The interpretation of diagnostic tests: A primer for physiotherapists

Megan Davidson
La Trobe University

This paper outlines a practical approach to assist physiotherapists to interpret the results of diagnostic or screening tests. Diagnostic tests are used during clinical assessment to increase or decrease the clinician's estimate of the likelihood that a client has a particular condition. A negative result for a test that is 100% sensitive can rule a condition out (SnOUT), and a positive result for a test that is 100% specific can rule a condition in (SpIN). However, tests are rarely 100% accurate, and false positive and false negative results can occur. The examining therapist needs to estimate the probability that a client has a particular condition (the pre-test probability), then estimate the extent to which they are more or less certain given a positive or negative test result (the post-test probability). The likelihood ratio, which combines the information provided by a test's sensitivity and specificity, is the most useful tool for the clinical interpretation of test results. **[Davidson M (2002): The interpretation of diagnostic tests: A primer for physiotherapists. *Australian Journal of Physiotherapy* 48: 227-233]**

Key words: Diagnosis; Intervention Studies; Likelihood Functions; Sensitivity and Specificity

During the patient interview, a physiotherapist develops hypotheses about possible causes or diagnoses for the presenting problem. These hypotheses are then tested during the objective assessment, or physical examination, using clinical tests. A diagnostic test seeks to determine whether or not a person has or does not have a particular condition. The evidence-based practitioner needs to be able to locate and evaluate the quality of research papers that report on the accuracy of diagnostic tests (Greenhalgh 1997, Sackett et al 2000). A preliminary step in becoming an evidence-based practitioner is to acquire a thorough understanding of the characteristics of diagnostic tests. Clinicians need to appreciate the extent to which a positive or negative test result can confirm or disprove their diagnostic hypothesis. The aim of this paper is to provide physiotherapists with an understanding of diagnostic test characteristics and how these can be interpreted in the clinical setting.

Sensitivity and specificity Tests are rarely 100% accurate, so false positives and false negatives can occur. The findings of a test are generally plotted against the actual diagnosis in a "two by two" or "truth" table (Figure 1). The characteristics of a diagnostic test, defined in Table 1, are calculated from the truth table.

Where sensitivity or specificity is extremely high (98-100%), interpretation of test results is simple. If the sensitivity is extremely high, we can be sure that a negative test will rule the disorder out. This is because there can be very few false negatives (Cell c). If the specificity is extremely high we can be sure that a positive test will rule the disorder in. This is because there can be very few false positives (Cell b). These rules can be recalled using the mnemonics SnOUT and SpIN:

SnOUT : If **Sn**sitivity is high, a negative test will rule the disorder **OUT**.

SpIN : If **Sp**ecificity is high, a positive test will rule the disorder **IN**.

Table 2 shows some SpINs and SnOUTs of interest to physiotherapists. These have been chosen on the basis of their high sensitivities and specificities. Unfortunately, it is rare for sensitivity and specificity to reach such giddy heights.

Sensitivity and specificity tell us how often a test is positive and negative in people who we already know have the condition or not. Clinically, however, we do not initially know whether or not our client has the condition. In this case, it is essential that we know how to interpret a positive or negative test result.

Positive and negative predictive values Predictive values tell us how likely it is that a person who tests positive has the disorder, and how likely it is that a person who tests negative does not have the disorder. Predictive values are also called "post-test probabilities" (Go 1998). Unfortunately the predictive values only apply when the clinical prevalence is identical to that reported in the study. Prevalence changes dramatically depending on where the test is being performed. For example, in a general physiotherapy outpatient department or practice, the prevalence of patients with anterior cruciate ligament (ACL) tears will be lower than in a sports clinic that specialises in knee injuries. Prevalence is also called the "pre-test probability" that a person has the disorder. The pre-test probability of a client having an ACL tear is higher in the sports clinic than in the general practice.

		Target Disorder				
		Present			Absent	Totals
Diagnostic	Positive	True Positives	a	b	False Positives	a+b
Test Result	Negative	False Negatives	c	d	True Negatives	c+d
	Totals		a+c	b+d		a+b+c+d

Figure 1. A two by two or "truth" table

Table 1. Definition and calculation of test characteristics.

Test characteristic	Definition	Calculation
Sensitivity	The proportion of people who have the disorder who test positive.	a/(a+c)
Specificity	The proportion of people who do not have the disorder who test negative	d/(b+d)
Positive Predictive Value (PPV)	The proportion of people who test positive who have the disorder (The probability that someone who tests positive has the disorder)	a/(a+b)
Negative Predictive Value (NPV)	The proportion of people who test negative who do not have the disorder (The probability that someone who tests negative does not have the disorder)	d/(c+d)
Accuracy	The proportion of people who were correctly identified as either having or not having the disorder.	(a+d)/(a+b+c+d)
Prevalence	The proportion of people in the sample who had the disorder.	(a+c)/(a+b+c+d)

Hint: Sensitivity and Specificity are read down the columns of the 2 x 2 table, while PPV and NPV are read across rows.

Table 2. Some SpINs and SnOUTs.

Target disorder	SpIN (and specificity) Positive test rules the disorder in	SnOUT (and sensitivity) Negative test rules the disorder out	Source
Cancer as a cause of low back pain	Previous history of cancer (98%)	Age ≥ 50, or history of cancer, or unexplained weight loss or failure of conservative therapy (100%)	Deyo, Rainville and Kent (1992)
Cervical spine fractures		Canadian C-Spine Rules (100%)	Stiell et al (2001)
Knee fractures		Ottawa Knee Rules (100%)	Emparanza and Aginaga (2001)
Ankle fractures		Ottawa Ankle Rules (99%)	Auleley et al (1998) Stiell et al (1995)
Pleural effusion	Auscultatory percussion note loud and sharp (100%)	Auscultatory percussion note soft or dull (96%)	Guarino and Guarino (1994)

Canadian Cervical Spine Rules: Question 1: Are there any high risk factors (age ≥ 65 or dangerous mechanism or parasthesias in extremities)? If YES – x-ray indicated, if NO - Question 2: Are there any low-risk factors that allow safe assessment of range of motion? (simple rear-end collision or sitting in Emergency Department or ambulatory since injury or delayed onset neck pain or absence of midline cervical spine tenderness) If NO – x-ray indicated, if YES – Question 3: Able to actively rotate neck 45° left and right. If NO – x-ray indicated, if YES x-ray not indicated.

Ottawa Knee Rules: aged ≥ 55, isolated patella tenderness, tenderness at the head of fibula, unable to achieve 90° flexion, or is unable to bear weight immediately and on examination.

Ottawa Ankle Rules: Ankle: Pain in the malleolar zone and tenderness of the poster edge or tip of the lateral or medial malleolus, or unable to bear weight immediately and on examination. Foot: Pain in the midfoot region, tenderness of base of 5th metatarsal or navicular, or unable to bear weight immediately and on examination.

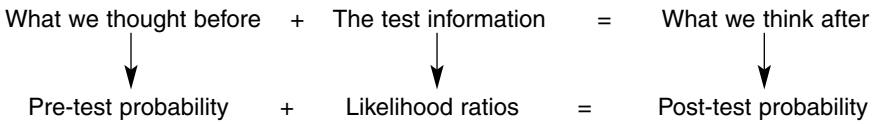

Figure 2. Adapted from Go 1988.

Because predictive values only apply to populations with the same prevalence, they are not very useful values. Consequently you can now almost disregard the positive predictive values and negative predictive values. A more useful tool for interpreting clinical tests is the likelihood ratio.

Likelihood ratios Likelihood ratios summarise the information contained in both sensitivity and specificity (Dujardin et al 1994). A likelihood ratio (LR) tells us how likely a given test result is in people with the condition, compared with how likely it is in people without the condition.

Calculation of LRs is simple:

- Likelihood ratio (test +ve) = sensitivity/(1-specificity)
- Likelihood ratio (test –ve) = (1-sensitivity)/specificity

However, it is easy to get confused when calculating LRs. To make the LR calculations work, sensitivity and specificity must be expressed as a decimal, ie 0.95. Alternatively, they may be expressed as percentages by changing the 1 in 1-specificity and 1-sensitivity to 100.

Interpreting the LR is also simple. The higher the positive LR, the more certain you can be that a positive test indicates the person has the disorder. The lower the negative LR, the more certain you can be that a negative test indicates the person does not have the disorder. If the LR is close to 1, then the test will not provide much information. That is, the likelihood that a person has, or does not have, a condition will not change at all if the LR is exactly 1.0 and the diagnostic hypothesis is no closer to being confirmed or rejected.

Likelihood ratios in a nutshell:

- A +ve LR of 10 or more is an indicator that a positive test will be very good at ruling the disorder IN.
- A –ve LR of 0.1 or less is an indicator that a negative test will be very good at ruling the disorder OUT.
- A LR close to 1.0 will provide little change in probability that a person has or does not have a disorder.

Now to put LRs into practice.

Estimating the pre-test probability Go (1998, p. 13) provides a helpful way of thinking about the contribution of test results to clinical decision-making (See Figure 2).

"What we thought before" (the pre-test probability) is how likely a clinician thinks it is, before doing the test, that a person has the disorder. Pre-test probabilities can be obtained from either published data or the clinician's subjective impression or personal experience (Elstein and Schwarz 2002, Go 1998, Sox et al 1988). Here are some examples of published prevalence data: in children aged between 10 and 16 the prevalence of idiopathic scoliosis is 2-4% (Reamy and Slakey 2001); peripheral neuropathy affects 2.4% of the population, peaking at 8% in older people (Hughes 2002); the prevalence of persistent asthma in childhood is 9% (Woolcock et al 2001). More commonly, a clinician will, during the subjective assessment or patient interview, form a hypothesis that a person may have a particular disorder. Based on the information provided by the client, the clinician may think it very likely or unlikely a particular disorder is present, or may be uncertain. To use this information in the most efficient way, it is necessary to put a figure on the level of uncertainty. This establishes the pre-test probability that the person has the disorder ("what we thought before"), so that the test result can then be used to arrive at "what we think after". Note that the clinician nominates pre-test probability for a particular condition after commencing the assessment. The clinician may have already ruled out rare but serious conditions by standard "screening" or "red flag" questions, and the patient's history has already provided some diagnostic information. Once the most likely diagnosis is identified, a conscious expression of the probability of that condition should be made (Sox et al 1988).

Consider the following cases and answer the following questions before proceeding:

Case 1: A 28 year old female sustained a knee injury while skiing. she reports an audible pop, immediate swelling, haemarthrosis (blood was aspirated), and a feeling of instability.

Case 2: A 55 year old female presents with numbness and tingling in the right hand affecting the thumb, index and middle fingers, worse at night. She also reports a tendency to drop objects held in the right hand.

For each case, what is your provisional diagnosis? What is the probability that this person has this disorder? (Express the probability as a value between 0 and 1.0 or as a percentage.)

Did you identify the most likely diagnosis in Case 1 as a torn ACL, and in Case 2 carpal tunnel syndrome? Estimating pre-test probabilities relies on clinical experience, so you might have opted for anything from

Table 3. Likelihood ratios for ACL and meniscus tests.

ACL Tests	+LR	-LR
Anterior draw test	3.8	0.3
Lachman's test	42	0.1
Meniscal Tests		
McMurray test	1.3	0.8
Medial-lateral grind test	4.8	0.4

Source: Solomon et al (2001)

60% to 90% as your pre-test probability. There is no absolutely correct answer.

Now use the pre-test probability for the knee injury in Case 1 (say it was 70%) to see how the LRs for common clinical tests (Table 3) can help to confirm or reject the initial diagnosis of an ACL tear, and to rule out meniscal damage. The sensitivity and specificity of tests are taken from Solomon et al (2001). The likelihood ratio nomogram allows us to quickly estimate a post-test probability for a positive or negative test result (Figure 3). For the primary diagnosis of ACL tear, the best test is Lachman's test because it has a much higher positive LR than the anterior draw test. Drawing a line from a pre-test probability of 0.70 through a LR of 42 results in a post-test probability of 0.99. With one test you have moved from being 70% to 99% certain that the patient has an ACL tear.

Meniscal tears often accompany ACL tears, and although the patient does not report any locking or catching sensations, you want to exclude a meniscal tear. Your pre-test probability is 0.3, and you select the test with the lowest negative LR (Table 3). A negative medial-lateral grind test moves your pre-test probability of 0.3 to a post-test probability of 0.15 or a 15% chance the patient has meniscal damage. You would like to be even more certain, so you select the next best test for ruling out meniscal tears, the McMurray test. The post-test probability for the first test now becomes the pre-test probability for the second test. A negative McMurray test moves your pre-test probability of 15% to 12%. This is a very small change because the LR was close to 1.

Points to remember:

- Tests with the highest positive LR will provide the most information in the event of a positive test.
- Tests with the lowest negative LR provide the most information in the event of a negative test.
- When using a sequence of tests the post-test probability of the first test then becomes the pre-test probability for the next test.

An alternative method of calculating the post-test probability is to use simple maths (Go 1998):

1 Estimate the pre-test probability (say 70% or 0.70).

2 Convert this to pre-test odds by dividing probability by 1-probability (odds = 0.70/(1-0.70) = 2.3).

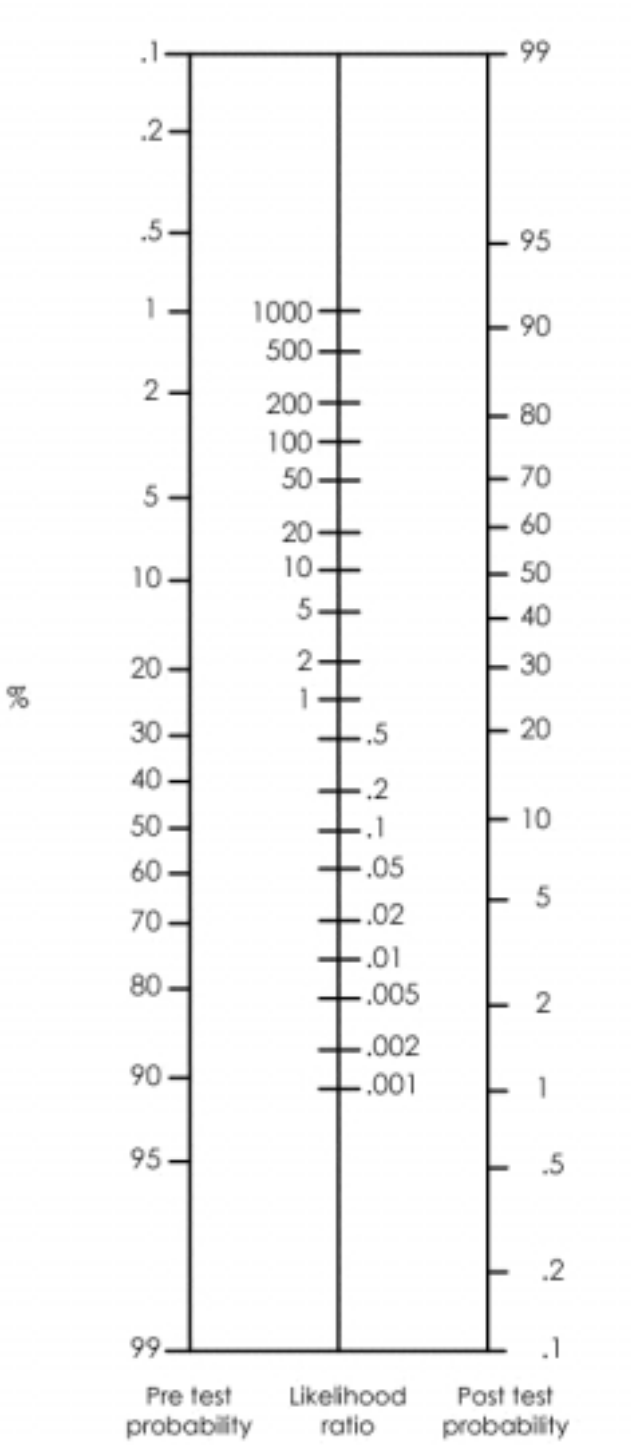

Figure 3. Likelihood ratio nomogram (adapted from Fagan 1975). Note that probabilities are expressed as percentages. To use the nomogram, proceed as follows: Step 1: Mark your pre-test probability on the left of the nomogram. Step 2: Mark the LR of your selected test on the middle line. Step 3. Draw line through these 2 points to determine the post-test probability.

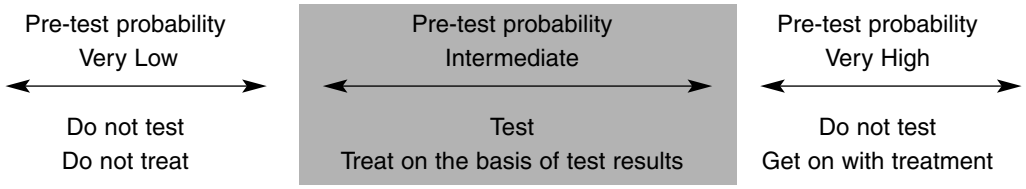

Figure 4. To test or not to test (adapted from Sackett et al 2000, p. 84).

3. Multiply the pre-test odds by the LR (let LR = 42, then 2.3 **x** 42 = 96.6).
4. Convert the post-test odds to post-test probability by dividing odds by odds +1 (post-test probability = 96.6/(96.6 + 1) = 0.99 or 99%.

The mathematical solution is usually a little more precise than the nomogram.

Key points for testing diagnostic hypotheses:

- For the primary diagnostic hypothesis, ie the single most likely explanation for the patient's presenting problem, choose the test(s) with the highest specificity and largest positive LR to confirm the diagnosis.
- For the secondary diagnostic hypothesis, ie a credible alternative explanation, choose the test(s) with the highest sensitivity and smallest negative LR to exclude this disorder.

Although the examples given show tests applied in series, some tests can be applied in parallel. For example the Ottawa Knee Rules have a sensitivity of 100% (SnOUT) for knee fractures (Emparanza and Aginaga 2001). A knee x-ray is only indicated if a patient with a knee injury is: aged 55 or more; or has isolated tenderness of the patella; or has tenderness at the head of fibula; or is unable to achieve 90 degrees flexion; or is unable to bear weight immediately and on examination. If any of the five signs are positive the patient is referred for an x-ray. If all five clinical signs are negative there is 0% post-test probability that the patient has a knee fracture.

Most tests provide only two possible results: condition present or absent; test result normal or abnormal. Multi-level tests have more than two possible results and LRs can be calculated for each level. For example, the results of a ventilation-perfusion test (lung scan) for suspected pulmonary embolism is reported in four categories: normal/near-normal, low, intermediate and high probability (Jaeschke et al 1994). The LRs for these categories are 0.1, 0.36, 1.2 and 18.3 respectively.

A word of warning about low and high pre-test probabilities:

Consider the case of a 32 year old male who presents with sudden onset low back pain, severe pain to both legs, difficulty initiating micturition and "saddle" numbness. What is the most likely diagnosis? Cauda equina syndrome caused by a large postero-central disc herniation is the immediate and strong suspect. The pre-test probability is extremely high (let's say 99%). Are any clinical tests required to confirm this diagnosis? Go to the nomogram and plot a post-test probability for a negative straight leg raise (SLR) test (-ve LR = 0.2). The chance of the person having a disc herniation has reduced from 99% to 95%. There is still a 95% chance the disorder is present. That means that it is most likely the negative test result is a false negative. In practice, you would arrange an urgent medical referral and not bother to do any further testing. You were already 99% certain of the diagnosis and a positive SLR would only have made you 100% sure.

Consider another case scenario. A 32 year old male presents with gradual onset low back pain, ie generalised deep pain in the low back that refers occasionally into the left buttock. As a disc herniation is very unlikely in the absence of leg pain (Deyo et al 1992) the probability that this patient has a disc herniation is very low, say 5%. A positive crossed SLR test (+ve LR = 2.9) moves the probability that the client has a disc herniation from 5% to 10%. There is still a 90% chance the person does not have the disorder. It is most likely the positive test result is a false positive. Conversely, a negative result would only have moved your certainty from 5% to 4%; a disc herniation was almost certainly not the cause of the problem, so why test for it?

The key rules to remember are:

- When pre-test probability is very high, negative test results are usually false negatives.
- When pre-test probability is very low, positive test results are usually false positives.

Figure 4 provides a decision-making aid about whether or not to test based on the pre-test probability that the condition is present. The probability threshold below which a clinician decides not to do a particular test for a particular condition depends on a variety of factors including the seriousness of the condition, the cost, unpleasantness and risks of the test, and the patient's need for reassurance. An over-riding principle is that a test should be done

only if the result could change treatment decisions (Sox et al 1988).

Finally, it has to be said that clinicians require valid studies of diagnostic tests from which they can calculate and apply likelihood ratios. Recent systematic reviews suggest that the methodological quality of such studies is often poor (Massy-Westropp et al 2000, Solomon et al 2001). The key features of a valid study are the selection of a sufficient sample of consecutive patients who are suspected of having the target condition, and the comparison (for all subjects) of the test with a "gold standard" test using blinded testers. That is, the person or persons performing the gold standard test should be blind to the results of the diagnostic test and vice versa (Deeks 2001, Greenhalgh 1997, Sackett and Haynes 2002). Studies that do not have these features will usually over-estimate diagnostic accuracy (Deeks 2001).

Summary

Clinical tests can assist clinicians to increase their level of certainty about whether or not a suspected condition is present. An understanding of the characteristics of diagnostic tests is essential to their clinical interpretation. A very sensitive test rules a condition out (SnOUT), and a very specific test rules a condition in (SpIN). Where sensitivity and specificity are less than perfect, the likelihood ratio nomogram is a useful aid in quantifying the probability that a client does or does not have a particular condition given a positive or negative test result. The clinician first must quantify the pre-test probability of a suspected condition. The likelihood ratio can then be used to calculate the post-test probability that the condition is present. Tests can be used in series with the post-test probability of the first test becoming the pre-test probability of the next test, and so on, until the condition is ruled in or out. Where the pre-test probability is either very high or very low, testing is not recommended, as results can provide little additional certainty as to whether or not a condition is present or absent.

Correspondence Megan Davidson, School of Physiotherapy, La Trobe University, Bundoora, Victoria 3086. E-mail: m.davidson@latrobe.edu.au.

References

Auleley GR, Kerboull L and Durieux P (1998): Validation of the Ottawa ankle rules in France: a study in the surgical emergency department of a teaching hospital. *Annals of Emergency Medicine* 32: 14-18.

Deeks JJ (2001): Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 323: 157-162.

Deyo RA, Rainville J and Kent DL (1992): What can the history and physical examination tell us about low back pain? *JAMA* 268: 760-765.

Dujardin B, Van den Ende J, Van Gompel A, Unger JP and Van der Stuyft P (1994): Likelihood ratios: A real improvement for clinical decision making? *European Journal of Epidemiology* 10: 29-36.

Elstein AS and Schwarz A (2002): Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 324: 729-732.

Emparanza JI and Aginaga JR (2001): Validation of the Ottawa knee rules. *Annals of Emergency Medicine* 38: 364-368.

Fagan TJ (1975): Nomogram for Baye's Theorem. *New England Journal of Medicine* 293: 257.

Go AS (1998): Refining probability: an introduction of the use of diagnostic tests. In Friedland DJ, Go AS, Davoren JB, Shlipak MG, Bent SW, Subak LL and Mendelson T (Eds): Evidence-Based Medicine: A Framework for Clinical Practice. Stamford, Connecticut: Appleton and Lange, pp. 11-33.

Greenhalgh T (1997): How to read a paper. Papers that report diagnostic or screening tests. *BMJ* 315: 540-543.

Guarino JR and Guarino JC (1994): Auscultatory percussion: a simple method to detect pleural effusion. *Journal of General Internal Medicine* 9: 71-74.

Hughes RAC (2002): Regular review. Peripheral neuropathy. *BMJ* 324: 466-469.

Jaeschke R, Guyatt GH and Sackett DL (1994): Users' guides to the medical literature III: How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? *JAMA* 271: 703-707.

Massey-Westropp N, Grimmer K and Bain G (2000): A systematic review of the clinical diagnostic tests for carpal tunnel syndrome. *Journal of Hand Therapy* 25A: 120-127.

Reamy BV and Slakey JB (2001): Adolescent idiopathic scoliosis: review and current concepts. *American Family Physician* 64: 111-116.

Sackett DL and Haynes RB (2002): Evidence base of clinical diagnosis. The architecture of diagnostic research. *BMJ* 324: 539-541.

Sackett DL, Straus SE, Richardson WS, Rosenberg W, and Haynes RB (2000): Evidence-Based Medicine. How to Practice and Teach EBM. Edinburgh: Churchill Livingstone.

Solomon DH, Simel DL, Bates, DW, Katz JN and Schaffer JL (2001): Does this patient have a torn meniscus or ligament of the knee? *JAMA* 286: 1610-1620.

Sox HC, Blatt, MA, Higgins MC and Marton KI (1988): Medical Decision-Making. Boston: Butterworths.

Stiell IG, Wells GA, Laupacis A et al (1995): Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. *BMJ* 311: 594-597.

Stiell IG, Wells GA and Vandemheen KL (2001): The Canadian C-Spine rule for radiography in alert and stable trauma patients. *JAMA* 286: 1841-1848.

Woolcock AJ, Bastiampillai SA, Marks GB and Keena ZA (2001): The burden of asthma in Australia. *Medical Journal of Australia* 175: 141-145.