests. Research Synthesis Methods 2014; 5: 294-312.

Application: Cervical Intraepithelial Neoplasia



Veroniki AA, Tsokani S, Praskevaidis E, Mavridis D. Evaluating multiple diagnostic tests: An application to cervical cancer. HJOG. 2021;20 (1); 11-24.