Chapter 1

ANALYTICAL AND DIAGNOSTIC CONCEPTS: GETTING IT RIGHT

Kristen R. Friedrichs, Michael A. Scott, and Steven L. Stockham

Fundamentals of Veterinary Clinical Pathology, Third Edition.
Edited by Steven L. Stockham and Michael A. Scott.

Companion website: https://www.wiley.com/go/fundamentals/vet

Table 1.1. Abbreviations and symbols in this chapter	
$[fT_4]_{ed}$	Free thyroxine concentration by equilibrium dialysis
[x]	x concentration (x = analyte)
BHB	β-Hydroxybutyrate
Ca^{2+}	Calcium
CBC	Complete blood count
CLSI	Clinical Laboratory Standards Institute
CV	Coefficient of variation
CV_A	Analytical variation
CV_G	Inter-individual (between-subject) biological variation
CV_I	Intra-individual (within-subject) biological variation
EDTA	Ethylenediaminetetraacetic acid
fCa^{2+}	Free ionized calcium
FN	False negative
FP	False positive
fT_4	Free thyroxine
II	Index of individuality
K_2EDTA	Dipotassium salt of ethylenediaminetetraacetic acid
K_3EDTA	Tripotassium salt of ethylenediaminetetraacetic acid
LRL	Lower reference limit
Na_2EDTA	Disodium salt of ethylenediaminetetraacetic acid
NIST	National Institute of Standards and Technology
NPV	Negative predictive value (predictive value of a negative test)
PPV	Positive predictive value (predictive value of a positive test)
QA	Quality assurance
QC	Quality control
QCM	Quality control material
RBC	Red blood cell, erythrocyte
RCV	Reference change value
RI	Reference interval
ROC	Receiver operating characteristic
SD	Standard deviation
SI	Système international d'unités (International System of Units)
tCa^{2+}	Total calcium
TE_a	Total error allowable (allowable total error)
TE_o	Total error observed (observed total error)
TN	True negative
TP	True positive
tT_4	Total thyroxine
U	International unit
URL	Upper reference limit
USD	Usual standard deviation
WBC	White blood cell, leukocyte
WRI	Within reference interval

Note: See Table 1.4 for abbreviations of units of measurement and the figure legends for abbreviations unique to figures.

CLINICAL PATHOLOGY

I. What is clinical pathology?
   A. Definitions
      1. *Pathology* is the study of disease. It deals with understanding the structural and functional changes causing or caused by disease.
      2. *Clinical pathology*, a subtype of pathology, is the study of disease in the clinical environment by use of laboratory assays. It deals with the use of laboratory testing (clinical chemistry, hematology, …) for the diagnosis and management of disease.
   B. As currently certified by the American College of Veterinary Pathologists, veterinary clinical pathologists are specialists in the disciplines of basic pathology, hematology (study of blood and bone marrow, including hemostasis), clinical chemistry (study of physiologic and biochemical reactions in blood or other body fluids), cytologic pathology (study of cells in tissues or body fluids), surgical pathology (study of disease via microscopic analysis of tissue samples obtained during surgery), and principles of laboratory medicine (equipment and methods, test properties, QA).
   C. Veterinary clinical pathologists, medical laboratory scientists (medical technologists, clinical laboratory scientists), medical laboratory technicians, and veterinary technicians often work in a clinical laboratory that limits its assay "menu" (offered tests) to hematologic assays including hemostasis tests, clinical chemical assays, urinalysis, and cytologic examinations. Other assays or diagnostic laboratory procedures may be offered but are often provided by other specific laboratories (e.g., microbiology, histopathology, and toxicology) that are commonly supervised by microbiologists, histopathologists, and toxicologists, respectively.

II. Laboratory tests should be used purposefully, judiciously, and along with other diagnostic procedures. Before laboratory tests are used to pursue a possible diagnosis, two diagnostic procedures are imperative: (i) obtain a complete history and (ii) perform a complete physical examination. With knowledge gained from these two basic procedures, a diagnostician can select appropriate diagnostic procedures to clarify or classify identified problems prior to developing or refining a differential diagnosis and pursuing further (or definitive) testing. Veterinarians frequently use laboratory assays in conjunction with other diagnostic methods to identify or classify pathologic states that develop in domestic mammals. Some body systems (e.g., integument, nervous, skeletal, and cardiovascular) are relatively easily evaluated via visual or imaging methods (physical examination, radiography, and ultrasonography), whereas other body systems (e.g., hemic, immune, urinary, and endocrine) are better evaluated by laboratory tests. Consideration of the following questions and associated material in this chapter and elsewhere[1] will help optimize selection of cost-effective tests that benefit patient management:
   A. Why is the test being ordered? Is it needed? How will the test results influence patient management or outcome?
   B. How well does the test discriminate between health and disease? How well does it discriminate among various disorders?
   C. How are the test results interpreted? Are appropriate reference values available?

III. What are the major reasons for analyzing patient samples via laboratory procedures?
   A. To detect an unidentified pathologic state
   B. To define, classify, or confirm a pathophysiologic disorder or disease state
   C. To eliminate a possible cause of the animal's illness
   D. To assess changes in a pathologic state either due to natural progression of the disease or because of medical or surgical intervention

SAMPLES

I. Whole blood, plasma, and serum
   A. Most clinical laboratory assays are designed to detect or quantify substances or cells in blood samples; the substance or cell of interest is called the *analyte* while the quantity that is measured is called the *measurand*. Obtaining useful results for the analyte requires appropriate samples. Whenever there is doubt about the appropriate sample for a particular test at a particular laboratory, the laboratory should be contacted prior to sample collection. (See Quality of Laboratory Results, Section I.B.)
   B. Whole blood (blood)
      1. Blood and its major components are frequently used as samples for laboratory assays. Blood must be collected and processed properly so that assay results reflect the true composition of blood rather than artifactual changes.
      2. Blood is composed of cells (RBCs, platelets, and five major WBC types) and plasma. Blood withdrawn from a blood vessel must immediately be mixed with an anticoagulant to prevent initiation of clot formation and to maintain cells and other components in suspension.
      3. Analysis or processing of whole blood must be relatively rapid because the cells deteriorate or die within a few hours, and thus a sample becomes unacceptable for analysis. What constitutes adequate sample handling varies with what is to be quantified or evaluated; occasionally, samples must be analyzed within minutes, usually within hours, rarely within days.
   C. Plasma
      1. *Plasma* is the fluid component of blood that is harvested after centrifugation of an anticoagulated blood sample. In vitro, plasma contains an anticoagulant that can interfere with some assays. Unlike serum, plasma contains fibrinogen.
      2. Anticoagulants used for blood sample collection
         a. Calcium-binding agents prevent Ca^{2+} from participating in the formation of a blood clot.
            (1) EDTA (as K_3EDTA, K_2EDTA, or Na_2EDTA)
               (a) EDTA is the preferred anticoagulant for almost all routine hematologic tests, including the complete blood count (CBC) assays.
               (b) EDTA chelates Ca^{2+} and other divalent cations (Mg^{2+}, Cu^{2+}, and Pb^{2+}) but the other anticoagulants do not. A *chelating agent* forms at least two bonds with the metallic ion it chelates. EDTA attaches to Ca^{2+} in six places.
               (c) Blood and anticoagulant should be mixed promptly with at least eight gentle and complete inversions of the tube; one inversion is defined as tipping the tube 180° and back again.
            (2) Citrate (as sodium citrate)
               (a) Citrate is the preferred anticoagulant for most tests of hemostasis. During coagulation tests, a specific amount of Ca^{2+} is added to the citrated plasma to override the effects of citrate and enable coagulation enzymes to function. Citrate's anticoagulant activity is achieved by its formation of a single ionic bond with Ca^{2+}.
               (b) Because it has low toxicity, citrate is also preferred for collection of whole blood to be used for transfusions.
               (c) Blood and anticoagulant should be mixed promptly with at least four gentle and complete inversions of the tube; one inversion is defined as tipping the tube 180° and back again.

(3) Oxalates (potassium salt is the most common)
(a) Oxalate is used for a few laboratory tests. It is the anticoagulant in sodium fluoride tubes that can be used for glucose and lactate assays. Generally, oxalates distort morphologic features of WBCs and RBCs and thus are unsuitable for hematologic samples.
(b) Oxalate's anticoagulant activity is achieved by its formation of an ionic bond with Ca^{2+}.

b. Heparin (as lithium, ammonium, potassium, or sodium salts) activates antithrombin (also called antithrombin III), which then inhibits the activity of coagulation enzymes, including thrombin. It also forms an ionic bond with Ca^{2+}, but its major action is through antithrombin.
(1) Heparin is used for several special laboratory assays (such as blood gas analysis) and for many clinical chemistry assays.
(2) Disadvantages
(a) May cause pink staining of cells and extracellular regions of blood smears, but it is commonly used for CBC samples from many nonmammalian species
(b) Allows minor clotting to occur over time as effects are slowly overridden by the coagulation system
(c) Allows or promotes platelet (thrombocyte) clumps to form
(3) Blood and anticoagulant should be mixed promptly with at least eight gentle and complete inversions of the tube; one inversion is defined as tipping the tube 180° and back again.

c. Some anticoagulant tubes contain a polymer gel that begins in the bottom of the tube but re-forms at the cell-plasma interface with centrifugation. This creates a barrier to decrease cell effects on plasma, facilitate pipetting plasma without cells, and increase plasma yield.

3. Plasma has two major components.
a. Water: about 92–95 % of plasma volume; 100 mL of plasma contains 92–95 mL of H_2O.
b. Solids: about 5–8 % of plasma volume; most solids are proteins on a weight per volume (weight/volume) basis. Other solids are glucose, urea, electrolytes, and other chemicals.

4. Generally, the chemical composition of plasma is very similar to interstitial fluid in most tissues. Within the body, plasma and interstitial fluid are the extracellular fluids, intravascular and extravascular, respectively.

D. Serum
1. Serum is the fluid component of blood that is harvested after centrifugation of a coagulated (clotted) blood sample in a blood tube commonly referred to as a clot tube. As described in Chapter 5, blood clotting involves platelets and coagulation proteins. To get the maximal amount of serum from a clotted sample, centrifugation should not be started prior to the retraction of the clot; this typically takes at least 30 min (45 min for equine samples) if a clot activator is not present in the tube. If samples are centrifuged prior to clot retraction, a soft fibrin clot resembling a veil may extend into the serum layer and reduce the serum yield; this is most often seen with equine samples.
2. Serum has essentially the same composition as plasma except serum does not contain most of the coagulation proteins. The major protein (on a weight/volume basis) that is absent in serum but present in plasma is fibrinogen. Fibrinogen is the only protein that contributes significantly to differences in paired serum and plasma total protein

concentrations, but such differences often occur for other reasons, including use of different methods (refractometry for plasma versus chemical methods for serum).

3. During the clotting process, substances released from cells may alter analyte concentrations in serum. For example, platelets release K^+, and thus serum [K^+] is greater than plasma [K^+] (see Chapter 9).
4. If the tube contains a clot activator, the tube should be gently inverted five times; one inversion is defined as tipping the tube 180° and back again.
5. Some clot tubes contain a polymer gel that begins in the bottom of the tube but re-forms at the cell-plasma interface with centrifugation. This creates a barrier to decrease ongoing cell effects on serum, facilitate pipetting serum without cells, and increase serum yield.

E. Order of blood sample collection for routine tubes: Blood samples are generally collected in the following order to avoid cross-contamination of tube additives and subsequent preanalytical error in results, but there can be reasons to alter the order for specific situations. Consider similar factors if collecting blood into other tube types.

1. Blood culture tube or bottle: to prevent contamination with potentially growth-inhibiting additives in other tubes
2. Sodium citrate tube (light blue cap): collected early to avoid (i) contamination by anticoagulants or clot activators from other blood tubes and (ii) activation of coagulation prior to sample anticoagulation
3. Serum tubes, including those with clot activator and gel (red, gold, or mottled caps): to avoid contamination with anticoagulants that may alter results (e.g., contaminating K_3EDTA may falsely increase [K^+] and falsely decrease [tCa^{2+}])
4. Sodium or lithium heparin tube, with or without gel (green, light green, or speckled green cap): to avoid contamination with EDTA that may alter results as noted for serum
5. EDTA tube (lavender or pink cap)

II. Urine samples

A. Other than blood, urine is the most common sample analyzed by laboratory assays. As with blood, urine must be collected and processed properly so that the assay results reflect the true composition of the product of the urinary system.

B. To prevent artifactual changes in urine, it should be processed soon after collection. General guidelines for the collection and processing of urine for routine analyses are described in Chapter 8.

III. Other body fluid samples

A. Pleural fluid, peritoneal fluid, pericardial fluid, synovial fluid, and cerebrospinal fluid samples are collected to characterize body cavity effusions, joint diseases, and central nervous system disorders, respectively.

B. The processing and handling of cavitary fluids are described in Chapter 19, and synovial fluid and cerebrospinal fluid analyses are described in Chapters 20 and 21, respectively.

IV. Nonfluid samples for cytologic or histologic evaluation

A. Air-dried cytologic preparations can be obtained from any tissue in the body to characterize various processes and lesions. Cells are collected by needle, scraping, swabbing, or imprinting, depending on the nature of the lesion. Such sampling of lymph nodes and bone marrow is described in Chapter 6.

B. Antemortem histologic samples of lesions are collected by various incisional or excisional techniques followed by fixation and processing for sectioning. Collection of bone marrow histologic samples (cores) is described in Chapter 6.

COMMON TYPES OF CLINICAL PATHOLOGY LABORATORY ASSAYS

I. Many laboratory tests or assays involve the analysis of body fluids (whole blood, serum, plasma, urine, peritoneal fluid, pleural fluid, pericardial fluid, cerebrospinal fluid, and synovial fluid), tissue samples, or feces. Most clinical laboratory procedures fall into one of the three large groups below (examples follow subdivisions); many procedures could be classified into more than one group.
   A. Clinical hematology assays: Most assays are completed on whole blood samples.
      1. Quantitation of cell concentrations in blood: total WBC concentration (count), [RBC], and [platelet]
      2. Semiquantitation of cell concentrations: calculated [WBC] and platelet estimate from blood film examination
      3. Defining or classifying cells by microscopic features: toxic neutrophils, reactive lymphocytes, polychromatophilic RBCs, poikilocytes, microcytes, hypochromic RBCs, and leukemic cells
      4. Assessing the hemostatic properties of blood: clotting times and platelet function assays
      5. Assessing hematopoiesis and bone marrow health: bone marrow aspirate and core biopsies
   B. Clinical chemistry assays: Most assays are completed on serum or plasma samples.
      1. Detecting or quantifying the concentration of a chemical substance
         a. Quantitative analysis: Results are close to the true concentration (e.g., serum concentrations of glucose, sodium, protein, creatinine, and urea) and typically are reported with a specific numerical value (see the Significant Digits section).
         b. Semiquantitative analysis: Results are "within the ballpark" (e.g., urine glucose, protein, and bilirubin concentrations by reagent pad chemistry assays) and may be reported as approximate numerical values or in a categorical scale (e.g., 1+, 2+, or 3+) that represents ranges of numerical values.
         c. Qualitative analysis: Results indicate that a substance is or is not present (e.g., fat detected microscopically in pleural fluid with a Sudan stain) and are reported as "present" or "absent" or as "positive" or "negative."
      2. Detecting or quantifying the activity of a chemical substance
         a. Quantitative analysis: Results are close to true activity (e.g., measured activities of serum enzymes such as alanine aminotransferase, alkaline phosphatase, lactate dehydrogenase, and creatine kinase).
         b. Qualitative analysis: Results indicate that activity is or is not present (e.g., heme's peroxidase activity or leukocyte esterase activity in urine).
   C. Clinical microscopy
      1. Clinical cytology: the study of cell populations and their microscopic features in an attempt to define or classify abnormal tissue or fluid (e.g., fine-needle biopsy smears from lymph nodes to diagnose lymphoma or histoplasmosis, fine-needle biopsy smears of a skin mass to determine whether it is an inflammatory or neoplastic lesion, and analysis of peritoneal fluid to determine if it is a transudate, exudate, or other type of effusion)
      2. Surgical histopathology: the study of frozen or fixed tissue to define or classify abnormal tissue (i.e., inflammatory, neoplastic, and toxic disorders) and perhaps establish an etiologic diagnosis
      3. Urine sediment analysis: microscopic examination of wet or air-dried urine preparations to detect or semiquantify the presence of WBCs, RBCs, casts, bacteria, crystals, or other structures
      4. Clinical parasitology: microscopic analysis of fecal, urine, blood, or other samples to detect ova, larvae, or other microscopic forms of parasites
      5. Clinical hematology: noted above in Section I.A

II. Actual descriptions of the numerous laboratory methods are beyond the scope of this textbook. However, an understanding of basic principles and methods is frequently needed to interpret results of a laboratory assay. Such principles are located in several parts of the textbook, and specific methods can be provided by laboratory personnel or manuals.
   A. Chapters 2–5 contain the basic principles and concepts of the common hematologic assays.
   B. Chapter 6 contains the guiding principles and overview of the analysis of bone marrow and lymph nodes, the two major tissues involved in hematopoiesis.
   C. Chapters 7–18 have short sections that describe analytical principles that apply to other individual analytes.
   D. Chapters 19–21 contain basic principles pertaining to the analysis of cavitary effusions, synovial fluid, and cerebrospinal fluid.

SIGNIFICANT DIGITS (SIGNIFICANT FIGURES)

I. Results of many laboratory assays consist of a number, often calculated, accompanied by a unit of measure and RIs. Although many reporting systems do not allow it, the number should be reported to the appropriate number of digits by following three basic rules for reporting significant digits.
   A. Retain only as many significant digits in a result as will give only one uncertain digit (Table 1.2). A digit is uncertain when the associated measurement cannot reliably

Table 1.2. Examples of significant digits

Number	Significant digits	Reason for the given number of significant digits
124	3	All digits imply significance.
124.0	4	The zero was not needed for the whole number. Its addition indicates that it is a significant digit.
120	2 or 3	If the zero is just preserving space, it is not a significant digit. If values are usually reported to the ones place in a particular setting (e.g., 121), then the zero is a significant digit.
120.0	4	The zero in the tenths space indicates it is a significant digit. Thus, the zero in the ones place is also a significant digit.
0.12	2	The zero is only preserving a space and is not a significant digit. It is often written by convention such that the "." is recognized as a decimal point and not a period, fly dirt, or other extraneous material.
0.02	1	The zero in the tenths place is just filling space and is not a significant digit.
0.020	2	The zero in the thousandths place indicates it is significant.

Note: The number of significant digits is determined by starting with the left-most nonzero digit and counting until reaching the last significant digit. If numbers are properly reported, nonzero digits are always significant, as are all zeros between other significant digits and any zeros that are to the right of the decimal point and terminal. If no decimal point is present, any terminal zeros after the right-most nonzero digit may or may not be significant (such as those in 17,000).

differentiate the value in that digit from one above or one below it. Zero is not a significant digit if it only preserves a space in the number.

B. When reporting results from multiplications or divisions that use two or more numbers with appropriate significant digits, the final calculated result should have no more significant digits than the number(s) with the fewest significant digits.
   1. For $1.23 \times 2.4 = 2.952$, the product should be reported as 3.0 because 2.4 has only two significant digits.
   2. Multiplying 0.03 and 0.12, which have one and two significant digits, respectively, yields 0.0036, but this should be reported to one significant digit: 0.004.
   3. If MCV = 71 fL and [RBC] = 3.8×10^9/L, the analyzer's Hct = $71 \times 3.8 / 10 = 26.98$ %, which is 27 % when reported appropriately to two significant digits.

C. When reporting results from additions or subtractions that use two or more numbers with appropriate significant digits, the final calculated result should have no more decimal places than the number(s) with the fewest decimal places. For $1.23 + 2.4 = 3.63$, the sum should be reported as 3.6.

II. As in the previous examples, it is frequently necessary to round numbers so that only significant digits are reported. There are various rules for rounding numbers when the last digit is 5, some being to round up the preceding digit whether it is even or odd. However, other rules are used to remove upward or downward rounding bias in data sets, as follows.

A. Add 1 to the last retained digit if it is odd (round up, example C.1).

B. Do not add 1 to the last retained digit if it is even (round down, example C.2).

C. Examples with rounding to two significant digits: The third example (C.3) does not require the rounding rule because 4 is the digit to be rounded – not 5 – when rounding to two significant digits; scientific notation is used to clarify that the zeroes following the significant digits in 1100, 1200, and 1300 are not significant digits.
   1. $1150 \rightarrow 1.2 \times 10^3$
   2. $1250 \rightarrow 1.2 \times 10^3$ (would round up to 1.3×10^3 with some rules)
   3. $1145 \rightarrow 1.1 \times 10^3$

III. Using the rules of significant digits for reporting results is one way that laboratory personnel can communicate the known precision and measurement uncertainty of the laboratory's assay result. The first rule of significant digits (see Significant Digits, Section I.A) is to retain only one uncertain digit. The following illustrates how that uncertain digit can be determined.

A. One sample is analyzed many times (e.g., 20). The mean and SD of the data are calculated.

B. If, for a given assay, the mean value was 125.345 and the SD was 0.578, then the interval that represents the mean ± 2 SD is 124.189–126.501. With this degree of analytical precision, the assay can readily distinguish between 100 and 200 and between 120 and 130, but it may not be able to distinguish consistently between 125 and 126. Thus, the value in the ones place is uncertain and the measured value should be reported to the nearest whole number (i.e., the mean = 125).

C. If, for a second assay, the mean value was 125.345 and the SD was 9.111, then the interval that represents the mean ± 2 SD is 107.123–143.567. With this degree of analytical precision, the assay can readily distinguish between 100 and 200, but it probably cannot reliably distinguish between 120 and 130. Thus, the tens value is uncertain, and the measured value should be reported to the nearest tens unit, making the mean value 120 or 130, depending on the rounding rule used for 5s (see preceding section).

D. Understanding significant digits becomes important when laboratory data are reported and interpreted. If one understands that we are uncertain about the last reported significant digit, then it is less likely that a 120 in today's sample and a 130 in tomorrow's sample will be considered a true biological change. Unfortunately, there are several reasons that all results are not reported this way:
   1. It may not be the policy for some laboratories.
   2. Analyzers and data management programs may not follow rules of significant digits and may not be modifiable.
   3. The appropriate number of places after the decimal point may vary for a particular analyte, depending on the magnitude of the value, but analyzers and software may not allow such variability. For example, precision is usually worse at low concentrations, so the appropriate number of significant digits may be greater at higher concentrations. However, for any analyte, most laboratories report all values to the same apparent degree of precision relative to the decimal point. This may lead to overestimation of the precision of smaller measurements or underestimation of the precision of larger measurements. An example is shown in Figure 1.1.

Measured values

WBC concentration = 21.5 × 10^9/L

WBC diff. count: 73 % neutrophils, 12 % lymphocytes, 9 % monocytes, 4 % eosinophils, 2 % basophils

Calculations and resultant significant figures

Neutrophils: (**21.5** × 10^9/L) × 0.**73** = 15.695 × 10^9/L = **16** × 10^9/L

Lymphocytes: (**21.5** × 10^9/L) × 0.**12** = 2.58 × 10^9/L = **2.6** × 10^9/L

For these results: 3 *significant figures* × 2 *significant figures* = 2 *significant figures*

Monocytes: (**21.5** × 10^9/L) × 0.0**9** = 1.935 × 10^9/L = **2** × 10^9/L

Eosinophils: (**21.5** × 10^9/L) × 0.0**4** = 0.86 × 10^9/L = 0.**9** × 10^9/L

Basophils: (**21.5** × 10^9/L) × 0.0**2** = 0.43 × 10^9/L = 0.**4** × 10^9/L

For these results: 3 *significant figures* × 1 *significant figure* = 1 *significant figure*

However, most computer-based reporting systems require reporting analytes' values consistently to a defined decimal point and thus significant figure rules will not be consistently applied.

	Results using significant figure rules	Results using consistent decimal points
Total WBC	21.5 × 10^9/L	21.5 × 10^9/L
Neutrophils	16 × 10^9/L	16.0 × 10^9/L
Lymphocytes	2.6 × 10^9/L	2.6 × 10^9/L
Monocytes	2 × 10^9/L	2.0 × 10^9/L
Eosinphils	0.9 × 10^9/L	0.9 × 10^9/L
Basophils	0.4 × 10^9/L	0.4 × 10^9/L

Figure 1.1. Significant digits. The significant digits of the measured and calculated values of a leukogram are provided. When two values are multiplied, the result should have no more significant digits than the number with the fewest significant digits (see I.B in this section). However, computer-based reporting systems may not enable the flexibility to report only significant digits.
diff., differential

UNITS

I. SI units versus nonSI (conventional) units

A. For several decades, there has been an attempt to switch to a metric system of units throughout the world. Other than the United States of America, the conversion is mostly complete in a modified practical form. During the 1970s and 1980s, several organizations in the United States attempted to convert members of the medical communities to SI units but had limited success. Because there is a lack of consistent use of the SI unit system, veterinary medical professionals need to be familiar with both SI and nonSI units.

B. In the context of units used for laboratory data, the basic units of measurement are listed in Table 1.3. Many clinical laboratory and professional organizations have agreed to use liter as the preferred unit for volume instead of the SI unit of cubic meter because the latter is rarely clinically relevant (1 m^3 = 1000 L). Even the use of liter for a volume unit has limited relevance in the clinical laboratory when the sample volume for many assays is less than 0.1 mL.

C. Because the reported units for amount or concentration (a concentration is an amount per volume) of substances vary considerably, a veterinary medical professional should know the common abbreviations for the major units (Table 1.4). A common error is to translate dL to decaliter (10 L) rather than deciliter (0.1 L).

D. The National Institute of Standards and Technology (NIST) provides rules and style conventions for the use of the SI units to reduce ambiguity in scientific communications. Examples of the NIST unit conventions used in this book are in Table 1.5.

E. Table 1.6 contains the formulas for the conversion of analyte values from nonSI units to SI units; only analytes presented in this textbook are included. The table contains two types of formulas: (i) formulas that show the simple conversion factor that is used to calculate the numerical value of the SI unit,[2] and (ii) formulas that show the conversion of the numerical value and units.

Table 1.3. Examples of measurement in SI units and conventional units

	SI units	Conventional units
Amount of substance	mole (mol)	gram (g)
Length	meter (m)	yard, foot, inch
Mass	kilogram (kg)	pound (lb), grain
Time	second (s)	minute (min), hour (h)
Volume	cubic meter (m^3)	liter (L)

Table 1.4. Common units and abbreviations for laboratory values

Number[a]		Volume		Mass	
—	—	—	—	kg	kilogram, 10^3 g
mol	mole	L	liter	g	gram
—	—	dL	deciliter, 10^{-1} L	—	—
mmol	millimole, 10^{-3} mol	mL	milliliter, 10^{-3} L	mg	milligram, 10^{-3} g
µmol	micromole, 10^{-6} mol	µL	microliter, 10^{-6} L	µg	microgram, 10^{-6} g
nmol	nanomole, 10^{-9} mol	nL	nanoliter, 10^{-9} L	ng	nanogram, 10^{-9} g
pmol	picomole, 10^{-12} mol	pL	picoliter, 10^{-12} L	pg	picogram, 10^{-12} g
fmol	femtomole, 10^{-15} mol	fL	femtoliter, 10^{-15} L	fg	femtogram, 10^{-15} g

[a] Based on moles, 1 mole being Avogadro's number ($\sim 6.022 \times 10^{23}$) of particles of a substance

Table 1.5. Examples of NIST style for writing units compared to other styles[a]

Rules	NIST	NonNIST
Symbols unaltered in the plural	20 pg 4 h	20 pgs 4 hrs
Symbols not followed by period unless end of sentence	2 L 6 yr	2 L. 6 yr.
There is a space between the number and the unit	10 % 37 °C 15 g/dL	10% 37°C, 37° C 15g/dL
No words are inserted between the number and the unit	3000/µL	3000 WBC/µL

[a] The NIST system has two styles for a range of numbers, e.g., 5 % to 10 % or (5 to 10) %. To save space and to conform to common use in veterinary medicine, the numbers are displayed as follows in this textbook: 5–10 %.

II. Amount versus concentration: One important basic concept for interpreting laboratory data is having a clear understanding of what a laboratory test result represents. Besides knowing what is really being measured, it is important to understand what the numbers and units represent. A number without a unit is meaningless. The following examples illustrate the concepts:
   A. A dog acutely lost a large amount of blood because of an injury. Because whole blood including RBCs was lost, the number of RBCs in the body is decreased. However, because plasma was lost with the RBCs, the [RBC] (number of RBCs per volume of blood) in the dog initially is not decreased, and thus the dog is not initially anemic. After fluid shifts restore plasma volume, the dog will have fewer RBCs in its body and a lower [RBC] in its blood.
   B. You are told that a cat's serum sodium level was increased. Does this mean the cat has more sodium in its body? Well, it might. However, the increased serum [Na^+] might be due to less water in the body, and the amount of sodium may not be increased. In fact, the total amount of sodium in the body could be decreased if there was relatively more water loss than sodium loss.
   C. You are told that a horse's serum enzyme level was decreased. Does this mean the horse has less of that enzyme? It might. However, it could be that the amount of enzyme (the protein) was not decreased, but the enzyme's activity was inhibited, or the structure of the enzyme was defective.
   D. You are told that a cat's reticulocyte percentage is increased. Does this mean that the cat has more reticulocytes in its blood? It might or might not. A percentage is always relative; the same number in the numerator (e.g., number of reticulocytes counted) and a smaller number in the denominator (e.g., total number of RBCs counted) results in an increased percentage, so there might be fewer RBCs rather than more reticulocytes.
   E. You are told that the myeloid to erythroid ratio in a cat's marrow is increased. Does the increased ratio mean the cat's marrow contains more myeloid cells, fewer erythroid cells, or both? Or is the ratio increased because there are more myeloid and more erythroid cells, but the number of myeloid cells is increased more than the increase in erythroid cells? A calculated ratio is always a relative number and must be interpreted accordingly.
   F. You are told that a dog's urine has an increased [protein]. Because the concentrations of all substances in urine depend on the conservation of water by the kidneys, the increased [protein] could have resulted from increased water conservation and not increased protein loss via the urinary system.

Table 1.6. Conversion of nonSI units to SI units

Analyte[a]	To convert		Multiply by[b]	Complete conversion formulas	Increment[c]
	From	To			
Acetoacetate	mg/dL	µmol/L	99	mg/dL × 9.9 µmol/mg × 10 dL/L = µmol/L	—
ACTH	pg/mL	pmol/L	0.2202	pg/mL ÷ 4541 pg/pmol × 1000 mL/L = nmol/L	1 pmol/L
Albumin	g/dL	g/L	10	g/dL × 10 dL/L = g/L	1 g/L
Aldosterone	ng/dL	pmol/L	27.74	µg/dL ÷ 360.5 µg/µmol × 1000 nmol/µmol × 10 dL/L = nmol/L	10 pmol/L
β-Hydroxybutyrate	mg/dL	µmol/L	97	mg/dL × 9.7 µmol/mg × 10 dL/L = µmol/L	—
Bile acid (total)	mg/L	µmol/L	2.547	mg/L ÷ 392.6 µg/µmol × 1000 µg/mg = µmol/L	0.2 µmol/L
Bile acid (total)	mg/mL	mmol/L	2.547	mg/mL ÷ 392.6 mg/mmol × 1000 mL/L = mmol/L	0.2 mmol/L
Bt	mg/dL	µmol/L	17.10	mg/dL ÷ 584.8 mg/mmol × 1000 µmol/mmol × 10 dL/L = µmol/L	2 µmol/L
Cholesterol	mg/dL	mmol/L	0.02586	mg/dL ÷ 386.7 mg/mmol × 10 dL/L = mmol/L	0.05 mmol/L
Cl^-	mEq/L	mmol/L	1	mEq/L × 1 mmol/mEq = mmol/L	1 mmol/L
Cl^-	mg/dL	mmol/L	0.2817	mg/dL ÷ 35.5 mg/mmol × 10 dL/L = mmol/L	1 mmol/L
COP	mmHg	pascals	133.322	mmHg × 133.322 pascals/mmHg = pascals	—
Cortisol	µg/dL	nmol/L	27.59	µg/dL ÷ 362.45 µg/µmol × 1000 nmol/µmol × 10 dL/L = nmol/L	10 nmol/L
Creatinine	mg/dL	µmol/L	88.4	mg/dL ÷ 113.1 mg/mmol × 1000 µmol/mmol × 10 dL/L = µmol/L	10 µmol/L
Cyanocobalamin	pg/mL	pmol/L	0.7378	pg/mL ÷ 1355 pg/pmol × 1000 mL/L = pmol/L	10 pmol/L
fCa^{2+}	mg/dL	mmol/L	0.2495	mg/dL ÷ 40.08 mg/mmol × 10 dL/L = mmol/L	0.01 mmol/L
fCa^{2+}	mEq/L	mmol/L	0.5	mEq/L × 0.5 mmol/mEq = mmol/L	0.01 mmol/L
Fe	µg/dL	µmol/L	0.1791	µg/dL × 0.01791 µmol/µg × 10 dL/L = µmol/L	1 µmol/L
Ferritin	ng/mL	µg/L	1	ng/mL × 1 µg/1000 ng × 1000 mL/L = µg/L	10 µg/L
Fibrinogen	mg/dL	g/L	0.01	mg/dL × g/1000 mg × 10 dL/L = g/L	0.1 g/L
Fibrinogen	mg/dL	µmol/L	0.0294	mg/dL × 0.00294 µmol/mg × 10 dL/L = µmol/L	—
Folate	ng/mL	nmol/L	2.266	ng/mL ÷ 441.3 ng/nmol × 1000 mL/L = nmol/L	2 nmol/L
fT_4	ng/dL	pmol/L	12.87	ng/dL ÷ 777 ng/nmol × 1000 pmol/nmol × 10 dL/L = pmol/L	1 pmol/L
Globulins	g/dL	g/L	10	g/dL × 10 dL/L = g/L	1 g/L
Glucose	mg/dL	mmol/L	0.05551	mg/dL ÷ 180.1 mg/mmol × 10 dL/L = mmol/L	0.1 mmol/L

Continues

Table 1.6. Conversion of nonSI units to SI units (Continued)

Analyte[a]	To convert		Multiply by[b]	Complete conversion formulas	Increment[c]
	From	To			
HCO_3^-	mEq/L	mmol/L	1	mEq/L × 1 mmol/mEq = mmol/L	1 mmol/L
Hgb	g/dL	g/L	10	g/dL × 10 dL/L = g/L	1 g/L
IRG	pg/mL	ng/L	1	pg/mL ÷ 1000 pg/ng × 1000 mL/L = ng/L	10 ng/L
IRI	µU/mL	pmol/L	7.175	µU/mL ÷ 139.4 µU/pmol × 1000 mL/L = pmol/L	5 pmol/L
IRI	µg/L	pmol/L	172.2	µg/L ÷ 5.807 µg/µmol × 1000 pmol/µmol = pmol/L	5 pmol/L
K^+	mEq/L	mmol/L	1	mEq/L × 1 mmol/mEq = mmol/L	0.1 mmol/L
K^+	mg/dL	mmol/L	0.2564	mg/dL ÷ 39 mg/mmol × 10 dL/L = mmol/L	0.1 mmol/L
L-Lactate	mg/dL	mmol/L	0.112	mg/dL × 1.11 mmol/mg × 10 dL/L = mmol/L	—
MCHC	g/dL	g/L	10	g/dL × 10 dL/L = g/L	—
Na^+	mEq/L	mmol/L	1	mEq/L × 1 mmol/mEq = mmol/L	1 mmol/L
Na^+	mg/dL	mmol/L	0.4348	mg/dL ÷ 23 mg/mmol × 10 dL/L = mmol/L	1 mmol/L
NEFA	mEq/L	mmol/L	1	mEq/L × mmol/mEq = mmol/L	—
NH_3	µg/dL	µmol/L	0.5871	µg/dL ÷ 17.03 µg/µmol × 10 dL/L = µmol/L	5 µmol/L
NH_4^+	µg/dL	µmol/L	0.5543	µg/dL ÷ 18.04 µg/µmol × 10 dL/L = µmol/L	5 µmol/L
Pi	mg/dL	mmol/L	0.3229	mg/dL ÷ 30.97 mg/mmol × 10 dL/L = mmol/L	0.05 mmol/L
Platelet	#000/µL	# × 10^9/L	10^9	#000/µL × 10^6 µL/L = # × 10^9/L	10 × 10^9/L
RBC	# × 10^6/µL	# × 10^{12}/L	10^6	# × 10^6/µL × 10^6 µL/L = # × 10^{12}/L	—
SDMA	µg/dL	µmol/L	0.0495	µg/dL ÷ 202.25 µmol/µg × 10 dL/L = µmol/L	—
T_3	ng/dL	nmol/L	0.01536	ng/dL ÷ 651 ng/nmol × 10 dL/L = nmol/L	0.1 nmol/L
T_3	pg/dL	nmol/L	15.36	pg/dL ÷ 651 pg/pmol × 1000 pmol/nmol × 10 dL/L = nmol/L	0.1 nmol/L
tCa^{2+}	mg/dL	mmol/L	0.2495	mg/dL ÷ 40.08 mg/mmol × 10 dL/L = mmol/L	0.02 mmol/L
tCa^{2+}	mEq/L	mmol/L	0.5	mEq/L × 0.5 mmol/mEq = mmol/L	0.02 mmol/L

TIBC	µg/dL	µmol/L	0.1791	µg/dL × 0.01791 µmol/µg × 10 dL/L = µmol/L	1 µmol/L
tMg^{2+}	mg/dL	mmol/L	0.4114	mg/dL ÷ 24.31 mg/mmol × 10 dL/L = mmol/L	0.02 mmol/L
tMg^{2+}	mEq/L	mmol/L	0.5	mEq/L × 0.5 mmol/mEq = mmol/L	0.02 mmol/L
Total protein	g/dL	g/L	10	g/dL × 10 dL/L = g/L	1 g/L
Triglyceride	mg/dL	mmol/L	0.01129	mg/dL ÷ 885.7 mg/mmol × 10 dL/L = mmol/L	0.05 mmol/L
TSH	µg/dL	ng/mL	10	µg/dL × 1000 ng/µg ÷ 100 mL/dL = ng/mL	—
tT_4	µg/dL	nmol/L	12.87	µg/dL ÷ 777 µg/µmol × 1000 nmol/µmol × 10 dL/L = nmol/L	1 nmol/L
tT_4	ng/mL	nmol/L	1.287	ng/mL ÷ 777 ng/nmol × 1000 mL/L = nmol/L	1 nmol/L
UN	mg/dL UN	mmol urea/L	0.3570	mg/dL ÷ 28.01 mg/mmol × 10 dL/L = mmol of urea/L	0.5 mmol/L
Urea	mg/dL urea	mmol urea/L	0.1665	mg/dL ÷ 60.06 mg/mmol × 10 dL/L = mmol of urea/L	0.5 mmol/L
WBC	#000/µL	× 10^9/L	0.001	#000/µL × 10^6 µL/L = #000 × 10^6/L = # × 10^9/L	—
Xylose	mg/dL	mmol/L	0.06661	mg/dL ÷ 150.1 mg/mmol × 10 dL/L = mmol/L	0.1 mmol/L

[a] ACTH, adrenocorticotropic hormone; Bt, total bilirubin; Cl^-, chloride; COP, colloidal osmotic pressure; Fe, iron; HCO_3^-, bicarbonate; Hgb, hemoglobin; IRG, immunoreactive glucagon; IRI, immunoreactive insulin; K^+, potassium; MCHC, mean cell hemoglobin concentration; Na^+, sodium; NEFA, nonesterified fatty acids; NH_3, ammonia; NH_4^+, ammonium; Pi, inorganic phosphorus; SDMA, symmetrical dimethylarginine; T_3, triiodothyronine; TIBC, total iron-binding capacity; tMg^{2+}, total magnesium; TSH, thyroid-stimulating hormone; and UN, urea nitrogen

[b] *Source*: Lundberg et al.[2]

[c] Recommended smallest reporting interval of the SI unit,[2] which may differ with magnitude, species, and assay precision; a reporting interval or increment of 1 g/L means that results would be reported to the nearest 1 g/dL with no decimal places to suggest less uncertainty than is actually present

ANALYTICAL AND BIOLOGICAL VARIATION

I. Appropriate use of laboratory results to monitor disease or differentiate disease from health requires an understanding of analytical variation, biological variation, and the potential contribution of various types of error to variations in results.

II. Variation in a set of data from an individual or population can be expressed as a standard deviation (SD, σ) or as a coefficient of variation (CV).
   A. SD describes the dispersion of a dataset relative to its mean and is calculated as the square root of the variance, which is the average of the squared differences of each value from the mean. The greater the SD, the more the data are spread from the mean.
   B. The CV of a set of measures is the SD reported as a percentage of the mean (Eq. 1.1). The greater the CV, the greater the variation in the data. Analytical and biological variation are commonly described in terms of CV.

$$\text{CV (expressed as a percentage)} = \frac{\text{standard deviation}}{\text{mean}} \times 100 \qquad (\mathbf{1.1})$$

III. Analytical variation is the variation in results caused by imprecision or inaccuracy of analytical methods.
   A. *Imprecision*, commonly denoted as the CV of repeated measures, is the unavoidable random error that occurs during analysis because of variations in reagent mixing, temperatures, electrical supply, operator, timing, pipetting, etc.
   B. *Bias*, or *systematic error*, is the inaccuracy of results caused by differences in instruments or methods rather than random variations and chance. When results from different methods or instruments are compared, method and instrument bias contribute to analytical variation.

IV. *Biological variation*, the inherent variation of repeated measures within a healthy individual or population, impacts many fundamental areas of laboratory medicine including method selection, quality goals (TE_a), QC testing, result reporting, and use of clinical decision limit guidelines.[3,4] There are different types.
   A. *Intra-individual (within-subject) variation*, the variation of a test result from an individual, is commonly denoted CV_I. This fluctuation around an individual's homeostatic set point occurs because of physiologic changes. Predictable physiologic variations such as circadian, reproductive, or seasonal changes are typically addressed separately by appropriate timing of sample collection and interpretation based on expected physiologic differences or a similar reference population.
   B. *Inter-individual (between-subject) variation*, the variation of test results within a population or group, is commonly denoted CV_G. This variation is a result of different homeostatic set-points among individuals of the group. Effects of age, sex, breed, reproductive status, and geographic location can contribute. The more broadly a population is defined, the greater the expected variation. Consider the anticipated inter-individual variation in blood values among all dogs (larger CV_G) compared to the inter-individual variation in blood values among laboratory-bred beagles (smaller CV_G). Inter-individual variation is reflected by the range (high minus low) of population-based RIs.

C. Intra-individual and inter-individual variation are typically determined by repeated measurement of an analyte in defined populations, and because a sample and a measurement are involved, the measurement of biological variation is affected by preanalytical and analytical variations. Preanalytical variations arise from preparation of the individual, sample collection, and the conditions from collection to testing, all factors that can be controlled by adherence to defined processes. Analytical variation for a particular assay is the imprecision of that assay's measurement (CV_A).

D. *Index of individuality (II or IoI, lol!)*: If assay imprecision (CV_A) is relatively small compared to CV_I and CV_G, which is the case for high-quality, well-controlled analyses, then the II of an analyte is the ratio of its intra-individual biological variation to its inter-individual biological variability: CV_I/CV_G.
   1. A high II (e.g., > 1.4) means that there is more intra-individual variation than inter-individual variation, so there is low individuality (individuals do not differ a lot from the group).
   2. A low II (high inverse II) means there is less intra-individual variation than inter-individual variation, so each individual's value fluctuates relatively mildly around a set point that varies more among individuals. A low II therefore reflects greater individuality.
   3. Note: The *inverse index of individuality*, which is determined by switching the numerator and denominator in the II calculation, is more intuitive, because a greater inverse II reflects greater individuality ($CV_I < CV_G$), and vice versa.

E. Procedures for determining CV_I, CV_G, II, and homeostatic set points are available,[3,5] and estimates of biological variation of various analytes in various species have been reported (www.vetbiologicalvariation.org). However, estimates of CV_I, CV_G, CV_A, and II vary considerably, thus limiting clinical application.

V. Impact of variation on data interpretation

A. Detecting disease by laboratory tests is based on recognizing when test results exceed expected degrees of variation, i.e., values are higher (increased) or lower (decreased) than expected. Recognition of unexpected findings is based on RIs that describe the combined biological variation in the reference population and the analytical variation of the assay. Unexpected values may reflect disease, an unusual biological variation, or an error (see Quality of Laboratory Results, I. Sources of Error).

B. Population-based RIs (see Reference Intervals, next section), which reflect the variation in analyte results across a group of subjects within a reference population, are commonly used to identify unusual results in individuals. These RIs work well for analytes with low individuality (individuals do not differ a lot from the group). However, population-based RIs are problematic for analytes with a high degree of individuality, because clinically important changes for an individual can be inapparent when values are compared to the broad population-based RIs that reflect the relatively large variation of the group (CV_G). For analytes with high individuality (e.g., creatinine), it is better to compare results to the individual's prior results, ideally by establishing a homeostatic set point and reference change value.[5,6] The number of samples needed to do so depends upon an analyte's CV_I and CV_A.

C. Reference change value (RCV) (also known as critical difference):[3,5,7] A RCV is the magnitude of change between serial results that is great enough to be considered of potential clinical significance, i.e., the change is greater than expected for estimates of biological and analytical variation. The value assumes negligible preanalytical variation. Changes in an analyte that exceed a provided RCV may support that a patient's status is improving, stable, or declining. Use of RCVs is limited.

REFERENCE INTERVALS

I. Reference intervals and their purpose
   A. Results of laboratory tests (laboratory data) on patient samples would be very difficult to interpret without *reference intervals*, which are usually population-based and reflective of the results we expect to find in most healthy animals of that population. These intervals are primarily used to help detect pathologic states. Other terms that are used as less-desirable synonyms include *normals*, *normal values*, *normal range*, and *reference range*.
   B. The following terms and definitions have been approved by the International Federation of Clinical Chemistry, the World Health Organization, and other groups.[8]
      1. *Reference individual*: an animal selected by using defined criteria
      2. *Reference population*: the group of all possible reference individuals
         a. Usually, the number of such individuals is unknown.
         b. In the case of captive wild animals, the total number of animals may be known.
      3. *Reference sample group*: an adequate number of reference individuals selected to represent the reference population
      4. *Reference value*: a value (result) obtained by observation or measurement of a particular substance in a reference individual
      5. *Reference distribution*: the distribution of reference values, which is sometimes Gaussian (a bell-shaped curve) and sometimes skewed
      6. *Reference limits*: the lowest value (LRL) and the highest value (URL) of the RI, as derived from a reference distribution
      7. *Reference interval*: an interval between and including the two reference limits (when only an upper limit is provided, e.g., < 3.5 µg/L, the lower limit of the RI is 0 or the detection limit of the assay)
      8. *Observed value*: a value obtained by observation or measurement that is to be compared to the RI
   C. Use of the term *reference range* is common but discouraged for two reasons.
      1. Statistically, a *range* is the difference between highest and lowest observations rather than the interval itself, e.g., the range is 40 if the highest observation was 50 and the lowest was 10.
      2. It may be ambiguous whether "reference range" refers to the RI or includes low-to-high values for all the reference individuals, about 5 % of which are excluded from typical RIs.
   D. Using the terms *normal* and *abnormal* to describe laboratory test results is common but can be misleading and is discouraged.
      1. A laboratory result can be WRI but still reflect a pathologic process, so it would not reflect normality in that individual. For example, a serum [Na^+] that is WRI in a dehydrated animal indicates that the animal has lost both water and sodium from its body.
      2. Healthy animals are expected to have some laboratory results outside of the RI because typical RIs exclude approximately the upper and lower 2.5 % of reference values. Such excluded values are outside the RI but do not reflect an abnormality.
      3. Because RIs are determined from a subset of the reference population, reference limits are merely a statistical estimate of the limits of health, but they do not define health or normality.
      4. It is difficult to define "normal" because many variations that may appear to be "abnormal" are caused by physiologic, dietary, environmental, or other nonpathologic factors.

II. Establishment of reference intervals: Establishing de novo (from scratch) RIs is a major laboratory undertaking that should not be approached without consideration of the time, expertise, and financial implications for the laboratory. The testing alone can add up to many thousands of dollars, and hundreds of hours of professional time are dedicated to establishing RIs for four major veterinary species before even considering the variety of other species (e.g., camelids) and subgroups (e.g., juveniles and geriatrics). The following major steps in the process help limit variation in RI results to interindividual variation within a given reference population that is assessed by given methods. Readers are referred to other publications for further details.[9,10]

A. Select and record inclusion and exclusion criteria for reference individuals. Criteria include species, age, sex, and method of determining health status.

1. Inclusion criteria should reflect the demographics and diversity of the population for which the RIs will be used.
2. Exclusion criteria should ensure that unrepresentative and ill subjects are excluded from the reference sample group so as not to contaminate the RI data. Neonates and young animals are typically excluded from routine RIs because of age-related differences in some analytes; separate RIs may be generated for these groups. Certain breeds with atypical results for certain analytes are excluded for the affected analytes (e.g., [platelet] in greyhounds).
3. In domestic species, determination of health may be as limited as a history and physical exam or as thorough (and expensive) as a full medical evaluation with screening blood work and imaging. How would you assess and document health in free-ranging wildlife?
4. Establishment of inclusion and exclusion criteria should include consideration of potentially clinically relevant differences that occur because of variations in breed, physiologic state (including pregnancy and lactation), geographic location (including altitude), nutritional state (including diet), and subclinical disease (e.g., retroviral infections). Should a reference sample group for dairy cattle include milking cows, dry cows, steers, and bulls?

B. Establish an appropriate size of the reference sample group, preferably to have at least 60 animals that meet selection criteria.[11] This typically requires sampling of several more individuals than the goal because some will be excluded when found to have subclinical disease or because of sample problems or other factors.

1. Appropriate sample size depends on reference population homogeneity and the distribution of the reference data. A smaller sample can be representative of a species that is fairly homogeneous (red fox), but a larger sample is required for a species that is heterogeneous (domestic dogs).
2. Some RI guidelines[10] call for at least 120 reference individuals, 120 being needed to calculate 90 % confidence intervals around nonparametric reference limits. However, reference sample groups of that size are often not obtained in veterinary medicine, and a more realistic number of 60 representative individuals may be sufficient with appropriate statistical analysis.
3. Obtaining representative samples from the desired number of qualified reference individuals is often a challenge. Consider the strengths and weaknesses of the following potential reference sample groups:
   a. Animals presented for yearly vaccinations, elective surgical procedures, or as part of a blood donor program; these groups may overrepresent young animals or large-breed dogs
   b. Collecting many samples from one kennel, one cattery, one stable, or one herd; the similar genetics, environment, and nutrition of such groups may yield results that are not representative of the population of interest

C. Collect and process the samples. The sample collection and processing protocol should be defined and aligned with what is used for clinical samples. Samples that do not meet collection and processing requirements should be excluded. Consideration should be given to the following:
   1. The need for fasted versus nonfasted samples, and the duration of fasting
   2. Method(s) and site(s) of collection (e.g., jugular versus cephalic vein versus indwelling catheter, vacuum tube versus syringe); should collection be uniform or varied to the degree that occurs with samples to be evaluated?
   3. Type of sample tube/container or anticoagulant and volume of the sample
   4. Storage time and temperature of samples prior to testing
   5. Sample quality issues such as substantial hemolysis, lipemia, or any clots in an anticoagulated sample

D. Measure or determine the reference values. Analyze the sample for the substance(s) of interest (analyte) in the same manner as patient samples, including use of appropriate QC measures (see Quality of Laboratory Results, III.B).
   1. Concurrent assessment of routine blood test panels (e.g., CBC, chemistry, blood gas) has financial and efficiency benefits, including health screening and determination of multiple RIs from a single reference sample group.
   2. Test results are inspected to identify reference subjects that may have subclinical disease or sample results with preanalytical or analytical error. This is ideally done as results are generated so that correctible errors can be addressed, and further testing can be done to evaluate for disease. Whether done real-time or later, data from individuals with evidence of disease are excluded in toto or in part, the latter if abnormalities have a known cause that does not affect other values.

E. Determine the reference distribution. Begin by displaying the data in a histogram and assessing by visual inspection. Apply statistical methods, such as the Shapiro–Wilk goodness-of-fit (normality) test, to determine whether data have a Gaussian (normal) distribution.[12]
   1. Many analyte results fit a Gaussian distribution, especially those that are tightly regulated by physiologic systems (e.g., concentrations of glucose, Na^+, K^+, and fCa^{2+}).
   2. Many analyte results (e.g., serum enzyme activities) do not fit a Gaussian distribution; these typically have positive (tail to the right) or negative (tail to the left) skewness.

F. Identify and remove outlier values, as these inappropriately affect the reference limits.
   1. Outliers are data that do not fit into the rest of the distribution. They may result from inadvertent inclusion of clinically or subclinically ill subjects or from preanalytical, analytical, or postanalytical errors (see Quality of Laboratory Results, I. Sources of Error).
   2. Statistical outlier tests are available to assess single or multiple (one at a time or sequentially) outlying values at one or both ends of the reference distribution.[13] Some are parametric tests requiring mathematical transformation of nonGaussian to Gaussian data.
   3. The cause of outliers should be investigated. They might result in exclusion of the single result or the entire reference subject because of illness, or they might be corrected after identifying an error. There should be a clear and convincing rationale for concluding that a value is not representative of health and therefore should be excluded; values should not be excluded simply to render the distribution Gaussian, as this will inappropriately affect reference limits.
   4. After outliers are excluded, the distribution of the data should again be determined.

G. Determine reference limits and RIs. The most commonly used RIs represent exactly or approximately the central 95 % of reference values.[14]

1. Limits may be defined by excluding the top and bottom 2.5 % of values. If 2.5 % of the values are > 150, and 2.5 % of values are < 50, the reference limits are 50 (LRL) and 150 (URL) and the RI is 50–150. This nonparametric approach has been recommended whenever the sample size is at least 120. It can be used with as few as 40 samples. With < 40 reference samples, reference limits are represented by the highest and lowest reference values. However, 90 % confidence limits of the reference limits cannot be calculated nonparametrically with fewer than 120 samples. When fewer than 120 reference samples are available, 90 % confidence limits for nonparametric RI must be calculated using bootstrap (resampling) methods.
2. When data have been shown to fit a Gaussian distribution, parametric methods may be used to establish reference limits based on the mean ± 2 SD of values (the central 95.54 % of values rounded to appropriate significant digits). Confidence intervals of the reference limits can be determined by parametric and bootstrap (resampling) methods.
3. When the distribution is not Gaussian and the sample number is less than 120, the data may be transformed into a Gaussian distribution, e.g., using the Box–Cox method for right-skewed data,[15] and assessed parametrically.
4. An iterative or resampling method called the robust method can be used to determine reference limits for large and small data sets. This method performs best if the data are symmetrically distributed, but Gaussianity is not required.[16]
5. Figure 1.2 shows the differences between Gaussian and skewed distributions and the RIs obtained from such distributions. Box–Cox transformation of the skewed distribution might convert it to Gaussianity.

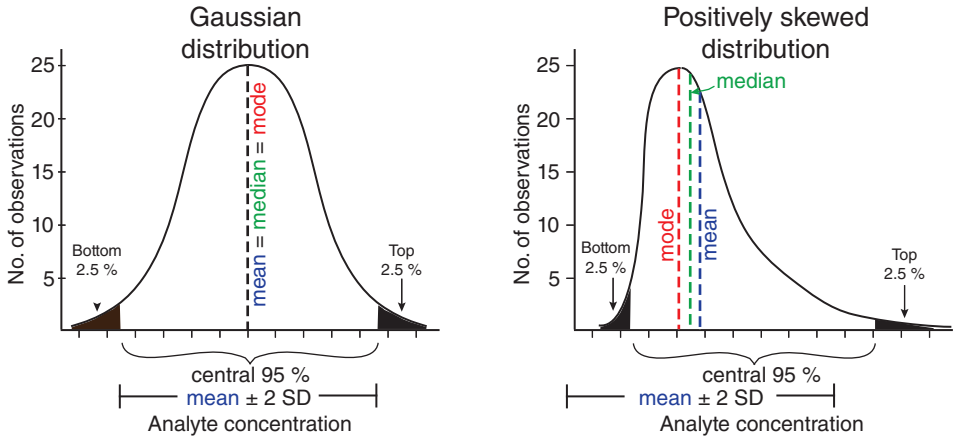


Figure 1.2. Reference distribution.

- In the left graph, the reference values conformed to a Gaussian distribution. If data have this distribution, then mean ± 2 SD represents the central 95 % of the reference values and thus the RI. If values have a Gaussian distribution, the mean, median, and mode values are equal.
- In the right graph, the reference values had positive skewness. If data have this distribution, nonparametric methods are used to determine the central 95 % of the values. If the interval represented by mean ± 2 SD were calculated, the LRL would be below the lowest reference value, and the URL would exclude more than the top 2.5 % of the reference values. Therefore, mean ± 2 SD would not be an accurate representation of the values expected in healthy animals. The mean, median, and mode values will be different if the reference values have a nonGaussian distribution. Statistical transformation of these data might yield a Gaussian curve that could be used to calculate reference limits.

H. Calculate 90 % confidence intervals (CIs) of the upper and lower limits.[17] Confidence intervals around the limits describe the precision of the limit estimates.
I. Document details of RI studies in a comprehensive report that can be made available to users upon request.
J. RIs should be verified every 3–5 yr.[9] Several procedures for verification are available, including the simple method described in the following section regarding transference. Methods using archived laboratory results also are described.[18]

III. Transference (adoption) of existing RIs
A. Transference of existing RIs from one laboratory to another laboratory or from one analyzer to another analyzer is a quicker and less expensive alternative to establishing de novo RIs. This may be done when analyzers are updated to new models or new methods are added, but it should be limited to a single occurrence, especially if the RIs require adjustment for bias between methods or analyzers. Once the method or analyzer is established in the laboratory, results from healthy subjects can be collected over time for de novo RIs.
B. Sources of RIs for transference may include the laboratory itself, another reputable laboratory, or the manufacturer of the analyzer or method.
C. Adherence to the following steps helps ensure that transferred RIs are appropriate.
1. Confirm that the transferred RIs match the adopting laboratory's patient population and collection methods. This requires availability and review of the donating laboratory's comprehensive RI study document.
2. If RIs are not transferred between the same methods and type of analyzer, perform a comparison of methods study (see Evaluating and Validating Laboratory Methods, Section IV) and adjust reference limits if there are biases that exceed predetermined quality goals (e.g., TE_a).
3. Confirm similar precision and accuracy of the analyzers and methods.
4. Verify the transferred RIs using samples from the laboratory's patient population. This can also be done to verify factory RIs provided for an analyzer. The simplest procedure is as follows:
a. Collect and analyze samples from 20 healthy subjects using the same inclusion and exclusion criteria established for de novo RIs.
b. Generally, accept the RI if two or fewer results fall outside it – theoretically, 95 % (19 of 20) would be WRI – but if no results fall outside the RI and they are clustered in a limited region, the RI might be too broad.
c. Reject the RI if five or more results fall outside it.
d. If three or four results fall outside the proposed RI, collect and assess an additional 20 samples from healthy subjects. Accept the RI if two or fewer results fall outside it; otherwise, reject it.

IV. Use of RIs
A. It is important to understand that most RIs represent results expected in approximately 95 % of healthy animals (i.e., in 19 of 20 healthy animals).
1. Therefore, 1 of 20 healthy animals is expected to have a measured value outside of the RI. The value that is outside the RI but still represents health is expected to be close to a reference limit. A marked difference from the RI probably represents a pathologic state, although it could also represent an erroneous value.
2. When examining multiple test results for one animal, it is important to remember that this 95 % chance of a value from a healthy animal falling WRI applies to each result

separately. (Note: The chance is theoretically 95.54 % for parametric methods, but not exactly given rounding to significant digits.)

a. As shown below, this means that in a panel of 20 assays, each with a 5 % chance of being outside of the RI even without pathologic change, there is a 64 % chance for at least one value to be outside the RI. With a panel of five results, there is about a 23 % chance. However, such values should be close to the lower or upper reference limit. This concept is particularly important when laboratory tests are requested for clinically healthy animals for a yearly health profile or a geriatric profile. Also, it would be unexpected for all CBC and chemistry results in a study's healthy control group to be WRI, if RIs were appropriate.

b. These percentage chances are derived from probability calculations. Because there is about a 95 % (0.95) probability in healthy animals that each result is WRI, the probability that all n values are WRI is 0.95^n, and the probability that not all values are WRI (i.e., at least one is not WRI) is $1 - 0.95^n$ (where n is the number of observations). Thus, for a panel of 20 test results, $100 \times (1 - 0.95^{20}) = 64$ %, so there is a 64 % chance that at least one value is outside the RI.

B. Not all RIs accurately reflect expected findings in a population of interest. This may be because of such problems as an inadequate procedure of establishing the values, a difference between reference and test populations, or a change in methods or instrumentation by the laboratory without an appropriate change in RIs.

C. It is appropriate to ask a laboratory for the basis of their RIs. This enables evaluation of how likely it is that the values reflect health in the population of interest. Such requests are particularly appropriate when first contemplating use of a laboratory. Requested information might include the following:

1. Number of reference individuals
2. Ages, breeds, and genders of reference individuals
3. Criteria for inclusion as a reference individual
4. Methods for establishing RIs, including de novo versus transference

D. Use of RIs published in textbooks is discouraged but may be necessary to provide guidelines for certain populations (e.g., neonates, wildlife, zoo animals). Such RIs may not be applicable because of preanalytical and analytical variations, or because the reference population does not represent the individuals of interest. Unfortunately, for many textbook RIs, important details may not be documented.

V. Reference limits versus decision thresholds (limits)

A. As defined above (see Reference Intervals, Section I.B.6), reference limits are the highest value and lowest value of the RI and thus are descriptive of a reference distribution; a RI includes values expected in about 95 % of the healthy animals. These values are used diagnostically to help detect pathologic states in ill animals. A patient value that is above or below the reference limits might indicate the presence of a pathologic state.

B. A *decision threshold* (clinical decision limit, cutoff point, or cutoff value) is a value that is used to classify a result as positive or negative for a disease, or to decide whether to treat or not to treat.[19] In other words, decision limits help *discriminate* between subjects with or without a particular disorder. As described in the section on Diagnostic Properties and Predictive Value of Laboratory Assays, a decision threshold is not necessarily a reference limit. Decision thresholds can also be used to estimate a cost-to-benefit ratio for diagnostic methods. Choosing a higher or lower decision limit alters rates of false positive or false negative conclusions, thus affecting costs related to additional diagnostics, inappropriate patient management, and increased morbidity or mortality.

C. The term *decision limit* is also used in the context of a *critical decision limit* or critical value, which is a value that is used for making critical therapeutic decisions. For example, a critical decision limit for serum [tCa^{2+}] might be 15.0 mg/dL, meaning that aggressive therapy is initiated whenever a serum [tCa^{2+}] is greater than 15.0 mg/dL. In contrast, the upper reference limit might be 12.0 mg/dL, and a decision threshold for a particular hypercalcemic disorder might be 12.5 mg/dL.

QUALITY OF LABORATORY RESULTS

I. Sources of error
   A. When results are unexpected, a clinician might say, "I don't believe these results; there must have been a lab error." It is important to remember that there are many potential reasons for an erroneous laboratory test result, and identifying the source requires recognition of where errors can occur throughout the laboratory cycle (Figure 1.3). Errors anywhere along the cycle may jeopardize patient management, and many do not occur in the lab. Errors may be preanalytical, analytical, or postanalytical.
   B. *Preanalytical errors* are common (Table 1.7) and arise from problems prior to sample analysis. They may occur before or after the sample reaches the laboratory, and include errors in patient preparation, test selection, sample collection, and sample processing.[20] These errors can be minimized with laboratory guidance and attention to the following:
      1. Patient preparation: Properly prepare the patient, including fasting many small animal patients and minimizing excitement prior to sampling.

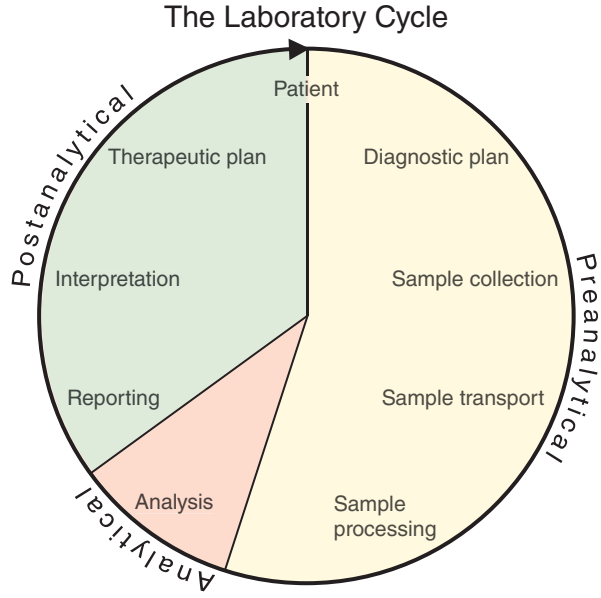


Figure 1.3. The laboratory cycle. Laboratory testing begins with the patient (*top*). After establishing a diagnostic plan, samples are collected, transported, and processed for analysis. Generated results are reported, and their interpretation leads to a therapeutic plan that is applied to the patient. Errors can occur at any step in this cycle, preanalytical errors occurring prior to analysis (*yellow region*), during analysis (*pink region*), or after analysis (*green region*). When using accredited laboratories, the fewest errors occur during analysis.

Table 1.7. Examples of preanalytical laboratory errors[a]

Prelaboratory	Prelaboratory (continued)
Test selection	Sample transport
Intentionally or accidently choosing wrong test	Delayed transport to laboratory
Sample collection	Sample held at improper temperature
Sample collected after treatment instituted	Sample not protected from breakage
Traumatic venipuncture	Regulatory noncompliance for biologicals
Prolonged venipuncture into syringe	Unstained smears exposed to formalin
Catheter or port improperly flushed	Submitted information
Blood tube underfilled	Unnecessary test requested
Urine collected from contaminated surface	Request not clear
Cytologic sample not representative	Species not provided
Collected too superficially	Fax, phone, address not provided
Fluid or necrotic center collected	For cytology: lesion not described
From single region of large mass	Reason for testing not indicated
Bone marrow aspirate hemodiluted	Multiple samples not clearly identified
Sample container	Bone marrow sample lacks CBC results
Urine collected into contaminated container	
Wrong blood tube used	
Tube past expiration date	
Sample processing[b]	
Blood tube not inverted promptly and properly	**In the laboratory**
Tubes and slides not labeled	Delayed processing (e.g., urinalysis)
Fresh blood smears not made (CBC)	Sample accepted that should be rejected
Blood smears thick or otherwise problematic	Sample mix-up or misidentification
Fluid sent out without fresh preparations	Centrifugation force/time inappropriate
No concentrated preps of low-cellularity fluids	Serum removed before thorough clotting
Cytologic sample sprayed on slide	Plasma and serum mixed
Cytologic sample not spread on slide	Cell contamination when serum harvested
Bone marrow smears thick or lack particles	Processed at inappropriate temperature

[a] Many examples apply whether testing is in-clinic or send-out; some apply only to send-out. If testing is in-clinic, clinic personnel must avoid errors; when sending samples out, trained laboratory personnel will help prevent or recognize problems.

[b] See "In the laboratory" for examples of other processing errors that may also occur prior to sending samples out.

2. Sample collection: Use proper collection order and technique to collect appropriate sample types and amounts for testing and potential retesting or additional testing (seek lab guidance as needed).
   a. Sample directly to vacuum tubes when possible; collection into a syringe with subsequent and delayed filling of tubes promotes clotting, platelet activation, and problematic samples.
   b. To obtain 1 mL of serum, at least 3 mL of blood is usually needed.
   c. Fill citrate tubes enough to yield a 9:1 blood to anticoagulant ratio.

3. Sample container: Use appropriate container and contents.
4. Sample processing: Mix samples as recommended, label samples, follow appropriate processing guidelines, make good smears of fresh samples, and avoid or minimize testing delays.
5. Sample transport: Transport promptly at appropriate temperature, maintain separation of cells and serum, package to avoid breakage, and keep formalin fumes separated from unstained smears.
6. Sample submission: Request appropriate test, provide all relevant client information, clearly identify all samples, and provide appropriate patient information.
7. Appropriate laboratory processes: Process samples appropriately, reject unsuitable samples, and confirm accurate sample identification throughout.

C. *Analytical errors* occur during analysis of the sample (Table 1.8) and should be minimized by attention to the following:
1. Using and maintaining quality instruments and equipment: Generally, "you get what you pay for," but a quality instrument remains a quality instrument only if it is properly maintained.
2. Minimizing operator errors: Have personnel properly trained and use written standard operating procedures to help avoid person-to-person variation.
3. Using instruments and methods appropriate for the species: An instrument calibrated to measure the relatively large human RBCs may not provide accurate measurement of smaller RBCs from domestic mammals.
4. Using appropriate reagents and consumables: Store and use according to instructions, including disposal of expired reagents.
5. Using a quality assurance plan (QA program): Laboratory personnel should adhere to internal assessment procedures to minimize, detect, and resolve errors to help ensure that all parts of the quality analysis of the sample are maintained daily.[21]
6. Recognition and, when possible, avoidance of interferents of laboratory results for the methods being used: Interferents include various therapeutics or their metabolites, hemoglobin from hemolysis, lipemia, and bilirubin.[22]
    a. Drugs may alter laboratory results because of their physiologic or pharmacologic effects, but some drugs or their metabolites cause direct analytical interference

Table 1.8. Examples of analytical errors

Instrument or equipment malfunction	Reagent problem
Physical problem	Reagents expired
Electronic problem	Bad lot of reagents
Out of calibration	Reagents improperly stored/handled
Operator error (consistent or sporadic)	Results accepted without acceptable control results
Untrained operator mistake	Interferences
Tired or distracted operator mistake	Lipemia
Operator short-cuts/alters procedure	Hemolysis
No standard procedure to follow	Hyperbilirubinemia
Wrong/old SOP[a] followed	Certain treatments
Wrong procedure	
SOP contains error(s)	

[a] SOP, standard operating procedure

(e.g., potassium bromide treatment falsely increases [Cl^-] and falsely decreases [glucose] by certain methods). Collect samples pretreatment when possible.

b. Hemoglobin may interfere with certain wavelengths of light transmission in spectrophotometric assays and yield false increases or decreases, depending on assay design. Avoid causing in vitro hemolysis.

c. Bilirubin in icteric serum or plasma may interfere with certain wavelengths of light transmission and alter results of some spectrophotometric assays.

d. Lipoproteins in serum, plasma, or other fluids can cause turbidity and therefore interfere with light transmission used for many assays. They can also alter protein conformation, which may affect some immunoassay and electrophoresis results, and they increase the volume percentage of solids in samples, thereby falsely decreasing electrolyte concentrations measured by indirect potentiometry.[23] Postprandial lipemia can be avoided by appropriate fasting.

D. *Postanalytical errors* occur after analysis, either within the laboratory or outside it (Table 1.9). They can be minimized with attention to the following:

1. Accurate entering and transferring of data: Limit need for data transcription and ensure care during transcription of results to a laboratory report or during keyboard entering of data into computers.
2. Verifying results and reports: Inspect reports before release.
3. Prompt, clear, and complete reporting to the correct destination: Minimize hand-written reports.
4. Appropriate use of RIs: Recognize that provided RIs may not be appropriate for a particular sample, e.g., one from a neonate.
5. Appropriate interpretation of results: Do not assume unexpected results are laboratory errors; use resources and continuing education opportunities.

II. Analytical properties of assays

A. Analyzers and their maintenance impact the quality of laboratory results. From the analytical perspective, the best clinical laboratory assay is one that consistently measures the true concentration or activity of the substance and at concentrations and activities that are clinically relevant. When evaluating and comparing laboratory assays, five analytical properties can be assessed.[24]

Table 1.9. Examples of postanalytical laboratory errors

In the laboratory	After the laboratory
Transcription incorrect	Results or recommendations are ignored
Result validation/verification	Results are misread or misunderstood
No system for validation of results	Flagged analyzer results are not ignored
Error missed while validating	Results are present but not noticed
Reporting	Results are misinterpreted
Reported to wrong person/place	Laboratory not consulted for clarification
Report delayed	
Report illegible	
Report incomplete	
Reference interval error	
No reference values provided	
Poor reference intervals provided	

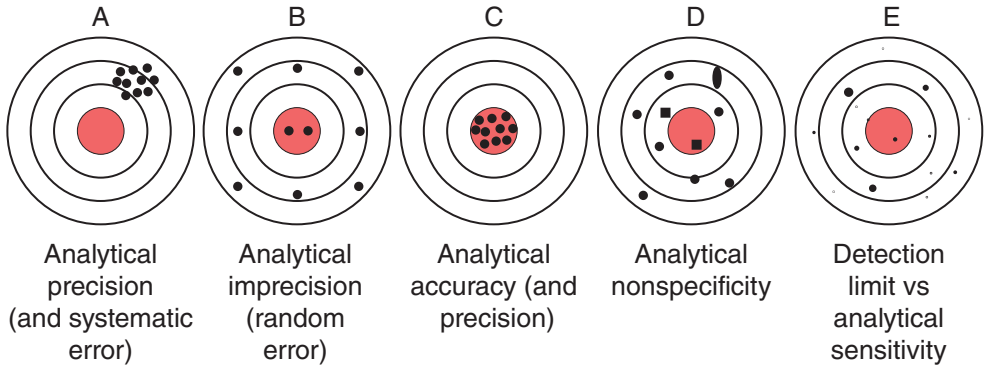


Figure 1.4. Illustrations of analytical properties of assays.

A. *Analytical precision (systematic error)*: Because the 10 holes in the target are tightly clustered, the target shooter was precise even though consistently (systematically) inaccurate.

B. *Analytical imprecision (random error)*: Because the holes in the target are evenly distributed, the average of 10 shots is exactly in the middle of the target area. Thus, the shooter was statistically accurate but imprecise.

C. *Analytical accuracy (and precision)*: Because all 10 holes in the target are tightly clustered in the middle of the target, the shooter was accurate and precise.

D. *Analytical nonspecificity*: Of the 10 holes in the target, 7 are round, 2 are square, and 1 is oval. Of the 10 observations, the shooter probably created 7 and 3 were created by other factors. Thus, the presence of holes in the target is not specific for the shooter's actions.

E. *Detection limit vs analytical sensitivity*: How many holes can you see in the target? The smallest hole that you can reliably detect is the detection limit of your eyes. Analytical sensitivity in this context is the smallest change in hole sizes that your eyes can reliably differentiate. If your eyes can reliably detect the different sizes of all 15 holes, then the sensitivity limit has not been reached. If your eyes can differentiate the changes in the four largest holes but not the other holes, the sensitivity limit is between the fourth and fifth holes. Each shot was made with a projectile whose diameter was 75 % of the previous shot.

B. *Analytical precision*: ability of an assay to get the same result if a sample is analyzed several times; this is also called repeatability, reproducibility, or random analytical error (Figure 1.4A)
   1. The need for analytical precision depends on the degree of variation that can be accepted as random variation (error). Typically, manual assays have more random error than do automated methods because people generally cannot reproduce their work as well as an instrument can. When management of a case requires that small changes in concentration reflect a biological change and not analytical variation, an assay needs to have high precision. However, if a wide range of measured values can be accepted, e.g., creatine kinase activity, then high precision is not needed.
   2. Precision is frequently expressed in clinical laboratories by the method's CV (Eq. 1.1). A method's CV is determined from replicate analysis of a sample or QCM within an assay run and between different runs of the same assay.[25]
      a. A *within-assay (within-run, intra-assay) CV* (also called repeatability) represents the random error that is expected when one sample is analyzed multiple times in one run of an assay. If an assay has a poor within-assay CV, the sample's analyte concentration should be determined by analyzing the sample in duplicate or triplicate and then calculating a mean concentration. However, repeated measures are not economical for routine testing in clinical pathology laboratories, so routine assays

generally have sufficient precision to provide reliable results when testing only once and monitoring with an effective QA program.

b. A *between-assay (inter-assay) CV* (also called reproducibility) represents the random error over multiple runs of the assay when using the same sample, which must be stable for the duration of the assessment. The within-assay CV is expected to be smaller than the between-assay CV; the opposite pattern suggests a problem with the precision studies.

c. The clinical relevance of an assay's CV is determined by many factors.

(1) If critical clinical decisions are made when changes in an analyte's concentration are minimal, then the assay must have a very low CV (high precision). Otherwise, the diagnostician would not know if the change in analyte concentration is due to a true change that occurred in the animal or if the change represents analytical error.

(2) An assay's CV typically varies with the analyte's concentration. Higher CV values may be found at the lower and upper limits of the assay's analytical range. Within the analytical range, CV values are typically higher at lower analyte concentrations because CV values are expressed as percentages. This is most apparent for analyte quantities that vary the most.

(a) If an analyzer's [platelet] measurement has a SD of 3000/µL at all concentrations, the assay's CV values would be 50 % at a mean concentration of 6000/µL, 5 % at a mean of 60,000/µL, and 0.5 % at a mean of 600,000/µL.

(b) If a [platelet] assay has a CV of 5 % at all concentrations, then the assay's SD would be 300/µL at an analyte concentration of 6000/µL, 3000/µL at an analyte concentration of 60,000/µL, and 30,000/µL at an analyte concentration of 600,000/µL. Depending on the analyte and the amount of its variation within an individual (biological variation), a CV of 10 % can be unacceptable or acceptable analytical variation (see TE_a in III.B.5 and III.B.8 of Quality of Laboratory Results).

(3) Knowing the amount of analytical variation for an assay is helpful when trying to determine whether a change in a patient's data represents analytical variation or a true biological change. One way to approach this decision is to use an assay's USD. *USD* is an average of SD values from three to six consecutive months of QA values. If a change in a patient's data is less than twice the USD, the change may be only analytical variation. If a change in patient data is greater than three times an assay's USD, then the change is probably truly reflecting a biological rather than analytical cause. The cumulative average SD representing USD varies considerably across assay methods but is often accessible to laboratorians through analyzers' onboard monitoring systems.

(a) If the USD for $[Na^+]$ at 150 mmol/L is 1.5 mmol/L and a patient's $[Na^+]$ is 150 mmol/L on day 1 and 148 mmol/L on day 2, then the change of 2 mmol/L may be due to analytical variation since it is less than twice the USD. However, a value of 145 mmol/L on day 2 (a change greater than three times the USD) probably represents a true biological change.

(b) If the USD for blood [neutrophil] is 1000/µL at a [neutrophil] of 10,000/µL, then the [neutrophil] in a second sample needs to be < 7000/µL or > 13,000/µL to be reliably considered a biological change if the first sample had a [neutrophil] of 10,000/µL.

(4) If a change in serial values appears to be biological, that does not mean it is of clinical significance because it might be within expected limits of biological variation. As discussed in the Analytical and Biological Variation section, V.C, RCVs or critical differences might become useful for assessing clinical significance when appropriate data are available. Other approaches to determining and conveying measurement uncertainty for laboratory data include measurement uncertainty (MU), expanded measurement uncertainty (EMU), and dispersion.[26–29]

d. An assay's precision (CV_A) has an impact on how results are reported, including the number of significant digits (see Significant Digits) used and the reporting interval. *Reporting interval* is the increment used to report laboratory results for an analyte, most often in decimal units of 0.01, 0.1, 1, or 10, although nondecile increments can be used.[29] Use of appropriate significant digits and reporting units helps communicate the uncertainty of a result. Rules for selecting clinically appropriate reporting intervals based on analytical variation and critical differences have been published.[26,30]

C. *Analytical accuracy*: closeness of the agreement between the measured value of an analyte and its "true" value (Figure 1.4C)

1. Methods of establishing "true" values vary considerably. At times, the true value is established by a reference method (e.g., a method developed by the NIST). If such standards have not been developed, the "true" value might represent the mean concentration determined by numerous observations using an assay that is accepted as the best available.
2. Typically, accuracy of a clinical assay is assessed by comparison of its results to the results of an accepted reference method, a method that has been accepted by a standardization group as providing a true value. Optimally, the assay that is accepted as the reference method is also precise so that only a few observations are needed to determine the true value (Figure 1.4C).
3. If method comparison reveals biases and results that deviate from those of a good reference method, the assay is not accurate. Use of a more accurate method or analyzer would be indicated, if available. Establishing reasons for the inaccuracies may allow assay modification, but these may need to be done by a company as part of analyzer development. Although adjustment of RIs from inaccurate assays is commonly done to allow interpretation of inaccurate data, this practice contributes to interlaboratory variations and impairs use of broadly acceptable RIs and medical decision limits.
4. Standard solutions
   a. The accuracy of the clinical assay may be assessed by measuring an analyte's concentration in a *reference standard solution* whose concentration was certified after determination by a reference method.
   b. Reference standard solutions are not the same as calibration standard solutions. *Calibration standard solutions* are commercially prepared and used to calibrate an instrument or method for a particular analyte. They should closely agree with reference standard solutions.
   c. In contrast to calibration standards, *quality control materials* (QCM) are analyzed every day or shift to monitor analytical performance by comparing measured values to expected results. They contain one or more analytes at concentrations that may be below, above, or within RIs. QCM may be "assayed" or "unassayed;" "assayed" QCMs have assigned analyte concentration intervals determined by the manufacturer, usually on several analyzers. For "unassayed" QCM, analyte concentrations are established through repeated measurement by the user.

D. *Analytical specificity*: ability of an assay to detect only the substance of interest (analyte); freedom from interfering substances (Figure 1.4D)
   1. Analytical specificity is related to analytical accuracy because an assay cannot be accurate if a nonspecific reaction is occurring.
   2. The need for analytical specificity varies directly with the likelihood of interfering substances. A serum glucose assay may be designed to react with all hexoses. Such specificity may be acceptable if glucose is the only hexose in serum that is at a sufficient concentration to be detected by the assay. Other glucose assays may be designed to react only with glucose even if other hexoses might be present in the sample.
   3. Substances might interfere with an assay in many ways:
      a. The substance might be very similar chemically, so the assay reacts with either the analyte or the interfering substance (e.g., Br^- detection falsely increases $[Cl^-]$).
      b. The substance might produce the same response that is detected in the assay system.
         (1) The presence of glucose might be detected when a chemical reaction produces hydrogen peroxide (H_2O_2). A substance can interfere with the assay by having the same oxidizing properties as H_2O_2.
         (2) In spectrophotometric assays, the presence of lipids, bilirubin, or hemoglobin can interfere with light transmission through a sample, and thus the results of the assay are changed by artifactual spectral changes and not by a chemical reaction.
   4. Depending on how a substance interferes with an assay, it might lead either to falsely increased or to falsely decreased concentrations, i.e., positive interference or negative interference, respectively. The degree and direction of interference can vary with the method, even for a single analyte.

E. *Detection limit* (definition of the International Union of Pure and Applied Chemistry, IUPAC): the smallest concentration or quantity of an analyte that can be detected with reasonable certainty for a given analytical range (Figure 1.4E)
   1. An assay's detection limit involves the ability of the assay to differentiate background "noise" from a true change because of the presence of an analyte. The detection limit defines the lowest value of an assay's analytical range; *analytical range* is the range of values over which the assay can provide reliable results.
   2. If an analyte is relatively abundant (e.g., serum Na^+), then a detection limit of 100 mmol/L may be adequate. For substances that are relatively rare (e.g., aldosterone), a detection limit of 100 pmol/L may be needed to be clinically useful. One pmol is a billionth of 1 mmol.
   3. Note: Several different terms and definitions are used for this general property.

F. *Analytical sensitivity* (definition of the International Union of Pure and Applied Chemistry, IUPAC): the slope of the calibration curve and the degree to which an analytical procedure produces a change in the signal for a defined change of the quantity (however, some organizations and many people consider analytical sensitivity to reflect detection limit)
   1. In other terms, an assay's analytical sensitivity is how much change of the analyte (concentration, property, etc.) is needed for the assay to detect the change. For example, an assay that can differentiate 50 mg/dL from 51 mg/dL has better analytical sensitivity than an assay that can only differentiate 5.0 g/dL (5000 mg/dL) from 5.1 g/dL (5100 mg/dL).
   2. Analytical sensitivity should not be confused with detection limit. Both relate to small changes in concentration, but, as defined by IUPAC, analytical sensitivity applies to changes within an assay's analytical range, whereas detection limit applies to the lowest limit of the analytical range.

III. Quality assurance

A. Every clinical laboratory (from large reference laboratories to small in-house laboratories) should have a QA program to monitor laboratory performance and ensure that reliable results are generated.[31] The Quality Assurance and Laboratory Standards Committee of the American Society for Veterinary Clinical Pathology is dedicated to the education and communication of quality laboratory topics and has published guidelines for the monitoring of laboratory procedures (www.asvcp.org/page/QALS_Guidelines).[32–40] Quality assurance programs encompass all nonstatistical and statistical procedures implemented to minimize error in laboratory results. When suitable control materials are not available (e.g., for urinalysis sediment examination), one must rely on nonstatistical QA methods.

1. Nonstatistical QA components include:
   a. Written standard operating procedures (SOPs)
   b. Maintenance routines with log-sheet documentation for all equipment
   c. Personnel with appropriate training and continuing education
   d. A responsible laboratory supervisor
   e. Internal or external proficiency testing
2. Statistical QA processes are often referred to as *quality control* (QC) testing and encompass testing of QCMs to ascertain presence or absence of analytical errors caused by problems with (i) the operator, (ii) the reagents, (iii) other consumables, or (iv) the instrument. To detect relevant operator errors, QC testing should be done by the person testing patient samples. Detectable analytical errors can be random or systematic.
   a. *Random error* is imprecision caused by factors that randomly affect the measurements, such as variations in dispensed volume of reagent or sample (Figure 1.4B).
   b. *Systematic error* (*bias*) is reproducible inaccuracy that consistently results in values that are too high or too low; causes include deterioration of a reagent or a light source (Figure 1.4A).
   c. QC testing does not assess for analytical error caused by interferents in patient samples, and it does not assess for preanalytical or postanalytical errors. If a so-called electronic QC-check is done rather than assessment of an actual QCM sample, that can verify only that an instrument's circuitry is functioning; it does not assess the entire operation of the analyzer from sample application to reporting a result.

B. Quality control testing

1. Quality control testing should be designed to detect random and systematic analytical errors of clinical significance before patient results are determined or reported. To this end, QCMs (see Quality of Laboratory Results, II.C.4.c) are generally assessed for each test at the beginning of each day or shift prior to testing patient samples. QCMs may contain analytes at quantities that are below, within, or above RI to verify that the assay can measure values that are decreased, WRI, or increased; these QCMs can be tested on a rotating schedule with only one per analyte each day.
2. If single and serial results for a QCM are within defined limits, then the assay appears to be "in control" and performing correctly, thus patient results are probably valid and can be generated and reported. If QCM results are outside the defined limits, then the assay is "out of control!" and troubleshooting is indicated to identify a source of unacceptable analytical error, i.e., operator, reagents/consumables, or analyzer. Although care should be taken to manage QCMs so that they are reliable, it is also possible that the QCM is the problem, so assessment of a new QCM might be indicated if other errors are not

found. Reliable patient results cannot be generated until corrective actions resolve the error, as demonstrated by subsequent QCM results being within defined limits.

3. *Westgard rules*, developed by James O. Westgard, Ph.D., form a system of rules that are applied singly or in combinations to warn of unwanted random or systematic error in QCM results (www.westgard.com).[41] Example rules follow; "s" stands for standard deviation.
   a. 1_{2s} rule: The assay is "out of control" when a single control measurement exceeds the mean plus 2s or the mean minus 2s of previous control sample values, or it serves as a warning system to initiate additional inspection of control data. Variations in results that are caused by random error should have a Gaussian distribution (Figure 1.2). With that distribution, 95.45 % of the results should fall within the interval formed by the mean ± 2 SD. Therefore, if a QCM is analyzed 20 times, about 19 of 20 results are expected to be within that interval, but it is important to recognize that 1 of 20 results is expected to be outside of that interval, even when the assay is performing well!
   b. 1_{3s} rule: The assay is "out of control" when a single control measurement exceeds the mean plus 3s or the mean minus 3s of previous control sample values. This rule is primarily sensitive to random error.
   c. 2_{2s} rule: The assay is "out of control" when two consecutive control measurements exceed the mean plus 2s or the mean minus 2s of previous control sample values. This rule is primarily sensitive to systematic error.
   d. R_{4s} rule: The assay is "out of control" when one control measurement in a group exceeds the mean plus 2s and another exceeds the mean minus 2s of previous control sample values. This rule is primarily sensitive to random error.
   e. 4_{1s} rule: The assay is "out of control" when four consecutive control measurements exceed the mean plus 1s or the mean minus 1s of previous control sample values. This rule is primarily sensitive to systematic error.
   f. 10_x rule: The assay is "out of control" when 10 consecutive control measurements fall on one side of the mean of previous control sample values (this rule is sometimes modified to either an 8_x rule or a 12_x rule). This rule is primarily sensitive to systematic error.
4. Laboratory personnel can determine whether a random error or systematic error occurred based on which control rule was violated. With this information, the search for the cause of the error can be narrowed to the most likely analytical problems.
5. Selection of appropriate Westgard rules to control an assay depends on the amount of error inherent in the assay as well as the magnitude of change in a measurement that is considered to be biologically significant. An assay's inherent error is the TE_o (observed imprecision and inaccuracy) determined during assay validation or verification; in most clinical situations, RIs adjust for assay inaccuracies, so it is the degree of random error (imprecision) that is relevant. Biologically significant changes in an analyte are described by TE_a, a quality goal determined by a consensus of experts and defined as the greatest magnitude of variation in a result that does not alter clinical interpretation of the results.
   a. The TE_a for an assay depends on biological variations in health and magnitude of change in analyte measurements with disease. If intra-individual and inter-individual biological variations (see Analytical and Biological Variation, Section IV) of a particular analyte are small, and clinical decisions are made when there is only a minor change in a laboratory test result, then the random error of the assay needs to be relatively small. However, if clinical decisions are made only when there is a marked change in a laboratory test result, a larger random error might be acceptable. Two examples illustrate these concepts:

(1) Serum [Na^+] in most healthy mammals is near 150 mmol/L. If a patient has a serum [Na^+] of 150 mmol/L on day 1 and 160 mmol/L on day 2, most clinicians will conclude that there has been change for biological, pathologic, or therapeutic reasons. Therefore, for the serum Na^+ assay to be clinically valuable, its acceptable random error definitely needs to be < 10 mmol/L. Otherwise, erroneous decisions may be made because of random analytical error.

(2) Serum [glucose] in most healthy mammals is near 100 mg/dL. If a patient has a serum [glucose] of 100 mg/dL on day 1 and 110 mg/dL on day 2, most clinicians will conclude that such changes are within expected biological variation, i.e., there has not been a change for biological, pathologic, or therapeutic reasons. Therefore, for the serum glucose assay to be clinically valuable, its acceptable random error can be larger than the random error of the serum Na^+ assay.

b. The lesser the TE_o of an assay or the greater the TE_a, the less stringent the Westgard control rules need to be to help ensure reliable results. Tests done in most current chemistry analyzers perform with high precision and accuracy; therefore, less stringent Westgard rules, such as 1_{3s}, are sufficient to monitor performance and detect errors that indicate performance problems. Coagulation analyzers that employ manual pipetting of sample or reagent are more prone to analytical errors and thus require use of combined and more stringent Westgard rules that monitor both random and systematic error.[42]

6. Each Westgard rule or multirule (combination of rules), when applied to an assay with a predetermined TE_o, has a predictable probability of error detection (P_{ed}) and a predictable probability of false rejection (P_{fr}). The simplest rule or multirule that is capable of detecting errors ≥ 90 % of the time with ≤ 5 % false rejections is preferred. Performance of a rule or multirule can be increased by increasing the number of results (n) that are assessed in a "control run." An n of two can be achieved either by measuring one QCM twice or by measuring each of two QCMs once. Selecting QC rules that appropriately and efficiently monitor assay performance is called *QC validation*. The use of control charts (OPSpecs Charts) and software programs facilitate QC selection and validation (see www.westgard.com).[43,44]
7. A common method of monitoring the results of the control samples is to plot them in an electronic or paper Levey–Jennings control chart (Figure 1.5). This graphic display of serial QCM results facilitates assessment of Westgard rules. Prerequisites to using Levey–Jennings control charts include selection of a quality goal (TE_a), measurement of assay performance (TE_o), calculation of the mean and SD for each QCM, and validation of control rules.
8. Dr. Westgard has also addressed QA using concepts of total quality management and Six Sigma Quality Management (see www.westgard.com).[45] These concepts have been accepted by some international groups as the preferred QA method. The understanding and application of the Six Sigma Quality Management system requires considerable effort; a few key concepts introduce the topic:
   a. *Sigma* refers to the Greek letter σ, which stands for standard deviation, a statistical index used to express random deviation from the mean (central tendency).
   b. A *sigma* classification for a test refers to the inherent variability of its results relative to TE_a (includes random and systematic error), which is the greatest magnitude of variation in a result that does not alter clinical interpretation. The TE_a must be determined at specified clinical decision thresholds. For some analytes, critical clinical

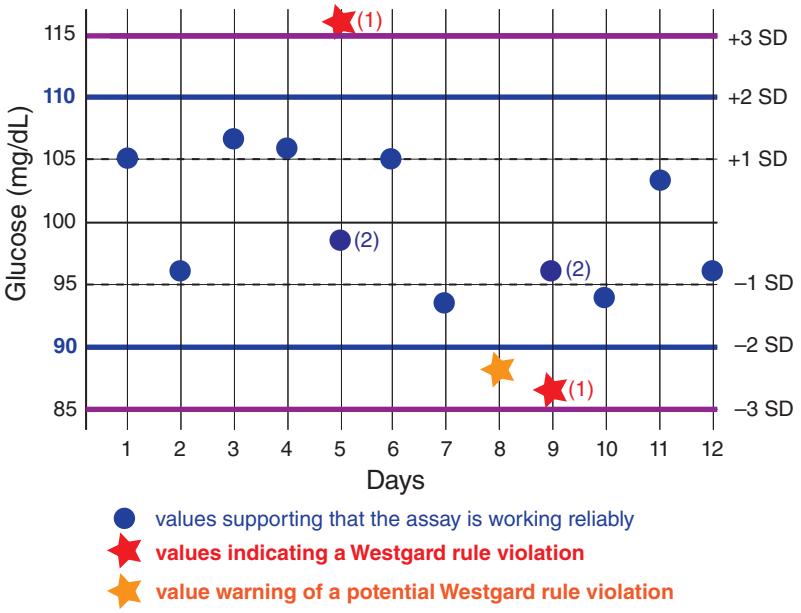


Figure 1.5. Levey–Jennings control chart for a glucose control solution. Quality assurance data for 12 d of testing are displayed; each day, the measured [glucose] in the QCM is plotted on the chart. Previous analyses established the mean (100 mg/dL) and SD (5 mg/dL) for the QCM's [glucose], allowing display of lines for ± 1, 2, and 3 SD.

- On day 5, the first measured concentration (1) was > 3 SD above the historical mean, thus violating the 1_{3s} Westgard rule and suggesting an analytical error. After corrective actions were taken to determine and correct the problem, a second measured concentration (2) was within acceptable limits, so values measured in subsequent patient samples could be considered reliable.
- On day 8, the QCM's [glucose] was more than 2 SD from the mean, representing a 1_{2s} warning that an error might be present, but this is expected to occur for 5 % of QCM results (1 in 20), so it could be a false alarm. In this case, the control value on day 9 was again > 2 SD below the mean, thus violating the 2_{2s} Westgard rule and suggesting an analytical error, likely systematic and not random. After corrective actions were taken, a second measured concentration (2) supported that the assay was again "in control."

For some assays, at least two control solutions are analyzed with each set of patient samples, and results of each control solution are plotted in separate Levey–Jennings control charts for each analyte. Control solutions are typically analyzed with each shift (e.g., day and night) or each day with the primary goal of detecting unacceptable assay performance before a patient's result is used for diagnostic or therapeutic decisions.

decisions are made at both increased and decreased quantities, so two TE_a values might be required.

c. Sigma is calculated as follows: $\sigma = (TE_a - |bias|) / CV_A$, where each value is expressed as a percentage to normalize σ over a wide range of quantities, and bias is expressed as the absolute value, so the direction of bias (positive or negative) is not a factor. In this equation, TE_a is a predetermined quality goal whereas bias and CV are determined during method-validation or verification studies (see Evaluating and Validating Laboratory Methods).

d. For theoretical assay A with a TE_a of 10 %, analysis of QCM with an analyte concentration near the decision threshold (100 mg/dL) reveals a SD of 5 mg/dL and

therefore a CV of 5 %. Assuming there is no systematic error (bias), this assay would be called a *two-sigma* assay because σ = 10 % / 5 % = 2 (and using concentrations instead of percentages, the TE_a divided by its SD is 2 also). When designing a QA program for this assay, there would be strict criteria established with several QCMs so that unacceptable errors could be detected when present. By random error alone, about 1 of 20 values is expected to be unacceptable.

e. For theoretical assay B, also with a TE_a of 10 %, analysis of QCM with an analyte concentration near the decision threshold (100 mg/dL) reveals a SD of 2 mg/dL and therefore a CV of 2 %. Assuming there is no systematic error (bias), this assay would be called a *five-sigma* assay because σ = 10 % / 2 % = 5 (and using concentrations instead of percentages, the TE_a divided by its SD is also 5). When designing a QA program for this assay, there would be relatively loose criteria established with two QCMs and one control rule (1_{3s}) because it is very unlikely that its random error would create changes that are clinically relevant (i.e., theoretically, less than one unacceptable value per million assays).

f. In the Six Sigma system, the ideal assay is a *six-sigma* assay because an unacceptable error would theoretically occur only 0.002 times per million assays (1 in 500 million). In a *six-sigma* assay, six standard deviations of variation are within the acceptable tolerance limits of the process.

VARIATIONS IN RESULTS ACROSS ASSAYS AND LABORATORIES

I. Not only do veterinarians face challenges of species variations, they also must deal with varied results generated by different laboratories and methods that are not all calibrated to a universal standard. There are many reasons for varied results; some of the differences are due to unpreventable random errors and others are caused by systematic errors (bias) identifiable and measurable by using method comparison studies.

II. Examples of variations

A. Table 1.10 and Figure 1.6 contain results of surveys completed by the Veterinary Laboratory Association (VLA) as part of an external QA program.[34] For such surveys, the VLA distributes aliquots of pooled samples of serum or whole blood to laboratories that subscribe to its QA program. After analysis of the distributed sample, each subscribing laboratory sends its results to the association for compilation and analysis. The association then distributes the results of the analyses. The type of program is sometimes referred to as a *proficiency testing* program and can be used to determine the validity of a laboratory's assay. Ideally, the results of an individual laboratory could be compared to the results established by a laboratory that analyzed the same sample by using a recognized reference method. Without such a value, the *proficiency testing* provides a method of periodically monitoring an assay's performance using *peer comparison*, the comparison of results to results of others who are using the same methods and types of analyzers.

B. Table 1.10 and Figure 1.6 illustrate marked differences in the results generated by laboratories, even for samples with few abnormalities (this dog's sample revealed only mild hypocalcemia, mildly increased alanine transaminase activity, and a 4.5× increase in alkaline phosphatase activity). Consequently, results from different assays or instruments within a laboratory or among laboratories (including multisite referral laboratories) should not be considered interchangeable unless this has been established by method comparison studies.

Table 1.10. Variation in serum chemistry results among labs using aliquots of one canine serum sample[a]

Analyte[b]	Number tested	Mean[c]	Unit	Low–high[c]	Range	Low to high as % change from mean	TE_a (%)[d]
Sodium	169	141	mmol/L	131–152	21	–7 to +8	5
Potassium	169	4.9	mmol/L	4.5–5.5	1.0	–8 to +12	10
Chloride	166	107	mmol/L	99–116	17	–7 to +8	5
tCa^{2+}	168	8.5	mg/dL	6.6–10.0	3.4	–22 to +18	10
Phosphorus	168	4.7	mg/dL	3.3–6.3	3.0	–30 to +34	20
tMg^{2+}	128	2.1	mg/dL	1.4–2.5	1.1	–33 to +19	20
Urea	165	21	mg/dL	17–29	12	–19 to +38	15
Creatinine	169	1.2	mg/dL	0.7–2.1	1.4	–42 to +75	20
Glucose	166	95	mg/dL	77–110	33	–19 to +16	20
Bilirubin, total	165	0.2	mg/dL	0.1–1.3	1.2	–50 to +550	30
Amylase	131	725	U/L	494–923	429	–32 to +27	25
ALP	164	459	U/L	311–1082	771	–32 to +136	25
CK	158	62	U/L	21–86	65	–66 to +39	30
AST	160	24	U/L	3–50	47	–88 to +108	30
ALT	164	83	U/L	52–112	60	–37 to +35	25
Total protein	165	6.1	g/dL	4.2–7.9	3.7	–31 to +30	10
Albumin	166	3.2	g/dL	2.8–4.1	1.3	–13 to +28	15
Cholesterol	162	264	mg/dL	221–294	73	–16 to +11	20

[a] Results from similar and different assay methods analyzed as part of one Veterinary Laboratory Association survey during 2022

[b] ALT, alanine transaminase; ALP, alkaline phosphatase; AST, aspartate transaminase; CK, creatine kinase; tMg^{2+}, total magnesium

[c] Mean values of all results for each analyte were provided in the survey report, while low and high values were extracted from graphs within the survey's report. All analyte values were rounded to reportable significant digits.

[d] TE_a, allowable total error, is the estimated error that is not expected to affect interpretation of results. These values are from TE_a guidelines of the American Society for Veterinary Clinical Pathology.[35] For each analyte, the difference between the lowest and highest values exceeds TE_a and could be clinically relevant if the values were obtained from the same animal over time. Similarly, for most analytes, the smaller changes from the lowest value to the mean and from the highest value to the mean could be considered clinically relevant, though three of each did not exceed the TE_a (see [tMg^{2+}], [glucose], [albumin], and [cholesterol]. Lack of agreement among laboratories emphasizes the need for establishment of reference intervals by each laboratory.

It is beyond the scope of this textbook to describe in detail the possible reasons for the differences; in general, results may vary because of one or more of the following:

1. Assay methods and instruments vary.
2. Commercial standard solutions vary across manufacturers and may be poorly calibrated against an appropriate reference. Although there are strict requirements for development of human assays, the same requirements for veterinary assays do not exist, in part because of species diversity.

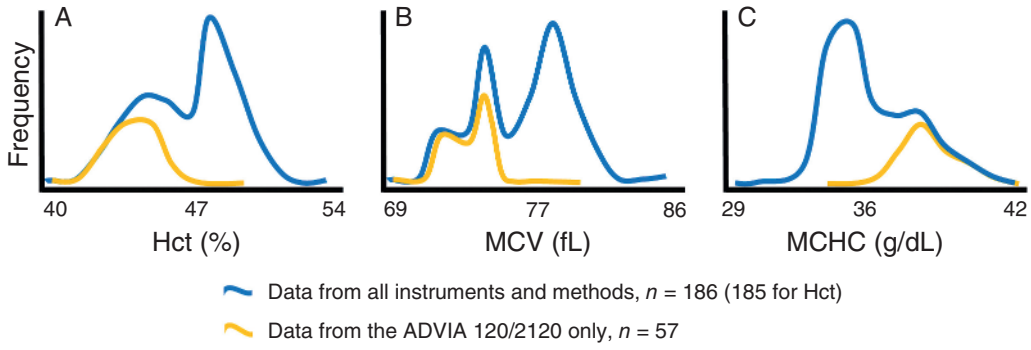


Figure 1.6. Interlaboratory variation in hematology results. These distribution curves reflect results from similar and different assay methods that analyzed aliquots of one canine whole blood sample as part of one survey completed by the Veterinary Laboratory Association (VLA) during 2022. These curves are close approximations of the histogram frequency distributions for Hct (A), MCV (B), and MCHC (C), three of the routine RBC analytes provided within the survey's report.

- *Blue curves*: These show the distribution of results for all participating laboratories, including all the different methods and analyzers used for these analytes. There is a wide variation in results. As for chemistry results of a similar survey in Table 1.10, the difference between the lowest and highest values would be considered clinically relevant if they represented values obtained from the same animal but different samples.
- *Gold curves*: These show the distribution for only those results generated by ADVIA analyzers. These results have narrower ranges than results from all laboratories because they were generated by the same methods and type of analyzer. Even for just this subset of results, the difference between the lowest and highest values would be considered clinically relevant if they represented values obtained from the same animal but different samples.
- Comparison of blue (all) and gold (ADVIA) curves also shows that the ADVIA results are lower (Hct and MCV) or higher (MCHC) than for other methods. The paired means for All and ADVIA results in this survey were, respectively: Hct, 47 % and 44 %; MCV, 77 fL and 73 fL; and MCHC, 35 g/dL and 38 g/dL.
- Lack of agreement among analyzers and methods necessitates establishment of RIs by each laboratory.

Hct, hematocrit; MCHC, mean cell hemoglobin concentration; MCV, mean cell volume

3. For assays that measure enzyme activity or depend on enzymatic activity within the assay, substantial differences in results can occur because of assay variations that include differences in substrates (reagents), reaction temperatures, or pH of assay systems (see Chapter 12).
4. Even though two laboratories might have the same instrument and assay method, there is inherent variability in analyzers, and the laboratories might operate them in different environments (e.g., altitudes) and with different maintenance schedules, or purchase reagents from different companies. Even if the reagents have the same constituents, they might be at different concentrations.

C. Because of the potential for marked differences in results determined by different laboratories, several recommendations are made:
1. Sample collection, processing, and transport should follow policies that minimize deterioration of sample quality.
2. Samples should be submitted to veterinary laboratories that adhere to robust QA guidelines.
3. A patient's results should be compared against appropriate RIs established for the same assay method in the same laboratory, when available. RIs established for another assay can be very misleading. Establishing de novo RIs and validation of transferred RIs from a manufacturer are described in the section titled Reference Intervals.

4. If comparing serial results across different analyzers, using the magnitude of change relative to the URL of each analyzer's RI may be more appropriate than directly comparing results. For example, if a dog with acute hepatic necrosis had a serum ALT activity of 1200 U/L (RI 75–120) at an emergency clinic and a second result of 600 U/L (RI 21–60) from a sample collected the next day at the patient's home clinic, the second result is half the first and might suggest improvement, but each is 10 times its respective URL, supporting that there is actually no change in hepatocellular damage. This presumes that both analyzers are performing well and that RIs were established appropriately.
5. A good diagnostician not only identifies but also confirms abnormalities that are at all questionable. If a laboratory result seems inconsistent with other findings for a case, a clinician should ask the laboratory to verify assay performance and, if possible, repeat the assay on the same sample and/or potentially use another analyzer that uses a different method.
6. Clinicians should establish strong professional relationships with their laboratory's personnel. A reputable laboratory strives to provide high-quality and timely results. As trusted healthcare providers, veterinarians should strive to provide their laboratories with high-quality samples.

WHICH LABORATORY SHOULD ONE USE?

I. Many factors must be considered before deciding how and where to obtain laboratory data for your patients.[46,47] Basically, there are three major options: (i) in-house laboratory, (ii) veterinary reference laboratory, and (iii) laboratory in local human hospital. Four major questions must be considered:
   A. How will the results affect the care of the patient?
   B. Can the laboratory consistently provide quality analysis of the sample? This includes access to expertise in veterinary clinical pathology.
   C. How much will it cost to analyze the sample?
   D. How important is a short turnaround time?

II. In-house laboratory (in a veterinary clinic or hospital)
   A. Advantages
      1. May allow 24-h access to laboratory data
      2. Shorter turnaround times, which are sometimes required for emergency cases (however, experts at outside laboratories may be particularly useful for sick patients)
      3. Fresh samples; thus fewer concerns about sample deterioration
      4. Client satisfaction because of fewer call-backs and repeat appointments (assuming no inaccurate results that complicate patient management)
      5. Potential for increased practice productivity, efficiency, and income
   B. Disadvantages
      1. Capital expenditure of thousands of dollars; equipment depreciates rapidly
      2. Need to maintain inventory of reagents and supplies
      3. Need sufficient personnel and meticulous training to proficiently operate and maintain equipment, recognize inappropriate samples, and process samples appropriately
      4. Need to follow a QA program and comply with any existing regulatory agency requirements to ensure high-quality results and minimize errors (Tables 1.7–1.9)
      5. Typically need high sample volume to generate enough income to cover expenses

6. RIs typically provided by the manufacturer of instruments or assay may or may not have been established appropriately or be relevant to the clinic's patient population or environment.
7. May need to ship a malfunctioning instrument to the manufacturer for repairs and thus need a replacement instrument or backup methods
8. Quality and value of the results are often inferior to those provided by veterinary reference laboratories, particularly for various microscopic assessments.

III. Veterinary reference laboratory
   A. Advantages
      1. Laboratory personnel are typically trained to provide quality analysis of veterinary samples and recognize samples that may lead to inaccurate results.
      2. Diagnostic support may be available from veterinary clinical pathologists in the laboratory, and veterinary specialists may be available for clinical consultations.
      3. RIs should be appropriate if derived from participating clinics' patient populations.
      4. Most laboratories have a quality management system in place to help ensure quality results. The system often includes proficiency testing by an external group such that performance can be compared to that of peers.
      5. Many more diagnostic assays are available than are available in an in-house laboratory.
      6. The cost for sample analysis is more clearly determined and thus can be charged to the client on a sample-by-sample basis.
      7. Some veterinary diagnostic companies manufacture in-clinic analyzers as well as operate reference laboratories with analyzers that allow higher throughput but that may employ methods and results similar to the in-clinic benchtop analyzers. Connectivity of in-house equipment with the manufacturer enhances monitoring of performance and troubleshooting.
   B. Disadvantages
      1. Not all analytes or samples are stable, and thus some deteriorate during shipment (e.g., urine, blood cells). Some analytes require special shipping, e.g., shipped frozen or in special tubes (collection tubes with added antiproteases).
      2. Turnaround times vary with location. Some are available the same day, and most are available the next day except for special tests.

IV. Laboratory in local human hospital
   A. An advantage is that turnaround times can be within hours.
   B. Disadvantages
      1. Quality RIs for veterinary samples may not be established by the laboratory.
      2. Assay methods may not be appropriate for veterinary samples. This is especially true for protein-based analytes (including enzymes and hormones) and immunologic-based assays.
      3. Technicians and technologists may not be trained to correctly identify species variations or diseases that are unique to veterinary samples.
      4. Pathologists who specialize in human patients might be very interested in providing consultation but lack training in the diseases of domestic animals and variations seen in veterinary samples.
      5. The fees charged by laboratories may be relatively high.
   C. Testing in local laboratories that specialize in human medicine should generally be avoided.

EVALUATING AND VALIDATING LABORATORY METHODS[24]

I. Reasons for evaluating laboratory methods
   A. New or revised laboratory methods may be evaluated for clinical use when the following are true: (i) recent research findings suggest a better approach, (ii) newer assays may be more accurate, more precise, more economical, or more efficient, or (iii) an instrument was purchased to replace an outdated or malfunctioning instrument.
   B. Before adopting new or revised methods, several questions should be asked. These include: Is the method practical? What are the equipment and space needs? Are personnel trained for the method? Will it improve patient care at a reasonable cost? Has there been adequate study of the method to prove its clinical value, or is the method still in the development stage? Are the requirements for sample collection, processing, and handling practical?

II. What are sources of analytical error?
   A. Imprecision: Each assay system has its own inherent random error causing imprecision (see Quality of Laboratory Results, Section II.B). Typically, the more precise the assay is, the better it is, but sometimes the most precise methods are too expensive or time consuming.
   B. Inaccuracy (bias): Besides random error, a method may not provide accurate results because of systematic error (bias). For example, the mean concentration of a new method may be consistently 5 mmol/L too high compared to the mean value determined by a reference method. In theory, *accuracy* is truth, but in reality, analytical accuracy is a relative term and sometimes difficult to assess in a clinical laboratory. An assay's result may be compared against one of two "true" values:
      1. The concentration determined by the reference method (a gold standard)
      2. The concentration determined by numerous analyses with comparative methods

III. Acceptable analytical performance, also called a quality goal
   A. Quality goals are often expressed as TE_a or the amount of error that does not change the interpretation and does not exceed the amount of error due to imprecision and biological variation. TE_a for an assay should be greater than TE_o, the amount of error measured (observed) in any assay.
   B. An assay does not have to be the most accurate, precise, specific, or sensitive assay to be clinically useful. It simply needs to fulfill predetermined requirements for acceptable performance.
   C. Criteria for acceptable performance of clinical assays have been proposed, and they vary considerably among analytes. A hematocrit method might be considered acceptable if it provides results within 10 % of the target value (e.g., 36–44 % for a hematocrit of 40 %). A creatinine method might be considered acceptable if creatinine concentrations are within 20 % of target values.[32,35]

IV. Analytical validation process
   A. Three stages of validation for analytical performance have been recommended with the goal of quantifying inherent analytical error (variation). This is in contrast to the validation of an assay's diagnostic performance described in the next section, a process pertaining to distinguishing between patients with and without a particular disorder. Detailed instructions for planning and carrying out rigorous method-validation studies have been published.[24]

B. Familiarization: This stage includes establishment of a working procedure, initial assessment of the analytical range, and calibration. Importantly, repetitive use serves as practice to "learn the assay," thereby promoting identification and reduction of errors and sources of variation.

C. Preliminary validation: In this stage, several studies are completed.
   1. Within-run precision (repeatability): to determine the degree to which promptly repeated analysis of the same samples yields the same results; if repeatability is poor, further validation may not be fruitful
   2. Linearity: to determine if signals (results) are proportional to analyte concentrations or activities, and over what analyte concentrations or activities they are linear and useful
   3. Interference studies: to determine the effects of common (e.g., hemolysis, bilirubin, lipemia) or potential relevant interferents; if clinically significant degrees of error (inaccuracy) are caused by clinically encountered concentrations of an interferent, results must be interpreted accordingly
   4. Recovery studies, if purified forms of the analyte are available: to determine if the assay is sufficiently accurate and can detect (recover) all the analyte added to samples (spiking) at relevant concentrations

D. Detailed validation: If initial results are satisfactory, then final validation includes the following components.
   1. Between-run precision (reproducibility) studies: to assess the reproducibility of results from a single sample tested over extended periods of time (e.g., 20 workdays, if the analyte is stable that long) to assess for imprecision caused by factors that vary over time (e.g., environment, operator)
   2. Comparison of methods with statistical analyses: to assess for accuracy compared to an accepted assay (see next section, Comparison of assays) in addition to, or instead of, a spiking and recovery study; patient specimens are the preferred samples
   3. Determination of acceptable performance criteria
   4. Establishment of RIs

E. Implementation phase
   1. If results of the validation procedures indicate that the assay is acceptable for use in the laboratory, then it needs to be incorporated into the daily routine of the laboratory, including equipment maintenance, reagent inventory, quality control procedures, and determination or validation of decision thresholds.
   2. When the assay is finally ready for clinical use, clinicians are notified of its availability, cost, and expected precision at the URL and LRL or major decision thresholds.

V. Comparison of assays: It is sometimes necessary to compare the results of two assays to determine the degree of analytical agreement between them, e.g., as part of validating a test by comparing it to a gold standard. Assays *agree* when they produce essentially the same results from the same samples and thus can be used interchangeably in making clinical decisions. Importantly, results may be strongly correlated but not in agreement. *Perfect correlation* means that the sample results for one assay always differ from those of another assay by the same percentage and direction; if results of one assay were always twice those of the other assay, the assays would correlate perfectly but they would not agree. The assays would have a *constant systematic bias* (difference between the results that is constant over all values). In contrast, if differences between assay results changed proportionately with the magnitude of the result, there would be a *proportional systematic bias*. Methods of assessing assay agreement and guidelines for completing and interpreting method comparison studies have been reviewed.[48,49] Veterinarians can benefit from awareness of what information the four major methods provide.

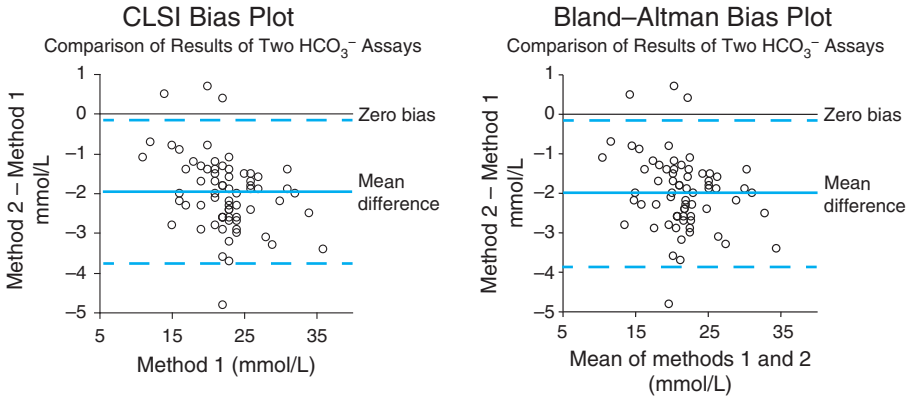


Figure 1.7. CLSI and Bland–Altman bias plots. The results were tabulated and entered in a software program (Analyse-it) to generate the bias plots.

- In the CLSI bias plot, the differences between HCO_3^- concentrations measured by two assays (method 2 – method 1 in this example) are plotted on the *y*-axis; the concentrations measured by method 1 are plotted on the *x*-axis. The mean difference for all samples is calculated. In this comparison, the mean difference is 2 mmol/L, so method 2 has a bias of –2 mmol/L compared to method 1. Lines representing ± 1.96 SD of the differences are also displayed as *limits of agreement*, the boundary for 95 % of the differences. The CLSI bias plot is the preferred plot when a second or test method is being compared to a reference method or an established (comparative) method.
- The Bland–Altman bias plot is the same as the CLSI bias plot except the average of the two measured values (average of method 1 and method 2 concentrations) are plotted on the x-axis. The Bland–Altman bias plot is used when neither method is considered to be more accurate.
- When results show a bias, a relatively flat distribution indicates a constant bias and a sloped distribution supports a proportional bias (this distribution tends to run from upper left to lower right, but more data would be needed to be sure). Determining if the bias is tolerable and manageable involves consideration of a quality goal for the assay (e.g., TEa), the number of results outside the limits of agreement, and the type of bias (proportional or constant).

HCO_3^-, bicarbonate

Choice of method is beyond the scope of this book, but depends on the type of result (continuous, discrete, ordinal), the range of data, the degree of correlation, and the type of any proportional or constant bias that is present.

A. Bias plots (Figure 1.7) for continuous data of narrow or wide range are particularly useful for analytes with a narrow range of data (e.g., electrolytes) and a correlation coefficient $r < 0.95$.
   1. CLSI bias plot (Note: CLSI was previously NCCLS, National Committee for Clinical Laboratory Standards, so this was called a NCCLS bias plot.)
      a. The CLSI plot displays the degree of agreement between paired values generated by a new method and the old method (or reference method). The values determined by the old/reference method are plotted on the *x*-axis, and the difference between values determined by the new and old method is plotted on the *y*-axis. Besides plotting the paired data points, the mean and SD of each difference is also plotted.
      b. The graph displays a positive bias (mean difference > 0) or a negative bias (mean difference < 0) and may indicate a *constant bias* (the plotted differences form a flat line significantly above or below the line of zero bias), a *proportional bias* (differences form a straight line that slopes up or down significantly compared to the line of zero bias), or both biases.

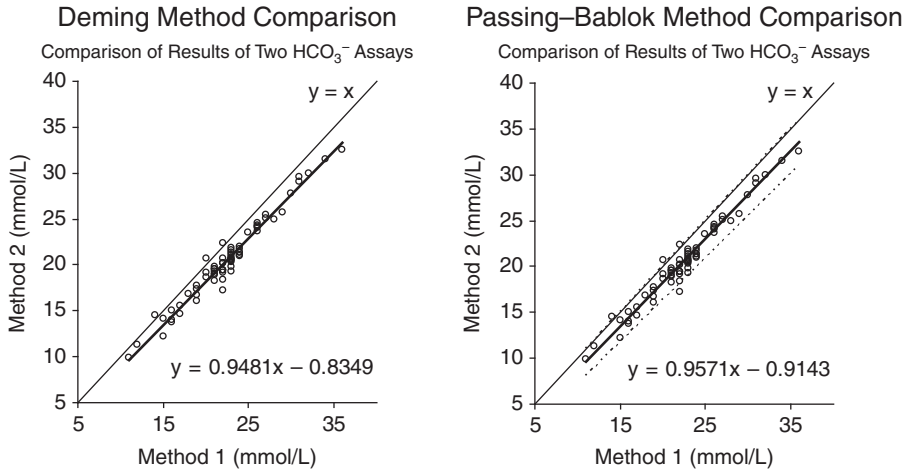


Figure 1.8. Deming and Passing–Bablok method comparisons. The results were tabulated and entered in a software program (Analyse-it) to generate the comparison graphs. The analysis of the two graphs reveals a mild proportional bias, i.e., the difference between results for the two methods increases at higher concentrations.

- In the Deming method, the results of the two assays are plotted on the *y*-axis and the *x*-axis, and a best-fit regression line is drawn and compared to the identity line ($y = x$). The equation for the best-fit line ($y = mx + b$) and associated confidence intervals are used to detect constant bias or proportional bias. In this method of comparison, imprecision of the assays should be normally distributed.
- The Passing–Bablok method comparison is very similar to the Deming method, but the imprecision need not be normally distributed and can have nonconstant variance over the sampling range. The 95 % confidence interval of the best-fit line is also displayed (dashed lines).

HCO_3^-, bicarbonate

2. The Bland–Altman bias plot is similar but with the following difference: the average of the two values is plotted on the *x*-axis and the difference between the two values is plotted on the *y*-axis. This method does not assume that either the new or the old method is more accurate.

B. Deming method comparison (Figure 1.8) for continuous data with wide range (e.g., many enzyme activities)
   1. In the Deming method comparison, the values obtained for a given sample are plotted with the new method on the *y*-axis and the old (or comparison) method on the *x*-axis. A regression equation for the best-fit line ($y = mx + b$) and the 95 % confidence intervals for the slope and *y*-intercepts are calculated.
      a. A *proportional bias* is detected if the 95 % confidence interval for the slope does not include 1 (i.e., slope of the identity line).
      b. A *constant bias* is detected if the 95 % confidence interval for the *y*-intercept does not include 0 (i.e., *y*-intercept of the identity line).
      c. Perfect agreement would yield a slope of 1 and a *y*-intercept of 0.
   2. The Deming method allows comparison of two independent variables which in this case are test methods, each with imprecision. As with any assessment of agreement, the comparison is improved with duplicate measurements by both methods because this reduces the effect of measurement imprecision.

3. In contrast to Deming regression, simple linear regression assumes that the *x* value is an independent variable with no error and the *y* value is dependent on *x*. A correlation coefficient (*r*) is computed to assess the strength and direction of the linear relationship between the two variables; $r = 0$ when there is no correlation and $r = 1$ when the correlation is perfect. A strong positive linear relationship is expected if two assays are designed to measure the same analyte concentration.
   a. Although method comparison studies involve *two independent* variables, linear regression statistics are used to assess agreement when errors in a reference method are small and the correlation coefficient is very high (e.g., $r > 0.99$).
   b. The correlation coefficient is very dependent on the range of data in the sample.[50, 51] A wide range of data tends to have a higher correlation than a narrow range. This is true even when most values are clustered and a single value stretches the range more widely. If such a value is aberrant, it can erroneously affect the relationship, suggesting more or less of a relationship than actually exists. Data sets should be broad and without large gaps between values.
   c. Figures 1.7 and 1.8 depict data with excellent correlation ($r = 0.98$) between data sets, but the bias plots illustrate there is a bias between the assays.

C. Passing–Bablok method comparison (Figure 1.8) for continuous data with a wide range
   1. The Passing–Bablok method is similar to the Deming method, but it is a nonparametric approach that does not require normally distributed measurement errors.
   2. This method estimates a linear regression line and tests whether the intercept is 0 and the slope is 1 as would be true of a line of identity.

D. Cohen's kappa test can be used to assess the agreement of two assays that generate results on ordinal scales, e.g., 1+, 2+, 3+, and 4+, or trace, mild, moderate, and marked. Details and caveats are beyond the scope of this book.
   1. The calculated kappa value provides a guide to the degree of agreement. Generally, agreement is considered poor with a kappa < 0.2, fair with kappa ≥ 0.2 and < 0.4, moderate with kappa ≥ 0.4 and < 0.6, good with kappa ≥ 0.6 and < 0.8, and very good with kappa of 0.8–1.0.
   2. The size or degree of disagreement is not considered with the standard kappa method, but it is with the weighted kappa method.

E. Interpretation of assay comparison studies
   1. After the quantitative aspects of the assays are compared, a decision is made as to whether the degree of agreement is acceptable such that the methods can be used interchangeably. Acceptability should be based on a performance goal, often TE_a, defined prior to the study.
   2. Visual inspection of graphed data may reveal that the results differ sufficiently so that the two assays cannot be considered the same. Thus, there would need to be different RIs and decision thresholds for the two assays.
   3. What would be the results of a comparison study if the new assay is actually better than the existing gold standard? The new assay could appear to perform poorly and might be rejected when it would actually be an improvement. This scenario occurs as better new tests are sought to replace poorer old tests. To validate accuracy of a new test in such a situation, spiking and recovery of purified analyte may be more useful than a method comparison.
   4. Because no assay is perfectly precise, differences between assays are expected even if they have excellent agreement. The amount of inherent difference can be assessed by calculating a combined inherent CV.[48]
   5. The acceptability of a new assay can also be assessed using a medical decision chart.[48]

F. Error grid analysis provides a visual assessment of agreement and is a nonparametric graphical approach for comparing the potential harm caused by differences in results between two measurement methods.[52] After acquiring approximately 100 paired measurements or more, data are plotted on a graph with the reference method measurement on the *x*-axis and the new method on the *y*-axis as for regression analysis. A so-called error grid is applied to the graph to denote three zones: A – the desirable region near the line of identity ($x = y$) that denotes no risk to the patient and therefore acceptable total error of the new test relative to the existing test, B – an intermediate zone outside of zone A, and C – a zone of unacceptably high errors likely to cause patient harm. For an appropriate replacement test, the vast majority of results should fall within zone A, a small proportion in zone B, and very few in zone C.

DIAGNOSTIC PROPERTIES AND PREDICTIVE VALUE OF LABORATORY ASSAYS

I. After ensuring that the analytical performance of an assay is acceptable for clinical use, valid RIs can be used to identify atypical results, but several disorders might be responsible for an increased or decreased result. If an assay is to be used to identify a particular disorder, the assay's diagnostic performance should be assessed to determine if the assay can distinguish between patients that have a disorder (disease positive) and patients that do not.[49] In so doing, it is important that the group of patients without the disorder includes patients with similar clinical findings and disorders that must be differentiated from the disorder of interest; if it includes only healthy patients, the observed diagnostic performance will be falsely optimistic and not reflective or predictive of actual performance in a clinical setting.

II. There are four classifications of test results relative to the presence or absence of a disease:
   A. TP (true positive): a result that correctly identified a patient as having a specified disease
   B. TN (true negative): a result that correctly identified a patient as not having a specified disease
   C. FP (false positive): a result that incorrectly identified a patient as having a specified disease
   D. FN (false negative): a result that incorrectly identified a patient as not having a specified disease

III. To classify test results into one of the four categories, two factors must be known:
   A. What is the diagnostic threshold (clinical decision limit) used to discriminate between individuals with the disease of interest (disease positive) and individuals with other disorders or no disorder (disease negative)?
      1. Is the decision limit either an upper or lower reference limit? Although RIs describe the expected location and spread of values in health, they generally reflect only the central 95 % of healthy individuals and are typically not useful to identify any particular disorder.[19]
      2. Or is the clinical decision limit within or outside of the RI? Clinical decision limits may be empirically determined (see ROC section) and applied, e.g., for pancreatic lipase immunoreactivity to diagnose canine pancreatitis, or for calf serum [IgG] to assess for transfer of passive immunity.
   B. What determines whether an animal has the disease of interest? That is, what is the gold standard that allows us to say that the animal definitely does or does not have the disease? For many spontaneous diseases, there may not be a gold standard, either a single test or specific constellation of findings. Diagnosis might be based on investigators' own criteria or consensus criteria from experts, which might require a preponderance of evidence from various tests, long-term follow-up, or even necropsy. Conclusions about the diagnostic value of a test must be interpreted relative to the quality of the selected diagnostic criteria.

IV. After test results are appropriately classified as TP, TN, FP, or FN, several calculations are made in an attempt to characterize the diagnostic properties or predictive value of the assay. It is important to differentiate diagnostic performance (e.g., diagnostic sensitivity, specificity, and accuracy), as discussed here, from analytical performance (analytical sensitivity, specificity, and accuracy), as discussed in the section on Quality of Laboratory Results (II.C, D, and F.)[49]

A. *Diagnostic sensitivity*

1. Definition: the frequency with which a test is positive in patients that have the disorder of interest (Eq. 1.2a)

In the following equations, multiply results by 100 to convert from decimals to percentages (e.g., 0.8 × 100 = 80 %); # stands for number.

$$\text{Diagnostic sensitivity} = \frac{\#\text{ of true positive}}{\#\text{ with specified disease}} = \frac{\text{TP }\#}{\text{TP }\# + \text{FN }\#} \quad (\mathbf{1.2a})$$

$$\text{Diagnostic specificity} = \frac{\#\text{ of true negative}}{\#\text{ without specified disease}} = \frac{\text{TN }\#}{\text{TN }\# + \text{FP }\#} \quad (\mathbf{1.2b})$$

$$\text{Diagnostic accuracy} = \frac{\#\text{ correctly classified}}{\#\text{ of animals in study}} = \frac{\text{TP }\# + \text{TN }\#}{\text{TP }\# + \text{FP }\# + \text{TN }\# + \text{FN }\#} \quad (\mathbf{1.2c})$$

$$\text{PPV} = \frac{\#\text{ of true positive}}{\text{all positive results}} = \frac{\text{TP }\#}{\text{TP }\# + \text{FP }\#} \quad (\mathbf{1.2d})$$

$$\text{NPV} = \frac{\#\text{ of true negative}}{\text{all negative results}} = \frac{\text{TN }\#}{\text{TN }\# + \text{FN }\#} \quad (\mathbf{1.2e})$$

$$\text{PPV} = \frac{\text{sensitivity} \times \text{prevalence}}{(\text{sensitivity} \times \text{prevalence}) + [(1 - \text{specificity}) \times (1 - \text{prevalence})]} \quad (\mathbf{1.2f})$$

$$\text{NPV} = \frac{\text{specificity} \times (1 - \text{prevalence})}{[(1 - \text{sensitivity}) \times \text{prevalence})] + [\text{specificity} \times (1 - \text{prevalence})]} \quad (\mathbf{1.2g})$$

2. A test that has high diagnostic sensitivity is a good one for screening for the presence of a disease because the test has very few FN results. If the animal has the disease, there is a high probability that the test will be positive. Conversely, if the test is negative, it is likely that the animal does not have the disorder, although this depends on other factors, including the prevalence of the disorder in the population.
3. A mnemonic for the concept of a high-sensitivity test helping to *rule out* a disorder is SnNOut (high **Sen**sitivity **N**egative rules disease **Out**).[53]

B. *Diagnostic specificity*

1. Definition: the frequency with which a test is negative in patients that do not have the disorder of interest (Eq. 1.2b)

2. A test that has high diagnostic specificity can be a good one for confirming (ruling in) that an animal has a disorder because the test has very few FP results. If the result is positive, there is a high probability that the animal will have the disorder, although this depends on other factors, including the prevalence of the disorder in the population. A mnemonic for the concept of a high-specificity test helping to rule in a disorder is SpPIn (high **Sp**ecificity **P**ositive rules disease **In**).[53]

C. *Diagnostic accuracy*
1. Definition: the frequency with which a test correctly classifies an animal as having or not having the disorder (Eq. 1.2c)
2. A test has high diagnostic accuracy when it has relatively few FP and FN results compared to TP and TN results.

D. *Positive predictive value (PPV)*, the predictive value of a positive test
1. Definition: the probability that a positive test result indicates that the animal has the disease (Eq. 1.2d)
2. A test that has a high PPV is one that has very few FP compared to TP results. Thus, a positive test result strongly suggests the presence of the disease.

E. *Negative predictive value (NPV)*, the predictive value of a negative test
1. Definition: the probability that a negative test result indicates that the animal does not have the disease (Eq. 1.2e)
2. A test that has a high NPV is one that has very few FN compared to TN results. Thus, a negative test result strongly suggests the absence of the disease.

V. Basic concepts of predictive values: Three major questions are considered when the diagnostic properties of an assay are evaluated.

A. What criteria are used to establish the presence or absence of disease? Definitive diagnoses are sometimes very difficult to reach for spontaneous diseases, and errors in disease state classification affect predictive value results.

B. What is the decision threshold chosen to indicate a particular disorder is present? Extensive evaluation of an assay in a carefully planned study is needed to find the best clinical decision limit, and this should be done prior to assessment of diagnostic performance. The effects of changing the decision threshold on diagnostic performance are illustrated in Figure 1.9. Different decision thresholds may be used for making different clinical decisions for the same disorder or for the same analyte for different disorders.

C. What is the prevalence of the disease in the studied population? The effect of prevalence is illustrated in Figure 1.10. Basically, when the disease prevalence is very low, there are relatively few TP results and a relatively greater percentage of FP results, so the PPV will be lower. Conversely, when the disease prevalence is very high, it is more likely that there will be FN results; the NPV will be lower and the PPV will be higher compared to a population with a lower prevalence of the disease.[54] Unlike predictive values, diagnostic sensitivity and specificity are calculated based only on test and disease positivity, and test and disease negativity, respectively; therefore, prevalence does not affect them.

VI. Application and interpretation of predictive value concepts

A. The clinical value of the calculated diagnostic properties is influenced by the prevalence of a disease. For examples A and B in Figure 1.10, assume the diagnostic sensitivity of a test is 90 % and its diagnostic specificity is 80 %.
1. In example A of Figure 1.10, with a disease prevalence of 30 %:
   a. The PPV is 66 %; for all positive test results, 66 % are TP results.
   b. The NPV is 95 %; for all negative test results, 95 % are TN results.

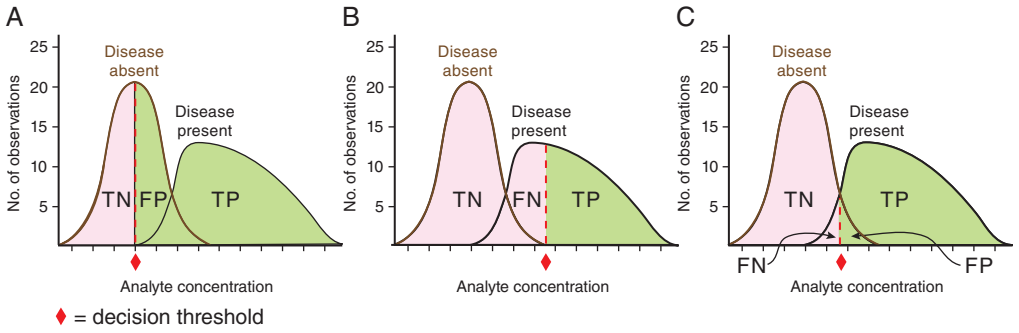


Figure 1.9. Effects of different decision thresholds on classifying test results. In these examples, the distribution of observed values in the animals without the disease appears to be Gaussian; such a distribution may or may not be true in real studies. The distribution of data in the diseased group is not Gaussian; typically, such data are not Gaussian but may not be skewed as shown in this example.

A. The decision threshold is near the mean analyte concentration found in the animals without the disease. With such a decision threshold:
   - The diagnostic sensitivity is 100 % because there are no FN results.
   - The diagnostic specificity is 50 % because there are equal numbers of TN and FP results.
   - The diagnostic accuracy is poor because of the many FP results.
   - The PPV is poor because of the many FP results
   - The NPV is 100 % because there are no FN results.

B. The decision threshold is at the highest value found in the animals without the disease.
   - The diagnostic sensitivity is poor (~ 60 %) because there are relatively many FN results.
   - The diagnostic specificity is 100 % because there are no FP results.
   - The diagnostic accuracy is poor because of the many FN results.
   - The PPV is 100 % because there are no FP results.
   - The NPV is poor because there are many FN results.

C. The decision threshold is at a concentration with the least overlap between the groups.
   - This decision threshold represents a compromise to obtain the best combination of diagnostic sensitivity and diagnostic specificity, and provides the best diagnostic accuracy because there are relatively few FP and FN results.
   - The values for PPV and NPV are high but not 100 %.

2. In example B of Figure 1.10, with a disease prevalence of 1 %:
   a. The PPV is 4.3 %; for all positive test results, 4.3 % are TP results.
   b. The NPV is 99.9 %; for all negative test results, 99.9 % are TN results.
3. By comparing the results of the two examples, these conclusions are formed:
   a. As the prevalence of the disease dropped from 30 % to 1 %, the PPV decreased from 66 % to 4.3 %. Thus, the PPV is less when the prevalence of the disease is lower.
   b. As the prevalence of the disease dropped from 30 % to 1 %, the NPV increased from 95 % to 99.9 %. Thus, the NPV is greater when the prevalence of the disease is lower.

B. The predictive value concepts are used to compare the diagnostic value of two laboratory tests. To illustrate this application, data were extracted from an article that compared the diagnostic value of serum $[tT_4]$ and $[fT_4]_{ed}$ for diagnosing feline hyperthyroidism.[55] As described in Chapter 17, serum $[tT_4]$ and $[fT_4]_{ed}$ may increase in sera of cats with hyperthyroidism, but other factors can influence $[tT_4]$ and $[fT_4]_{ed}$.

Example A – 30 % disease prevalence

Step 1	Disease present		Disease absent	
Positive test	(TP)	270	(FP)	140
Negative test	(FN)	30	(TN)	560
Totals		300		700

Step 2	Disease present		Disease absent		Totals
Positive test	(TP)	270	(FP)	140	410
Negative test	(FN)	30	(TN)	560	590
Totals		300		700	1000

Step 3

$$\text{Predictive value of positive test (as \%)} = \frac{\text{TP \#}}{\text{TP \# + FP \#}} \times 100 = \frac{270}{410} \times 100 = 66\ \%$$

$$\text{Predictive value of negative test (as \%)} = \frac{\text{TN \#}}{\text{TN \# + FN \#}} \times 100 = \frac{560}{590} \times 100 = 95\ \%$$

Example B – 1 % disease prevalence

Step 1	Disease present		Disease absent	
Positive test	(TP)	9	(FP)	198
Negative test	(FN)	1	(TN)	792
Totals		10		990

Step 2	Disease present		Disease absent		Totals
Positive test	(TP)	9	(FP)	198	207
Negative test	(FN)	1	(TN)	792	793
Totals		10		990	1000

Step 3

$$\text{Predictive value of positive test (as \%)} = \frac{\text{TP \#}}{\text{TP \# + FP \#}} \times 100 = \frac{9}{207} \times 100 = 4.3\ \%$$

$$\text{Predictive value of negative test (as \%)} = \frac{\text{TN \#}}{\text{TN \# + FN \#}} \times 100 = \frac{792}{793} \times 100 = 99.9\ \%$$

Figure 1.10. Examples of diagnostic properties of assays. These examples demonstrate the effect of prevalence on predictive values. For each example, the values for diagnostic sensitivity and specificity are 90 % and 80 %, respectively.

- *In example A*, we discovered (via a gold standard) that 30 % of 1000 dogs have the disease. Based on the prevalence, what is the test's positive predictive value? What is the test's negative predictive value?
  - *Step 1*: Construct a table from the available information. Because 30 % of the dogs have the disease, 300 dogs have the disease and 700 dogs do not. Given the diagnostic sensitivity of 90 %, 90 % (270) of the 300 diseased dogs will have a positive test result (TP) and 30 will have a negative test result (FN). Because the diagnostic specificity is 80 %, 80 % (560) of 700 dogs will have negative results (TN) and 140 will have a positive result (FP).
  - *Step 2*: Add the values for the number of positive and negative results.
  - *Step 3*: Calculate the positive predictive value and negative predictive value by using Eqs. 1.2d and e.
- *In example B,* we discovered (via a gold standard) that 1 % of 1000 dogs have the disease. Based on the prevalence, what is the test's positive predictive value? What is the test's negative predictive value?
  - *Step 1*: Construct a table from the available information. Because 1 % of the dogs have the disease, 10 dogs have the disease and 990 dogs do not. Given the diagnostic sensitivity of 90 %, 90 % (9) of the 10 diseased dogs will have a positive test result (TP) and 10 % (1) will have a negative test result (FN). Because the diagnostic specificity is 80 %, 80 % (792) of 990 dogs will have negative results (TN) and 198 will have a positive result (FP).
  - *Step 2*: Add the values for the number of positive and negative results.
  - *Step 3*: Calculate the positive predictive value and negative predictive value by using Eqs. 1.2d and e.
- When prevalence and diagnostic sensitivity/specificity data are known, positive and negative predictive values can be calculated from Eq. 1.2f and 1.2g. Plugging in different prevalence values to those equations demonstrates the effect of prevalence on predictive values.

1. Gold standard for this study
   a. Cats were classified as having hyperthyroidism by using the following clinical or laboratory findings:
      (1) Clinical signs consistent with hyperthyroidism
      (2) Palpable thyroid nodule
      (3) Good clinical response to treatment for hyperthyroidism
      (4) Basal [tT_4] increased or basal [tT_4] not increased, but positive results of triiodothyronine-suppression or thyrotropin-releasing hormone stimulation test
   b. Cats were classified as not having hyperthyroidism if they did not meet hyperthyroidism criteria. These cats had clinical signs suggestive of hyperthyroidism (e.g., weight loss, vomiting, diarrhea, and polyuria), but none had a palpable thyroid mass, none had increased [tT_4], and all had a diagnosis other than hyperthyroidism. Note that healthy individuals were not included, as is appropriate for studies of diagnostic accuracy involving tests intended to distinguish a specific disorder, in this case hyperthyroidism, from other disorders with similar clinical findings. Inclusion or use only of healthy subjects can inappropriately augment assay diagnostic performance.
2. Classification of results
   a. tT_4 results were classified as positive if [tT_4] was > 48 nmol/L. fT_4 results were classified as positive if $[fT_4]_{ed}$ was > 51 pmol/L. Decision thresholds represented the URLs determined from 172 healthy cats.
   b. Results were classified as negative if they did not meet positive criteria.
3. Based on these criteria for the 1138 cats:
   a. 917 cats had hyperthyroidism. Of these, 837 had an increased [tT_4] and 903 had increased $[fT_4]_{ed}$.
   b. 221 cats did not have hyperthyroidism. Of these, none had increased [tT_4] and 14 had increased $[fT_4]_{ed}$.
4. From the data provided, tables were constructed to show the classification of test results. From the tabulated data, the diagnostic properties and predictive values of [tT_4] and $[fT_4]_{ed}$ were calculated (Figure 1.11).
5. Based on the evaluation of reported data and application of the aforementioned gold standard and decision thresholds, these conclusions can be drawn:
   a. Serum $[fT_4]_{ed}$ had better diagnostic sensitivity (98 %) than did serum [tT_4] (91 %) for detecting feline hyperthyroidism. Thus, serum $[fT_4]_{ed}$ would be a better screening test for hyperthyroidism, i.e., more cats with hyperthyroidism have increased $[fT_4]_{ed}$ than increased [tT_4].
   b. Serum [tT_4] had better diagnostic specificity (100 %) than did serum $[fT4]_{ed}$ (94 %), i.e., [tT_4] had fewer FP results. Thus, an increased serum [tT_4] is more indicative of feline hyperthyroidism than is an increased serum $[fT_4]_{ed}$.
   c. Serum $[fT_4]_{ed}$ had better diagnostic accuracy (98 %) than did serum [tT_4] (93 %). Thus, if only [tT_4] or $[fT_4]_{ed}$ can be determined, serum $[fT_4]_{ed}$ has a better chance of correctly classifying the cat as having or not having hyperthyroidism. However, measuring $[fT_4]_{ed}$ is more expensive than measuring [tT_4].
   d. The PPV for increased [tT_4] was 100 %, and it was 98 % for $[fT_4]_{ed}$.
   e. The NPVs for [tT_4] and for $[fT_4]_{ed}$ were 73 % and 94 %, respectively. These results indicate that a $[fT_4]_{ed}$ that is WRI would strongly suggest that a cat does not have hyperthyroidism because there were very few FN results.

[tT_4] in feline sera

	Hyperthyroidism present		Hyperthyroidism absent		Totals
Positive test	(TP)	837	(FP)	0	837
Negative test	(FN)	80	(TN)	221	301
Totals		917		221	1138

$$\text{Diagnostic sensitivity (as \%)} = \frac{\text{TP \#}}{\text{TP \#} + \text{FN \#}} \times 100 = \frac{837}{917} \times 100 = 91\ \%$$

$$\text{Diagnostic specificity (as \%)} = \frac{\text{TN \#}}{\text{TN \#} + \text{FP \#}} \times 100 = \frac{221}{221} \times 100 = 100\ \%$$

$$\text{Diagnostic accuracy (as \%)} = \frac{\text{TP \#} + \text{TN \#}}{\text{TP \#} + \text{FP \#} + \text{TN \#} + \text{FN \#}} \times 100 = \frac{1058}{1138} \times 100 = 93\ \%$$

$$\text{Predictive value of positive test (as \%)} = \frac{\text{TP \#}}{\text{TP \#} + \text{FP \#}} \times 100 = \frac{837}{837} \times 100 = 100\ \%$$

$$\text{Predictive value of negative test (as \%)} = \frac{\text{TN \#}}{\text{TN \#} + \text{FN \#}} \times 100 = \frac{221}{301} \times 100 = 73\ \%$$

[fT_4]$_{ed}$ in feline sera

	Hyperthyroidism present		Hyperthyroidism absent		Totals
Positive test	(TP)	903	(FP)	14	917
Negative test	(FN)	14	(TN)	207	221
Totals		917		221	1138

$$\text{Diagnostic sensitivity (as \%)} = \frac{\text{TP \#}}{\text{TP \#} + \text{FN \#}} \times 100 = \frac{903}{917} \times 100 = 98\ \%$$

$$\text{Diagnostic specificity (as \%)} = \frac{\text{TN \#}}{\text{TN \#} + \text{FP \#}} \times 100 = \frac{207}{221} \times 100 = 94\ \%$$

$$\text{Diagnostic accuracy (as \%)} = \frac{\text{TP \#} + \text{TN \#}}{\text{TP \#} + \text{FP \#} + \text{TN \#} + \text{FN \#}} \times 100 = \frac{1110}{1138} \times 100 = 98\ \%$$

$$\text{Predictive value of positive test (as \%)} = \frac{\text{TP \#}}{\text{TP \#} + \text{FP \#}} \times 100 = \frac{903}{917} \times 100 = 98\ \%$$

$$\text{Predictive value of negative test (as \%)} = \frac{\text{TN \#}}{\text{TN \#} + \text{FN \#}} \times 100 = \frac{207}{221} \times 100 = 94\ \%$$

Figure 1.11. Analysis of the diagnostic properties of serum [tT_4] and [fT_4]$_{ed}$. The source of the data for this figure is explained in the text. Conclusions are also presented in the text.

- *[tT4] data*
  - From the data provided, a table was constructed to show the classification of test results.
  - From the tabulated data, the diagnostic properties and predictive values of serum [tT_4] were calculated.
- *[fT_4]$_{ed}$ data*: the same procedures were completed.

f. Note that the results of such studies may vary with different assays, different populations, different gold standards, and different decision thresholds.
   (1) $[tT_4]$ was used to determine the presence or absence of hyperthyroidism. The gold standard used (or method of establishing the presence of hyperthyroidism) in the study is considered an excellent method of establishing the presence of feline hyperthyroidism, but it would have been interesting to learn whether the diagnostic value data for $[tT_4]$ would change if $[tT_4]$ had not been used to help determine the presence or absence of hyperthyroidism.
   (2) Most cats in the study had been referred to specialists for treatment of hyperthyroidism. Thus, there was a high prevalence of hyperthyroidism in the study's population. A high prevalence increases the PPV of diagnostic tests and lowers the NPV. Accordingly, the predictive values of $[tT_4]$ and $[fT_4]_{ed}$ would probably be different in a nonreferral veterinary practice.

VII. Application of the aforementioned methods to evaluate and compare diagnostic methods requires careful planning, appropriate choices of disease-present and disease-absent populations, and availability of an excellent gold standard.[56]
   A. A deficiency in the diagnostic properties and predictive value theories is that a positive result is given the same weight or importance if the value is only slightly increased or if it is extremely increased. In practice, the degree to which a result is abnormal does affect clinical decisions.
   B. After a diagnostic test (assay) has been thoroughly studied in a specific clinical context, findings regarding its diagnostic performance provide guidance for its expected performance in other laboratories with different patient populations and clinicians. However, six criteria should be fulfilled before the diagnostic properties (predictive value, diagnostic accuracy, etc.) can be considered completely transferrable to another clinical setting:[57]
      1. The definition of the disease is constant, and the diagnostic criteria are applied in the same way such that two animals with the same abnormalities in two different settings would be diagnosed with the same disorder.
      2. The same test is used. However, as described in earlier sections, complete agreement between two analytical assays may not be present even if the same instrument and reagents are used.
      3. The decision thresholds between categories of test results are constant and clear, i.e., two different observers should agree on results that are *positive* and on results that are *negative*.
      4. The distribution of test results in the disease group is constant, i.e., the data should have the same average and same distribution curve. However, as for RIs, decision limits are specific for the demographics of the population in which they are determined, so it may be difficult to fulfill this criterion. The demographics of the patient population in a location to which a decision limit is transferred might not fully replicate the original patient population in terms of severity or stage of disease. Data for animals with mild disease probably differ from those with severe disease.
      5. The distribution of test results in the disease-absent group is constant, i.e., the data should have the same average and same distribution curve. It may be difficult to fulfill this criterion. Certain disorders may be rare in a primary-care clinic, yielding more FP results, but very common in a referral clinic, resulting in more FN results, a lower NPV, and a higher PPV.

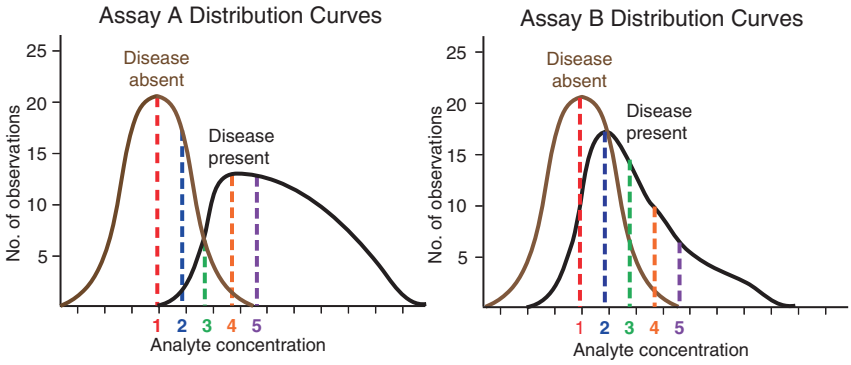
6. The ratio of disease-present to disease-absent individuals (pretest probability) is constant. Again, this criterion may be difficult to fulfill because certain disorders may be rare in a primary-care clinic but very common in a referral clinic.

C. **STA**ndards for **R**eporting of **D**iagnostic accuracy studies (STARD) are guidelines that were created to improve the completeness and transparency of diagnostic accuracy studies and reports.[58,59] A STARD checklist facilitates compliance with these guidelines and may be recommended or required by a journal's editorial board. Improved assessment and reporting leads to better diagnostic assays and improved patient care.

RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES[60,61]

I. ROC curves were originally developed to assess the ability of radar images to detect enemy aircraft in World War II, and therefore to assess the ability to detect true signals from background noise. In the context of laboratory tests, the ROC curves display the relationship between a TP rate and a FP rate.
   A. *TP rate* is equal to diagnostic sensitivity expressed as a decimal (sometimes reported as a %). For example, when the diagnostic sensitivity is 90 %, the TP rate is 0.9; results are positive in 9 of 10 disease-present animals.
   B. *FP rate* is equal to 1 minus the diagnostic specificity expressed as a decimal (sometimes reported as a %). For example, when the diagnostic specificity is 70 %, the FP rate is 1 – 0.7 = 0.3, and 3 of 10 disease-absent animals would have a FP result. When the diagnostic specificity is 100 %, there would be no FP results and thus the FP rate would be 0.0.
   C. Figure 1.12 shows theoretical results from the comparison of two assays, assay A and assay B.

II. The clinical value of the comparison of diagnostic procedures by ROC curves depends on many factors. Major issues to be addressed include the following:
   A. The assays should be analytically valid and applicable to clinical investigations.
   B. Selection of the comparison groups should be clinically relevant. Both groups should have similar clinical features (e.g., both have polyuria, vomiting, or anemia) so that the ability of the assays to differentiate relevant disorders is evaluated. Conversely, comparison of a sick group versus a healthy group is probably not appropriate or needed (the animals could typically be defined as healthy without this laboratory testing).
   C. The accuracy of the comparison between diagnostic procedures is dependent on the accuracy of the differentiation between disease-present and disease-absent groups. Results of the comparison need to be interpreted accordingly.
   D. ROC curves can be used to compare the diagnostic accuracy of assays.
      1. A visual inspection of Figure 1.12 reveals that the ROC curve for assay A is closer to the top left corner than assay B, and thus assay A has better diagnostic accuracy than assay B.
      2. A more objective comparison can be accomplished by calculating the area under the curve for each assay. Generally, the greater the area under the curve, the greater the diagnostic accuracy of the assay. However, these values should be interpreted carefully if the shapes of the ROC curves differ.
   E. ROC curves can be used to help establish a decision threshold for diagnosing a disorder (Figure 1.13). This is most commonly based on the point in the ROC curve that maximizes the sum of sensitivity and specificity, i.e., the point closest to the upper left corner of the plot. Other than being used to establish decision thresholds, Figure 1.13 also provides information that should be considered when interpreting the test results of assay C.



Assay A Diagnostic Properties

Decision threshold	1	2	3	4	5
Dx sensitivity	1.00	0.98	0.90	0.80	0.60
Dx specificity	0.50	0.60	0.90	0.97	1.00
1 – Dx specificity	0.50	0.40	0.10	0.03	0.00

Assay B Diagnostic Properties

Decision threshold	1	2	3	4	5
Dx sensitivity	0.90	0.70	0.50	0.35	0.20
Dx specificity	0.52	0.70	0.90	0.97	1.00
1 – Dx specificity	0.48	0.30	0.10	0.03	0.00

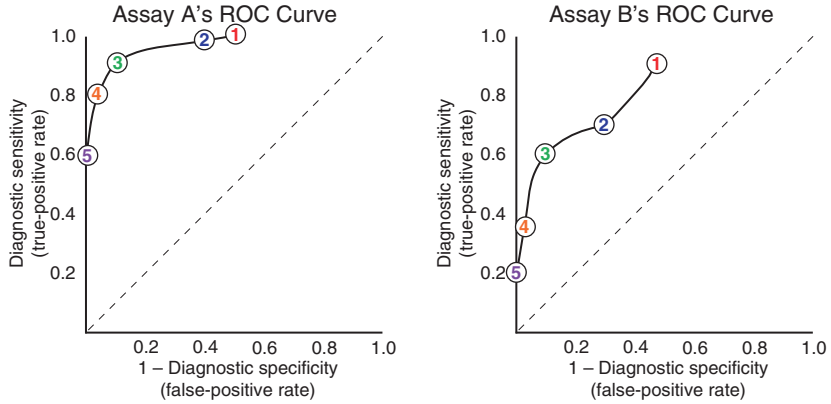


Figure 1.12. Comparison of diagnostic value of two theoretical assays by ROC curves.

- The initial step of the evaluation is the analysis of samples from two groups of animals (disease present and disease absent) by the two assays. The presence or absence of disease is established by a gold standard procedure.
- The data are plotted to obtain the distribution curves (*top curves in figure*). The assays generate different results and therefore curves for the disease group. To gather data for the ROC curves (*lower curves*), multiple decision thresholds (analyte concentrations) are selected that will provide different diagnostic sensitivity and specificity values. The decision thresholds are then used to classify actual measured concentrations as being TP, FP, TN, or FN results. For the illustration in the top graphs, five decision thresholds were selected, and the dashed lines represent the separation of positive and negative results at each decision threshold.
- From the classified data, the diagnostic sensitivity (TP rate) and specificity (TN rate) values are calculated for each decision threshold. (For this illustration, the number of animals in both groups was estimated from the graphs with an assumption that the total number in each group was equal.) The FP rate is calculated by subtracting diagnostic specificity from 1.
- The decimal fractions for TP rate and FP rate are plotted in the bottom graph. The 45° dashed line represents the ROC curve that would be obtained by random classification (e.g., flipping a coin to classify animals as disease present or disease absent). The best ROC curve approaches the top left corner of the graph where nearly all positive results are TP results, and the area under the curve is greater (not calculated here). In this comparison, assay A is a better diagnostic procedure than assay B for detecting the assessed disease.

Dx, diagnostic

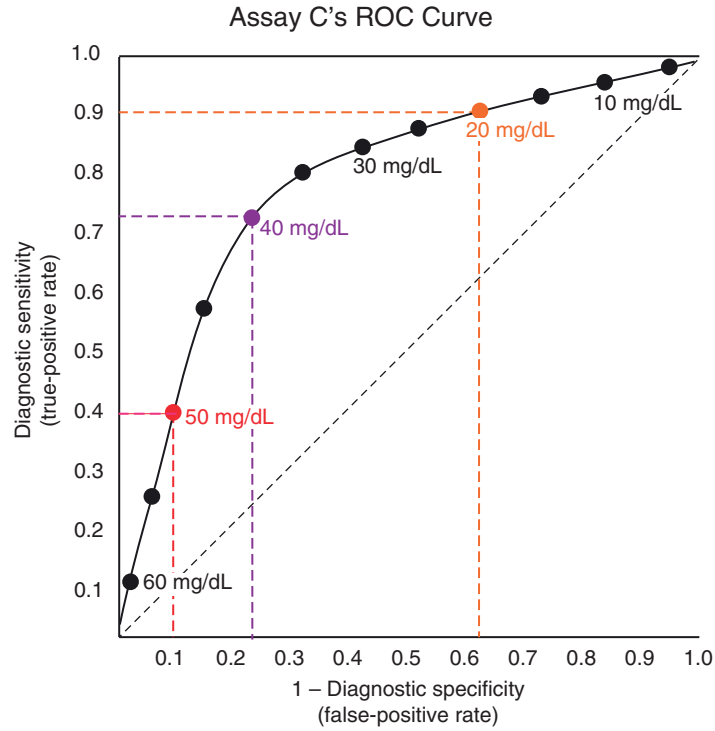


Figure 1.13. Selecting decision thresholds with ROC curves. In this theoretical assay C, the analyte concentrations are plotted on the ROC curve for assay C.

- The top-leftmost point (*purple*) provides the best balance between diagnostic sensitivity and diagnostic specificity and is often chosen as a decision threshold to conclude that a patient has a particular condition. Values of > 40 mg/dL would be considered evidence of the disease. With this decision threshold, the assay would have a diagnostic sensitivity of 73 % and a diagnostic specificity of 78 %.
- For use as a screening test, one might choose to set the decision threshold for disease at 20 mg/dL, accepting only 10 % FNs (i.e., 90 % diagnostic sensitivity). Values of > 20 mg/dL would suggest the presence of the disease. With this decision threshold, the assay would have a high diagnostic sensitivity (90 %), but a diagnostic specificity of only 37 %.
- For a confirmatory test, the decision threshold for disease may be set at 50 mg/dL to accept only 10 % FPs (i.e., 90 % diagnostic specificity, or an FP rate = 1.0 – 0.9 = 0.1). Values of > 50 mg/dL suggest that the animal has the disease. With this decision threshold, the diagnostic specificity would be high (90 %), but the diagnostic sensitivity would be only 40 %.
- One could also use the ROC curve to determine upper and lower clinical decision thresholds. A value of 20 mg/dL might be chosen as the lower decision threshold, so animals with values of < 20 mg/dL would be considered likely not to have the disease. A value of 50 mg/dL might be chosen as the upper decision threshold, so animals with values of > 50 mg/dL would be considered likely to have the disease. Based on this theoretical example, what would be the decision thresholds if you decide to accept 5 % FNs and 5 % FPs?

Answer: < 10 mg/dL – likely not to have the disease; > 55 mg/dL – likely to have the disease

1. Consider a patient with clinical signs suggestive of the disease of interest and an assay C result of 40 mg/dL.
   a. At 40 mg/dL, assay C has a FP rate near 0.22, or a diagnostic specificity of 78 %. In other words, of 100 dogs that do not have the disease of interest, about 22 will have an analyte concentration of > 40 mg/dL (22 % of those without the disease

classified as diseased). Therefore, if the result is > 40 mg/dL, the patient may or may not have the disease, but disease is more likely than not.

b. At 40 mg/dL, assay C has a TP rate near 0.73, or a diagnostic sensitivity near 73 %. So, of 100 dogs that do have the disease of interest, 73 will have an analyte concentration of > 40 mg/dL (73 % of those with the disease classified as diseased). Therefore, despite a value that is not > 40 mg/dL, the patient may still have the disease because 27 % of patients with the disease have values of 40 mg/dL or less.

c. This example emphasizes that for many diagnostic tests, animals with and without the disease of interest may have the same assay result.

2. Another concept related to this illustration is that each ROC curve represents the results obtained from one sampling of the defined populations by using one assay as it compares to a diagnosis based on one set of criteria. Another sampling using the same selection criteria, the same assay, and the same diagnostic criteria probably will not produce the same ROC curve.

F. Factors other than the probability of correctly classifying an animal as having or not having a disease can influence where decision thresholds are established:

1. If the mortality rate is very high for a disease that is not treated promptly, or if the disease is known to cause pain or unreasonable discomfort, the decision threshold may be changed so that there are fewer false negatives (greater diagnostic sensitivity).
2. If the treatment is known to have severe side effects or the financial cost of a certain diagnosis is great, the decision threshold may be changed so that there are fewer false positives (greater diagnostic specificity).

HERD-BASED TESTING FOR CATTLE

I. Other than using results of laboratory assays to help detect or establish the presence of pathologic states in individual animals, clinical laboratory assays also are used to evaluate the nutritional and metabolic status of groups of cattle. These evaluations should complement other management tools such as ration evaluation, milk composition analysis, body condition scoring, and monitoring disease incidence (retained placenta, displaced abomasum, milk fever, and ketosis). The goal of herd-based testing (metabolic profiling) is to detect subclinical laboratory abnormalities that reflect correctible nutritional management practices so that changes can be instituted prior to adverse clinical and economic consequences. Metabolic profiling may be most applicable to large herds where (i) adequate sample sizes are available, (ii) costs are diluted across more animals, and (iii) correcting a suboptimal nutritional state could translate into considerable economic savings.

II. The approach to group testing differs in several ways from individual animal testing with diagnostic profiles.[62–65]

A. Emphasis is placed on evaluating and monitoring subclinical rather than clinical disease. Clinically ill cattle are excluded from herd-based testing because their results may be indicative of an individual disease problem rather than a herd problem.

B. Individuals to be tested must belong to defined groups of similar age and lactation stage under similar environmental conditions and nutritional management (e.g., prefreshening cows and postcalving cows). Specific group trends or abnormalities may not be apparent if inclusion criteria are too broad, thus yielding more physiologic variation.

C. Useful information requires adequate group representation and therefore an adequate sample size for the group of interest. Sampling multiple animals from a defined group helps reduce variation unrelated to nutritional state.

1. Sample sizes of 7–12 have been advocated as reasonable pools to provide useful information for most purposes, regardless of herd size. However, greater numbers of test individuals provide results with tighter confidence limits and may be necessary to adequately represent target groups in large herds.
2. Herd size may limit the number of animals in a relevant group that can be tested at one time (e.g., too few prefresh cows). However, useful information may be obtained in these groups by testing over time as animals enter the group. Interpretation must then include considerations of temporal variables.
3. Economic pressures encourage smaller test groups and pooling of samples, but pooling may be undesirable.
   a. In one study, 5 of 21 analytes (including [tCa^{2+}] and [BHB]) had some pooled mean values that differed from mean values calculated from results for the individual animals.[66]
   b. Unexpected and marked abnormalities in one or more samples may bias the pooled results, especially if unequal volumes of sample from each individual are used.
   c. Individual animal data cannot be generated, so only mean values rather than ranges of values or proportions of abnormal results above or below a decision threshold can be assessed (see Section III).

D. Analytes selected for herd-based metabolic profiling tests should reflect metabolic/nutritional status. Typically, these analytes are not tightly regulated by hormonal or other physiologic factors. For example, metabolic pathways attempt to maintain [glucose] within a certain range even if there is a marked difference in the intake of carbohydrates; thus, serum [glucose] is not a good analyte for assessing carbohydrate intake or digestibility. In contrast, plasma [BHB] is not tightly regulated, and increased concentrations reflect a deficient energy status. Even if analytes meet this criterion, it is also important that analytical variation is small compared to changes caused by altered nutritional states. Examples of tests and results that may be useful include the following:
   1. Low blood hematocrit values in the first 8 wk of lactation suggest the need to evaluate energy and protein nutrition.
   2. Serum [urea] at any lactation stage is an assessment of ruminal balance of energy and available protein.
   3. High serum [BHB] at about 5–50 d into lactation is associated with suboptimal energy and glucose balance and, clinically, reduced milk production, increased clinical ketosis, displaced abomasums, and reduced fertility. Samples should be collected at a consistent time relative to feeding, e.g., 4–5 h after the first meal of the day, when rumen butyric acid production is high.
   4. High serum or plasma nonesterified fatty acid (NEFA) concentrations in prefresh dry cows in the last month of gestation (especially 2–14 d prior to calving) are evidence of negative energy balance. Samples should be collected shortly before feeding time. One can draw and freeze multiple samples and send those that fulfill the 2- to 14-d precalving criterion once this is known.
   5. Low serum [tCa^{2+}] 12–24 h after calving by multiparous cows indicates clinical or subclinical parturient hypocalcemia.

III. Interpretation

A. Abnormalities in herd-based test results must be interpreted with attention to all the variables that can affect them, not just nutritional factors. Other factors include preanalytical sample variation, analytical variation, biological variation, circadian/prandial

variation, seasonal variation, variation in physiologic states, occurrence of pathological states, and other environmental and management factors. Most of these variations are controlled by careful group selection criteria and appropriate procedures for sample collection, processing, and analysis. Preanalytical factors are detailed in specific chapters where analytes are discussed.

B. Most herd-based test results fall within diagnostic RIs because the sampled individuals lack clinical signs of disease. Therefore, a different set of interpretive guidelines is necessary to detect subclinical abnormalities. Interpretive guidelines are suggested recommendations by experts in the field based on experience and research, but it is not clear how widely applicable the guidelines are given interlaboratory variability in assay results.

C. Data have been evaluated in several ways.
   1. Mean herd values may be compared to mean values of reference herds. However, these data are not readily available because of the expense of testing and difficulties identifying suitable reference herds.
   2. Mean herd values may be compared to an established expected mean result for the analyte and group of interest. For example, it has been suggested that mean urine pH should be 6.0–7.0 in prefresh cows fed anions to help prevent milk fever. A mean pH of 6.0–7.0 supports appropriate acidification of the group. When the herd's mean ± an uncertainty interval does not include 6.0–7.0, a herd's urine pH is considered inappropriate. When the herd's mean value is not 6.0–7.0, but the mean ± uncertainty interval overlaps this target, the results are considered borderline. A 75 % confidence interval has been suggested as a useful guide for metabolic profiling and a reasonable compromise between 95 % confidence and practicality.
   3. The percentage of individual results above or below an established decision threshold can be calculated and compared to what is considered an acceptable percentage of high or low values. A percentage (± uncertainty interval) greater than the accepted percentage signals a problem. A percentage and confidence interval overlapping with the accepted percentage is a borderline result. Decision thresholds may be somewhat artificial in that the risk for some conditions increases continuously with the magnitude of the analyte (e.g., concentration of nonesterified fatty acids and displaced abomasum).
   4. Others have suggested using process control charts to graphically monitor changes of an analyte in a particular group over time. A running graph of mean values and high-low ranges of individual values can be plotted over time to evaluate for group changes that signal problems. Fluctuations must be assessed relative to established limits to determine when action should be taken.

D. Just as with population-based RIs, herd-based profiling has limitations, but useful information can be acquired for making management decisions.

References

A complete set of references is available online. Please use the below link to access the References files.

https://www.wiley.com/go/fundamentals/vet