Diagnostic systematic reviews and meta-analyses

Mariska Leeflang
m.m.leeflang@amc.uva.nl

Disclaimer!

Cochrane

Trusted evidence.
Informed decisions.
Better health.

What is a diagnostic systematic review?

- What outcome measures?
- What kind of designs?

What is a diagnostic systematic review?

A diagnostic systematic review can include randomized controlled trials, with risk differences as the outcome measure.

TRUE or FALSE?

Cochrane Database of Systematic Reviews

Rapid diagnostic tests versus clinical diagnosis for managing people with fever in malaria endemic settings

Review | Intervention

John Odaga, David Sinclair, Joseph A Lokong, Sarah Donegan, Heidi Hopkins, Paul Garner

First published: 17 April 2014

Editorial Group: Cochrane Infectious Diseases Group

DOI: 10.1002/14651858.CD008998.pub2 View/save citation

Cited by: 6 articles Citation tools

Am score 29

Abstract

English | French

Background

In 2010, the World Health Organization recommended that all patients with suspected malaria are tested for malaria before treatment. In rural African settings light microscopy is often unavailable. Diagnosis has relied on detecting fever, and most people were given antimalarial drugs presumptively. Rapid diagnostic tests (RDTs) provide a point-of-care test that may improve management, particularly of people for whom the RDT excludes the diagnosis of malaria.

Objectives

Systematic reviews of diagnostic test accuracy (1)

- **P**atients: patients who will be tested in practice with the test you are interested in
- **I**ndex test: the test or tests you are interested in
- **C**omparison: an alternative test (NOT necessarily the reference standard)
- **O**utcome: target condition yes/no, as defined by the reference standard

Systematic reviews of diagnostic test accuracy (2)

- Outcome measures: sensitivity, specificity
  - Also predictive values, odds ratios, likelihood ratios

- Study design: typically cross-sectional
  - Index test results and reference standard result are measured at the same time
  - Diagnosis is about the ***current*** status of the patient

Cochrane Database of Systematic Reviews

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

Comment | Review | Diagnostic

Katharine Abba ✉, Jonathan J Deeks, Piero L Olliaro, Cho-Min Naing, Sally M Jackson, Yemisi Takwoingi, Sarah Donegan, Paul Garner

First published: 6 July 2011

Editorial Group: Cochrane Infectious Diseases Group

DOI: 10.1002/14651858.CD008122.pub2 View/save citation

Cited by: 12 articles Citation tools

Am score 4

Abstract

English | Spanish; Castilian | French

Background

Rapid diagnostic tests (RDTs) for *Plasmodium falciparum* malaria use antibodies to detect either HRP-2 antigen or pLDH antigen, and can improve access to diagnostics in developing countries.

Objectives

To assess the diagnostic accuracy of RDTs for detecting *P. falciparum* parasitaemia in persons living in endemic areas who present to ambulatory healthcare facilities with symptoms suggestive of malaria by

Steps of a Cochrane systematic review

1. define the review question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Steps of a Cochrane systematic review

1. define the review question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Quality assessment: QUADAS-2

	Patient Selection	Index Test(s)	Reference Standard	Flow and Timing
Risk of Bias	Consecutive sample No case control All inclusive	Blinding Pre-specified cut-off	Likely to correctly classify the disease Blinding	Appropriate time interval Partial Verification Differential verification Missings
Concerns regarding applicability	Representative sample?	Index test same as in practice?	Appropriate target condition?	N/A

Review Article

Systematic Review into Diagnostics for Post-Kala-Azar Dermal Leishmaniasis (PKDL)

Emily R. Adams,[1,2] Inge Versteeg,[2] and Mariska M. G. Leeflang[3]

[1] *Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK*
[2] *Royal Tropical Institute, KIT Biomedical Research, 1105 AZ Amsterdam, The Netherlands*
[3] *Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands*

Correspondence should be addressed to Emily R. Adams; e.adams@liv.ac.uk

Received 21 January 2013; Revised 22 April 2013; Accepted 8 May 2013

Academic Editor: Abul Faiz

Identification of post-kala-azar dermal leishmaniasis (PKDL) is important due to the long and toxic treatment and the fact that PKDL patients may serve as a reservoir for visceral leishmaniasis (VL). We summarized the published literature about the accuracy of diagnostic tests for PKDL. We searched Medline for eligible studies investigating the diagnostic accuracy of any test for PKDL.

Example PKDL

DAT

Study	TP	FP	FN	TN	Reference standard	Study design	Control group
Das 2007	10	15	0	10	Skin slit smear	Case control	Healthy endemic controls
Harith 1996	57	0	0	11	Skin slit smear	Case control	Healthy endemic controls
Hassan 1992	16	0	1	10	Skin slit smear	Case control	Leprosy patients
Kumar 2008	19	0	1	19	Microscopy otherwise	Case control	Healthy endemic controls
Singh 2005	64	0	4	12	Response to treatment	Case control	Healthy endemic controls

rK39 lateral flow

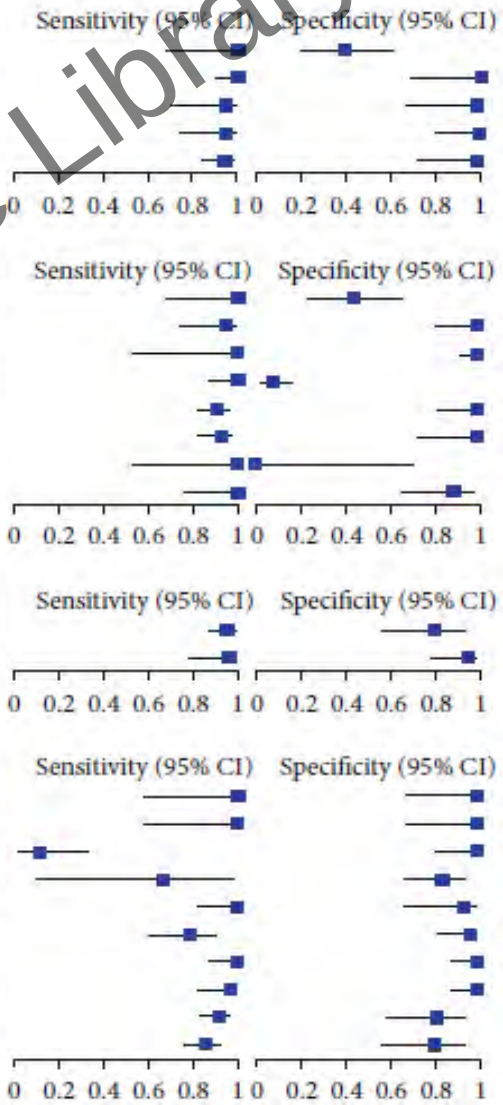
Study	TP	FP	FN	TN	Reference standard	Study design	Control group
Das 2007	10	14	0	11	Skin slit smear	Case control	Healthy endemic controls
Kumar 2008	19	0	1	19	Microscopy otherwise	Case control	Healthy endemic controls
Mathur 2004	6	0	0	50	Microscopy otherwise	Case control	Differential diagnosis controls
Nasreen 2012	32	72	0	6	Skin slit smear	Consecutive	N/A
Salotra 2001 a	81	0	8	20	Microscopy otherwise	Case control	Healthy endemic controls
Singh 2005	63	0	5	12	Response to treatment	Case control	Healthy endemic controls
Singh 2012	6	3	0	0	Skin slit smear	Consecutive	N/A
Uranw 2011	14	2	0	17	Skin slit smear	Consecutive	N/A

rK39 ELISA

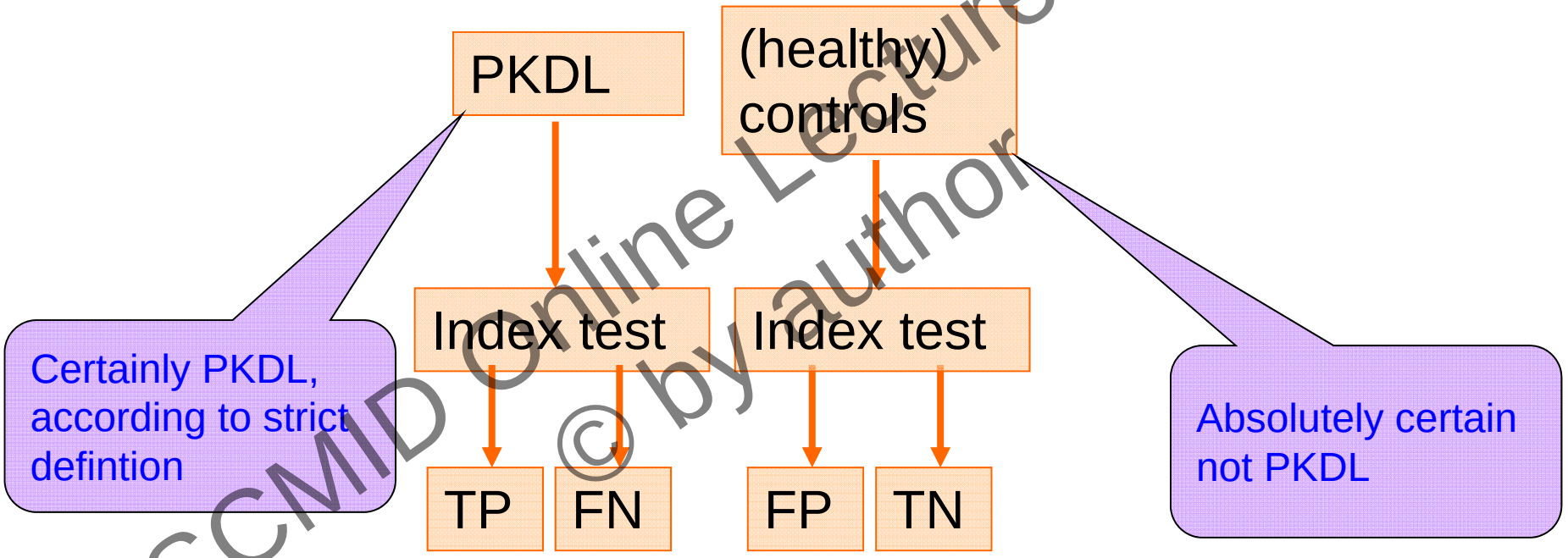
Study	TP	FP	FN	TN	Reference standard	Study design	Control group
Salotra 2002 1	84	4	4	17	Response to treatment	Case control	Healthy endemic controls
Salotra 2003	24	1	1	24	Response to treatment	Case control	Leprosy patients

Other serological tests

Study	TP	FP	FN	TN	Reference standard	Study design	Control group
Chaterjee 1998	17	0	0	10	Microscopy otherwise	Case control	Healthy endemic controls
Chaterjee 1999	27	0	0	10	Microscopy otherwise	Case control	Healthy endemic controls
Kumar 2008	2	0	18	19	Microscopy otherwise	Case control	Healthy endemic controls
Mathur 2004	2	5	1	27	Microscopy otherwise	Case control	Differential diagnosis controls
Saha 2011	20	1	0	14	Skin slit smear	Case control	Healthy endemic controls
Salotra 1999 1	25	1	7	29	Response to treatment	Case control	Healthy controls
Salotra 1999 2	32	0	0	30	Response to treatment	Case control	Healthy controls
Salotra 1999 3	31	0	1	30	Response to treatment	Case control	Healthy controls
Salotra 2002 1	81	4	7	18	Response to treatment	Case control	Healthy endemic controls
Salotra 2002 2	76	4	12	17	Response to treatment	Case control	Healthy endemic controls

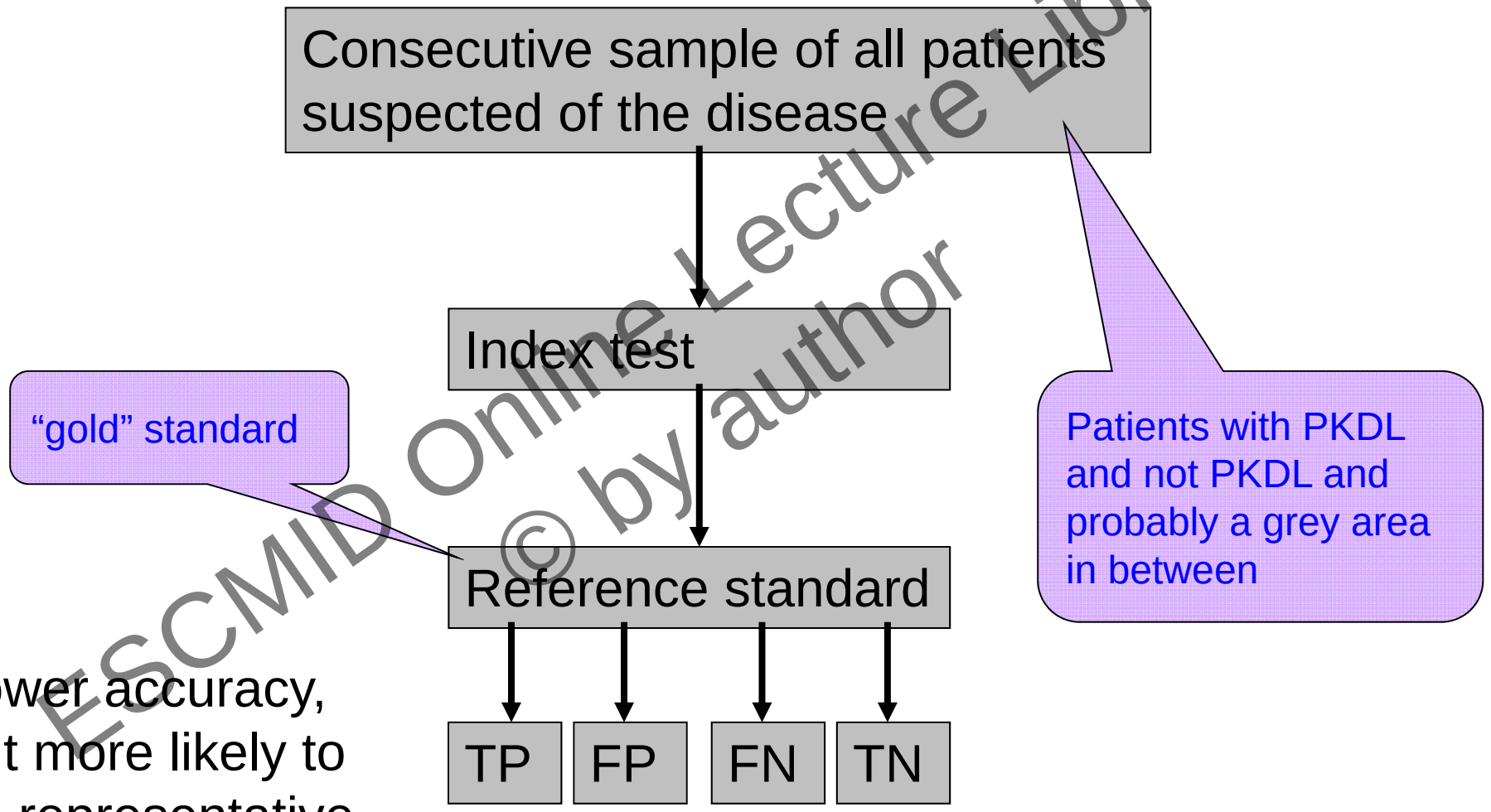


Diagnostic case control



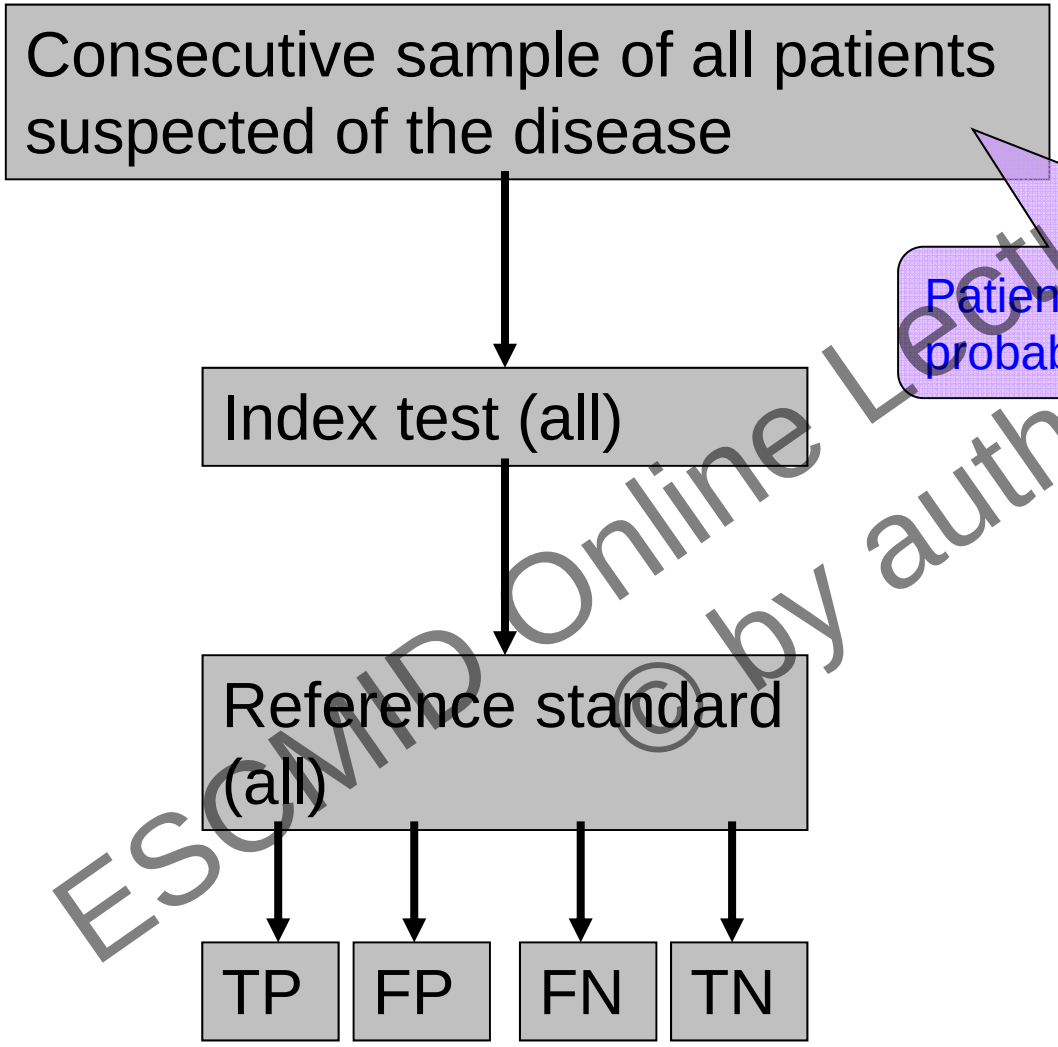
Nice results, but not representative for practice

Ideal accuracy design



Lower accuracy, but more likely to be representative

Ideal accuracy design



Patients with PKDL and not PKDL and probably a grey area in between

Participants may be recruited from:

-Screening program including asymptomatic patients

-based on (self-reported) signs and symptoms

-After initial screening with an ELISA

Leeflang *et al. BMC Infectious Diseases* (2016) 16:140
DOI 10.1186/s12879-016-1468-4

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

The diagnostic accuracy of serological tests for Lyme borreliosis in Europe: a systematic review and meta-analysis

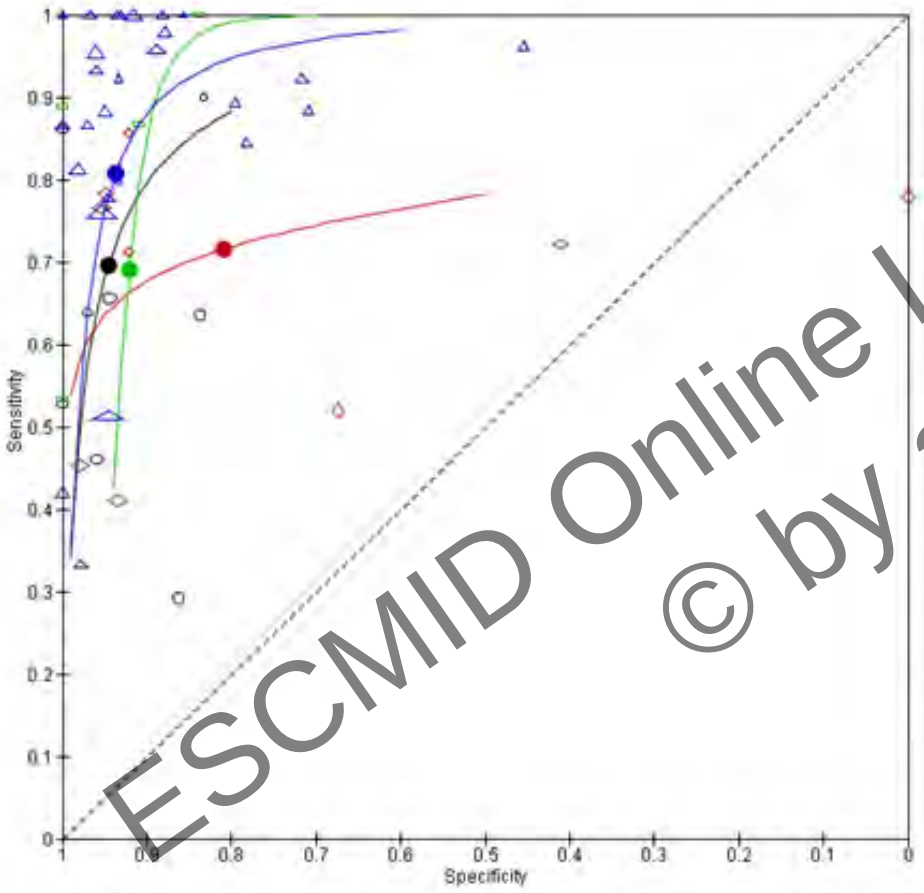
M. M. G. Leeflang[12*], C. W. Ang[1], J. Berkhout[2], H. A. Bijlmer[3], W. Van Bortel[4], A. H. Brandenburg[5], N. D. Van Burgel[6], A. P. Van Dam[7], R. B. Dessau[8], V. Fingerle[9], J. W. R. Hovius[10], B. Jaulhac[11], B. Meijer[13], W. Van Pelt[3], J. F. P. Schellekens[13], R. Spijker[14], F. F. Stelma[15], G. Stanek[16], F. Verduyn-Lunel[17], H. Zeller[4] and H. Sprong[3]

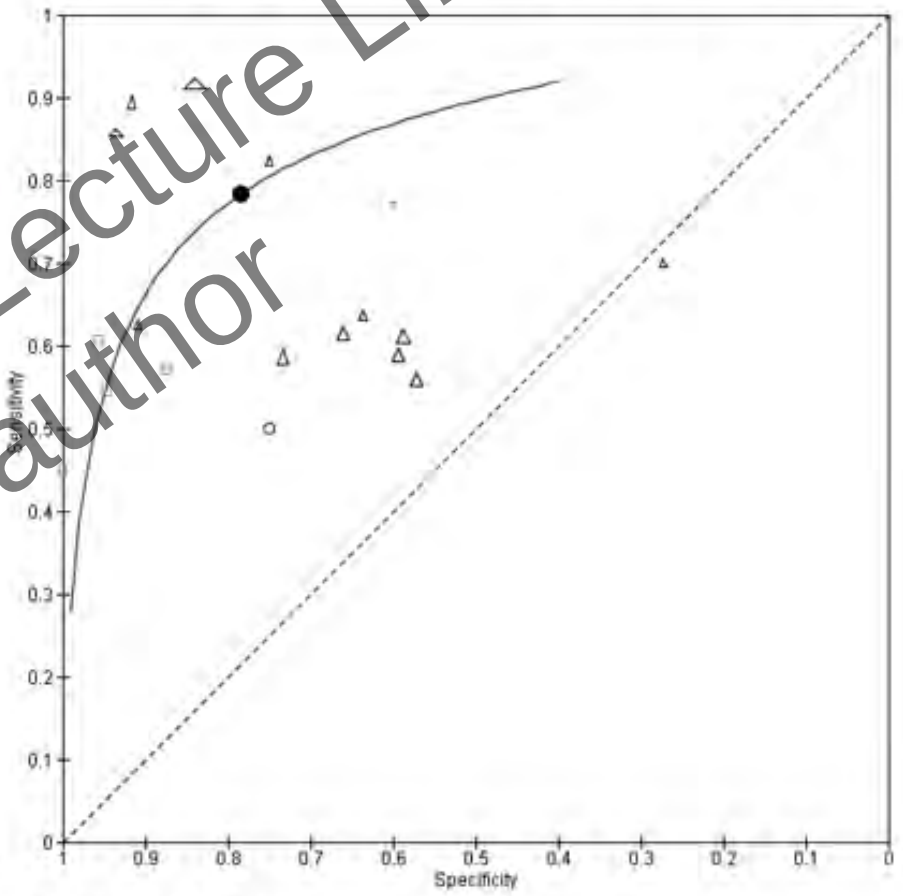
Of the 80 included studies, 60 were case-control design. Most of these included healthy controls or cross-reacting controls.

serological assays for Lyme borreliosis in Europe were eligible. Study selection and data-extraction were done by

Example NB

3a: neuroborreliosis case-control studies including healthy controls. **3b:** neuroborreliosis cross-sectional studies.





Triangle=commercial EIA; Diamond=in house EIA; Square=commercial IB; Circle=in house IB.

Quality assessment: QUADAS-2

	Patient Selection	Index Test(s)	Reference Standard	Flow and Timing
Risk of Bias	Consecutive sample No case control All inclusive	Blinding Pre-specified cut-off	Likely to correctly classify the disease Blinding	Appropriate time interval Partial Verification Differential verification Missings
Concerns regarding applicability	Representative sample?	Index test same as in practice?	Appropriate target condition?	N/A

Reference standard

Is not the same as the gold standard!

Is not necessarily inferior to the gold standard!

“Gold standard”

Example trypanosomiasis

Although microscopy is not gold, you don’t want to treat if parasite has not been seen

So microscopy guides therapy

Practical solution: define target condition in terms of the reference standard

Example malaria

For PCR, the proportion of microscopy positives detected ("sensitivity") was 98% [91-99 CI] and the proportion microscopy negatives, with a negative PCR ("specificity"), was 65% [59-72 CI]. PCR may seem to miss fewer patients with peripheral *P. falciparum* parasites than an RDT, but does indicate a number of women without parasites detected by microscopy as having malaria. Whether these are cases that were missed by microscopy or whether these were false positive PCR results, resulting in low specificity, needs to be further investigated.

Other solutions

- Latent class analysis
- Include expert opinion in reference standard
- Reference standard based on therapy
- Reference standard based on transmission patterns
- Adjust for assumed accuracy reference standard

… but each method has its own pitfalls…

Partial / Differential verificiation

Sometimes the ideal reference standard is too costly or too burdensome or just not possible

E.g. biopsy is not possible if initially no lesion found

E.g. operating everyone for appendicitis is not done

In that case, researchers have two options:

Verifiy only part of the patients

Partial verification

Verify the other part with another (less ideal) reference standard

Differential verification

This may lead to biased results

Partial verification?

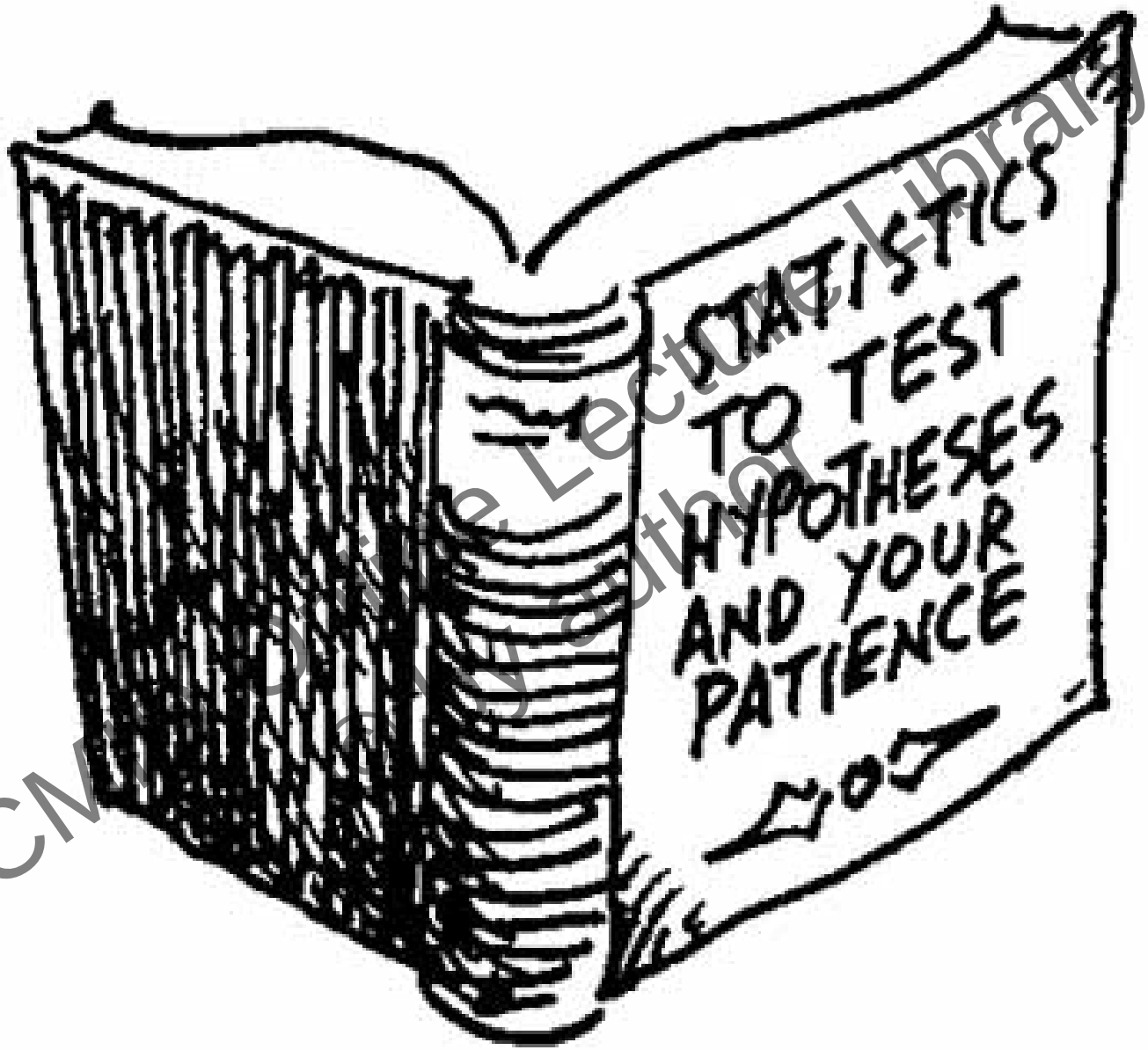
Is a form of missing data

Missing data may cause biased results

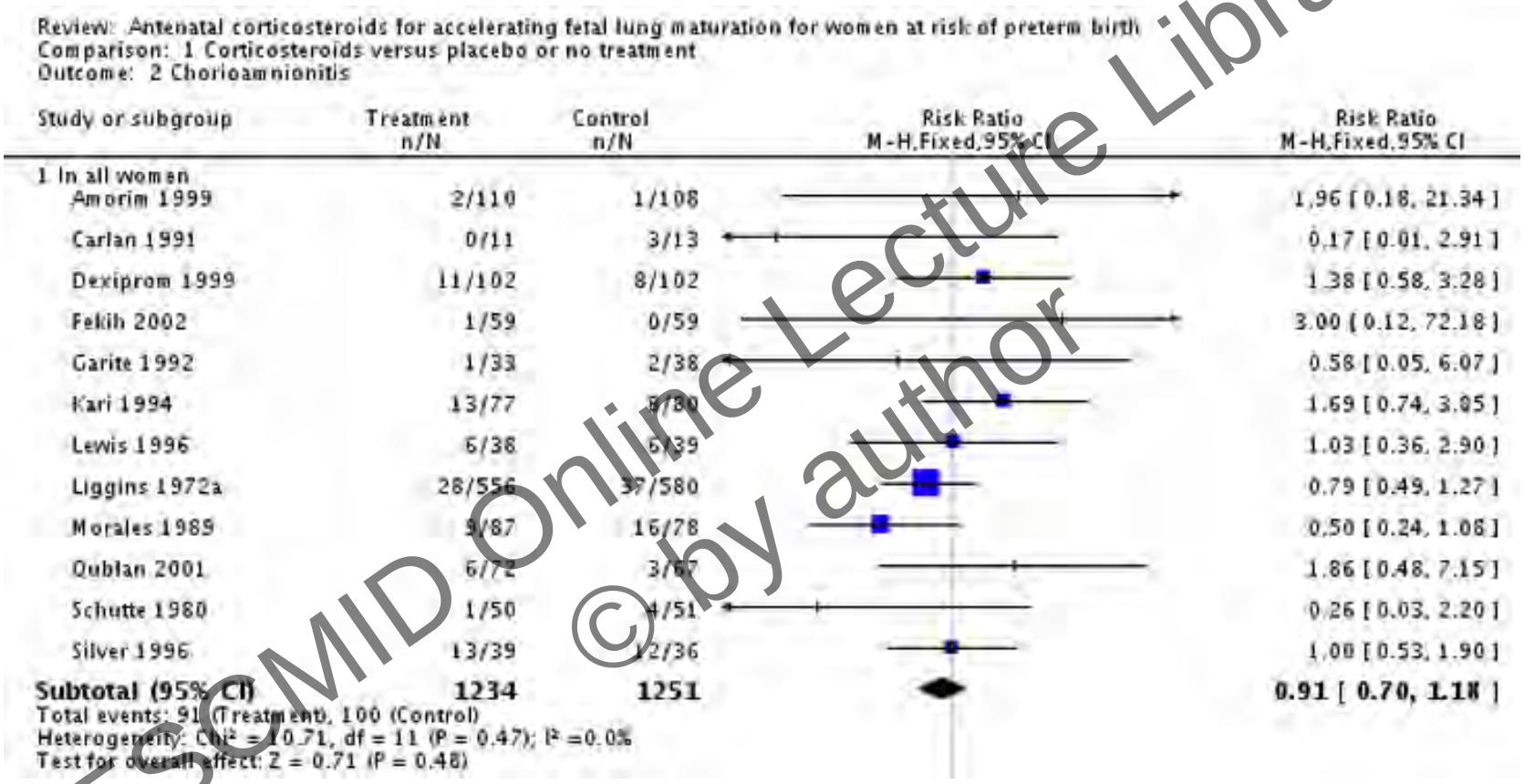
What if a study reports the inclusion of 200 patients and 2x2 tables also include 200 patients, but you know that some patients cannot have had the reference standard?

Steps of a Cochrane systematic review

1. define the review question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review



Intervention Review



Risk Ratio = 0.91 → When using corticosteroids, the chances of developing chorioamionitis are 0.91 times the chances without any treatment.

DTA meta-analysis: challenging

Outcome measure: paired.

- Sensitivity and specificity
- ROC curve

Threshold problems

- Explicit
- Implicit

Heterogeneity is rule rather than exception

- Issues of bias
- Different study designs
- Random effects models default

Figure 4. Forest plot of the included studies. TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative. Between brackets the 95% confidence intervals (CI) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% confidence interval (black horizontal line).

Study	TP	FP	FN	TN	Sensitivity	Specificity
Adam 2004	1	41	1	175	0.50 [0.01, 0.99]	0.81 [0.75, 0.86]
Allan 2005	0	1	1	123	0.00 [0.00, 0.97]	0.99 [0.96, 1.00]
Becker 2003	6	12	7	62	0.46 [0.19, 0.75]	0.84 [0.73, 0.91]
Bialek 2002	1	8	0	8	1.00 [0.03, 1.00]	0.50 [0.25, 0.75]
Bretagne 1998	14	5	4	18	0.78 [0.52, 0.94]	0.78 [0.56, 0.93]
Buchheidt 2004	3	1	6	167	0.33 [0.07, 0.70]	0.99 [0.97, 1.00]
Busca 2006	2	12	0	60	1.00 [0.16, 1.00]	0.83 [0.73, 0.91]
Challier 2004	20	9	6	35	0.77 [0.56, 0.91]	0.80 [0.65, 0.90]
Doermann 2002	10	4	2	407	0.83 [0.52, 0.98]	0.99 [0.98, 1.00]
Florent 2006	8	39	4	116	0.67 [0.35, 0.90]	0.75 [0.67, 0.81]
Foy 2007	6	7	6	102	0.50 [0.21, 0.79]	0.94 [0.87, 0.97]
Herbrecht 2002	31	49	67	650	0.32 [0.23, 0.42]	0.93 [0.91, 0.95]
Hovi 2007	1	4	1	111	0.50 [0.01, 0.99]	0.97 [0.91, 0.99]
Kallel 2003	4	7	1	62	0.80 [0.28, 0.99]	0.90 [0.80, 0.96]
Kawazu 2004	11	23	0	115	1.00 [0.72, 1.00]	0.83 [0.76, 0.88]
Lai 2007	11	14	3	161	0.79 [0.49, 0.95]	0.92 [0.87, 0.96]
Machetti 1998	3	3	1	15	0.75 [0.19, 0.99]	0.83 [0.59, 0.96]
Maertens 2002	11	7	2	80	0.85 [0.55, 0.98]	0.92 [0.84, 0.97]
Marr 2004	20	11	4	32	0.83 [0.63, 0.95]	0.74 [0.59, 0.86]
Moragues 2003	2	1	2	49	0.50 [0.07, 0.93]	0.98 [0.89, 1.00]
Pereira 2005	1	9	0	29	1.00 [0.03, 1.00]	0.76 [0.60, 0.89]
Pinel 2003	17	17	17	756	0.50 [0.32, 0.68]	0.98 [0.97, 0.99]
Rovira 2004	9	15	2	123	0.82 [0.48, 0.98]	0.89 [0.83, 0.94]
Scotter 2005	3	1	2	19	0.60 [0.15, 0.95]	0.95 [0.75, 1.00]
Suankratay 2006	16	13	1	20	0.94 [0.71, 1.00]	0.61 [0.42, 0.77]
Ulusakarya 2000	16	11	0	108	1.00 [0.79, 1.00]	0.91 [0.84, 0.95]
Weisser 2005	16	41	4	100	0.80 [0.56, 0.94]	0.71 [0.63, 0.78]
White 2005	0	2	3	100	0.00 [0.00, 0.71]	0.98 [0.93, 1.00]
Williamson 2000	6	8	1	89	0.86 [0.42, 1.00]	0.92 [0.84, 0.96]
Yoo 2005	12	25	2	89	0.86 [0.57, 0.98]	0.78 [0.69, 0.85]

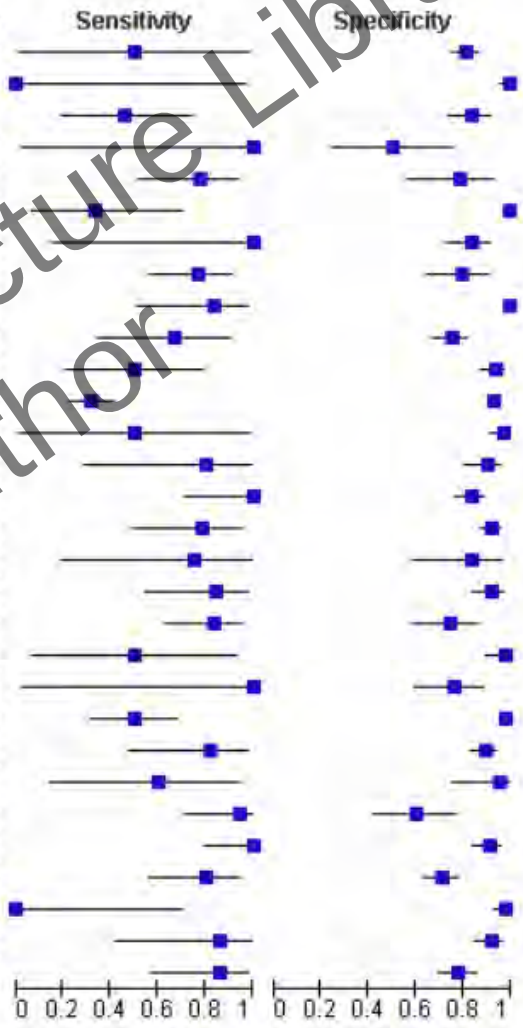
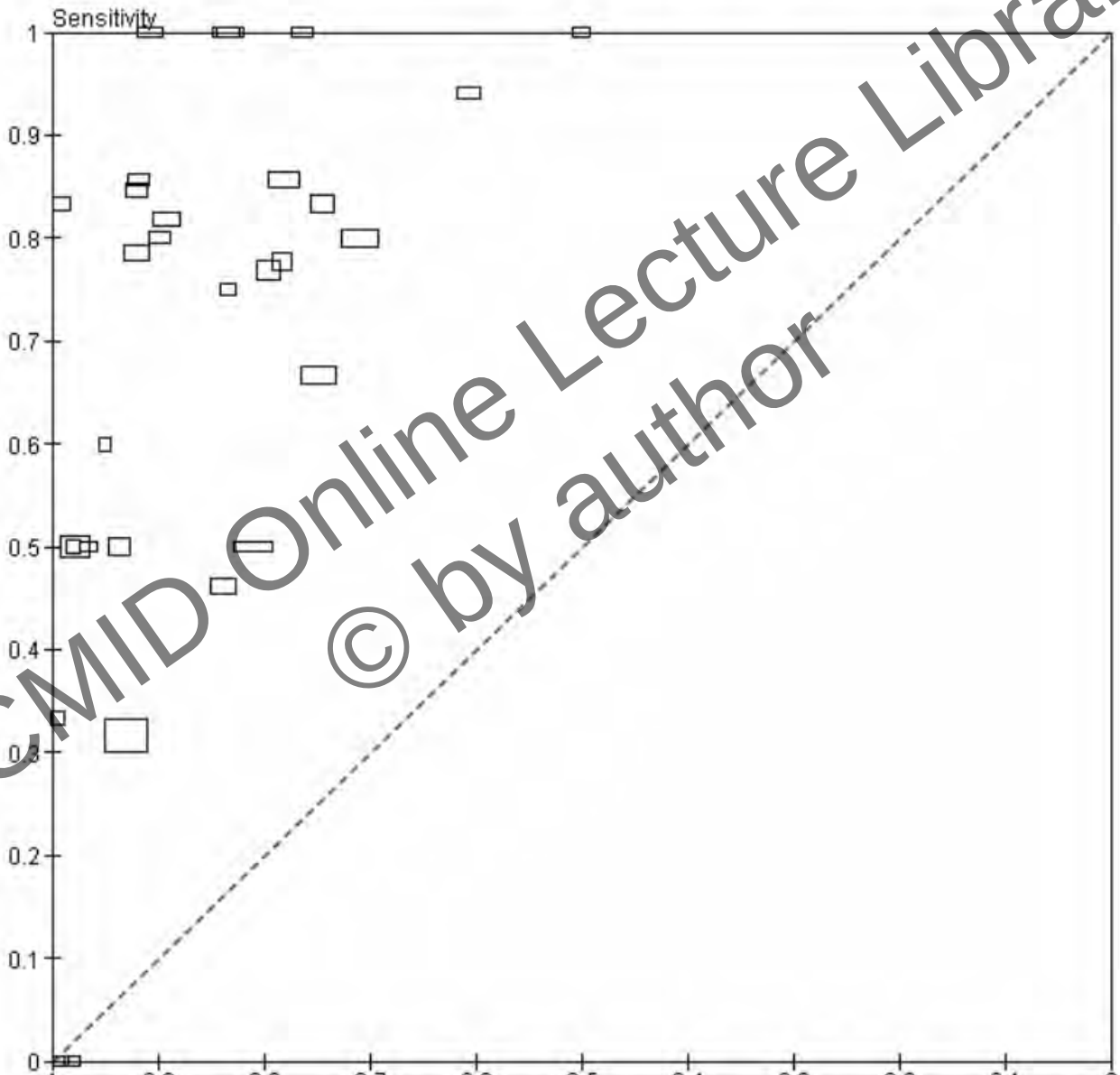


Figure 5. Plot of sensitivity versus specificity for all 30 studies, irrespective of cut-off value. The width of the blocks is proportional to the inverse standard error of the specificity in every study and the height of the blocks is proportional to the inverse standard error of the sensitivity.



Advanced software needed

Review Manager or SPSS are not capable of doing advanced multilevel modelling

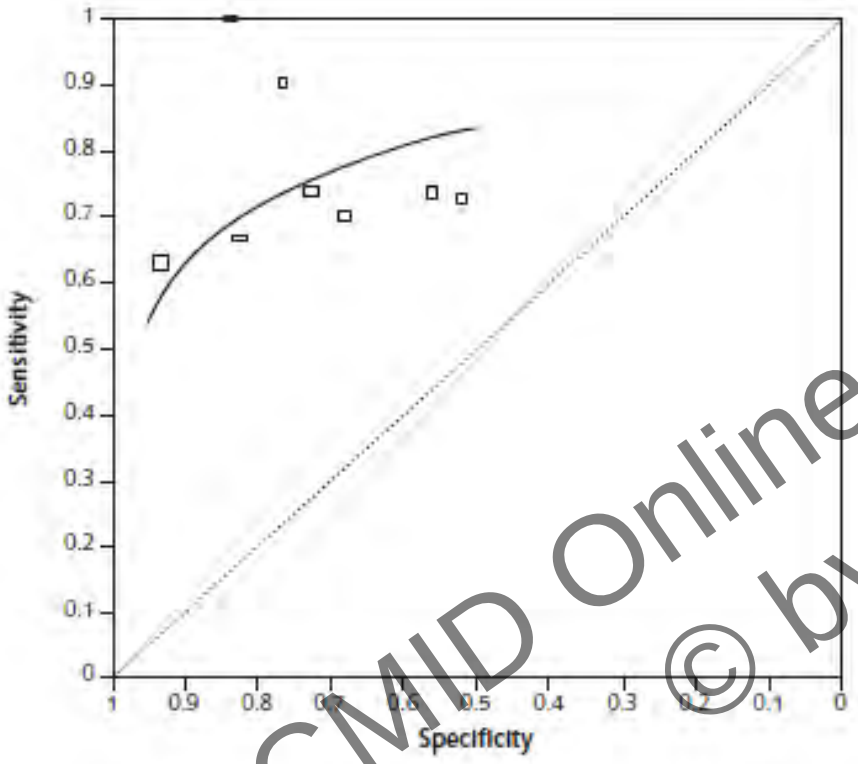
These models are needed to get a valid estimate of sensitivity and specicity

- Bivariate model
- Hierarchical SROC model

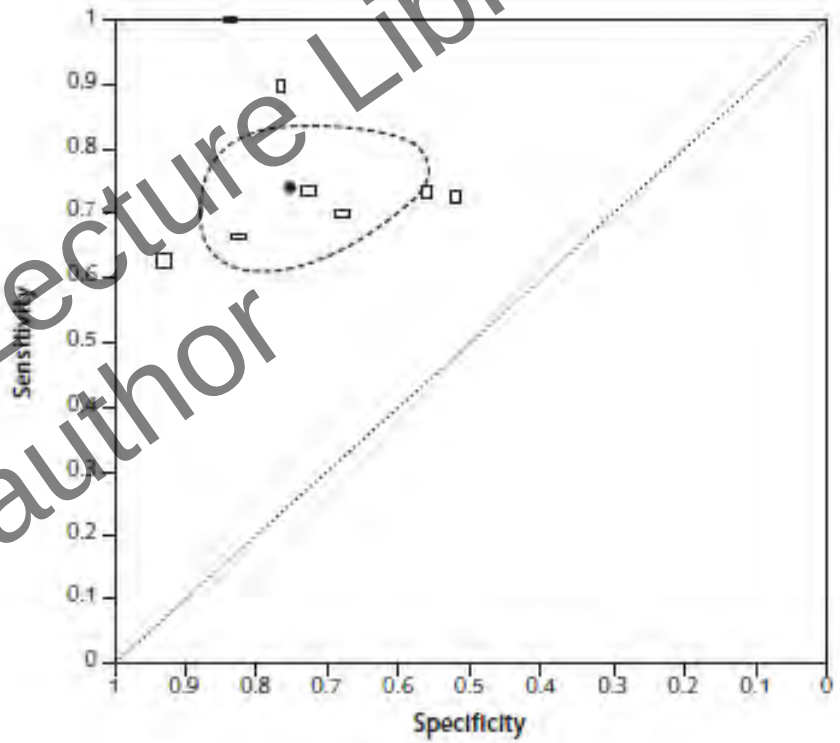
External statistical software is needed

- SAS, STATA, WinBugs, R

Meta-analysis: SROC plots

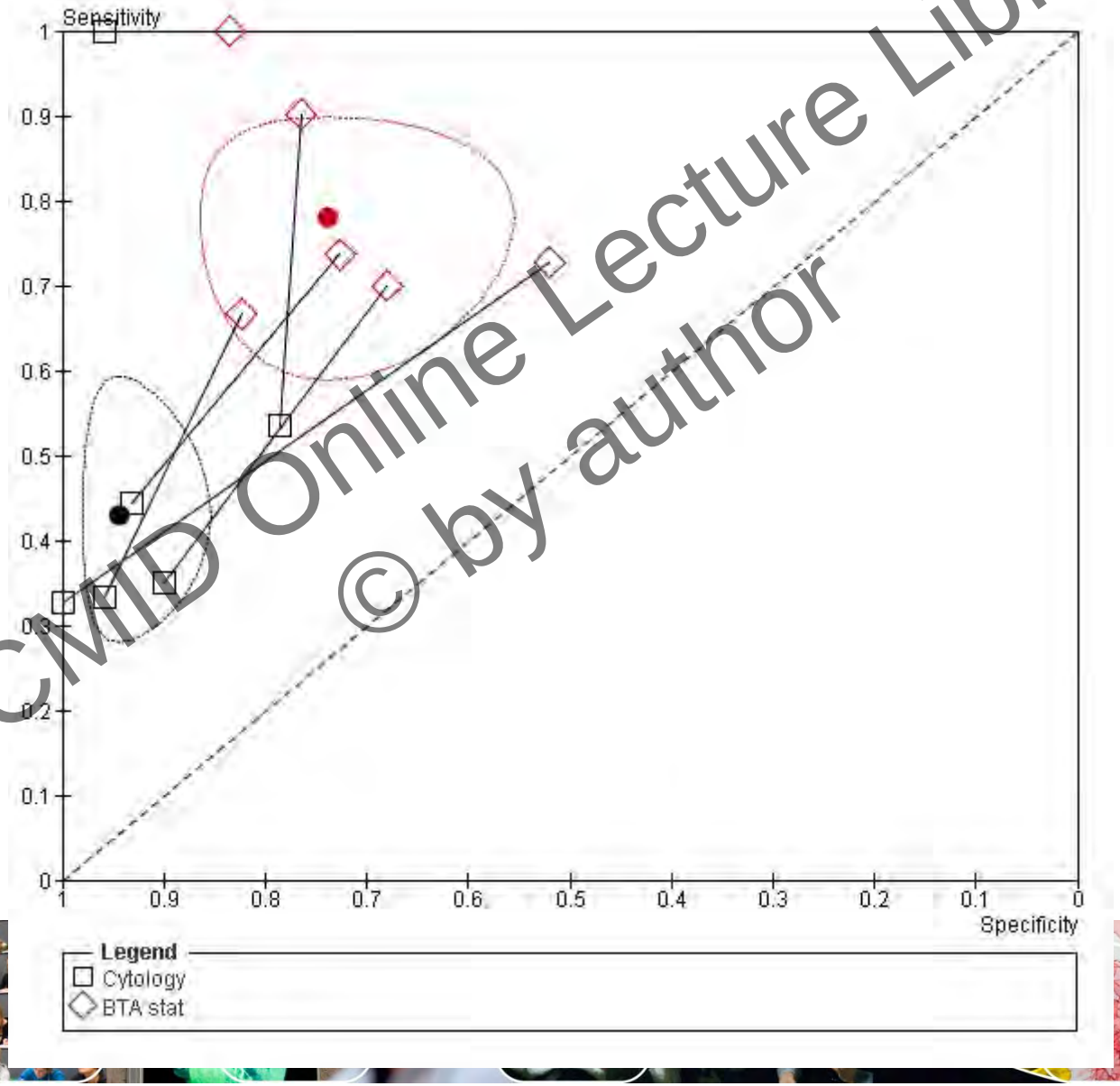


HSROC model
(curve)



Bivariate model
(point estimate + confidence region)

Comparative sROC plot



Steps of a Cochrane systematic review

1. define the review question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Interpretation / Conclusions

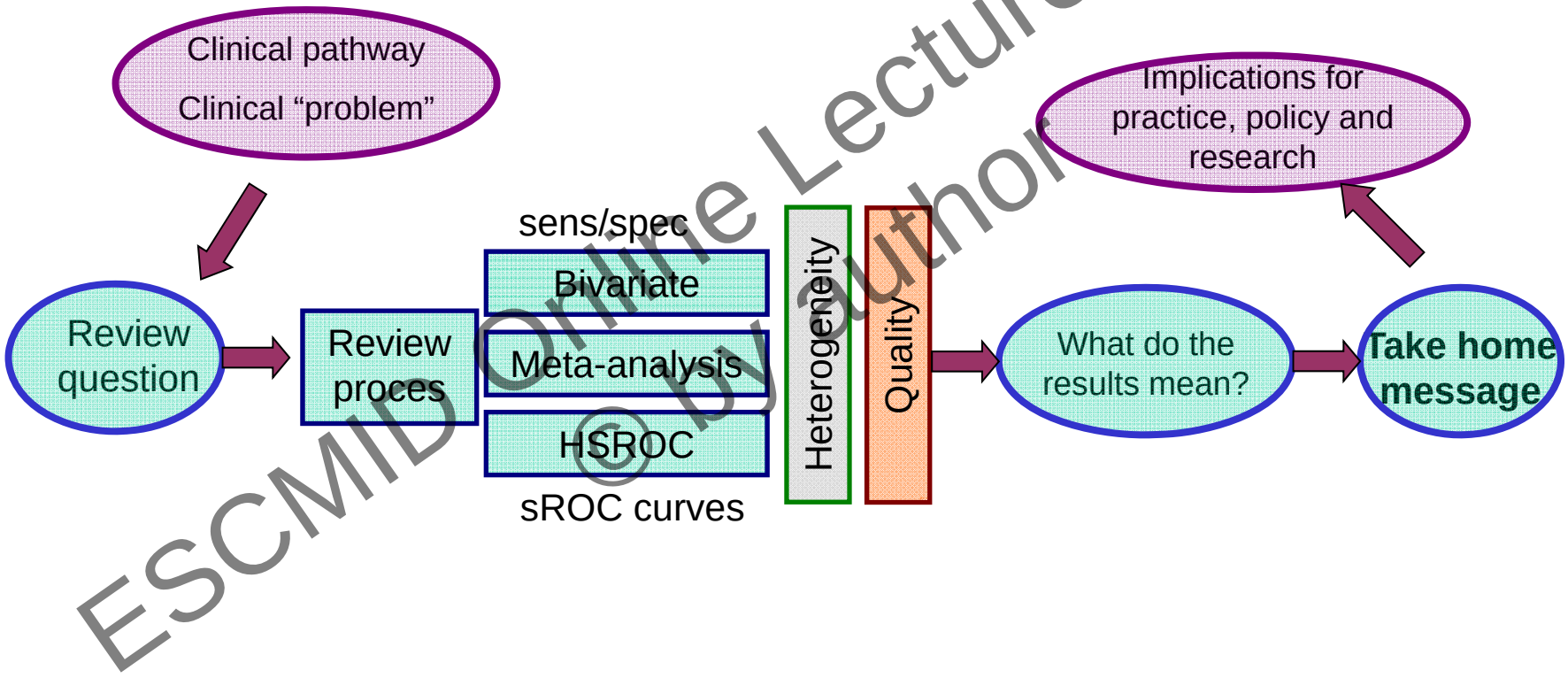
Even a perfect test can be worthless

Sensitivity and specificity are difficult concepts

Rather DOR, Predictive values or LRs?

Conseqences of testing: what will be the next steps?

Interpreting the results of a DTA review



Clinical pathway…

All immunocompromized patients

OR

All immunocompromized patients with fever

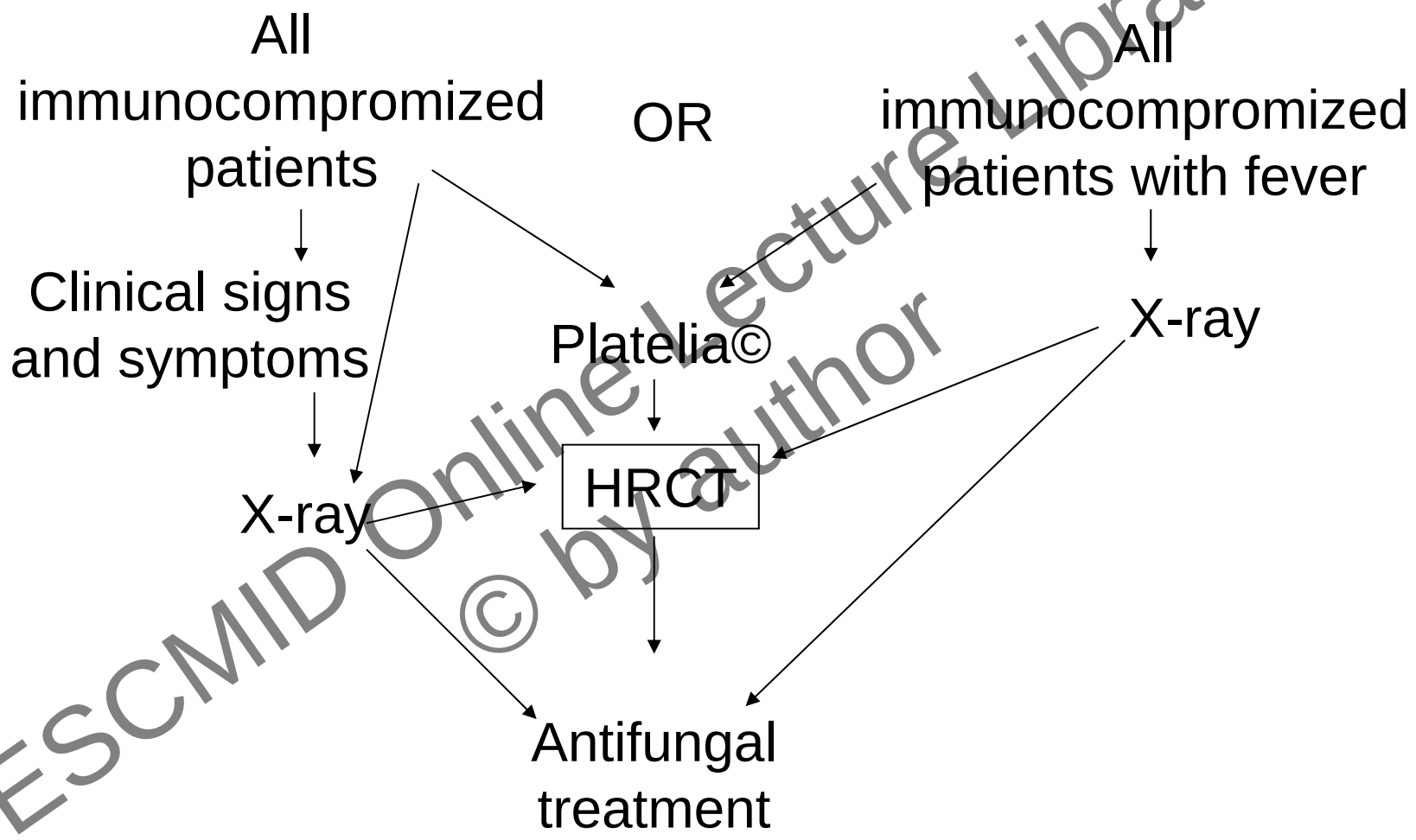
Platelia©

pos

neg

Antifungal treatment

Clinical pathway…



Platelia is done: (1) once a week; (2) twice a week; (3) only if there is a suspicion of IA.

Implications for practice

The value of the galactomannan test will depend on the role that the results of this test will play in clinical decisions about starting therapy for aspergillosis.

Will it be positive before signs and symptoms become suggestive?

If GM test positive, will HRCT be positive as well (or will HRCT become positive at a later stage)?

Will GM positive patients profit from treatment?

Leeflang, Cochrane Library, 2008

Steps of a Cochrane systematic review

1. define the review question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Common “mistakes”

- Role of index test not clearly stated
- Too many index tests involved
- RCTs to be included (-> intervention review)
- QUADAS items not operationalised
- Out-of-date statistical methods
- Testing for heterogeneity or for publication bias

Make sure that the author team is capable of doing a DTA review

Summary

Diagnostic test accuracy is only one piece of the puzzle called evidence for diagnostics

Consequences of testing are important

→ Clinical pathway!

Can we please stop interpreting case-healthy control designs as if they are representative for practice?

More info on: http://srdta.cochrane.org