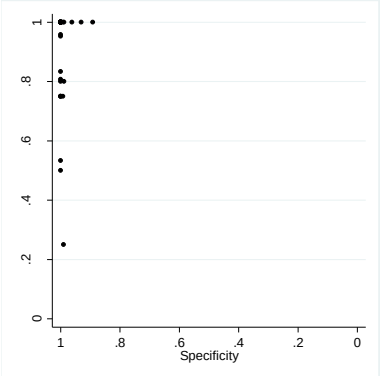
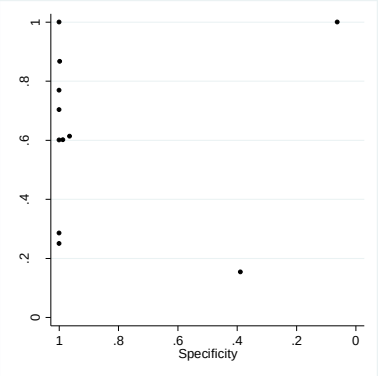
Supplemental Material (Online Only)

Supplementary Figure 1 ROC plots for reviews causing extreme underestimation of DOR with ML models (using unadjusted data)

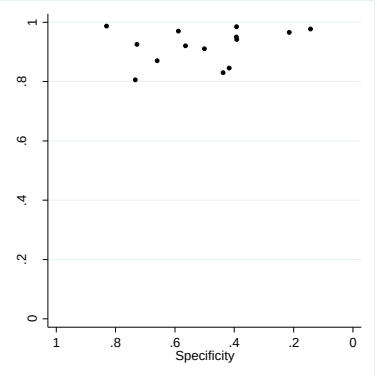
a. Dijkhuizen et al [25]



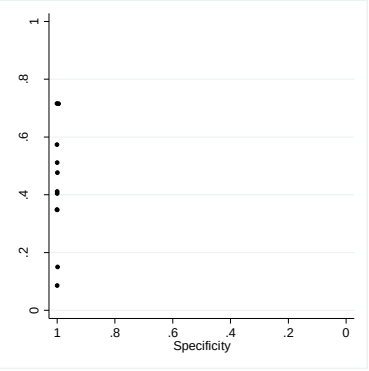
b. Medical Services Advisory Committee [28]



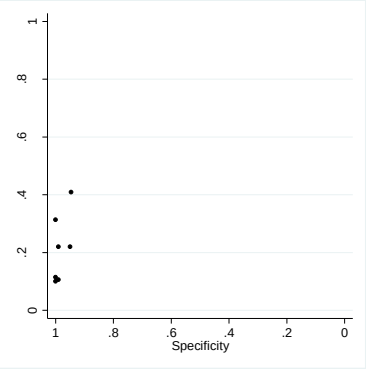
c. Nallamothu et al [24]



Bricker et al [26]



Eden et al [27]



Supplemental Material (Online Only)

Supplementary Table 1 Review details

Review details				Review characteristics			
Author Year	**Target disorder**	**Test**	**Covariates investigated (no. studies)**	**No. studies**	**Median sample size (SD)**	**Range in 'S'***	**Zero cells (%)†**
Balk 2001[1]	acute cardiac ischemia	presentation myoglobin	Hospitalised patients (14) vs Emergency Department patients (18)	32	101 (355)	7.9	3%
Bricker 2000[2]	pregnancy	ultrasound	Tertiary (4) vs Primary/Secondary care (7) Second trimester (6) vs Any trimester (5) Low risk (4) vs Unselected (7)	11	7,575 (9,324)	5.0	2%
Buchanan 2001[3]	dangerous severe personality disorder	clinical assessment	Prison release (8) vs Community/hospital discharge (13) Time at risk <=20 months (10) vs >20 months (8)	21	293 (880)	8.0	0%
Chapell 2002[4]	carpal tunnel syndrome	distal motor latency: symptoms/presented patient groups	Possible age bias (4) vs no bias or not reported (9) Possible bias to easy cases (5) vs no bias (8) ‡ Symptomatic/presented (8) vs unspecified diagnosis (5)	13	85 (115)	3.9	15%
Delgado 2003[5]	detection of primary tumours in patients with metastasis	F18-FDG PET	Any unknown primary tumour (8) vs unknown primary tumour with cervical adenopathies or intra-/extra-cranial metastases (7)	15	20 (12)	6.4	18%
Dijkhuizen 2000[6]	endometrial carcinoma	endometrial sampling	Pre- and post-menopausal women (22) vs post-menopausal women only (7) Asymptomatic only (20) or symptomatic women included (13)	33	120 (174)	7.8	34%
Eden 2001[7]	thyroid cancer screening	palpation	Environmental exposure (3) vs medical exposure or unexposed (4)	7	102 (781)	3.5	11%
Flemons 2003[8]	sleep apnoea	sleep monitors	Home setting (13) vs sleep laboratory setting (36) ‡ <75% male (10) vs 75-100% male (29) mean Apnea Hypopnea Index <=30 (15) vs >30 (17)	49	71 (129)	10.1	7%
Flobbe 2002[9]	breast cancer	mammography	mean body mass index <=30 (9) vs >30 (25)	22	213 (478)	5.8	0%
Gifford 2000[10]	potentially reversible causes of dementia	clinical assessment	age <=70 (3) years >70 (8) dementia/memory clinic setting (5) vs other (6) ‡ diagnostic criteria met (6) vs referrals (5)	11	202 (108)	6.5	9%
Glas 2003[11]	primary bladder cancer	cytology	<30% Grade 1 tumours (14) vs >30% (6) <30% Grade 2 tumours (6) vs >30% (14) <30% Grade 3 tumours (8) vs >30% (12) ‖ 100% Urological controls (4) vs nonurological patients and/or healthy controls (6) [case-control studies only]	26	107 (76)	7.7	9%
Gould 2001[12]	lung cancer	FDG-PET	>=70% men (14) vs <70% men (14) <60 years old (7) versus >=60 years (17)	35	46 (27)	7.6	15%
Gould 2003[13]	mediastinal staging of non small cell lung cancer	PET	>=70% men (12) vs <70% men (10) <60 years old (4) versus >=60 years (21)	33	49 (44)	4.5	8%

Review details				Review characteristics			
Author Year	**Target disorder**	**Test**	**Covariates investigated (no. studies)**	**No. studies**	**Median sample size (SD)**	**Range in 'S'***	**Zero cells (%)†**
Gray 2000[14]	oral cancer	toludine blue dye in visual screening	Clinical suspicion/lesions (10) vs cancer history (4)	14	85 (301)	6.5	9%
Ioannidis[15]	acute myocardial infarction	out-of-hospital ECG	Symptoms suggestive of acute cardiac ischaemia (4) vs chest pain (6) Age <65 years (3) vs >=65 (4) <65% men (3) vs >=65% men (4) §	10	295 (439)	10.8	5%
Kittler 2002[16]	melanoma	dermoscopy	Non-melanocytic lesions excluded (4) versus Included (9)	13	172 (890)	7.6	2%
Koelemay 2001[17]	peripheral arterial disease - aortoiliac tract	MRA	Age <65 years (9) vs >=65 (7) <70% men (7) vs >=70% (11) ‖ <65% intermittent claudication (5) vs >=65% (10)	19	96 (71)	7.2	13%
MSAC 2002[18]	fragile X syndrome	cytogenetic tests	<50% male (6) vs >=50% male (6) Fragile X pedigree/families (8) vs definite/suspected/prenatal (4)	12	77 (176)	13.8	21%
Nallamothu 2001[19]	coronary artery disease	electron beam computed tomography	Age <55 years (5) vs >=55 years (9) ‡ <65%male (7) vs >=65% male (7)	14	104 (63)	5.1	0%
Patwardhan 2004[20]	Alzheimer disease dementia	PET	Age <70 (11) vs >=70 years (5) Healthy controls (13) vs diseased controls (6)	19	43 (31)	8.0	5%
Romagnuolo 2003[21]	bilary disease - detection of stones	MRI cholangiopancreatography	Range possible diagnoses (11) vs stones or cancer diagnoses (35)	46	63 (53)	6.4	15%
Sauerland 2004[22]	pelvic fractures	clinical examination	Adults (10) vs children (3)	13	219 (577)	7.9	8%
Sotiriadis 2003[23]	Down syndrome	intracardiac echogenic foci	Age <=30 years (4) vs >30 (8) High risk (7) vs low risk/routine (5)	12	4,308 (4,642)	7.9	0%
Varonen 2000[24]	acute maxillary sinusitis	ultrasound	ENT clinic setting (3) vs general clinic (4)	7	156 (74)	4.5	0%
Visser 2000[25]	peripheral arterial disease	Duplex ultrasound	<=60% men (8) vs >60% (8) Age <=65 years (8) vs >65 (8) ‖ N America (14) vs other location (7)	21	404 (739)	6.8	4%
Whitsel 2000[26]	autonomic failure in diabetes	Bazett's heart rate-corrected QT interval (QTc)	Age <=40 years (8) vs >40 years (8) <=50% men (5) vs >50% men (11) ‖ <=50% type 1 diabetes (5) vs >50% (10) Mean duration <=10 years (10) vs >10 years (4)	17	58 (772)	6.7	4%

* range in 'S' is based on values for 'S' from ML model, where *S* = logit(sensitivity) + logit(1 - specificity). 1 – range 3 to <6; range 6 to <8; range 8+

† number of zero false positive and false negative cells as a percentage of the total number of cells per analysis. 1 - <5%; 2 – 5 to 10%; 3 - >10%

‡ model would not converge for HSROC model either with parallel curves or non-parallel curves

§ model would not converge for HSROC model with parallel curves

‖ model would not converge for HSROC model either with non-parallel curves