NIPH
Norwegian Institute of Public Health

The reality is more than effectivness research

Challenges and possibilities for systematic reviews and meta-analysis

Research questions and design

We have all sorts of questions

- You work with children with Attention Deficit Hyperactivity Disorder (ADHD)
- Your boss suggests summarizing relevant evidence in a systematic review
- You argue his question is rather vague
- What kind of questions might be relevant?

Research questions

Different questions, different needs

Core topic	Question	Preferred design
Efficacy	What should we to to prevent or treat this problem	RCT
Causes/ risk factors	Why do some people exeperience this problem	Cohort/ case control
Prevalence	How many people experience this problem?	Cross sectional
Prognosis	How are these people doing in the long run?	Cohort
Diagnosis	Who has a problem?	Cross sectional/ RCT
Experiences	How do you experience living with the problem	Qualitative

The road to a systematic review

GRADE

Analyse/ compile

Data extraction

Risk of bias

Selection

Search

Question

Efficacy

Systematic reviews about effect of intervention are most common

- **Design:** Randomised controlled trials provide best evidence
- **Question:** PICO
- **RoB:** Traditional risk of bias assessment
- **Analysis:** Meta-analysis compare groups (OR, RR, MD, SMD)

- **Challenges** will arise if we choose to include observational studies:
  - Risk of bias, e.g. ROBINS-I tool*
  - Analysis, e.g. adjusting for confounding

*Risk of Bias In Non-randomised Studies of Interventions
https://sites.google.com/site/riskofbiastool/welcome/home

NIPH
Norwegian Institute of Public Health

Risk factors and causes

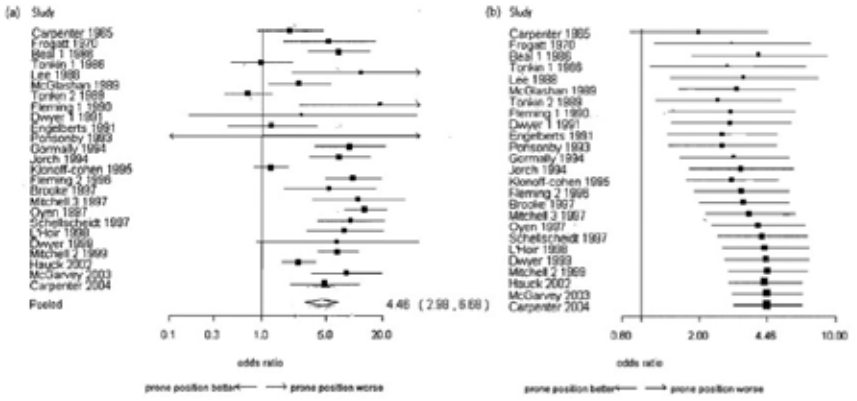
Why does it happen?

Risk factors and causes

- **Design:** Cohort or case control studies
- **Question:** PECO
- **RoB:** Additional questions regarding selection of participants, matching and adjustment for confounding to assess risk of bias adequately*
- **Analysis:** Paired comparison in exposed versus unexposed individuals

* E.g. ROBIN-S, Newcastle-Ottawa Scale

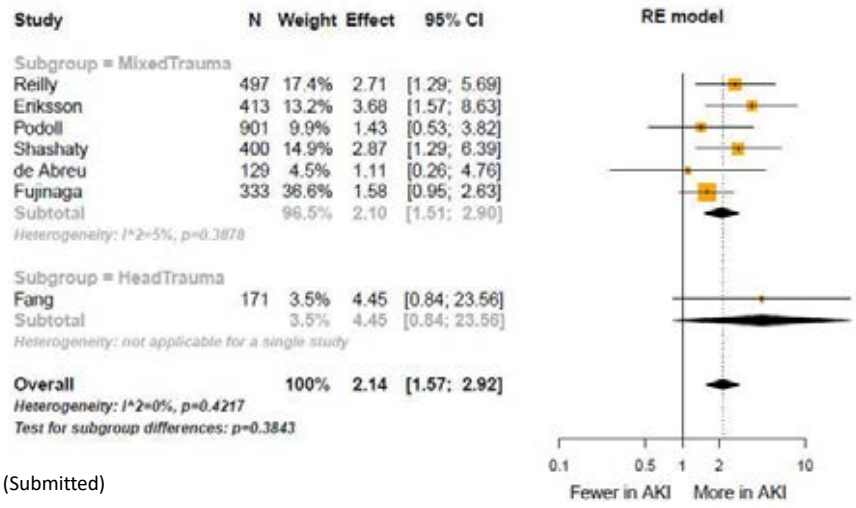
Meta-analyses of case control studies



Gilbert et at, Infant sleeping position and the sudden infant death syndrom, International journal of epidemiology (2005) 34:874-87

Diabetes and risk of acute kidney injury (AKI)

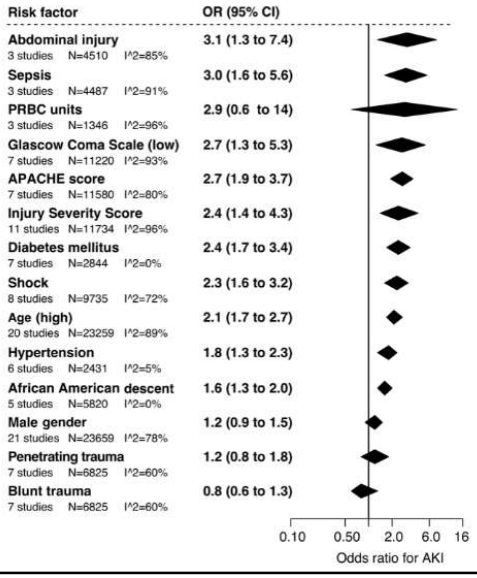
Patient admitted to intensive care unit



(Submitted)

Risk factor associated with acute kidney injury

Summarized in a super forest plot



NIPH
Norwegian Institute of Public Health

Prognosis

How can you expect this to affect you?

Prognostic reviews

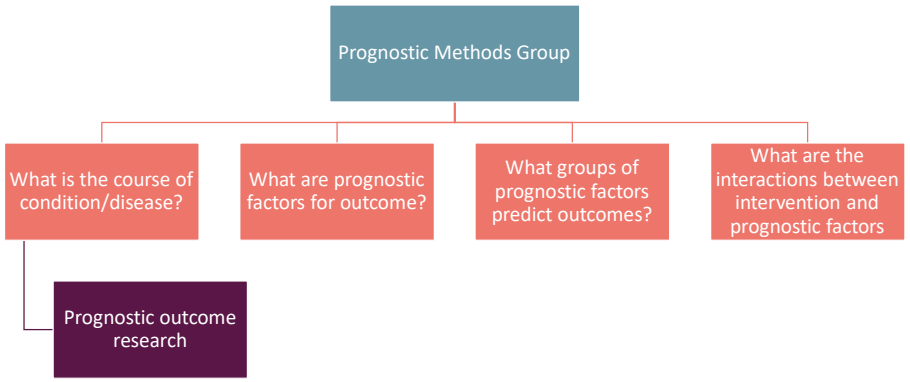
Information about prognosis is important for patients, stratification and informed decission making

- **Design**: Often cohort studies
- **Question**: PO/PECO/PICO
- **RoB**: Lack agreed standards for reporting and critical appraisal
- **Analysis**: Large variability in the use of analytic methods
- **Challenges** are many, but there is progress
  - Studies often small and imprecise
  - Many primary studies AND they are harder to to find
  - Many studies suffer from poor methodological quality

Altman DG, Systematic reviews of evaluations of prognostic variables, BMJ (2001), 323;224-8

Prognostic studies

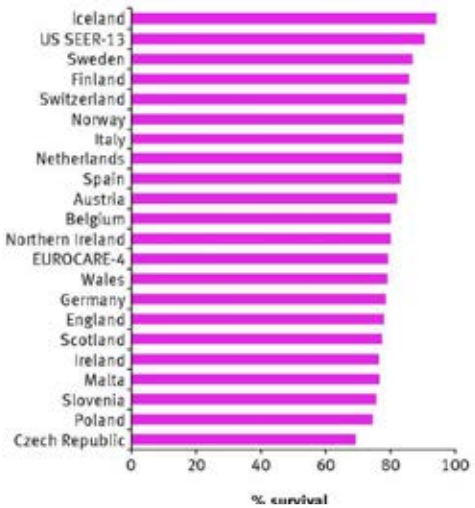
Different purposes



http://prognosismethods.cochrane.org

Fundamental prognosis research

What is the course of condition/disease?

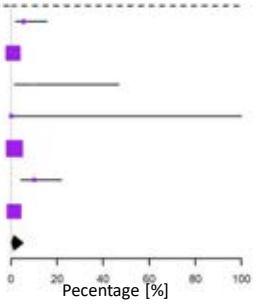


Hemingway et al, Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes, *BMJ* 2013;346:e5595 doi: 10.1136/bmj.e5595

Survival without impairment

Extremely premature infants (GA22)

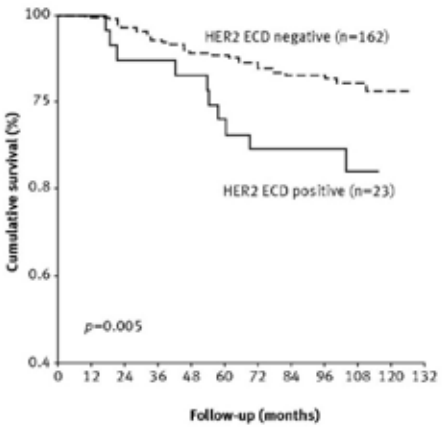
Study	n/N	Percentage (95% CI)	Weight
Zayek (E1)	3/55	5.5 (1.8-15.6)	6
Younge (E1)	2/241	0.8 (0.2-3.3)	26.3
Kutz	1/10	10 (1.4-46.7)	1.1
Serenius	0/51	0 (0-100)	5.6
Younge (E2)	4/274	1.5 (0.5-3.8)	29.9
Zayek (E2)	5/50	10 (4.2-21.9)	5.5
Younge (E3)	3/234	1.3 (0.4-3.9)	25.6
Total	**18/915**	**1.2 (0.3-5.2)**	**100**



- Many small an imprecise studies
- No comparison of groups
- Non-parametric data (asymmetry)
- Need for transformation to pool

Myrhaug et al, Pediatrics (in press)

Pognostic factor research



Hemingway et al, Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes, *BMJ* 2013;346:e5595 doi: 10.1136/bmj.e5595

Groups of prognostic factors

Table 2. Astro-CHARM Multivariable Predictors of Atherosclerotic Cardiovascular Disease

Parameter	β Coefficient	χ^2	*P* Value	Hazard Ratio	95% CIs	
					Lower	Upper
Age	0.019227	3.7	0.06	1.2*	1.00	1.34
Male sex	0.514818	14.8	<0.001	1.7	1.3	2.2
Race						
Black	0.289896	3.4	0.06	1.3	0.98	1.8
Hispanic	0.319984	2.6	0.10	1.4	0.94	2.0
Other	−0.03008	0.01	0.91	0.97	0.55	1.7
Total cholesterol	0.000405	0.09	0.8	1.01*	0.92	1.12
High-density lipoprotein cholesterol	−0.00407	0.66	0.4	0.94*	0.81	1.1
Systolic blood pressure	0.019908	40.1	<0.001	1.4*	1.3	1.6
Hypertension medication	0.073609	0.3	0.6	1.1	0.8	1.4
Smoking	0.797946	35.0	<0.0001	2.2	1.7	2.9
Diabetes mellitus	0.866738	28.4	<0.001	2.4	1.7	3.3
Family history myocardial infarction	0.46861	11.1	<0.001	1.6	1.2	2.1
High-sensitivity C-reactive protein	0.022105	6.9	0.009	1.1*	1.0	1.2
Coronary artery calcium score, natural log	0.026688	63.4	<0.001	1.5*	1.4	1.7

*Hazard ratio per 1 SD unit of continuous predictor variable; SD for age=7.5; total cholesterol=37.5; high-density lipoprotein cholesterol=14.7; systolic blood pressure=17.5; high-sensitivity C-reactive protein=4.8; and coronary artery calcium score, natural log=1.95.
Astro-CHARM indicates Astronaut Cardiovascular Health and Risk Modification.

Circulation 2018;138:1819–1827. DOI: 10.1161/CIRCULATIONAHA.118.033505

Groups of prognostic factors

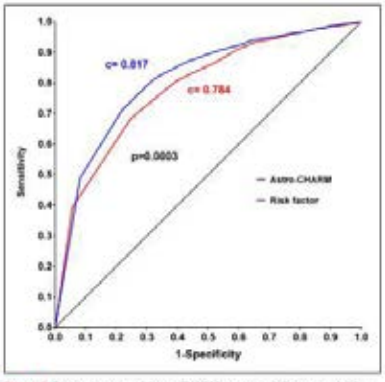
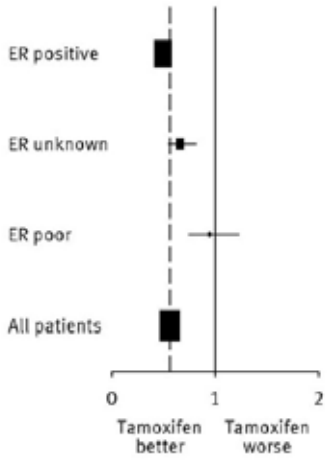


Figure 1. Discrimination of Astro-CHARM versus risk factor model for ASCVD.
The areas under the receiver operating curves for prediction of atherosclerotic cardiovascular disease (ASCVD) events are presented for the Astro-CHARM and risk factor–only models, with significant improvement using Astro-CHARM (*P*=0.0003).

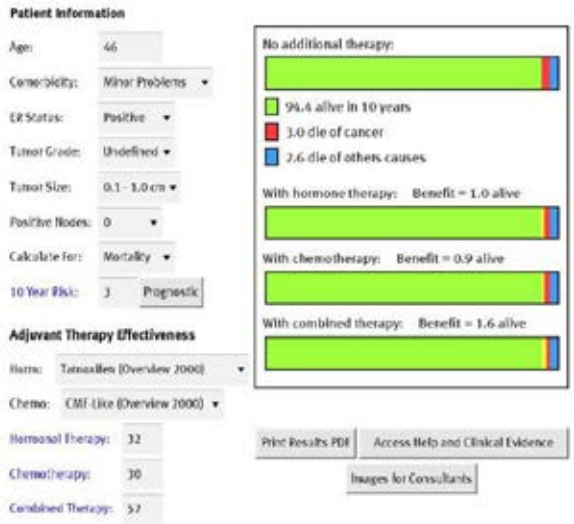
Circulation 2018;138:1819–1827. DOI: 10.1161/CIRCULATIONAHA.118.033505

Stratified medicine research



Hemingway et al, Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes, *BMJ* 2013;346:e5595 doi: 10.1136/bmj.e5595

Prognostic model research



Hemingway et al, Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes, *BMJ* 2013;346:e5595 doi: 10.1136/bmj.e5595

Risk of Bias: QUIPS

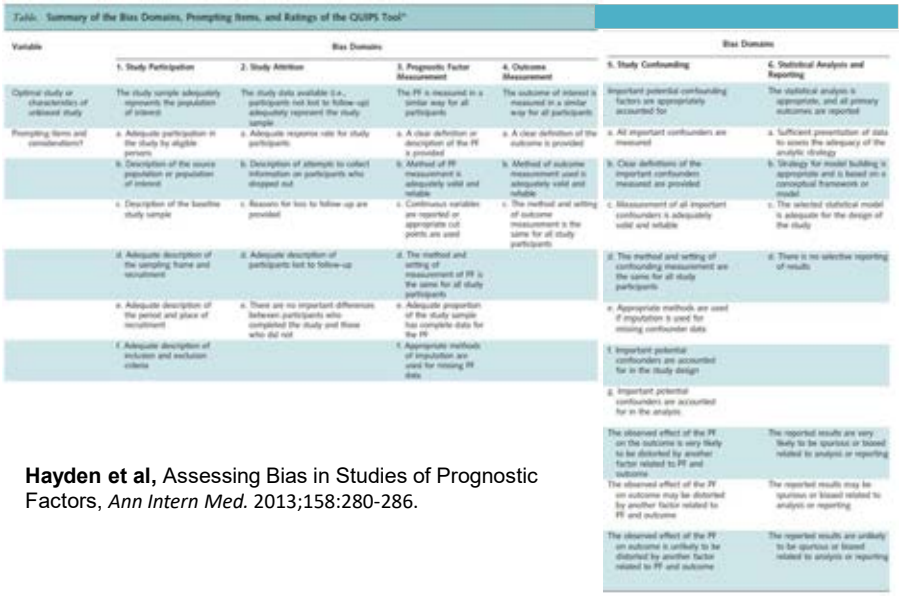


Table. Summary of the Bias Domains, Prompting Items, and Ratings of the QUIPS Tool*

Variable	Bias Domains					
	1. Study Participation	2. Study Attrition	3. Prognostic Factor Measurement	4. Outcome Measurement	5. Study Confounding	6. Statistical Analysis and Reporting
Optimal study or characteristics of unbiased study	The study sample adequately represents the population of interest	The study data available (i.e., participants not lost to follow-up) adequately represent the study sample	The PF is measured in a similar way for all participants	The outcome of interest is measured in a similar way for all participants	Important potential confounding factors are appropriately accounted for	The statistical analysis is appropriate, and all primary outcomes are reported
Prompting items and considerations†	a. Adequate participation in the study by eligible persons	a. Adequate response rate for study participants	a. A clear definition or description of the PF is provided	a. A clear definition of the outcome is provided	a. All important confounders are measured	a. Sufficient presentation of data to assess the adequacy of the analytic strategy
	b. Description of the source population or population of interest	b. Description of attempts to collect information on participants who dropped out	b. Method of PF measurement is adequately valid and reliable	b. Method of outcome measurement used is adequately valid and reliable	b. Clear definitions of the important confounders measured are provided	b. Strategy for model building is appropriate and is based on a conceptual framework or model
	c. Description of the baseline study sample	c. Reasons for loss to follow-up are provided	c. Continuous variables are reported or appropriate cut points are used	c. The method and setting of outcome measurement is the same for all study participants	c. Measurement of all important confounders is adequately valid and reliable	c. The selected statistical model is adequate for the design of the study
	d. Adequate description of the sampling frame and recruitment	d. Adequate description of participants lost to follow-up	d. The method and setting of measurement of PF is the same for all study participants		d. The method and setting of confounding measurement are the same for all study participants	d. There is no selective reporting of results
	e. Adequate description of the period and place of recruitment	e. There are no important differences between participants who completed the study and those who did not	e. Adequate proportion of the study sample has complete data for the PF		e. Appropriate methods are used if imputation is used for missing confounder data	
	f. Adequate description of inclusion and exclusion criteria		f. Appropriate methods of imputation are used for missing PF data		f. Important potential confounders are accounted for in the study design	
					g. Important potential confounders are accounted for in the analysis	
					The observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome	The reported results are very likely to be spurious or biased related to analysis or reporting
					The observed effect of the PF on outcome may be distorted by another factor related to PF and outcome	The reported results may be spurious or biased related to analysis or reporting
					The observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome	The reported results are unlikely to be spurious or biased related to analysis or reporting

Hayden et al, Assessing Bias in Studies of Prognostic Factors, *Ann Intern Med.* 2013;158:280-286.

NIPH
Norwegian Institute of Public Health

Prevalence

Prevalence

It is important to know how frequent a problem is

- **Design**: cross sectional studies
- **Analysis:**
  - Prevalence estimates will often vary between studies
  - Normal distribution is not always to be expected
  - Systematic reviews useful to put numbers in perspectives and explore why estimates vary
- **Risk of bias**
  - *How are responses collected?*
  - *Are questionnaires validated?*
  - *How many people responded?*
  - *How to deal with 'non-response'?*

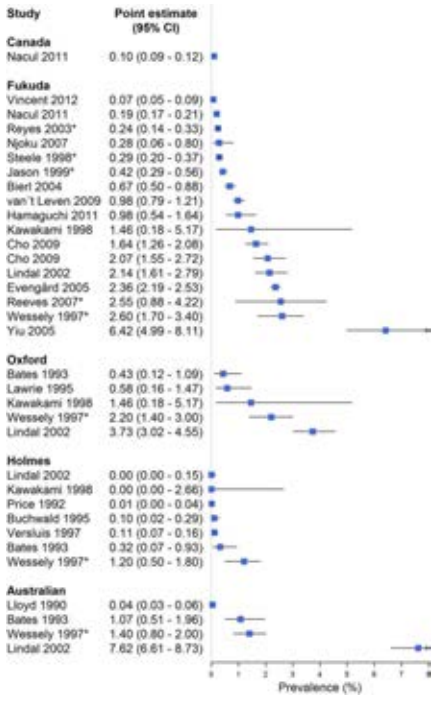
Hoy D. et al, Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement, Journal of clinical epidemiology (2012), 65 (9): 934-9

Prevalence of CFS/ME

Large variability – why?

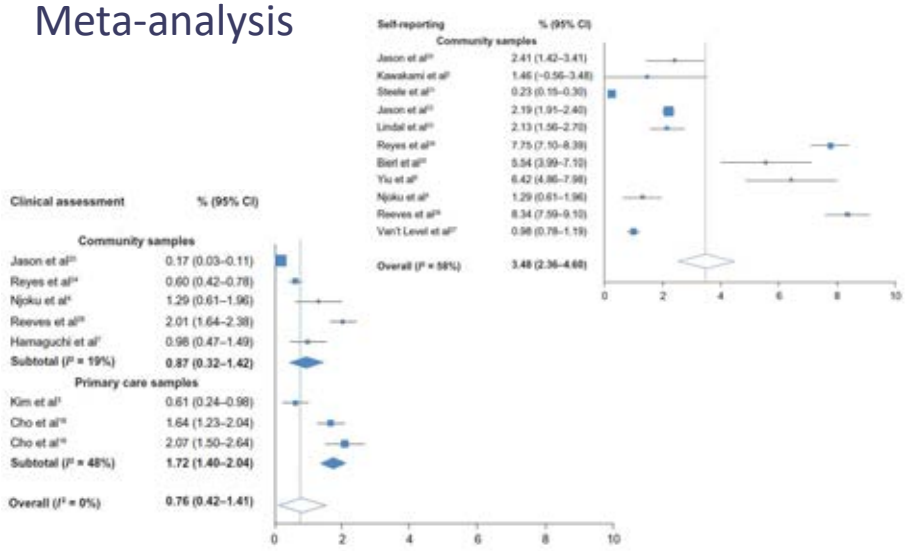
Many differences in:

- Diagnostic criteria
- Recruitment strategies
- Non-responders
- Application of criteria



BMJ Open 2014;4:e003973

Meta-analysis



Johnston et al, The prevalence of chronic fatigue syndrome/myalgic encephalomyelitis: a meta-analysis, Clinical Epidemiology 2013:5 105–110

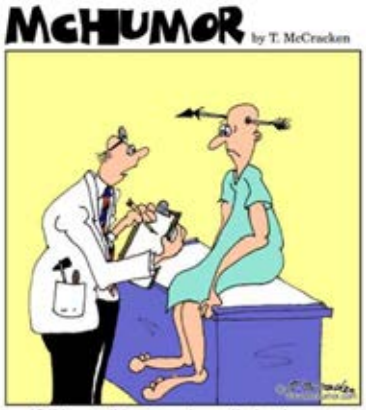
NIPH
Norwegian Institute of Public Health

Diagnostic tests

Research and reviews

What do we use diagnostic tests for?

- Assess the likelihood for absence or presence of a disease
- Inform about seriousness and prognosis
- Monitoring disease progression
- Treatment planning
- Buy time… (by postponing decissions)



"Off hand, I'd say you're suffering from an arrow through your head, but just to play it safe, I'm ordering a bunch of tests."

NIPH
Norwegian Institute of Public Health

Should you trust this test?

Validation against a reference standard

From PICO to PIRO

- P: Patient spectrum (setting)
- I: Index test
- R: Reference test (gold standard)
- O: Outcome

Diagnostic phase I studies

Does the test dicriminate between sick and healthy individuals?

Table 1 Answering a phase I question: do patients with left ventricular dysfunction have higher concentrations of B-type natriuretic peptide (BNP) precursor than normal individuals?

	Patients known to have disorder	Normal controls
Median (range) concentration of BNP precursor (pg/ml)	493.5 (248.9-909.0)	129.4 (53.6-159.7)

BMJ 2002; 324: 539-41

Diagnostic phase II studies

Is heart disease more likely in patients with certain test results?

Table 2 Answering a phase II question: are patients with higher concentrations of B-type natriuretic peptide (BNP) more likely to have left ventricular dysfunction than patients with lower concentrations?

	Patients known to have target disorder	Normal controls
High BNP concentration	39	2
Normal BNP concentration	1	25

Test characteristics (95% CI):
Sensitivity=98% (87% to 100%)
Specificity=92% (77% to 98%)

Good to go?

BMJ 2002; 324: 539-41

Diagnostic phase III studies

Does the test discriminate between patients with and without heart failure when applied to representative sample?

Table 3 Answering a phase III question: among patients in whom it is clinically sensible to suspect left ventricular dysfunction (LVD), does the concentration of B-type natriuretic peptide (BNP) distinguish patients with and without left ventricular dysfunction?

	Patients with LVD on echocardiography	Patients with normal results on echocardiography
Concentration of BNP:		
High (>17.9 pg/ml)	35	57
Normal (<18 pg/ml)	5	29
Prevalence (pretest probability) of LVD	40/126=32%	

Test characteristics (95% CI):
Sensitivity=88% (74% to 94%)
Specificity=34% (25% to 44%)

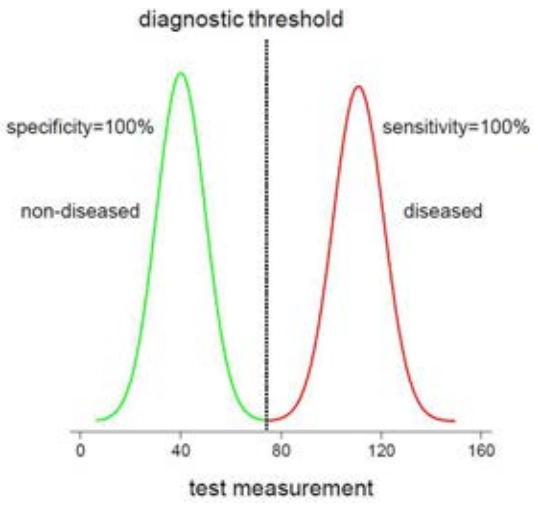
BMJ 2002; 324: 539-41

Test accuracy

What proportion of those with the disease does the test correctly identify? (sensitivity)

What proportion of those without the disease does the test correctly exclude? (specificity)

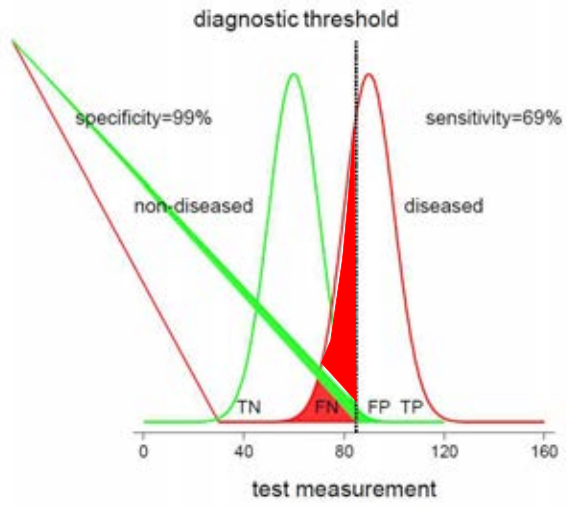
In the dreams of every diagnostic researcher



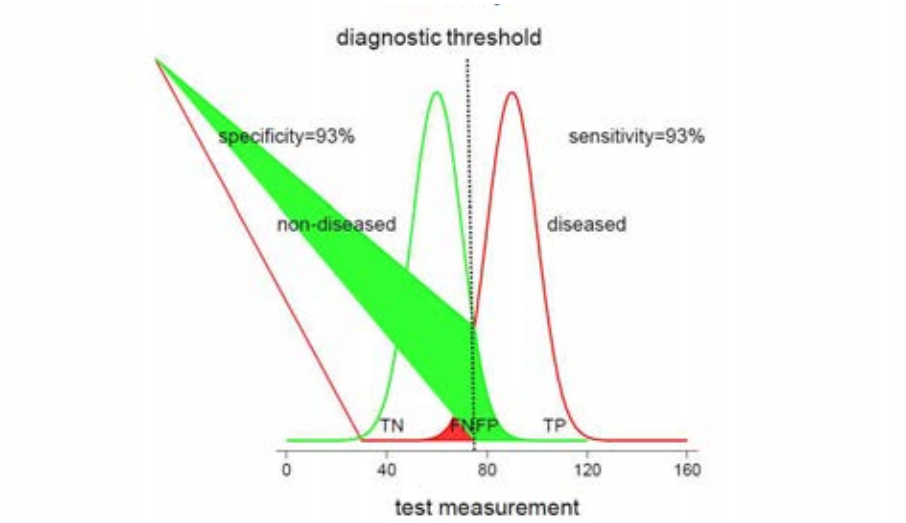
Reality I

- It is amazingly normal to be abnormal

Reality II



Reality III



Sensitivity and Specificity

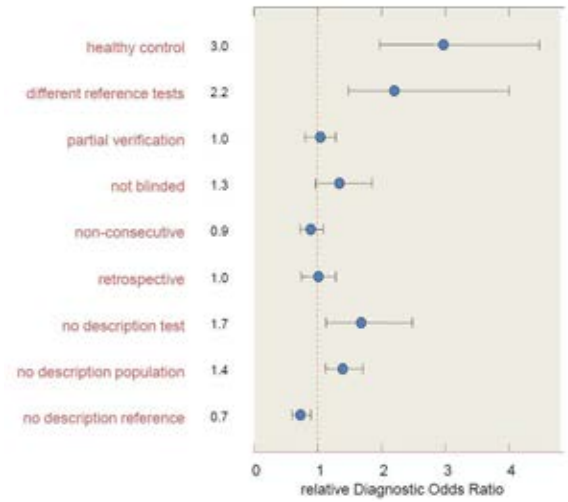
	Reference standard positive	Reference standard negative	
Index test positive	TP	FP	TP + FP
Index test negative	FN	TN	FN + TN
	TP + FN	FP + TN	TP +FN+FP+TN

Sensitivity = TP / (TP+FN)

Specificity = TN / (TN+FP)

kunnskapssenteret

Kilder til bias

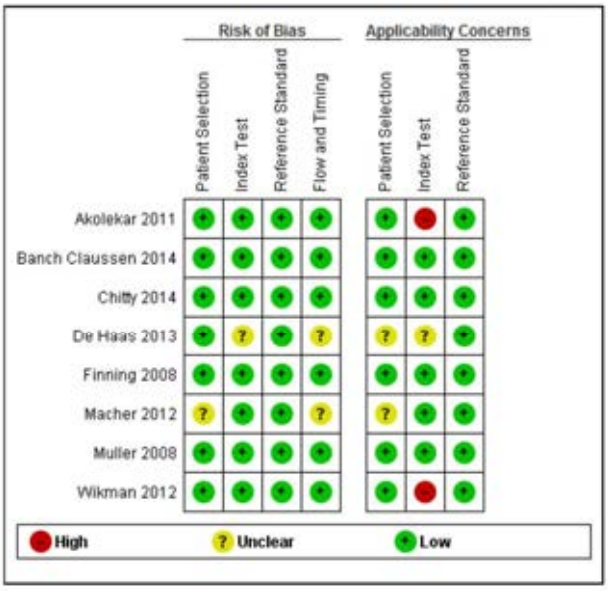


Lijmer et al, JAMA (1999) 282; 1061-1066

QUADAS

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
Description	Describe methods of patient selection Describe included patients (previous testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table (refer to flow diagram) Describe the interval and any interventions between index tests and the reference standard
Signaling questions (yes, no, or unclear)	Was a consecutive or random sample of patients enrolled? Was a case–control design avoided? Did the study avoid inappropriate exclusions?	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it prespecified?	Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?	Was there an appropriate interval between index test and reference standard? Did all patients receive a reference standard? Did all patients receive the same reference standard? Were all patients included in the analysis?
Risk of bias (high, low, or unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns about applicability (high, low, or unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Whiting P et al, QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies, Ann Intern Med (2011): 155: 529-36



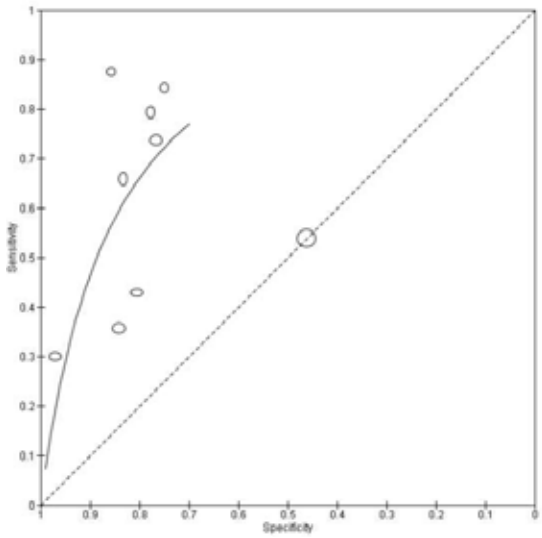
Figur 2 *Risiko for systematiske skjevheter og utfordringer med overførbarhet i de inkluderte studiene*

Challenges

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

HSROC analysis (curve estimate)

Stroke Driver Screening Assessment



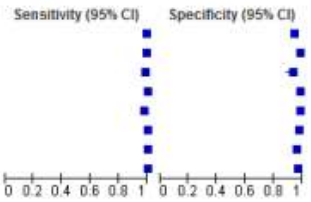
Smedslund et al. Screening tools for cognitive function and driving (2015)

Bivariate analysis (point estimate)

Rhesus typing based on fetal DNA

Sn 99,8% (99,7 to 99,9)
Sp 96,9% (95,5 to 97,8)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Finning 2008	1153	43	3	670	1.00 [0.99, 1.00]	0.94 [0.92, 0.96]
Muller 2008	657	6	2	357	1.00 [0.99, 1.00]	0.98 [0.96, 0.99]
Akolekar 2011	404	12	6	164	0.99 [0.97, 0.99]	0.93 [0.88, 0.96]
Macher 2012	619	7	0	386	1.00 [0.99, 1.00]	0.98 [0.96, 0.99]
Wikman 2012	2242	24	55	1331	0.98 [0.97, 0.98]	0.98 [0.97, 0.99]
De Haas 2013	11274	157	8	6944	1.00 [1.00, 1.00]	0.98 [0.97, 0.98]
Chitty 2014	935	31	0	696	1.00 [1.00, 1.00]	0.96 [0.94, 0.97]
Banch Claussen 2014	7819	132	11	4706	1.00 [1.00, 1.00]	0.97 [0.97, 0.98]



Arentz-Hansen et al Determination of fetal rhesus D status from maternal plasma of rhesus negative women (2014)

Predictive values

Add a new perspective

- Positive predictive value: For a person who tests positive, what is the the probability that the person is sick?
- Negative predictive value: For a person who tests negative, what is the probability that the person don't have the disease

Likelihood ratio

Yet another perspective

- Positive likelihoodratio (LR+): Sn/(1-Sp)
  - How much more likely is it that you get a positive test result in a patient with the disease compared to a healthy person?
- Negative likelihood ratio (LR-): (1-Sn)/Sp
  - How much more likely is it that you get a negative test result from a healthy person compared to a patient with the disease

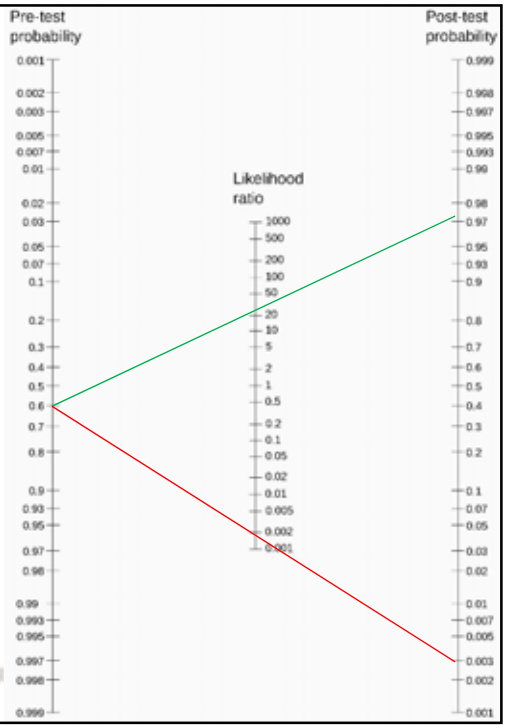
LR magic

Sensitivitet = 99,84 %

Spesifisitet = 96,86 %

Pretest = ~61 %

- LR+=99,84/(100-96,86)=31,8
- LR-=(100-38)/94=0.002



Diagnosis versus prognosis

1. Do we identify the right patients?
2. Does it matter, i.e. can we help them?
3. Do patients care about Sn and Sp?
4. How often do we have reference standard, anyway?

Croft P et al, The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. BMC Med 2015;13:20

Reitsma et al, A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard, Journal of clinical epidemiology (2009) 62 797-806

Diagnostic phase IV studies

Powerful and fun, but rare

- Does not require the existence of a reference
- Focus on patient centred outcomes
  - Mortality
  - Morbidity
  - Quality of life