Systematic reviews and meta-analyses of test accuracy: developing methods that meet practitioners' needs

by

Dr Clare Frances Davenport

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

[Volume 2]

Department of Public Health, Epidemiology and Biostatistics
School of Health and Population Sciences
College of Medical and Dental Sciences
University of Birmingham
Submitted April 2012

Volume 2 of 2

Appendices. The main text is in a separate file.

VOLUME II

Appendices

Appendices to chapter 2:

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010)

Appendix 2.2: Characteristics and results of included empirical test accuracy studies

Appendix 2.3: Characteristics and results of included empirical risk communication literature

Appendices to chapter 3:

Appendix 3.1: Pragmatic search filters created for use with HTA, DARE and ARIF databases

Appendix 3.2: Flow of references from the HTA, DARE, Medion, C-EBLM and ARIF databases

Appendix 3.3: Pro-forma for coding review references according to Title and Abstract

Appendix 3.4: Yield from searches of combinations of databases excluding primary research, research not concerned with test accuracy and duplicates

Appendix 3.5: Characteristics of specialist reviews databases

Appendices to chapter 4:

Appendix 4.1: Consistency rules for inclusion of reviews on the basis of test

Appendix 4.2: Demographic details of included reviews

Appendix 4.3: Details of review question formulation

Appendix 4.4: Reporting of review findings

Appendix 4.5: Use of outcome measures in reviews

Appendix 4.6: Detail of contextualisation of review findings

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation.

Appendices to chapter 5:

Appendix 5.1: Paper based version of questionnaire used for the survey of understanding and application of test accuracy information
Appendix 5.2: Details of web based resources cited as sources of test accuracy information by survey respondents (accessed 27-07-11)

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: Cochrane Library (Wiley) 2010 Issue 2 Methodology Database: Search strategy 1
Search Date: 5th May 2010

#1 understand*

#2 interpret*

#3 comprehen*

#4 inform*

#5 convey*

#6 access*

#7 select*

#8 choose

#9 choice*

#10 decision*

#11 order*

#12 present*

#13 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12)

#14 screen*

#15 result*

#16 test*

#17 diagnos*

#18 accuracy

#19 MeSH descriptor Sensitivity and Specificity explode all trees

#20 (#14 OR #15 OR #16 OR #17 OR #18 OR #19)

#21 (#13 AND #20)

#22 physician* or gp* or practitioner* or doctor* or consultant* or student* or decision* or policy or medic*

#23 (#21 AND #22)

5787 refs

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: Cochrane Library (Wiley) 2010 Issue 2 Methodology Database: Search strategy 2
Search date : 19th April 2010

#1 diagnos* near/2 test*

#2 diagnos* near/2 accura*

#3 laboratory next test*

#4 blood next test*

#5 (#1 OR #2 OR #3 OR #4)

#6 understand*

#7 interpret*

#8 comprehen*

#9 inform*

#10 convey*

#11 access*

#12 select*

#13 choose

#14 choice*

#15 decision*

#16 order*

#17 (#6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16)

#18 (#5 AND #17)

#19 clinician* or doctor* or physician* or medic* or practitioner*

#20 (#18 AND #19)

280 refs

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations
Search Date: 4th May 04 2010

1 (physician$ or gp$ or practitioner$ or doctor$ or consultant$ or student$ or decision-maker$ or policy-maker$ or student$ or medic$).mp.

2 (explain$ or convey$ or express$ or interpret$ or understand$ or present$ or inform$ or infer$ or assess$ or estimate$ or summary or summarise or decide$ or decision$).ti.

3 (diagnos$ or screen$ or test$ or accura$ or result$).ti.

4 1 and 2 and 3

5 from 4 keep 1-176 (176)

Database: Ovid MEDLINE(R)
Search Date: 1950 to April Week 3 2010

1 (physician$ or gp$ or practitioner$ or doctor$ or consultant$ or student$ or decision-maker$ or policy-maker$ or student$ or medic$).mp.

2 (explain$ or convey$ or express$ or interpret$ or understand$ or present$ or inform$ or infer$ or assess$ or estimate$ or summary or summarise or decide$ or decision$).ti.

3 (diagnos$ or screen$ or test$ or accura$ or result$).ti.

4 1 and 2 and 3

5 limit 4 to "therapy (sensitivity)" 1280

Database: EMBASE
Search Date: 1980 to 2010 Week 17

1 (physician$ or gp$ or practitioner$ or doctor$ or consultant$ or student$ or decision-maker$ or policy-maker$ or student$ or medic$).mp.

2 (explain$ or convey$ or express$ or interpret$ or understand$ or present$ or inform$ or infer$ or assess$ or estimate$ or summary or summarise or decide$ or decision$).ti.

3 (diagnos$ or screen$ or test$ or accura$ or result$).ti.

4 1 and 2 and 3

5 limit 4 to "treatment (2 or more terms high sensitivity)"

6 limit 4 to "treatment (1 term high sensitivity)"

7 from 6 keep 1-550 (550)

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: ERIC (CSA)
Search Date: 7th May 2010

Query: TI=(physician* or gp* or practitioner* or doctor* or consultant* or

student* or decision-maker* or policy-maker* or student* or medic*) and

TI=(explain* or convey* or express* or interpret* or understand* or

present* or inform* or infer* or assess* or estimate* or summary or

summarise or decide* or decision*) and TI=(diagnos* or screen* or test*

or accura* or result*)

545 refs

Database: ISI Proceedings Web of Science with conference proceedings
Search date: 5th May 2010

Title=(diagnostic or diagnosis or test or tests or accuracy or screening) AND Title=(understand* or comprehend* or interpret* or inform* or convey* or access* or choose or choice* or decision* or order*)

Refined by: Subject Areas=(MEDICINE, GENERAL & INTERNAL)

Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S.

1093 refs

Database: PsycINFO (Ovid) 1967 to April Week 2 2010
Search date : 5th May 2010

1 (physician$ or gp$ or practitioner$ or doctor$ or consultant$ or student$ or decision-maker$ or policy-maker$ or student$ or medic$).mp.

2 (explain$ or convey$ or express$ or interpret$ or understand$ or present$ or inform$ or infer$ or assess$ or estimate$ or summary or summarise or decide$ or decision$).ti.

3 (diagnos$ or screen$ or test$ or accura$ or result$).ti.

4 1 and 2 and 3

5 limit 4 to "0400 empirical study" (1910 refs)

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: ZETOC (British Library)
Search date: 30th April 2010

Diagnostic and clinician

Diagnostic and health professional

Diagnostic and policy

Understanding and diagnos* and risk

Understanding and diag* tests

Choosing and diag* tests

Comprehending and tests

43 refs

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: Ovid MEDLINE(R)
Search Date: 1950 to March Week 3 2007

1 diagnostic test$.mp. or exp Diagnostic Tests, Routine/

2 general practitioner$.mp. or exp Physicians, Family/

3 primary care.mp. or exp Primary Health Care/

4 doctor$.mp.

5 exp Family Practice/ or GP$.mp.

6 or/2-5

7 1 and 6

8 understand$.mp.

9 interpret$.mp.

10 or/8-9

11 7 and 10 (101)

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: Ovid MEDLINE(R)
Search Date: 1950 to April Week 1 2007

1 diagnostic test$.mp. or exp Diagnostic Tests, Routine/

2 "Laboratory Techniques and Procedures"/ or laboratory test$.mp.

3 (test$ adj3 order$).mp.

4 or/1-3

5 general practitioner$.mp. or exp Physicians, Family/

6 general practice.mp. or exp Family Practice/

7 primary care.mp. or exp Primary Health Care/

8 or/5-7

9 4 and 8

10 understand$.mp.

11 interpret$.mp.

12 information.mp.

13 exp "Predictive Value of Tests"/ or predictive value.mp.

14 sensitivity.mp. or exp "Sensitivity and Specificity"/

15 or/10-14

16 9 and 15

17 or/10-12

18 9 and 17

19 understand$.ti.

20 interpret$.ti.

21 inform$.ti.

22 or/19-21

23 16 and 22

24 from 23 keep 1-47

25 survey$.mp.

26 9 and 25

27 from 26 keep 1-297 (297)

28 from 25 keep 1-47 (47)

Database: Ovid MEDLINE(R)
Search Date: 1950 to March Week 3 2007

1 diagnostic test$.mp. or exp Diagnostic Tests, Routine/

2 general practitioner$.mp. or exp Physicians, Family/

3 primary care.mp. or exp Primary Health Care/

4 doctor$.mp.

5 exp Family Practice/ or GP$.mp.

6 or/2-5

7 1 and 6

8 survey$.mp.

9 order$.mp.

10 request$.mp.

11 usage.mp.

12 user$.mp.

13 questionnaire$.mp. or exp Questionnaires/

14 or/8-13

15 7 and 14

16 primary care.ti.

17 diagnostic test$.ti.

18 16 and 17

19 test$.ti.

20 16 and 19

21 15 or 18 or 20 (634)

24 or/21-23 (85998)

25 18 and 24 (47)

26 from 25 keep 1-47 (47)

27 survey$.mp. (234954)

28 11 and 27 (297)

29 from 28 keep 1-297 (297)

Appendix 2.1: Search strategies employed for test accuracy and risk communication literature reviews (2010; 2007; 2005)

Database: MEDLINE, CINAHL, EMBASE
Search Date: 2005 version

1 diagnostic test$.ti.

2 understand$.ti.

3 interpret$.ti.

4 convey$.ti.

5 information.ti.

6 accessib$.ti.

7 comprehen$.ti.

8 inform$.ti.

9 or/2-8

10 1 and 9 (130)

11 remove duplicates from 10 (94)

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Berwick 1981[(ETA2)] USA -Medical undergraduates -1y care (practising and academic) - 2y care (practising and academic)	-Convenience sample, self selected. -Hypothetical -Cross-sectional study. Results for a minority of questions presented. All questions MCQs -N=281	-2x2 (U) D&U, % or formula: -False –ve rate (defined as1-sensitivity) -False +ve rate (defined as1-specificity) -Sensitivity -Specificity **-FINDINGS:** Performance worse for diagnostic compared to effectiveness terms. Lack of consensus over definitions of false +ve rate and false-ve rate (confused with 1-PPV and 1-NPV respectively). 'High levels of error' for definitions of sensitivity and specificity. Practising clinicians performed less well than academic clinicians and medical undergraduates. Performance inversely correlated with yrs since training.			**-PPV** (semi-quantitative %), given prevalence AND false +ve rate AND false –ve rate). **-FINDINGS:** Correct responses: 32% (range 24% practising clinicians; 33% medical students / house officers; 73% academic clinicians). Base rate neglect Inverse association between years since qualification and correct responses within practising clinician group.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Borak 1982[ETA3] USA -1y academic clinicians -2y practising clinicians and nurses -Hospital clerical and maintenance workers	-Purposive sample, self-selected. -Controlled trial -Quality assessment precluded by poor reporting -Hypothetical -169	-Sensitivity (%) (U) -Specificity (%) (U)			**-PPV** (%), given prevalence AND sensitivity AND specificity. **-Findings:** Base rate neglect Correct responses: 34% academic 1y care; < 2% other groups
Cahan 2003[ETA5] Israel -2y care clinicians	-Convenience, self-selected (response rate 64%) -Cross sectional -Hypothetical -N=84		-'Pre-test' probability semi-quantitative (%) of ≥ 5 competing diagnoses given clinical examination and ECG findings. **-Findings:** Severe, less probable diagnoses overestimated (availability heuristic). 65% respondents sum of probabilities > 100%; 20% respondents sum of probabilities < 100%.		

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Casscells 1978[ETA6] USA -Medical undergraduates -2y care clinicians	-Opportunistic, self-selected (response rate 89%) -Hypothetical -Cross sectional -N=60	-False +ve rate (1-specificity) (U) (%)			**-PPV** (%), given prevalence AND false +ve rate) **-Findings:** Base rate neglect. Correct responses: 18% Inverse association between years since qualification and correct responses.
Christensen-Szalanski 1983[ETA7] USA -Army clinicians (general outpatient setting)	-Recruitment unclear -Ecological -Cross-sectional study -N=9				**-PPV** (semi-quantitative %) following clinical history and examination. No pre-test probability or test accuracy information provided. **-Findings:** Positive but not negative test results were used to estimate post-test probability. Authors suggest 'insensitvity' to negative results due to emphasis on presence of symptoms in clinical encounters and epidemiological terminology (NPV= absence of disease).

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Curley 1990[ETA9] USA (1) -Medical undergraduates -2y care clinicians	-Recruitment unclear -Hypothetical Cross sectional study -N=52	-Sensitivity (verbal description) (U) -Specificity (verbal description) (U) **-Findings:** Sensitivity and specificity formulae correctly identified by 98% respondents.			
Dolan 1986[ETA11] USA -2y care clinicians	-Convenience sample, self-selected. Response rate 71% Cross sectional study -Hypothetical -N=104		-'Pre-test' probability estimation (semi-quantitative %) given clinical history only. **-Findings:** Mean overestimation in 6/7 scenarios. Pre-test probability estimates varied by 80-90% No association between years since qualification and accuracy or variability of estimates.		**PPV** estimation (semi-quantitative %) based on respondent's own pre-test probability estimation and a test result (+ve or –ve). **-Findings**: Test result modified estimated disease probability in 3/7 scenarios. Confirmatory bias. No association between years since qualification and degree of probability revision.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Egglin 1996[ETA12] USA -2y care clinicians	-Recruitment unclear -Controlled trial -Hypothetical -Poor reporting precluded assessment of selection or attrition bias. Participants blind to study objectives. Standardised outcome measurement. Uncertain diagnoses included in analysis -N=6		-Investigation of varying pre-test probability (inferred from abnormal CXR) on sensitivity and specificity. **-Findings:** Pre-test probability of target disorder positively correlated with changes in sensitivity (based on subjective CXR interpretation) as a result of re-classification of indeterminate CXR results. No association of specificity with pre-test probability.		
Gigerenzer 1998[ETA14] USA -Public health clinicians -Social workers	-Purposive sample -Qualitative covert observational study -Ecological N=20	-Sensitivity (D,E & U) -Specificity (D, E&U) -False+ve rate (1-specificity) (D&U). -False –ve rate (1-sensitivity) (all open) **-Findings:** 15/20 respondents identified relationship between sensitivity and false –ves; 6/20 respondents between specificity and false+ves	-Pre-test probability (prevalence) estimation (open). **-Findings:** Estimates of pre-test probability varied between 0.0075-6.0 (%)		**-PPV** (semi-quantitative, open), based on respondents own pre-test probability, sensitivity and specificity estimation. **Findings:** 5/20 respondents provided reasonable (%) estimates of the PPV of the HIV test. Reference class confusion (sensitivity confused with PPV).

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Heller 2004(ETA18) UK -1y care clinicians -2y care clinicians -Clinical post-graduate examiners	-Random sample from professional registers and convenience sample for examiners (response rate 56% first stage and 29% second stage).41% of 2y care had a research qualification compared to 7% GPs. (? examiners). -Hypothetical -Cross sectional study -N =3102 first stage and 154 second stage.	Second stage: -Effect of prevalence on false +ves (1-specificity) (verbal description of direction of change) -Effect of prevalence on false –ves (1-sensitivity) (verbal description of direction of change. **-Findings:** Correct estimation of direction of test errors: Post-graduate examiners (64-71%) AND 2y care (67-72%)> 1y care (39-41%).	First stage: -'Pre-test' probability estimation (%) given clinical history. **-Findings:** Overestimation of pre-test probability compared to literature based estimates. Estimates varied widely from 5-100% and did not appear to differ between groups.		
Hoffrage 1998(ETA20) Germany -2y care	-Recruitment unclear -Cross sectional study -Hypothetical -N = 48	-Sensitivity (U) -False –ve rate (defined as 1-specificity) (U) (probabilities versus natural frequencies)			**-PPV** (semi-quantitative % or natural frequency), given prevalence AND sensitivity AND false +ve rate **-Findings:** Base rate neglect with probabilities Test accuracy neglect with natural frequencies Correct: 10% for probabilities; 46% for natural frequencies.
HEALTH PROFESSIONALS					

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

Study ID	-Design, Quality -External validity -Hypothetical / Ecological	FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases		FULLY RATIONAL Motivational biases	ABLE TO COMPUTE ACCURATELY Cognitive biases
		Test accuracy	**Pre-test probability**		
Houben 2010 (ETA21) Netherlands -1y care	-Recruitment unclear -Cohort study -Ecological -N=87		-'Pre-test' probability (verbal) given clinical history and examination results. **-Findings:** 43% 'definitely or 'probably' no disease 29% 'maybe' disease 28% 'probably' or 'definitely' disease. This reflects general practice being a low prevalence setting.	-Reason for ordering test from 5 possible categorical variables: 1) exclude disease (reduce clinicians uncertainty); 2) confirm diagnosis and determine treatment; 3) reassure patient / at patient's request; 4) screen or monitor; 5) other. -Findings: 20% to exclude disease and reassure patient 62% to exclude disease and reassure physician -19% confirm diagnosis and commence treatment	**-PPV** estimation (verbal) (based on respondent's own verbal pre-test probability estimation and a +ve or –ve test result in the absence of test accuracy information. **-Findings:** Estimation of post-test probability 'low' for 50% of patients concurring with pre-test probability estimates suggesting confirmatory bias (only 8.8% of abnormal results resulted in further investigations).

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Lyman 1994(a) (ETA22) USA *Journal of Cancer Education* -2y care clinicians -Nurses -Pharmacists -Medical undergraduates	-Recruitment unclear -Hypothetical -Cross sectional study -N=50	-Sensitivity (open) (E &U) -Specificity (open) (E&U) **-Findings:** The value of +ve and -ve test results both overestimated.	-'Pre-test' probability estimation (%) provided clinical history and examination findings. Age was varied in the 2 scenarios presented. **-Findings:** Estimates of pre-test probability appropriately increased with age. Mean estimates of pre-test probability did not show large variation (18%-20% younger age and 50-59% older age. The magnitude of between individual estimates was not reported.		**-PPV and 1-NPV** (%) (based on respondents own pre-test probability , sensitivity and specificity estimation and a +ve or –ve test result in the absence of test accuracy information). **-Findings:** Correct direction of change in probability revision for +ve test result 47/50 respondents. Correct direction of change in probability revision for –ve test result 34/50 respondents.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Lyman 1994(b) (ETA23) USA *Journal of Gen Internal Medicine* -2y care clinicians -Nurses -Pharmacists -Medical undergraduates	-Recruitment unclear -Hypothetical -Cross sectional study -N=50	-Sensitivity (%) (U) -Specificity (%) (U)			**-PPV and 1-NPV** (%), given pre-test probability (%) AND sensitivity AND specificity. One of either prevalence, sensitivity or specificity were varied across scenarios **-Findings:** Base rate neglect negatively associated with pre-test probability. Consistent overestimation of PPV and 1-NPV. Lack of understanding of -ve test results (in derivation of 1-NPV) compared to +ve test results (in derivation of PPV)
Noguchi 2002(ETA25) Japan -Medical undergraduates.	-Recruitment unclear (response rate 96-99%) -Respondents completed a minimum of 1 session on Bayes' theorem. -Hypothetical -Cross sectional study -N=224	-Sensitivity (%)(E&U) -Specificity (%)(E&U) Findings: Under-estimation of test accuracy compared to literature based estimates (specificity (21%) >sensitivity (3%).	-'Pre-test 'probability estimation (%) given clinical history. **-Findings:** Overestimation of pre-test probability more exaggerated for atypical presentations. Relative magnitude of estimates across scenarios was correct.		**PPV & 1-NPV** (%), given respondents' estimates of pre-test probability, sensitivity specificity, and a +ve or –ve test result. **Findings:** Overestimation of PPV and 1-NPV which was greatest if test result concurred with the clinical picture. –Confirmatory bias

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Phelps 2004[ETA26] USA -2y care clinicians	-Convenience sample, self selected. -Hypothetical -Cross sectional study -N=61		-'Pre-test' probability (E) (semi-quantitative %) given clinical history and examination findings. **-Findings:** -Wide variation in pre-test probability estimates for each of 4 different clinical scenarios: (range 70%-95%)		
Poses 1995[ETA27] USA -1y care -2y care	-Recruitment unclear -Ecological &Hypothetical -Non-concurrent controlled trial. -Poor reporting precluded quality assessment. -N=14		-'Pre-test' probability (E) (semi-quantitative %) given clinical history and examination findings. **-Findings:** Educational intervention reduced over -estimation from an average of 23-26% across scenarios to ~3%.-Improvement in accuracy of pre-test probability did not affect treatment decisions.		

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Puhan 2005[ETA28] Switzerland. -1y care 2y care	-Convenience sample attending CME conference. Response rate 37% -Hypothetical -RCT -Random allocation -Allocation concealment -Selection: (2y care over-represented in LR group) -Unclear if Blinding -Attrition not reported -N=183	- Sensitivity (%) (U) -Specificity ((%)) (U) -LRs (verbal description) (U) -Inexact graphic (U)	-'Pre-test' probability estimation (semi-quantitative %) given clinical history and examination findings. **-Findings:** Individual estimates of pre-test probabilities varied 30% and 50% across 7 different disease scenarios.		**-PPV and1-NPV** (semi-quantitative %) based on respondents' semi-quantitative estimates of pre-test probability AND (literature based estimates of test accuracy presented as either sensitivity & specificity (%) OR LRs OR an inexact graphic. **-Findings:** Errors in post-test probability estimates were not significantly different across test accuracy presentation formats.
Reid, C.M. 1998 [ETA29] USA. -Iy care clinicians -2y care clinicians	-Random sample, representing practising clinicians (>90% patient contact). 69% formal training in diagnostic methods nos. Response rate 91%. -Ecological -Qualitative semi-structured interviews -N= 300	-ROC curves (U) -LRs (U) -Sensitivity & specificity(U) -PPV & NPV (D&U) -False+ves & false –ves (D&U) **-Findings:** 1% use ROC & LRs 4% use sensitivity & specificity 80% use PPV and NPV Test errors emphasised			**-Use of Bayesian calculations.** **-Findings:** 3% use formal Bayesian calculations Reference class confusion PVs with sensitivity and specificity: 80% Estimates of test accuracy based on experience rather than the published literature.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Sassi 2008[ETA30] UK. 2y care clinicians	-Purposive sample (64% response rate) -Hypothetical -Cross sectional study supplemented with semi-structured interviews N=64	-False +ves (%) (U) -False –ves (%) (U) **-Findings:** Importance attached to false +ves greater than that predicted by normative decision theory due to emphasis on short-term outcomes in this context. A relatively large gain in test accuracy was required to compensate for an increase in test associated morbidity. 2/30 clinicians were willing to risk sacrificing test accuracy for the opportunity to use innovative technology.			
Schwartz 2003[ETA31] USA -Medical undergraduates	-Convenience sample in an exam setting. -Hypothetical -RCT -Random allocation -Unclear if Allocation concealment -Selection: Unclear if groups comparable -Unclear if Blinding -Attrition: 3% -N=159	-Sensitivity (%) (U) -Specificity (%) (U) **-Findings:** 70% correct use of sensitivity and specificity for management decisions. Internal validity of accuracy evidence did not moderate management decision.			

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Sox 2006(a) (ETA32) USA. -2y care	-Random sample of practising paediatricians. Response rate 50%. -Hypothetical -RCT -Random allocation -Allocation concealment -Selection: groups comparable -Blinding -Attrition not reported -N= 653	Sensitivity and specificity (%) (U) -Sensitivity and specificity (normalised frequencies) (U) -False +ve rate (%) (defined as -1PPV) (U) -False +ve rate (normalised frequency (defined as 1-PPV) (U)			**-Management decisions** assumed to be based on derivation of post-test probability following a test result, based on pre-test probability (%) AND respondents' estimates of test accuracy OR provided with sensitivity and specificity (%) OR false +ve (%) OR sensitivity and specificity (normalised frequency) OR false +ve (normalised frequency). **-Findings:** No difference in management decisions between respondents provided with no test accuracy information; provided with sensitivity and specificity %; false +ve rates as %; sensitivity and specificity as normalised frequencies; false +ve rates as normalised frequencies.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Steurer, J. 2002[ETA33] Switzerland -1y care clinicians	-Convenience sample attending an educational event. Response rate unclear. -Hypothetical -RCT - Random allocation unclear -Allocation concealment -Selection: unclear if groups comparable -Blinding unclear -Attrition:0-4% -N= 251-263	-Sensitivity (%)(D&U) -Specificity (%) (U) -PPV (D) -LR (verbal description) (U) (Note definitions were presented as MCQs) **-Findings:** -76% chose correct definition of sensitivity -61% chose correct definition of PPV	-'Pre-test' probability estimation, given clinical history. -Findings: - 48% of respondents correctly adjusted pre-test probability with info about patient age.		**-Post-test probability** (%) given prevalence, sensitivity and specificity (MCQ format) **-PPV** (%) given pre-test probability AND a test result (+ve or –ve) OR sensitivity and specificity OR LR (Non MCQ format). **- Findings:** 22% of respondents chose correct estimate of post-test probability from MCQ. PPV overestimated: no test accuracy info > sensitivity & specificity (%) > LR in plain language. Base rate neglect.
Van den Ende 2006[ETA34] Belgium -Medical undergraduates -Clinicians nos.	-Recruitment medical undergraduates unclear. Clinicians attending an educational event. Medical undergraduates had had training in Bayes theorem. -Hypothetical & Ecological -Cross sectional study	-LR (U) -Sensitivity) (verbal; %) (U) -Specificity (verbal; %) (U) **-Findings:** No respondents used the terms sensitivity, specificity or LR to describe how they used test accuracy	-Pre-test probability (D&U) **-Findings:** 96% clinicians stated they could estimate pre-test probability: 53% verbal expression; 27% normalised or natural frequency; 20% as %. 76% clinicians do not use pre-test probability		**-PPV** (quantitative %) given pre-test probability (%), sensitivity and specificity. **-Findings:** 37% of medical undergraduates responded correctly

	-N=66	information.	estimation in practise.		
HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Zaat 1992[ETA36] Netherlands -1y care clinicians	-Recruitment unclear. Response rate 65%. - Hypothetical & Ecological Cross-sectional and cohort studies. -N=75			-Self-reproach -Risk -avoidance -Risk -preference -Association between self-reproach, risk -avoidance, risk- taking and laboratory test use. **-Findings :** GPs were more self -reproachful for missing serious disease compared to less serious disease. There was no clear relationship between laboratory test use and self-reproach, risk avoidance or risk preference.	

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Adab 2003[ETA1] UK -Women attending for cervical screening	-Recruitment of practices unclear. Response rate of women 94% -Hypothetical -RCT -Random allocation -Allocation concealment unclear Selection: groups comparable -Attrition: 3% -N=283	-False +ve rate (1-specificity) (U) -False -ve rate (1-sensitivity) (U) -Sensitivity (U) (normalised frequencies) Women also provided with information on lifetime disease risk; disease specific mortality rates; absolute risk reduction; cost of test **-Findings:** Women provided with additional information were significantly less likely to demonstrate an intention to attend for screening. Study design does not allow differentiation between effects of information on risk and that of test accuracy on intention to attend for screening.			

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Cosmides 1996[ETA8] USA (Exp 1-8) -University undergraduates who participated in all 8 experiments	-Paid volunteers. 'Little or no training in statistics'. -Cross sectional study -Hypothetical -N=75	-True +ve rate (sensitivity)(U) -False +ve rate (1-specificity) (U) (% or normalised frequencies)			**-PPV** (% or normalised frequencies), given prevalence (% or normalised frequencies) AND true +ve rate AND false +ve rate **-Findings:** Correct responses: normalised frequencies > % Partitioning information > no partitioning Base rate neglect:% Test accuracy neglect: normalised frequencies.
Davey 2003[ETA10] Australia -Female population sample	-Random incentivised sample until saturation reached. Biased sample with respect to older age and education. -Qualitative interviews -Hypothetical -N=37		-Pre-test probability (% and normalised frequencies)(U) **-Findings:** No clear preference for % or normalised frequentist expression.		**-PPV & NPV** (verbal; normalised frequency; %; graphic)(U) -**Findings:** Wide range (10-90%) of values assigned to verbal descriptors of probability. % and normalised frequencies perceived as about 'other people'. No preference graphic or verbal expression. Verbal descriptions facilitate numeric comprehension.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Evans 2000[(ETA13)] UK (Exp1) -Undergraduate students	-Recruitment unclear -Cross sectional study -Hypothetical -N=255	-False +ve rate (1-specificity) (U) -True positive rate (sensitivity) (U) (%, normalised frequency same denominator, normalised frequency, different denominator).			**-PPV** (% or normalised frequency), given prevalence, true +ve rate and false +ve rate **-Findings**: Correct responses: normalised frequency, same denominator > % > normalised frequency different denominators. PPV as normalised frequency> PPV % Base rate neglect: % -Test accuracy neglect: normalised frequencies.
Evans 2000[(ETA13)] UK (Exp 2) -Undergraduate students	-Recruitment unclear -Cross sectional study -Hypothetical -N=144	-False +ve rate)1-specificity) (U) -True positive rate (sensitivity) (U			**PPV** (% or normalised frequency), given prevalence AND, true +ve rate AND false +ve rate (% or normalised frequency). Prevalence and false +ve rate varied **-Findings** Correct responses: normalised frequencies >%, regardless of requirement for answer as % or normalised frequency. Base rate neglect: %

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Evans 2000[ETA13] UK (Exp 3) -Undergraduate students	-Recruitment unclear -Cross sectional study -Hypothetical -N=103	-False +ve rate (U) -True positive rate (U (%, normalised frequency, natural frequency)			**-PPV** (semi-quantitative, open), given prevalence AND true +ve rate AND false +ve rate (% or natural frequencies). Values of prevalence and false +ve rate varied. Explication and non-explication of subsets **-Findings:** Correct responses: Natural frequency(48%) > % (13%) Explication of subsets> non-explication in both % and natural frequency formats. Test accuracy neglect when information presented as natural frequencies.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Girotto 2001[ETA15] France (Exp 1-7) Undergraduate students	-Recruitment unclear -Cross sectional study -Hypothetical -N=32-160 (Median=40)	-True +ve rate (U) -True –ve rate (U) -False +ve rate (U) (natural frequencies, normalised frequencies and %).			**PPV** (quantitative or semi-quantitative), given prevalence AND true +ve rate AND (false +ve rate OR true -ve rate (natural frequencies, normalised frequencies or %). **-Findings**: Correct responses: partitioning > non partitioning for all representations. Correct responses: natural frequencies > % or normalised frequencies with partitioning. Reference class confusion for all representations without partitioning.
Hamm 1998[ETA17] USA -Patients in 1y care	-Convenience sample. ? response rate -Hypothetical -Cross sectional study -N=184	-Sensitivity (E) -Specificity (E) **Findings:** Estimates of sensitivity and specificity were = and clustered regardless of test.	-Pre-test probability (E) **Findings:** Estimates of pre-test probability clustered between 37% and 50% regardless of disease.		**-PPV** (E,? format), based on respondents' estimates of pre-test probability, sensitivity and specificity. -**Findings:** PPV overestimated and clustered, regardless of diseases or test

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Hinsz 2005[ETA19] USA -Undergraduate students	-Recruitment unclear. Incentivised. -Hypothetical -Cross sectional study N= 32	-Test accuracy (verbal) (U)			**-PPV** (E, %), given pre-test probability) AND test accuracy (normalised frequency. **-Findings:** Pre-test probability and test accuracy positively correlated with PPV. Increasing test accuracy neglect as pre-test probability decreased. Increasing base rate neglect with increased test accuracy. Increasing confidence in judgements with increasing pre-test probability.

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

MIXED HEALTH AND NON HEALTH PROFESSIONALS					
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological**	**FULLY INFORMED - D (define) - E (estimate) - U (use) -Cognitive biases**		**FULLY RATIONAL Motivational biases**	**ABLE TO COMPUTE ACCURATELY Cognitive biases**
		Test accuracy	**Pre-test probability**		
Bramwell 2006[ETA4] UK -Patients -2y care clinicians -Midwives	-Convenience, self-selected (response rate 62%). Health professionals attending educational events. -RCT -Random allocation -Allocation concealment -Selection: unclear if groups comparable -Blinding unclear -Attrition: 2% -Hypothetical -N=100	-Sensitivity (U) -Specificity (U) (%)			**-PPV** (semi-quantitative % OR semi-quantitative natural frequencies), given prevalence AND sensitivity AND specificity as % or natural frequencies. **-Findings:** Base rate neglect with probabilities. Test accuracy neglect with frequencies. Correct responses: Natural frequencies> % Obstetricians 34% > non health professionals 15% > midwives 0% (average 14%).

NOTES TO TABLE:
E: estimate
D: define
U: use (hypothetical or in practice depending on study design (Hypothetical or Ecological).Note for Davey 2003[ETA10] U='preference for information format'
Semi-quantitative estimation: Respondents' indicate correct answer on a sliding or interval scale or answers approximating to the correct answer accepted as correct.
NOS: not otherwise specified
'Explication of subsets' or 'partitioning': explicit linking of probabilities to their reference class as occurs in natural frequency format (see chp.2.3.3)
Open: format for expressing probability left to respondent
Natural frequencies and relative (%; decimal; normalised) frequencies (see table below and chp.2.3.3).

Appendix 2.2: Characteristics and Results of Included Empirical Test Accuracy Studies

Natural Frequency Expression	Normalised frequency expression	Probabilistic expression	Test accuracy expression
In a population of 100, 10 individuals will have disease X and 90 will be unaffected by disease.	In a population of 100, 10 individuals will have disease X and 90 will be unaffected by disease.	The prevalence of disease is 10% (0.1).	Pre-test probability.
Of the 10 individuals with disease, 8 will test positive with test A.	Of every 100 individuals with disease 80 will test positive with test A.	The probability of testing positive with test A if you have disease X is 80% (0.8).	The true positive rate (sensitivity).
Of the 90 individuals without disease, 80 will test negative with test A but 10 will test positive.	Of every 100 individuals without disease, 89 will test negative. *AND* Of every 100 individuals without disease 11 will test positive	The probability of testing negative with test A if you do not have disease X is 89% (0.89). *AND* The probability of testing positive with test A even if you do not have disease is 11% (0.11).	The true negative rate (specificity). The false positive rate (1-sensitivity).
How many patients who test positive will have disease?	How many patients who test positive will have disease?	What is the probability of having disease X if you test positive with test A ?	Positive predictive value or post test probability given a +ve test result.
Answer: 8/ (8+10) = 8/18.	Answer: $\frac{(80/100) \times (10/100)}{((80/100) \times (10/100)) + ((11/100) \times 90/100))}$	Answer: $\frac{(0.8) \times (0.1)}{(0.8 \times 0.1)+(0.11 \times 0.9)}$	

Examples of scenarios:

Percentages and natural frequencies (Bramwell 2006[(ETA4)]):

*Version 1: **percentages**: The serum test screens pregnant women for babies with Down's syndrome. The test is a very good one but not perfect. Roughly 1% of babies have Down's syndrome. If the baby has Down's syndrome, there is a 90% chance that the result will be positive. If the baby is unaffected, there is still a 1% chance that the result will be positive. A pregnant woman has been tested and the result is positive. What is the chance that her baby actually has Down's syndrome?*

*Version 2: **natural frequencies**: The serum test screens pregnant women for babies with Down's syndrome. The test is a very good one but not perfect. Roughly 100 babies out of 10 000 have Down's syndrome. Of these 100 babies with Down's syndrome, 90 will have a positive test result. Of the remaining 9900 unaffected babies, 99 will still have a positive test result. How many pregnant women who have a positive result to the test actually have a baby with Down's syndrome?*

Normalised frequencies (Girotto 2001[(ETA15)])

4 out of 100 people were infected

75 out of 100 infected people had a positive reaction to the test

25 out of 100 uninfected people also had a positive reaction to the test

Among people who have a positive reaction to the test, the proportion that has the infection will be equal to ------ out of ------?

Partitioning information into the reference class to which it belongs (Cosmides 1992[(ETA8)])

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease the test comes out positive (ie the true positive rate is 100%). But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease (ie the false positive rate is 5%).

Imagine that we have assembled a random sample of 1000 Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people. Given the information above:

On average

1) *How many of these 1000 people will have the disease?*
2) *How many of the 1000 people will have the disease and test positive for it?*
3) *How many of the 1000 people will be healthy and test positive for the disease?*
4) *How many of the 1000 people will tests positive for the disease whether they have the disease or not?*
5) *How many people who test positive for the disease will actually have the disease? ---------- out of ----------.*

Estimation of 'pre-test' probability given clinical history (Heller 2004[(ETA18)])

Atypical angina example:

A 65 year old man presents having had two episodes of retrosternal chest pain today, both precipitated by exertion but lasting approximately 2h despite rest. Before obtaining the rest of the history or performing a physical exam, what would you estimate his true risk of ischaemic heart disease to be?

Estimation of pre-test probability and test accuracy in order to derive PPV and decide on subsequent management (Lyman 1994 (a))

A breast lump is found at the time of a routine physical examination of an otherwise healthy 30-yr old woman.

What is the probability that the lump is malignant?
What is the sensitivity of mammography for detecting malignancy in this setting?
What is the specificity of mammography for detecting malignancy in this setting?
What is the probability that the lump is malignant if the mammogram is positive?
Should a biopsy be performed?

Questions about use of test accuracy information in practice (semi-structured telephone survey)(Reid 1998[(ETA29)])
Some authorities recommend that diagnostic decisions be made first by obtaining a test's sensitivity and specificity, estimating the prevalence of disease then calculating a positive or negative predictive value. Do you perform these calculations when you make diagnostic decisions? If no, can you tell me why you do not do them?
As above for ROC curves, LRs.

Do you use test sensitivity and specificity values when you order tests or interpret results? Can you tell me in what way you use them?

When you use sensitivity and specificity where do you get your values from?
Do you prefer to use published values for sensitivity and specificity, or values based on your own clinical experience?

Do you use positive and negative predictive accuracies when you interpret test results?
Do you use any other methods to help you determine the effectiveness or accuracy of the tests you use in practice?

Measurement of attitudes to risk (Zaat 1992[(ETA36)])
Risk avoidance measures:

a) *Adapted from Grol 1990*[(ETA16)]

1) *When in doubt it is better to refer to a specialist rather than wait and see.* 2) *General practitioners must do all they can to find the cause(s) of somatic complaints* 3) *A general practitioner should always stay on the safe side* 4) *A general practitioner should always keep in mind that any complaint could be the beginning of a serious disorder* 5) *A general practitioner may not take any risks with respect to somatic problems*

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
McGettigan 1999 (ER16) (Search date unclear) The effects of information framing on the practices of physicians.	**-Review** **-Effect of risk presentation:** *RR versus AR versus NNT versus verbal* **-Framing effects** SS;SF;PICO 12 trials included (7 RCTs and 5 cross-over trials). Generally of poor quality -Country of origin of included studies not reported. 7/12 2y care; 3 /12 1y care; 2/12 undergraduates; 1/12 unclear setting. 4 /12 probable selection bias as a result of sampling from educational events. **-Hypothetical** **-Intervention and testing risks**	**FINDINGS:** **Comprehension:** -Numerical presentation > consistency in ratings of effectiveness across specialities compared to verbal presentation. **Accuracy of perception:** -RR magnifies perceptions of effects (+ve or –ve) compared to AR or NNT **Behaviour change (intended):** Risk of benefits: **-**RR > +ve effect on treatment uptake compared to AR or NNT. -Numerical presentation > consistency across specialities compared to verbal presentation.	**FINDINGS:** ***Positive versus negative framing:*** -Treatments perceived as more beneficial and were more likely to be chosen for use when +ve outcomes rather than –ve outcomes presented. ***Modifiers of framing effects:*** Effects of comparative risk information format and positive versus negative framing modified or nullified by public vs private providers; risk aversion; experience of clinician; decision type (intervention of treatment; magnitude of risk; cost.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Albada 2009(ER1) (Search date 2007) Tailored information about cancer risk and screening: a systematic review.	**-Review** **-Effect of risk presentation:** *content tailored versus non-content tailored risk information.* -**SS;SF;PICO;QA*(for RCTs only)*;DD** RCTs (N=28) and observational studies (N=12). RCT quality: 2/28= high; 7/28=moderate; 19/28=low. -39/40 studies conducted in USA; 1/40 in the UK. 30/40 interventions aimed at low risk individuals (screening) for cancer; 5/40 high risk; 5/40 self-selected or workplace. **-Ecological** **-Testing risks**	**FINDINGS:** **Comprehension:** (1 high quality study) -Content-tailored significant > compared to non-content tailored. **Accuracy of perception:** -2 trials (moderate – low quality): Content-tailored significant > non-content tailored. 2 studies (quality not assessed): Content-tailored no significant difference to non-content tailored. **Behaviour change:** Risk of harms (risk of developing disease): 3 studies (quality not assessed) -Inconsistent effects on screening uptake of content vs non-content tailored risk information.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2000[ER5] (Search date 1996). The effectiveness of one to one risk communication interventions in health care: a systematic review.	**-Review -Effect of clinical topic; intervention or testing context; healthcare setting; theoretical model underpinning intervention; mode of delivery; presentation of information -SS;SF;PICO;QA;DD** 96 studies met inclusion criteria. 84 studies with quantitative data included in meta-regression. RCTs and 'other (comparative) designs' including before-after studies. Study quality described as variable. Publication bias present using funnel plot analysis. Significant heterogeneity suggests caution in interpreting statistical tests of significance. -Country of origin of included studies not reported. **-Hypothetical> Ecological -Intervention and testing risks**	Comprehension, perception, anxiety, behavioural outcomes pooled for meta-regression; behavioural outcomes favoured if reported by included studies. **FINDINGS:** In addition to study design factors, (non-RCT> effect compared to RCT and continuous > effect compared to binary outcome measures): -***Content tailored*** estimates of risk associated with greater effects on outcomes compared to other risk presentation formats	**FINDINGS:** ***Attitudes to risk:*** Treatment choices associated with greater effects on outcomes compared to preventive behaviour or test choices.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2001[ER6] (Search date 1999) Presenting risk information – a review of the effects of framing and other manipulations on patient outcomes.	**-Review** **-Effect of risk presentation:** *-Numerical vs graphical* *-More vs less explanation of data* *-Numerical vs verbal* *-RR vs AR vs NNT* *-Vivid vs abstract (vignette descriptions)* *-Lay vs medical terminology* **-Manipulation of base rates** **-Framing** -SS;PICO;QA;DD 24 included studies Range of study designs (nos). Quality: poor to moderate. Small numbers of studies for each of 9 sub-group comparisons. -19/24 included studies conducted in USA; 1 in Canada, 1 in the UK; 1 in Belgium, 1 in Australia, and 1 in NZ. **-Hypothetical and Ecological -Intervention and testing risks**	**FINDINGS:** **Comprehension**: -Numerical > verbal presentation. -No difference personalised vs generic information. - Lay> medical terminology. - Base rate neglect when denominators of AR and frequencies 1/n manipulated. **Accuracy of perception**: -No difference for numerical vs graphical presentation. **Preference:** -No difference in perceived value of information for vivid versus abstract information. **Behaviour change (intended and actual)** Risk of benefits and harms: +ve effect on treatment uptake when presenting treatment benefit: more explanation data > less explanation. No effect when presenting treatment harms. -Treatment uptake when presenting treatment harms: verbal presentation> numerical presentation.	**FINDINGS:** ***Negative vs positive framing*:** -No consistent effects on risk taking observed for negative vs positive framing (unclear treatment or testing choices). ***Loss vs gain framing:*** -Loss framing> acceptance to accept screening compared to gain framing. ***Anxiety:*** -Verbal>numerical when communicating risks of harm. - No difference vivid versus abstract information. - Lay> medical terminology when communicating risks of harm.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2001(CONTINUED) (Search date 1999) Presenting risk information – a review of the effects of framing and other manipulations on patient outcomes.		**Behaviour change (intended and actual)** Risk of benefits and harms: - RRR > +ve effect on treatment uptake and screening uptake compared to AR > NNT when presenting treatment / screening benefits. This was the largest effect observed in this review -No difference in preventive behaviour uptake for vivid compared to abstract vignette presentation of risk of developing disease. -Lay> -ve effect on treatment uptake compared to medical terminology when presenting treatment harms.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2006(a) [ER7] (Search date 2005) Personalised risk communication for informed decision making about screening tests (review).	**-Review** **-Effects of content tailored risk presentation** *(risk score, numerical risk estimate, verbal risk estimate or description of risk factors)* **versus non-content tailored risk presentation.** 4/22 studies used a calculated numerical risk estimate; 3/22 studies categorised individuals' risk verbally (high, medium, low risk) and 15/22 studies used listing of risk factors pertinent to an individual **-SS;SF;PICO;QA;DD** 22 RCTs included. Studies described of variable quality but generally good. -5/22 studies in populations at higher risk than 'average'. The majority of included studies were undertaken in the USA. **-Ecological.** **-Testing risks**	**FINDINGS:** **Comprehension:** (3/22 studies) -Content- tailored> non content-tailored (heterogeneity precluded meta-analysis). **Accuracy of perception:** (3/22 studies) -Content-tailored > non-content- tailored (fixed effects OR 1.46 95% CI 1.13 to1.88; random effects OR 1.65 95% CI 0.96-2.81) **Behaviour change:** Risk of harms (risk of developing disease): (13/22 studies) -Content- tailored >uptake screening compared to non-content- tailored (fixed effects OR 1.13 95% CI 1.02 to1.24; random effects OR 1.31 95% CI 0.98 to 1.77 -Within the overall increase in uptake with content-tailored information numerical < screening uptake compared to verbal presentation.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2008 (ER9) (Search date 2006) Interventions to improve risk communication in clinical genetics.	**-Review** Outcomes of relevance to this review: **-Effect of risk presentation:** *-Point estimate versus a range of estimates versus both.* *-Numerical versus verbal format.* *-Tailored versus non-tailored (tailoring nos)* *-Natural frequencies versus probabilities.* **-SS;SF;PICO;DD** Experimental designs (RCTs and non-RCTs included). Quality assessment undertaken but results not reported. 4/28 included studies relevant to this review (3 RCTs and 1 quasi-experimental study). -2 studies conducted in the USA and 2 in the UK. **-3 Ecological and 1 Hypothetical studies** **-Intervention and testing risks.**	**FINDINGS:** **Comprehension:** -Numerical presentation significantly > compared to verbal format. -Tailored information significantly > improvement in knowledge compared to non-tailored information. **Accuracy of perception:** -Point or range estimates> accuracy of risk perception in the short-term (6 months) compared to no numerical information. **Preference:** -No preference for either point or range estimates of risk.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Julian-Reynier 2003 (ER12) (Most recent reference 2002) Risk communication strategies: state of the art and effectiveness in the context of cancer genetic services.	**-Review** **-Effect of risk presentation:** *Numerical versus verbal versus graphical presentation.* *Tailoring information (content and presentation)* **-Framing** -SS; PICO;* 13 reviews and 10 primary studies. Authors note considerable heterogeneity of interventions, outcomes and study designs. -Country of origin of included studies not reported. **-Unclear Hypothetical / -Ecological** **-Intervention, population and testing risks.**	**FINDINGS:** **Comprehension:** Frequencies > probabilities -Tailoring inconsistent effects. **Accuracy of perception:** -Graphical + verbal+ numerical >accuracy compared to single metrics -Tailoring information had inconsistent effects. -Accuracy of perception modified by respondent numeracy **Preference:** -Graphical+ verbal +numerical >numerical > verbal. -Tailoring information had inconsistent effects. **Behaviour change (intended):** Risk of harm (developing disease): +ve effect on preventative behaviour change verbal > numerical presentations. -Content-tailored risk information >uptake of screening compared to non-content- tailored.	**FINDINGS:** ***Gain versus loss framing*** -Loss framing observed to increase of preventative behaviours and screening.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS NON-HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Kuhberger 1998 (ER13) (Search date 1997) The influence of framing on risky decisions: a meta-analysis.	**-Review** **-Effect of positive and negative framing effects** on choice or rating of risk; distinction made between explicit labelling (eg loss, gain) and implicit labelling by the nature of the choice. -**Effect of characteristics of the risk manipulation** *(nature of risk; risk versus risk and risk versus no risk)* **on choice or rating of risk.** **-SS; PICO** 136 studies (>30000 participants) included. Range of study designs. In the face of considerable heterogeneity assessment of study design features rather than methodological quality undertaken. -Country of origin of included studies not reported. **-Hypothetical** **-Intervention and non-medical risks**.		**FINDINGS:** ***Modifiers of framing effects:*** Overall effect sizes for different contextual modifiers small: **-** Framing effects modified by reference point (types of outcome) used. -Risk versus riskless choices> framing effects compared to risk versus risk choices. -Choice between risks > framing effects compared to rating of risks. -Explicit loss, gain wording > framing effect compared to implicitly implied loss or gain. -Business and gambling domains > framing effect compared to social /clinical domains.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Carling 2009 (ER3) The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomised controlled trial. Norway and North America	**-Randomised on-line survey** of the general population. -**Effect of risk presentation:** *-RRR; AR; NNT; event rates; TNT; natural frequencies.* Participants presented with the risks of CHD with and without statins. Block (100) randomisation to 1 of 6 risk presentation formats ;Al;?B;At? Underpowered -N=2978 Self-selected adult volunteers. 63% ≥17 years of education and 71% assessed as numerate. **-Hypothetical.** **-Intervention risks**	**FINDINGS:** **Comprehension:** (5-point Likert scale) -Natural frequencies> other presentation formats. **Preference:** -Natural frequencies (31%)>RRR (30%)> (AR) (20%)> NNT (10%)>ARR (5%)> TNT (3%). **Behaviour change (intended):** Risk of benefits: - RR 21%> effect on treatment uptake compared to all other (absolute summary statistic) presentations.	**FINDINGS:** ***Attitudes to risk:*** Intended behaviour sensitive to values placed on outcomes by respondents across all presentation formats.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Cuite 2008 (ER4) A test of numeric formats for communicating risk probabilities. USA	**-On-line randomised survey** of visitors to a cancer –related internet site. **-Effect of risk presentation:** *(%; normalised frequencies; frequencies 1/n)* on semi-quantitative mathematical operation required: compare, triple, halve, add, sequence and trade off. -Quality assessment precluded by poor reporting. Groups were probably not comparable at baseline. -N=16133 (response rate 36.1% did not vary by risk presentation format or type of mathematical operation required). 82% white; 67% female; 77% college education. **-Hypothetical** **-Population and treatment risks.**			**FINDINGS:** -Mean accuracy across operations 57% for %; 55% normalised frequencies (constant denominator); 45% frequencies 1/n. -Accuracy by operation: compare > halve > triple>trade-off > sequence>add. -Education level and ethnicity significantly positively associated with accuracy.

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Edwards 2006(b) (ER8) Presenting risk information to people with diabetes: evaluating effects and preferences for different formats by a web-based randomised controlled trial. UK	**-On-line survey and qualitative analysis** of free text (50% of respondents). Random assignment (2 x 2 factorial design). **-Effect of risk presentation:** *(no risk information or detailed numerical information (AR, RR, NNT) +/- graphical displays +/- anchoring to familiar risks (normalised frequencies)).* **-R;Al;S;?B;At: 0%** Positive and negative frames used to avoid the effect of loss/gain framing effects. Adequately powered. Qualitative content analysis included illustrative quotes by allocation groups. -N=508 (response rate 72%) for quantitative analysis. N= 256 for qualitative analysis. No significant difference between responders and non-responders. 61% female; 68% received further education beyond school. **-Ecological** **-Intervention risks**	**FINDINGS:** **Comprehension:** (using a decision conflict scale). -No significant difference according to presentation format. **Preference:** -No significant difference in satisfaction according to presentation format. -Qualitative findings: Bar charts were reported as helpful. Multiple numerical presentations and anchoring were reported as unhelpful. Pictographs and thermometers received mixed positive and negative comments.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Hembroff 2004 (ER11) Treatment decision-making and the form of risk communication: results of a factorial survey. USA	**-Randomised telephone survey.** Random assignment to one of 8 risk presentation formats (2x4factorial design). **Effect of risk presentation:** *RR (verbal) or AR (normalised frequency).* **-Quality assessment precluded by poor reporting.** -N= 952 (response rate 54.3%). 57% > high school education. **-Hypothetical.** **-Intervention risks**	**FINDINGS:** **Behaviour change (intended):** Risk of treatment benefit and harms: -RR significantly >+ve effect on treatment uptake compared to AR when communicating about risks of benefit. -RR significantly > -ve effect on treatment uptake when communicating about risks of harm.		**FINDINGS:** -Requesting respondents to combine information about increases in one risk of harm and decrease in a different risk of harm led to > risk aversive behaviour when information presented as RR compared to AR.
Lobb 2003 (ER15) Womens' preferences and consultants' communication of risk in consultations about familial breast cancer. Australia	**- Cross-sectional questionnaire** of consecutive women attending familial cancer clinics. **-Effect of risk presentation:** *(normalised frequencies and % risk information)* -N=193 for preference measures (response rate 84%) and N= 158 for perception (response rate 68%). **-Ecological** **-Population risk**	**FINDINGS:** **Comprehension:** (semi-quantitative MCQ) - 70% accurate; 10% over and 20% underestimation. - No association between preference for presentation format and accuracy of risk comprehension. **Accuracy of Perception:** (semi-quantitative, normalised frequencies or %) - 50%accurate; 24% over and 26% overestimation. Accuracy and educational achievement +ve association.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Lobb 2003 (CONTINUED) Womens' preferences and consultants' communication of risk in consultations about familial breast cancer. Australia		**Preference:** -Words and numbers > numbers)> no preference. -Lifetime or short-term (10 years) risk information> no time preference> before age 50> multiple time frames. -Percent (%) > proportions >no numerical preference > combination of measures >odds. - AR >RR > no preference/ RR + AR.		
Sheridan 2003 (ER18) A randomised comparison of patients' understanding of number needed to treat and other common risk reduction formats. USA	**-Randomised cross sectional survey comparing -Effect of risk presentation:** *(RRR (%); ARR (normalised frequencies); NNT and a combination of all 3 presentations)* **-R; Al;S;?B; At 10% Underpowered** -Non-consecutive out-patient sample aged 50-80. (N=407). Response rate 74%. **-Hypothetical. -Intervention risks**			**FINDINGS:** -Overall 44% correctly identified most effective of 2 treatments - RRR (60% correct)> combination of effect measures (43%)> ARR(42%) >NNT (30%). -Overall 13% of respondents correctly calculated treatment effect. - RRR (21 correct)> ARR (17%), a combination of effect measures (7%) and NNT (6%). -Base rate neglect evident. -Numeracy and accuracy significant +ve association.

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Siegrist 2008 (ER19) The effect of graphical and numerical presentation of hypothetical prenatal diagnosis results on risk perception. Switzerland	**- Randomised cross sectional surveys.** Downs Syndrome screening test result: for low and high risk results separately. **-Effect of risk presentation:** *Study 1: Risk ratio numerator constant (1/n) versus Risk ratio denominator constant (normalised frequencies) versus Paling perspective scale versus pictograph.)* *Study 2: Paling perspective scale versus presentation of multiple risks in numerical (risk ratio)*. **-Quality assessment precluded by poor reporting.** -Volunteer female university students. Men age 24. Study 1: (N=400). Study 2: (N=200). Response rate not reported **-Hypothetical** **-Testing risks**	**FINDINGS:** **Accuracy of perceived risk:** (6 point Likert scale) -Respondents correctly perceived difference between high and low risk. - Paling perspective scale significantly> magnitude of perceived risks compared with numerical ratios alone or pictograph alone. -No significant difference in risk perception frequency1/n compared to normalised frequencies. - Pictographs alone significantly< magnitude of perceived risk compared to other presentation formats **Preference:** (usefulness 6 point Likert scale) -Pictographs >Paling perspective scale> normalised frequencies and frequencies 1/n. The only significant difference in preference was between pictograms and frequencies (1/n).	**FINDINGS:** **Affect:** -Paling perspective scale> negative affect compared to frequencies 1/n> Pictographs and normalised frequencies.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Quality, Question -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Young 2006 (ER20) Different methods of presenting risk information and their influence on medication compliance intentions: results of three studies. USA	- **Cross sectional face-face questionnaires** **-Effect of risk presentation:** *(verbal versus %)* **Single blind but further quality assessment precluded by poor reporting.** -Randomly sampled participants from the general population (N=40) (study 1). Psychology students (N=31) (study 2). Participants randomly sampled from the general population (N=120) (study 3). Response rate and characteristics of participants not reported. **-Hypothetical** **-Intervention risks**	**FINDINGS:** **Perception:** (semi-quantitative) -Verbal presentation of risk resulted in a significantly higher perception of risk compared to numeric (%). **Behaviour (intended):** Risk of harms (side effects): - numeric (%)> +ve effect on treatment uptake compared to verbal presentation.	**FINDINGS:** ***Anxiety:*** -Numeric (%) presentation significantly < anxiety compared to verbal presentation of risk.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

PRIMARY STUDIES NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Zigmund-Fischer 2008 (ER21) Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. USA.	**-Cross sectional online survey** presenting side effects of taking prophylactic tamoxifen compared to the same effects in women not taking tamoxifen. –**Effect of risk presentation** *(RR versus AR; pictograph (dot graphic) versus numeric information (normalised frequencies and % together); denominator 100 or denominator 1000.)* -N= 631 (84% response rate). Self-selected sample of women high risk for breast cancer (incentivised to take part). 95% Caucasian. 66% respondents had a University qualification and numeracy scores were correlated with comprehension. **-Unclear Ecological / Hypothetical.** **-Intervention risks.**	**FINDINGS:** **Comprehension:** (Identification of group at risk: those taking tamoxifen; those not taking treatment; both groups equally at risk). -Larger denominator (1000) significantly > correct identification of at risk group compared to smaller denominator (100). -RR significantly> correct identification of at risk group compared to AR. - Pictographs > correct identification of at risk group compared to numeric (normalised frequencies and % together). **Accuracy of perception:** (5 point Likert scale: not at all likely to extremely likely): -RR significantly > perceived risk compared to AR. -No difference denominator 100 versus denominator 10000. - No difference pictographs versus numeric (normalised frequencies and %).	**FINDINGS:** ***Anxiety:*** -AR significantly< anxiety compared to RR.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS: HEALTH PROFESSIONALS and NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Epstein 2004 (ER10) (Searches stopped 2003) Communicating evidence for participatory decision making.	**-Review** **-Effect of risk presentation:** *(RR vs AR vs NNT vs multiple metrics* *Order of presentation of survival benefits of 2 competing treatment choices.* *Graphical presentation of risks.* *Frequencies vs probabilities for communication of risk.)* **-PICO; unclear DD**** 6 studies included Range of study designs including qualitative focus groups nos. -Country of origin not reported **-Hypothetical** **-Intervention risks**	**FINDINGS:** **Comprehension:** -Patients: Confusion with interpretation of absolute risk measures. -Medical students: Overestimation RR compared to AR and NNT. **Accuracy of perception** (patients): -Stick figures overestimation compared to other graphical presentations (less educated only). -Frequencies perceived as attributed to self and probabilities to others. **Preference** (patients)**:** -Pictographs > bar charts for presentation of single risks -Vertical bar charts> pictographs for communication of multiple probabilities. **Behaviour (intended):** Risk of benefits (treatment choice): -Order of presentation affects treatment choices, particularly in older and less educated patients.		

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS: HEALTH PROFESSIONALS and NON HEALTH PROFESSIONALS				
Study ID	**-Design, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Lipkus 1999 (ER14) Most recent reference 1998 The visual communication of risk.	**-Review -Effect of risk presentation:** *(graphical vs numerical vs verbal vs mixed)* **-SS;PICO;*** Unclear number of included studies. Authors note considerable heterogeneity of intervention, outcome assessment and study design. -Country of origin of included studies not reported. **-Hypothetical and Ecological -Intervention and population risks**	**FINDINGS: Comprehension:** -Limited evidence to support improvement in comprehension with multiple presentation formats or whether the use of particular graphics has relatively greater effect on comprehension when used as part of a multi-presentation format. -Risk ladders effective at conveying range and magnitude of risk based on positioning on the ladder, independent of any numerical information provided. -Comprehension of low probabilities is consistently poor, even with the aid of graphics. **Accuracy of perception:** -Combining graphics with numerical and written information in the form of advice > either presentation format alone.	**FINDINGS: *Attitudes to risk:*** Pictographs or histograms, + numerical presentation >risk aversion compared to pictographs or histograms alone.	

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS: HEALTH PROFESSIONALS and NON HEALTH PROFESSIONALS				
Study ID	**-Design, Question, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Ancker 2006 [ER2] (described as an update of Lipkus 1999) Most recent reference 2005. Design features of graphs in health risk communication: a systematic review.	**-Review** **-Effect of risk presentation:** *(Features of graphic design:-part-whole relationships (eg stacked bars; pie charts)* *-features exploiting graphical perception (eg comparing lengths on a common scale; segments in a pie chart;* *-differences in colour; shape size)* *(Numerical presentation format: frequencies; proportions; common denominators).* **-SS; PICO; unclear DD** 24 studies included. Country of origin not reported. Heterogeneity of interventions, outcome measures and study designs which largely precluded distinction between different graphical presentation formats. **-Hypothetical; unclear ecological** **-Intervention risks; non-medical risks; population risks.**	**FINDINGS:** **Comprehension:** -Ability to express numerical probability using icon arrays inconsistent. -No consistent difference in comprehension face or stick pictographs vs other graphical presentation. -Sequential>random block icon displays(pictographs) -No difference risk ladders and numerical presentation. - Frequencies 1/n < normalised frequencies. -Part-whole presentation >non part-whole representation. -Interpretation of survival curves inaccurate, even in highly educated samples. **Preference:** -Smaller denominators preferred. -Low income samples: bar charts with ordinal scale (low, medium, high) > pictographs or % or bar charts with RR scales. -Preference not correlated with understanding.	**FINDINGS:** ***Positive and Negative framing effects:*** -Probability of long-term survival expressed as mortality (negative frame) leads to risk aversion (aversion to risky treatment) in the short term whereas probabilities of long-term survival expressed in terms of survival lead to risk taking (acceptance of risky treatment) in the sort-term. -Survival curves may reduce emphasis given to short-term survival by drawing attention to longer term outcomes.	**FINDINGS:** -Pictographs to illustrate part-whole relationships >performance on probability problems compared to numerical presentation (probabilities or proportions).

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

REVIEWS: HEALTH PROFESSIONALS and NON HEALTH PROFESSIONALS				
Study ID	**-Design, Quesiton, Quality -External validity -Hypothetical / Ecological -Intervention risks /Testing risks/Population risks**	**FULLY INFORMED - Comprehension -Accuracy of perception -Preference - Intended or actual behaviour**	**FULLY RATIONAL Motivational biases (attitudes to risk; anxiety/ affect; framing effects)**	**ABLE TO COMPUTE ACCURATELY Manipulation of risks Comparison ≥ 2 risks**
Ancker 2006 (CONTINUED) (described as an update of Lipkus 1999) Most recent reference 2005. Design features of graphs in health risk communication: a systematic review.		**Behaviour change (intended):** Risk of harms (non-medical risks): -No difference part-whole relationships compared to numerical representation. -Pictographs or bar charts>+ve effect on preventive behaviour compared to numerical presentation.		
PRIMARY STUDIES HEALTH AND NON-HEALTH PROFESSIONALS				
Schwartz 2005 (ER17) Can patients interpret health information? An assessment of the medical data interpretation test. USA	**-Cross sectional survey. -Information needed to derive and compare risks. -Effect of risk presentation:** *(natural frequency, normal frequency or %.)* -General population and out-patients, (incentivised) (N=174) and physicians described as having 'strong' critical reading skills (N=15). 52% non-health professionals had a college or post-graduate degree. **-Hypothetical -Intervention risks**	**FINDINGS: Comprehension** (non health-professionals)**:** -Information to calculate risk (75% correct). - Contextual information to compare risks across different socio-demographic groups (47-62%) - Denominators to compare risks in 2 groups (45% correct) -Correct comparison of risks: natural frequencies (85%)> normalised frequencies (61%). Health professionals> non-health professionals correct.		**FINDINGS: -**RR and baseline risk for AR (63% correct) -RRR and baseline risk for post-treatment risk (87% correct) **-** RRR and baseline risk for ARR (80% correct) - 2 RRs for RRR (52% correct) - 2 ARs for ARR (77% correct) -AR and sample size for number of events (72% correct). Health professionals> non-health professionals correct.

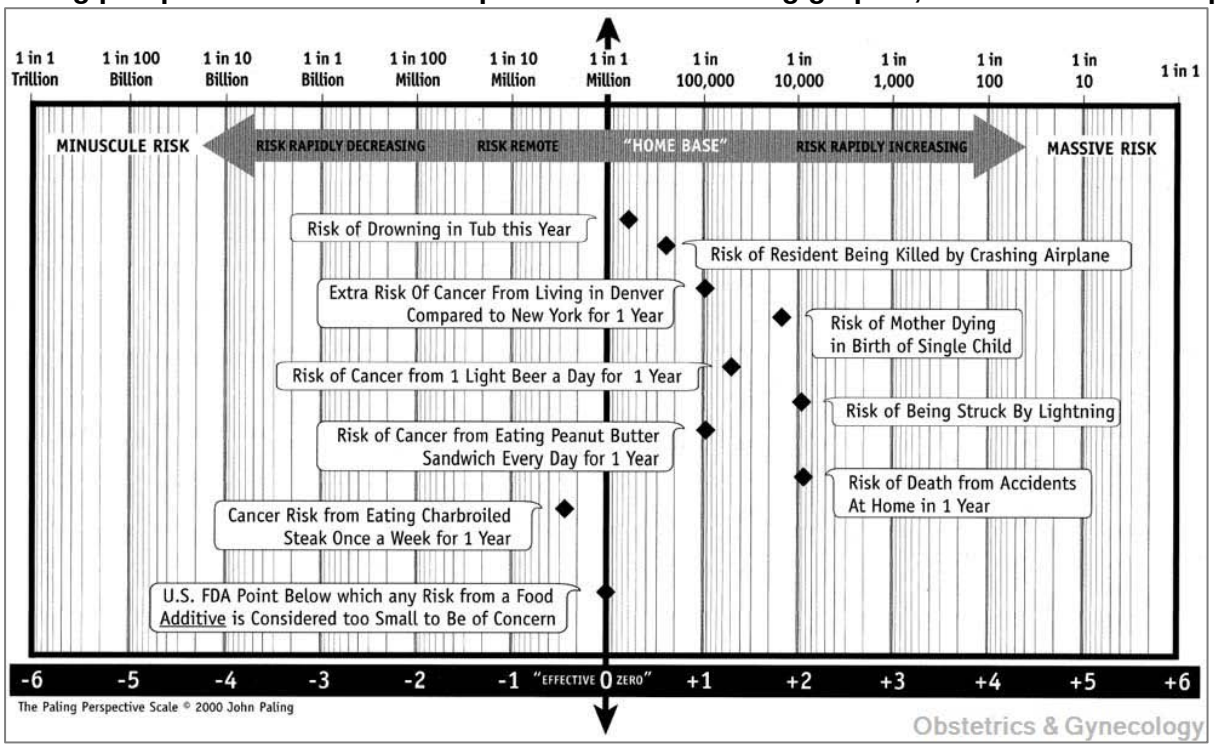
APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

-Quality assessment Reviews:
Search **S**trategy includes> 1 bibliographic database and≥1 of reference checks, contact with experts, hand searching.
Study **F**low presented
PICO
Quality **A**ssessment
Double **D**ata extraction

-Quality assessment RCTs:
R: random allocation met
AI: allocation concealment met
S: groups comparable
B: blinding achieved
At: attrition

-Abbreviations:
NOS: not otherwise specified
RR: relative risk
RRR: relative risk reduction
AR: absolute risk
ARR: absolute risk reduction
NNT: Number needed to treat
TNT: tablets needed to take
Comparative risk measures: RR; RRR; AR; ARR; NNT;TNT
1/n: eg1/312. Frequencies presented in isolation, with a constant numerator but varying denomiator and without reference to a class, in contrast to natural frequencies.
-Natural frequencies (see 2.3.3)
Normalised frequencies: frequencies with a constant denominator (see table X, 2.3.3)
Frequencies 1/n: frequencies with a constant numerator
Graphical methods include bar charts, histograms, pie charts, risk ladders, pictographs (faces, stick men, dot graphics).
Part-Whole relationships: representation of a proportion of a larger value; for example stacked bar charts, pie charts and pictographs.
Tailored and personalised information: Tailored information is defined as tailored to individuals' risk factors (content) or tailored to individuals' preferences and cognitive processing styles (presentation). The term personalised information is used interchangeably with tailored information in the literature.
Anchoring: comparing unfamiliar (medical) risks to familiar (often non-medical) risks
Loss / gain framing: emphasising either losses or gains
Positive /negative framing: stating effects in terms of either positive or negative outcomes

Paling perspective scale: an example of a tool combining graphic, numeric and verbal presentation formats.



Stallings, Shawn P.; Paling, John E. Obstetrics & Gynecology. 98(2):345-349, August 2001.

APPENDIX 2.3 Characteristics and Results of Included Empirical Risk Communication Literature

Risk comprehension (Sheridan 2003) *(ER18)*
"Imagine that 40 out of a 1000 people just like you will develop disease Y over the next 5 years. Treatment A reduces the chance that you will develop disease Y by 25%. Treatment B reduces the chance that you will develop disease Y by 10%.

Which treatment is more effective?
What is the chance you will develop disease Y after treatment A?"

Risk perception and affect: (Siegrist 2008) *(ER19):*
"We describe a fictitious scenario about a pregnant woman. Sandra is 35 years old and she is pregnant for the first time. Her gynaecologist utilises a blood test to assess whether her child might have Downs syndrome. The physician informs Sandra as follows: based on the test, the probability of having a Downs syndrome child is 1:112"

"How do you assess the risk of Sandra delivering a Downs syndrome child? " 6 point Likert scale, small to large.
"What affect would you experience confronted with such a test result? 26 point Likert scale, negative to positive.

Behaviour change (Hembroff 2004) *(ER11):*
"Suppose you had a friend who was told that she was very likely to get a bone disease that would make her crippled. Suppose the doctor said there was a medication she could take on a daily basis that would greatly reduce her chances of getting the bone disease. By taking the medication she would also double her risk of breast cancer. Would you recommend the friend take the treatment?"

Behaviour change (Carling 2009):
"Imagine that you have just found out that you have elevated cholesterol and you are given the option of taking pills called statins that will lower your cholesterol and your risk of developing heart disease over the next 10 years. The pills must be taken one each day, they are usually well tolerated and the side effects, if any, are usually mild and temporary. You need to decide whether to take the pills. Among 50 people that take the pills for the next 10 years, they will swallow a total of 182 500 pills and there will be one additional person who will not get heart disease during that time."

Risk manipulation (Cuite 2008) [ER4]***:***

Sequence: *"When people like you try the drug, about 30% have a negative side effect. Many of these effects are minor, but 10% of the people who experience side effects need to be hospitalised. If you take the drug, what is your chance of being hospitalised?"*

Trade-off: *"Your risk of getting cancer C is 1 in 20 and your risk of getting cancer D is 1 in 100. A new drug would cut your risk of cancer C in half. Unfortunately it would double your risk of cancer D. Would taking the drug: decrease your total cancer risk; increase your total cancer risk; leave your risk unchanged; I wouldn't know?"*

Add: *"There is a 30% chance that the treatment will cure your cancer and a 4% chance that* it will not cure it but will keep it from getting worse. *What is the chance that the treatment will benefit you by either curing your cancer or keep it from getting worse?"*

Compare: "*Your risk of cancer A is 1 in 360 and your risk of cancer B is 1 in 25. Which risk is greater?"*

Halve: *"Your risk of cancer is 8 in 1000 but a new drug would cut that risk in half. What would your new risk be?"*

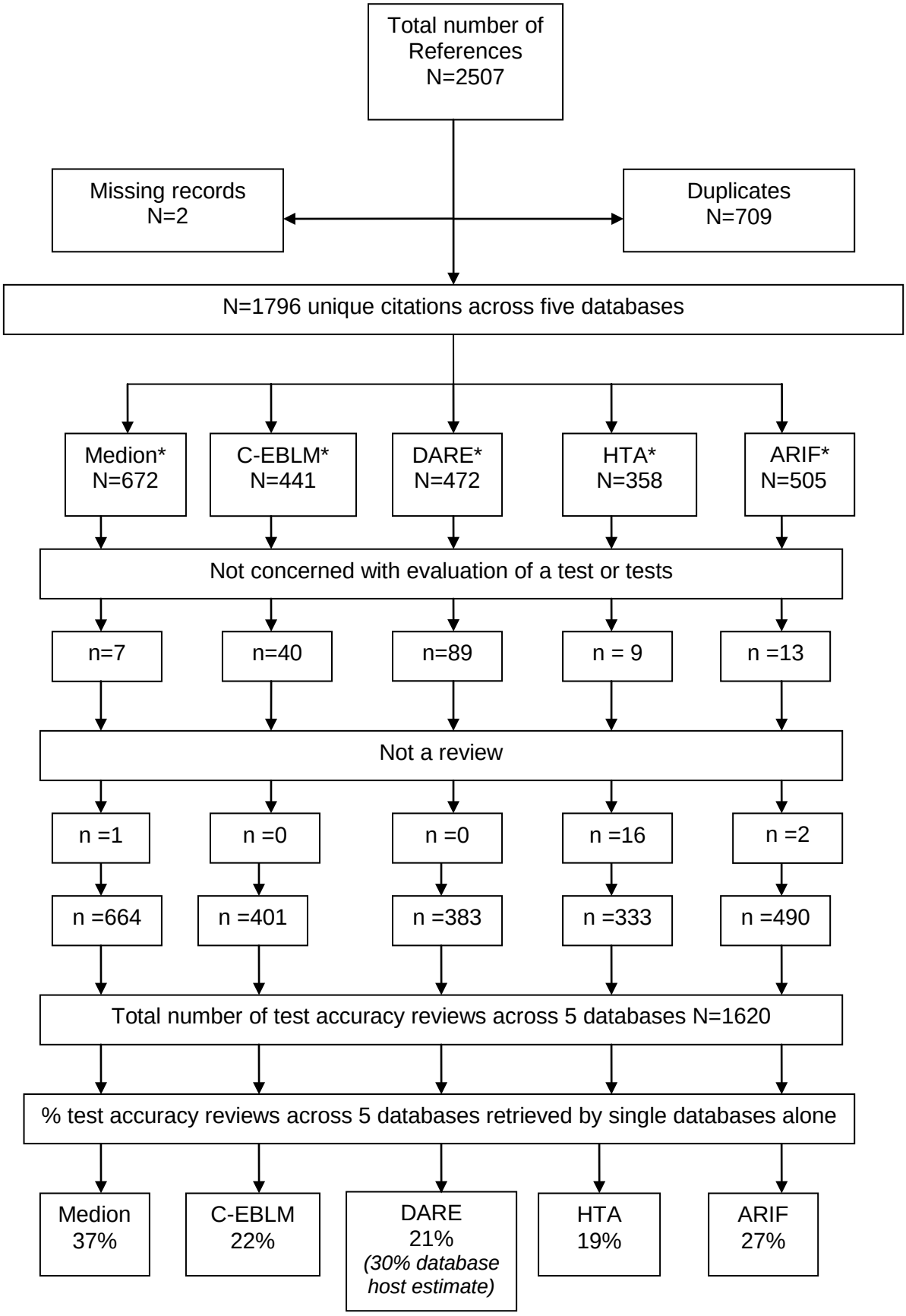
Triple: *"Your risk of cancer is 12 in 1000 but smoking high tar cigarettes would triple that risk. What would your risk be if you smoked these cigarettes?"*

Appendix 3.1: Pragmatic search filters created for use with HTA, DARE and ARIF databases

Pragmatic search filters created for use with HTA, DARE and ARIF in-house databases

	DARE	HTA	ARIF	MEDION	C-EBLM
Indexing term	MeSH Exp Sensitivity or specificity/	MeSH Exp Sensitivity or specificity/	Diagnosis tag	No strategy required	No strategy required
Boolean operator	OR	OR	OR		
Indexing term	MeSH Exp Mass screening/	MeSH Exp Mass screening/	Screening tag		

Appendix 3.2 Flow of References from the HTA, DARE, Medion, C-EBLM and ARIF databases



*total reviews in each individual database will not add up to 1796 due to overlap across databases. Thus one single record in reference manager could be indexed in more than one database.

Appendix 3.3 Pro-Forma for Coding Review References according to Title and Abstract

3.3.1 Testing Setting

Setting	Working Definition
Screening	National screening programmes where asymptomatic individuals are invited to attend for screening or where certain members of the population known to be at increased risk (eg family history of cancer) are called for screening AND Opportunistic screening programmes where individuals attending a health facility or undertaking a particular activity unrelated to the screening test being offered, are offered screening eg blood pressure measurement; random alcohol testing of drivers, oral cancer screening; patient requested screening.
Over The Counter	Self-testing. Not to included home monitoring of chronic disease; this should be coded according to setting in which monitoring initiated as an indicator of disease spectrum.
Community	Application of a test to diagnose a suspected condition in the community setting; includes chiropractice, school based testing.
Primary (1y)	Application of a test to diagnose a suspected condition presenting in primary care.
Secondary (2y)	Application of a test to diagnose a suspected condition presenting in secondary care.
Multiple	Title or abstract clearly indicates that review covers more than one setting eg primary and secondary care; community and primary care.
Unclear	Testing setting unclear from review title and abstract

Appendix 3.3 Pro-Forma for Coding Review References according to Title and Abstract

3.3.2 Disease Topic Area

Disease category	**Examples (particularly disorders that might fall into more than one disease category)**
Anaesthetics	Tests performed whilst under anaesthesia or in preparation for anaesthesia
Breast	Breast screening
Cardiovascular	Diagnosis MI
Cerebrovascular	Stroke
Dental	Screening oral cancers in dental setting
Dermatology	Melanoma
Endocrinology	Thyroid disease
ENT	Hearing impairment; rhinosinusitis; newborn hearing impairment
genetics	Cystic fibrosis; fragile x; genetic predisposition to disease
GI	Irritable bowel; bowel cancer
GU	UTI; renal; microalbuminurea
Haematology	DVT; PE
Head and Neck	Investigation oral cancers; screening oral cancers outside dental setting
Health promotion	Obesity; Smoking
Immunology	C-reactive protein
Infectious dx	HPV; pneumonia; Hep B; Hep C; TB including complications such as pleural effusions; neonatal sepsis
Mental Health	Domestic violence; alcohol; dementia; child abuse; autism
Metabolism	Inborn errors of metabolism
Neurology	Meningitis; syncope; brain injury; head trauma
Non specific	Symptom based diagnosis; diagnosis of a range of disorders
Obs & Gynae	Cervical pathology; HPV; IVF; gynaecological cancers
Opthalmology	Diabetic retinopathy
Musculoskeletal	Includes rheumatology; orthopaedics; lumbar spinal stenosis; carpal tunnel; SLE; soft tissue disease
Renal	Renal dysfunction
Respiratory	Sleep apnoea
Unclear	

Not:e where clinical setting not specified in reference title or abstract but on the basis of clinical experience a setting category could be assigned then studies were coded eg IVF 2y; lymph node metastases 2y.

Appendix 3.3 Pro-Forma for Coding Review References according to Title and Abstract

3.3.3 Review Purpose

Category	Notes
Test accuracy review 1 index test	TA1
Test accuracy review 2 index tests	TA2
Test accuracy ? number of index tests	TA not specified
Effectiveness	E
Cost-effectiveness	CE
Costs	C
Methodological diagnostic review	Method
Other aspects diagnostic tests eg consequences of test error; acceptability; uptake; organisation; descriptive / morphological studies; concerned with existing diagnostic criteria; concerned with differential diagnosis of a symptom; test execution (eg effective collection of endocervical cells during cervical smears); descriptive overview of existing tests; discussion concerning promising diagnostic markers; description testing strategies/work up; indications for testing.	Other
Test review but purpose unclear	Unclear
Multiple testing purpose Eg accuracy and diagnostic impact; effectiveness and cost-effectiveness.	Multiple
Not concerned with tests	Not test
Primary research	Not review

Note: TA refers to studies addressing test accuracy at all stages of test development including development of algorithms. It includes assessment of the accuracy of single tests and testing strategies for diagnosis, prognosis, monitoring and investigation of the accuracy of surrogate end-points.

Appendix 3.4 Yield from searches of combinations of databases excluding primary research, research not concerned with test accuracy and duplicates

Database combination	**Yield (N)** ***Pragmatic search filter (denominator 1620 & DARE contributes 383)***	**Yield %** ***Pragmatic search filter (denominator 1620 & DARE contributes 383)***	**Yield (N)** ***DARE search (denominator 1779 & DARE contributes 542)****	**Yield %** ***DARE search (denominator 1779 & DARE contributes 542)****
C-EBLM OR Medion OR in-house ARIF	1232	76	1232	69
C-EBLM OR Medion OR HTA	1227	76	1227	69
C-EBLM OR Medion OR DARE	1020	63	1179	66
DARE OR HTA OR Medion	1011	62	1170	66
Medion OR in-house ARIF	970	60	970	56
Medion OR HTA	952	59	952	54
Medion OR C-EBLM	948	59	948	53
C-EBLM OR DARE OR in-house ARIF	923	57	1082	61
C-EBLM OR DARE OR HTA	887	55	1046	59
DARE OR HTA OR in-house ARIF	868	54	1027	58
C-EBLM OR in-house ARIF	812	50	812	46
Medion OR DARE	750	46	909	51
HTA OR in-house ARIF	747	46	747	42
C-EBLM OR HTA	688	42	688	39
C-EBLM OR DARE	630	39	789	44
DARE OR in-house ARIF	622	38	781	44
DARE OR HTA	561	35	720	40

*Yield based on DARE database host within-house searching for diagnostic reviews; a facility not accessible on the public database interface at the time of searching

Appendix 3.5: Characteristics of specialist reviews databases

Characteristic	Medion	DARE	HTA	*IFCC's C-EBLM	ARIF
Document source	-Majority sourced from MEDLINE	-Hand-searching -Scanning databases	-Scanning websites -INAHTA project submissions 6 monthly	-MEDLINE and other resources using filters -Contact with experts -HTA sites	-Scanning -Alerting services
Study types	-Systematic reviews. -Separate methodology and genetics databases	-Systematic reviews	-Systematic reviews -Health technology assessments -Primary studies	-Systematic reviews	-Systematic reviews. -Separate methodology database
Overall size of databases (Jan 2007)	1380: -664 test accuracy reviews -597 methods -119 genetics	4539 -542 test accuracy reviews *(database producer estimate)*	6175 -333 test accuracy reviews	555 -401 test accuracy reviews	8670 -491 test accuracy reviews
(Oct 2011)	1650: -1650 test accuracy reviews *(no additions since 2010)*	15950: -1000 test accuracy reviews *(database producer estimate)*	1060 test accuracy systematic reviews & health technology assessments	NO LONGER AVAILABLE	15000 -1800 test accuracy reviews
Dataset	Au, ti, source	Au, ti, source	Au, ti, source	Au, ti, source	Au, ti, source
Abstract	√	√	√	√	×
Textword search	√	√	√	×	√
Indexed by	Keywords IPCP codes	MeSH	MeSH	Keywords	Keywords
Links to text	×	√	√	×	×
Advanced search	√	√	√	×	×
Quality assured	×	√	×	×	×
Updating	Periodic	Monthly	Monthly	Twice per year	Weekly
Gives no of hits	√	√	√	√	√
Sort facility	√	√	√	√	×
Disadvantages	-Website sometimes inaccessible -No help facility	-Test accuracy reviews not tagged	-Test accuracy studies not tagged	-Potential language bias (English reviews only)	-Currently not publicly accessible (plans to make it accessible via ARIF website).
Advantages	- Most unique references	-Detailed indexing: "Reference standard against which new test was compared"	-International focus. -Good resource for reviews concerned with screening	-Laboratory medicine focus -Well indexed. -Large number of unique references	-Currency (updated weekly) -Diagnostic reviews tagged

*C-EBLM website no longer available

Appendix 4.1 Consistency rules for inclusion of reviews on the basis of test

Test	Setting (1y care = included)
Acute Thoracic Dissection	2y
ADHD	1y
Angiography (other than coronary)	2y
Anti CCP antibody in rheumatoid arthritis	1y
Appendicitis: clinical examination	1y
Appendicitis: laboratory and imaging	2y
Biomarkers in gastric inflammation	1y
Blood pressure / hypertension	1y
Breast cancer detection: mammography; MRI; genetic testing	Screening
Breast cancer detection: self-examination	1y
Bronchiolitis	2y
Chlamydia screening	Screening
Chlamydia testing	1y
Cholecystitis	1y
Chrohns disease	2y
Clinical diagnosis / symptoms various	1y
Coagulation testing and bleeding risk	2y
Coeliac disease and serology	1y
Cognitive ability tests	1y
Compartment syndrome	1y
Constipation	1y
Coronary artery disease (all presentations and all tests including stress and exercise)	1y
C-reactive protein and general population; and marker of bacterial infection	1y
C-reactive protein and neonatal sepsis; and acute appendicitis	2y
Dating of bruises in children	1y
Dementia - Diagnosis spect	2y
Dementia – early detection / screening	1y ;Screening
Doppler ultrasound in pregnancy	2y; Screening
Deep Vein Thrombosis: examination; d-dimer	1y
Deep Vein Thrombosis: sonography; asymptomatic; MRI; venography	2y
Dystonia	2y
ECG (including ambulatory)	1y
Ectopic pregnancy	2y
EEG	1y
EEG	2y
Endometriosis	2y
Epilepsy	1y
Evaluation carpal tunnel syndrome (all tests)	1y
Faecal occult blood	Screening
Fractures	2y
Head injury (major and minor)	2y
Headache	1y
Hearing impairment	1y
Heart failure	1y
Helicobacter pylori testing	1y
Hirschsprungs	2y
HIV testing adults	1y
HIV testing infants	2y
Home HPV testing	1y
Home pregnancy testing	1y

Appendix 4.1 Consistency rules for inclusion of reviews on the basis of test

Imaging temporo-mandibular disc	2y
Impotence	1y
Investigation mental retardation	1y
Irritable bowel	1y
LFTs problem drinkers	1y
Low back pain	1y
Lumbar disc herniation	1y
Macroscopc haematuria and urological cancers	1y
Malignant extradural spinal cord compression; lumbar spinal stenoss: diagnosis	2y
Melanoma dermoscopy; dermatoscopy; PET; immunoscintigraphy; integrated diagnosis	2y
Melanoma physical examination primary care	1y
Meningitis (acute presentation)	1y
MI	2y
Microalbuminuria testing	1y
MRI joints	1y
MRI Multiple Sclerosis	1y
Neisseria gonorrhoeae testing	1y
Oral cancers investigation	2y
Oral cancers screening	Other
Osteoporosis; bone mineral density measurement	1y; Screening
Ovarian cancer: symptoms	1y
Pancreatic function tests	1y
Parasight trade mark-F test	2y
Parathyroid disease	2y
Pulmonary embolism	2y
Plain radiograph accuracy various	1y
Prostate cancer monitoring	2y
Prostate cancer testing (PSA)	Screening; 1y
Protein: creatinine ratio random urine	2y
Renal artery stenosis	2y
Rheumatoid arthritis testing	1y
Rhinosinusitis; acute maxillary sinusitis	1y
School entry medical	Screening; 1y
Seizures / epilepsy	2y
Skin prick testing	1y
Sleep apnoea	1y
Stroke prevention (eg detection of high BP)	1y
Stroke-treatment	2y
Syncope	1y
Tarsal tunnel	1y
Testing for von Willebrand’s disease in menorrhagia	2y
Thyroid (hospitalised patients)	2y
Thyroid disease screening	Screening; 1y
Thyroid disease testing	1y
Tuberculosis	2y
Tumour markers	2y
Unstable angina	2y
Urine incontinence	1y
Urine markers for bladder cancer surveillance	2y
Vision screening	Screening; 1y
Vision testing	Other
X-ray lower respiratory tract infection children	1y

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Agency for Healthcare Research and Quality 2003[TAR1]	USA	Cardiovascular	Systematic review of research on diagnosis and treatment of coronary heart disease in women: Sub-question: Are there accurate non-invasive approaches to evaluating suspected coronary disease in women?	4	1	≥55	≥11447	√	6
Anderson 2004[TAR2]	Sweden	Gastro-Intestinal	Meta-analysis of the clinical and laboratory diagnosis of appendicitis	32	?	24	40192	√	1
Appel 1993[TAR3]	USA	Cardiovascular	Ambulatory blood pressure monitoring and blood pressure self-management in the diagnosis and management of hypertension	3	NS	58	NS	√	1
Austin 2003[TAR4]	Australia	Mental Health; Obstetrics & Gynaecology	Antenatal screening for postnatal depression: a systematic review	13	9	16	22664	√	3
Barlow 1998[TAR5]	UK	Applicable to multiple disease areas	Systematic review of the school entry medical examination	2	NS	16	17996	√	4
Bastian 1998[TAR6]	USA	Obstetrics & Gynaecology	Diagnostic efficacy of home pregnancy test kits	16	1	5	620	√	7
Battaglia 2006[TAR7]	Switzerland	Cardiovascular	Accuracy of B-type natriuretic peptide tests to exclude congestive heart failure	3	3	19	9093	√	7
Becker 1996[TAR8]	USA	Haematology	D-Dimer testing and acute thrombo-embolism	4	6	13	1853	√	4
Berger 2000[TAR9]	Netherlands	Gastro-Intestinal	Abdominal symptoms: Do they predict gallstones?	7	3	24	36 302	√	3
Berry 2003[TAR10]	UK	Genito-Urinary; Endocrinology	Micro-albuminuria testing in diabetes: is a dipstick as effective as laboratory tests.	1	3	4	3168	√	5
Brietzke 2004[TAR11]	USA	Respiratory	Can history and physical examination reliably diagnose pediatric obstructive sleep apnoea/hypopnea syndrome? A systematic review of the literature.	6	1	12	>782	√	4

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Chen 2001[TAR12]	USA	Dermatology	A comparison of dermatologist's and primary care physician's accuracy in diagnosing melanoma.	2	2	NS	NS	√	6
Chunn 2004[TAR13]	USA	Cardiovascular	Bedside diagnosis of coronary artery disease: a systematic review	4	4	64	>17000	√	3
Conde-Agudelo 2004[TAR14]	Columbia	Obstetrics & Gynaecology	WHO Systematic review of screening tests for pre-eclampsia	39	1	87	211369	√	6
Cook 2005[TAR15]	USA	Genito-Urinary	Systematic review: noninvasive testing for chlamydia trachomatis and neisseria gonorrhoeae	2	2	29	NS	√	5
de Bruyn 2001[TAR16]	USA	Gastro-Intestinal	A systematic review of the diagnostic accuracy of physical examination for the detection of cirrhosis	19	1	12	1895	√	5
Deville 2004[TAR17]	Netherlands	Genito-Urinary	The urine dipstick to rule out infections. A meta-analysis of the accuracy.	3	1	72	84396	√	7
Dinnes 2003[TAR18]	UK	Musculoskeletal	The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. (HTA)	4	5	73	5553	√	7
Dodd 2006[TAR19]	UK	Applicable to multiple disease areas	In a systematic review, infrared ear thermometry for fever diagnosis in children finds poor sensitivity.	1	2	23	4098	√	5
Doust 2004[TAR20]	Australia	Cardiovascular	A systematic review of the diagnostic accuracy of natriuretic peptides for heart failure	3	2	20	11564	√	4
Fancher 2004[TAR21]	USA	Haematology	Combined use of d-dimer testing and estimation of clinical probability in the diagnosis of deep vein thrombosis: systematic review	1	3	12	5431	√	7
Fiellin 2000[TAR22]	USA	Mental Health	Screening for alcohol problems in Primary Care	26	23	38	NS	√	4
Flemons 2003[TAR23]	USA	Respiratory	Home diagnosis of sleep apnea: a systematic review of the literature	1	1	51	5901	√	6
Fowler-Brown 2004[TAR24]	USA	Cardiovascular	Exercise tolerance testing to screen for coronary heart disease: a systematic review for the technical support for the U.S. preventative task force. AHRQ.	18	2	40	159359	√	4

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Fransen 2004[TAR25]	Netherlands	Gastro-Intestinal	Meta-analysis: the diagnostic value of alarm symptoms for upper gastrointestinal malignancy	6	1	NS	31683	√	3
Garber 1999[TAR26]	USA	Cardiovascular	Cost-effectiveness of alternate test strategies for the diagnosis of coronary artery disease.	5	1	NS	26592	√	2
Gianrossi 1990[TAR27]	USA	Cardiovascular	Cardiac fluroscopy for the diagnosis of coronary artery disease: a meta-analytic review	1	1	13	3765	√	3
Gisbert 2001[TAR28]	Spain	Gastro-Intestinal	Diagnosis of helicobacter pylori infection by stool antigen determination: a systematic review	1	6	60	6847	√	1
Goodacre 2005(a) [TAR29]	UK	Haematology	Meta-analysis: The value of clinical assessment in the diagnosis of deep venous thrombosis	3	3	51	NS	√	9
Goodacre 2005(b) [TAR30]	UK	Haematology	Variation in the diagnostic performance of D-dimer for suspected deep vein thrombosis	1	5	97	NS	√	8
Gorelick 1999[TAR31]	USA	Genito-Urinary	Screening tests for urinary tract infection in children: a meta-analysis	6	1	26	17096	√	3
Harris 2003[TAR32]	USA	Endocrinology	Screening adults for type 2 diabetes: a review of the evidence for the U.S. Preventive Task Force	3	3	7	NS	√	6
Heim 2004[TAR33]	USA	Haematology	D-Dimer testing for deep venous thrombosis: a meta-analysis	1	2	23	3985	X	5
Hobbs 1997[TAR34]	UK	Applicable to multiple disease areas	A review of near patient testing in primary care. HTA.	13	?	92	?	√	9
Huicho 1996[TAR35]	Peru	Infectious disease	Fecal screening tests in the approach to acute infectious diarrhoea: a scientific overview	4	1	25	>19016	√	6
Huicho 2002[TAR36]	Peru	Genito-Urinary	Meta-analysis of urine screening tests for determining the risk of urinary tract infection in children	6	1	48	>31070	√	7
Ioannidis 2001[TAR37]	USA	Ear, Nose and Throat	AHRQ: Technical Report: Evidence for the diagnosis and treatment of acute uncomplicated sinusitis in children (HTA)	5	3	4	508	√	4
Jarvik 2002[TAR38]	USA	Musculoskeletal	Diagnostic evaluation of low back pain with emphasis on imaging	37	>2?	NS	NS	√	3

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Jorm 1997[TAR39]	Australia	Mental Health	Methods of screening for dementia: A meta-analysis of studies comparing an informant questionnaire with a brief cognitive test	6	5	10	2230	?	3
Kearon 1998[TAR40]	Canada	Haematology	Non-invasive diagnosis of deep venous thrombosis. McMaster Diagnostic Imaging Practice Guidelines Initiative.	5	1	40	NS	√	4
Kim 2001[TAR41]	USA	Cardiovascular	Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis	4	1	82	7995	√	6
Kotler 1990[TAR42]	USA	Cardiovascular	Exercise thallium-201 scintigraphy in the diagnosis and prognosis of coronary artery disease	1	1	122	NS	√	2
Kwok 1999[TAR43]	USA	Cardiovascular	Meta-analysis of exercise testing to detect coronary artery disease in women	3	1	22	4113	√	5
Law 1998[TAR44]	UK	Speech and Language	Screening for speech and language delay: a systematic review of the literature. (HTA)	26	24	45	NS	√	5
Lee 2006[TAR45]	USA	Genito-Urinary	A meta-analysis of the performance characteristics of the free prostate-specific antigen test	1	1	41	19643	√	2
Lewis 2006[TAR46]	UK	Gastro-Intestinal	Systematic review: The use of serology to exclude or diagnose coeliac disease (a comparison of the endomysial and tissue transglutaminase antibody tests)	3	1	NS	14247	√	3
Linzer 1997[TAR47]	USA	Respiratory	Diagnosing syncope: Part1: Value of history, physical examination and electrocardiography. Guideline: Clinical Efficacy Assessment Project (CEAP) of the American College of Physicians.	12	0	>28	NS	√	1
Loy 1996[TAR48]	Australia	Gastro-Intestinal	Do commercial serological kits for helicobacter pylori infection differ in accuracy? A meta-analysis	1	>4?	21	NS	√	4
Maguire 2005[TAR49]	UK	Applicable to multiple disease areas	Can you age bruises accurately in children? A systematic review	>1?	1	3	95	√	6

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Mant 2004[TAR50]	UK	Cardiovascular	Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care. (HTA)	4	7	213	NS	√	7
Marshall 1996[TAR51]	Sweden	Musculoskeletal	Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. (Commissioned by the Swedish Council on Technology Assessment in Health Care: HTA)	4	1	19	48802	√	5
Marx 2005[TAR52]	UK	Infectious disease	Meta-analysis: Accuracy of rapid tests for malaria in travellers returning from endemic areas	2	3	25	5747	√	7
McGowan 2003[TAR53]	UK	Cardiovascular	Reliability of reporting left ventricular systolic dysfunction by echocardiography: a systematic review of 3 methods	1	2	43	>2400	√	2
Mohseni-Bandpei 2000[TAR54]	UK	Musculoskeletal	Application of surface electromyography in the assessment of low back pain: a literature review	1	?	38	2360	√	4
Mourad 2003[TAR55]	Canada	Applicable to multiple disease areas	A comprehensive evidence-based approach to fever of unknown origin	6	0	27	NS	√	3
Nayak 2006[TAR56]	USA	Musculoskeletal	Meta-analysis: Accuracy of quantitative ultrasound for identifying patients with osteoporosis	1	1	25	9061	√	7
Nelson 2006[TAR57]	USA	Speech and Language	US Preventative Task Force. Screening for Speech and Language Delay in Pre-School Children: Systematic Evidence Review to the US Preventative Task Force (HTA).	16	14	22	>7521	√	6
Numans 2004[TAR58]	Netherlands	Gastro-Intestinal	Short-term treatment with proton pump inhibitors as a test for gastro-esophageal reflux disease	1	3	15	2793	√	6
Oei 2003[TAR59]	Netherlands	Musculoskeletal	MR Imaging of the menisci and cruciate ligaments: a systematic review	1	1	29	>3000	√	7

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Ogilvie 2005[TAR60]	Canada	Genito-Urinary	Diagnostic accuracy of self-collected vaginal specimens for human papillomavirus compared to clinician collected human papillomavirus specimens: a meta-analysis	1	1	12	4212	√	6
O'Meara 2006[TAR61]	UK	Endocrinology; cardiovascular	Systematic review of methods to diagnose infection in foot ulcers in diabetes	3	2	3	198	√	4
Oosterhuis 2000[TAR62]	Netherlands	Haematology	Diagnostic value of the mean corpuscular volume in the detection of vitamin B12 deficiency	1	>8?	47	NS	?	6
Owens 1996[TAR63]	USA	Infectious disease	Polymerase chain reaction for the diagnosis of HIV infection in adults	1	3	141	>14668	√	5
Pasternack 2003[TAR64]	Finland	Musculoskeletal	Magnetic resonance imaging findings in respect to carpal tunnel syndrome.	1	3	13	780	√	6
Peters 2003[TAR65]	UK	Mental Health	Systematic review of instruments designed to predict child maltreatment during the antenatal and postnatal periods	8	3	8	22496	?	6
Petersen 2001[TAR66]	USA	Mental Health	Practice parameter: Early detection of dementia: mild cognitive impairment (en evidence based review). Report of the quality standards subcommittee of the American Academy of Neurology.	13	5	24	14653	√	3
Pignone 2002[TAR67]	USA	Mental Health	Screening for depression in adults: a summary of the evidence for the U.S. preventative task force	11	4	14	7739	√	5
Pirozzo 2003[TAR68]	Australia	Ear, Nose and Throat	Whispered voice test for screening for hearing impairment in adults and children: systematic review.	1	1	7	1006	?	6
Price 2005[TAR69]	USA	Genito-Urinary	Use of protein: creatinine ratio measurements on random urine samples for prediction of significant proteinurea: a systematic review	1	1	16	1781	?	5
Ramsay 2002[TAR70]	UK	Mental Health	Should health professionals screen women for domestic violence? Systematic review	1	NS	10	35 603	√	6

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Rappeport 1996[TAR71]	Denmark	Musculoskeletal	MR imaging before arthroscopy in knee joint disorders?	1	1	33	4221	√	1
Reed 1996[TAR72]	USA	Infectious disease; respiratory	Sputum gram's stain in community-acquired pneumococcal pneumonia: a meta-analysis	1	4	12	1322	√	3
Reuchlin-Vrocklage 2005[TAR73]	Netherlands	Gastro-Intestinal	Diagnostic value of abdominal radiography in constipated children	2	2	6	485	√	7
Riedemann 2005[TAR74]	USA	Musculoskeletal	The use of second generation anti-CCP anitbody (anti-CCP2) testing in rheumatoid arthritis - a systematic review	1	?	16	7069	√	2
Rietveld 2003[TAR75]	Netherlands	Opthalmology	Diagnostic impact of signs and symptoms in acute infectious conjunctivitis: systematic literature search.	NS	1	0†	0†	√	4
Rodgers 2006[TAR76]	UK	Genito-Urinary	Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation. (HTA)	25	16	105	>30000	√	7
Ross 1999[TAR77]	USA	Respiratory	Systematic review of the literature regarding the diagnosis of sleep apnoea	4	1	71	7572	√	0
Schmitt 2005[TAR78]	USA	Heamatology	Screening primary care patients for hereditary haemochromatosis with transferrin saturation and serum ferritin level: systematic review for the American College of Physicians	1	2	3	7315	√	5
Scholten 2001[TAR79]	Netherlands	Musculoskeletal	The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis	4	3	13	1826	√	6
Schuijf 2006[TAR80]	Netherlands	Musculoskeletal	Meta-analysis of comparative diagnostic performance of magnetic resonance imaging and multi-slice computed tomography for non-invasive coronary angiography	2	1	52	2203	√	2
Scott 2002[TAR81]	UK	Endocrinology	Screening for gestational diabetes: a systematic review and economic evaluation. (HTA)	8	1	135	>36049	√	2

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
Scouller 2000[TAR82]	Australia	Gastro-Intestinal; Mental health	Should we use carbohydrate-deficient transferrin instead of gamma-glutamyltransferase for detecting problem drinkers? A systematic review and meta-analysis	2	2	110	NS	?	4
Selley 1997[TAR83]	UK	Genito-Urinary	Diagnosis, management and screening of early, localised prostate cancer. HTA	6	>1	51	105743	√	2
Singer 1992[TAR84]	USA	Endocrinology; Opthalmology	Screening for diabetic retinopathy	4	?	8	4583	√	1
Siu 1991[TAR85]	USA	Mental Health	Screening for dementia and investigating its causes	19	>2	35	NS	√	3
Stein 2004[TAR86]	USA	Haematology	D-Dimer for the exclusion of acute venous thrombosis and pulmonary embolism	7	11	108	16076	√	6
Stein 2006[TAR87]	USA	Cardiovascular	Multi-detector computed tomography for the diagnosis of coronary artery disease: a systematic review	1	1	33	1606	√	4
Storgaard 1994[TAR88]	Denmark	Mental Health	The validity of the Michigan Alcoholism Screening Test (MAST)	1	8	20	4433	√	2
Takata 2003[TAR89]	USA	Ear, Nose and Throat	Evidence assessment of the accuracy of methods of diagnosing middle ear effusion in children with otitis media with effusion (HTA).	7	3	33	10599	√	7
Tamariz 2004[TAR90]	USA	Haematology	Usefulness of clinical prediction rules for the diagnosis of venous thrombo-embolism: a systematic review	1	5	23	10519	√	6
Tu 2005[TAR91]	USA	Musculoskeletal	Musculoskeletal causes of chronic pelvic pain: A systematic review of diagnosis: Part I	5	2	6	2909	√	2
Tugwell 1997[TAR92]	Canada	Infectious disease	Laboratory evaluation in the diagnosis of Lyme Disease	8	?	9	NS	√	3
van den Hoogen 1995[TAR93]	Netherlands	Musculoskeletal	On the accuracy of history, physical examination and erythrocyte sedimentation rate in diagnosing low back pain in practice	50	3	36	NS	√	3

Appendix 4.2 Demographic Details of Included Reviews

Study ID	Country of origin	Disease Topic category	Title	Index tests (N)	Ref. tests (N)	Incl studies (N)	Patients (N)	Clinician	Quality score/ 9
van der Meer 2005[TAR94]	Netherlands	Immunology; Respiratory	Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review	1	2	17	2980	√	7
Wang 2005[TAR95]	China	Gastro-Intestinal	Is proton pump inhibitor testing an effective approach to diagnose gastroesophageal reflux disease in patients with non cardiac chest pain.	1	2	6	220	√	6
Waugh 2004[TAR96]	UK	Obstetrics & Gynaecology	Accuracy of urinalysis dipstick techniques in predicting significant proteinuria in pregnancy	1	1	7	1841	√	7
Whiting 2005[TAR97]	UK	Genito-Urinary	Rapid tests and urine sampling techniques for the diagnosis of urinary tract infection (UTI) in children under 5 years: a systematic review.	21	2	70	NS	√	7
Whiting 2006[TAR98]	UK	Neurology	Accuracy of magnetic resonance imaging for the diagnosis of mulitple sclerosis	1	1	29	5287	√	7
Wiese 2000[TAR99]	USA	Genito-Urinary	A meta-analysis of the papanicolaou smear and wet mount for the diagnosis of vaginal trichomoniasis	2	2	30	9501	√	8
Zintzaras 2006[TAR100]	Greece	Gastro-Intestinal	Performance of antibodies against tissue transglutamase for the diagnosis of celiac disease: meta-analysis	1	1	21	4457	?	6

Notes to table:
?: unclear
NS: not stated
†: no studies found
Quality assessment items (9): Search>1 bibliographic database; one or more of handsearching, contact with experts, reference checking; quality assessment; explicit inclusion / exclusion citeria; data extraction in duplicate; study flow documented; discussion of review limitations; level of agreement inclusion; level of agreement quality assessment.

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Agency for Healthcare Research and Quality 2003[TAR1]	Diagnosis	Unclear	X	X	X	X	X	X	X	X	gender
Anderson 2004[TAR2]	Diagnosis	Replace	√	2y	X	√	X	√	Unclear	X	none
Appel 1993[TAR3]	Diagnosis; Monitoring; Prognosis	Unclear	√	X	Unclear	Unclear	X	X	X	X	none
Austin 2003[TAR4]	Screening	Add	Unclear	> 1 unclear	X	X	X	X	X	X	none
Barlow 1998[TAR5]	Screening	Replace	√	Community	X	X	X	√	X	X	socio-economic; ethnicity
Bastian 1998[TAR6]	Diagnosis	Replace	Unclear	Community	X	X	X	X	X	X	none
Battaglia 2006[TAR7]	Screening; Diagnosis*	Add	Unclear	1y and 2y	√	√	X	X	X	X	none
Becker 1996[TAR8]	Diagnosis	Triage	Unclear	X	X	X	X	X	X	X	none
Berger 2000[TAR9]	Screening; Diagnosis*	Replace	√	1y and 2y	X	√	X	√	√	X	gender
Berry 2003[TAR10]	Screening	Replace	√	Unclear	√	Unclear	X	X	X	X	none
Brietzke 2004[TAR11]	Diagnosis	Replace	√	Unclear	X	√	X	√	X	X	none
Chen 2001[TAR12]	Diagnosis	Replace	Unclear	Unclear	√	√	X	X	√	X	none

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Chunn 2004[TAR13]	Diagnosis; Prognosis	X	Unclear	Unclear	√	√	X	√	X	X	none
Conde-Agudelo 2004[TAR14]	Prognosis	Unclear	Unclear	Unclear	X	√	X	√	X	√	none
Cook 2005[TAR15]	Screening	Replace	√	Unclear	X	√	X	X	X	X	none
de Bruyn 2001[TAR16]	Diagnosis	Replace	X	X	X	X	X	X	X	X	none
Deville 2004[TAR17]	Screening; Diagnosis*	Replace	X	>1: unclear	X	√	X	√	X	√	country of origin
Dinnes 2003[TAR18]	Diagnosis	Unclear	√	1y and 2y	Unclear	√	X	√	X	√	none
Dodd 2006[TAR19]	Diagnosis	Replace	X	X	X	X	X	√	X	√	none
Doust 2004[TAR20]	Diagnosis	Unclear	X	Unclear	X	X	X	X	X	X	none
Fancher 2004[TAR21]	Diagnosis	Triage	√	Unclear	X	√	X	X	X	X	none
Fiellin 2000[TAR22]	Screening	Add	Unclear	1y	√	√	X	√	√	√	gender
Flemons 2003[TAR23]	Diagnosis	Replace	√	Community	X	√	X	√	X	X	gender
Fowler-Brown 2004[TAR24]	Screening; Prognosis	Add	√	X	√	√	X	X	X	X	none
Fransen 2004[TAR25]	Diagnosis	Add	√	X	X	X	X	X	X	X	none

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Garber 1999(TAR26)	Diagnosis	Triage	√	Unclear	Unclear	√	Unclear	√	√	X	gender
Gianrossi 1990(TAR27)	Diagnosis; Prognosis unclear†	Unclear	Unclear	X	X	√	X	√	√	X	gender
Gisbert 2001(TAR28)	Diagnosis	Unclear	Unclear	X	X	X	X	X	X	X	none
Goodacre 2005(a) (TAR29)	Diagnosis	Replace	Unclear	>1: unclear	√	X	X	√	√	√	gender
Goodacre 2005(b) (TAR30)	Diagnosis	Unclear	X	Unclear	√	√	X	X	X	√	none
Gorelick 1999(TAR31)	Diagnosis	Unclear	Unclear	1y and 2y	X	X	X	√	X	X	none
Harris 2003(TAR32)	Screening	Add	√	1y	X	√	X	X	X	X	none
Heim 2004(TAR33)	Diagnosis	Triage	√	>1: unclear	X	√	X	X	X	X	none
Hobbs 1997(TAR34)	Diagnosis; Prognosis; Monitoring	Unclear	Unclear	1y	X	X	X	X	X	X	none
Huicho 1996(TAR35)	Screening; Diagnosis	Add	√	>1: unclear	√	√	X	√	X	X	none
Huicho 2002(TAR36)	X	Unclear	Unclear	Unclear	X	√	X	√	X	X	none
Ioannidis 2001(TAR37)	Diagnosis	X	X	X	√	√	X	√	√	X	none

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Jarvik 2002[TAR38]	Diagnosis	X	√	1y	Unclear	√	X	√	√	√	none
Jorm 1997[TAR39]	X	Add	X	Unclear	X	X	X	X	X	X	none
Kearon 1998[TAR40]	Diagnosis	Replace; Add	Unclear	X	√	√	X	X	√	√	none
Kim 2001[TAR41]	Diagnosis	Replace	X	Unclear	X	X	X	X	√	√	none
Kotler 1990[TAR42]	Diagnosis; Prognosis	Add	Unclear	Unclear	Unclear	Unclear	X	X	X	X	none
Kwok 1999[TAR43]	Diagnosis	Unclear	X	Unclear	X	Unclear	X	√	Unclear	X	gender
Law 1998[TAR44]	Diagnosis	Add	Unclear	Community and 1y	X	√	X	√	X	√	none
Lee 2006[TAR45]	X	Unclear	Unclear	X	X	X	X	X	X	√	none
Lewis 2006[TAR46]	Screening	Replace	X	Unclear	X	X	X	√	X	X	none
Linzer 1997[TAR47]	Diagnosis	Unclear	X	1y and 2y	√	√	X	√	X	X	none
Loy 1996[TAR48]	Diagnosis	Replace	X	X	X	X	X	X	X	X	none
Maguire 2005[TAR49]	Diagnosis	X	Unclear	Unclear	√	X	X	√	X	X	none
Mant 2004[TAR50]	Diagnosis	Unclear	√	1y	√	√	X	√	X	√	none
Marshall 1996[TAR51]	Screening; Prognosis	Add	Unclear	X	X	X	X	√	X	√	gender

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Marx 2005[TAR52]	Diagnosis	Replace	√	Unclear	√	√	X	X	X	X	none
McGowan 2003[TAR53]	Diagnosis‡	Replace	X	Unclear	X	X	X	√	X	X	none
Mohseni-Bandpei 2000[TAR54]	X	X	Unclear	>1: unclear	√	X	X	X	X	X	none
Mourad 2003[TAR55]	Diagnosis	Unclear	Unclear	Unclear	√	√	X	√	√	√	country of origin
Nayak 2006[TAR56]	Screening	Add	√	X	X	X	X	X	X	X	none
Nelson 2006[TAR57]	Screening	Unclear	√	1y	X	X	X	√	X	√	none
Numans 2004[TAR58]	Diagnosis	Unclear	√	Unclear	X	√	X	√	√	√	none
Oei 2003[TAR59]	Diagnosis	Triage	Unclear	2y	X	X	X	√	X	X	none
Ogilvie 2005[TAR60]	Screening	Replace	√	1y and 2y	X	X	X	X	√	X	none
O'Meara 2006[TAR61]	Diagnosis	Unclear	√	X	X	X	X	√	X	X	none
Oosterhuis 2000[TAR62]	Screening; Diagnosis	Add	√	>1: unclear	X	√	X	X	X	X	none
Owens 1996[TAR63]	Diagnosis	Replace	Unclear	X	√	Unclear	X	√	X	X	none
Pasternack 2003[TAR64]	Diagnosis	Triage	Unclear	>1:unclear	X	√	X	X	√	√	none
Peters 2003[TAR65]	Screening	Add	√	Unclear	X	√	X	√	X	X	none

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Petersen 2001[TAR66]	X	Unclear	X	Unclear	X	X	X	X	X	X	none
Pignone 2002[TAR67]	Screening	Add	Unclear	1y	Unclear	√	X	√	X	Unclear	none
Pirozzo 2003[TAR68]	Screening	Unclear	√	Unclear	X	X	X	√	X	X	none
Price 2005[TAR69]	X	Unclear	X	>1:unclear	X	X	X	X	X	X	none
Ramsay 2002[TAR70]	Screening	Add	Unclear	>1:unclear	X	X	X	√	X	X	gender
Rappeport 1996[TAR71]	Diagnosis; Prognosis	Unclear	√	X	√	√	X	X	X	X	none
Reed 1996[TAR72]	Diagnosis	Unclear	X	Unclear	X	Unclear	X	X	X	X	none
Reuchlin-Vrocklage 2005[TAR73]	Diagnosis	Unclear	√	X	X	√	X	√	X	√	none
Riedemann 2005[TAR74]	Diagnosis; Prognosis	Add	Unclear	Unclear	X	Unclear	X	X	Unclear	Unclear	none
Rietveld 2003[TAR75]	Diagnosis	Unclear	√	Unclear	√	√	X	√	X	√	none
Rodgers 2006[TAR76]	Diagnosis	Unclear	√	1y and 2y	X	√	X	√	X	X	none
Ross 1999[TAR77]	Diagnosis	Replace	√	X	X	√	X	√	√	√	gender
Schmitt 2005[TAR78]	Screening	Add	√	1y	Unclear	X	X	X	X	X	none
Scholten 2001[TAR79]	Diagnosis	Unclear	Unclear	1y and 2y	X	X	X	X	X	X	none

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
Schuijf 2006[TAR80]	Diagnosis	Triage	Unclear	Unclear	X	X	X	√	X	X	gender
Scott 2002[TAR81]	Screening	Unclear	√	Unclear	√	Unclear	X	X	X	X	none
Scouller 2000[TAR82]	Diagnosis	Replace	X	X	X	X	X	X	X	√	gender
Selley 1997[TAR83]	Screening; Diagnosis	Unclear	√	Unclear	X	√	X	X	X	X	none
Singer 1992[TAR84]	Screening	Unclear	Unclear	X	X	X	X	X	X	X	none
Siu 1991[TAR85]	X	Add	X	1y	X	X	X	X	X	√	none
Stein 2004[TAR86]	Diagnosis	Triage	X	X	√	√	X	X	√	√	none
Stein 2006[TAR87]	Diagnosis	Replace	Unclear	X	X	X	X	X	X	X	none
Storgaard 1994[TAR88]	Screening	Add	X	Unclear	X	X	X	X	X	X	gender
Takata 2003[TAR89]	Diagnosis	Unclear	X	X	X	Unclear	X	√	X	X	none
Tamariz 2004[TAR90]	Diagnosis	X	√	X	X	X	X	X	X	X	none
Tu 2005[TAR91]	Diagnosis	X	X	X	X	√	X	X	X	X	none
Tugwell 1997[TAR92]	Diagnosis	Unclear	√	Unclear	√	√	X	X	√	X	none
van den Hoogen 1995[TAR93]	Diagnosis	Add; Replace	√	>1: unclear	√	√	X	√	√	X	gender

Appendix 4.3: Detail of Review Question Formulation

Study ID	Index applic.	Index role	Prior tests	Population / Presentation							
				Index Setting	Spectrum: Acute/ Chronic	Spectrum: Symptoms	Spectrum: Prevalence	Spectrum: Age	Spectrum: Severity	Spectrum: co-morbidity	Spectrum: other
van der Meer 2005[TAR94]	Diagnosis	X	√	1y	√	√	X	X	X	X	none
Wang 2005[TAR95]	Diagnosis	Replace	Unclear	Unclear	√	√	X	√	X	√	none
Waugh 2004[TAR96]	Diagnosis	Replace	Unclear	Unclear	X	Unclear	X	X	√	√	none
Whiting 2005[TAR97]	Diagnosis	Unclear	X	Unclear	Unclear	√	X	√	X	X	none
Whiting 2006[TAR98]	Diagnosis	Add	Unclear	Unclear	X	√	X	√	X	X	none
Wiese 2000[TAR99]	Diagnosis	Unclear	Unclear	X	X	X	X	X	X	X	none
Zintzaras 2006[TAR100]	Diagnosis	Replace	Unclear	X	X	√	X	√	X	X	gender

Notes to table: *Additional application included at formulation stage
†Diagnosis and prognosis discussed as part of review background. Application specified as 'prediction' at question formulation
‡ Index test discussed in a diagnostic application but utility evaluated using measurement of agreement
X: not specified
Unclear: some information was presented / discussed but lack of clarity precluded judgement
Index test: test(s) being evaluated
Comparator test(s): Current practice
Replace: index test being evaluated as a replacement for current practice (comparator test(s)
Triage: index test being evaluated as a screen for further testing
Add: index test being evaluated as an addition to current practice
OTC: Test available Over The Counter
Community: Test to be made available in the community via community healthcare personnel or free access to the public
1y: Test to be made available following contact with a general practitioner / family physician
2y: Test to be made available following contact with a secondary care provider (encompassing 2y and 3y care).

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Agency for Healthcare Research and Quality 2003[TAR1]	X	X	√	Excl	X	X	X	√	X	X	X	X	X
Anderson 2004[TAR2]	X	X	X	Hetero	√	√	Range: 27-61%	X	√	X	2y	X	√
Appel 1993[TAR3]	X	X	X	X	X	X	X	X	√	X	X	X	X
Austin 2003[TAR4]	X	X	X	X	X	X	Range: 6 -32%	X	X	√	populn	√	?
Barlow 1998[TAR5]	Poor Quality	√	√	X	X 1y	X	X	√	X	X	comm	X 1y	?
Bastian 1998[TAR6]	X	√†	√	Excl	√	X	X	X	X	X	OTC	X	X
Battaglia 2006[TAR7]	Variable	√	√ Good	Hetero	√	√	Range: 2-72%	√	√	X	1y & 2y	?	?
Becker 1996[TAR8]	Poor spectrum	√ Sub-group	X	Excl	X	X	X	X	X	X	X	?	?
Berger 2000[TAR9]	Poor spectrum	X	√ Poor	Hetero	X	√	Range: 5-44%	√	√	X	populn & 2y	√	√
Berry 2003[TAR10]	X	X	√ Good	X	X	X	X	X	X	X	1y & 2y	X	√
Brietzke 2004[TAR11]	X	X	√ Good / Excellent	X	X	√	X	X	X	X	X	X	√

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Chen 2001[TAR12]	Poor spectrum & index test.	√	√	X	√	√	X	X	X	X	X	X	X
Chunn 2004[TAR13]	X	X	X	X	√	√	X	X	X	√	2y	?	?
Conde-Agudelo 2004[TAR14]	X	√ Sub-group	√ Variable	Hetero	X	√	Range: 0.6-34%	√	X	√	2y	X	X
Cook 2005[TAR15]	Moderate / good	X	√	X	X	√	Range: 1.2-24%	X	X	X	1y & 2y	√	√
de Bruyn 2001[TAR16]	Poor spectrum & index test	√	√	Hetero	√	?	X	√	X	√	2y	√	X
Deville 2004[TAR17]	Poor spectrum & quality	X	√Moderate	Hetero	X	√	Range: 1-68%	√	X	√	comm & 1y & 2y	?	X
Dinnes 2003[TAR18]	Poor spectrum & index test	√	√ Poor	Hetero	√	√	Range: 21-90%	√	X	√	1y & 2y	X 1y	X 1y
Dodd 2006[TAR19]	Poor spectrum & index test	X	√ Poor	X	X	X	X	X	X 1y	X 1y	X	X	X
Doust 2004[TAR20]	X	X	√ Good	X	√	√	Range: 0.6-71%	X	X	?	comm & 1y & 2y	X	X
Fancher 2004[TAR21]	X	√	X	Hetero	X	√	Range: 0.8-43%	√	X	X	?	√	√

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Fiellin 2000[TAR22]	Poor spectrum	√	√ Variable	X	√	√	Range: 1-44%	√	√	√	1y & 2y	√	√
Flemons 2003[TAR23]	Poor index test	√†	√ Variable	X	X	√	Range: 22-94%	X	√	√	comm & 2y	X	√
Fowler-Brown 2004[TAR24]	X	√	√Moderate - good	X	X	√	Range: 0.1-86	√	√	√	?	√	√
Fransen 2004[TAR25]	Poor spectrum	X	X	X	X	X	Range: 0.5-10%	√	√	X	X 1y	?	√
Garber 1999[TAR26]	X	X	X	X	X	√	Range: 41-74%	X	√	X	X	X	√
Gianrossi 1990[TAR27]	Poor index test & spectrum	X	√ Good	Hetero	X	√	Range: 30-95%	√	√	X	X	X	?
Gisbert 2001[TAR28]	X	X	X	X	X	X	X	X	X	X	X	X	X
Goodacre 2005(a) [TAR29]	Poor quality & index test	√	X	Hetero	X	X	Range: 10-70%	√	X	X	1y & 2y	?	?
Goodacre 2005(b) [TAR30]	X	X	√	Hetero	√	√	Range: 2-78%	√	X	√	2y	?	X
Gorelick 1999[TAR31]	X	X	X	X	X	X	Range: 3-69%	√	X	X	1y & 2y	X	X
Harris 2003[TAR32]	X	X	X	Excl	X	√	X	X	X	X	comm & 1y	X	√
Heim 2004[TAR33]	Poor reference standard	X	X	Excl	X	√	Range: 20-69%	√	X	√	2y	√	√

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Hobbs 1997[TAR34]	Poor spectrum & index test	X	√ Good Sub-group	X	X	X	X	X	X	?	comm & 1y & 2y	?	?
Huicho 1996[TAR35]	Poor spectrum	√	√ Variable	X	√	X 1y	Range: 10-89%	X 1y	X 1y	X 1y	X	X	X
Huicho 2002[TAR36]	X	√	√ Variable	Hetero	X	√	Range: 0.5 - 60%	√	√	X	1y & 2y	X	X
Ioannidis 2001[TAR37]	Poor spectrum	X	X	X	√	X	X	√	X	X	X	X	X
Jarvik 2002[TAR38]	X	X	√	X	X	√	X	X	X	X	X	X	X
Jorm 1997[TAR39]	X	X	X	X	X	X	X	√	X	X	comm & 2y	X	X
Kearon 1998[TAR40]	X	X	X	Excl	√	√	X	X	√	√	1y & 2y	√	√
Kim 2001[TAR41]	X	√	√ Moderate - good	Hetero	√	X	Range: 27-100%	√	√	X	X	X	X
Kotler 1990[TAR42]	X	X	X	X	X	√	X	X	√	X	X	X	√
Kwok 1999[TAR43]	X	√	X	Hetero	X	?	Range: 18-75%	√	?	X	X	X	X
Law 1998[TAR44]	X	X	X	Hetero	X	√	Range: 5-23%	√	X	X	comm & 1y & 2y	?	?
Lee 2006[TAR45]	X	X	X	Hetero	X	X	Mean: 36%	X	X	X	2y	X	X

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Lewis 2006[TAR46]	Poor reference standard & quality	X	√	X	X	√	X	√	X	√	X	X 1y	X
Linzer 1997[TAR47]	X	X	X	X	X	√	X	X	X	√	1y & 2y	X	?
Loy 1996[TAR48]	Poor quality	√	√	Hetero	X	√	X	√	X	X	X	X	X
Maguire 2005[TAR49]	X	X	X	X	√	X	X	?	?	?	1y & 2y	?	?
Mant 2004[TAR50]	X	√	√ Moderate	X	√	√	X	√	X	X	1y & 2y	X 1y	X 1y
Marshall 1996[TAR51]	X	√†	X	Hetero	X	X	Range: 2-66%	√	X	X	?	X	X
Marx 2005[TAR52]	X	√	√ Variable	Hetero	X	√	IQR: 13-33%	X	X	X	X	X	X
McGowan 2003[TAR53]	Poor reference standard	X	X	X	X	X	X	X	X	√	X	X	X
Mohseni-Bandpei 2000[TAR54]	X	X	√	X	√	√	X	X	X	X	X	X	X
Mourad 2003[TAR55]	X	X	√ Poor - moderate	X	X	X	X	X	X	X	X	X	X

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Nayak 2006[TAR56]	Poor spectrum, quality & reference standard	√	√	Excl	X	X	Range: 7 -59%	√	X	X	comm & 1y & 2y	X	√
Nelson 2006[TAR57]	Poor spectrum	√†	X	X	X	X	X	√	X	X	comm &1y & 2y	X	X
Numans 2004[TAR58]	X	X	X	X	√	√	Range: 0-100%	√	X	X	1y & 2y	?	?
Oei 2003[TAR59]	X	√	X	Hetero	X	X	X	√	X	X	2y	X	X
Ogilvie 2005[TAR60]	X	X	X	X	X	X	X	X	√	X	comm & 1y & 2y	√	X 1y
O'Meara 2006[TAR61]	Poor quality	√	√	X	√	X	X	X	X	√	X	X	?
Oosterhuis 2000[TAR62]	X	√†	X	Hetero	X	√	X	√	X	X	?	√	√
Owens 1996[TAR63]	X	X	√	Hetero	√	X	X	X	X	X	X	X	X
Pasternack 2003[TAR64]	Poor spectrum & reference standard	X	√	X	X	√	X	√	X 1y	√	X	√	?
Peters 2003[TAR65]	X	X	√	X	X	√	Range: 0.6- 22%	√	X	X	comm & 2y	X	√

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Petersen 2001[TAR66]													
Pignone 2002[TAR67]													
Pirozzo 2003[TAR68]	Poor quality	√	√ Moderate	X	X	X	Range: 26-61% adults; 9-31% children	√	X	X	comm & 2y	X 1y	X
Price 2005[TAR69]	X	X	X	X	X	X	?	X	X	X	X	X	X
Ramsay 2002[TAR70]	Poor	√	√ Poor	X	√	X	Range: 0-3%	√	X	X	comm & 1y & 2y	√	√
Rappeport 1996[TAR71]	X	X	√	X	X	X	X	X	X	X	X	?	?
Reed 1996[TAR72]	Poor index test	√	X	Hetero	X	X	X	X	X	X	X	X	X
Reuchlin-Vrocklage 2005[TAR73]	X	√†	X	Hetero	X	√	X	√	X	√	2y	√	√
Riedemann 2005[TAR74]	X	X	X	X	√	√	?	√	X	?	X	?	X
Rietveld 2003[TAR75]	X	X	√ Poor	Excl	X	X	X	X	X	X	X	X	X
Rodgers 2006[TAR76]	Poor quality	√Sub group	√ Sub group	Hetero	√	√	X	X	X	X	comm & 2y	X 1y	X 1y

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Ross 1999[TAR77]													
Schmitt 2005[TAR78]													
Scholten 2001[TAR79]	Poor spectrum & index tests	X	√	Hetero	√	?	Range: 13-84%	√	X 1y	X 1y	2y	X	X
Schuijf 2006[TAR80]	X	X	X	X	X	X	Range: 53-100%	√	X	X	X	X	X
Scott 2002[TAR81]	Poor spectrum	X	X	X	√	√	Range: 0-27%	√	X 1y	√	X	√	√
Scouller 2000[TAR82]	Poor	X	√ Variable	Hetero	X	X	X	√	√	√	comm & 1y & 2y	X	X
Selley 1997[TAR83]	X	X	√	X	X	√	Range: 0.2-31%	?	X	√	comm & 1y & 2y	?	√
Singer 1992[TAR84]	X	X	X	X	X	X	X	X	√	X	?	X	X
Siu 1991[TAR85]	X	X	√	X	X	?	X	X	X	X	comm & 1y & 2y	X	√
Stein 2004[TAR86]	X	X	X	Excl Hetero	√	√	Range: 8-78%	√	X	√	?	?	?
Stein 2006[TAR87]	X	X	X	Hetero	?	?	X	X	?	X	X	X	?
Storgaard 1994[TAR88]	X	X	X	X	√	X	Range: 4-100%	X	X	√	comm & 2y	X	X
Takata 2003[TAR89]	Poor index test	√	√ sub group	X	X	X	Range: 8-74%	√	X	X	X	X	X

Appendix 4.4. Reporting of Review Findings

Study	Quality of reporting of primary studies	Methodological quality of primary studies			Study Characteristics Reported in Review Findings								
		Tabulation by study	Discuss txt	Excl / Hetero	Chronicity	Presentation: Symp /Asymp	Prevalence	Age	Severity	Co-morbidity	Setting	Prior tests	Current practice
Tamariz 2004[TAR90]	Poor spectrum	X	√ Good	X	X	X	X	√	X	√	2y	X 1y	?
Tu 2005[TAR91]	Poor	X	X	X	X	√	X	X	X 1y	X	X	X	X
Tugwell 1997[TAR92]	X	X	X	Excl	√	√	X	X	√	X	X	√	√
van den Hoogen 1995[TAR93]	Poor index test	√	√	Excl	X	√	X	X	√	X	1y & 2y	X	√
van der Meer 2005[TAR94]	X	√	√ Poor	Hetero	√	√	X	√	X	X	comm & 1y & 2y	√	√
Wang 2005[TAR95]	Good	X	√	Hetero	√	√	Range: 33-76%	X	√	√	X	√	?
Waugh 2004[TAR96]	Moderate	√	X	Hetero	X	X	Range: 5-81%	X	X	√	2y	X	X
Whiting 2005[TAR97]	Poor quality & spectrum	√	X	Hetero	X	√	Range: 3-73%	√	X	X	?	√	√
Whiting 2006[TAR98]	X	√	√ Poor	X	X	√	Range: 0-80%	√	√	X	?	?	?
Wiese 2000[TAR99]	Good	√ Sub group	X	X	X	√	Range: 6-73%	X	X	X	1y & 2y	?	X
Zintzaras 2006[TAR100]	Poor spectrum Good index test	X	X	Hetero	X	√	X	√	X 1y	X	X	X	X

Appendix 4.4. Reporting of Review Findings

Notes to table:
Comm: Test to be made available in the community via community healthcare personnel or free access to the public
Populn: Test applied at population level (eg screening)
OTC: Test available Over The Counter
1y: Test to be made available following contact with a general practitioner / family physician
2y: Test to be made available following contact with a secondary care provider (encompassing 2y and 3y care).
?: some information was presented / discussed but lack of clarity precluded judgement
X: not reported by review authors, no reason given
X 1y: not reported by review authors; explicit poor reporting by included studies
QR: interquartile range
†: Quality scores tabulated for each study but not individual quality items
Sub-group: Review authors provide quality assessment by sub-groups of studies to assist with interpretation
Methodological quality of primary studies:
Discuss txt: √: authors discuss quality in text. Where overall assessment provided this is given eg ‘good’; ‘poor’.
Tabulation: Graphical or tabular summary of quality items for each study
Excl / hetero: Quality assessment used in decisions about inclusion of studies and / or in the investigation of heterogeneity

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Agency for Healthcare Research and Quality 2003[TAR1]	-LR+ and LR- -Sensitivity and specificity	X	-Results of individual studies reported	X	X
Anderson 2004[TAR2]	-LR+ and LR-† -AUC†	√	-Meta-analysis	X	X
Appel 1993[TAR3]	-Correlation coefficients -Range of mean differences	X	-Results of individual studies reported	√	√
Austin 2003[TAR4]	-PPV and NPV -Sensitivity and specificity -LR+ and LR- -Test positives (TP and FP) -FN	X	-Results of individual studies reported	√	√
Barlow 1998[TAR5]	-Test positives (TP+FP) -Test positive rate (TP+FP ÷all tested)	√	-Results of individual studies reported	X -Limitations1y studies Screening (only test +ve results available) -Multiple target disorders -“Diagnostic yield”	√
Bastian 1998[TAR6]	-Sensitivity and specificity † -Effectiveness score† (Hasselblad 1995 id 3975)	√	-Results of individual studies reported -Meta-analysis	√	√
Battaglia 2006[TAR7]	-TP, TN, FP, FN -LR+ and LR-† -Sensitivity and specificity -Pre-post test probability† -sROC curve†	√	-Results of individual studies reported -Meta-analysis	√	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Becker 1996[TAR8]	-Sensitivity and specificity -ROC space plot -FN, FP	X	-Results of individual studies reported	√	√
Berger 2000[TAR9]	-ROC space plot -DOR† -PPV and NPV	√	-Results of individual studies reported -Meta-analysis	X	X
Berry 2003[TAR10]	-Sensitivity and specificity -PPV and NPV -ROC space plot	X	-Results of individual studies reported	√	√
Brietzke 2004[TAR11]	-Sensitivity and specificity -PPV†	√	-Results of individual studies reported -Meta-analysis	√	√
Chen 2001[TAR12]	-sROC curve† -Sensitivity and specificity	X	-Results of individual studies reported -Meta-analysis	X	X
Chunn 2004[TAR13]	-Sensitivity and specificity -LR+ and LR-†	√	Results of individual studies reported -Meta-analysis	√	X
Conde-Agudelo 2004[TAR14]	-Pre-post test probability † -LR+ and LR-† -Test positives (TP+FP) -Sensitivity and specificity	√	-Results of individual studies reported -Meta-analysis	√	X
Cook 2005[TAR15]	-TP, TN, FP, FN -Sensitivity and specificity† -LR+ and LR-	√	-Results of individual studies reported -Meta-analysis	√	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
de Bruyn 2001[TAR16]	-Sensitivity and specificity† -sROC curve† -Q†	√	-Results of individual studies reported -Meta-analysis	√	X
Deville 2004[TAR17]	-Sensitivity and specificity† -DOR † -Pre-post test probability † -PPV and NPV†	√	-Meta-analysis	√	√
Dinnes 2003[TAR18]	-LR+ and LR-† -Sensitivity and specificity† -1- specificity (FPR) -ROC space plot	√	-Results of individual studies reported -Meta-analysis	√	√
Dodd 2006[TAR19]	-sROC curve† -Sensitivity and specificity† ‡ -DOR†	√	-Meta-analysis	√	√
Doust 2004[TAR20]	-AUC† -TP, TN, FP, FN† -LR+ and LR-† -DOR †	√	-Results of individual studies reported -Meta-analysis	X	X
Fancher 2004[TAR21]	-Sensitivity and specificity † -LR-†	√	-Meta-analysis	√	√
Fiellin 2000[TAR22]	-Sensitivity and specificity	X	-Results of individual studies reported	X	X
Flemons 2003[TAR23]	-Sensitivity and specificity -LR+ and LR- -TP, TN, FP, FN	√	-Results of individual studies reported	√	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Fowler-Brown 2004[TAR24]	-Sensitivity -PPV -TP+FP÷ all tested (termed diagnostic yield) -PPV÷NPV (termed relative risk of target condition)	√	-Results of individual studies reported	X -Limitations 1y studies -Screening (only test +ve results available)	X
Fransen 2004[TAR25]	-PPV and NPV† -Sensitivity and specificity† -TP, TN, FP, FN	√	-Results of individual studies reported -Meta-analysis	√	√
Garber 1999[TAR26]	-Sensitivity and specificity † -TP, TN, FP, FN	X	-Results of individual studies reported -Meta-analysis	X	√
Gianrossi 1990[TAR27]	-Sensitivity and specificity†	√	-Results of individual studies reported -Meta-analysis	X	X
Gisbert 2001[TAR28]	-PPV and NPV† -Sensitivity and specificity†	√	-Results of individual studies reported -Meta-analysis	X	X
Goodacre 2005(a) [TAR29]	-Sensitivity and specificity † -LR+ and LR-†	√	-Meta-analysis	√	√
Goodacre 2005(b) [TAR30]	-Sensitivity and specificity † -ROC space plot	√	-Results of individual studies reported -Meta-analysis	√	√
Gorelick 1999[TAR31]	-FPR(1-specificity)† -TPR (sensitivity) † -sROC curve† -LR+ and LR-† -ROC space plot	X	-Results of individual studies reported -Meta-analysis	X	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Harris 2003[TAR32]	-Sensitivity and specificity -PPV	X	-Results of individual studies reported	X	√
Heim 2004[TAR33]	-Sensitivity and specificity -PPV and NPV -rDOR†	X	-Results of individual studies reported -Meta-analysis	√	√
Hobbs 1997[TAR34]	-LR+ and LR- -PPV and NPV -Sensitivity and specificity -TP+FP÷ all tested (termed test positive rate) -Correlation coefficient -Mean difference index test and reference standard	√	-Results of individual studies reported	Some included tests only	Some included tests only
Huicho 1996[TAR35]	-PPV and NPV -Sensitivity and specificity -sROC curve†	X	-Results of individual studies reported -Meta-analysis	X	X
Huicho 2002[TAR36]	-PPV and NPV -Sensitivity and specificity -sROC curve†	X	-Results of individual studies reported -Meta-analysis	X	X
Ioannidis 2001[TAR37]	-Sensitivity and specificity -PPV and NPV -LR+ and LR- -Average (unweighted) sensitivity	X	-Results of individual studies reported	X	X
Jarvik 2002[TAR38]	-AUC -LR+ and LR- -Sensitivity and specificity	X	-Results of individual studies reported	X	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Jorm 1997[TAR39]	-Effectiveness score† (Hasselblad 1995 id 3975) -Sensitivity and specificity† -Mean difference index test and reference standard	√	-Results of individual studies reported -Meta-analysis	X	X
Kearon 1998[TAR40]	-Sensitivity and specificity† -PPV and NPV†	√	-Results of individual studies reported -Meta-analysis	√	√
Kim 2001[TAR41]	-Sensitivity and specificity† -sROC curve†	√	-Meta-analysis	X	X
Kotler 1990[TAR42]	-PPV and NPV -Test accuracy (TP+TN÷all tested) -Sensitivity and specificity -Average (unweighted) sensitivity and specificity	X	-Results of individual studies reported	X	X
Kwok 1999[TAR43]	-Sensitivity and specificity† -LR+ and LR-† -sROC curve†	√	-Results of individual studies reported -Meta-analysis	X	X
Law 1998[TAR44]	-Sensitivity and specificity -LR+	X	-Results of individual studies reported	√	√
Lee 2006[TAR45]	-Sensitivity and specificity† -AUC† -LR+†	√	-Meta-analysis	√	X
Lewis 2006[TAR46]	-Sensitivity and specificity † -LR+ and LR-†	√	-Results of individual studies reported -Meta-analysis	√	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Linzer 1997[TAR47]	-TP+FP÷ all tested (termed yield)	X	-Results of individual studies reported	X -Limitations 1y studies -Multiple target disorders -"Diagnostic yield"	X
Loy 1996[TAR48]	-sROC curve† -Sensitivity and specificity† -Difference in sensitivity and difference in specificity†	X	-Meta-analysis	X	X
Maguire 2005[TAR49]	-Test accuracy (TP+TN÷all tested) -Correlation coefficient	X	-Results of individual studies reported	X Test accuracy	X
Mant 2004[TAR50]	-LR+ and LR-†	√	-Results of individual studies reported -Meta-analysis	√	√
Marshall 1996[TAR51]	-Sensitivity and specificity† -PPV†	√	-Meta-analysis	X	√
Marx 2005[TAR52]	-ROC space plot -Sensitivity and specificity† -LR+ and LR-†	√	-Results of individual studies reported -Meta-analysis	√	√
McGowan 2003[TAR53]	-Correlation coefficient -Limits of agreement -Inter-observer variability -Intra-observer variability	√	-Results of individual studies reported	X Limits of agreement	√
Mohseni- Bandpei 2000[TAR54]	-Correlation coefficient -Mean difference index test and reference standard -Inter-observer variability -Intra-observer variability -Sensitivity and specificity	X	-Results of individual studies reported	X	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Mourad 2003[TAR55]	- TP+FP÷ all tested (termed yield) -Sensitivity and specificity -FN -LR+	X	-Results of individual studies reported	X -Limitations 1y studies -Multiple target disorders -"Diagnostic yield"	X
Nayak 2006[TAR56]	-TP, TN, FP, FN -AUC † -sROC curve † -Sensitivity and specificity †	√	-Results of individual studies reported -Meta-analysis	X	√
Nelson 2006[TAR57]	-Sensitivity and specificity	X	-Results of individual studies reported	X	X
Numans 2004[TAR58]	-TP, TN, FP, FN -Sensitivity and specificity † -PPV and NPV -LR+ -sROC curve †	X	-Results of individual studies reported -Meta-analysis	√	√
Oei 2003[TAR59]	-Sensitivity and specificity † -DOR † -sROC curve † -rDOR†	√	-Results of individual studies reported -Meta-analysis	X	X
Ogilvie 2005[TAR60]	-Sensitivity and specificity † -DOR † -LR+ and LR- -Q † -sROC curve † -AUC †	√	-Results of individual studies reported -Meta-analysis	X	X
O'Meara 2006[TAR61]	-Sensitivity and specificity	X	-Results of individual studies reported	X	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Oosterhuis 2000[TAR62]	-PPV and NPV -Sensitivity †	X	-Results of individual studies reported -Meta-analysis	√	√
Owens 1996[TAR63]	-Sensitivity and specificity †∞ -FN, FP †∞ -sROC curve † -Pre – post test probability †∞	√	-Results of individual studies reported -Meta-analysis	√	√
Pasternack 2003[TAR64]	-Sensitivity and specificity	X	-Results of individual studies reported	X	X
Peters 2003[TAR65]	-Sensitivity and specificity -PPV and NPV -ROC space plot	√	-Results of individual studies reported	√	√
Petersen 2001[TAR66]	-Sensitivity and specificity -PPV and NPV	X	-Results of individual studies reported	X	X
Pignone 2002[TAR67]	-TP+FP÷ all tested -PPV presented as: TP÷(TP+FP) - (TP+FP)÷TP	√	-Results of individual studies reported	X	√
Pirozzo 2003[TAR68]	-LR+ and LR- -Sensitivity and specificity -sROC curve†	√	-Results of individual studies reported -Meta-analysis	X	√
Price 2005[TAR69]	-Sensitivity and specificity -LR+ and LR- † -sROC curve †	√	-Results of individual studies reported -Meta-analysis	√	X
Ramsay 2002[TAR70]	-TP÷ all tested -sensitivity	√	-Results of individual studies reported	X	X
Rappeport 1996[TAR71]	-Sensitivity and specificity -PPV and NPV -(TP+TN)÷ all tested (test accuracy)	X	-Results of individual studies reported	√	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Reed 1996[TAR72]	-Sensitivity and specificity -ROC space plot -sROC curve †	X	-Results of individual studies reported -Meta-analysis	√	√
Reuchlin-Vrocklage 2005[TAR73]	-Sensitivity and specificity -LR+ -ROC space plot	√	-Results of individual studies reported	X	X
Riedemann 2005[TAR74]	-Sensitivity and specificity -LR+ and LR- -PPV and NPV -DOR	√	-Results of individual studies reported	X	X
Rietveld 2003[TAR75]	No studies included	NA	NA	NA	NA
Rodgers 2006[TAR76]	-Sensitivity and specificity -TP, TN, FP, FN -DOR † -LR+ and LR- † -ROC space plot	√	-Results of individual studies reported -Meta-analysis	√	X
Ross 1999[TAR77]	-Sensitivity and specificity -sROC curve †	√	-Results of individual studies reported -Meta-analysis	X	X
Schmitt 2005[TAR78]	-PPV	X	-Results of individual studies reported	X -Limitations 1y studies -Invasive reference standard	X
Scholten 2001[TAR79]	-Sensitivity and specificity †β -LR+ and LR-β -PPV and NPV†β -sROC curve †	X	-Results of individual studies reported -Meta-analysis	X	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Schuijf 2006[TAR80]	-Sensitivity and specificity † -sROC curve † -PPV and NPV†	√	-Results of individual studies reported -Meta-analysis	X	X
Scott 2002[TAR81]	-Sensitivity and specificity -PPV -LR+ and LR- -FN, FP	X	-Results of individual studies reported	X	X
Scoulller 2000[TAR82]	-DOR † -sROC curve †	√	-Meta-analysis	X	X
Selley 1997[TAR83]	-Sensitivity and specificity -PPV and NPV -FP rate -FN rate -TP÷ all tested (termed detection rate)	X	-Results of individual studies reported	X	√
Singer 1992[TAR84]	-Sensitivity and specificity	X	-Results of individual studies reported	X	X
Siu 1991[TAR85]	-Sensitivity and specificity -LR+ and LR-	√	-Results of individual studies reported	X	√
Stein 2004[TAR86]	-Sensitivity and specificity † -LR+ and LR-†	√	-Results of individual studies reported	√	√
Stein 2006[TAR87]	-Sensitivity and specificity † -LR+ and LR-†	√	-Results of individual studies reported -Meta-analysis	X	X
Storgaard 1994[TAR88]	-Sensitivity and specificity -PPV and NPV	X	-Results of individual studies reported	√	X
Takata 2003[TAR89]	-Sensitivity and specificity † -Pre-post test probability † -ROC space plot † (pooled sensitivity and 1-specificity plotted in ROC space)	√	-Meta-analysis	X	X

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Tamariz 2004[TAR90]	-PPV and NPV -LR+ and LR- -AUC †	√	-Results of individual studies reported -Meta-analysis	√	√
Tu 2005[TAR91]	-Sensitivity and specificity	X	-Results of individual studies reported	X	X
Tugwell 1997[TAR92]	-Sensitivity and specificity † -LR+ and LR- †	X	-Results of individual studies reported -Meta-analysis	√	√
van den Hoogen 1995[TAR93]	-Sensitivity and specificity -ROC space plot	X	-Results of individual studies reported	X	X
van der Meer 2005[TAR94]	-Sensitivity and specificity -PPV and NPV -sROC curve † -AUC †	√	-Results of individual studies reported -Meta-analysis	√	X
Wang 2005[TAR95]	-Sensitivity and specificity † -PPV and NPV -DOR † -AUC † -sROC curve †	√	-Results of individual studies reported -Meta-analysis	X	X
Waugh 2004[TAR96]	-Sensitivity and specificity -LR+ and LR- † -AUC † -Post-test probability	√	-Results of individual studies reported -Meta-analysis	X	X
Whiting 2005[TAR97]	-Sensitivity and specificity -LR+ and LR- † -ROC space plot	√	-Results of individual studies reported -Meta-analysis	√	√
Whiting 2006[TAR98]	-Sensitivity and specificity -TP, TN, FP, FN -LR+ and LR- -ROC space plot -DOR † -AUC †	√	-Results of individual studies reported -Meta-analysis	√	√

Appendix 4.5: Use of Outcome Measures in Reviews

Study ID	Test Accuracy Outcomes	CI	Synthesis	Two dimensions test accuracy distinguished	Consequences of test results discussed
Wiese 2000[TAR99]	-Sensitivity and specificity † -LR+ and LR- † -PPV	√	-Results of individual studies reported -Meta-analysis	√	√
Zintzaras 2006[TAR100]	-Sensitivity and specificity † -Q † -sROC curve † -AUC †	√	-Results of individual studies reported -Meta-analysis	X	X

Notes to table:
LR: likelihood ratio
AUC: area under the receiver operator characteristic curve
PPV: positive predictive value; NPV: negative predictive value
ROC: Receiver operator characteristic
TP: true positive; TN: true negative; FP: false positive; FN: false negative
†Pre-post test probability: Plot of pre-post test probability using pooled estimate of test accuracy and therefore assuming constancy of test accuracy
‡ In this study sensitivity and specificity were pooled. In addition sensitivity and specificity were derived from the DOR derived from a sROC curve.
∞ In this study PPV and sensitivity and specificity were derived from 'Q'.
B In this study average sensitivity and specificity were derived from a sROC curve and PPV and NPV, LR+ and LR- were derived from average sensitivity and specificity.
NA: not applicable no studies included

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Agency for Healthcare Research and Quality 2003[TAR1]	-Index test	Diagnosis	**Role unclear:** Indirect comparison competing index tests	**-Not addressed**	X	X
Anderson 2004[TAR2]	X	Diagnosis	**Replace:** Indirect comparison of competing index tests. Significance testing.	**-Inclusion restricted** to 2y care **-Sub-grouping**†	X	X
Appel 1993[TAR3]	X	-Diagnosis -Monitoring -Prognosis *Distinguished*	**Role unclear**: Indirect comparison (Bland Altman plots of blood mean differences in blood pressure measurement) of competing index tests	**-Not addressed**	√	√
Austin 2003[TAR4]	-Index test	Screening	**Add** (no current test) Indirect comparison of competing index tests.	**-Pre to post-test probability**	√	√
Barlow 1998[TAR5]	-Spectrum	Screening	**Replace**: Direct and indirect comparison of competing index tests Interpretation assisted by CI.	**-Inclusion restricted** to community.	X -Limitations1y studies Screening(only test +ve results available) -Multiple target disorders -"Diagnostic yield"	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Bastian 1998(TAR6)	-Index test	Diagnosis	**Replace**: Indirect comparison competing index tests.	**-Narrative** distinction of results according to setting	√	√
Battaglia 2006(TAR7)	-Prevalence	-Screening -Diagnosis *Distinguished*	**Add** (existing tests): Indirect comparison of competing index tests. Significance testing. Interpretation assisted by CI. Incremental accuracy not assessed	**-Meta-regression** † **-Pre to post-test probability**	√	√
Becker 1996(TAR8)	-Spectrum	-Diagnosis	**Triage:** Single index test In addition indirect comparison of index tests followed by the reference test for test +ves or proceeding straight to the reference test.	-**Sub-grouping** **-Narrative** discussion of potential impact of setting	√	√
Berger 2000(TAR9)	-Spectrum	-Screening -Diagnosis *Distinguished*	**Replace**: Indirect comparison of competing index tests. Interpretation assisted by CI.	**-Meta-regression**† **-Pre to post-test probability**	X	X
Berry 2003(TAR10)	X	Screening	**Replace** (reference standard) Single index test	**-Narrative** distinction of results according to setting	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Brietzke 2004[TAR11]	X	Diagnosis	**Replace** (reference standard) Indirect comparison of competing index tests. Interpretation assisted by CI	-**Not addressed**	√	√
Chen 2001[TAR12]	-Index test	Diagnosis	**Replace** Indirect and direct comparison of competing index tests.	**-Sub-grouping** **-Limited by primary studies -** construction of artificial case series.	X	X
Chunn 2004[TAR13]	X	-Diagnosis -Prognosis *Distinguished*	**Role unclear** Indirect comparison competing index tests compared to a reference standard. Interpretation assisted by CI for some outcomes	-**Not addressed**	√	X
Conde-Agudelo 2004[TAR14]	-Spectrum	-Prognosis	**Role unclear**: Indirect comparison competing index tests compared to a reference standard Interpretation assisted by CI.	**-Sub-grouping**†	√	X
Cook 2005[TAR15]	-Prevalence -Spectrum	-Screening	**Replace:** Direct comparison competing index tests. Interpretation assisted by CI.	**-Sub-grouping**†	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
de Bruyn 2001(TAR16)	-Spectrum -Index test	-Diagnosis	**Replace (reference standard)** Indirect and direct comparison competing index tests. Interpretation assisted by CI.	**-Meta-regression**† **-Narrative** discussion of potential impact of setting. **-Limited by primary studies**	√	X
Deville 2004(TAR17)	-Spectrum -Index test	-Screening -Diagnosis *Distinguished*	**Replace (reference standard)** Single index test. Interpretation assisted by CI.	**-Sub-grouping**† **-Meta-regression**	√	√
Dinnes 2003(TAR18)	-Spectrum -Prevalence	-Diagnosis	**Multiple Index Tests** : **Replace** Direct and indirect comparison of competing index tests **Add** (existing tests) Assessment of incremental accuracy using direct and indirect comparisons across primary studies. Interpretation assisted by CI.	**-Sub-grouping**†	√	√
Dodd 2006(TAR19)	X	-Diagnosis	**Replace (reference standard)** Single index test. Interpretation assisted by CI.	-**Not addressed.**	√	√
Doust 2004(TAR20)	X	-Diagnosis -Screening	**Add** (existing tests) Direct and indirect comparison of competing index tests. Interpretation assisted by CI. Incremental accuracy not assessed	**-Sub-grouping**†	X	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Fancher 2004[TAR21]	-Index test	-Diagnosis	**Triage** Single index test Interpretation assisted by CI.	**-Meta-regression** †	√	√
Fiellin 2000[TAR22]	-Spectrum -Index test	-Screening	**Add** (existing tests) Direct and indirect comparison of competing index tests Incremental accuracy not assessed	**-Sub-grouping**†	X	X
Flemons 2003[TAR23]	-Prevalence -Spectrum -Index test	-Diagnosis	**Replace (reference standard)** Single index test Interpretation assisted by CI.	**-Narrative** discussion of results according to setting	√	X
Fowler-Brown 2004[TAR24]	-Prevalence -Spectrum	-Screening -Prognosis *Distinguished*	**Add** (no current test): Indirect comparison of competing index tests.	**-Narrative** – discussion of differences in results for high risk and low risk / unselected populations.	X -Limitations 1y studies -Screening (only test +ve results available)	X
Fransen 2004[TAR25]	-Prevalence -Spectrum	-Diagnosis	**Add** (existing tests): Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Sub-grouping**.	√	√
Garber 1999[TAR26]	-Prevalence -Index test	-Diagnosis	**Triage**: Indirect comparison of competing index tests. Effectiveness of triage versus no triage.	-**Sub-grouping**†	X	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Gianrossi 1990(TAR27)	-Spectrum	-Diagnosis	**Replace (reference standard)** Single index test Interpretation assisted by CI.	**-Meta-regression** † **-Sub-grouping** **-Other** (scatter plot prevalence and sensitivity)	X	X
Gisbert 2001(TAR28)	-Spectrum	-Diagnosis	**Role unclear:** Single index test compared to a reference standard.	**-Sub-grouping**†	X	X
Goodacre 2005(a) (TAR29)	-Prevalence -Spectrum	-Diagnosis	**Replace**: Indirect and direct comparison of competing index tests. Interpretation assisted by CI.	**-Meta-regression**† **-Sub-grouping** **-Pre-post-test probability**	√	√
Goodacre 2005(b) (TAR30)	-Prevalence -Spectrum -Index test	-Diagnosis	**Role unclear:** Single index test compared to a reference standard. Interpretation assisted by CI.	**-Meta-regression**† **-Sub-grouping**	√	√
Gorelick 1999(TAR31)	-Prevalence -Spectrum -Index test	-Diagnosis	**Role unclear:** Single index test compared to a reference standard. Significance testing.	**-Meta-regression** † **-Sub-grouping**	X	X
Harris 2003(TAR32)	X	-Screening	**Add** (no current test) Indirect comparison of competing index tests.	**-Inclusion restricted**	X	√
Heim 2004(TAR33)	-Prevalence -Spectrum	-Diagnosis	**Triage:** Single index test	**-Meta-regression** †	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Hobbs 1997(TAR34)	-Spectrum -Index test	-Screening -Diagnosis -Monitoring *Distinguished*	**Multiple Index tests: roles unclear**	**-Narrative** discussion of results according to setting	Some included tests only	Some included tests only
Huicho 1996(TAR35)	-Spectrum	-Diagnosis -Screening	**Add** (existing tests): Indirect comparison of competing index tests Incremental accuracy not assessed	**-Meta-regression**† **-Sub-grouping**	X	X
Huicho 2002(TAR36)	X	Unclear	**Role unclear:** Indirect comparison of competing index tests compared to a reference standard	**-Not addressed**	X	X
Ioannidis 2001(TAR37)	X	-Diagnosis	**Role unclear:** Indirect comparison of competing index tests compared to a reference standard	**-Not addressed**	X	X
Jarvik 2002(TAR38)	X	-Diagnosis	**Multiple index tests: roles unclear**	**-Not addressed**	X	X
Jorm 1997(TAR39	X	-Screening	**Add** (no current test) Indirect comparison of competing index tests. Interpretation assisted by CI.	**-Not addressed**	X	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Kearon 1998[TAR40]	-Spectrum	-Diagnosis	**Replace (reference standard)** Direct and indirect comparison of competing index tests **Add** (existing tests): Assessment of incremental accuracy using indirect comparisons across primary studies and Bayesian updating by review authors.	**-Sub-grouping**†	√	√
Kim 2001[TAR41]	-Spectrum	-Diagnosis	**Replace:** Indirect comparison of competing index tests.	**-Meta-regression**† **-Sub-grouping**	X	X
Kotler 1990[TAR42]	X	-Diagnosis -Prognosis *Distinguished*	**Add** (existing tests): Single index test Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Sub-grouping -Pre to post test probability**	X	X
Kwok 1999[TAR43]	X	-Diagnosis	**Role unclear:** Indirect comparison of competing index tests compared to a reference standard.	**-Sub-grouping**† **-Pre to post test probability**	X	X
Law 1998[TAR44]	-Prevalence -Spectrum	-Diagnosis	**Add** (no current test) : Indirect comparison of competing index tests.	**-Sub-grouping**†	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Lee 2006[TAR45]	X	To be determined by review	**Role unclear** (to be determined by review): Single index test compared to a reference standard.	**-Sub-grouping**†	X	X
Lewis 2006[TAR46]	-Prevalence -Index test	-Screening -Diagnosis	**Replace:** Direct and indirect comparison of competing index tests	**-Sub-grouping**† **-Narrative:** discussion of the potential impact of prevalence on post-test probability.	√	√
Linzer 1997[TAR47]	X	-Diagnosis	**Add** (existing tests): Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Not addressed**	X -Limitations 1y studies -Multiple target disorders -"Diagnostic yield"	X
Loy 1996[TAR48]	X	-Diagnosis	**Replace:** Direct comparison of competing index tests. Significance testing.	**-Meta-regression**† **-Pre to post test probability**	X	X
Maguire 2005[TAR49]	-Spectrum	-Diagnosis	**Role unclear:** Single index test compared to a reference standard.	**-Not addressed**.	X Test accuracy	X
Mant 2004[TAR50]	-Spectrum -Index test	-Diagnosis	**Multiple Index tests: roles unclear**	**-Not addressed**.	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Marshall 1996[TAR51]	-Spectrum	-Screening -Prognosis *Distinguished*	**Add** (existing tests)**:** Single index test Incremental accuracy assessed using Bayesian updating by review authors.	**-Sub-grouping**† **-Pre to post test probability**	X	√
Marx 2005[TAR52]	X	-Diagnosis	**Replace:** Direct and indirect comparison of competing index tests. Interpretation assisted by CI. Significance testing.	**-Meta-regression**† **-Sub-grouping** **-Pre to post test probability**	√	√
McGowan 2003[TAR53]	-Spectrum -Index test	-Diagnosis	**Replace (reference standard):** Single index test Interpretation assisted by CI.	**-Not addressed**	X Limits of agreement	√
Mohseni-Bandpei 2000[TAR54]	-Index test	-Diagnosis -Screening -Monitoring *Distinguished*	**Role** (early test evaluation) Single index test compared to a reference standard.	**-Not addressed** (early test development)	X	X
Mourad 2003[TAR55]	X	-Diagnosis	**Role unclear:** Indirect comparison of competing index tests.	**-Not addressed**	X -Limitations 1y studies -Multiple target disorders -“Diagnostic yield”	X
Nayak 2006[TAR56]	-Prevalence -Spectrum -Index test	-Screening	**Add** (no current test) Indirect comparison of competing index tests Interpretation assisted by CI.	**-Sub-grouping** † **-Pre to post test probability**	X	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Nelson 2006[TAR57]	-Spectrum -Index test -Organisation of healthcare	-Screening	**Role unclear**: Indirect comparison of competing index tests.	**-Narrative** distinction by age and discussion of potential impact of setting	X	X
Numans 2004[TAR58]	-Index test	-Diagnosis	**Add** (existing tests): Single index test. Incremental accuracy not assessed	**-Not addressed**	√	√
Oei 2003[TAR59]	X	-Diagnosis	**Triage:** Indirect comparison of competing index tests.	**-Sub-grouping**† **-Meta-regression**	X	X
Ogilvie 2005[TAR60]	-Spectrum -Index test	-Screening	**Replace (reference standard)** Single index test. Interpretation assisted by CI.	**-Sub-grouping**†	X	X
O'Meara 2006[TAR61]	-Spectrum -Index test	-Diagnosis.	**Role unclear :** Indirect comparison of competing index tests compared to a reference standard.	**-Not addressed**	X	√
Oosterhuis 2000[TAR62]	-Spectrum -Index test	-Screening -Diagnosis	**Add** (no current screening test) Single index test compared to a reference standard. **Add** (existing tests for diagnosis): Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Sub-grouping**†	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Owens 1996[TAR63]	-Spectrum -Index test	-Diagnosis.	**Replace (reference standard)** Single index test. Interpretation assisted by CI.	**-Not addressed**	√	√
Pasternack 2003[TAR64]	-Spectrum	-Diagnosis	**Triage:** Single index test	**-Not addressed** (early test development)	X	X
Peters 2003[TAR65]	-Spectrum	-Screening	**Add** (no current test) Indirect comparison of competing index tests. Interpretation assisted by CI.	**-Not addressed**	√	√
Petersen 2001[TAR66]	X	To be determined by review	**Add** (no current test) Indirect comparison of competing index tests.	**-Narrative** distinction of results according to setting	X	X
Pignone 2002[TAR67]	X	-Screening	**Add** (no current test) Indirect comparison of competing index tests. Assessment of incremental effectiveness by comparing 'new test and treat and 'no test and treat'.	**-Inclusion restricted** to 1y care	X	√
Pirozzo 2003[TAR68]	-Spectrum	-Screening	**Add** (existing tests)**:** Single index test Incremental accuracy not assessed	**-Not addressed**	X	√
Price 2005[TAR69]	-Spectrum -Index test	To be determined by review	**Replace (reference standard)** Single index test. Interpretation assisted by CI.	**-Not addressed**	√	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Ramsay 2002(TAR70)	-Spectrum -Index test -Organisation of healthcare	-Screening	**Add** (no current test) Indirect comparison of competing index tests.	**-Not addressed**	X	X
Rappeport 1996(TAR71)	-Index test	-Diagnosis	**Triage**: Single index test.	**-Not addressed**	√	√
Reed 1996(TAR72)	-Spectrum -Index test	-Diagnosis	**Add** (existing tests) Single index test Incremental accuracy not assessed	**-Sub-grouping**†	√	√
Reuchlin-Vrocklage 2005(TAR73)	-Spectrum -Index test	-Diagnosis	**Role unclear** Single index test compared to a reference standard. Interpretation assisted by CI.	**-Not addressed**.	X	X
Riedemann 2005(TAR74)	-Prevalence -Spectrum	-Diagnosis -Prognosis *Distinguished*	**Add** (no current test) Single index test	**-Pre to post test probability**	X	X
Rietveld 2003(TAR75)	NA (no included studies)	-Diagnosis	**- NA** (no included studies)	**- NA** (no included studies)	NA	NA
Rodgers 2006(TAR76)	X	-Screening -Diagnosis	**Multiple index tests: roles unclear**	**-Not addressed**	√	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Ross 1999[TAR77]	-Spectrum -Index test	-Diagnosis -Screening	**Replace**: Indirect comparison of competing index tests. Interpretation assisted by CI.	-**Not addressed**	X	X
Schmitt 2005[TAR78]	X	-Screening	**Add** (no current test). Single index test.	**-Inclusion restricted**	X -Limitations 1y studies -Invasive reference standard	X
Scholten 2001[TAR79]	-Spectrum -Organisation of healthcare	-Diagnosis	**Role** (to be determined by review) Direct and indirect comparison of competing index tests. Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Meta-regression**† **-Pre to post-test probability**	X	X
Schuijf 2006[TAR80]	X	-Diagnosis	**Triage** Indirect comparison of 2 competing index tests. Interpretation assisted by CI. Significance testing	**-Meta-regression** †	X	X
Scott 2002[TAR81]	-Prevalence -Spectrum -Index test	-Screening	**Add** (no current test) Indirect and direct comparison of competing index tests.	**-Not addressed**	X	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Scouller 2000(TAR82)	-Spectrum	-Diagnosis	**Replace** Indirect and direct comparison of competing index tests Interpretation assisted by CI. Significance testing	**-Sub-grouping**†	X	X
Selley 1997(TAR83)	X	-Screening -Diagnosis *Distinguished*	**Role unclear** Indirect comparison index tests. Assessment of incremental accuracy restricted to analysis in primary studies	**-Narrative** discussion of results according to setting	X	√
Singer 1992(TAR84)	X	-Screening	**Add** (no current test) Indirect comparison of competing index tests.	**-Not addressed**	X	X
Siu 1991(TAR85)	X	-Unclear	**Add** (no current test) Indirect comparison of competing index tests alone and in combination where addressed by primary studies. Interpretation assisted by CI.	**-Pre to post-test probability**	X	√
Stein 2004(TAR86)	-Spectrum -Organisation of healthcare	-Diagnosis	**Triage** Single index test Interpretation assisted by CI.	**-Meta-regression**† **-Narrative** discussion of potential impact of setting.	√	√
Stein 2006(TAR87)	-Spectrum	-Diagnosis	**Replace** (reference standard): Single index test Interpretation assisted by CI.	**-Sub-grouping**†	X	X

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Storgaard 1994[TAR88]	-Spectrum	-Screening	**Add** (no current test) Indirect comparison of competing index tests.	**-Sub-grouping**	√	X
Takata 2003[TAR89]	-Index test	-Diagnosis	**Role unclear** . Indirect comparison of competing index tests. Interpretation assisted by CI.	**-Pre to post-test probability**	X	X
Tamariz 2004[TAR90]	-Spectrum	-Diagnosis	**Replace** Indirect comparison of competing index tests. Interpretation assisted by CI. **Add** (existing tests) Assessment of incremental accuracy using indirect comparisons across primary studies. Interpretation assisted by CI.	**-Not addressed**	√	√
Tu 2005[TAR91]	-Spectrum	-Diagnosis	**Role unclear** Indirect comparison competing index tests compared to a reference standard	**-Not addressed**	X	X
Tugwell 1997[TAR92]	-Spectrum	-Diagnosis	**Role unclear**. Indirect comparison competing index tests compared to a reference standard.	**-Sub-grouping** **-Pre to post-test probability**	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
van den Hoogen 1995[TAR93]	-Spectrum -Index test	-Diagnosis.	**Add** (existing tests). Assessment of incremental accuracy using indirect comparisons across primary studies. **Replace** (existing tests) Indirect comparison of competing index tests	**-Not addressed**	X	X
van der Meer 2005[TAR94]	-Prevalence -Spectrum	-Diagnosis	**Add** (no current test) Single index test	**-Sub-grouping**† **-Narrative:** discussion of the potential impact of prevalence on post-test probability.	√	X
Wang 2005[TAR95]	-Index test	-Diagnosis	**Replace** (reference standard): Single index test Interpretation assisted by CI.	**-Meta-regression**† **-Narrative** discussion of potential impact of setting.	X	X
Waugh 2004[TAR96]	-Spectrum	-Diagnosis	**Replace** (reference standard): Single index test Interpretation assisted by CI.	**-Pre to post test probability**	X	X
Whiting 2005[TAR97]	-Spectrum	-Diagnosis	**Role** (to be determined by review): Indirect comparisons of multiple index tests compared to a reference standard. Interpretation assisted by CI.	**-Meta-regression**† **-Pre to post test probability**	√	√

Appendix 4.6: Detail of Contextualisation of Review Findings

Study ID	Consideration of applicability of included studies	Test application (if >1 do authors distinguish between alternatives)	Synthesis accounts for test role:	Synthesis accounts for characteristics of tested population (spectrum / prevalence):	Synthesis accounts for outcomes of testing -Two test dimensions distinguished -Consequences of test results discussed	
					Two dimensions	Con-sequences
Whiting 2006[TAR98]	-Spectrum -Index test	-Diagnosis	**Add** (existing tests): Assessment of incremental accuracy using indirect comparisons across primary studies.	**-Pre to post test probability** **-Sub-grouping**†	√	√
Wiese 2000[TAR99]	X	-Diagnosis	**Role unclear** Indirect comparison competing index tests compared to a reference standard Interpretation assisted by CI.	**-Pre to post test probability**	√	√
Zintzaras 2006[TAR100]	X	-Diagnosis	**Replace (reference standard)** Indirect comparison of competing index tests. Interpretation assisted by CI.	**-Not addressed**	X	X

Notes to table X:

Applicability of included studies: the extent to which review authors have related the characteristics of included studies to a specified target context or contexts. Test: variation in testing technology including experience / skill of operator and threshold. Organisation of healthcare: aspects of the organisation of healthcare that may impact on testing pathways and the setting in which tests are used. Spectrum: characteristics of tested population. Prevalence: prevalence of target disorder.

Test application: Distinguished: Authors distinguish between test applications by sub-grouping, meta-regression or narratively.

Test role: Assessment of incremental accuracy restricted to analysis in primary studies: Review authors rely on primary studies that have assessed the accuracy of existing tests with the addition of the index test.

Synthesis accounts for spectrum / prevalence:

-Restrict inclusion: Authors restrict inclusion according to spectrum or prevalence

-Sub-grouping: Authors sub-group results according to spectrum or prevalence

-Meta-regression: spectrum or prevalence used as a covariate in meta-regression.

-Pre-post-test probability: Authors illustrate the impact of prevalence on post-test probability

-Narrative: Authors distinguish results according to spectrum or prevalence in reporting of results or discussion

-Limited by primary studies: Review authors report being limited by poor reporting or quality of included studies
-†a priori statement of potential heterogeneity variables

Synthesis accounts for outcomes of testing:

-Two test dimensions distinguished: The ability of a test to increase or decrease the probability of the condition being tested for is clearly distinguished and linked to test accuracy outcomes, eg LR+ increase probability, LR- decrease probability; sensitivity and false negatives, specificity and false positives.
-Consequences of test results discussed: Authors include some discussion about the consequences of index test results (some or all of true positives, false positives, true negatives, false negatives).
- Diagnostic yield defined variably in these studies as true positives ÷ all tested or (true positives + false positives) ÷ all tested.

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Agency for Healthcare Research and Quality 2003[TAR1]	√	?	X	1	X	X	X	√	1	X	√	X	1	6
Anderson 2004[TAR2]	√	√	√	3	X	√	√	X	2	X	X	X	0	1
Appel 1993[TAR3]	√	√	√	3	X	X	X	X	0	X	X	X	0	1
Austin 2003[TAR4]	√	√	?	2	X	X	√	X	1	X	√	X	1	3
Barlow 1998[TAR5]	√	√	√	3	Poor reporting 1y studies	X	X	X	1	√	X	X	1	4
Bastian 1998[TAR6]	√	√	?	2	X	X	X	√	1	X	√	X	1	7
Battaglia 2006[TAR7]	√	√	?	2	X	√	√	√	3	X	X	√	1	7
Becker 1996[TAR8]	√	√	?	2	X	X	X	√	3	√	X	X	1	4
Berger 2000[TAR9]	√	√	√	3	√	√	√	X	3	√	X	X	1	3
Berry 2003[TAR10]	√	√	√	3	X	X	X	X	0	X	X	X	0	5
Brietzke 2004[TAR11]	√	√	√	3	X	X	X	X	0	X	X	X	0	4

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Chen 2001[TAR12]	√	√	?	2	X	√	X	√	2	X	√	X	1	6
Chunn 2004[TAR13]	√	X	?	1	?	√			1	X	X	X	0	3
Conde-Agudelo 2004[TAR14]	√	?	?	1	X	√	√	√	3	√	X	X	1	6
Cook 2005[TAR15]	√	√	√	3	√	√	√	√	4	√	X	√	2	5
de Bruyn 2001[TAR16]	√	√	X	2	√	?	X	√	2	√	√	X	2	5
Deville 2004[TAR17]	√	√	X	2	?	√	√	√	3	√	√	X	2	7
Dinnes 2003[TAR18]	√	?	√	2	Poor reporting 1y	√	√	√	4	√	X	√	2	7
Dodd 2006[TAR19]	√	√	X	2	X	X	X	√	1	X	X	X	0	5
Doust 2004[TAR20]	√	?	X	1	X	√	√	√	3	X	X	X	0	4
Fancher 2004[TAR21]	√	√	√	3	√	√	√	X	3	X	√	X	1	7
Fiellin 2000[TAR22]	√	√	?	2	√	√	√	√	4	√	√	X	2	4
Flemons 2003[TAR23]	√	√	√	3	X	√	√	X	2	√	√	√	3	6

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Fowler-Brown 2004[TAR24]	√	√	√	3	√	√	√	X	3	√	X	√	2	4
Fransen 2004[TAR25]	√	√	√	3	Unclear	X	X	X	0	√	X	√	2	3
Garber 1999[TAR26]	√	√	√	3	X	√	√	X	2	√	X	√	2	2
Gianrossi 1990[TAR27]	√	?	?	1	X	√	√	√	3	√	X	X	1	3
Gisbert 2001[TAR28]	√	?	?	1	X	X	X	X	0	√	X	X	1	1
Goodacre 2005(a) [TAR29]	√	√	?	2	?	X	√	√	2	√	X	√	2	9
Goodacre 2005(b) [TAR30]	√	?	X	1	?	√	√	√	3	√	√	√	3	8
Gorelick 1999[TAR31]	√	?	?	1	X	X	√	X	1	√	√	√	3	3
Harris 2003[TAR32]	√	√	√	3	X	X	X	X	0	X	X	X	0	6
Heim 2004[TAR33]	√	√	√	3	√	√	√	X	3	√	X	√	2	5
Hobbs 1997[TAR34]	√	?	?	1	?	X	X	√	1	√	√	X	2	9

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Huicho 1996[TAR35]	√	√	√	3	X	X	X	√	1	√	X	X	1	6
Huicho 2002[TAR36]	X	?	?	0	X	√	√	√	3	X	X	X	0	7
Ioannidis 2001[TAR37]	√	X	X	1	X	X	X	X	0	X	X	X	0	4
Jarvik 2002[TAR38]	√	X	√	2	X	√	X	√	2	X	X	X	0	3
Jorm 1997[TAR39]	X	√	X	1	X	X	X	X	0	X	X	X	0	3
Kearon 1998[TAR40]	√	√	?	2	√	√	X	X	2	√	X	X	1	4
Kim 2001[TAR41]	√	√	?	2	X	X	√	√	2	√	X	X	1	6
Kotler 1990[TAR42]	√	√	?	2	X	√	X	X	1	X	X	X	0	2
Kwok 1999[TAR43]	√	?	X	1	X	?	√	√	2	X	X	X	0	5
Law 1998[TAR44]	√	√	?	2	?	√	√	X	2	√	X	√	2	5
Lee 2006[TAR45]	X	?	?	0	X	X	√	X	1	X	X	X	0	2
Lewis 2006[TAR46]	√	√	X	1	X	√	X	√	2	X	√	√	2	3
Linzer 1997[TAR47]	√	?	X	1	X	√	X	X	1	X	X	X	0	1

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Loy 1996[TAR48]	√	√	X	2	X	√	X	√	2	X	X	X	0	4
Maguire 2005[TAR49]	√	X	?	1	?	X	X	X	0	√	X	X	1	6
Mant 2004[TAR50]	√	?	√	2	Poor reporting 1y	√	X	√	3	√	√	X	2	7
Marshall 1996[TAR51]	√	√	?	2	X	X	√	√	2	√	X	X	1	5
Marx 2005[TAR52]	√	√	√	3	X	√	√	√	3	X	X	X	0	7
McGowan 2003[TAR53]	√	√	X	2	X	X	X	X	0	√	√	X	2	2
Mohseni-Bandpei 2000[TAR54]	X	X	?	0	X	√	X	√	2	X	√	X	1	4
Mourad 2003[TAR55]	√	?	?	1	X	X	X	√	1	X	X	X	0	3
Nayak 2006[TAR56]	√	√	√	3	X	X	X	√	1	√	√	√	3	7
Nelson 2006[TAR57]	√	?	√	2	X	X	X	√	1	√	√	X	2	6
Numans 2004[TAR58]	√	?	√	2	?	√	√	X	2	X	X	X	0	6
Oei 2003[TAR59]	√	√	?	2	X	X	X	√	1	X	X	X	0	7

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Ogilvie 2005[TAR60]	√	√	√	3	X	X	X	√	1	√	√	X	2	6
O'Meara 2006[TAR61]	√	√	√	3	√	X	X	√	1	√	√	X	2	6
Oosterhuis 2000[TAR62]	√	√	?	2	X	X	X	√	1	√	√	X	2	5
Owens 1996[TAR63]	√	√	?	2	√	√	X	√	3	√	X	X	1	6
Pasternack 2003[TAR64]	√	√	√	3	X	√	√	√	3	√	X	X	1	6
Peters 2003[TAR65]	X	?	X	0		√	?	√	2	X	X	X	0	3
Petersen 2001[TAR66]	√	√	?	2	√	√	X	√	3	X	X	X	0	5
Pignone 2002[TAR67]	√	?	√	2	Poor reporting 1y	X	√	√	3	√	X	X	1	6
Pirozzo 2003[TAR68]	X	?	X	0		X	?	X	0	√	√	X	2	5
Price 2005[TAR69]	√	√	?	2	√	X	√	√	3	√	√	X	2	6
Ramsay 2002[TAR70]	√	√	?	2	√	X	√	√	3	√	√	X	2	6
Rappeport 1996[TAR71]	√	?	√	2	?	√	X	√	2	X	√	X	1	1

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Reed 1996[TAR72]	√	?	X	1	X	X	X	√	1	√	√	X	2	3
Reuchlin-Vrocklage 2005[TAR73]	√	?	√	2	√	√	X	√	3	√	√	X	2	7
Riedemann 2005[TAR74]	√	√	?	2	?	√	?	X	1	√	X	X	1	2
Rietveld 2003[TAR75]	√	?	√	2	X	X	X	√	1	X	X	X	0	4
Rodgers 2006[TAR76]	√	?	√	2	Poor reporting 1y	√	X	√	3	X	X	X	0	7
Ross 1999[TAR77]	√	√	√	3	Poor reporting 1y studies	X	X	√	2	√	√	X	2	0
Schmitt 2005[TAR78]	√	√	√	3	Poor reporting 1y studies	Poor reporting 1y studies	Poor reporting 1y studies	√	4	√	√	X	2	5
Scholten 2001[TAR79]	√	?	?	1	X	?	√	√	2	√	X	X	1	6
Schuijf 2006[TAR80]	√	√	?	2	X	X	√	X	1	X	X	X	0	2
Scott 2002[TAR81]	√	?	√	2	√	√	√	X	3	√	√	√	3	2
Scouller 2000[TAR82]	√	√	X	2	X	X	X	√	1	√	X	X	1	4

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
Selley 1997[TAR83]	√	?	√	2	?	√	√	√	3	X	X	X	0	2
Singer 1992[TAR84]	√	?	?	1	X	X	X	X	0	X	X	X	0	1
Siu 1991[TAR85]	X	√	X	1	X	?	X	√	1	X	X	X	0	3
Stein 2004[TAR86]	√	√	X	2	?	√	√	X	2	√	X	X	1	4
Stein 2006[TAR87]	√	√	?	2	X	?	X	X	0	√	X	X	1	6
Storgaard 1994[TAR88]	√	√	X	2	X	X	√	X	1	√	X	X	1	2
Takata 2003[TAR89]	√	?	X	1	X	X	√	√	2	X	√	X	1	7
Tamariz 2004[TAR90]	√	X	√	2	Poor reporting 1y	X	X	√	2	√	X	X	1	6
Tu 2005[TAR91]	√	X	X	1	X	√	X	X	1	√	X	X	1	2
Tugwell 1997[TAR92]	√	?	√	2	√	√	X	X	2	√	X	X	1	3
van den Hoogen 1995[TAR93]	√	√	√	3	X	√	X	√	2	√	√	X	2	3

Appendix 4.7: Reporting of study characteristics, consideration of applicability of review findings and methodological quality of reviews reporting the most complete question formulation

Study ID	Question formulation				Reporting study characteristics					Applicability of Primary Studies considered				Review Quality Score /9
	Index applic	Index role	Prior tests	TOTAL /3	Prior tests	Presentation	Prevalence target disorder	Quality included studies	TOTAL /4	Spectrum	IndexTest	Prevalence	TOTAL /3	
van der Meer 2005[TAR94]	√	X	√	2	√	√	X	√	3	√	X	√	2	7
Wang 2005[TAR95]	√	√	?	2	√	√	√	√	4	X	√	X	1	6
Waugh 2004[TAR96]	√	√	?	2	X	X	√	√	2	√	X	X	1	7
Whiting 2005[TAR97]	√	?	X	1	√	√	√	√	4	√	X	X	1	7
Whiting 2006[TAR98]	√	√	?	2	?	√	√	√	3	√	√	X	2	7
Wiese 2000[TAR99]	√	?	?	1	?	√	√	√	3	X	X	X	0	8
Zintzaras 2006[TAR100]	√	√	?	2	X	√	X	X	1	X	X	X	0	6

Notes to table X:

? = unclear

Quality assessment items (9): Search>1 bibliographic database; one or more of hand-searching, contact with experts, reference checking; quality assessment; explicit inclusion / exclusion criteria; data extraction in duplicate; study flow documented; discussion of review limitations; level of agreement inclusion; level of agreement quality assessment.

DIAGNOSTIC TEST ACCURACY GP SURVEY

Intro1

Doctors.net.uk invites you to participate in a survey commissioned by a leading UK University. The aim of this survey is to identify which methods of communicating information about screening and diagnostic test accuracy are likely to be most useful in clinical practice. If you have any queries about this research please contact Dr Clare Davenport, Clinical Research Fellow, MB ChB MSc FFPH, C.F.Davenport@bham.ac.uk The survey will take around 20 minutes to complete. All eligible members completing the survey will receive 4,000 eSR points. Please read the following text, which explains the intent of this research.Doctors.net.uk would like to reassure you that:Doctors.net.uk will comply with all UK laws protecting your personal data and the British Healthcare Business Intelligence Association and Market Research Society guidelines. Your responses will be used by us and the sponsoring University for social research and public relations purposes only. Your responses will be collated with other respondents and presented to the sponsor in aggregated or anonymised form. Your responses will be confidential and will not be used for any other purposes or disclosed to any third party without your approval. Please confirm that you have read and understood this information.

❍ Yes (1)
❍ No (2)

Intro2

You are about to enter a social research interview. We are now being asked to pass on to our client details of adverse events and / or product complaints that are raised during the course of social research interviews. Although this is an on-line social research interview and how you respond will, of course, be treated in confidence, should you raise an adverse event and / or product complaint, we will need to report this, even if it has already been reported by you directly to the company or the regulatory authorities using the MHRA's 'Yellow Card' system. In such a situation you will be contacted to ask whether or not you are willing to waive the confidentiality given to you under the market and social research codes of conduct specifically in relation to that adverse event and / or product complaint. Everything else you contribute during the course of the interview will continue to remain confidential. Are you happy to proceed with the interview on this basis?

❍ I would like to proceed and protect my anonymity (1)
❍ I would like to proceed and give permission for my contact details to be passed on to the Drug Safety department of the company if an adverse event is mentioned by me during the survey (2)
❍ I don't want to proceed and end the interview here (3)

We will now ask you 2 questions to check that this survey is relevant to you.

S1 Cohort

Please indicate which of the following applies to you:

❍ GP Principal (1)
❍ GP Non-principal (2)
❍ GP Locum (3)
❍ Primary Care Nurse (4)
❍ Other (5)

S2 Region

Where are you currently practicing?

❍ North East SHA (1)
❍ North West SHA (2)
❍ Yorkshire & Humber SHA (3)
❍ West Midlands SHA (4)
❍ East Midlands SHA (5)
❍ East of England SHA (6)

- ❍ South West SHA (7)
- ❍ South Central SHA (8)
- ❍ South East Coast SHA (9)
- ❍ London SHA (10)
- ❍ Scotland (11)
- ❍ Wales (12)
- ❍ Northern Ireland (13)
- ❍ Retired (14)
- ❍ Not working in the UK (15)

Thank you. We are now confident that this survey is relevant to you. The rest of the survey explores which methods of communicating information about screening and diagnostic test accuracy are likely to be most useful in clinical practice.

Q1 Sources of Information about Test Accuracy (1)

Please estimate how often you use the following test accuracy information sources of as part of your clinical work

	Always use source (1)	Often use source (2)	Rarely use source (3)	Never use source (4)	Cannot estimate how often use source (5)
Own clinical experience and training (1)	☐	☐	☐	☐	☐
Colleagues (2)	☐	☐	☐	☐	☐
Textbooks (3)	☐	☐	☐	☐	☐
Research papers (4)	☐	☐	☐	☐	☐
Guidelines (5)	☐	☐	☐	☐	☐
Manufacturer's information (6)	☐	☐	☐	☐	☐
Normal range (eg blood test) (7)	☐	☐	☐	☐	☐
Other source of information (8)	☐	☐	☐	☐	☐

Please specify the 'Other source of information' that you use

Q2a Sources of Information about Test Accuracy (2)

Please indicate which statements apply to the following test accuracy information sources

Please select all that apply

	I don't know how to access the source (1)	I can't access the source at the time I need information (2)	The source uses terminology I don't understand (3)	The source does not contain information relevant to my practice (4)	None of the statements apply (6)
Textbooks (1)	☐	☐	☐	☐	☐
Research papers (2)	☐	☐	☐	☐	☐
Guidelines (3)	☐	☐	☐	☐	☐
Manufacturer's information (4)	☐	☐	☐	☐	☐

Q2b Sources of Information about Test Accuracy (3)

Please use the space below to make any additional comments about these sources of test accuracy

Q3 Sensitivity and specificity (1)

Have you heard of the measures 'sensitivity' and 'specificity'?

❍ Yes (1)
❍ No (2)

Q4 Sensitivity and specificity (2)

How confident would you be in defining sensitivity and specificity to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Sensitivity (1)	☐	☐	☐	☐
Specificity (2)	☐	☐	☐	☐

Q5 Sensitivity and specificity (3)

Do you use sensitivity and specificity in clinical practice?

❍ Yes (1)
❍ No (2)

Q6 Sensitivity and specificity (4)

Please comment on how you use the measure sensitivity in practice

Q7 Sensitivity and specificity (5)

Please comment on how you use the measure specificity in practice

Q8 Positive predictive value (PPV) and negative predictive value (NPV) (1)

Have you heard of the measures 'positive predictive value (PPV)' and 'negative predictive value (NPV)'?

❍ Yes (1)
❍ No (2)

Q9 Positive predictive value (PPV) and negative predictive value (NPV) (2)

How confident would you be in defining positive predictive value (PPV) and negative predictive value (NPV) to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Positive predictive value (PPV) (1)	☐	☐	☐	☐
Negative predictive value (NPV) (2)	☐	☐	☐	☐

Q10 Positive predictive value (PPV) and negative predictive value (NPV) (3)

Do you use the positive predictive value (PPV) and negative predictive value (NPV) in clinical practice?

❍ Yes (1)
❍ No (2)

Q11 Positive predictive value (PPV) and negative predictive value (NPV) (4)

Please comment on how you use the measure positive predictive value (PPV) in practice

Q12 Positive predictive value (PPV) and negative predictive value (NPV) (5)

Please comment on how you use the measure negative predictive value (NPV) in practice

Q13 Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) (1)

Have you heard of the measures 'positive likelihood ratio (LR+)' and 'negative likelihood ratio (LR-)'?

❍ Yes (1)
❍ No (2)

Q14 Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) (2)

How confident would you be in defining positive likelihood ratio (LR+) and negative likelihood ratio (LR-) to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Positive likelihood ratio (LR+) (1)	☐	☐	☐	☐
Negative likelihood ratio (LR-) (2)	☐	☐	☐	☐

Q15 Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) (3)

Do you use the positive likelihood ratio (LR+) and negative likelihood ratio (LR-) in clinical practice?

❍ Yes (1)
❍ No (2)

Q16 Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) (4)

Please comment on how you use the measure positive likelihood ratio (LR+) in practice

Q17 Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) (5)

Please comment on how you use the measure negative likelihood ratio (LR-) in practice

Q18 Diagnostic table (1)

Have you ever seen a table like the one below?

❍ Yes (1)
❍ No (2)

Example of a 2x2 Diagnostic Table

	DISEASE PRESENT *Diagnosis verified by the most accurate test available (the reference standard test)*	**DISEASE ABSENT** *Diagnosis verified by the most accurate test available (the reference standard test)*
POSITIVE TEST RESULT *of the test under evaluation (the index test)*	**TRUE POSITIVES**	**FALSE POSITIVES**
NEGATIVE TEST RESULT *of the test under evaluation (the index test)*	**FALSE NEGATIVES**	**TRUE NEGATIVES**

Q19 Diagnostic table (2)

How confident would you be in explaining the 2x2 diagnostic table to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
2x2 diagnostic table (1)	☐	☐	☐	☐

Q20 Diagnostic table (3)

Do you use the 2x2 diagnostic table in clinical practice?

❍ Yes (1)
❍ No (2)

Q21 Diagnostic table (4)

Please comment on how you use the 2x2 diagnostic table in practice

Q22 Receiver operator characteristic (ROC) curve (1)

Have you heard of the 'receiver operator characteristic (ROC) curve'?

❍ Yes (1)
❍ No (2)

Q23 Receiver operator characteristic (ROC) curve (2)
How confident would you be explaining the receiver operator characteristic (ROC) curve to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Receiver operator characteristic (ROC) (1)	☐	☐	☐	☐

Q24 Receiver operator characteristic (ROC) curve (3)

Do you use the receiver operator characteristic (ROC) curve in clinical practice?

❍ Yes (1)
❍ No (2)

Q25 Receiver operator characteristic (ROC) curve (4)

Please comment on how you use the receiver operator characteristic (ROC) curve in practice

Q26 Diagnostic odds ratio (1)

Have you heard of the measure 'diagnostic odds ratio (DOR)'?

❍ Yes (1)
❍ No (2)

Q27 Diagnostic odds ratio (2)

How confident would you be in defining the diagnostic odds ratio (DOR) to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Diagnostic odds ratio (DOR) (1)	☐	☐	☐	☐

Q28 Diagnostic odds ratio (3)

Do you use the diagnostic odds ratio (DOR) in clinical practice?

❍ Yes (1)
❍ No (2)

Q29 Diagnostic odds ratio (4)

Please comment on how you use the diagnostic odds ratio (DOR) in practice

Q30 AREA UNDER the receiver operator characteristic CURVE (AUC) (1)

Have you heard of the measure 'Area Under the receiver operator characteristic Curve (AUC)'?

❍ Yes (1)
❍ No (2)

Q31 AREA UNDER the receiver operator characteristic CURVE (AUC) (2)

How confident would you be in defining the Area Under the receiver operator characteristic Curve (AUC) to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Area Under the receiver operator characteristic Curve (AUC) (1)	☐	☐	☐	☐

Q32 AREA UNDER the receiver operator characteristic CURVE (AUC) (3)

Do you use the Area Under the receiver operator characteristic Curve (AUC) in clinical practice?

❍ Yes (1)
❍ No (2)

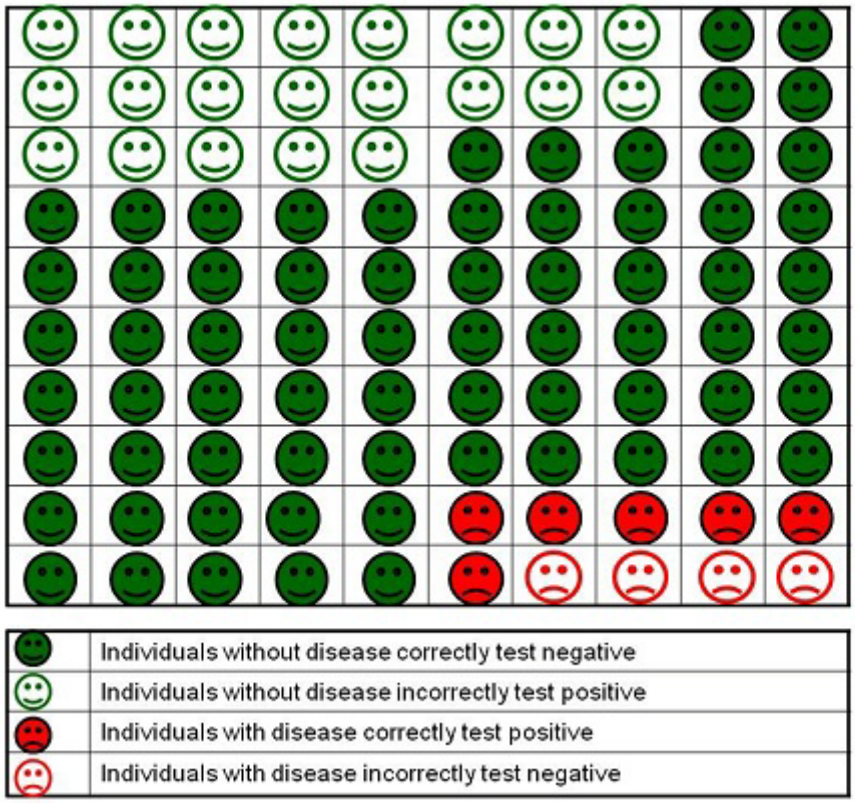
Q33 AREA UNDER the receiver operator characteristic CURVE (AUC) (4)

Please comment on how you use the Area Under the receiver operator characteristic Curve (AUC) in practice

Q34 Diagrammatic methods for representing test accuracy (1)

Have you ever seen a diagram like the one below?

❍ Yes (1)
❍ No (2)



Q35 Diagrammatic methods for representing test accuracy (2)

How confident would you be in explaining the meaning of such a diagram to a colleague?

	Very confident (1)	Moderately confident (2)	Not very confident (3)	Could not define (4)
Test Accuracy Diagram (1)	☐	☐	☐	☐

Q36 Diagrammatic methods for representing test accuracy (3)

Do you use diagrammatic methods of presenting test accuracy, similar to the one displayed on the previous page, in clinical practice?

❍ Yes (1)
❍ No (2)

Q37 Diagrammatic methods for representing test accuracy (4)

Please comment on how you use diagrammatic methods of presenting test accuracy in practice

Q38a Scenario I (1)

The questions that follow will ask you to apply test accuracy information to clinical scenarios. Each scenario reflects a primary care setting where the prevalence of ovarian cancer in asymptomatic, post-menopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. The marker has a sensitivity of 76% and a specificity of 98%. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q38b Scenario I (2)

Please use this space for any comments about the scenario or your response above.

Q39a Scenario II (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. Of every 100 women who test positive with the marker, 54 will have ovarian cancer but 46 will not. Of every 100 women who test negative with the marker 99 will not have ovarian cancer but 1 will have ovarian cancer and be missed. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q39b Scenario II (2)

Please use this space for any comments about the scenario or your response above.

Q40a Scenario III (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. The marker has a positive likelihood ratio (LR+) of 38 and a negative likelihood ratio (LR-) of 0.2. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q40b Scenario III (2)

Please use this space for any comments about the scenario or your response above.

Q41a Scenario IV (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. The following table represents the accuracy of a new blood test for the detection of ovarian cancer when evaluated in a cohort of 1360 asymptomatic, postmenopausal women. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and requests the test.

	Women with confirmed ovarian cancer (based on surgery and long term clinical follow up)	**Women confirmed free of ovarian cancer (based on surgery and long term clinical follow up)**	
New blood test for detecting ovarian cancer: POSITIVE RESULT	**31** women with ovarian cancer correctly test +ve with the new blood test **(TRUE +VES)**	**26** women without ovarian cancer incorrectly test +ve with the new blood test **(FALSE +VES)**	
New blood test for detecting ovarian cancer: NEGATIVE RESULT	**10** women with ovarian cancer incorrectly test –ve with the new blood test **(FALSE -VES)**	**1293** women without ovarian cancer correctly test –ve with the new blood test **(TRUE -VES)**	
	41 women, in total, with confirmed ovarian cancer	**1319** women, in total, confirmed free of ovarian cancer	**1360** women tested in total

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q41b Scenario IV (2)

Please use this space for any comments about the scenario or your response above.

Q42a Scenario V (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. In low prevalence populations the marker has a positive predictive value (PPV) of 54% and a negative predictive value (NPV) of 99%. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q42b Scenario V (2)

Please use this space for any comments about the scenario or your response above.

Q43a Scenario VI (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. A positive test result raises the probability of having ovarian cancer from ~3% to ~54%. A negative test result lowers the probability of having ovarian cancer from ~3% to ~0.6%. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q43b Scenario VI (2)

Please use this space for any comments about the scenario or your response above.

Q44a Scenario VII (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. The Diagnostic Odds Ratio (DOR) is a way of expressing test accuracy. A DOR of 1 indicates a useless test and values between 1 and infinity indicate an increasingly accurate test. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. The marker has a DOR of 190. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q44b Scenario VII (2)

Please use this space for any comments about the scenario or your response above.

Q45a Scenario VIII (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. A new biological marker for ovarian cancer has been identified and is available as a blood test for use in primary care. Of every 100 women with ovarian cancer, 76 would test positive (be detected by the test) but 24 would test negative (be missed). Of every 100 women without ovarian cancer, 98 would test negative (receive a correct diagnosis) but 2 would test positive (be falsely labelled as having cancer). A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.

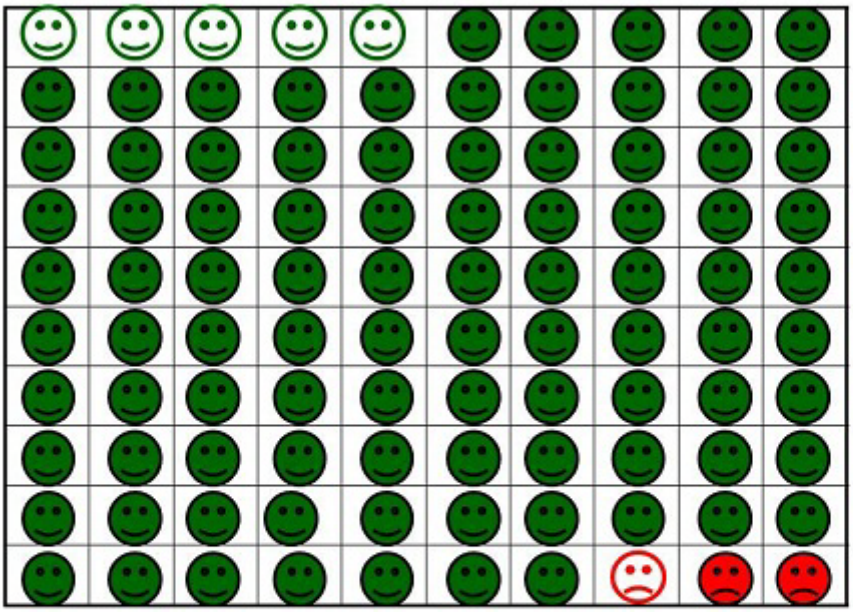
	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q45b Scenario VIII (2)

Please use this space for any comments about the scenario or your response above.

Q46a Scenario IX (1)

This scenario is set in a primary care setting where the prevalence of ovarian cancer in asymptomatic, postmenopausal women is ~3%. The following diagram represents the accuracy of a new blood test for ovarian cancer in a female population with a prevalence of ovarian cancer of 3%. The blood test is available for use in primary care. A 57 year old asymptomatic woman presents to you concerned about her risk of ovarian cancer and you perform the blood test at her request.



	Women without ovarian cancer correctly test negative with the new blood test
	Women without ovarian cancer incorrectly test positive with the new blood test
	Women with ovarian cancer correctly test positive with the new blood test
	Women with ovarian cancer incorrectly test negative with the new blood test

	Yes (1)	No (2)	Don't know (3)
If the test came back positive would you refer the woman for further investigation? (1)	☐	☐	☐
If the test came back negative would you be confident not to investigate further at this point in time? (2)	☐	☐	☐

Q46b Scenario IX (2)

Please use this space for any comments about the scenario or your response above.

Q47a Acceptable levels of test errors (1)

When tests are applied in a screening / triaging context, asymptomatic individuals who test positive undergo further definitive testing (for example following referral to secondary care). Individuals who test negative usually do not receive any further testing unless they re-present with new symptoms. The clinical significance of false positive test errors (individuals without disease who test positive) depends on the risks associated with further investigation and the clinical significance of false negative test errors (individuals with disease who test negative) depends on the risks associated with missed, untreated disease.If a test was being used to screen asymptomatic individuals for a potentially serious condition, such as cancer, please indicate on the scale below an acceptable level of missed disease (% false negative test results) and an acceptable level of healthy individuals wrongly labelled as having disease (% false positive test results) that you would tolerate from the test before you would consider it accurate enough to be used for this purpose.

	≤1% (1)	≤5% (2)	≤10% (3)	≤15% (4)	≤20% (5)	≤25% (6)	≤30% (7)	≤35% (8)	≤40% (9)	≤45% (10)	≤50% (11)
Acceptable % of diseased individuals missed by the test (false negatives) (1)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Acceptable % of healthy individuals wrongly labelled as having disease by the test (false positives) (2)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Q47b Acceptable levels of test errors (2)

Please use this space for any comments concerning the acceptability of test errors as indicated above.

D1 Years since qualification

And finally a few questions about you. When did you qualify in this specialty?

[ ]

D2 Gender

Are you...

❍ Male (1)
❍ Female (2)

D3 Mode of work

Do you currently work full-time or part-time?

❍ Full-time (1)
❍ Part-time (2)

D4 Responsibilities/roles

Please select all work responsibilities/roles that apply to you:

❑ Clinical primary care (1)
❑ GP registrar (2)
❑ GP with a special interest (3)
❑ GP trainer/course organiser (4)
❑ Academic position associated with a University/Deanery (5)
❑ Local, regional or national policy/guideline development (6)
❑ Other (please specify) (7)_____

D5 Training

Have you undertaken any training that included interpretation of test accuracy measures in the last 3 years?

❍ Yes (1)
❍ No (2)

D6 Feedback

Thank you for taking the time to complete this questionnaire. If you have any questions about this research please contact Dr Clare Davenport MBChB MSc FFPH, c.f.davenport@bham.ac.uk

Please use this space to provide any feedback specifically about this questionnaire or about research on tests more generally.

GPnotebook (www.gpnotebook.co.uk)

GPnotebook is a reference guide conceived initially by UK doctors. The content of GP notebook is based on clinical practice in the United Kingdom and provides a clinical reference guide for general practitioners and medical students. The system content is continually updated and expanded and is included as a resource on the UK National Electronic Library of Health Virtual Branch Libraries in Primary Care and Emergency Care. Content is guided by a 'pragmatic' approach to searching based on topical issues and relying on hand searching selected journals as well as clinical experience and guidelines published by national and international bodies. Peer review is internal to the authors.

Patient UK (www.patient.co.uk)

Patient UK was launched in 1997 by PiP (Patient Information Publications). The aim of this website was to provide non-medical people in the UK with good quality information about health and disease. It started as a directory of UK websites which provided information on health, disease and related issues. Lead authors and editors are general practitioners practising in the UK. Content is 'clinically peer-reviewed'.

In December 2002, Patient UK was re-launched as a joint venture between PiP and EMIS (Egton Medical Information Systems) aimed at the UK general public although the PatientPlus section is suggested to be of particular interest to health professionals.

EMIS (www.emis-online.com/mentor-)

EMIS was developed by UK general practitioners. EMIS develops, supplies and supports General Practice computing systems and has 50% coverage of practices in the UK. EMIS aims to facilitate timely access to 'appropriate' medical information. Information leaflets aimed at patients and carers are included on the website. EMIS Mentor is a reference library although it is unclear what proportion of reference articles are peer reviewed. Information

pertaining to testing appears limited to 'meaningful and relevant differential diagnosis for symptoms and test results' rather than quantitative test accuracy information.

Doctors net (http://www.doctors.net.uk/)

Doctors.net.uk is available to UK-registered doctors in primary and secondary care. It is a secure service offering a professional e-mail facility, clinical and non-clinical forums, medical news and free accredited education. Content is peer reviewed by clinical advisors.

Clinical Evidence (http://clinicalevidence.bmj.com)

Clinical Evidence is described as a decision-support resource underpinned by internationally peer reviewed systematic reviews which are regularly updated and integrated with a range of additional evidence based medicine resources. Clinical evidence is owned by the BMJ Publishing Group. Content is driven by questions rather than by the availability of research. Clinical evidence also includes EBM training including training concerned with test evaluation.

NHS Clinical Knowledge Summaries (formerly PRODIGY) (http://www.cks.nhs.uk/home)

The NHS Clinical Knowledge Summaries (CKS) (formerly PRODIGY) were commissioned by the National Institute for Health and Clinical Excellence (NICE). However as of March 2011 the content of CKS is no longer being maintained. CKS is aimed at healthcare professionals working in primary and first-contact care.