Fig. 1.1. Significant figures. The significant figures of the measured and calculated values of a leukogram are provided. However, computer-based reporting systems may not enable the flexibility to report only significant figures. diff., differential; and WBC, white blood cell (leukocyte).

Fig. 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 will represent the central 95 % of the reference values and thus the reference interval. If values have a Gaussian distribution, the mean, median, and mode values will be 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 lower reference limit would be below the lowest reference value, and the upper reference limit 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 non-Gaussian distribution.

Fig. 1.3. Illustrations of analytical properties of assays.

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

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

c. *Analytical precision and accuracy*: Because all ten holes in the target are tightly clustered in the middle of the target, the shooter was accurate and precise.

d. *Analytical specificity*: Of the ten holes in the target, seven are round, two are square, and one is oval. Of the ten observations, the shooter probably created seven, and three were created by other factors. Thus, the presence of holes in the target is not specific for the shooter's actions.

e. *Detection limit and 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 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 bullet whose diameter was 75 % of the previous shot.

Fig. 1.4. 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 control solution is plotted on the chart. Previous analyses established that the acceptable limits for the control solution are 100 ± 10 mg/dL (mean ± 2 sd). On days 5 and 9, the first measured concentration (1) was outside acceptable limits. After corrective actions were taken (e.g., recalibrate, new reagent set, and clean pipette), a second measured concentration (2) was within acceptable limits so values measured in patient samples could then be considered reliable. For some assays, at least two control solutions are analyzed with each set of patient samples, and results of each control solutions are plotted in separate Levey-Jennings control charts. For other assays, control solutions might be analyzed with each shift (e.g., day and night) or each day. The frequency of analyzing control solutions is determined by several factors but with the primary goal of detecting unacceptable assay performance before a patient's result is used for diagnostic or therapeutic decisions.

Fig. 1.5. Results from similar and different assay methods that analyzed aliquots of one canine serum sample as part of one survey completed by the Veterinary Laboratory Association (VLA) during 2000. Ranges for most measured analyte concentrations or enzyme activities were extracted from graphs within the survey's report, but data for ALP, AMS, and LPS activities were obtained from the VLA because graphed bars were too small to be reliably interpreted.

- For nearly every measured analyte, 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.
- The greatest differences occur when different assay methods are used, but frequently there were large differences when the same assay method was used on different machines (e.g., ALT, ALP, AMS, cholesterol, CK, glucose, sodium, and tCa^{2+}).
- For the data used in this figure, the "results of the same method on same instrument" represent data for the same dry chemistry reagent system on the same type of instrument. Even so, the lack of agreement emphasizes the need for the establishment of reference intervals by each laboratory.

ALT, alanine transaminase; ALP, alkaline phosphatase; AMS, amylase; Cl^-, chloride; CK, creatine kinase; tCa^{2+}, total calcium; and tCO_2, total carbon dioxide.

Fig. 1.6. NCCLS and Altman-Bland bias plots. The results were tabulated and entered in a software program (Analyse-it) to generate the bias plots.

- In the NCCLS 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 average difference for all samples is calculated; in this comparison, the average difference is 2 mmol/L (bias, − 2 mmol/L). Lines representing ± 2 sd of the differences are also displayed. The NCCLS bias plot is the preferred plot when a second method is being compared to a reference method or an established method.
- The Altman-Bland bias plot is the same as the NCCLS 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 advantage of the Altman-Bland bias plot is that neither method is considered to be the more accurate method.

HCO_3^-, bicarbonate.

Fig. 1.7. 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; that is, the absolute difference between 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 and 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.

HCO_3^-, bicarbonate.

Fig. 1.8. Kappa-agreement data for urine heme reaction. After dipping the reaction pad in 100 urine samples, the color changes in a heme-reaction pads were assessed by two methods (reflectance photometry and visual examination) and graded as negative, trace, 1+, 2+, or 3+. The results were tabulated and entered in a software program (Analyse-it) to calculate a weighted kappa value.

Fig. 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 would be 100 % because there are no FN results.

- The diagnostic specificity would be 50 % because there are equal numbers of TN and FP results.
- The diagnostic accuracy would be poor because of the many FP results.
- The PV(+) would be poor because of the many FP results.
- The PV(–) would be 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 would be poor (about 60 %) because there are relatively many FN results.
- The diagnostic specificity would be 100 % because there are no FP results.
- The diagnostic accuracy would be poor because of the many FN results.
- The PV(+) would be 100 % because there are no FP results.
- The PV(–) would be poor because there are many FN results.

C. The decision threshold is at a concentration where the least overlap between the groups occurs.
- 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 PV(+) and PV(–) would be high but not 100 %.

Fig. 1.10. Examples of diagnostic properties of assays. 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. As the diagnostic sensitivity is 90 %, then 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 %, then 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 the Eq. 1.1 formulas.
- *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, ten dogs have the disease and 990 dogs do not. Because the diagnostic sensitivity is 90 %, 90 % (nine) of the ten diseased dogs will have a positive test result (TP) and 10 % (one) 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 the Eq. 1.1 formulas.

Fig. 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. Also, conclusions from the analyses are presented in the text.
- $[tT_4]$ 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.

Fig. 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). To gather data for the ROC curve, multiple decision thresholds 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 values are calculated for each decision threshold. (For this illustration, the number of animals in both groups were 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 (bottom graphs). 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. In this comparison, assay A is a better diagnostic procedure than assay B for detecting a certain disease.

Dx., diagnostic.

Fig. 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 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 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?