Studios of the	accuracy of	tasts to r	ula in or r	ula out d	100000
Studies of the	accuracy of			uie out u	150050

Spectrum of patientsStudyStudyStudyDiagnostic tests are designed to resolveenrolled in the studyconsists ofconsists ofconsists ofresolvestudypatients likely patients likely to receive the differentialpatients whose differentialpatients whodiagnostic uncertainties; if the targetdiagnostic the positive testfest in clinical diagnosisdiagnosisthe targetthe positive test subjects havesubjects have advanceddifferential diagnosisbesides theinformation, target disease, alreadydisease basedsubjects have advancedincludes the but alsobut in whomwho are clearly who are clearlybe biased upwards; if the negative testincludesalreadylow likelihood of having the target diseaseon already availablesubjects are clearly haveincludesalreadyon already availableon already target diseaseclearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive	Criterion	Green	Yellow	Red	Comments
patients enrolled in the studypopulation consists of patients likely to receive the differentialpopulation consists of patients whose differentialpopulation consists of patients whose differentialare designed to resolvestudypatients likely to receive the differentialpatients whose differentialpatients whose differentialmore tainties; if the targetdiagnosis diagnosistest in clinical diagnosisincludes other diseasesdisease based on availableuncertainties; if the targetdiagnosis diagnosisbesides the but in whom target disease, includesinformation, and patientsdisease, the sensitivity willbut also includeslikely to be already alreadyalready availableon already availableon already availableof having the target disease the specificity of the target disease needs to be differentiatedon already availableof having the target disease the specificity of the test will be biased upwards; this bias is reduced when consecutive	Spectrum of	Study	Study	Study	Diagnostic tests
enrolled in the studyconsists of patients likely to receive the test in clinical practice; the differential diagnosisconsists of patients whose differential diagnosisresolve diagnosis the target diseasesdiagnosis practice; the differential diagnosisconsists of patients whose differential diseasesconsists of patients whose clearly have the target disease based on available includes the target disease, but alsoresolve diagnosis target disease, but in whom target disease, but alsoconsists of patients whose clearly have target disease, but in whom target disease, on already available includesresolve diagnosic uncertainties; if the target diseases, the sensitivity will be biased upwards; if the negative test subjects are clearly have disease the sensitivity will be biased upwards; if the negative test subjects are clearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive	patients	population	population	population	are designed to
studypatients likely to receive the test in clinical practice; the differentialpatients whose differential diagnosispatients who clearly have the target disease based on available information, advanced disease, the sensitivity will be biased target disease, but alsopatients whose differential diseases, but in whom target disease, already already on already similarly, from which the target disease needs to be differentiatedpatients whose clearly have the target information, and patients who are clearly healthy and have a very low likelihood of having the target disease the specificity of the test will be biased upwards; this bias is reduced when consecutive	enrolled in the	consists of	consists of	consists of	resolve
to receive the test in clinical practice; the differentialdifferential diagnosisclearly have the target disease based on availableuncertainties; if the positive test subjects have advanceddiagnosis diagnosisbesides the target disease, but alsoincludes the the diagnosis is likely to be alreadyinformation, and patients who are clearly healthy and have a very low likelihood of having the target disease needs to be differentiateduncertainties; if the positive test subjects have advancedto receive the test in cludesbesides the target disease, alreadyinformation, and patients who are clearly healthy and have a very low likelihood of having the target disease needs to be differentiateduncertainties; if the target advanced disease based on available information, available	study	patients likely	patients whose	patients who	diagnostic
test in clinical practice; the differentialdiagnosis includes other diseasesthe target disease basedthe positive test subjects have advanceddiagnosisbesides the besides theinformation, and patientsdisease, the sensitivity willreasonably includes the target disease, but alsotarget disease, but alsoand patients healthy and likely to besensitivity will be biasedbut alsolikely to be alreadyhave a very low likelihoodnegative test subjects are clearly healthy, the specificity of the test will be biased upwards; the specificity of the test will be biased upwards; this bias is reduced when consecutive		to receive the	differential	clearly have	uncertainties; if
practice; the differentialincludes other diseasesdisease based on availablesubjects have advanceddiagnosisbesides the information,information, disease, the sensitivity willdisease, the sensitivity willreasonably includes the target disease, but alsotarget disease, the diagnosis is likely to be alreadyand patients healthy and low likelihoodsensitivity will be biasedbut alsolikely to be alreadyhave a very low likelihoodupwards; if the negative testmay present similarly, from target disease needs to be differentiatedavailable informationof having the target diseaseclearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive		test in clinical	diagnosis	the target	the positive test
differential diagnosisdiseases besides the target disease, but in whomon available information, and patientsadvanced disease, the sensitivity will be biased upwards; if the negative testincludes target disease, but alsothe diagnosis is likely to be alreadyhealthy and have a very low likelihood of having the target diseaseupwards; if the negative testincludes diseases which may present similarly, from target diseaseapparent based on already available informationof having the target diseaseclearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive		practice; the	includes other	disease based	subjects have
diagnosisbesides the target disease, but in whominformation, and patientsdisease, the sensitivity willreasonablytarget disease, but in whomupwards; if the negative testtarget disease, but alsothe diagnosis is likely to be alreadyhealthy and have a veryupwards; if the negative testdiseases which may present similarly, from which the target diseaseapparent based on already availableof having the target diseaseclearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive		differential	diseases	on available	advanced
reasonably includes the target disease, but in whomand patients who are clearly healthy and have a very low likelihoodsensitivity will be biased upwards; if the negative test subjects are clearly healthy, traget disease subjects are clearly healthy, traget disease may present similarly, from which the target disease needs to be differentiatedtarget disease, but in whom target disease, alreadyand patients who are clearly healthy and have a very low likelihood of having the target diseasesensitivity will be biased upwards; if the negative test subjects are clearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive		diagnosis	besides the	information,	disease, the
includes the target disease, but alsobut in whom the diagnosis is likely to be alreadywho are clearly healthy and have a verybe biased upwards; if the negative testincludes diseases which may present similarly, from target diseasealready on already availablelow likelihood of having the target diseasesubjects are clearly healthy, the specificity of the test will bewhich the target disease needs to be differentiatedinformation target diseasesis reduced when consecutive		reasonably	target disease,	and patients	sensitivity will
target disease, but alsothe diagnosis is likely to be alreadyhealthy and have a veryupwards; if the negative testincludesalreadylow likelihoodsubjects are clearly healthy, the specificity of the test will bediseases which may presentapparent based on alreadyof having the target diseaseclearly healthy, the specificity of the test will bewhich the target disease needs to be differentiatedinformationinformation consecutivethe specificity of the test will be		includes the	but in whom	who are clearly	be biased
but alsolikely to behave a verynegative testincludesalreadylow likelihoodsubjects arediseases whichapparent basedof having theclearly healthy,may presenton alreadytarget diseasethe specificity ofsimilarly, fromavailableinformationbiased upwards;target diseaseneeds to beinformationthis bias isneeds to beinferentiatedclearly healthy,		target disease,	the diagnosis is	healthy and	upwards; if the
includesalreadylow likelihoodsubjects arediseases whichapparent basedof having theclearly healthy,may presenton alreadytarget diseasethe specificity ofsimilarly, fromavailableinformationbiased upwards;target diseaseneeds to beinformationthis bias isneeds to beinferentiatedclearly healthy,		but also	likely to be	have a very	negative test
diseases which may present similarly, from target diseaseapparent based on already available information target diseaseof having the target diseaseclearly healthy, the specificity of the test will be biased upwards; this bias is reduced when consecutive		includes	already	low likelihood	subjects are
may present similarly, from which the target diseaseon already available informationtarget disease biased upwards; this bias is reduced when consecutive		diseases which	apparent based	of having the	clearly healthy,
similarly, from available information target disease needs to be differentiated when consecutive		may present	on already	target disease	the specificity of
which the target disease needs to be differentiatedinformation informationbiased upwards; this bias is reduced when consecutive		similarly, from	available		the test will be
target disease needs to be differentiated this bias is reduced when consecutive		which the	information		biased upwards;
needs to be differentiated consecutive		target disease			this bias is
differentiated consecutive		needs to be			reduced when
		differentiated			consecutive
patients who					patients who
would be					would be
candidates for					candidates for
the test are			· *		the test are
enrolled, and					enrolled, and
increased when a					increased when a
case-control					case-control
design is used					design is used
Evaluation of The interpreter The test results The test results If the test is	Evaluation of	The interpreter	The test results	The test results	If the test is
test results is of the test are interpreted are interpreted interpreted under	test results is	of the test	are interpreted	are interpreted	interpreted under
done under results has the with only part under highly artificial	done under	results has the	with only part	under	highly artificial
circumstances same kind of of the circumstances circumstances,	circumstances	same kind of	of the	circumstances	circumstances,
which closely information information which would the study may	which closely	information	information	which would	the study may
resemble the that would be which would rarely be seen inaccurately	resemble the	that would be	which would	rarely be seen	inaccurately
circumstances available to a be available to in practice describe how the	circumstances	available to a	be available to	in practice	describe how the
under which clinician using a clinician (interpreter test will perform	under which	clinician using	a clinician	(interpreter	test will perform
iney would be the test in daily using the test in has never seen in the real world;	iney would be	the test in daily	using the test in	the net int	in the real world;
evaluated in practice (has daily practice the patient) this is NOT to be	evaluated in	practice (has	daily practice	the patient)	this is NOT to be
everyday seen the confused with	everyday	seen the			confused with
practice patient, taken a naving the test	practice	patient, taken a			naving the test
Instory, done a results		nistory, done a			interpreted

Criterion	Green	Yellow	Red	Comments
	examination,			blinded to the
	seen the routine			results of the
	laboratory			gold standard
	tests, etc)			(see below)
Description of	Sufficient	Partial	Insufficient	It is important to
the test	information	information is	information	have enough
	about the test	given about	about the	description of
	equipment and	how the test is	execution of	test protocols to
	execution is	executed	the test is given	allow results to
	provided to			be compared
	permit			between studies.
	replication of			and to decide
	the test			whether the test
	the test			technique being
				studied is the
			CY	same as the test
				being considered
				for a guideline
				recommendation.
		, A	Y	it is acceptable to
			2	have technical
				details furnished
				in a separate
				document
		A Y		provided that the
				reference section
				noint the reader
		x P		to the source of
				to the source of
Departing of		Desitive		The frequency
Reporting of	All test results	Positive,	Only positive	The frequency
results	for all patients	negative, and	and negative	with which the
1	are reported,		results are	test does not
	including the	results are	reported and	return a definite
	number of	reported, but	used for	result is required
	positive,	the number of	calculation of	for estimation of
	negative,	uninterpretable	sensitivity and	its performance
	indeterminate,	results is not	specificity	in practice
	and	reported		
	uninterpretable			
	results			
Reference	There is a	There is a	There is no	The readily
standard (gold	recognized	recognized	gold standard	applicable gold
standard)	gold standard	gold standard	for the disease	standard test may
	which provides	for the disease,		be the exception
	a definitive test	but it is not		rather than the

Criterion	Green	Yellow	Red	Comments
	of the presence	practical to		rule; if it is an
	of the disease,	apply to all		invasive or
	and which can	patients		expensive test,
	be applied to	undergoing the		application to all
	all patients	diagnostic test		patients in a
	undergoing the	being evaluated		study may be
	diagnostic test	C		impractical or
	being evaluated			unethical. It is
	e			acceptable to
				apply the gold
				standard to those
				who test
				positive, and to
				follow up those
				who test negative
			C	for subsequent
				developments.
				when the gold
				standard test is
		A	× ×	not practical
Gold standard	All patients	Some patients	The gold	If the gold
applied to all	who had the	who had the	standard was	standard test is
patients who	test being	test being	applied in a	invasive or
underwent the	evaluated, or a	evaluated did	manner which	expensive, it
test being	random sample	not have the	is influenced	need not be
evaluated, or to	of such	gold standard	by factors	applied to those
a random	patients, also	test, but there is	which may be	with a negative
sample of	received the	no indication	associated with	result on the test
patients	test for the gold	that the	the condition	being evaluated;
1	standard	performance of	being	follow-up and
	10	the gold	diagnosed	continued
•		standard test	0	observation may
1		was influenced		be substituted
		by factors		
		which may		
		predict its		
		result		
Withdrawals	There is	Some	The patients	It is necessary to
	sufficient	ambiguity	who	know how many
	information to	exists	participated at	patients who
	determine	concerning	the various	received the gold
	whether all	what happened	stages of the	standard also
	patients who	to all of the	study are not	received the test
	entered the	patients who	reported	under
	study are	entered the		consideration,

Criterion	Green	Yellow	Red	Comments
	accounted for,	study; some		and vice versa; if
	including how	patients are not		many patients
	many patients	accounted for		withdrew after
	participated in	at the end of		participating in
	each phase of	the study		only one phase
	the study (flow	-		of the study, it is
	diagrams with			necessary to
	numbers of			describe and
	patients at each			account for them
	stage of the			
	study are ideal)			
Test thresholds	Clearly defined	Same criteria,	Cutoff points	This applies only
	cutoff points	but with area	are unclear, or	when the test
	are given	under ROC	area under	returns a
	which	curve of 0.7 to	ROC curve is	continuous
	distinguish the	0.8	less than 0.7	result, and the
	difference			tradeoff of
	between a			sensitivity and
	positive and a			specificity can be
	negative test	A	y	expected to be
	result: when	3		displayed
	multiple cutoff			graphically
	points are			8 m
	possible, the			
	sensitivity and			
	specificity are			
	reported for			
	each, and a			
	Receiver			
	Operating			
	Characteristic			
•	(ROC) curve is			
	given, with			
	area under the			
	curve of 0.8 or			
\Box	more			
Blinding of test	It is clearly	There is	Blinding of the	Large biases are
interpreters	stated that the	ambiguity	interpreters is	introduced when
	interpreters of	about whether	not clear, or	test
	the test under	the interpreters	was not done;	interpretation is
	evaluation were	of one test were	sequence of	influence by
	not aware of	aware of the	tests cannot be	knowledge of the
	the results of	results of the	determined	results of other
	the gold	other test; it is		tests; if tests are
	standard test,	clear whether		strictly

Criterion	Green	Yellow	Red	Comments
	and that the	the gold		numerical
	interpreters of	standard or the		readings of
	the gold	test under		instruments, this
	standard test	evaluation was		criterion is less
	were unaware	applied first		important
	of the results of	11		1
	the test under			
	evaluation; it is			
	clear which test			
	was applied			
	first			
Inter-rater	The	The	The	Kappa may be
reliability	interpretation	interpretation	interpretation	biased if the
	of the test is	of the test is	of the test is	prevalence of the
	done by two or	done by two or	done by two or	disease in the
	more assessors	more assessors	more assessors	study population
	working	working	working	is close to zero
	independently.	independently.	independently.	or is close to
	and there is a	and there is a	and there is a	100%: this
	good	fair agreement	slight or poor	should not
	agreement	between them	agreement	happen if there is
	between them	(Kappa is 0.4	between them	an appropriate
	(Kappa is 0.6	to 0.6	(Kappa is less	spectrum of
	or greater)		(120) than $(0, 4)$, or	patients in the
	or grower)		there was no	study sample
			report of inter-	
			rater reliability	
Test settings	The test has	The test has	The test has	Test
	been applied in	been applied in	been applied in	performance
	a wide variety	only a few	only one	may vary with
	of settings	settings	setting	different settings,
	(primary care,			and a wide
	specialty care,			variety of
	tertiary care,			settings is
	high and low			necessary for
	prevalence of			assessing its
	the disease)			usefulness in
				clinical practice
Test	Point estimates	Point estimates	Test	Sensitivity and
performance	are given for	are given for	performance is	specificity are
measures are	sensitivity and	sensitivity and	not clear from	the core
presented with	for specificity,	for specificity,	the data in the	performance
measures of	together with	with	study	measures;
uncertainty	95%	confidence		predictive values
(e.g., 95%	confidence	intervals, but		depend on

Criterion	Green	Yellow	Red	Comments
confidence	intervals for	cutoff points		population
intervals)	both measures,	are either		characteristics
	and are	lacking or are		and are
	presented for	unclear		optionally
	two or more			reported
	well-described			1
	cutoff points			
Likelihood	LR+ is 10 or	LR+ is between	LR+ is less	Likelihood ratios
ratios (LR+)	greater	5 and 10	than 5	are measures of
for a positive				how much more
test (true				probable a
positive				positive test is in
rate/false				a person with a
positive rate)			A 1	disease than in a
are likely to				person without
produce useful			C	the disease, and
shifts in the				are a useful
estimate of the			Sh	summarv
probability of				measure of the
the presence of			Y	impact of the test
the disease.		3		result on the
with the				odds that a
potential to				patient has the
alter clinical				disease
decisions				aiseuse
Likelihood	LR- is less than	LR- is between	LR- is greater	As with LR for
ratios (LR-) for	0.1	0.1 and 0.2	than 0.2	positive tests, a
a negative test		y • · - • · - • • · -		low LR- can
(false negative				alter clinical
rate/true				decisions
negative rate)	10			regarding
are likely to				whether to
produce useful				consider a
shifts in the				diagnosis
estimate of the				improhable
probability of				enough to look
the presence of				to other
the disease				diagnoses of the
the disease				clinical condition
Diagnostic	DOR of greater	DOR less than	DOR less than	DOR unlike
odde ratio	than 20	201 1855 utall	20	positive and
(DOP) con bo	nrafarably aven	20	20	positive allu
(DOK) call De	greater			negative
$(I \mathbf{P} + /I \mathbf{P}) + \mathbf{h}_{\mathbf{P}}$	gicaldi			is relatively
(LK+/LK-) UIC				is icialively
пкеппооа				independent of

Criterion	Green	Yellow	Red	Comments
ratios positive				prevalence of the
and negative				disease; it is
				sensitive to the
				spectrum of
				patients enrolled
				in the study
Characteristics	Test	There is some	Information	Test
of test	interpreters are	information	about the test	interpretation
interpreters	well	about the test	interpreters is	may involve
	characterized in	interpreters, but	vague or	subjective
	terms of	they are not	missing	judgment, and a
	specialty	fully described		learning curve
	training,	in their		may be involved
	experience, and	expertise and	~1	in reading or
	expertise with	training		executing the test
	executing and			
	reading the test			
Benefits of	Test results	Test results	Test results	More than one
receiving the	clearly change	successfully	make no	type of study
test	patient	diagnose the	difference in	may be required
	management in	target disease,	management or	to make this
	ways that lead	but there is	outcome	determination; a
	to fewer	equivocal		randomized
	complications,	benefit from		clinical trial is
	faster recovery,	the changes in		the most robust
	and better final	that result from		design to
	to the making	making the		compare outcomes of
	of diagnosas	diagnosis		patients who do
	with different	ulagilosis		and do not have
	treatment			the test
•	strategies			uie test
Incremental	The test is	The test has	The test adds	Clinical
value of test	clearly shown	hetter	nothing to what	investigations
value of test	to have an	diagnostic	is already	are expected to
	advantage over	performance	available for	result in useful
	simpler or	than simpler or	diagnostic	changes in
	cheaper tests.	cheaper tests.	investigations	management, not
	in having	but there is no		simply additional
	higher	evidence that		information
	likelihood	doing it leads		
	ratios, or in	to better		
	leading to	outcomes		
	better outcomes			
	for patients			

Criterion	Green	Yellow	Red	Comments
	who get the test			
Purpose of test	There is a clear	The setting and	The setting and	Sensitivity is
	description of	purpose are not	purpose are not	crucial for
	the setting in	stated, but may	apparent	screening tests
	which the test	be inferred by		but not for
	is to be used,	the reader		confirmatory
	and the			tests; specificity
	purposes to			is crucial for
	which it is			confirmatory but
	intended			not for screening
				tests

Reference for likelihood ratios and diagnostic odds ratios:

Fisher JE, Bachmann LM, Haesche R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med 2003;29:1043 -1051