1

Requesting and interpreting tests

COPYRIGHTED MATERIAL

Learning objectives

To understand:

- ✔ how sample handling, analytical and biological factors can affect test results in health and disease and how these relate to the concept of a test reference range;
- ✔ the concepts of accuracy, precision, test sensitivity, test specificity in the quantitative assessment of test performance.

Introduction

Biochemical tests are crucial to modern medicine. Most biochemical tests are carried out on blood using plasma or serum, but urine, cerebrospinal fluid (CSF), faeces, kidney stones, pleural fluid, etc. are sometimes required. Plasma is obtained by collecting blood into an anticoagulant and separating the fluid, plasma phase from the blood cells by centrifugation. Serum is the corresponding fluid phase when blood is allowed to clot. For many (but not all) biochemical tests on blood, it makes little difference whether plasma or serum is used.

There are many hundreds of tests available in clinical biochemistry but a core of common tests makes up the majority of tests requested. These core tests are typically available from most clinical laboratories throughout the 24-h period. Tests are sometimes brought together in profiles, especially when a group of tests provides better understanding of a problem than a single test (e.g. the liver function test profile). More specialist tests may be restricted to larger laboratories or specialist centres offering a national or regional service.

In dealing with the large number of routine test requests, the modern clinical biochemistry laboratory depends heavily on automated instrumentation linked to a laboratory computing system. Test results are assigned to electronic patient files that allow maintenance of a cumulative patient record. Increasingly, test requests can be electronically booked at the ward, clinic or in General Practice via a terminal linked to the main laboratory computer. Equally, the test results can be displayed on computer screens at distant locations, removing the need to issue printed reports.

In this first chapter, we set out some of the principles of requesting tests and of the interpretation of results. The effects of analytical errors and of physiological factors, as well as of disease, on test results are stressed. Biochemical testing in differential diagnosis and in screening is discussed.

Collection of specimens

Test requests require unambiguous identification of the patient (patient's name, sex, date of birth and, increasingly, a unique patient identification number), together with the location, the name of the

Clinical Biochemistry Lecture Notes, Tenth Edition. Peter Rae, Mike Crane and Rebecca Pattenden.
© 2018 John Wiley & Sons Ltd. Published 2018 by John Wiley & Sons Ltd.
Companion website: www.lecturenoteseries.com/clinicalbiochemistry

requesting doctor and the date and time of sampling. Each test request must specify the analyses requested and provide details of the nature of the specimen itself and relevant clinical diagnostic information. This may be through a traditional request form and labelled specimen or be provided electronically in which case only the sample itself need be sent to the laboratory with its own unique identifier (typically a bar code which links it to the electronic request).

Clinical laboratories have multiple procedures at every step of sample processing to avoid errors. Regrettably, errors do occur and these arise at different stages between the sample being taken and the result being received:

- Pre-analytical. These arise prior to the actual test measurement and can happen at the clinical or laboratory end. Most errors fall into this category (see Table 1.1).
- Analytical. Laboratory based analytical errors are rare but may occur, e.g. reagent contamination, pipetting errors related to small sample volumes, computing errors.
- Post-analytical. These are increasingly rare because of electronic download of results from the analyser but include, for example, transcription errors when entering results from another laboratory into the computer manually; results misheard when these are telephoned to the clinician.

Despite the scale of requesting of biochemical tests, errors are fortunately very rare. However, occasional blunders do arise and, if very unexpected results are obtained, it is important that the requesting doctor contacts the laboratory immediately to check whether the results are indeed correct or whether some problem may have arisen. Occasionally this reveals that more than one problem has occurred, for example two samples were labelled with each other's details on the ward, so querying the results can have wider benefits.

The use of clinical biochemistry tests

Biochemical tests are most often *discretionary*, meaning that the test is requested for defined diagnostic purposes. The justification for discretionary testing is well summarised by Asher (1954):

1. Why do I request this test?
2. What will I look for in the result?
3. If I find what I am looking for, will it affect my diagnosis?

Table 1.1 Some more common causes of pre-analytical errors arising from use of the laboratory.

Error	Consequence
Crossover of addressograph labels between patients	This can lead to two patients each with the other's set of results. Where the patient is assigned a completely wrong set of results, it is important to investigate the problem in case there is a second patient with a corresponding wrong set of results.
Timing error	There are many examples where timing is important but not considered. Sending in a blood sample too early after the administration of a drug can lead to misleadingly high values in therapeutic monitoring. Interpretation of some tests (e.g. cortisol) is critically dependent on the time of day when the blood was sampled.
Sample collection tube error	For some tests the nature of the collection tube is critical, which is why the Biochemistry Laboratory specifies this detail. For example, using a plasma tube with lithium–heparin as the anti-coagulant is not appropriate for measurement of a therapeutic lithium level. Electrophoresis requires a serum sample rather than plasma so that fibrinogen does not interfere with the detection of any monoclonal bands. Topping up a biochemistry tube with a haematology (potassium ethylenediamine tetraacetic acid [EDTA]) sample will lead to high potassium and low calcium values in the biochemistry sample.
Sample taken from close to the site of an intravenous (IV) infusion	The blood sample will be diluted so that all the tests will be correspondingly low with the exception of those tests that might reflect the composition of the infusion fluid itself. For example, using normal saline as the infusing fluid would lead to a lowering of all test results, but with sodium and chloride results that are likely to be raised.

4 How will this investigation affect my management of the patient?
5 Will this investigation ultimately benefit the patient?

The main reasons for this type of testing are summarised in Table 1.2. Tests may also be used to help evaluate the future risk of disease (e.g. total cholesterol and HDL-cholesterol levels contribute to assessment of an individual's risk of cardiovascular disease), or in disease prognosis (e.g. biochemical tests to assess prognosis in acute pancreatitis or liver failure), or to screen for a disease, without there being any specific indication of its presence in the individual (e.g. maternal screening for foetal neural tube defects).

Screening may take several forms:

- In well-population screening a spectrum of tests is carried out on individuals from an apparently healthy population in an attempt to detect pre-symptomatic or early disease. It is easy to miss significant abnormalities in the large amount of data provided by the laboratory, even when the abnormalities are highlighted in some way. For these and other reasons, the value of well-population screening has been called into question and certainly should only be initiated under certain specific circumstances (Table 1.3).

Table 1.2 Test selection for the purposes of discretionary testing.

Category	Example
To confirm a diagnosis	Serum free T4 and thyroid-stimulating hormone (TSH) in suspected hyperthyroidism
To aid differential diagnosis	To distinguish between different forms of jaundice
To refine a diagnosis	Use of adrenocorticotrophic hormone (ACTH) to localise Cushing's syndrome
To assess the severity of disease	Serum creatinine or urea in renal disease
To monitor progress	Plasma glucose and serum K^+ to follow treatment of patients with diabetic ketoacidosis (DKA)
To detect complications or side effects	Alanine aminotransferase (ALT) measurements in patients treated with hepatotoxic drugs
To monitor therapy	Serum drug concentrations in patients treated with anti-epileptic drugs

Table 1.3 Requirements for well-population screening.

- The disease is common or life-threatening
- The tests are sensitive and specific
- The tests are readily applied and acceptable to the population to be screened
- Clinical, laboratory and other facilities are available for follow-up
- Economics of screening have been clarified and the implications accepted

Table 1.4 Examples of tests used in case-finding programmes.

Programmes to detect diseases in	Chemical investigations
Neonates	
Phenylketonuria (PKU)	Serum phenylalanine
Hypothyroidism	Serum TSH
Adolescents and young adults	
Substance abuse	Drug screen
Pregnancy	
Diabetes mellitus in the mother	Plasma glucose
Open neural tube defect (NTD) in the foetus	Maternal serum α-foetoprotein
Industry	
Industrial exposure to lead	Blood lead
Industrial exposure to pesticides	Serum cholinesterase activity
Elderly	
Malnutrition	Serum vitamin D levels
Thyroid dysfunction	Serum TSH and thyroxine

- In case-finding screening programmes appropriate tests are carried out on a population sample known to be at high risk of a particular disease. These are inherently more selective and yield a higher proportion of useful results (Table 1.4).

Point of care testing (POCT)

These are tests conducted close to the patient, for example in the emergency department, an outpatient clinic, or a general practitioner's surgery.

Table 1.5 Examples of POCT that are in common use.

Common POCT in blood	Common POCT in urine
Blood gases	Glucose
Glucose	Ketones
Urea and creatinine	Red cells/haemoglobin
Na, K and Ca	Bilirubin
Bilirubin	Protein
Alcohol	hCG

The instrumentation used is typically small and fits on a desk or may even be handheld. This approach can be helpful where there is a need to obtain a result quickly (e.g. blood gas results in the emergency department in a breathless patient), or where a result can be used to make a real-time clinical management decision (e.g. whether to adjust someone's statin dose on the basis of a cholesterol result). A further attraction is the immediate feedback of clinical information to the patient. POCT can be used to monitor illness by the individual patient and help identify if a change in treatment is needed (e.g. blood glucose monitoring in a diabetic patient). There is also an increasing number of urine test sticks that are sold for home use (e.g. pregnancy and ovulation testing by measuring human chorionic gonadotrophin (hCG) and luteinising hormone (LH), respectively). Table 1.5 shows examples of POCT tests in common use.

The introduction of POCT methodology requires attention to cost, ease of use, staff training, quality, health and safety as well as need. The advantages and disadvantages of POCT are summarised in Table 1.6.

Table 1.6 Advantages and disadvantages of point-of-care testing (POCT).

Advantages	Disadvantages
Rapid results on acutely ill patients	More expensive than centralised tests
Allows more frequent monitoring	Wide staff training may be needed
Immediate patient feedback	Nontrained users may have access, with potential for errors
Available 24 h if required	Calibration and quality control may be less robust
	Health and Safety may be less well monitored
	Results less often integrated into patient electronic record

Interpretation of clinical biochemistry tests

Most reports issued by clinical biochemistry laboratories contain numerical measures of concentration or activity, expressed in the appropriate units. Typically, the result is interpreted in relation to a reference range (see Chapter 1: Reference ranges) for the analyte in question. Results within and outside the reference range may be subject to variation caused by a number of factors. These include analytical variation, normal biological variation, and the influence of pathological processes.

Sources of variation in test results

Analytical sources of variation

Analytical results are subject to error, no matter how good the laboratory and no matter how skilled the analyst. The words "accuracy" and "precision" have carefully defined meanings in this context.

An *accurate* method will, on average, yield results close to the true value of what is being measured. It has no systematic bias. Lack of accuracy means that results will always tend to be either high or low.

A *precise* method yields results that are close to one another (but not necessarily close to the true value) on repeated analysis. If multiple measurements are made on one specimen, the spread of results will be small for a precise method and large for an imprecise one. Lack of precision means that results may be scattered, and unpredictably high or low.

A 'dartboard' analogy is often used to illustrate the different meanings of the terms accuracy and precision, and this is illustrated in Figure 1.1.

The standard deviation (SD) is the usual measure of scatter around a mean value. If the spread of results is wide, the SD is large, whereas if the spread is narrow, the SD is small. For data that have a Gaussian distribution, as is nearly always the case for analytical errors, the shape of the curve (Figure 1.2) is completely defined by the mean and the SD, and these characteristics are such that:

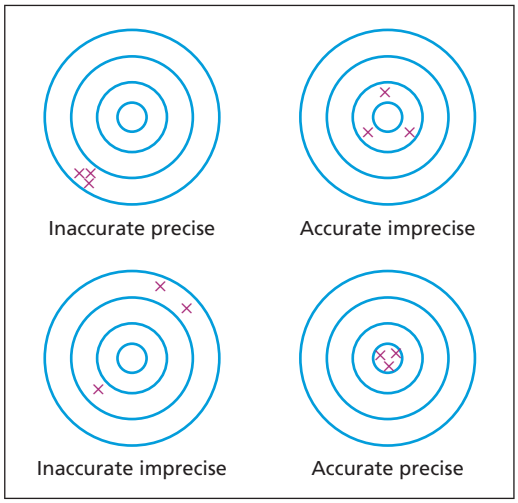

Figure 1.1 The 'dartboard' analogy can be used to illustrate accuracy and precision.

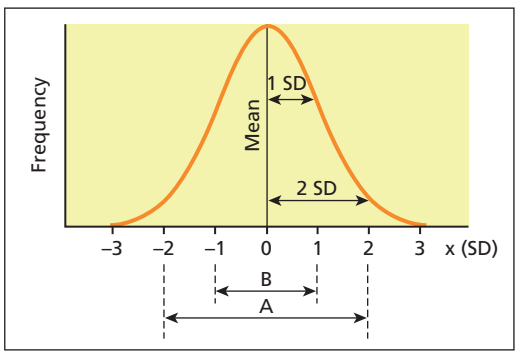

Figure 1.2 Diagram of a Gaussian (normal or symmetrical) distribution curve. The span (A) of the curve, the distance between the mean ± 2 SD, includes about 95% of the 'population'. The narrower span (B), the distance between the mean ± 1 SD, includes about 67% of the 'population'.

- About 67% of results lie in the range mean ± 1 SD.
- About 95% of results lie in the range mean ± 2 SD.
- Over 99% of results lie in the range mean ± 3 SD.

Blunders are grossly inaccurate results that bear no constant or predictable relationship to the true value. They arise, for instance, from mislabelling of specimens at the time of collection, or transcription errors when preparing or issuing reports (see Table 1.1).

If different results for the same test are obtained on two or more occasions on the same patient, then an important question that arises is whether that difference is due to analytical imprecision or to a true change in the patient's clinical condition. Statistically, if the results differ by more than 2.8 times the analytical SD then there is a chance of over 95% that a genuine change in concentration of the substance has occurred.

Biological causes of variation

Test results also show biological variation in both health and disease. The concentrations of all analytes in blood vary with time due to diverse physiological factors *within* the individual. There are also differences *between* individuals.

The following may be important causes of within-individual variation:

1. *Diet:* Variations in diet can affect the results of many tests, including serum triglyceride, the response to glucose tolerance tests and urinary calcium excretion.
2. *Time of day:* Several plasma constituents show diurnal variation (variation with the time of day), or a sleep/wake cycle. Examples include iron, adrenocorticotrophic hormone (ACTH) and cortisol concentrations.
3. *Posture:* Proteins and all protein-bound constituents of plasma show modest differences in concentration between blood collected from upright individuals and blood from recumbent individuals. Examples include serum calcium, cholesterol, cortisol and total thyroxine concentrations.
4. *Muscular exercise:* Recent exercise, especially if vigorous or unaccustomed, may increase serum creatine kinase (CK) activity and blood lactate, and lower blood pyruvate.
5. *Menstrual cycle:* Several substances show variation with the phase of the cycle. Examples include serum iron, and the serum concentrations of the pituitary gonadotrophins, ovarian steroids and their metabolites, as well as the amounts of these hormones and their metabolites excreted in the urine.
6. *Drugs:* These can have marked effects on chemical results. Attention should be drawn particularly to the many effects of oestrogen-containing oral contraceptives on serum constituents (Chapter 10: Steroid contraceptives).

Even after allowing for known physiological factors that may affect plasma constituents and for analytical imprecision, there is still considerable residual individual

CASE 1.1

A 52-year-old man taking a statin drug to reduce his cholesterol level attended for a routine follow-up appointment. He was well, and had recently started training for a half-marathon as part of his determination to get fitter and reduce his risk of future cardiovascular problems. Statins sometimes cause muscle side effects, so among his other blood tests a creatine kinase (CK) was checked and it was very high. Should his statin be stopped?

Comments: Muscular exertion, especially if unaccustomed or severe, can give rise to high CK results. He was asked to refrain from training for a few days, and on repeat a CK level was normal.

variation (Table 1.7). The magnitude of this variation depends on the analyte, but it may be large and must be taken into account when interpreting successive values from a patient.

Differences between individuals can affect the concentrations of analytes in the blood. The following are the main examples:

1. *Age:* Examples include serum phosphate and alkaline phosphatase (ALP) activity, and serum and urinary concentrations of the gonadotrophins and sex hormones.
2. *Sex:* Examples include serum creatinine, iron and urate concentrations, and serum and urinary concentrations of the sex hormones.
3. *Race:* Racial differences have been described for serum cholesterol and protein. It may be difficult to distinguish racial from environmental factors, such as diet.

Reference ranges

When looking at results, we need to compare each result with a set of results from a particular defined (or reference) population. This reference range is determined, in practice, by measuring a set of reference values from a sample of that population, usually of healthy individuals. The nature of the reference population should be given whenever reference ranges are quoted, although a healthy population is usually assumed. Even age-matched and sex-matched reference ranges are often difficult to obtain, since fairly large numbers of individuals are needed.

When results of analyses for a reference population are analysed, they are invariably found to cluster around a central value, with a distribution that may be symmetrical (often Gaussian, Figure 1.3a) or asymmetrical (often log-Gaussian, Figure 1.3b). However, reference ranges can be calculated from these data without making any assumptions about the distribution of the data, using nonparametric methods.

Because of geographical, racial and other biological sources of variation between individuals, as well as differences in analytical methods, each laboratory should ideally define and publish its own reference ranges. By convention, these encompass the central 95% of the results obtained for each analysis from the reference population.

Analytical factors can affect the reference ranges for individual laboratories. If an inaccurate method is used, the reference range will reflect the method bias. If an imprecise method is used, the reference range will be widened, that is, the observed span of results (reflected in the SD) will be greater. In statistical terms, the observed variance (i.e. the square of the SD) of the population results will equal the sum of the true or biological variance of the population plus the analytical variance of the method.

How do results vary in disease?

Biochemical test results do not exist in isolation, and when laboratory tests are requested, the clinician will often have made a list of differential diagnoses based on the patient's history, symptoms and signs, and may have a provisional diagnosis that is the likeliest possibility from this list

Table 1.7 Residual individual variation of some serum constituents (expressed as the approximated day-to-day, within-individual coefficient of variation). CV = coefficient of variation.

Serum constituent	CV (%)	Serum constituent	CV (%)
Sodium	1	ALT activity	25
Calcium	1–2	AST activity	25
Potassium	5	Iron	25
Urea	10		

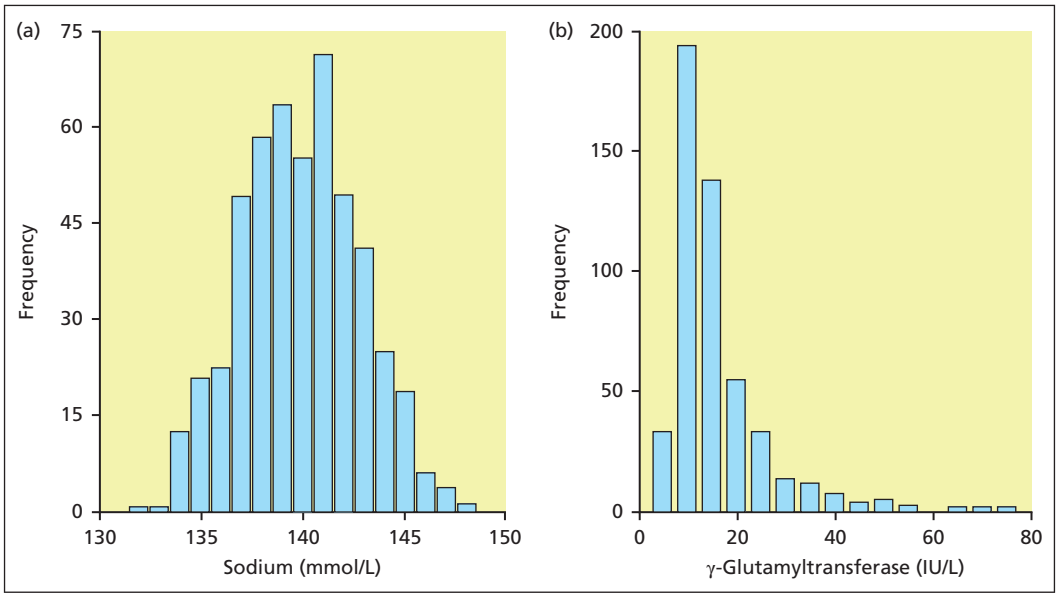

Figure 1.3 Histograms showing the relative frequency with which results with the values indicated were obtained when serum Na^+ and γ-glutamyltransferase (GGT) activities were measured in a reference population of healthy adult women. (a) The sodium data are symmetrically distributed about the mean whereas (b) the GGT data show a log-Gaussian distribution.

For example, in a patient with severe abdominal pain, tenderness and rigidity, there may be several diagnoses to consider including acute pancreatitis, perforated peptic ulcer and acute cholecystitis. In all three conditions, the serum amylase activity may be raised above the upper reference value for healthy adults. We need to know how the serum amylase activity might vary in the clinically likely diagnoses. It would be useful to know, for instance, that very high serum amylase activities are more likely to be associated with one of these diagnostic possibilities (pancreatitis), than with the other two.

To summarise, to interpret results on patients adequately, we need to know:

- the reference range for healthy individuals of the appropriate age range and sex;
- the values to be expected for patients with the disease, or diseases, under consideration;
- the prevalence of the disease, or diseases, in the population to which the patient belongs.

The assessment of diagnostic tests – sensitivity and specificity

In evaluating and interpreting a test, it is necessary to know how it behaves in health and disease. Central to this understanding are the terms "sensitivity" and "specificity." These define how well a test performs in the diagnosis of a disease, but in order to calculate them it is necessary to know whether the disease is present or not by some method (this could be some other definitive test, or may be a diagnosis made later once the clinical course has made this more obvious).

- Test sensitivity refers to how effective the test is in detecting individuals who have the disease in question. It is expressed as the percentage of true positives in all the individuals who have disease (all the individuals with disease will encompass the true positives (TP) and false negatives (FN)). So:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \times 100\%.$$

- Test specificity is a measure of how good the test is at providing a negative result in the absence of disease. It is expressed as the percentage of true negatives in all those without the disease (all the individuals without disease will encompass the true negatives (TN) and the false positives (FP)). So:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \times 100\%.$$

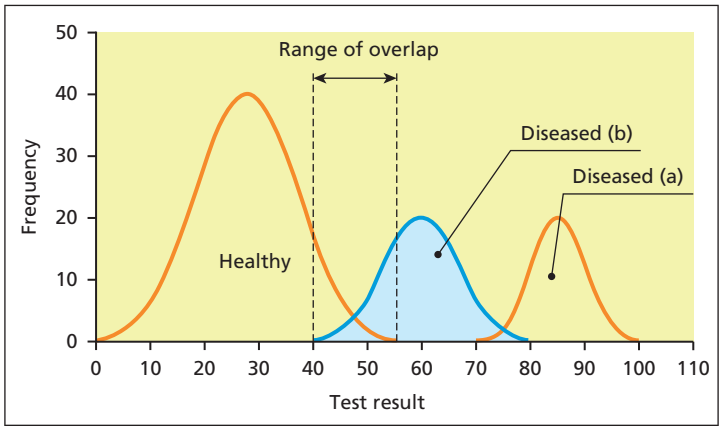

Figure 1.4 Diagrammatic representation of the distributions of results obtained with a test (a) that completely separates healthy people from people with a disease without any overlap between the distribution curves (i.e. an ideal test with 100% sensitivity and 100% specificity), and a test (b) that is less sensitive and less specific, in which there is an area of overlap between the distribution curves for healthy people and people with disease.

The ideal test is 100% sensitive (positive in all patients with the disease) and 100% specific (negative in all patients without the disease). We can illustrate this by means of the following hypothetical example shown diagrammatically in Figure 1.4a. This ideal is rarely achieved; there is usually overlap between the healthy and diseased populations (Figure 1.4b). In practice, we have to decide where to draw dividing lines that most effectively separate 'healthy' from 'diseased' groups, or disease A from disease B.

In evaluating tests for decision making, it is important to decide on the relative importance of sensitivity versus specificity in the context for which a test is used, and to compare the performance of different tests. In defining the presence or absence of a disease, a cut-off may be assigned to a test. Consider the situation where a high value for a particular test equates with the presence of a particular disease. A value above the cut-off would then define the presence of the disease and a value below the cut-off, the absence of disease. A cut-off set at a higher level will increase the test specificity at the expense of test sensitivity (more false negatives), while a cut-off set at a lower value will increase test sensitivity at the expense of test specificity (more false positives).

One way of comparing different tests and of visualising the balance between sensitivity and specificity at different test cut-offs is to plot the test sensitivity against specificity in a so-called 'receiver operating characteristic' (ROC) curve.

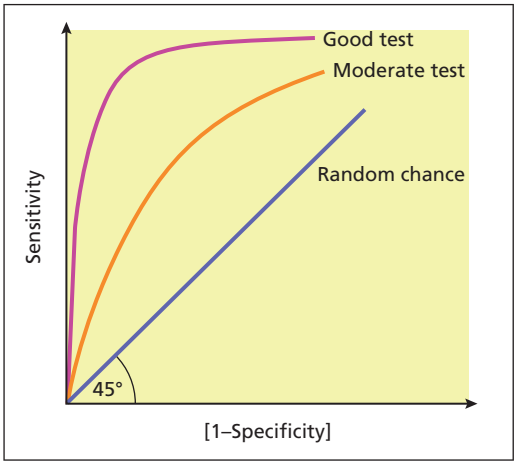

Figure 1.5 Schematic representation of a receiver operating characteristic (ROC) plot. A random test produces a straight line set at 45° to the axes. A discriminatory, good test produces a graph with a steep slope from the origin, displaying high sensitivity at high specificity. Less discriminatory tests produce curves at intermediate positions, as shown. (Adapted from: Roulston, J.E. and Leonard, R.F.C. (1993). Serological Tumour Markers: An Introduction. Reproduced with permission from Elsevier.)

These curves will highlight which test is best suited to which requirement and will also help to define which cut-off to select in order to balance specificity versus sensitivity. This is illustrated in Figure 1.5.

The assessment of diagnostic tests – positive and negative predictive values

The effectiveness of a test can also be defined in terms of the predictive value of a positive result and the predictive value of a negative result. These reflect the reality of clinical practice more than sensitivity and specificity do, in that the presence of a positive or negative test provides information about how likely a disease is to be present or not.

The positive predictive value is the proportion of the positive results that are true positives:

$$TP/(TP+FP)\times 100\%.$$

A test with a high positive predictive value will, by definition, have few false positives. This would be important in a situation where a high number of false positives would otherwise lead to extensive and costly further investigation.

The negative predictive value is the proportion of the negative results that are true negatives:

$$TN/(TN+FN)\times 100\%.$$

A test with a high negative predictive value would, by definition, have few false negatives. This would be particularly important, for example, in a test that was used for a screening programme where it is essential not to miss a case of the disease in question.

What is not immediately intuitive is that the predictive values depend not just on the sensitivity and specificity of a test but on the prevalence of the condition being tested for in the population of patients being tested.

To illustrate this, first consider screening neonates for an inherited metabolic disorder with an incidence of 1:5000; this is similar to that of some of the more common, treatable, inherited metabolic diseases such as phenylketonuria (PKU) or congenital hypothyroidism. Assume that we have a test with excellent performance, defined by sensitivity and specificity each of 99.5%, and that 1 000 000 neonates are tested (Table 1.8). The known incidence of the condition means that there will be 200 affected babies: 199 of these 200 babies will have a positive result (this is the sensitivity of 99.5%); 999 800 of the babies screened will not have the condition, and of these 994 801 will have a negative result (this is the specificity of 99.5%). However, there will be 4999 babies without the condition who receive a positive result, despite the apparently excellent specificity of the test, because the test has been performed on so many babies. This means that the predictive value of a positive result is as low as 3.8%, while the predictive value of a negative result is virtually 100%. For every neonate affected by the disorder who has a positive test result, there will be about 25 (4999/199) neonates who also have a positive test but who do not have the disease.

Table 1.8 A hypothetical set of results of a screening test for a relatively common inherited metabolic disorder in neonates.

Diagnostic category	Positive results	Negative results	Total
Disease present	199	1	200
Disease absent	4999	994 801	999 800
Total	5198	994 802	1 000 000
Predictive value	3.8%	100%	

Assumptions: sensitivity of the test 99.5%; specificity 99.5%; prevalence of the disorder 1:5000; 1 000 000 neonates screened. Note that the prevalence of PKU and of hypothyroidism in the UK is about 1:5000 live births, and that about 800 000 neonates in the UK are screened annually.

Table 1.9 A hypothetical set of results of a test for a myocardial infarction.

Diagnostic category	Positive results	Negative results	Total
Disease present	225	25	250
Disease absent	75	675	750
Total	300	700	1000
Predictive value	75%	96%	

Assumptions: sensitivity of the test 90%, specificity 90%, prevalence of the disorder 1:4; 1000 patients tested.

Next consider a new test for myocardial infarction (MI) that has been shown to have slightly lower (but still acceptable) sensitivity and specificity of 95%. This is used to test patients presenting at an emergency department with severe central chest pain and it has previously been established in this department that 25% of such patients have suffered an MI (Table 1.9). The known incidence of MI in this population means that there will be 250 affected patients: 225 of these 250 patients will have a positive result (this is the sensitivity of 95%); 750 of the patients tested will not have had an MI and will be suffering from some other cause of chest pain, and of these 675 will have a negative result (this is the specificity of 95%). There will be 75 patients without an MI who receive a positive

result, meaning that the predictive value of a positive result is 75% (225 out of 300 patients), while the predictive value of a negative result is 96% (675 out of 700 patients). A negative result has virtually excluded the possibility of an MI, while a positive result has made an MI more than likely.

The second of these examples has shown a much higher positive predictive value (75%) than the first (3.8%) despite the lower sensitivity of the test used. This is because the condition being tested for was much more likely to be present. This provides an important lesson about how laboratory investigations can be used to make diagnoses, showing that tests perform better when a diagnosis is at least a likely possibility, and less well when a test is performed speculatively looking for an unlikely diagnosis.

Two important points regarding screening tests (those used to look for a condition in a basically healthy population) follow on from this. First, tests with very high sensitivity and with very low false-positive rates are required. Secondly, a heavy investigative load will result from the screening programme, since all the false positives will have to be followed up to determine whether or not they indicate the presence of disease. The traditional 95% reference range is not relevant to screening for rare conditions, since the rate of false positives would be far too high. The cut-off value has to be altered to decrease the false-positive rate, at the probable expense of missing some patients who have the condition for which screening is being carried out.

CASE 1.2

A new test is marketed which claims to diagnose heart failure. The test has a specificity of 70% and a sensitivity of 95% at the manufacturer's recommended cut-off for diagnosis. The Admissions Unit decides to use the test as part of an admission profile on breathless patients over the age of 65 years admitted for further assessment in order to exclude heart failure. Assuming a prevalence of 20% for heart failure in this population, calculate how many false negatives would be recorded after the first 1000 patients meeting the testing criteria had passed through the unit. Given that other tests can be used to establish a diagnosis of heart failure, do you think that the cut-off selected is sensible? (Prevalence figures are for illustrative purposes only.)

Comment: This is best examined by constructing a table as follows:

	Positive results	Negative results	Totals
Heart failure present	190 TP	10 FN	200
Heart failure absent	240 FP	560 TN	800
Total	430	570	1000

Because the test has a relatively high sensitivity, the table shows that it identifies the majority of patients with heart failure which is what is required in a test to rule out heart failure. Because the test lacks specificity, it can also be seen from the table that it identifies a considerable number of patients with positive results who do not have heart failure. In fact, the test is positive on more occasions in patients who do not have heart failure than in those with heart failure. Because other tests are available to the clinician, the false-positive patients can be separated from the true-positive patients on the basis of these further investigations. The 560 patients where the result is a true negative would then not need to go through more expensive further investigations. In this example, the test has been valuable in ruling out patients who would not require further investigation but ruling in those who would benefit. Clearly, it is not a perfect test but it would potentially prevent costly further investigations in a significant number of patients and, provided that the test itself is not too expensive, ultimately be worthy of consideration in terms of health economics.

Clinical audit in laboratory medicine

Clinical audit is a process that monitors the procedures involved in patient care to improve the delivery of the service. The principles are applicable to all clinical and investigational specialties (e.g. radiology), as well as laboratory-based specialties such as clinical biochemistry. For example, the monitoring of laboratory performance may identify a delay in analysing samples from the emergency department. This would precipitate a review of the way tests are requested, how samples are delivered to the department, the possible need for these samples to be prioritised in some way, and the way results are communicated back to the clinicians. Any necessary changes would

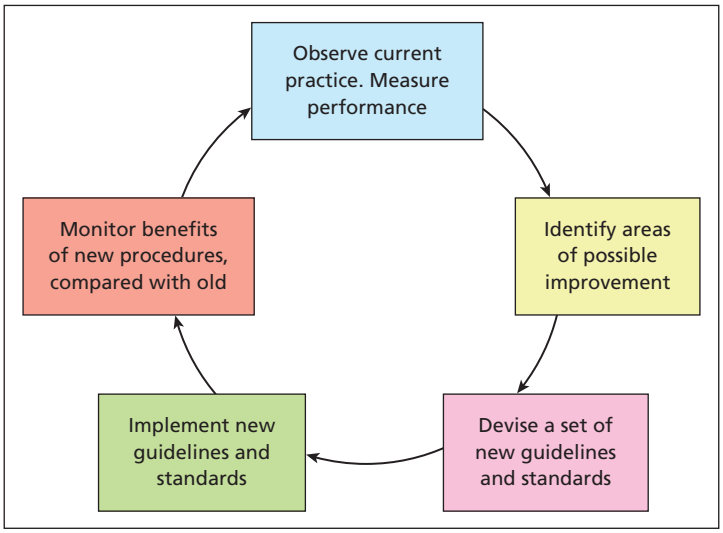

Figure 1.6 The audit cycle.

be instituted, and the process re-monitored to ensure that the original problem had been overcome.

The audit cycle

There is an essential sequence to auditing activities (Figure 1.6):

1. Identify an area of concern or interest, particularly if it is felt that there is room for improvement in the service, or if the same quality of service can be provided more economically.
2. Review and analyse the present procedures.
3. Identify specific aspects that might be capable of improvement.
4. Identify alternative procedures or standards that might lead to improvement.
5. Take the practical steps necessary to implement any changes proposed.
6. Compare the performance after the changes with that before them.

It must be emphasised that the final stage of analysis of the effects of any change is an integral part of the audit process; it is essential to know whether the measures taken have improved the service or made it more cost-effective. Sometimes, changes have no effect, or even have adverse effects.

FURTHER READING

Asher, R. (1954) Straight and crooked thinking in medicine. *British Medical Journal* **2**, 460–2.