Q J Med 1998; **91**:247–258

Commentary

QJM

Likelihood ratios: getting diagnostic testing into perspective

A. HALKIN, J. REICHMAN, M. SCHWABER, O. PALTIEL[1] and M. BREZIS

From the Department of Medicine, Hadassah University Hospital Mount Scopus and [1]Department of Social Medicine, Hadassah University Hospital Ein Kerem, Jerusalem, Israel

Received 16 September 1997

Summary

In modern medicine, sophisticated laboratory tests and imaging studies are often emphasized at the expense of history and physical examination, rather than complementing clinical assessment. Ancillary testing often fails to advance the diagnostic process, and increases patient risk and the expense of medical care. The relative value of clinical evaluation and technological methods is rarely considered, and the power of the clinical evaluation is therefore underestimated. The likelihood ratio (LR) is a semi-quantitative measure of the performance of diagnostic tests which indicates how much a diagnostic procedure modifies the probability of disease, and is calculated from the sensitivity and specificity of the test (or directly from the change in probability associated with the test result). We review the performance of frequently-used tests by their LRs, and compare them to the power of clinical assessment, with clinical cases to illustrate the application of LRs in the diagnostic process. The discriminative power of clinical assessment and ancillary tests is often similar, and the combination of the two greatly increases accuracy in the diagnostic process. Clinical assessment is indeed frequently more informative than current technical modalities. LRs assist in putting the value of testing in proper perspective. Practice in evaluating pre-test probabilities of disease and in the application of LRs should be enhanced in medical training.

Introduction

A frequent belief among practitioners in clinical medicine is that laboratory tests and imaging techniques are more definitive than medical history and physical examination. This may result from a conviction that tests are objective and reliable, from the permanence of test results, and from underestimating the information that can be obtained by history and physical examination.[1,2] Also detracting from physicians' clinical skills and their appreciation of clinical evaluation is the decline in bedside teaching in major medical centres.[80] Physicians thus tend to regard clinical assessment as inaccurate or insufficient, and rely on the results of sophisticated tests. For example, an echocardiogram is frequently ordered to diagnose cardiac failure, or a duplex to rule out deep venous thrombosis, with minimal attention paid to clinical findings. However, testing exposes patients to risks, compromises their well-being, increases costs, and may actually impede diagnosis when false-positive results occur.[72] In elderly subjects, excessive testing increases the chance of discovering incidental findings, of little or no relevance to patient suffering. The time and energy invested by physicians in obtaining tests often comes at the expense of communication with patients and their families. It is therefore not surprising that in major teaching hospitals, despite increasing use of new techniques (such as ultrasonography, computed tomography and endoscopy), rates of misdiagnosis among hospitalized patients have remained essen-

Address correspondence to Professor M. Brezis, Hadassah University Hospital Mount Scopus, Jerusalem, POB 24035, Israel

Table 1 LRs for frequently used diagnostic tests

Condition	Test	Sensitivity (%)	Specificity (%)	LR^+	LR^-
Deep vein thrombosis[7]	Ultrasonography (B mode or duplex):				
	symptomatic	97	97	32	0.03
	asymptomatic	59	98	29	0.4
Pulmonary embolism[8]	Ventilation/perfusion scan:				
	high probability*				18.3
	intermediate probability**				1.2
	low probability***				0.36
	normal/near normal				0.1
	Spiral CT†	93	88	8	0.08
Myocardial infarction[13,14]	ECG (new Q, ↑↓ ST, LBBB)	81	69	2.61	0.27
	CPK (single)	38	80	1.9	0.77
	CPK (serial)	98	67	2.96	0.03
	CPK MB (single > 4 hrs)	34	88	2.83	0.75
	CPK MB (serial)	100	98	50	0
Coronary artery disease[15,16]					
any vessel disease	Exercise ECG	61	80	3	0.49
	Thallium—visual assessment	84	87	6.46	0.18
	Thallium—quantitative analysis	90	75	4	0.13
	Thallium—tomography	96	83	5.29	0.05
detection of multivessel disease after MI	Exercise ECG	62	70	2.1	0.54
	Thallium scintigraphy	70	88	5.8	0.34
Renovascular hypertension[17–23]	Rapid sequence IVP	74	86	5	0.3
	Peripheral renin	65	84	4	0.4
	Captopril test	74	89	6.7	0.3
	Captopril renogram	92	95	18.4	0.08
	Ultrasonography (duplex)	98	98	49	0.02
	Magnetic resonance angiography‡	94	95	18.8	0.06
Phaeochromocytoma[24]	Urinary catecholamine excretion	82	95	16.4	0.19
	Abdominal CT scan	92	80	4.6	0.1
Renal colic[25,26]	Ultrasound for hydronephrosis	85	95	17	0.16

Table 1 (*Continued*)

Condition	*Test*	*Sensitivity (%)*	*Specificity (%)*	*LR*$^+$	LR$^-$
Anaemia					
Iron-deficiency[27]	MCV > 95				0.11
	91 < MCV ⩽ 95				0.34
	85 < MCV ⩽ 91				0.64
	74 < MCV ⩽ 85				1.35
	MCV ⩽ 74				8.82
	TS > 0.21				0.28
	0.08 < TS ⩽ 0.21				0.57
	0.05 < TS ⩽ 0.08				1.43
	TS ⩽ 0.05				16.5
	Ferritin > 100 microgram/L				0.13
	45 < Ferritin ⩽ 100				0.46
	18 < Ferritin ⩽ 45				3.12
	Ferritin ⩽ 18				41.5
Pernicious anaemia[28,29]	MCV > 105	70	93	10	0.32
	Serum cobalamin levels	90	60	2.2	0.17
Infectious diseases					
Infective endocarditis[30,31]	Blood cultures				
	1 set	80	97	27	0.21
	3 sets	99	97	33	0.01
	Transthoracic echocardiography				
	native valves	50	78	2.3	0.64
	prosthetic valves	17	94	2.8	0.88
	Transoesophageal echocardiography				
	native valves	100	89	9.1	0
	prosthetic valves	83	95	16.6	0.18
Pneumococcal pneumonia[32,33]	Sputum culture	50	65	1.42	0.77
	Sputum Gram stain	60	90	6	0.44
Meningitis[34–37]	CSF Gram stain	60–90	100	inf.	0.1–0.4
	CSF culture	70–85	100	inf.	0.15–0.3
Antibiotic-induced colitis[38,79]	*Clostridium difficile* toxin B	22–97.5	98	11–49	0–0.06
	Sigmoidoscopy	70	98	35	0.31
Colorectal adenocarcinoma[39]	Hemoccult II (without rehydration)	37	98	18.5	0.64
	Hemoccult II Sensa (with rehydration)	79	87	6.0	0.24
	HemeSelect (immunochemical test)	69	94	11.5	0.33

LR$^+$, LR$^-$, likelihood ratios for positive and negative results, respectively; CPK, creatine phosphokinase; CSF, cerebrospinal fluid; CT, computed tomography; ECG, electrocardiogram; inf., infinity; IVP, intravenous pyelography; LBBB, left bundle branch block; LR, likelihood ratio; MCV, mean corpuscular volume; TS, transferrin saturation. * Two or more near-segmental or larger defects without matching ventilation defects. ** Not fitting into other categories. *** Subsegmental perfusion defects or single moderate perfusion defects or matched ventilation-perfusion defects. † Mean of several studies on more than 300 patients[9–12]. ‡ Mean of several studies on more than 200 patients[19–23].

Table 2 Examples of LRs for clinical findings

	Sensitivity (%)	Specificity (%)	LR^+	LR^-
Clinical diagnosis of deep-vein thrombosis[41]			17	0.15
Clinical diagnosis of sinusitis[44]				
Maxillary toothache	18	93	2.5	0.9
No improvement with decongestants	41	80	2.1	0.7
Purulent discharge	51	76	2.1	0.7
Nasal speech	45	73	1.7	0.8
Abnormal transillumination	73	54	1.6	0.5
Cough	70	44	1.3	0.7
Summation of the above*			39	0.12
Overall clinical impression			4.7	0.4
Clinical diagnosis of Alzheimer's dementia[45]	98	97	33	0.02
Prediction of preserved left ventricular function after myocardial infarction[48]	58.3	97.7	25.3	0.43
Abdominal jugular test for cardiac failure[49]	86	98	43	0.14
Diagnosis of appendicitis[50]	63	82	3.5	0.45
Clinical diagnosis of joint injury or disease				
Clinical diagnosis of tears of the rotator cuff[51]	91	75	3.6	0.12
Accuracy of knee examination for single diagnosis vs. arthroscopy[52]	56	87	4.3	0.51
Accuracy of clinical examination for posterior cruciate ligament tears[53]	90	99	90	0.1
Lyme arthritis[54]			12	
Clinical signs of pneumonia in infants under 2 months[55]	91	97	30	0.02

LR^+, LR^-, likelihood ratios for presence or absence of finding(s). *LRs were calculated by multiplication of LR of all these clinical features, assuming presence of all of them and their independence from each other.

tially unchanged throughout the last four decades.[73,74] New modalities provided conclusive information in 30% of cases and were misleading in 10%, while the history and physical examination established the diagnosis in more than 60%, and were misleading in less than 2% of cases.[74]

As emphasis shifted to sophisticated testing, clinicians have neglected objective assessment of bedside evaluation as a diagnostic tool: in contrast to the abundant literature rating the performance of many tests, relatively few reports probe the performance of plain clinical evaluation in a quantitative manner. A convenient and increasingly adopted measure of the performance of diagnostic methods is the likelihood ratio (LR), which expresses the magnitude by which the probability of a diagnosis in a given patient is modified by the result of a test.[3–6] The LR for a test result is the ratio between the chance of observing that result in patients with the disease in question, and the chance of that result in subjects without the disease. This value is readily calculated from the sensitivity and specificity of the test (see Appendix I). The product of the LR and pre-test odds determines post-test odds. LRs range from zero to infinity. Therefore, a value of 1 means that the test provides no additional information; ratios above or below 1 respectively increase or decrease the likelihood of disease. Simple calculations convert odds to probability and vice versa (see Appendix II).[3,4,40] By using a normogram, these calculations may be avoided,[11] as illustrated in the examples and figures that follow. When several independent tests are performed, their joined LR is equal to the product of the LRs of the individual tests. Misuse and misconceptions regarding LRs have been described.[5] Among these is the erroneous assumption that tests can be compared linearly by their LRs (the power of a test with a LR of 100 is not ten times greater than that of a test with a LR of 10). LRs may be inaccurate if the studies defining test performance are of poor quality or involve verification bias. Differences between the population in which a test was characterized and the population in which it is applied may render the LR inapplicable.[6]

Rapid technological advances in medicine have rendered certain diagnostic methods for common diseases obsolete, and new tests have been introduced. In the current article we present updated LRs for tests frequently employed in daily hospital practice based on recent literature published since this topic was last reviewed (Table 1).[3,4] From published data, we also estimated LRs for findings obtained by interview and physical examination to assess the power of clinical evaluation as a diagnostic test. Since better information is certainly needed, we suggest that LRs be viewed as semi-quantitative values that can be used to compare the power of ancillary testing with that of clinical information,

mainly as a teaching exercise. This serves to illustrate the power of a refined clinical assessment, and demonstrates that for diagnosing commonly encountered medical conditions, its usefulness often does not fall below that of sophisticated tests. Using true cases from our department we demonstrate the relative significance of findings obtained by clinical evaluation and ancillary tests and how integrating these findings influences the diagnostic process.

Integrating clinical assessment and test results

The following example illustrates application of LRs in the diagnosis of deep venous thrombosis (DVT).

Example 1: A patient recovering from a knee operation complains of lower-limb swelling and pain. The left leg is swollen with tenderness along the medial thigh. The probability of DVT in this setting is approximately 85%.[41] Since the LR for a positive duplex study in this setting is 32 (Table 1), the post-test probability would be over 99% were the test positive, as illustrated in Figure 1**a** (for calculations, see Appendix II).

While the clinical diagnosis of DVT is considered unreliable, a recent study[41] showed that refined clinical assessment producing specific findings (such as a history of recent immobilization, cancer, paralysis, calf swelling and localized tenderness along the deep venous system) increases the chance of deep venous thrombosis from 25% to 85% (in the presence of three or more major points and no alternative diagnosis), equivalent to a LR of 17 (see Appendix II). Using a 25% prevalence of DVT in out-patients suspected to have this condition[41] as pre-test probability, a positive duplex would have yielded a post-test probability of 91.4%. In this case, a refined clinical assessment can be almost as powerful as a duplex study (Figure 1**b**) and, as described below, may sometimes be more accurate.

Example 1 (continued): The duplex is negative. Since the LR of a negative duplex study in a symptomatic patient is 0.03 (Table 1), the post-test probability is 14.5%, as illustrated in Figure 1**a** (for exact calculations, see Appendix II).

While following out-patients with suspected deep venous thrombosis and serial negative duplex studies appears safe[42] this patient obviously can not be treated similarly. Since the pre-test probability deter-

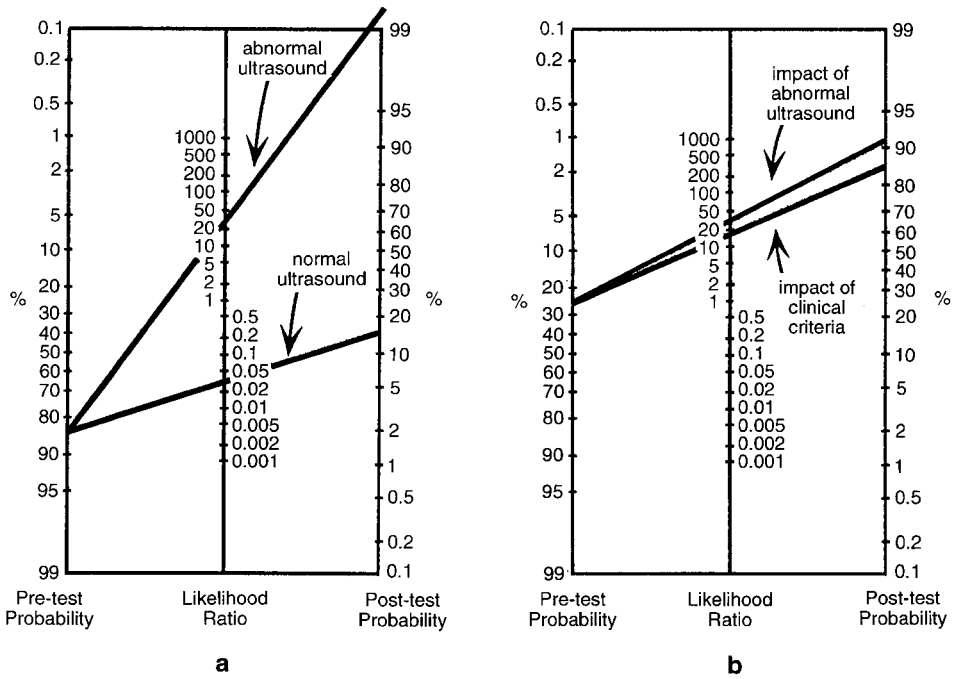

Figure 1. Magnitude of impact from a refined clinical assessment vs. that from an ultrasound in the diagnosis of deep venous thrombosis. **a** The probability that the patient described in example 1 has deep venous thrombosis is estimated as approximately 85%.[41] Given LRs of 32 and 0.04 for positive and negative results in a symptomatic patient (see Table 1), the lines drawn show post-test probabilities over 99% if the ultrasound is abnormal and 14.5% if the ultrasound is reported normal (for exact calculations, see Appendix II). **b** As illustrated in example 1 in the text, refined clinical assessment (noting recent immobilization, calf swelling and localized tenderness along the deep venous system and no alternative diagnosis) increases the chance of deep venous thrombosis from 25% to 85%.[41] An abnormal ultrasound without refined clinical assessment, using a 25% prevalence for such a population, would yield a post-test probability of 91.4%. The close proximity of the lines suggests that in this case, a refined clinical assessment is almost as powerful as an ultrasound.

mined clinically was so high, the negative duplex results in a post-test probability of 14.5%, which can not be ignored. This calls for further testing such as contrast venography.

Example 1 (continued): An MRI ordered to image the soft tissues discloses deep venous thrombosis, which is confirmed by a repeated ultrasound.

The bedside diagnosis of deep venous thrombosis was made with high probability in this patient, by clinical features and very low probability of alternative diagnoses, such as a ruptured Baker's cyst or cellulitis. The reliability of compression ultrasonography heavily depends on experience, which may result in sensitivities ranging from 0 (!) to 100%.[43] When using literature data on test performance, caution should be applied for operator-dependent factors.

Impact of pre-test probability

Clinical evaluation is also operator-dependent, and a frequently-asked question is how to determine the pre-test probability. In many diseases, quantitative information about the sensitivity and specificity of symptoms and signs is lacking and the pre-test probability is estimated by combining epidemiological data (about population prevalence) with clinical impressions based on prior education (similarity of the patient's picture to descriptions of diseases in textbooks), personal experience or pattern recognition (such as for the diagnosis of erythema nodosum). Medical students and interns are trained to recognize syndromes in a qualitative rather than in quantitative fashion. In our training program, residents asked to assign a pre-test probability of coronary ischaemia to the same patient presenting with chest pain, would often assign values ranging from 25 to 75%. If the correct value is 75%, a clinician using the value of 25%, even helped by a positive exercise test, would only reach a post-test probability of 50%, lagging behind the best clinical assessment, as illustrated in Figure 2**a**. The following examples further illustrate the importance of the correct determination of pre-test probability.

Example 2: A 45-year-old male tourist was admitted because of chest pain. He had no personal or family history of coronary disease, his past history being significant for chronic renal failure and renal transplantation. The pain was retrosternal and radiating to the back, having started while watching TV after a heavy supper and persisted for several hours. Patient's vital signs, cardiac examination, ECG and CPKs were normal. The physician in charge considered unstable angina, at an estimated probability of 75%. Because of ongoing intermittent pain despite several days on heparin, aspirin, nitrates, propranolol and bed rest, a cardiologist proposed a coronary angiogram. Beforehand, the clinical symptoms were reassessed: the pain was recurring after each meal, especially after lying down; the patient denied any pain at effort and could recall somewhat similar symptoms after the kidney transplantation, while on high doses of glucocorticoids, at which occasion an endoscopy had revealed reflux esophagitis resolved with omeprazole. A diagnosis of recurrent reflux oesophagitis was entertained, and the patient was started on omeprazole, with complete disappearance of his symptoms.

In this case, from an original estimate of 75%, a refined clinical assessment reassigned a probability of less than 5% for coronary disease as the cause of chest pain. Figure 2**b** illustrates the magnitude of the impact of refined clinical assessment, for the diagnosis of coronary disease, in comparison with an exercise test (which might have been ordered to evaluate the source of the chest pain before an angiogram), using positive and negative LRs of 3 and 0.49, respectively (see Table 1). The impact of refining clinical assessment was greater than the information from a stress test.

Example 3: A 40-year-old female non-smoker, using contraceptive pills, was admitted because of cough and shortness of breath. The physical examination, a chest X-ray and a duplex study of the legs, were all unremarkable. The resident suspected pulmonary embolism at a 50% pre-test probability and the patient was treated with heparin. A ventilation perfusion scan reported intermediate probability (LR 1.2, see Table 1), leaving the post-test probability of pulmonary embolism practically unchanged at 54%. While a pulmonary angiogram was being considered, a more detailed history revealed that similar symptoms had occurred last spring and that hay fever had been present for many years. A peak expiratory flow rate was low and increased after salbutamol. The patient's symptoms cleared on bronchodilators and she was discharged with the diagnosis of bronchial asthma.

The impact of refining clinical assessment was more substantial than the information from a ventilation perfusion scan. With the recognition of the alternative diagnosis of bronchial asthma, the probability of pulmonary embolism dropped to less than 1%, even more efficiently than if the scan had been normal (LR 0.1, see Table 1), which would have yielded a post-test probability of 9% (had the pre-test probability of 50% been retained). Clinical evaluation is not necessarily more precise than ancillary tests, but its accuracy is of marked consequence to the correct interpretation of test results. Practice in the evaluation of pre-test probability should be exercised in medical training, for instance, by comparing the estimates by interns or residents

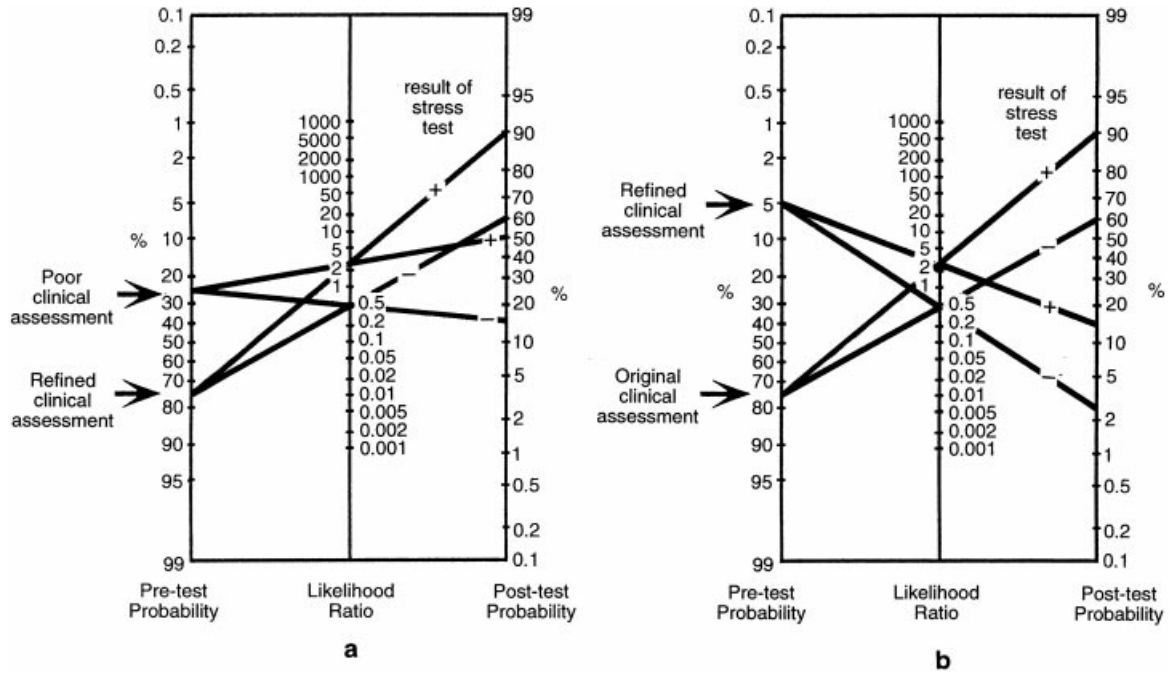

Figure 2. Magnitude of impact from a refined clinical assessment in the diagnosis of coronary disease. **a** Residents were asked to assign a pre-test probability of coronary ischaemia to the same patient presenting with chest pain, assign values ranging from 25 to 75%. If the correct value is 75%, a clinician using the value of 25%, even helped by a positive exercise test, would reach a post-test probability of only 50%, lagging behind the best clinical assessment. **b** In Example 2, refinement of the clinical assessment shifted the pre-test probability from 75 to 5%, while the impact of any (positive or negative) result of a stress test would have left the post-test probability in the range of 60–90% (for the high pre-test suspicion of 75%) or 2–15% (for the low pre-test probability of 5%). Refining clinical assessment can provide more substantial information than a stress test.

to those made by experienced clinicians, or by discussing discrepancies using published data on clinical performance.

Assessment of pre-test probability

In some diseases, the reliability of clinical data has been quantified, or specific criteria have been described to help the formulation of a pre-test probability. For instance, in the clinical evaluation of a patient with suspected sinusitis, the diagnostic value of symptoms and signs, such as maxillary toothache, purulent discharge or abnormal transillumination has been reported.[44] As shown in Table 2, summation of these clinical items (positive and negative LRs of 39 and 0.12) may lead to a predictive value greater than 90%,[44] better than overall clinical impression or than a CT scan, where incidental findings consistent with sinusitis are present in as many as 30–40% of asymptomatic patients.[56,57] Other examples listed in Tables 2 and 3 include the clinical diagnosis of diseases such as dementia, cardiac failure, renovascular hypertension or, as shown in Table 4, the prediction of risk after a myocardial infarction. For instance, the diagnosis of ankylosing spondylitis based on summation of specific characteristics of the back pain, as illustrated in Table 3, may be as informative as HLA-testing, if not more. The clinical evaluation referred to in Tables 2–4 is usually completed within minutes, at low cost and without risk to the patient. Moreover, enhanced communication between patient and physician improves understanding and trust. Additional studies are needed to expand our knowledge on sensitivity, specificity and reproducibility of symptoms and signs in many other diseases. This approach, i.e. attempting to quantify clinical performance, should be encouraged in medical textbooks, practice and teaching.

Relative values of clinical and technological assessment

Diagnostic tests are mostly valuable as complementary information to clinical assessment, particularly when the pre-test probability of disease is intermediate. Junior clinicians often forget this basic tenet and order tests 'to diagnose' or 'to rule out' conditions rather than to corroborate or challenge a clinical hypothesis. With this misconception in mind, a comparison of the diagnostic power of clinical assessment vs. that of ancillary testing might be useful to

Table 3 Comparison of the power of clinical evaluation vs. current tests in some common disorders

	LR^+	LR^-
Diagnosis of deep-vein thrombosis		
Clinical diagnosis[41]*	17	0.15
Ultrasound or duplex[7]		
symptomatic	32	0.03
asymptomatic	29	0.4
Diagnosis of streptococcal pharyngitis		
Clinical scoring[60]**	6.3	0.18
Rapid antigen test[61]	7	0.33
Culture[62]	9	0.11
Diagnosis of renovascular hypertension		
Clinical suspicion[17]***	66	
Captopril renogram[17]	18.4	0.08
Magnetic resonance angiography†	18.8	0.06
Duplex[18]	49	0.02
Diagnosis of ankylosing spondylitis‡		
Must leave bed during night	3.13	0.44
Pain not relieved by lying down	1.56	0.41
Duration >3 months	1.54	0.55
Morning stiffness >30 minutes	1.55	0.61
Age at onset <35 years	1.31	0.28
Insidious onset	1.07	0.93
Summation of clinical features§	16.3	0.01
HLA-B27 testing††	11.5	0.09

LR^+, LR^-, likelihood ratios for presence or absence of finding(s). * Presence of three or more major points (and no alternative diagnosis) from a list of points including recent immobilization, cancer, paralysis, calf swelling and localized tenderness along the deep venous system increases the chance of deep-vein thrombosis from 25% to 85%. ** Presence of tonsillar exudates, cervical lymphadenitis, fever and absence of cough. *** DBP >120 mmHg with progressive renal failure or refractory to aggressive treatment, accelerated or malignant hypertension (with grade III or IV retinopathy), increase in plasma creatinine with converting enzyme inhibitors, or asymmetry of the kidneys. † Using sensitivity and specificity of 94% and 95%, respectively, averaged from several studies on more than 200 patients.[19–23] ‡ Data are from reference 58. LRs were calculated from sensitivity and specificity. § LRs were calculated by multiplication of LR of all clinical features, assuming presence of all of them and their independence from each other. †† Data are for Whites, from reference 59, using sensitivity and specificity of 92% each.

Table 4 Prediction of future cardiac events in survivors of myocardial infarction

Test	LR^+	LR^-
Exercise EKG[16]	2	0.67
Thallium[15]	1.9	0.43
Ejection fraction*		
>50%	0.61	
35–49%	1.31	
<35%	3.23	
Coronary angiography*		
normal	0.33	
one vessel	0.89	
two vessels	1.78	
three vessels	3.56	
left main	3.82	
Clinical predictors[46]**	4.7	0.65
Bilateral ear creases[47]	1.8	0.49
Psychosocial stress***	2	0.4
Summation of above clinical data†	16.9	0.13
Summation of all non-invasive data†	109	0.05

LR^+, LR^-, likelihood ratios for presence or absence of finding(s). * Four-year mortality, from an average of about 10% in the CASS study[63]. ** Presence of five items from: prior angina, hypertension, diabetes, peripheral vascular disease, anterior wall infarction, cardiac failure. *** Depression or emotional distress, LRs calculated as average from several recent studies.[64–69] † LRs calculated by multiplication of LR of above features, assuming presence or absence of all of them and their independence from each other.

show the respective limitation and relative accuracy of each individual approach.

In many instances, the LRs for positive tests range between 2 and 30, and are often below 10 (Table 1). As illustrated in Tables 2 and 3, the LRs for clinical assessment of diseases, when such an estimate is available from the literature, is often in the same range or higher than 10. For the diagnosis of systemic lupus erythematosus, the presence of a malar rash has a LR of 7, whereas a fluorescent antinuclear antibody has a LR of only 3.2.[70] In attempting to exclude the diagnosis of acute myocardial infarction in the emergency room, a ECG has a negative LR of 0.27 and serial CPKs after 24 h have a negative LR of 0.03,[13] while a fully pleuritic character of the pain can almost immediately exclude the diagnosis.[71] While the clinical diagnosis of streptococcal pharyngitis is often viewed as unreliable, the LRs for a simple clinical score vs. that of microbiology tests are comparable (Table 3). Another illustrative example is the diagnosis of renovascular hypertension, where a positive captopril renogram increases the odds by approximately 18-fold. Certain clinical findings, such as a diastolic blood pressure of over 120 mmHg refractory to therapy, evidence for occlusive vascular disease elsewhere, an abdominal bruit or renal failure induced by administration of a converting enzyme inhibitor, can predict the probability of renovascular hypertension as over 25% (odds of 25 : 75 or 1 : 3).[17] If the prevalence of renovascular hypertension among unselected hypertensive patients is estimated at 0.5% (odds of 1 : 199), the LR generated by the clinical findings in this example is 66 (199 : 3). This compares favourably with the LR of some of the best diagnostic tests for this condition,

such as duplex or magnetic resonance angiography (see Table 3). For the prediction of an adverse prognosis in patients after a myocardial infarction, simple clinical features which are cheap and fast to obtain, may readily provide more powerful information than physiological or invasive tests (such as exercise ECG or coronary angiography), by multiplication of LRs for clinical findings, as illustrated in Table 4 and in Figure 3, which shows that combination of non-invasive data can be more informative than coronary angiography. These examples illustrate that judgment from clinical findings may be a predictor of disease as powerful as any of the tests available to confirm the diagnosis in question.

Power of combined clinical information and ancillary testing

Since clinical assessment and ancillary testing provide largely independent information, their combination allows multiplication of their LRs with a sharp increase in diagnostic accuracy. An example of this synergism is illustrated in Figure 4, for a patient who suffers from a disease with a prevalence of 5%, for which both the refined clinical evaluation and the best test have each a diagnostic power estimated by a LR of about 10. Application of either the refined clinical evaluation only (e.g. to save the cost of a test) or the test alone (by a physician who underestimates the power of clinical evaluation) would each yield the same post-test probability of disease of 34%. By contrast, the combination of the two modalities, which has now the joined LR of $10 \times 10 = 100$, raises the post-test probability of disease to 84%, illustrating the remarkable synergism achievable from combination of clinical assessment and ancillary modalities. Patient management with a probability of diagnosis of 84% is generally more practicable than with a suspicion of disease of 34%, which means in words: 'this diagnosis is likely to be wrong' and requires additional investigations. This powerful sequence of validation of a clinical concept by testing is one of the most important steps in the diagnostic process, akin to the probing of a scientific hypothesis by an experiment in basic research (similarity of reasoning may explain why research could make better doctors). LRs put the value of testing in proper perspective, showing the increment in diagnostic certainty expectable from a test or from a refined clinical evaluation and the remarkable synergism from these combined modalities.

Advances in diagnostic modalities may enhance

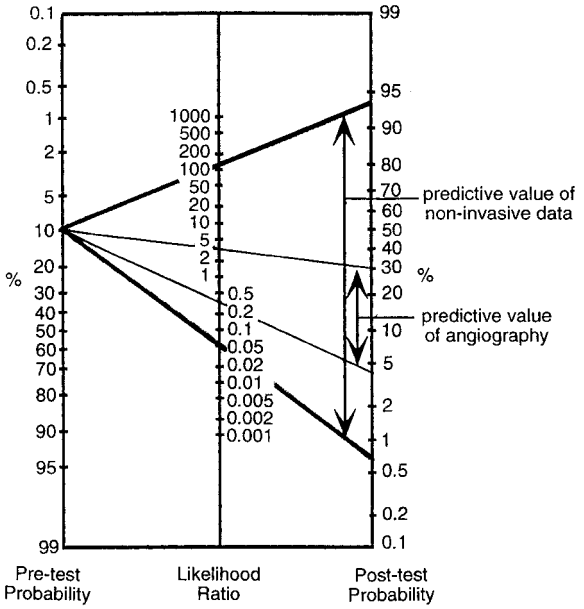

Figure 3. Prediction of future cardiac events in survivors from a myocardial infarction. For the prediction of an adverse prognosis in patients after a myocardial infarction, clinical features combined with non-invasive data (Table 4) may provide more powerful prognostic information than coronary angiography (sometimes considered to obtain a more 'precise' assessment of disease severity).

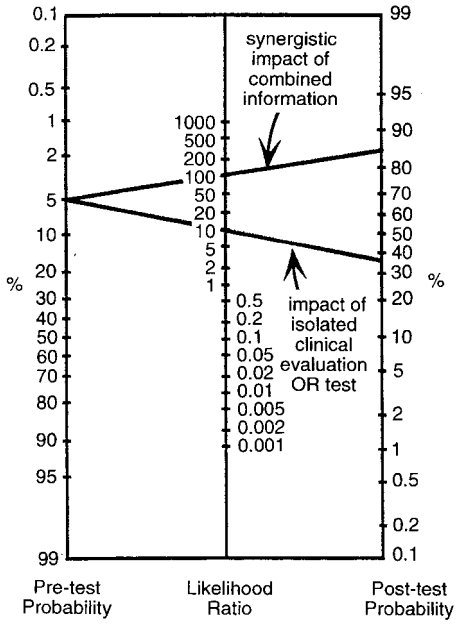

Figure 4. Synergistic impact of clinical assessment and testing. For a disease with a prevalence of 5%, for which the refined clinical evaluation and the best ancillary test have each a LR of about 10 (for instance, as an approximation for the LRs in ankylosing spondylitis, see Table 3), application of either the clinical evaluation or the test would each yield a post-test probability of disease of 34%. By contrast, the combination of the two modalities, which has the joined LR of $10 \times 10 = 100$, raises the post-test probability of disease to 84%, illustrating the remarkable synergism achievable from combined clinical assessment and ancillary modalities.

patient care only if guided by careful clinical judgement: before ordering a test, one must consider potential benefit and risks, know its performance to define how the result may help reaching or rejecting the diagnosis and whether the result may alter management.[75,76] An ancillary test should be used to probe the clinical hypothesis and increase confidence but not to obtain certainty in the diagnosis[77] as often-implied in defensive medicine. To that effect, probabilistic thinking should be enhanced in clinical teaching.[78]

In summary, LRs collected from the literature show that refined clinical evaluation may be equivalent or even superior to ancillary testing, and that the combination of the two approaches sharply increases accuracy in the diagnostic process. Putting the value of testing in proper perspective is one of the most important challenges to the teaching of modern medicine.

Acknowledgements

The authors are grateful to Professor J. Benbassat for his incisive criticism.

References

1. Goodwin J. The importance of clinical skills. *Br Med J* 1995; **310**:1281–2.
2. Fletcher RH, Fletcher SW. Has medicine outgrown physical diagnosis? *Ann Intern Med* 1992; **117**:786–7.
3. Department of Cinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont. Interpretation of diagnostic data: 5. How to do it with simple maths. *Can Med Assoc J* 1983; **129**:947–54.
4. Sox HCJ, Blatt M, Higgins M, Marton KI. *Medical Decision Making*. Stoneham MA, Butterworths, 1988.
5. Dujardin B, Van, den, Ende, J, Van GA, Unger JP, Van, der, Stuyft, P. Likelihood ratios: a real improvement for clinicdecision? *Eur J Epidemiol* 1994; **10**:29–36.
6. Sox HC. The evaluation of diagnostic tests. *Ann Rev Med* 1996; **47**:463–71.
7. Weinmann EE, Salzman E. Deep-vein thrombosis. *N Engl J Med* 1994; **331**:1630–41.
8. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990; **263**:2753–9.
9. van Rossum AB, Treurniet FE, Kieft GJ, Smith SJ, Schepers BR. Role of spiral volumetric computed tomographic scanning in the assessment of patients with clinical suspicion of pulmonary embolism and an abnormal ventilation/perfusion lung scan. *Thorax* 1996; **51**:23–8.
10. Cauvain O, Remy JM, Remy J, Petyt L, Beregi JP, Steinling M, Duhamel A. Spiral CT angiography in the diagnosis of central pulmonary embolism: comparison with pulmonary angiography and scintigraphy. *Rev Mal Respir* 1996; **13**:141–53.
11. Remy JM, Remy J, Deschildre F, Artaud D, Beregi JP, Hossein FC, Marchandise X, Duhamel A. Diagnosis of pulmonary embolism with spiral CT: comparison with pulmonary angiography and scintigraphy. *Radiology* 1996; **200**:699–706.
12. van Rossum AB, Pattynama PM, Ton ER, Treurniet FE, Arndt JW, van Eck B, Kieft GJ. Pulmonary embolism: validation of spiral CT angiography in 149 patients. *Radiology* 1996; **201**:467–70.
13. Lee TH, Goldman L. Serum enzyme assays in the diagnosis of acute myocardial infarction. Recommendations based on a quantitative analysis. *Ann Intern Med* 1986; **105**:221–33.
14. Schweitzer P. The electrocardiographic diagnosis of acute myocardial infarction in the thrombolytic era. *Am Heart J* 1990; **???**:642–54.
15. Kotler TS, Diamond GA. Exercise thallium–201 scintigraphy in the diagnosis and prognosis of coronary artery disease. *Ann Intern Med* 1990; **113**:684–702.
16. Zaret BL, Wackers FJ. Nuclear cardiology (1). *N Engl J Med* 1993; **329**:775–83.
17. Mann SJ, Pickering TG. Detection of renovascular hypertension. State of the art: 1992. *Ann Intern Med* 1992; **117**:845–53.
18. Olin JW, Piedmonte MR, Young JR, DeAnna S, Grubb M, Childs MB. The utility of duplex ultrasound scanning of the renal arteries for diagnosing significant renal artery stenosis. *Ann Intern Med* 1995; **122**:833–8.
19. Debatin JF, Spritzer CE, Grist TM, Beam C, Svetkey LP, Newman GE, Sostman HD. Imaging of the renal arteries: value of MR angiography. *Am J Roentgenol* 1991; **157**:981–90.
20. Silverman JM, Friedman ML, Van AR. Detection of main renal artery stenosis using phase-contrast cine MR angiography. *Am J Roentgenol* 1996; **166**:1131–7.
21. Kaufman JA, Geller SC, Petersen MJ, Cambria RP, Prince MR, Waltman AC. MR imaging (including MR angiography) of abdominal aortic aneurysms: comparison with conventional angiography. *Am J Roentgenol* 1994; **163**:203–10.
22. Grist TM. Magnetic resonance angiography of renal artery stenosis. *Am J Kidney Dis* 1994; **24**:700–12.
23. Kent KC, Edelman RR, Kim D, Steinman TI, Porter DH, Skillman JJ. Magnetic resonance imaging: a reliable test for the evaluation of proximal atherosclerotic renal arterial stenosis. *J Vasc Surg* 1991; **13**:311–18.
24. Pauker SG, Kopelman RI. Interpreting hoofbeats: can Bayes help clear the haze? *N Engl J Med* 1992; **327**:1009–13.
25. Sinclair D, Wilson S, Toi A, Greenspan L. The evaluation of suspected renal colic: ultrasound scan versus excretory urography. *Ann Emerg Med* 1989; **18**:556–9.
26. Middleton WD, Dodds WJ, Lawson TL, Foley WD. Renal calculi: sensitivity for detection with US. *Radiology* 1988; **167**:239–44.
27. Guyatt GH, Patterson C, Ali M, Singer J, Levine M, Turpie I, Meyer R. Diagnosis of iron-deficiency anemia in the elderly. *Am J Med* 1990; **88**:205–9.
28. Stabler SP, Allen RH, Savage DG, Lindenbaum J. Clinical spectrum and diagnosis of cobalamin deficiency. *Blood* 1990; **76**:871–81.
29. Matchar DB, Feussner JR, Millington DS, Wilkinson RHJ, Watson D, Gale D. Isotope-dilution assay for urinary methylmalonic acid in the diagnosis of vitamin B12 deficiency. *Ann Intern Med* 1987; **106**:707–10.

30. Aronson MD, Bor DH. Blood cultures. *Ann Intern Med* 1987; **106**:246–53.

31. Lowry RW, Zoghbi WA, Baker WB, Wray RA, Quinones MA. Clinical impact of transesophageal echocardiography in the diagnosis and management of infective endocarditis. *Am J Cardiol* 1994; **73**:1089–91.

32. Pomilla PV, Brown RB. Outpatient treatment of community-acquired pneumonia in adults. *Arch Intern Med* 1994; **154**:1793–802.

33. Barrett CE. The nonvalue of sputum culture in the diagnosis of pneumococcal pneumonia. *Am Rev Respir Dis* 1971; **103**:845–8.

34. Hyslop NJ, Swartz MN. Bacterial meningitis. *Postgrad Med* 1975; **58**:120–8.

35. Martin WJ. Rapid and reliable techniques for the laboratory detection of bacterial meningitis. *Am J Med* 1983; **75**:119–23.

36. Marton KI, Gean AD. The spinal tap: a new look at an old test. *Ann Intern Med* 1986; **104**:840–8.

37. Tunkel AR, Scheld WM. Acute meningitis. In: Mandell GL, Bennett JE, Dolin R, eds. *Mandell, Douglas and Bennett's principles and practice of infectious diseases*. New York, Churchill-Livingstone, 1995:831–65.

38. Doern GV, Coughlin RT, Wu L. Laboratory diagnosis of Clostridium difficile-associated gastrointestinal disease: comparison of a monoclonal antibody enzyme immunoassay for toxins A and B with a monoclonal antibody enzyme immunoassay for toxin A only and two cytotoxicity assays. *J Clin Microbiol* 1992; **30**:2042–6.

39. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. *N Engl J Med* 1996; **334**:155–9.

40. Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975; **293**:257.

41. Wells PS, Hirsh J, Anderson DR, Lensing AW, Foster G, Kearon C, Weitz J, D'Ovidio R, Cogo A, Prandoni P, Giralomi A, Ginsberg JS. Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 1995; **345**:1326–30.

42. Heijboer H, Buller HR, Lensing AW, Turpie AG, Colly LP, ten CJ. A comparison of real-time compression ultrasonography with impedance plethysmography for the diagnosis of deep-vein thrombosis in symptomatic outpatients. *N Engl J Med* 1993; **329**:1365–9.

43. Garino JP, Lotke PA, Kitziger KJ, Steinberg ME. Deep venous thrombosis after total joint arthroplasty. The role of compression ultrasonography and the importance of the experience of the technician. *J Bone Joint Surg Am* 1996; **78**:1359–65.

44. Williams JJ, Simel DL, Roberts L, Samsa GP. Clinical evaluation for sinusitis. Making the diagnosis by history and physical examination. *Ann Intern Med* 1992; **117**:705–10.

45. Monsch AU, Bondi MW, Salmon DP, Butters N, Thal LJ, Hansen LA, Wiederholt WC, Cahn DA, Klauber MR. Clinical validity of the Mattis Dementia Rating Scale in detecting Dementia of the Alzheimer type. A double cross-validation and application to a community-dwelling sample. *Arch Neurol* 1995; **52**:899–904.

46. Kornowski R, Goldbourt U, Zion M, Mandelzweig L, Kaplinsky E, Levo Y, Behar S. Predictors and long-term prognostic significance of recurrent infarction in the year after a first myocardial infarction. SPRINT Study Group. *Am J Cardiol* 1993; **72**:883–8.

47. Elliott WJ, Powell LH. Diagonal earlobe creases and prognosis in patients with suspected coronary artery disease. *Am J Med* 1996; **100**:205–11.

48. Silver MT, Rose GA, Paul SD, O'Donnell CJ, O'Gara PT, Eagle KA. A clinical rule to predict preserved left ventricular ejection fraction in patients after myocardial infarction. *Ann Intern Med* 1994; **121**:750–6.

49. Ewy GA. The abdominojugular test: technique and hemodynamic correlates. *Ann Intern Med* 1988; **109**:456–60.

50. Wade DS, Marrow SE, Balsara ZN, Burkhard TK, Goff WB. Accuracy of ultrasound in the diagnosis of acute appendicitis compared with the surgeon's clinical impression. *Arch Surg* 1993; **128**:1039–44.

51. Lyons AR, Tomlinson JE. Clinical diagnosis of tears of the rotator cuff. *J Bone Joint Surg* 1992; **74**:414–15.

52. Oberlander MA, Shalvoy RM, Hughston JC. The accuracy of the clinical knee examination documented by arthroscopy. A prospective study. *Am J Sports Med* 1993; **21**:773–8.

53. Rubinstein RJ, Shelbourne KD, McCarroll JR, VanMeter CD, Rettig AC. The accuracy of the clinical examination in the setting of posterior cruciate ligament injuries. *Am J Sports Med* 1994; **22**:550–7.

54. Blaauw I, Dijkmans B, Bouma P, van der Linden S. Rational diagnosis and treatment in unclassified arthritis: how clinical data may guide requests for Lyme serology and antibiotic treatment. *Ann Rheum Dis* 1993; **52**:206–10.

55. Singhi S, Dhawan A,KatarS, Walia BN. Clinical signs of pneumonia in infants under 2 months. *Arch Dis Child* 1994; **70**:413–17.

56. Havas TE, Motbey JA, Gullane PJ. Prevalence of incidental abnormalities on computed tomographic scans of the paranasal sinuses. *Arch Otolaryngol Head Neck Surg* 1988; **114**:856–9.

57. Flinn J, Chapman ME, Wightman AJ, Maran AG. A prospective analysis of incidental paranasal sinus abnormalities on CT head scans. *Clin Otolaryngol* 1994; **19**:287–9.

58. Gran JT. An epidemiological survey of the signs and symptoms of ankylosing spondylitis. *Clin Rheumatol* 1985; **4**:161–9.

59. Khan MA, Khan MK. Diagnostic value of HLA-B27 testing ankylosing spondylitis and Reiter's syndrome. *Ann Intern Med* 1982; **96**:70–6.

60. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. *Med Dec Making* 1981; **1**:239–46.

61. Reed BD, Huck W, French T. Diagnosis of group A beta-hemolytic Streptococcus using clinical scoring criteria, Directigen 1-2-3 group A streptococcal test, and culture. *Arch Intern Med* 1990; **150**:1727–32.

62. Kellogg JA. Suitability of throat culture procedures for detection of group A streptococci and as reference standards for evaluation of streptococcal antigen detection kits. *J Clin Microbiol* 1990; **28**:165–9.

63. Mock MB, Ringqvist I, Fisher LD, Davis KB, Chaitman BR, Kouchoukos NT, Kaiser GC, Alderman E, Ryan TJ, Russell ROJ, Mullin S, Fray D, Killip T, *et al*. Survival of medically treated patients in the coronary artery surgery study (CASS) registry. *Circulation* 1982; **66**:562–8.

64. Berkman LF, Leo SL, Horwitz RI. Emotional support and survival after myocardial infarction. A prospective, population-based study of the elderly. *Ann Intern Med* 1992; **117**:1003–9.

65. Linden W, Stossel C, Maurice J. Psychosocial interventions

for patients with coronary artery disease: a meta-analysis. *Arch Intern Med* 1996; **156**:745–52.

66. Frasure SN, Lesperance F, Talajic M. Depression following myocardial infarction. Impact on 6-month survival. *JAMA* 1993; **270**:1819–25.
67. Denollet J, Sys SU, Stroobant N, Rombouts H, Gillebert TC, Brutsaert DL. Personality as independent predictor of long-term mortality in patients with coronary heart disease. *Lancet* 1996; **347**:417–21.
68. Ladwig KH, Roll G, Breithardt G, Budde T, Borggrefe M. Post-infarction depression and incomplete recovery 6 months after acute myocardial infarction. *Lancet* 1994; **343**:20–3.
69. Allison TG. Identification and treatment of psychosocial risk factors for coronary artery disease. *Mayo Clin Proc* 1996; **71**:817–19.
70. Clough JD, Elrazak M, Calabrese LH, Valenzuela R, Braun WB, Williams GW. Weighted criteria for the diagnosis of systemic lupus erythematosus. *Arch Intern Med* 1984; **144**:281–5.
71. Lee TH, Cook EF, Weisberg M, Sargent RK, Wilson C, Goldman L. Acute chest pain in the emergency room. Identification and examination of low-risk patients. *Arch Intern Med* 1985; **145**:65–9.
72. Sisson JC, Schoomaker EB, Ross JC. Clinical decision analysis. The hazard of using additional data. *JAMA* 1976; **236**:1259–63.
73. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three medical eras. *N Engl J Med* 1983; **308**:1000–5.
74. Kirch W, Schafii C. Misdiagnosis at a university hospital in 4 medical eras. *Medicine (Baltimore)* 1996; **75**:29–40.
75. Weinstein MC, Fineberg HV, Elstein AS, *et al. Clinical Decision Analysis*. Philadelphia, WB Saunders, 1980:141.
76. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; **302**:1109–17.
77. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med* 1989; **320**:1489–91.
78. Walter SD, Mitchell A, Southwell D. Use of certainty of opinion data to enhance clinical decision making. *J Clin Epidemiol* 1995; **48**:897–902.
79. Lamont JT. Bacterial infections of the colon. In: Yamada T, Alpers DH, Owyang C, Powell DW, Silverstein FE, eds. *Textbook of gastroenterology*. Philadelphia. Lippincott, 1991:1749–68.
80. LaCombe MA. On bedside teaching. *Ann Intern Med* 1997; **126**:217–20.

Appendix I: Calculation of LR

(1) LR for a positive test result = sensitivity/(1–specificity)
(2) LR for a negative test result = (1–sensitivity)/specificity

Appendix II: Interconversion between probability and odds

Information from reference 77.
(1) Odds = probability/(1–probability)
(2) Probability = odds/(1 + odds).

Conversion of pre-test probability to pre-test odds:

In example 1, the pre-test probability of deep venous thrombosis is 85%, thus the pre-test odds is 85/15 = 5.667

Conversion of pre-test odds to post-test odds after the test (LR of 32 for a positive duplex scan, 0.03 for a negative test)

Post-test odds = pre-test odds × LR. In example 1, if the test is positive, the post-test odds = 5.667 × 32 = 181.333. For a negative result, the post-test odds = 5.667 × 0.03 = 0.17.

Conversion of post-test odds to post-test probability

In example 1, for a positive result, the post-test probability = 181.333/182.333 = 99.4%. For a negative test, the post-test probability = 0.17/1.17 = 14.5%. To avoid the calculations, see the use of the normogram in Figure 1**a**.

Conversion of an increase in probability of disease into a LR

In example 1, an increase in the probability of disease from 25% to 85% (by a refined clinical assessment), is an increment in odds from 25/75 (0.333) to 85/15 (5.667), equivalent to an LR of 5.667/0.333, or 17.