1

# Biochemical investigations in clinical medicine

#### INTRODUCTION

A central function of the chemical pathology or clinical chemistry laboratory is to provide biochemical information for the management of patients. Such information will be of value only if it is accurate and relevant, and if its significance is appreciated by the clinician so that it can be used appropriately to guide clinical decisionmaking. This chapter examines how biochemical data are acquired and how they should be used.

#### USE OF BIOCHEMICAL INVESTIGATIONS

Biochemical investigations are used extensively in medicine, both in relation to diseases that have an obvious metabolic basis (e.g. diabetes mellitus, hypothyroidism) and those in which biochemical changes are a consequence of the disease (e.g. renal failure, malabsorption). The principal uses of biochemical investigations are for diagnosis, prognosis, monitoring and screening (Fig. 1.1).

#### Diagnosis

( )

Medical diagnosis is based on the patient's history, if available, the clinical signs found on examination, the results of investigations and sometimes, retrospectively, on the response to treatment. Frequently, a confident diagnosis can be made on the basis of the history combined with the findings on examination. Failing this, it is usually possible to formulate a differential diagnosis, in effect a short-list of possible diagnoses. Biochemical and other investigations can then be used to distinguish between them.

Investigations may be selected to help either confirm or refute a diagnosis, and it is important that the clinician appreciates how useful the chosen investigations are for these purposes. Making a diagnosis, even if incomplete, such as a diagnosis of hypoglycaemia without knowing its cause, may allow treatment to be initiated.

# **Prognosis**

Investigations used primarily for diagnosis may also provide prognostic information, while others are used specifically for this purpose. For example, serial measurements of plasma creatinine concentration in progressive renal disease are used to indicate when dialysis may be required. Investigations can also indicate the risk of developing a particular condition. For example, the risk of coronary artery disease increases with increasing plasma cholesterol concentration. However, such risks are calculated from epidemiological data and cannot give a precise prediction for a particular individual.

# Monitoring

A major use of biochemical investigations is to follow the course of an illness and to monitor the effects of treatment. To do this, there must be a suitable analyte, for instance glycated haemoglobin in patients with diabetes mellitus. Biochemical investigations can also be used to detect complications of treatment, such as hypokalaemia during treatment with diuretics, and are extensively used to screen for possible drug toxicity, particularly in trials, but also in some cases when a drug is in established use.

#### Screening

Biochemical investigations are widely used to determine whether a condition is present subclinically. The bestknown example is the mass screening of all newborn babies for phenylketonuria (PKU), which is carried out in many countries, including the UK and the USA. This is an example of **population screening**: other types include **selective screening** (e.g. of older people for carcinoma of colon using the detection of faecal occult blood); **individual screening** (e.g. as part of a 'health check-up') and **opportunistic screening** (e.g. for hypercholesterolaemia in people found to have hypertension). The use of the 'biochemical profile', a battery of



**Figure 1.1** The principal functions of biochemical tests.

biochemical tests usually performed on a multichannel auto-analyser, is discussed later in this chapter.

#### **SPECIMEN COLLECTION**

#### The test request

۲

The specimen for analysis must be collected and transported to the laboratory according to a specified procedure if the data are to be of clinical value. This procedure begins with the clinician making a test request, either on a computer terminal or on paper. The completed request should include:

- patient's name, sex and date of birth
- hospital or other identification number
- ward/clinic/address
- name of requesting doctor (telephone/pager number for urgent requests)
- clinical diagnosis/problem
- test(s) requested
- type of specimen
- date and time of sampling
- relevant treatment (e.g. drugs).

It is essential that sufficient information be provided to identify the patient. In practice, vital information is often omitted and this may either cause delay in analysis and reporting or make it impossible to interpret the results.

Relevant clinical information and details of treatment, especially with drugs, are necessary to allow laboratory staff to assess the results in their clinical context. Drugs may interfere with analytical methods *in vitro* or

Factor	Example of variable affected
age	alkaline phosphatase, urate
sex	gonadal steroids
ethnicity	creatine kinase
pregnancy	urea
posture	proteins
exercise	creatine kinase
stress	prolactin
nutritional status	glucose
time	cortisol
drugs	triglycerides (alcohol) γglutamyl transferase (phenytoin)

**Figure 1.2** Examples of important factors that influence biochemical variables; these and others are discussed elsewhere in this book.

may cause changes *in vivo* that suggest a pathological process; for instance, some psychotrophic drugs increase plasma prolactin concentration.

#### The patient

Some analytes are affected by variables such as posture, time of day etc. and it may be necessary to standardize the conditions under which the specimen is obtained. Factors of importance in this respect are listed in Figure 1.2 and are discussed further in subsequent chapters.

Even when standardized conditions are used for sampling, the results of repeated quantitative tests (e.g. daily measurements of fasting blood glucose concentration) will themselves show a Gaussian distribution, clustering about the 'usual' value for the individual. Typically, the scatter, which can be assessed by determining the standard deviation (SD), is less for analytes subject to strict regulation (e.g. fasting blood glucose and plasma calcium concentrations) than for others (e.g. plasma enzyme activities). This **biological variation** can be

expressed as the coefficient of variation (CV) for repeated tests where  $CV = SD \times 100/mean$  value.

#### The specimen

The specimen provided must be appropriate for the test requested. Most biochemical analyses are made on serum or plasma, but occasionally whole blood is required (e.g. for 'blood gases'), and analyses of urine, cerebrospinal fluid, pleural fluid etc. can also be valuable. For most analyses on serum or plasma, either fluid is acceptable but in some instances it is of critical importance which of these is used; for example, serum is required for protein electrophoresis and plasma for measurement of renin activity. Haemolysis must be avoided when blood is drawn and, if the patient is receiving intravenous therapy, blood must be drawn from a remote site (e.g. the opposite arm) to avoid contamination. Haemolysis causes increases in plasma potassium and phosphate concentrations and aspartate aminotransferase activity, owing to leakage from red cells. If haemolysis is a consequence of a delay in centrifugation to separate blood cells from plasma, glucose concentration can fall. Other analytes may also be affected by haemolysis, depending on the analytical method used. The laboratory should always draw attention to potentially spurious results. It should be noted that leakage from cells in vitro can cause increases in plasma potassium and phosphate concentrations even in the absence of obvious haemolysis, particularly in patients with high white blood cell or platelet counts.

Collecting a blood specimen into the wrong container can lead to (usually obviously) erroneous results (Case History 1.1): citrate and EDTA, which are used as anticoagulants in containers used for some haematological tests, combine with calcium and cause low measured concentrations in the plasma; so does oxalate (the anticoagulant in containers for blood glucose measurement, which also contain fluoride to inhibit glycolysis), and it is clearly inappropriate to collect blood for lithium measurement into a container with lithium heparin as an anticoagulant. Laboratory handbooks should provide clear guidance on the types of specimen, and, where appropriate, the sampling conditions, for all laboratory tests. This should include guidance on the sequence in which individual specimen tubes are filled to avoid any possibility of contamination; for example, blood should be collected into 'plain tubes' (not containing an anticoagulant or other additive) before being collected into a tube containing, for example, EDTA.

# Case history 1.1

The laboratory staff were concerned when a serum specimen from an outpatient due to attend the diabetic clinic was analysed and the following results were found:

#### Investigations

Serum:	potassium	12.2 mmol/L
	sodium	140 mmol/L
	creatinine	84µmol/L
	calcium	0.34 mmol/L
	phosphate	1.22 mmol/L

#### Comment

The potassium and calcium concentrations are not compatible with life. Investigation disclosed that a locum phlebotomist, who had taken the blood, had collected the original specimen into a tube containing (potassium) fluoride and oxalate, the correct container for an accurate blood glucose measurement, but then compounded his error by transferring the sample to a plain tube. Citrate acts as an anticoagulant by binding to calcium ions to form insoluble calcium oxalate.

All specimens must be correctly labelled and transported to the laboratory without delay. There should be a written protocol for discarding incorrectly collected or labelled specimens. The serum or plasma is then separated from blood cells and analysed. When analysis is delayed, or when specimens are sent to distant laboratories for analysis, degradation of labile analytes must be prevented by refrigerating or freezing the serum or plasma.

Equal care is needed with the collection and transportation of other specimens, such as urine and spinal fluid. All specimens should be regarded as potentially infectious and handled using appropriate precautions.

#### **Urgent requests**

Although laboratories should endeavour to generate results as quickly as possible, some requests will be urgent in that their results may have an immediate bearing on the management of the patient. Examples

include the measurement of serum paracetamol concentration in a patient who has taken a drug overdose, measurement of serum troponin concentration in a patient with chest pain, and measurement of serum potassium concentration in a patient with acute renal failure. Special provision must be made for such samples to be 'fast-tracked' through the analytical process, albeit in full accordance with procedures to ensure quality, and the results reported to the requesting clinician as soon has they have been validated.

#### **Repeat requesting**

When biochemical investigations are being used to monitor the progress of a patient's condition, serial analyses will be required, and the question arises how frequently these should be performed. This will depend on both physiological and pathological factors. For example, in patients being treated with thyroxine for hypothyroidism, it can take several weeks for the plasma concentration of thyroid stimulating hormone (TSH) to stabilize at a new value after a change in the dose of thyroxine: repeating thyroid function tests in a patient whose dose of thyroxine has been changed at an interval of <1 month may therefore provide misleading information, and could prompt a doctor who is not cognizant with the rate of response to make a further change of dose prematurely. In contrast, plasma glucose and potassium concentrations can change very rapidly in patients being treated for diabetic ketoacidosis, and it may be appropriate to make measurements as frequently as every 1–2h, at least initially.

# SAMPLE ANALYSIS AND REPORTING OF RESULTS

# Analysis

The ideal analytical method is accurate, precise, sensitive and specific. It gives a correct result (accurate: Fig. 1.3) that is the same if repeated (precise: Fig. 1.3). It measures low concentrations of the analyte (sensitive) and is not subject to interference by other substances (specific). In addition, it should preferably be cheap, simple and quick to perform. In practice, no test is ideal, but the pathologist must ensure that the results are sufficiently reliable to be clinically useful. Laboratory staff make considerable efforts to achieve this and analytical methods are subject to rigorous quality control and quality assurance procedures.



**Figure 1.3** Precision and accuracy of biochemical tests. Both graphs show the distribution of results for repeated analysis of the same sample by different methods. *Precision*: the mean value is the same in each case, but the scatter about the mean is less in method A than in method B. Method A is, therefore, more precise. *Accuracy*: both are equally precise, but in method D, the mean value differs from the true value. The mean for method C is equal to the true value. Both methods are equally precise, but method C is more accurate.

۲

۲

Nevertheless, there will always be a potential for some degree of imprecision or **analytical variation** in a result. The extent of this can be assessed by making repeated analyses (using exactly the same method) on the same sample (cf. biological variation, above). The results will cluster about a mean for which the SD can be calculated. The imprecision of the analysis can be expressed as the CV where  $CV = SD \times 100/mean$  result. An understanding of the concepts of both analytical and biological variation is essential to the informed interpretation of laboratory data.

It is important to appreciate that results obtained using different methods may not be interchangeable. When a comparison between two results is being made, the same analytical method should be used on both occasions.

It is often appropriate to perform a group of related tests on a specimen. For example, plasma calcium and phosphate concentrations and alkaline phosphatase activity all provide information that may be useful in the diagnosis of bone disease; several liver 'function' tests may usefully be grouped together. Such groupings are sometimes referred to as 'biochemical profiles'. Many currently available analysers can perform >20 assays simultaneously on a single serum specimen. However, although it may be tempting to perform all the assays on every specimen, this approach generates an enormous amount of information, some of which may be unwanted, ignored or misinterpreted (e.g. an elevated creatine kinase (CK) activity in someone who has recently undertaken severe exercise being construed as evidence of myocardial damage). Worst of all, it may actually prevent the clinician from discerning the important results. Discrete analysis, that is, performing only the necessary tests required to answer the clinical question (e.g. 'Is this patient's jaundice cholestatic or due to hepatocellular disease?'), is to be preferred.

#### **Reporting results**

Once analysis has been completed and the necessary quality control checks made and found to be satisfactory, a report can be issued. Cumulative reports, which show previous as well as current results, allow trends in the data to be picked out at a glance. It may be appropriate to add a comment to a report to assist the clinician with its interpretation. Results that indicate a need for urgent clinical intervention should be communicated to the requesting clinician as a matter of urgency.

# Point of care testing

Not all analyses need to be performed in a central laboratory. Reagent sticks for testing urine at the bedside or in the clinic have long been available. Various substances, including glucose, protein, bilirubin, ketones and nitrites (indicative of urinary tract infection), can be tested for using such sticks.

Testing of blood for analytes, such as glucose, and hydrogen ion and 'blood gases' at point of care has also been available for some time. Indeed, the availability of easily used instruments to measure glucose allows patients with diabetes to monitor their blood glucose concentrations at home. Increasingly, manufacturers are developing instruments that can perform a wide range of tests suitable for use at point of care. Such instruments may allow the more rapid provision of analytical results for patients (e.g. in intensive therapy units) than if samples have to be transported to a central laboratory. It is clearly desirable that such instruments should be capable of providing results that are as robust with regard to accuracy and precision as those provided by the main laboratory. These instruments are designed to be very simple to operate but it is nevertheless essential that individuals using them, who may include nurses and doctors, are properly trained in their use, and adhere to protocols designed to ensure quality. Both the training and quality issues should be supervised by trained laboratory staff.

Some analyses can be performed outside traditional healthcare settings and the results given directly to patients. An example is the measurement of plasma cholesterol concentration in retail pharmacies. Such analyses should be subject to appropriate quality assurance procedures and trained personnel should be available to advise patients on the significance of the results.

#### **SOURCES OF ERROR**

Erroneous results are at best a nuisance; at worst, they have potential for causing considerable harm. Errors can be minimized by scrupulous adherence to robust, agreed protocols at every stage of the testing process: this means a lot more than ensuring that the analysis is performed correctly. Errors can occur at various stages in the process:

• **pre-analytical**, occurring outside the laboratory, e.g. the wrong specimen being collected, mislabelling, incorrect preservation etc.

- **analytical**, occurring within the laboratory, e.g. human or instrumental error
- post-analytical, whereby a correct result is generated but is incorrectly recorded in the patient's record, e.g. because of a transcription error.

Analytical errors can be systematic (also known as bias: different analytical methods may produce results that are higher or lower (it is to be hoped only slightly so) than the definitive or reference method) or random. Many of the few errors that do occur even in good laboratories are detected by quality control procedures, including data-handling software or personal scrutiny of reports by laboratory staff. Some are so bizarre that they are easily recognized for what they are. More subtle ones are more likely to go undetected. Unfortunately, the risk of errors occurring can never be entirely eliminated.

#### **INTERPRETATION OF RESULTS**

When the result of a biochemical test is obtained, the following points must be taken into consideration:

• is it normal?

۲

- is it significantly different from previous results?
- is it consistent with the clinical findings?

#### Is it normal?

The use of the word 'normal' is fraught with difficulty. Statistically, it refers to a distribution of values from repeated measurement of the same quantity and is described by the bell-shaped Gaussian curve (Fig. 1.4). Many biological variables show a Gaussian distribution: the majority of individuals within a population will have a value approximating to the mean for the whole, and the frequency with which any value occurs decreases with increasing distance from the mean.

For some analytes, the distribution of values is skewed; an example is plasma bilirubin concentration. Such data can often be mathematically transformed to a normal distribution: data distributed with a skew to the right of the mean (as is the case with bilirubin) can often be transformed to a normal distribution if replotted on a semi-logarithmic scale.

If the variable being measured has a normal (Gaussian) distribution in a population, statistical theory predicts that approximately 95% of the values in the



**Figure 1.4** Gaussian distribution. The range of the mean  $\pm 2$  standard deviations (SDs) encompasses 95.5% of the total number of test results. The range of the mean  $\pm 3$  SDs encompasses 99.7% of the total number.

population will lie within the range given by the mean  $\pm$  2 SDs (Fig. 1.4); of the remaining 5%, half the values will be higher and half will be lower than the limits of this range.

When establishing the range of values for a particular variable in healthy people, it is conventional to first examine a representative sample of sufficient size to determine whether or not the values fall into a Gaussian distribution. The range (mean  $\pm 2$  SDs) can then be calculated; this, in statistical terms, is the 'normal range'. Several important points arise from this:

- Although it is assumed that the population is healthy, values from 5% of individuals by definition lie outside the normal range. This suggests that, if the measurements were to be made in a group of comparable individuals, 1 in 20 would have a value outside this range
- The specialized statistical use of the word 'normal' does not equate with what is generally meant by the word, that is, 'habitual' or 'usually encountered'
- The statistical 'normal' may not be related to another common use of the word, which is to imply freedom from risk. For example, there is an association between increased risk of coronary heart

disease and plasma cholesterol concentrations even within the normal range as derived from measurements on apparently healthy men.

Thus, the normal range for an analyte, defined and calculated as described, has severe limitations. It only identifies the range of values that can be expected to occur most often in individuals who are comparable with those in the population for whom the range was derived. It is not necessarily normal in terms of being 'ideal', nor is it associated with no risk of having or developing disease. Further, by definition, it will exclude values from some healthy individuals. In all cases, like must be compared with like. When physiological factors affect the concentration of an analyte (Fig. 1.2), an individual's result must be assessed by comparing it with the value expected for comparable healthy people. It may, therefore, be necessary to establish normal ranges for subsets of the population, such as various age groups, or males or females only.

To alleviate the problems associated with the use of the word 'normal', the term 'reference interval' (RI) (often called the reference range) has been widely adopted by laboratory staff, using numerical values (reference limits) generally based on the mean  $\pm$  2 SDs. Results can be compared with the RI without assumptions being made about the meaning of 'normal'. In practice, the term 'normal range' is still in general use outside laboratories. It is used synonymously with 'reference interval' in this book. Reference intervals for some common analytes are given in the Appendix: these are as used in one of the authors' laboratories, and are appropriate for the Case Histories, but may not apply to other laboratories because of differences in methodology and in the characteristics of the population on which the data are based.

In using RIs to assess the significance of a particular result, the individual is being compared with a population. Some analytes show considerable biological variation, but the combined analytical and biological variations will usually be less for an individual than for a population. For example, although the reference interval for plasma creatinine concentration is  $60-120 \mu mol/L$ , the day-to-day variation in an individual is less. Thus, it is possible for a test result to be abnormal for an individual, yet still be within the accepted 'normal range'.

An abnormal result does not always indicate the presence of a pathological process, or a normal result

its absence. However, the more abnormal a result, that is, the greater its difference from the limits of the reference interval, the greater is the probability that it is related to a pathological process.

In practice, there is rarely an absolute demarcation between normal values and those seen in disease: equivocal results must be supported by further investigation. If an important decision in the management of a patient is to be based upon a single result, it is vital that the cut-off point, or 'decision level', is chosen to ensure that the test functions efficiently. In screening for PKU, for example, the blood concentration of phenylalanine selected to indicate a positive result must include all infants with the condition; in other words, there must be no false negatives. This means that some normal children will be test positive (false positives) and will be subjected to further investigation. Generally, it is unusual to have to determine a patient's management on the basis of one result alone.

It has been explained that 5% of healthy people will, by definition, have a value for a given variable that is outside the reference interval. If a second and independent variable is measured, the probability that this result will be 'abnormal' is also 0.05 (5%). However, the abnormal results may not arise in the same individuals and the overall probability of an abnormal result from at least one test will be >5%. It follows that the more tests that are performed on an individual, the greater the probability that the result of one of them will be abnormal: for 10 independent variables, the probability is 0.4; in other words, at least one abnormal result would be expected in 40% of healthy people. For 20 variables, the probability is 0.64.

Although biochemical parameters are frequently, to some extent, interdependent (e.g. albumin and total protein), the use of multichannel auto-analysers to produce biochemical profiles inevitably risks generating a number of spuriously 'abnormal' results. Before any decision can be made on the basis of such results, some information is required about the probability that they are indicative of a pathological process. This topic is discussed on p. 9.

#### Is it different?

If the result of a previous test is available, the clinician will be able to compare the results and decide whether any difference between them is significant. This will depend upon the precision of the assay itself (a measure

Analyte	Analytical variation	Biological variation
sodium	1.1 mmol/L	2.0 mmol/L
potassium	0.1 mmol/L	0.19 mmol/L
bicarbonate	0.5 mmol/L	1.3 mmol/L
urea	0.4 mmol/L	0.85 mmol/L
creatinine	5.0 μmol/L	4.1 μmol/L
calcium	0.04 mmol/L	0.04 mmol/L
phosphate	0.04 mmol/L	0.11 mmol/L
total protein	1.0 g/L	1.66 g/L
albumin	1.0 g/L	1.44 g/L
aspartate transaminase	6.0 U/L	8.0 U/L
alkaline phosphatase	4.0 U/L	15.0 U/L

**Figure 1.5** Analytical and biological variation. Analytical variation: typical standard deviations for repeated measurements made using a multichannel analyser on a single quality control serum with concentrations in the normal range. *Biological* variation: means of standard deviations for repeated measurements made at weekly intervals in a group of healthy subjects over a period of ten weeks, corrected for analytical variation.

of its reproducibility) and the natural biological variation. Some examples of variation in common analytes are given in Figure 1.5.

The probability that the difference between two results is **analytically significant** at a level of p < 0.05 is 2.8 times the **analytical SD**. Thus, for plasma calcium concentration, with an analytical SD of 0.04 mmol/L, an apparent increase in calcium concentration from 2.54 mmol/L to 2.62 mmol/L (2 × SD) is within the limits of expected analytical variation, whereas an increase from 2.54 to 2.70 (4 × SD) is not. However, to decide whether an analytical change is **clinically significant** it is necessary to consider the extent of natural **biological variation**. The effects of analytical and bio-

۲

# Case history 1.2

A GP measured the serum creatinine concentration of a 41-year-old man newly diagnosed as having diabetes mellitus and hypertension. The result was 105μmol/L. Six months later, both conditions were well controlled and the test was repeated.

#### Investigation

Serum creatinine 118µmol/L

The patient was alarmed at the apparent increase, but the GP was uncertain as to whether this was a significant change.

#### Comment

The analytical variation for creatinine is  $5.0 \mu$ mol/L, the biological variation  $4.1 \mu$ mol/L (Fig. 1.5). The critical difference is:

$$2.8 \times \sqrt{4.1^2 + 5.0^2}$$

that is,  $18 \mu \text{mol/L}$ . Thus the apparent increase in creatinine is not significant at a level of p = 0.05.

logical variation can be assessed by calculating the overall standard deviation of the test, given by:

$$SD = \sqrt{SD_A^2 + SD_B^2}$$

where  $SD_A$  and  $SD_B$  are the SDs for the analytical and biological variation, respectively. If the difference between two test results exceeds 2.8 times the SD of the test, the difference can be regarded as of potential clinical significance: the probability of this difference being a result of analytical and biological variation is <0.05 (Case History 1.2). It should be appreciated, however, that setting the level of significance at a probability of <0.05 is arbitrary (albeit conventional). It does not mean that a difference of less than that equating to this probability cannot be of significance, nor that a greater difference necessarily is significant. If undertaking a major intervention depends on a result, it may be desirable only to make this decision if the probability that the change is not the result of innate variation is considerably greater.

# Is it consistent with clinical findings?

If the result is consistent with clinical findings, it is evidence in favour of the clinical diagnosis. If it is not consistent, the explanation must be sought. There may have been a mistake in the collection, labelling or analysis of the sample, or in the reporting of the result. In practice, it may be simplest to request a further sample and to repeat the test. If the result is confirmed, the utility of the test in the clinical context should be considered and the clinical diagnosis itself may have to be reviewed.

#### THE CLINICAL UTILITY OF LABORATORY INVESTIGATIONS

In using the result of a test, it is important to know how reliable the test is and how suitable it is for its intended purpose. Thus, the laboratory personnel must ensure, as far as is practicable, that the data are accurate and precise, and the clinician should appreciate how useful the test is in the context in which it is used. Various properties of a test can be calculated to provide this information.

## Specificity and sensitivity

Earlier in the chapter, the terms 'sensitivity' and 'specificity' were used to describe characteristics of analytical methods. The terms are also widely used in the context of the utility of laboratory tests. The specificity of a test is a measure of the incidence of negative results in persons known to be free of a disease, that is, 'true negative' (TN). Sensitivity is a measure of the incidence of positive results in patients known to have a condition, that is, 'true positive' (TP). A specificity of 90% implies that 10% of disease-free people would be classified as having the disease on the basis of the test result: they would have a 'false positive' (FP) result. A sensitivity of 90% implies that only 90% of people known to have the disease would be diagnosed as having it on the basis of that test alone: 10% would be 'false negatives' (FN).

Specificity and sensitivity are calculated as follows:

Specificity =  $\frac{\text{TN}}{\text{all without disease (FP+TN)}} \times 100$ 

Sensitivity =  $\frac{\text{TP}}{\text{all with disease (TP + FN)}} \times 100$ 

An ideal diagnostic test would be 100% sensitive, giving positive results in all subjects with a particular disease, and also 100% specific, giving negative results in all subjects free of the disease. Because the ranges of results in quantitative tests that can occur in health and in disease almost always show some overlap, individual tests do not achieve such high standards. Factors that increase the specificity of a test tend to decrease the sensitivity and vice versa. To take an extreme example, if it were decided to diagnose thyrotoxicosis only if the plasma free thyroxine concentration were at least 32 pmol/L (the upper limit of the reference range is 26 pmol/L), the test would have effectively 100% specificity; positive results (>32 pmol/L) would only be seen in thyrotoxicosis (an exception is a very rare condition in which patients are resistant to thyroid hormones). On the other hand, the test would have a low sensitivity in that many patients with mild thyrotoxicosis would be misdiagnosed. If a concentration of 20 pmol/L were used, the test would be very sensitive (all those with thyrotoxicosis would be correctly assigned) but have low specificity, because many normal people would also be diagnosed as having thyrotoxicosis. These concepts are illustrated in Figure 1.6.

Whether it is desirable to maximize specificity or sensitivity depends on the nature of the condition that the test is used to diagnose and the consequences of making an incorrect diagnosis. For example, sensitivity is paramount in a screening test for a harmful condition, but the inevitable FP results will have to be investigated further. However, in selecting patients for a trial of a new treatment, a highly specific test is more appropriate to ensure that the treatment is being given only to patients who have a particular condition. In some cases, this decision may not be straightforward, for example in the context of chest pain and suspected acute myocardial infarction, where the possible options are to identify all those who have had a myocardial infarction ('rule in') or to identify all those who have definitely not ('rule out'). The preferred option should depend on the relative outcomes of treatment and non-treatment for patients in the two groups.

One way of comparing the sensitivity and specificity of different tests is to construct 'receiver operating characteristic curves' (ROC curves). Each test is performed in each of a series of appropriate individuals.

۲

( )



**Figure 1.6** Because the ranges of values for a test result in health and disease overlap (a), some patients with disease will have results within the reference range (false negatives), while some individuals free of disease will have results outside this range (false positives). If the diagnostic cut-off value for a test is set too high (b), there will be no false positives, but many false negatives; specificity is increased but sensitivity decreases. If the diagnostic cut-off value is set too low (c), the number of false positives, and sensitivity, increases, at the expense of a decrease in specificity.



**Figure 1.7** ROC curves for three hypothetical tests, A, B and C. Examination of the curves shows that test A performs less well in terms of both sensitivity and specificity than tests B and C. Test B has better specificity than C, but C has better sensitivity.

The specificity and sensitivity are calculated using different cut-off values to determine whether a given result is positive or negative (Fig. 1.7). The curves can then be assessed to determine which test performs best in the specific circumstances for which it is required.

The specialized use of the terms 'sensitivity' and 'specificity' that has been discussed here in the context of the utility of laboratory tests sometimes causes confusion, since these terms are also used to describe purely analytical properties of tests. Readers should appreciate that, in this latter context, 'sensitivity' relates to the ability of a test to detect low concentrations of an analyte and 'specificity' to its ability to measure the analyte of interest and not some other (usually similar) substance.

#### Efficiency

The efficiency of a test is the number of correct results divided by the total number of tests. Thus efficiency is given by:

 $\frac{\text{TP} + \text{TN}}{\text{total number of tests}} \times 100$ 

10

۲

When sensitivity and specificity are equally important, the test with the greatest efficiency should be used.

#### **Predictive values**

Even a highly specific and sensitive test may not necessarily perform well in a clinical context. This is because the ability of a test to diagnose disease depends on the prevalence of the condition in the population being studied (prevalence is the number of people with the condition in relation to the population). This ability is given by the 'predictive value' (PV).  $PV_{+ve'}$  the PV for a positive result, is the percentage of all positive results that are TPs, that is:

$$PV_{+ve} = \frac{TP}{TP + FP} \times 100$$

If a condition has a low prevalence and the test is <100% specific, many FPs will result and the PV will be low.

A high predictive value for a positive test is important if the appropriate management of a patient with a TP result would be potentially dangerous if applied to someone with a FP result. However, when a test is used for screening, the appropriate management is to perform further (diagnostic) tests, and although these may cause inconvenience for subjects with FP results, they are unlikely to be dangerous.

In order not to miss cases, a screening test should have a very high PV<sub>-ve</sub>, the PV for a negative result; this is the percentage of all negative results that are TNs, that is:

$$PV_{-ve} = \frac{TN}{TN + FN} \times 100$$

This conclusion follows directly from the fact that the test must be highly sensitive.

For clarity, this discussion has centred on the use of single tests for diagnostic purposes, but in practice, the clinician will combine clinical information and, often, the results of several investigations to make the diagnosis. If the tests are used rationally, the PV of positive results will be higher since the tests will be used only in patients who have other features suggesting a particular diagnosis (the prevalence of the disease in question would be higher in a group of such people than in the general population). For example, although Cushing's disease is rare, making the PV of a positive test for the condition in the general population low, in practice one would only investigate patients suspected on clinical grounds of having the condition and in whom the prevalence will therefore be higher. This may be self-evident, but doctors frequently order tests on flimsy clinical grounds and fail to appreciate how unhelpful, or even misleading, the results may be.

#### **Likelihood ratios**

The concept of predictive values is an unfamiliar one for many people: it has no obvious parallel in our everyday lives. The concept of odds is a more familiar one. 'Likelihood ratios' (LRs) express the odds that a given finding (e.g. a particular result) would occur in a person with, as opposed to without, a particular condition. The LR for a positive result is given by:

$$LR_{+ve} = sensitivity/(1 - specificity)$$

The  $LR_{ve}$  (the odds that a negative test result would occur in a person with, as opposed to without, a particular condition) is given by:

$$LR_{-ve} = (1 - sensitivity)/specificity$$

LRs can be used to convert the probability of a condition being present before the test was done (in the case of a screening test, this is the prevalence) to the post-test probability of its being present. The greater the value of the LR, the more useful the test will have been.

#### **Evidence-based clinical biochemistry**

Most clinicians and pathologists use laboratory tests primarily on the basis of their own clinical expertise and interpret results intuitively. Ideally, tests should be chosen on the basis of evidence of their utility, and their results used on the basis of outcome measures. Such an approach is advocated as part of the practice of evidence-based medicine, and could be facilitated by the use of test characteristics such as have been discussed above. However, it remains the case that many wellestablished tests have been introduced into clinical practice without being properly evaluated, and few systematic reviews of existing tests have been performed. Furthermore, new tests are often introduced into laboratories' repertoires without a systematic assessment of their utility having been made, and their value and

( )

limitations may only become apparent in the light of experience of their day-to-day use.

#### **CLINICAL AUDIT**

Clinical audit is part of the process of ensuring quality - in this context, of ensuring the provision of a high quality laboratory service. In this respect, it is complementary to the other techniques of quality assurance, which in the main concentrate on the analytical aspects of the service, that is, the provision of precise and accurate results. Clinical audit is the process of systematically examining practice in order to ensure that it is efficient and beneficial to patients. It involves identifying an area of practice, setting standards or guidelines (e.g. a protocol for investigation of patients suspected of having a particular condition), implementing changes designed to achieve these and then examining compliance with them and the effects on patient care. The cycle is completed by revision of the standards in the light of this analysis and their modification as required. It should be followed by re-audit after an appropriate interval.

The term 'audit' is also applied to procedures used by some laboratory accreditation bodies to check on the internal functioning of laboratories. It is beyond the scope of this book to describe such procedures.

#### SCREENING

( )

Screening tests are used to detect disease in groups of apparently healthy individuals. Such tests may be applied to whole populations (e.g. the detection of PKU in the newborn), to groups known to be at risk (the detection of hypercholesterolaemia in the relatives of people with premature coronary heart disease), or to groups of people selected for other reasons (biochemical profiling of pre-operative patients, health screening for business executives and screening for common conditions in the elderly).

As previously discussed, high sensitivity is particularly desirable for screening tests but, to avoid unnecessary further tests of normal people, high specificity is also an important consideration. Screening tests for PKU are designed to maximize sensitivity but are also highly specific. However, PKU has a low incidence so that even with a sensitivity of 100% and specificity of 99.9%, the predictive value of a positive test is only 10%, that is, 9 out of 10 positive tests will be shown on further investigation to be false positives. These calculations are made as follows:

1. incidence of PKU = 1 in 10000 live births

2. sensitivity = 100% or 
$$\frac{111^{\circ}}{1 \text{ case of PKU}}$$
  
9990TN

- 3. specificity = 99.9% or  $\frac{333333}{9999}$  without PKU
- 4. number of positive tests per 10 000 infants tested =  $\frac{(100 99.9)}{1000} \times 10\,000 = 10$
- 5. numbers of TP and FP results: TP = 1, FP = 9
- 6. predictive value of a positive test =  $\frac{1}{10} \times 100 = 10\%$ .

On the other hand, the predictive value of a negative test will be 100%, confirming that no cases will be missed using the screening test.

Screening for specific conditions is discussed in other chapters of this book. Such screening is often based on the use of considerably less specific or sensitive tests and therefore has a low efficiency for detecting disease. Indiscriminate biochemical profiling is also inefficient. The more tests that are performed, the greater is the probability that an apparently abnormal result will arise that is not the result of a pathological process.

When multichannel analysers are used to generate biochemical data and an unexpected abnormality is found, a decision must be made as to what action to take. The abnormality may be considered insignificant in some clinical circumstances, but if it is not, further investigations must be made. Although these may be of ultimate benefit to the patient, their cost and economic consequences may be considerable. At the very least, the tests should be repeated to ensure that the abnormality was not due to analytical error.

The ready availability of an investigation often leads to it being used unnecessarily or inappropriately. Doctors should be encouraged to be selective in making test requests. They should also join with laboratory staff in critically examining all current tests and investigative techniques to ensure that they are using these tests to their best advantage in medical practice.

#### **SUMMARY**

i

۲

- Biochemical investigations are used for diagnosis, monitoring, screening and in prognosis
- **Specimens for analysis** must be collected and transported to the laboratory under appropriate conditions
- Analytical results are affected by both **analytical and biological variation**
- Results can be compared either with **reference intervals** or with the results of previous tests
- The utility of test results depends on many factors: an 'abnormal' result should not be assumed to indicate a pathological process, nor a 'normal' one to exclude disease or potential disease
- The utility of tests can be measured and described mathematically: applying this information can considerably enhance the value of laboratory test results in clinical practice.

# Plasma and serum

Plasma is the aqueous phase of blood and can be obtained by removal of blood cells from blood to which an anticoagulant has been added. Serum is the aqueous phase of blood that has been allowed to clot. For technical reasons, many biochemical measurements are more conveniently made on serum, but the concentrations of most analytes are effectively the same in both fluids. In this book, the term 'serum' is used only where actual measurements made in serum are referred to (e.g. in the Case Histories) and in the few instances where serum must be used for analysis.

