VetBooks.ir

Chapter 2

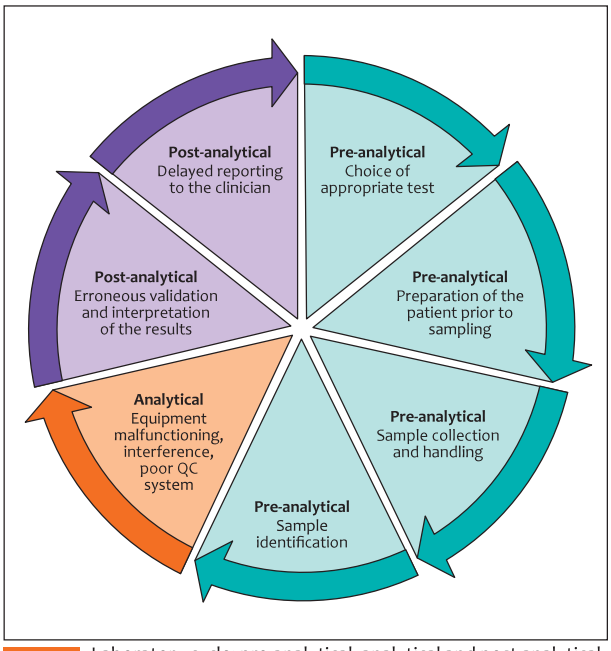
Quality assurance and interpretation of laboratory data

Paola Monti and Joy Archer

Laboratory test results form part of the database from which a clinical diagnosis may be made. History, clinical examination and ancillary tests (laboratory tests, radiographs, etc.) are interpreted in conjunction with each other to obtain the best possible diagnosis. Laboratory testing has an important role in the clinical work-up and monitoring of the therapy of veterinary patients. Hence, the care provided to patients is strongly dependent upon consistent and reliable laboratory data. Laboratory results should not be interpreted in isolation, but with an understanding of the laboratory methods used and the potential errors caused by inappropriate sample collection and handling.

Errors may be introduced into the diagnostic laboratory cycle at three main stages (Figure 2.1):

- **Pre-analytical:** inappropriate test request, patient preparation prior to sample collection or sample collection and handling; sample identification problems
- **Analytical:** equipment malfunction, interference, poor quality reagents and controls, poor quality control (QC) system
- **Post-analytical:** erroneous validation or interpretation of the results, delayed reporting to the clinician (excessive turnaround time).



2.1 Laboratory cycle: pre-analytical, analytical and post-analytical phases with the most common areas where errors can occur. QC = quality control.

In the last two decades, clinical laboratories have focused their attention on QC to minimize the number of errors that occur during the analytical process (analytical errors). This can be pursued by implementing routine internal checks and enrolling in external quality assessment programmes. However, recent surveys in human laboratory medicine have suggested that laboratory errors occur more frequently before or after the test has been performed (pre- and post-analytical errors).

Pre-analytical errors

Most errors affecting laboratory testing occur in the pre-analytical phase. Poor quality or inappropriate samples can lead to the generation of poor quality results. This can cause erroneous clinical interpretation, resulting in poor patient care.

According to the International Organization for Standardization (ISO) 15189 (2007) definition, the pre-analytical phase includes clinician request, preparation of the patient, collection of the sample and transportation to, and handling of, the sample in the laboratory, and ends when the analytical examination begins (Hawkins, 2012). Pre-analytical errors can be sub-classified as follows:

- Preparation of the patient prior to sampling, and patient variables
- Sample collection and handling
- Problems with identification.

Preparation of the patient prior to sampling and patient variables

The most common physiological changes or patient variables that can affect some test results are:

- Exercise or excitement/fear can cause changes in some haematology parameters due to the release of catecholamines. This leads to an increased neutrophil count and sometimes lymphocyte count due to their shift from the marginated to the circulating pool. These changes are referred to as physiological leucocytosis and are often observed in young cats
- Food consumption can affect biochemistry tests, in particular cholesterol, triglyceride, glucose and urea.

VetBooks.ir

Additionally, a postprandial sample may be lipaemic and this can, depending on the analytical method, affect other biochemical tests, especially total protein, albumin (values are elevated) and electrolytes (values are lowered). Unless postprandial samples are required (e.g. for dynamic bile acid measurement) an overnight (12-hour) fast is preferred before general biochemical testing. Lipaemia may also interfere with the spectrophotometric assay for haemoglobin (Hb), resulting in falsely high Hb and mean corpuscular haemoglobin concentration (MCHC). Large lipid droplets may be falsely counted as leucocytes or platelets by some analysers (e.g. Cell-Dyn)

- If samples are to be collected for monitoring drug therapy (e.g. thyroid supplementation, digoxin levels, etc.), collection times can be important and should be timed to correspond with peak and trough drug levels; the times should be carefully recorded. Special tests, such as glucose tolerance tests and hormone stimulation tests, should have protocols defining test substance dosage and times of administration and sample collection. Times should be carefully recorded on the sample containers.

Additionally, there are some variables intrinsic to the patient that, if ignored, could lead to incorrect interpretation of the results. The more common patient variables are breed and age related (see examples later in the text).

Sample collection

Incorrect sample collection and handling can lead to an unsuitable sample for analysis, potentially leading to inaccurate results and an incorrect clinical decision.

The sampling technique is influenced by the testing required. For example, urine for microbiology testing should be collected aseptically by cystocentesis, while for routine urinalysis an uncontaminated voided sample collected into a clean container is often appropriate. Urine from the cage floor is unsuitable for any analysis.

The most common reasons why a sample may not be suitable for analysis are listed below.

- **Incorrect test requested:** when choosing a laboratory test, the clinician should consider the diagnostic accuracy and the predictive value of the test for identifying the suspected disease. For example, if hyperadrenocorticism is clinically suspected, measuring the urine cortisol to creatinine ratio would not be the test of choice because it is poorly specific although highly sensitive. This means that it is a good test to rule out hyperadrenocorticism, but better tests are available to confirm this disease (e.g. the adrenocorticotropic hormone (ACTH) stimulation test).
- **Haemolysed sample:** with blood collection for haematology, biochemistry and special tests, venepuncture should be performed rapidly and as atraumatically as possible to reduce the potential for haemolysis. Unless a vacutainer system is used for blood collection, the needle should be removed from the syringe before the blood is transferred *gently* to the tube to avoid damage to the cells and to minimize haemolysis. The parameters that are more affected by haemolysis are creatine kinase (CK), aspartate aminotransferase (AST), phosphate and total protein, although the effect varies depending on the method being used. The interference occurs because free haemoglobin may absorb at the same wavelength as the coloured product of a reaction, or because the substance being measured is released from lysed red cells. Haemolysis falsely raises MCHC and lowers the packed cell volume (PCV) and the red cell count. In human laboratory medicine, haemolysis is the most common reason for sample rejection. Haemolysis is often caused by delayed sample separation, which can also lead to spurious elevations in potassium due to release from leucocytes and platelets. Blood replacement products prepared from bovine haemoglobin interfere with tests in a similar way to haemolysis (directly in a reaction or with colorimetric methods). The effects are dose dependent and persist for 48 hours or more after administration.
- **Clotted sample (micro and macro clots):** traumatic or delayed blood collection can cause platelet activation and secondary aggregation, leading to a spurious thrombocytopenia. The presence of micro or macro clots may also falsely decrease the white blood cell (WBC) count.
- **Under- or over-filling of blood tubes:** tubes should be filled to the correct volume and gently inverted to mix the blood with the pre-measured contents (e.g. ethylenediamine tetra-acetic acid (EDTA), sodium citrate, lithium heparin). It is important to collect an adequate volume of blood for the tests required, remembering that approximately 50–60% of the volume is plasma/serum. For routine haematology, the anticoagulant of choice is EDTA, potassium or sodium salt, because it preserves cell morphology. If the concentration of EDTA is excessive in relation to blood volume (tube under-filling), cells will shrink and falsely lower the PCV. EDTA tubes less than half full (>3.0 mg EDTA/ml blood) reduce the PCV by 5%. The calculated haematocrit (HCT) is unaffected because the red cells re-expand when they are mixed with the isotonic diluent used by the analyser. If liquid EDTA is used, this can add to the error by diluting the sample, thus further lowering cell counts. Conversely, insufficient EDTA in relation to blood will lead to clot formation. Small clots in the sample, which might be missed when visually inspecting the sample, can cause errors in machine-measured parameters, in particular the platelet count and white cell count. For the measurement of coagulation times, citrated plasma is used. The concentration of citrate in the sample affects the results, and maintaining a citrate to sample ratio of 1:9 is essential for an accurate result. If the tubes are under-filled, coagulation times will be falsely prolonged, while over-filling may lead to falsely shortened times. Sample handling is very important in haemostatic tests and is discussed in Chapter 6.
- **Contamination of the sample:** if a single sample is to be divided between several collecting tubes, it is good practice to collect it into a plain (serum) tube first, followed by tubes containing anticoagulant agents. This is to prevent possible contamination, especially with EDTA, which causes a false increase in potassium and a decrease in calcium, magnesium, CK and alkaline phosphatase (ALP). The Clinical and Laboratory Standards Institute (CLSI) has released a recommended order for collecting blood samples (Figure 2.2).

VetBooks.ir

1. Culture tubes or serum tubes with no additive
2. Citrate tubes
3. Gel separator tubes and clot activator tubes
4. Heparin tubes
5. EDTA tubes
6. Other additive tubes (e.g. fluoride/oxalate)

2.2 Clinical and Laboratory Standards Institute (CLSI) guidelines for sample collection into blood tubes in order to avoid sample contamination. Blood should be placed into sample tubes in this order. EDTA = ethylenediamine tetra-acetic acid.

Sample handling

Once the sample has been collected into the correct tube, it should be processed promptly. For haematology, it is always best to make one or more blood films from an EDTA sample close to the time of collection and air-dry them. Although EDTA preserves cell morphology, changes begin to appear within hours, especially in white cells. Samples should be held in the refrigerator before shipping and/or analysing but blood films should not.

At the pre-analytical stage, the greatest numbers of errors for haematology tests are introduced by the ageing of the sample. For example, after 12 hours from collection, the mean cell volume (MCV) and HCT (calculated from the MCV and red blood cells (RBCs)) can significantly increase, and consequently the MCHC decreases.

Coagulation factors are degraded *in vitro* within hours of sampling. Hence, citrate plasma should be separated within 30 minutes from collection. Prothrombin time (PT) and activated partial thromboplastin time (aPTT) are stable for 48 hours in separated plasma at room temperature, but plasma should be frozen if a longer delay is anticipated (see Chapter 6).

Samples for measurement of ionized calcium and magnesium must also be handled carefully; serum should be separated quickly and stored anaerobically.

Samples for glucose determination need to be separated promptly or placed in fluoride/oxalate. Glucose decreases at a rate of 10% per hour if unseparated samples are held at room temperature. Separation of the sample shortly after collection should be preferred when possible because fluoride/oxalate may induce haemolysis.

Ammonia is another labile analyte that requires special handling and should be analysed immediately after sampling. Within hours at room temperature, ammonia concentration can increase up to 2–3 times.

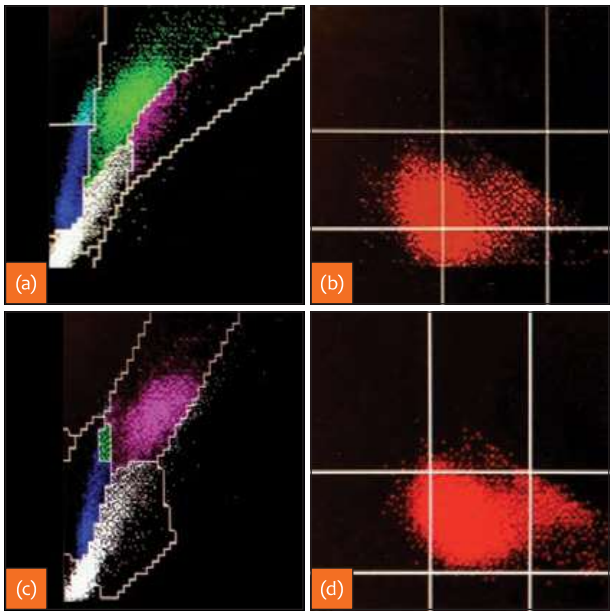
Some endocrinology tests such as those for endogenous ACTH, parathyroid hormone (PTH) and renin require special handling. The samples should be collected in EDTA and the plasma separated immediately and promptly frozen. The sample should then be sent frozen to the reference laboratory.

Identification problems

Examples of problems with the identification of the sample are:

- Specimens not labelled or incorrectly labelled (e.g. blood tubes, cytology slides, etc.)
- Mismatch between the sample's label and the submission form
- Incorrect information provided on the submission form (e.g. incorrect species, breed, age; incomplete or wrong clinical history, etc.; Figure 2.3).

Each sample should be clearly labelled with patient identification and date of collection, and the time of collection if relevant. Along with the samples, there should be a legible submission form which should indicate the tests requested, patient identification (name, number, species, age, breed and sex) and a brief history with clinical findings and information on any drug therapy or blood replacement products given.



2.3 Scatter plots obtained by analysing EDTA blood from a cat with (a–b) canine settings and (c–d) feline settings. An EDTA blood sample from a cat was submitted to a reference laboratory for haematology analysis. This sample was accompanied by a submission form that stated that the animal was a dog. (a–b) The analyser (Advia® 120) scatter plots show the leucocyte and red cell scatter plots, respectively, that were obtained when the sample was analysed with the canine setting. (c–d) Leucocyte and red cell scatter plots obtained when the sample was analysed using the correct feline setting. Using the wrong setting caused an erroneous gating of the erythrocytes and leucocytes, leading to a falsely low mean cell volume (MCV), mean cell haemoglobin concentration (MCHC) and neutrophil count.

Analytical errors

In the last two decades, advances in standardization, automation and technology have significantly decreased analytical errors, thus improving the reliability of laboratory results. Statistical QC activities have been introduced into the diagnostic laboratory to identify and subsequently correct analytical errors. These were first described by Levey and Jennings in 1950.

Analytical errors cannot be eliminated completely but only reduced. In order to guarantee reliable and clinically useful test results, the laboratory should set a total error that is allowable without compromising the quality of the results and the patient care. This is defined as *Total allowable error* (TEa) and is expressed as a percentage. The choice of the TEa is based on the clinical need for each test. In other words, the TEa is the maximum error allowed for a test in order to be able to describe medically important changes in test values. This is obtained based on the *clinical decision level* (TEa = [(clinical decision level – closest reference limit) x 100] / clinical decision level). The clinical decision level is a test value or a change of a test result that triggers additional clinical actions (e.g. further testing or treatment). Usually, the clinical decision level is set with a mutual agreement between the laboratory and clinicians.

VetBooks.ir

Example

Knowing the canine potassium reference interval and clinical decision level in a specific laboratory, the TEa can be calculated as follows:

- Potassium reference interval: 3.4–5.6 mmol/l
- Potassium clinical decision level: 6.0 mmol/l

TEa = [(6.0 – 5.6) x 100]/6.0 = 6.6%

However, in veterinary medicine, it is not always easy to set clinical decision levels for each analyte, mostly owing to the variance in test results and changes related to different species, breeds, age, sex, etc.

An alternative to calculating the TEa on the basis of the clinical decision limit is to follow the guidelines recommended by the American Society of Veterinary Clinical Pathologists (ASVCP) and Clinical Laboratory Improvement Amendments (CLIA) (Figure 2.4). The ASVCP provides one or two TEa values for each analyte: one to be used when the analyte has a concentration close to the lower reference interval and one for a concentration near the higher reference interval. This is because the clinical importance of being able to detect a change in concentration at these two levels may be different. For example, for potassium it is more important to identify an increase rather than a decrease in concentration, so the TEa for the high concentration is smaller.

The performance of an analytical test is defined by the accuracy and precision of the analyser (see below). The sum of these two variables gives the total calculated error (TEc). In order to estimate a 95% confidence interval for potential errors that may occur, the equation that is most commonly used for obtaining the total calculated error is: TEc = bias + 2CV (coefficient of variation). The laboratory should ensure that the TEc is kept below the predefined TEa, by defining QC and setting an internal quality control system.

Analytical accuracy and precision are inherent sources of variation in laboratory results and are defined below.

Accuracy (bias)

Accuracy is the degree of closeness of the measurements to the true value. This is a measure of the systematic error or bias (Figure 2.5).
Accuracy is obtained from the formula:

Accuracy = (mean $_{target}$ – mean $_{measured}$) x 100(%)

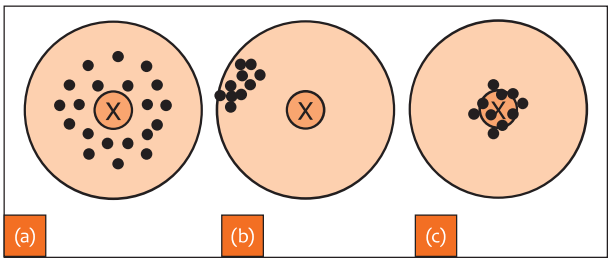
Precision (coefficient of variation)

Precision is the degree to which repeated measurements of the same sample under unchanged conditions give the same result. The closer these replicates are to each other, the more precise is the instrument or method. This is a measure of reproducibility or random error and is expressed as coefficient of variation (CV%). The coefficient of variation is obtained by dividing the standard deviation (SD) by the mean of the results (CV = SD/mean x 100%) (Figure 2.5).

Analyte	Low analyte value	Within RI	High analyte value	CLIA value
Total protein	10%	10%	10%	10%
Albumin	15%	15%	15%	10%
ALP	NCR	25% (20% desirable)	25% (20% desirable)	30%
ALT	NCR	25%	25%	20%
AST	NCR	30%	30%	20%
Bile acids	20%	20%	20%	Not found
GGT	NCR	20%	20%	15% (RCPA) 30% (CFX)
Total bilirubin	NCR	30% (25% desirable)	30% (25% desirable)	20%
Creatinine	20%	20%	20%	15%
Urea	15%	12%	12%	9%
Phosphorus	20%	15%	15%	10–23% (CAP)
Total calcium	10%	10%	10%	2% (BV) to 8% (CFX)
Sodium	5%	5%	5%	4 mmol/l
Chloride	5%	5%	5%	5%
Potassium	10%	5%	5%	0.5 mmol/l
Glucose	10%	20%	20%	6% low; 10% high
Amylase	NCR	25%	25%	30%
Cholesterol	20%	20%	20%	10%
Triglycerides	NCR	25%	25%	25%
CK	NCR	30%	30%	30%

2.4 American Society of Veterinary Clinical Pathologists (ASVCP) and Clinical Laboratory Improvement Amendments (CLIA) recommended TEa values for the most common chemistry tests. The low analyte values, within reference interval (RI) and high analyte values are TEa values recommended by the ASVCP, while the far right column refers to CLIA recommendations. The values vary depending on how near the value is to the clinical decision value (see text). ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BV = Spanish Society of Clinical Chemistry and Molecular Pathology (SEQC); CAP = College of American Pathologists Participant Summary (April 2004); CFX = Canadian Fixed Limits, The College of Physicians and Surgeons of Saskatchewan; CK = creatine kinase; GGT = gamma-glutamyl transferase; NCR = not clinically relevant; RCPA = Royal College of Pathologists of Australasia and the Australasian Clinical Biochemist Association Quality Assurance Program.
(Data from Harr *et al.*, 2013)

VetBooks.ir



2.5 In the three circles, X is the true value and the dots represent results obtained from sequential analyses of the sample. (a) An accurate but imprecise method: the dots are widely but evenly distributed around the true value and the mean of all the values is equal to the true value. (b) A precise but inaccurate method: the dots are closely clustered together, showing good repeatability, but the results are consistently biased. (c) A method that is both accurate and precise: all the dots are close to the true values and are clustered closely together.

Sources of analytical error

There are different sources of analytical error, and being able to differentiate between systematic and random errors is helpful for identifying the cause of the problem.

Systematic errors may be caused by problems with the instrument or the calibration of the test, including:

- The instrument pipette is dispensing the wrong amount of sample or reagent
- Contamination of the reagents or calibrators (e.g. interfering substance)
- Unstable or incorrectly prepared reagents or calibrators.

Random errors are usually related to:

- Instability of the instrument (e.g. in response to changes in temperature of the laboratory environment)
- Inconsistency among different operators performing the test (how the samples or reagents are prepared, pipetting technique, etc.)
- Random problems with the analysis (e.g. a bubble in the sample causing the aspiration of the incorrect amount of specimen or reagent).

To minimize the analytical errors as much as possible, procedures should be followed in order to guarantee the best performance of the instruments and reagents, calibrators and controls used:

- Detailed records of equipment maintenance according to the manufacturer's instructions should be kept and any failures of performance addressed
- Reagents and materials for calibration and control should be inventoried with dates of receipt and lot and batch numbers
- Reagents etc. should be stored under the conditions recommended by the manufacturers and discarded when outdated
- When a new batch or lot of reagents/calibrators/controls is started, its performance should be compared with the old batch and sample tests run in parallel to ensure that there are no significant changes in test performance.

Daily or more frequent checks on instrument and reagent performance are required to ensure correct analysis of patient samples. Manufacturers provide information on the expected performance of such materials. If the results obtained in the laboratory are outside these predetermined limits, patient samples should not be run until the cause has been addressed and corrected.

In addition, to minimize the imprecision caused by intraoperator differences, adequate training of the laboratory personnel and adoption of standard operating procedures (SOPs) are required.

Control and minimization of analytical error using quality control systems

There are two important QC systems in laboratory medicine:

- **Internal quality control:** to measure and monitor the precision and accuracy of the instrument (random and systematic errors)
- **External quality control programmes:** to measure and monitor the accuracy of the instrument (systematic error).

Internal quality control: The internal QC involves all the procedures used to monitor the laboratory operations continuously in order to guarantee that the performance of the instrument is good enough to produce reliable results. This can be performed by adopting a statistical QC programme able to detect errors that would invalidate patient results. This implies the use of daily checks to monitor that the results produced by the instrument remain reproducible and accurate.

Statistical QC consists of running one or more *quality control materials* (QCMs) for each analyte to monitor the performance of the analyser. The criteria that determine whether the QC results can be accepted or should be rejected (and consequently whether patient samples can be run or not) are expressed as *control rules*. The internal QC procedures primarily monitor the *bias* of data by using the QCMs and the *precision* of data by comparing multiple analyses of controls or samples.

Internal QC was first introduced by Levey and Jennings (1950), and was based on the assumption that multiple analyses of the same QCM and/or sample have a normal distribution (Gaussian distribution). This was then later developed by John Westgard (www.westgard.com), who developed a system of QC control rules (see below). The simplest of these is the 1_{2S} rule, which is commonly used as the rule for acceptance or rejection of an analytical run. This means that, if the result of one QCM (the figure '1' in the shorthand) is below or above the mean concentration of the QCM ± twice the standard deviation (SD; the '$_{2S}$' in the shorthand), the run should be rejected and patients' samples should not be tested. However, this rule has several disadvantages. First, it has a high rate of false rejection of approximately 1 out-of-control event every 20 runs, which equals a 5% rate of false rejection if only one control is used and 10% if two QCMs are analysed. This high rate of false rejection increases the amount of waste in terms of costs (control material and reagents) and time (delayed turnaround time). Additionally, the 1_{2S} rule is responsive only to random error and does not detect systematic errors.

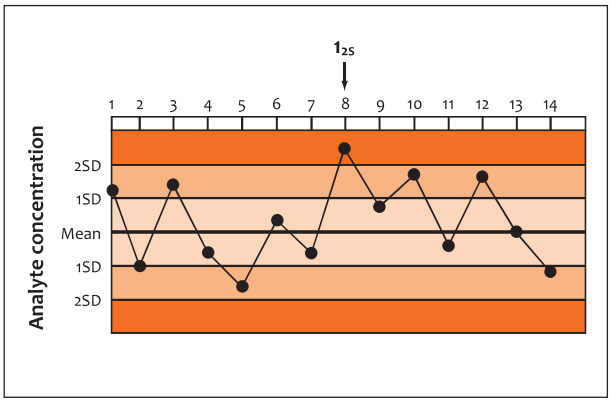
Years later, Westgard investigated the performance of different control rules by using computer simulations. With these, he was able to calculate two probabilities of detecting error: the probability of false rejection (P_{fr}) and the probability of error detection (P_{ed}). In this regard, he recommends a P_{fr} <5% and a P_{ed} ≥90% (i.e. a 90%

VetBooks.ir

chance of detecting a critical systematic shift that would cause a 5% risk of incorrect test results).

He also showed that QC procedures should usually consist of at least two control rules, one sensitive to random error and the other sensitive to systematic error. These rules are known as 'Westgard multirule control procedures' and have been introduced into the software of many instruments.

These rules can be visualized on Levey–Jennings control charts (Figure 2.6). These charts can be produced manually by plotting the daily QCM results, although most modern instrument software programs generate and store these charts and automatically perform statistical analysis of the data. Most charts are constructed to cover a period of 1 month or 30 days of data collection and analysis.

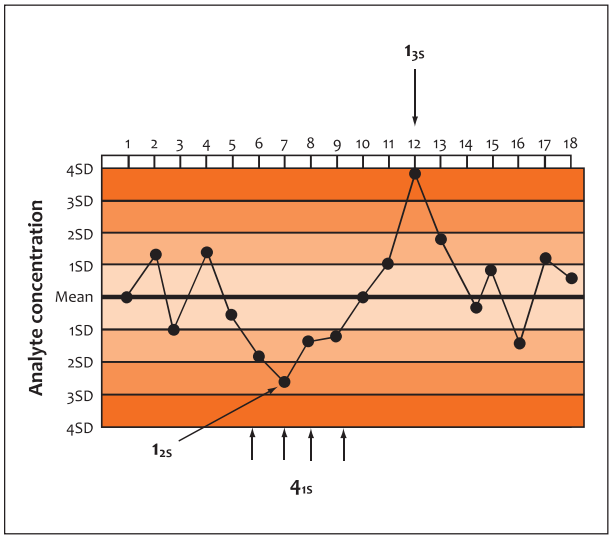


2.6 A Levey–Jennings plot system with an allowable error of ± 2SD from the mean (1_{2S}). The first seven values appear 'in control' and are close to the mean, evenly distributed either side of it. The eighth value is outside the control limits (as indicated by the arrow), but subsequent values are within control limits.

The most widely used Westgard rules are:

- The 1_{2S} rule (one control value exceeding mean ± 2SD) is an early warning for further testing using other control rules
- The 1_{3S} rule (one control value exceeding mean ± 3SD) is sensitive to random error (imprecision)
- The 2_{2S} rule (two consecutive values exceeding mean + 2SD or two consecutive values less than mean – 2SD) is sensitive to systematic error (bias)
- The R_{4S} rule rejects a run if one observation exceeds mean + 2SD and one observation is less than mean – 2SD within the same run. This rule can be applied when multiple QCMs are used, e.g. low and high QCM, but is sensitive to random error (imprecision)
- The 4_{1S} rule rejects a run if four consecutive control observations exceed mean + 1SD or are all less than mean – 1SD. This is sensitive to systematic error (bias)
- The 10X rule rejects a run if 10 consecutive control observations fall on one side of the mean. This is sensitive to systematic error (bias).

These rules help to classify the type of error into random or systematic error. Figure 2.7 shows a Levey–Jennings chart with several rules being violated. There are many control rules that can be used alone or in combination and various numbers of QCMs can be adopted for each analyte. The process of determining



2.7 A Levey–Jennings plot using the Westgard Multirules system. The first five readings are 'in control'. Readings 6–9 all exceed 1SD on the same side of the mean, violating the 4_{1S} rule. This is an indicator of systematic error (bias). In addition, reading 7 violates the 1_{2S} rule. Following recalibration of the analyser, readings 10 and 11 are 'in control', but reading 12 exceeds 3SD, violating the 1_{3S} rule, reflecting random error. Troubleshooting revealed an error in the test procedure. Subsequent readings are 'in control'.

which control rules and how many QCMs should be used to achieve a high probability of error detection (P_{ed}) and a low probability of false rejection (P_{fr}) is called QC validation.

Specifically, the number of QCMs and the number and type of QC rules to be used to obtain the best P_{ed} and P_{fr} are obtained using complicated statistical models (whose description is beyond the scope of this chapter). Westgard has released a commercially available software package specifically designed for this purpose (EZ Rules statistic package). In general, the choice of the QC rules should take into account:

- The QC goals of the laboratory (TEa)
- The stability of the method of each test; different assays have different stability of reagents or calibrators that, if ignored, could introduce systematic errors in the results
- The performance of the instrument at different analyte concentrations near to the clinical decision points; the precision and accuracy of a test vary depending on the concentration of the analyte. Hence, it is important to determine whether the instrument is performing accurately at these concentrations.

It is important to consider that, every time there is a change in the performance of a method or the quality requirements, the type and number of control rules and the total quality control (TQC) strategy should be re-assessed. Some large laboratories have recently adopted a more sophisticated control system called the Six Sigma Rules System, which was originally used for industrial manufacturing processes. Most veterinary laboratories however do not yet extend much beyond the control rules described above.

For every analyte that is measured in the laboratory, a specific QC procedure should be chosen in order to have the optimal compromise between the rate of error detection and false rejection.

VetBooks.ir

PRACTICAL APPROACH TO 'IN-CLINIC' INTERNAL QC

1 Define the quality requirements (TEa) for each analyte: the TEa can be estimated from the clinical decision limit (see earlier in text) or may be extrapolated from the literature (e.g. ASVCP Guidelines, CLIA)

Example

The TEa for 'low albumin' value recommended by the ASVCP is 15%

TEa = 15%

2 Evaluate the performance of the instrument and obtain the TEc for each analyte:

a. For each analyte, use two QCMs with concentrations similar to the clinical decision levels (e.g. low and high concentration) and analyse them every day for 5 days (in reference laboratories, 20 data points are used for this purpose)

Example

Albumin low QCM: manufacturer target value: 23 g/l
Daily results (QCM run once daily for 5 days): 22, 23, 22, 21, 23 g/l

b. Calculate the instrument bias and CV using the QCM data obtained

Example

Mean = 22.2 g/l

SD = 0.8 g/l

Bias = (mean target – mean measured)/mean target x 100 = (23 – 22.2)/23 x 100 = 3.5%

CV = (SD x 100)/mean = (0.8 x 100)/22.2 = 3.6%

c. Calculate the TEc by using the formula:
TEc = bias + 2CV

Example

TEc = bias (absolute number in %) + 2CV = 3.5% + (2 x 3.6) = 10.7%

d. The TEc is compared with the predefined TEa to deem the performance of the instrument to be adequate or not:
- If TEc < TEa, the instrument performance is adequate
- If TEc > TEa, the laboratory can:
  - Try to improve the performance of the analyser (e.g. more training for the operators, adjustments of the instrument, etc.)
  - Relax the initial quality requirements slightly, but only if there is room for additional error without compromising patient results
  - If no solution can be found, the instrument is not suitable for laboratory needs

Example

In this case the TEc (10.7%) < TEa (15%) and the instrument performance is considered adequate for the needs of this laboratory

3 Choose the number of QCMs to be run for each analyte for the daily internal QC procedure. Usually, for in-clinic laboratories, one or two levels of QCMs are sufficient. If possible, the QCMs should be chosen with concentrations similar to the clinical decision levels

Choose the QC rule(s) to be used to define the specific performance limits for a particular analyte. The choice of the QC rules (number and type) based on the P_{ed} and P_{fr} can be obtained by using specifically designed computer software or OPSpec charts (the description of these models is beyond the scope of this chapter and can be found on the Westgard website). For in-clinic laboratories, simple QC procedures are preferred such as the 1_{2S} rule (which may give rise to false rejections) and 1_{3S} rule. The latter may not detect all errors, so is more suitable for a test with good precision and accuracy and a higher TEa

4 For every QCM, prepare a *quality control chart* (e.g. Levey–Jennings chart). This can be started by using:
- The target mean and SD provided by the manufacturer of the QCMs
- Analysing the QCMs multiple times (at least 20 times) and then calculating the mean and SD. The use of in-house mean and SD is preferred because they adapt better to the instrument used. Once the mean for each QCM is known, the next step is to calculate decision limits. These limits are ± 1SD, ± 2SD and ± 3SD from the mean. These are drawn on the chart (see Figures 2.6 and 2.7)

5 Every day, before analysing patient samples, analyse the QCMs for each analyte and plot the result on the specific chart. By applying the adopted QC rule(s), accept or reject the run:

a. If the QCM data are within the ± 2SD or ± 3SD from the mean (depending on whether the QC rule in use is the 1_{2S} or 1_{3S}) then the run is accepted and the patient samples can be tested

b. If the QCM result exceeds the mean by ± 2SD or ± 3SD, the run is rejected. If this occurs, the source of error should be investigated. For example check whether:
- An adequate amount of sample or reagent has been aspirated
- There is any obvious problem with the instrument (e.g. leakage, tube blockage)
- Reagents/calibrators/QCMs have expired
- QCMs/reagents have been stored and reconstituted according to the manufacturers' guidelines
- The operator performing the test has been sufficiently trained, and follows the SOP

c. If no obvious problems are found, repeat the QC. If this is accepted, analyse the patient samples. If this is rejected again, contact technical support for the instrument

VetBooks.ir

External quality assessment: Most large veterinary laboratories participate in external QC programmes. The external quality assessment (EQA) is the process of controlling the accuracy of an analytical method by inter-laboratory comparison. In these programmes, an authorized agency prepares and sends sample materials to all the laboratories participating in the scheme. The laboratories analyse the samples and return the results to the agency. The agency then calculates the target values (consensus mean) and SDs. It also produces its own charts and statistics, the most important being the *standard deviation index* (SDI). The SDI shows the difference between the results of each laboratory and the consensus mean. These data are then sent back to the laboratories participating in the scheme. The simplest way to evaluate whether the EQA performance is acceptable is to verify that the results of the laboratory fall within ± 2SD from the consensus mean.

The most widely used external QC systems are based on human samples. These include the UK National External Quality Assessment Service (NEQAS) and the Randox International Quality Assessment Scheme (RIQAS) for haematology, biochemistry and endocrinology. A veterinary microbiology QC service is provided by the Animal Health Veterinary Laboratories Agency (AHVLA). A pilot EQA scheme for veterinary endocrinology has recently been set up by Dechra Specialist Laboratories in collaboration with the European Society of Veterinary Endocrinology (VEEEQAS or EVE-QAS).

Use of patient data in quality decisions: In human laboratories, QCMs are the primary samples used for the internal QC. Additionally, patient results can be used to supplement the QCMs, especially when the control products are very expensive or have a very short shelf life or when the QCM does not simulate the patient specimen accurately. For this purpose, samples from healthy subjects are usually used.

In veterinary laboratories, patient data are not easy to use for QC procedures, especially because the majority of samples are from sick animals. However, deviations from usual test result patterns can probably still be used to monitor performance. For example, a patient with very low calcium but no clinical signs attributable to hypocalcaemia would prompt a check on the assay performance.

Delta checks: A delta check is a flag (warning code) signalling a change in the patient's value for a test between one time and another. If the difference between two consecutive laboratory results exceeds a predefined limit, this should trigger further investigation to rule out an underlying error caused either by a pre-analytical or analytical error.

External accreditation services: Many large laboratories apply for accreditation to specific external agencies which provide assurance that the laboratory tests are performed and managed according to set standards. In the UK the major system is the United Kingdom Accreditation Service (UKAS), which follows international guidelines set down by the International Organization for Standardization (ISO) for laboratory performance.

Post-analytical errors

The ISO 15189 (2007) defines the post-analytical phase of all the procedures following the analysis of the sample, including formatting and interpretation of the result, authorization for release, and reporting and transmission of the results. Errors can occur as a result of reporting of incorrect values or ascribing the results to the wrong patient (Hawkins, 2012). Occasionally the incorrect reference values for the species may be provided. However, the majority of errors at this stage are related to the interpretation of the results. Error may occur because the person interpreting the results is a third party and is incompletely informed (e.g. incomplete history, including drug therapy) or because the clinician in charge of the case is unaware of certain changes that can occur in laboratory tests in certain conditions.

Units of measurement

In many countries, laboratory test results are reported in SI units (Système International d'Unites), while in the USA they are still widely reported in conventional units (non-SI units) based on mass gravimetric measurements. A few non-SI units have been retained in other countries, either because of the complexity of converting them into SI units or because of their widespread use.

A litre (l) is the designated measure of volume. SI units report the concentration of constituents in terms of the numbers of dissolved molecules, measured in moles (with decimal units mol, mmol, μmol, pmol). A *mole* of a chemical contains the number of grams equivalent to its molecular weight. Conventional units report concentrations of constituents in terms of the dissolved mass in grams (g, mg, μg, pg).

SI units are not used for total protein, for example, because this is a complex of molecules of different molecular weights, therefore, it is usually reported as g/l. Albumin is also reported in g/l (although it could be reported in μmol/l), largely because total protein and albumin are considered together when test results are evaluated and used to determine the globulin concentration (globulin = total protein – albumin).

The SI unit of enzyme activity is the *katal*, which is defined as the amount of enzyme that will catalyse the transformation of 1 mole of substrate per second in an assay system. This is the reporting unit accepted by the IUPAC (International Union of Biochemistry), but not for clinical tests, and the international unit (IU) continues to be used. There is a constant relationship between katal and IU when measured under identical conditions: 1 katal = 60 million IU.

Some conversion factors are shown in Figure 2.8. A more complete conversion table can be found in Appendix 7. In addition, there are various conversion tools available online which can be useful, particularly for some of the more unusual analyses (for example http://www.globalrph.com/conv_si.htm).

Gravimetric unit conversion	SI unit
g/100 ml x 10	g/l
g/100 ml x $\frac{10}{\text{mol wt}}$	mol/l
pg/ml x $\frac{10^3}{\text{mol wt}}$	pmol/l
mEq/l x $\frac{1}{\text{valency}}$	lmol/l

2.8 Conversion factors from gravimetric units to SI units (see also Appendix 7).

VetBooks.ir

Interpretation of test results

For sick patients, laboratory tests are used to help in the diagnosis of disease, or to monitor disease progression or the response to a treatment. In all cases, the results obtained by the laboratory are compared with species-specific reference intervals. However, when monitoring therapy or the course of a disease, useful information may be gained by comparing sequential laboratory results (e.g. before and after treatment).

For a correct interpretation of the laboratory data, it is essential to integrate the results with the patient history, physical examination, ancillary tests and the list of clinical differential diagnoses. The approaches in the interpretation of the laboratory results are:

- Comparison with predetermined reference values
- Comparison between two (or more) sequential results.

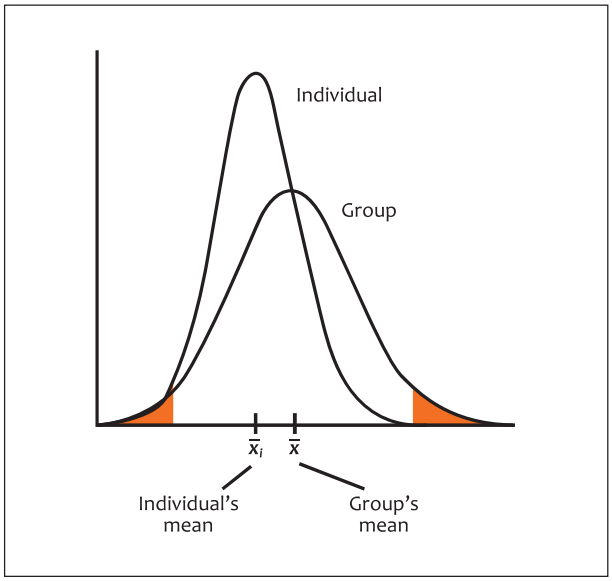
Comparison with predetermined reference intervals

A reference interval (RI) for a given analyte is a range of values expected to be found in healthy animals. This represents the interval between an upper and lower limit and commonly includes the central 95% of the values from the selected reference sample group, determined by statistical methods. Reference intervals may also be referred to as 'normal values', 'expected values' or, more commonly, as 'reference ranges'.

A 'reference range' is defined as the entire range of values (actual minimum to maximum measured values) obtained for a test on a reference sample group of healthy animals. The term 'normal value' should be avoided as this implies absence of disease, while sick patients may have some analytes that are within the reference intervals. Additionally, in some pathological conditions, finding some results within the RIs could be an indication of disease (e.g. a lymphocyte count within the RI in severely sick dogs may suggest an underlying hypoadrenocorticism; red cell parameters within RIs in a markedly dehydrated animal could mask an underlying anaemia).

Reference intervals may be classified as population-based or subject-based RIs. In the former, the RIs are obtained from a group of reference individuals selected (preferably randomly) from a reference population (see later). Subject-based RIs (or intra-individual RIs) are derived from sequential samples from a single individual.

The width of population based-RIs is wider than the width of subject-based RIs (Figure 2.9). Determining which type of RI is more appropriate for each analyte is very important when interpreting laboratory results. This is achieved by calculating the index of individuality for a given analyte, which is based on biological and analytical variation.



2.9 Illustration of the problems with population-based reference intervals (RIs) when applied to an individual. If the test of interest has marked individuality, a result may fall within the population-based RI even though it is too high for that individual.

Biological variation and index of individuality

Biological variation (BV) is the random inherent variation of analytes around a homeostatic set point. The inherent oscillation of the analyte's concentration leads to a variation within each individual (within-subject BV or BV_I) and between animals (between-subject BV or BV_G) at any particular time point.

There are three types of biological variation:

- Variation over the lifespan (age): HCT, total protein, globulin, ALP, calcium, phosphate, CK
- Predictable cyclical variation (daily, monthly, seasonal): cortisol, oestrogen, thyroxine
- Random variation: urea, creatinine.

Within-subject biological variation is represented by the mean coefficient of variation (CV) for consecutive values obtained from a single animal (CV_I). The between-subject biological variation is represented by the mean CV for values obtained from different individuals (CV_G). The CVs for commonly used analytes in dogs and cats are shown in Figure 2.10. The low CVs for electrolyte and protein levels reflect the regulation of these parameters within a small range in the body and the high precision of analytical tests for these parameters. Conversely, urea and creatinine have much higher CVs. Urea is affected by diet and creatinine is affected by muscle mass and exercise, and their levels are not so closely controlled in the body. Hormones have greater variability due to circadian rhythms and the test methods are not as accurate and precise as those for other biochemical tests.

Knowledge of the variation in CVs for different analytes is important when monitoring specific parameters. For example, in a dog with hypoadrenocorticism an increase in potassium of >3.3% (approximately 0.25–0.30 mmol/l) would reflect a 'real' increase and not be attributable to intra-patient biological variation or assay variation. However, a similar small increase in urea may simply reflect biological or assay variation, which could account for an oscillation of up to 19% from the homeostatic set point within one animal.

In 1974, the concept of index of individuality (II) was introduced in human laboratory medicine. The II is defined as the ratio between the intra-individual and between-individual BVs. The original equation included the analytical variation (CV_A) but this formula is often simplified to $II = CV_I/CV_G$.

As mentioned above, the II is used to investigate the utility of conventional population-based reference values compared with subject-based RIs. Analytes with $CV_I < CV_G$ have a low II and therefore have a marked individuality. This means that the variation that occurs in a

VetBooks.ir

Analyte	CV_G (%)	CV_I (%)	CV_A (%)	II
RBC	4.4	5.4	2.8	1.4
HCT	5.2	6.4	1.1	1.3
Hb	4.7	5.9	2.9	1.4
WBC	12.3	12.1	3.7	1.0
ALT	23.7	9.7	3.2	0.4
AST	10.9	11.4	3.3	1.1
ALP	34.4	8.6	1.7	0.3
Albumin	3.0	2.4	1.6	1.0
Total protein	3.1	2.6	1.1	0.9
Urea	35.1	16.1	3.8	0.5
Creatinine	12.9	14.6	2.9	1.2
Phosphorus	12.1	13.7	3.4	1.2
Cholesterol	15.1	7.3	3.0	0.5
Glucose	3.8	9.5	3.7	2.7
Potassium	3.6	3.3	0.1	0.9
aPTT	69.3	19.4	25.0	0.5
PT	4.6	1.4	2.0	0.8
(a) cPLI	49.5	193.8	8.4	3.9

Analyte	CV_G (%)	CV_I (%)
ALT	23.01	13.21
AST	14.97	14.65
ALP	33.72	12.45
Albumin	3.88	3.00
Total protein	13.95	8.31
Urea	15.54	10.43
Creatinine	11.43	5.97
Phosphorus	11.79	8.49
Cholesterol	22.35	10.74
Glucose	8.06	6.76
Potassium	4.91	3.63
Sodium	0.57	0.86
Chloride	1.16	1.17
Calcium	2.51	2.34
(b) Total bilirubin	94.31	87.46

2.10 (a) Data on biological variation (BV) and index of individuality (II) in dogs. (b) Data on biological variation (BV) in cats. BV values from other sources may vary. ALP = alkaline phosphatase; ALT = alanine aminotansferase; aPTT = activated partial thromboplastin time; AST = aspartate aminotransferase; cPLI = canine pancreatic lipase immunoreactivity; CV_A = analytical variation; CV_G = inter-subject BV; CV_I = intra-subject BV; Hb = haemoglobin; HCT = haematocrit; PT = prothrombin time; RBC = red blood cells; WBC = white blood cells. (a, Adapted from Walton, 2012; b, Adapted from Baral *et al.*, 2014)

single individual is smaller than the variation that occurs between different subjects. In dogs, examples of analytes with a marked individuality (low II) are ALP, alanine aminotransferase (ALT), cholesterol, thyroid-stimulating hormone (TSH) and coagulation times. For these tests, the use of population-based RIs is insensitive for detecting a change in the clinical status of the animal, and subject-based RIs or reference change values (RCVs – see below) would be preferred. In fact, in sick patients, these analytes may shift from their usual set point, but not enough to move the result outside the RIs (Figure 2.11). Conventionally, an analyte is considered markedly individual when its II is <0.6.

Conversely, for analytes with $CV_I > CV_G$ (II >1.4), the use of population-based RIs is adequate (e.g. glucose, fructosamine, canine pancreatic lipase immunoreactivity). For those analytes with II lying between 0.6 and 1.4 (e.g. total protein, albumin), population-based RIs can be used, but with caution, and the interpretation of the results may be aided by using RCVs. Canine indexes of individuality are shown in Figure 2.10. Unfortunately, in laboratory medicine, obtaining subject-based RIs for each patient is impractical and unfeasible, and population-based RIs are commonly used independently from the II of the analytes. An alternative to having subject-based RIs when interpreting the result of a test with low II is the use of reference change values (see below).

Determination of the reference intervals

As discussed above, the reference interval is the most widely used medical decision-making tool. Hence, the quality of the RIs plays as important a role in result interpretation as the quality of the result itself. Because reference values are affected by pre-analytical and analytical variables, these stages should be subjected to rigorous QC before establishment of reference values.

In recent years, working groups of both the American and European Societies of Veterinary Clinical Pathology (ASVCP and ESVCP) have been preparing guidelines for establishing reference intervals *de novo*. These guidelines mirror the recommendations published by CLSI. In practice, it is often impossible for any single laboratory to perform these studies and alternative processes can be adopted, such as the transference of reference intervals. This is especially true in veterinary medicine because of the large number of species and breeds encompassed.

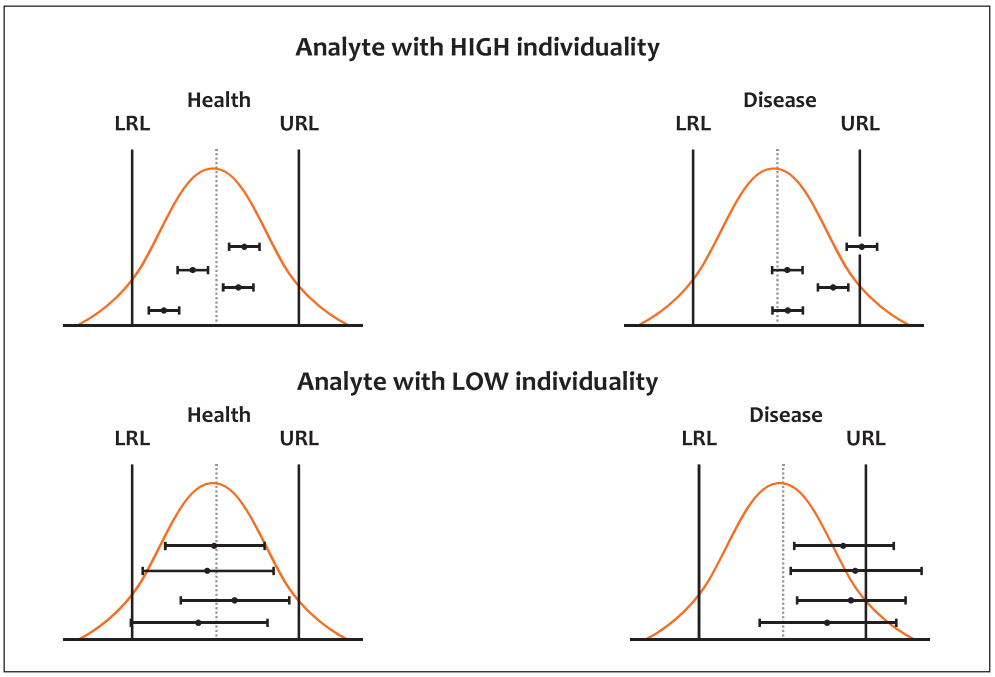
Definitions:

- **Reference individual:** a subject selected for testing based on stringent predefined inclusion and exclusion criteria.
- **Reference population:** the entire group of all selected reference individuals.
- **Reference sample group:** a subgroup of reference individuals, selected (preferably randomly) from the reference population that is used to determine the RIs.

Determination 'de novo' of the reference interval: Appropriate selection of the reference population to be used for the establishment of the RIs is essential. The reference population from which reference individuals are chosen should be predefined, and tight clinical parameters for 'healthy' must be established by defining the inclusion and exclusion criteria. Ideally, this population should be representative of the animals from whom samples are sent to the individual laboratory. In practice this is difficult to achieve. Bias may occur if a restricted group of animals is used, for example, values from a colony of young Beagle dogs or from cattery cats. The reference population should represent a general mix of breeds, sexes and ages, living in different environmental conditions. It is also important to consider the Veterinary Surgeons Act 1966 when establishing reference intervals. Collecting blood from healthy animals is not allowed for this purpose; however, excess blood from samples collected for the patient's benefit may be used, for example, surplus blood from pre-anaesthetic screens.

For the establishment of the RI, appropriate numbers of reference individuals should be randomly selected (reference sample group) from the reference population.

VetBooks.ir



2.11 Examples of two analytes with a high and a low index of individuality, respectively. The four horizontal bars represent the range of values in four individuals. In the case of high individuality, a significant change in the analyte concentration caused by the disease may be missed and the result may fall within the RI. In this case, the use of subject-based RIs or RCVs may be beneficial. LRL = lower reference limit; URL = upper reference limit.

The number of individuals needed for this purpose should be estimated based on the desired confidence interval (CI) of the RI. The CI reflects the probability that a reference limit derived from a sample group approximates the true reference limit from the entire reference population. Once the reference individuals have been selected, the values obtained from each of these animals are subjected to statistical analysis.

First, the data should be analysed for their distribution (normal distribution or not), preferably using graphical analysis (e.g. histograms; Figure 2.12) or by using goodness-of-fit statistical tests such as the Anderson–Darling test. This step also highlights potential outliers, which are reference values that do not belong to the underlying distribution. If these outliers are retained, they will widen the RIs, decreasing the sensitivity of the test.

The presence of outliers may be due to:

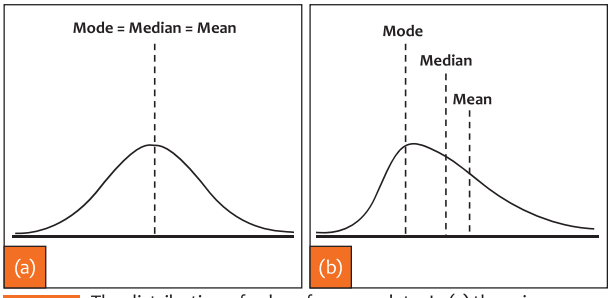
- Inclusion of non-healthy or non-representative subjects in the reference population
- Pre-analytical, analytical or post-analytical errors.

Unless these values are known to be the result of one of these possibilities, outliers should be retained. If not, after the outlier has been removed, retesting the remaining values for any additional outliers would be recommended. Specific statistical methods (e.g. Dixon's test and Horn's algorithm) can be adopted to identify outliers accurately.

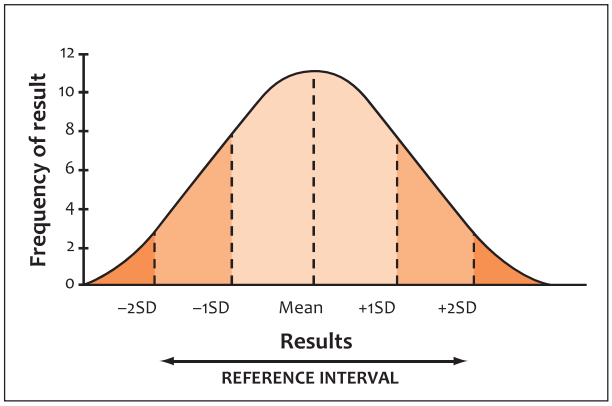
Different statistical methods can be used to define the RIs, and the choice of test to be used depends on the number of reference subjects that are available and on the distribution of the results (Gaussian or not). The more data used, the more likely it is that the established RIs will accurately reflect the entire population. Conventionally, at least 120 healthy animals are required to produce reliable RIs, but as few as 40 subjects may be used if necessary.

If the values are normally distributed the reference intervals are established on the basis of the mean ± 2SD (Figure 2.13).

If the data do not have a Gaussian distribution, a non-parametric test is required. This consists of ranking values and using percentile limits. The value at the 97.5th percentile is the upper reference limit and the value at the



2.12 The distribution of values for an analyte. In (a) there is a Gaussian symmetrical distribution and the mean, median and mode are in the same central position. These data could be analysed by parametric methods, calculating the mean and 2SD to produce reference values. (b) The data points are not in a symmetrical distribution and the mode, median and mean are different. These data would be analysed by non-parametric methods (usually using percentiles) to produce reference values.



2.13 Establishment of reference intervals by using the central values with exclusion of the lowest and highest 2.5% of the reference values.

VetBooks.ir

2.5th percentile is the lower reference limit. The values obtained from each reference individual are put into ascending order. If *n* is the number of samples, the position of the sample result lying on the 2.5th percentile is calculated using (*n* + 1) x 0.025. The 97.5th percentile is calculated using (*n* + 1) x 0.975. For example, if there are 78 dogs in the reference population the 2.5th percentile is calculated as (78 + 1) x 0.025 = 1.975 (approximately 2). The samples are ranked in ascending order from 1 to 78. The value of the second result in the ascending series is the lower limit of the reference interval. The 97.5th percentile is (78 + 1) x 0.975 = 77. The value of the 77th sample is the upper limit of the reference interval.

In both cases, the defined RIs represent 95% of the animals tested. Given this, it should be borne in mind that, using these methods, 5% of healthy animals would be classified as abnormal (2.5% below the limit for measured values and 2.5% above the limit for measured values). Therefore, for every 20 healthy animals there will be one animal with results outside the RI. For most tests and analytes this is acceptable, because truly diseased animals will be expected to have values far higher or lower than the RIs.

When evaluating a panel of test results from a single animal, there is a probability of $1 - 0.95^{n}$ (where *n* is the number of tests in the panel) that not all values will be within the set reference interval. Thus, for 20 results there is a 64% chance that one result will be 'abnormal' [$100 \times (1 - 0.95^{20})$]. This should be remembered when interpreting results from potentially clinically healthy animals for pre-anaesthetic screens or geriatric profiling.

A valuable aid in the statistical analysis of data for establishing the RIs is the use of the '*Reference Value Advisor*'. This is a set of macro instructions for Microsoft Excel™ that computes reference intervals using the standard and robust methods. This is available online for free download at http://www.biostat.envt.fr/spip/spip.php?article63.

Determination of the reference interval using patient data already analysed: The introduction and increased use of computer databases and laboratory information management systems (LIMs) have allowed stored patient data to be used to construct reference intervals. If a large proportion of the patient samples are from healthy individuals, computerized methods based on a combination of laboratory and diagnostic data can be used to select healthy patients to produce RIs. However, this approach does not guarantee that most of the data are derived from healthy subjects. Additionally, this type of data accumulation has inherently increased levels of error related to pre-analytical and analytical factors. For this reason, current guidelines do not endorse this method.

Transfer of reference intervals: This is another, more widely used and accepted, method to determine RIs. When new instrumentation or methods are introduced into a laboratory, reference intervals can be obtained from:

- An existing reference interval generated in the same laboratory on an old instrument or using a different method
- Values from another laboratory using the same instrument and/or method
- Values provided by the manufacturer.

When the method and instrument used are the same, RIs can be transferred directly. Otherwise a comparison of methods should be carried out. If a bias between the two methods or instruments is found, the reference limits may be adjusted using regression analysis.

Once the RIs have been transferred to a new instrument or method, these should be validated before being used in a clinical setting. A way to determine whether the RIs can be safely used is to measure the analyte(s) on 20 healthy animals and compare the results with the 95% CI provided. When ≤2 values exceed the interval, the RIs can be adopted. If 3 or 4 of the values lie outside the interval, an additional 20 healthy individuals can be tested and interpreted as above. If ≥3 are still outside the interval, the RIs should be rejected and new ones established.

It has to be noted that if the RIs to be transferred are inappropriately wide, this method will fail to identify unhealthy subjects accurately because there will be a greater chance that all samples will fall within the given interval.

Limitations of reference intervals

Most laboratories provide reference intervals that are based on a wide-ranging reference population. A narrower selection of healthy subjects partitioned into subgroups (e.g. age, breed, sex) would be ideal, but this is often impractical. If partitioned RIs are not available, it is important to be aware of the common deviations of specific subgroups from the ranges provided. A typical example is young animals, which have HCT, MCV, total protein, globulins, calcium, phosphorus and ALP values that differ from the adult concentrations. Likewise, certain hormone levels, electrolyte and protein values may vary outside the quoted values in pregnant (depending on the stage of gestation) and lactating animals. Specific breed-related differences should also be considered, including:

- Greyhounds and other sighthounds: higher HCT, RBC count, MCV, MCHC and Hb, higher creatinine and ALT; lower WBC, neutrophil and platelet counts, lower total calcium, total protein and globulin
- Japanese breed of dogs (e.g. Akitas, Chinese Shar Pei): microcytosis
- Cavalier King Charles Spaniels and Norfolk Terriers: macrothrombocytopenia.

Partitioned RIs based on age and some canine breeds have been published in the literature in the past few years. However, before adopting these RIs, they should be transferred and validated as described in the section above.

Interpretation of results by comparison between two sequential results

The comparison of two or more sequential laboratory results from the same patient is not as straightforward as it may seem. In fact, each result carries an inherent random variation so that the result is not a single number but a dispersion of numbers.

When two sequential results differ, their difference can be due to:

- An inherent source of variation: pre-analytical, analytical or biological variation
- Clinical improvement or deterioration of the patient.

A change in the condition of the patient is indicated when two consecutive results exceed a certain value known as the *reference change value* (RCV) (also called

VetBooks.ir

the significant change value) or, in other words, when the difference between the results is greater than the inherent variation of the test.

The inherent sources of variation of laboratory tests were described at the beginning of this chapter, when the pre-analytical (CV_P), analytical (CV_A) and biological variation (CV_I) were discussed. The total amount of inherent variation (CV_T) intrinsic in each laboratory result is given by the following formula: $CV_T = \sqrt{(CV_P^2 + CV_A^2 + CV_I^2)}$. Often, the pre-analytical variable is excluded from this equation when assuming that standardization of sample collection and handling and patient preparation has been adopted.

Due to the fact that the inherent variation is random, by definition this has a Gaussian distribution. As expected in a normal distribution, 99.7% of the values will fall within the range ± 3CV from the mean concentration, 95.5% within ± 2CV and 68.3 within ± 1CV. The multipliers 1, 2 and 3 are called *z-scores*. Therefore, every analytical value lies within ± *Z* x *total variation* with a probability appropriate to the *z*-score. When two consecutive results are compared, the total variation doubles. To be clinically relevant, the difference between two sequential results must be ≥ $Z \times \sqrt{[2 \times (CV_A^2 + CV_I^2)]}$. This value represents the reference change value (RCV) and is expressed as a percentage.

When monitoring a patient, the use of RCVs is especially important for all those analytes that have a marked individuality and for which the population-based RIs are not sensitive enough. The main limitation of the use of RCVs in sick patients is that the biological variation (BV) data are often obtained from healthy animals and therefore may not truly mirror the BV of each analyte in the presence of disease. In fact, in human medicine there is evidence that the BV of some analytes is higher in diseased patients than in healthy subjects. As a consequence of this, the use of RCVs determined for healthy individuals to interpret sequential results from sick patients may cause a false positive interpretation of the results. Ideally, the BV of each disease-associated analyte should be estimated in patients with specific diseases.

Example

A 6-year-old, male neutered Jack Russell Terrier is presented with a history of protein-losing enteropathy.

Analyte	Day 1	Day 14
Albumin (g/l)	17	14

Clinical data

Analyte	CV_A	CV_I
Albumin	1.3%	2.4%

Laboratory analytical variation (CV_A) and biological variation (CV_I)

Question: Is the difference between the results obtained on day 1 and day 14 clinically significant or does this just reflect an inherent variation of the tests?

Answer: The albumin reference change value should be calculated. If the difference between the results on day 1 and day 14 is greater than the RCV, it would mean that the patient has deteriorated:

$RCV = Z \times \sqrt{[2 \times (CV_A^2 + CV_I^2)]}$

Z = 2 gives a probability of 95.5%
Z = 3 gives a probability of 99.7% ▶

Albumin RCV = $2 \times \sqrt{[2 \times (1.3^2 + 2.4^2)]}$
= (95.5% probability)
= 7.7

= $3 \times \sqrt{[2 \times (1.3^2 + 2.4^2)]}$
= (99.7% probability)
= 11.58

The difference between the albumin concentrations on day 1 and day 14 is 3 g/l, which expressed as a percentage is 17.6% (3/17 x 100).

This difference is higher than the RCV and therefore this change is significant and reflects a deterioration of the patient.

Clinical decision limits

When interpreting a laboratory result, the final clinical decision must take into consideration not only the RIs but also the clinical information and the clinical significance of a laboratory test. As discussed above, RIs representing the central 95% of the distribution of the values can be established. However, the final choice of the reference limits (cut-off values) should take account of the sensitivity and specificity required for a given test, especially where there is an overlap in the results from healthy and diseased patients.

This involves setting cut-off values that minimize the number of false negatives or false positives for a particular test. Cut-off values are determined using the concepts of sensitivity, specificity and predictive value, based on the distribution of test results from healthy animals, animals with the disease of interest and, in certain situations, a third group of animals with a different pathological condition.

For example, if one uses the urine cortisol:creatinine ratio for the diagnosis of hyperadrenocorticism and sets a *low* cut-off value, the test will have close to 100% diagnostic sensitivity (there will be very few false negative results) but a low specificity with many false positive results. This can be interpreted clinically to mean that if the test result is negative then it is highly likely to be a true negative and the animal does not have hyperadrenocorticism. However, there will be many false positive results and so other diagnostic tests, such as an ACTH stimulation test, would be required to confirm the presence of disease. If a *high* cut-off value is set, the specificity will be increased to close to 100% with very few false positive results but sensitivity will decrease and more false negatives will be generated.

In cases where tests are affected by more than one disease, setting cut-off limits becomes difficult. For example, amylase and lipase are excreted by the kidney. To obtain high specificity for the diagnosis of pancreatitis in an animal with renal compromise, a high cut-off limit for the pancreatic enzyme tests would have to be set and false negatives would be more likely. In general, cut-off limits are set at levels that produce the highest diagnostic efficiency for a particular disease.

The selection of the appropriate laboratory test should take into consideration the clinical performance characteristics of a test and the purpose of the selected test. If a test is used to screen for a disease of low prevalence in a healthy population, then it must be very sensitive (to identify a high proportion of affected animals), while specificity is less important (animals which test positive can be subjected to further, more specific tests). Screening tests also need to be safe and inexpensive. For tests that are used to

VetBooks.ir

confirm a diagnosis, specificity is more important (the test should not incorrectly identify non-diseased animals), especially if the consequences of a positive result are serious (e.g. chemotherapy, surgery or even euthanasia).

Clinical performance characteristics of a test

Based on the presence or absence of disease, test results can be classified into:

- **True positive (TP):** a result that correctly identifies a patient as having a specified disease
- **True negative (TN)**: a result that correctly identifies a patient as not having a specified disease
- **False positive (FP)**: a result that incorrectly identifies a patient as having a specified disease
- **False negative (FN)**: a result that incorrectly identifies a patient as not having a specified disease.

Example

A population of 1000 dogs is tested for disease X. According to the gold standard test, 100 dogs are affected by the disease and 900 dogs are not (prevalence 10%) in case A; only 10 dogs are affected in case B whereas 990 are not affected (prevalence 1%). A new test for disease X is applied to this population and the results are as follows:

Case A

		Gold standard test		
		Dogs affected by disease X	*Dogs not affected by disease X*	
New test	Dogs testing positive with new test	90 (TP)	45 (FP)	PPV = 67% (90/135)
	Dogs testing negative with new test	10 (FN)	855 (TN)	NPV = 99% (855/865)
		SEN = 90% (90/100)	SPEC = 95% (855/900)	

Case B

		Gold standard test		
		Dogs affected by disease X	*Dogs not affected by disease X*	
New test	Dogs testing positive with new test	9 (TP)	49 (FP)	PPV = 16% (9/58)
	Dogs testing negative with new test	1 (FN)	941 (TN)	NPV = 99.9% (941/942)
		SEN = 90% (9/10)	SPEC = 95% (941/990)	

Using the ability (or inability) of a test to produce correct results, the clinical performance of each test can be calculated. The clinical performance of a test is described by the following: ▶

- **Diagnostic sensitivity (SEN):** the frequency of positive test results in animals that have the disease. The use of a test with a high diagnostic sensitivity is preferred when screening for the presence of a disease.
  SEN = TP/(TP + FN) x 100 = 90%
- **Diagnostic specificity (SPEC):** the frequency of negative test results in animals that do not have the disease. A highly specific test is used to confirm the presence of a disease.
  SPEC = TN/(TN + FP) x 100 = 95%
- **Positive predictive value (PPV):** probability that an animal with a positive test result has the disease; PPV = TP/(TP + FP) x 100 = 67%
- **Negative predictive value (NPV):** probability that an animal with a negative test result does not have the disease; NPV = TN/(TN + FN) x 100 = 99%
- **Prevalence:** estimate of the frequency of a disease in a population at a point in time.
  PREV = (TP + FN)/(TP + TN + FP + FN) = 10%

The above example shows how the prevalence is important in determining the predictive value of a test. While sensitivity and specificity reflect the *pre-test* probability of a test itself to correctly identify sick from healthy animals, the positive (and negative) predictive values represent a *post-test* probability that is determined by the amount of disease present in the population of interest. For a test with a diagnostic sensitivity and specificity of 95%, the predictive value of a positive test result (PPV) within a population with a disease prevalence of 50% is 95%. However, if the prevalence is only 5% then the predictive value of a positive test decreases to only 50%, causing the predictive value for the test to be no better than chance or flipping a coin.

Prevalence 50%, sensitivity and specificity 95%		
TP: 47.5%	FP: 2.5%	PPV = 47.5/50 x 100 = 95%
FN: 2.5%	TN: 47.5%	NPV = 47.7/50 x 100 = 95%

Prevalence 5%, sensitivity and specificity 95%		
TP: 4.75%	FP: 4.75%	PPV = 4.75/9.5 x 100 = 50%
FN: 0.25%	TN: 90.25%	NPV = 90.25/90.5 x 100 = 99.7%

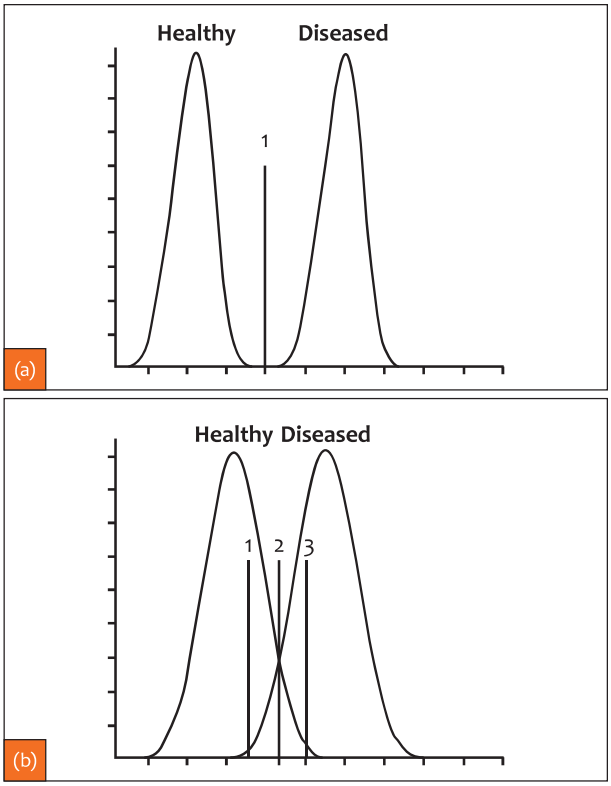
A test that has reasonably high sensitivity and specificity and is a good diagnostic test in a population with a high probability of having the disease will, therefore, be very poor in a population where disease prevalence is very low, i.e. when used as a screening test in a healthy population. The typical example used to illustrate the PPV is the in-house assay (snap test enzyme-linked immunosorbent assay (ELISA)) for feline immunodeficiency virus (FIV). This test has a diagnostic specificity that is below 100% (i.e. false positive results may occur). This becomes extremely important in populations with a low prevalence of FIV, such as in the UK, where the reported prevalence of this infectious disease is <1%. Using an assay with diagnostic specificity of 95%, the PPV would be approximately 16%. So, every cat that tests positive should be retested with a 'gold standard assay' e.g. Western blotting or polymerase chain reaction (PCR). Conversely, if a cat that had a positive result with the test comes from a high-risk area (where the prevalence is higher), this result is more likely to be a true positive result.

VetBooks.ir

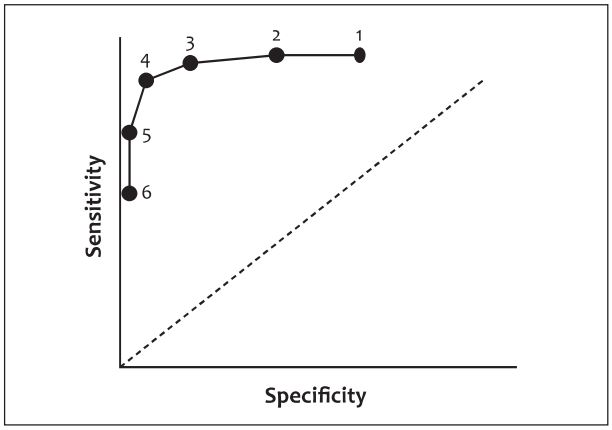
Generally the aim is to maximize both sensitivity and specificity, but no test is 100% sensitive and 100% specific. As one is maximized, the other is decreased. In the diagnostic situation a test with 100% sensitivity may generate unacceptable numbers of false positive test results. Likewise, when a test is 100% specific it may generate unacceptable numbers of false negative results. In the clinical diagnosis of disease, other medical decision limits can be used, and where results have a numerical value (i.e. not just positive or negative), cut-off values can be selected to maximize the discriminatory power of the test (Figure 2.14).

Receiver-operating characteristic curve analysis

This can be used to show graphically the ability of a test to discriminate between diseased and healthy animals or to compare the efficiency of two tests in the diagnosis of a disease. To produce a receiver-operating characteristic (ROC) curve, sensitivity (true positive rate) is plotted against 1 – specificity (false positive rate). Different cut-off values can then be applied to generate the best values for decisions about diagnosis (Dawson-Saunders and Trapp, 2000). A perfect diagnostic test would have 100% sensitivity and 100% specificity and be close to the top left corner of the graph. A diagonal line (lower left corner to upper right corner) would indicate a useless test. The point on the curve that is closest to the upper left corner is the cut-off value or decision limit that provides the greatest diagnostic accuracy (efficiency of the test) (Figure 2.15). The area under the curve (AUC) is a quantitative representation of the overall accuracy of the test and ranges between 0.5 and 1. The greater the AUC, the more accurate is the test in diagnosing the disease in question. Conventionally, values between 0.5 and 0.7 represent a test of low accuracy. If the AUC is >0.9, the test accuracy is high, while values in between (0.7–0.9) represent a test with moderate accuracy in the diagnosis of a specific disease.



2.14 Demonstration of the problem of trying to establish cut-off points for any test between healthy individuals and diseased individuals. In (a) the test results from healthy animals do not overlap the test results from diseased animals and so there is a clear cut-off indicated by line 1. This test has 100% sensitivity and specificity. In (b) the test results from healthy animals overlap those from diseased animals. If the cut-off is set at line 1 the test is sensitive but not specific, because a high proportion of healthy animals will have results above the cut-off. Conversely, if the cut-off is set at line 3 the test becomes more specific (very few non-diseased animals have results above the cut-off), but is much less sensitive (a significant proportion of diseased animals have results below the cut-off. If line 2 is selected as the cut-off, the test has moderate sensitivity and specificity.



2.15 The receiver-operating characteristic (ROC) curve is a graphical representation of diagnostic sensitivity and specificity for a test at varying selected cut-off values, in this example numbered 1–6. Cut-off level 1 is the lowest cut-off plotted and gives a high sensitivity but low specificity. Cut-off level 6 is the highest cut-off plotted. The specificity is much higher (close to the *y*-axis) but the sensitivity is lower. Cut-off 4 has the best compromise of sensitivity and specificity, lying closest to the top left corner of the graph. A good test has values close to the upper left corner of the plot. Test results around the diagonal dotted line would indicate a useless test.

Example

The following contingency table represents the performance of the 'mitotic index' (MI) as a diagnostic test to predict the risk of death of a patient with an intermediate grade mast cell tumour (MCT). The MI cut-off value is set at 5 mitoses per high power field at 40× (≤5 equals a negative result; >5 equals a positive result). Based on the data in the table, calculate the sensitivity (SEN), specificity (SPEC), positive predictive value (PPV) and negative predictive value (NPV) of the test, assuming that 49 dogs were included in the study and 10 dogs died of their disease.

		Clinical outcome	
		Dogs that died of intermediate grade MCT	*Dogs that did not die of intermediate grade MCT*
Mitotic Index	Positive MI (MI >5)	6 (TP)	2 (FP)
	Negative MI (MI <5)	4 (FN)	37 (TN)

Results

SEN = TP/(TP + FN) x 100 = 6/(6 + 4) x 100 = 60%
SPEC = TN/(TN + FP) x 100 = 37/(37 + 2) x 100 = 95%
PPV = TP/(TP + FP) x 100 = 6/(6 + 2) x 100 = 75%
NPV = TN/(TN + FN) x 100 = 37/(37 + 4) x 100 = 90%

VetBooks.ir

References and further reading

Baral RM, Dhand NK, Freeman KP, Krockenberger MB and Govendir M (2014) Biological variation and reference change values of feline plasma biochemistry analytes. *Journal of Feline Medicine and Surgery* **16(4)**, 317–325

Dawson-Saunders B and Trapp RG (2000) Evaluating diagnostic procedures. In: *Basic and Clinical Biostatistics, 3rd edn*, pp. 232–247. Appleton and Lange, Norwalk

Farr AJ and Freeman KP (2008) Quality control validation, application of sigma metrics, and performance comparison between two biochemistry analyzers in a commercial veterinary laboratory. *Journal of Veterinary Diagnostic Investigation* **20**, 536–544

Fraser CG (2001) *Biological Variation: from principles to practice*. AACC Press, Washington, DC

Friedrichs KR, Harr KE, Freeman KP *et al.* (2012) ASVCP reference interval guidelines: determination of *de novo* reference intervals in veterinary species and other related topics. *Veterinary Clinical Pathology* **41(4)**, 441–453

Harr KE, Flatland B, Nabity M and Freeman KP (2013) ASVCP guidelines: allowable total error guidelines for biochemistry. *Veterinary Clinical Pathology* **42(4)**, 424–436

Hawkins R (2012) Review article: managing the pre- and post-analytical phases of the total testing process. *Annals of Laboratory Medicine* **32(1)**, 5–16

Lester S, Harr KE, Rishniw M and Pion P (2013) Current quality assurance concepts and considerations for quality control of in-clinic biochemistry testing. *Journal of the American Veterinary Medical Association* **2(15)**, 182–192

Levey S and Jennings ER (1950) The use of control charts in the clinical laboratory. *American Journal of Clinical Pathology* **20(11)**, 1059–1066

Radford A and Dawson S (2005) Diagnosis of viral infections. In: *BSAVA Manual of Canine and Feline Clinical Pathology, 2nd edn*, ed. E Villiers and L Blackwood, pp. 410–423. BSAVA Publications, Gloucester

Rishniw M, Pion PD and Maher T (2012) The quality of veterinary in-clinic and reference laboratory biochemical testing. *Veterinary Clinical Pathology* **41(1)**, 92–109

Walton RM (2012) Subject-based reference values: biological variation, individuality, and reference change values. *Veterinary Clinical Pathology* **41(2)**, 175–181

Westgard JO (2000) *Basic Planning for Quality – Training in Analytical Quality Management for Healthcare Laboratories*. Westgard QC Publishing, Madison, WI

Useful websites

American Society for Veterinary Clinical Pathology
www.asvcp.org/pubs/index.cfm

Veterinary Biological Variation
http://vetbiologicalvariation.org

Westgard QC
www.westgard.com