Clinical Chemistry Data, Expectations and Realities

Holly Bender, *University of Blacksburg*
Carolyn S. McCoy, *University of Blacksburg*
Clinical Chemistry Data, Expectations and Realities

Holly S. Bender DVM, PhD
Carolyn S. McCoy MT (ASCP)

"A physician who depends on the laboratory to make his diagnosis is probably inexperienced; one who says that he does not need a laboratory is uninformed. In either instance the patient is in danger." - Wells and Halsted 1967

Have you ever noticed that people tend to regard information in print as ever so much more factual than the spoken word? The presence of numbers seems to add the perception of even more legitimacy and precision to the document. This quirk of human nature places an enormous (and often unrealistic) responsibility on clinical laboratories, as our reports are not only in print, they are composed of numerical data. The television commercials whine on about knowing your exact cholesterol number, and news reports strike fear into the public when the results of a test performed at two separate laboratories differ by a couple of mg/dl. It is important that you as diagnosticians, understand how these numbers are obtained, their benefits, as well as their limitations. This will help you to approach data interpretation rationally, using it to your advantage to practice sound veterinary medicine. At the same time, this perspective will help you to avoid becoming disappointed and disillusioned with the entire subject of clinical pathology in the event that the numbers do not reach up to unrealistic expectations.

Reference Intervals

Our objective as a clinical laboratory is to use such sophisticated techniques and instrumentation, that we can arrive at laboratory data which reliably approximate the "real values". With the current state of technology, we are getting fairly close to this goal. We also strive to obtain reference ranges that are gold standards with which to compare patient results, in order to differentiate diseased from healthy animals. Even in the finest and best funded laboratories, we can only partially reach this goal. This is largely because of the limitations inherent in the tests themselves as well as the natural variation of blood constituents in both normal and abnormal individuals.

After repeatedly looking up reference values for the various tests, you may be wondering how these ranges (intervals) are established. In an ideal situation (at least in a statistician's dreams), these intervals would be derived from the assayed blood values of all healthy animals in the world. Of course to be useful, these ranges would also exclude the values of all the sick animals of the world. If this were true, medicine would be quite straightforward. We could simply run hematology and chemistry profiles on all sick animals, enter the data in a computer system, and out would pop the name of the disease.

Unfortunately, our job is not that simple (or boring for that matter). First of all, there are not enough vacutainer tubes available to bleed all of these creatures, veterinarians to perform the physical examinations, nor funds to perform all of the tests (especially in the current fiscal climate). So we take a more practical approach, employ a little statistical creativity, and essentially do the best we can with the resources available.

The currently accepted method of establishing a reference range is to test a population of clinically healthy animals, somewhat smaller than all the individuals in the world. "Healthy" is defined by a lack of disease found on physical examination, clinical history and blood work for known subclinical diseases. At least 30 animals are required, but 100 are preferred. Ideally, to be a truly representative sample, this group should include only the animals that are of the specific breed, age, geographic area, housing, gender, stage of the reproductive cycle, and on the same diet as your current patient. It would be prohibitively expensive to determine specific ranges for all of these variables. Therefore, reference ranges are usually only specific for each species and are consequently broader than would be expected for a more exclusive population. It is important that the reference population used to create these numbers is composed of animals representing all of the ages, geographic areas, etc. that you may expect to see as patients. Even with all these parameters in mind, you would be wise to consider the emotional state of the animal at the time of sampling as it may markedly influence certain tests.

For many analysts, a graph of the frequency of individual values forms a bell-shaped curve, much like the familiar point spread obtained after a veterinary school exam. The bell-shape illustrates that most of the values are bunched up in the middle (class average), with fewer and fewer results occurring as the distance increases on either side of the mean. This bell-shaped pattern is called a normal or Gaussian distribution, and may be analyzed by a so-called parametric approach. Once the data is plotted and it is established that it indeed forms a normal distribution, calculations are performed to determine the standard deviation.

\[
\text{Standard Deviation} = \sqrt{\frac{\sum(X-\text{Mean})^2}{N-1}}
\]

\[X=\text{individual test score}\]

\[N=\text{number of data points (# of students)}\]

PROC. 9th ACVIM FORUM

NEW ORLEANS, LA, MAY 1991
The standard deviation is a mathematical expression of the amount of variability in data, that has been striking fear and dread in the hearts of veterinary clinicians for many years. The concept is really pretty simple if you understand why we use it. Again, I will use the example of the results of a veterinary school examination. Assume for simplicity that there were 100 students in your class and everyone shows up for the exam. Once the results are tabulated, the class average or mean is determined by adding up all scores and dividing by the total number of students (in this case, 100). Then the score for each individual test is subtracted from the mean and the result is squared. The resulting numbers are added up and divided by one less than the number of test results (99). The standard deviation is determined by calculating the square root of this number.

When the number of students receiving particular grades are plotted, the graph typically appears in a bell-shape, or normal distribution.

Notice that 68% of the students have scores within ±1 standard deviation from the mean of the curve, 95% of the test scores are within ±2 standard deviations, and 99.7% of the scores are within ±3 standard deviations from the mean. This is the reason we calculate the standard deviation of the data when establishing reference ranges. By convention, we accept data that is ±2 standard deviations from the mean as our reference range. This includes 95% of our clinically normal subjects.

Some laboratory tests do not yield a normal or Gaussian distribution when plotted as above. The plot of these tests generally forms a bell that is skewed to the right or left. Here, the standard deviation cannot be used to determine the reference range because the data is not evenly distributed around the mean. The nonparametric approach is used in these situations. With this method, results of the tests are placed in increasing order of magnitude. Then the 2.5 and 97.5 percentiles are calculated, which in turn yields the middle 95% of the data. This middle 95% becomes the reference interval.*

You may notice that whether the analysis is parametric or nonparametric, 2.5% of the so-called healthy reference population is by very definition, designated as abnormally high, and another 2.5% designated as abnormally low. In fact, on a 12 test chemistry panel, there is only a 54% chance that all of the results will be within the limits of our reference range! The occurrence of a result that is within the third standard deviation from the mean in a healthy individual is called a false positive. This test could categorize an otherwise healthy animal as sick. The occurrence of false positives is simply due to the way we figure reference intervals. Of course, nearly 100% of the values from our healthy reference animals fall within ±3 standard deviations from the mean. Then why don’t we figure the reference range as ±3 SD from the mean? Unfortunately, lab values from diseased individuals also tend to occur in a Gaussian distribution, and there is often considerable overlap in the ranges of healthy and unhealthy animals. That third standard deviation from the mean generally makes the reference range too wide to be clinically useful, because it results in too many tests results that are considered normal in sick animals (false negatives). Simply put, we sacrifice having a few false positives for eliminating many false negatives. This is where your skill as a clinician comes in. If you examine an animal and he has no history, signs nor earthly reasons to have an anemia, but has a low PCV, take a closer look at the data. If the PCV value is within the third standard deviation from the mean, perhaps he is a statistical outlier.

* Reference intervals used to be called normal values, but this terminology has become increasingly unpopular over the past 20 years because of the multiple meanings of normal. Normal could mean Gaussian, average or healthy. The term reference interval is largely replacing normal range because it is thought to be more precise, implying a representative set of values obtained from a comparable group of individuals.

PROC. 9th ACVIM FORUM 74  NEW ORLEANS, LA, MAY 1991
Quality Assurance

It is very important that the methods, instrumentation and conditions used to perform the tests for the reference population and the clinical population are the same. It is therefore advisable to use the reference values established by each individual laboratory. This is most important for enzymatic assays, where individual methodologies may be conducted under various conditions. Incubating enzymatic reactions at two different temperatures or using different substrates may yield widely disparate results. This is because enzymes are measured in the amount of activity, not by concentration. Obviously, enzyme activities are very temperature and substrate dependent. The values obtained from some colorimetric assays may also very greatly according to the method. Hematologic values are less subject to variation by method, as a cell is a cell no matter the temperature or suspending substrate. Electrolyte values are also generally reproducible between methods. However, It is essential that the units used for expressing concentration are identical when comparing any laboratory result.

Choosing a particular laboratory to analyze your samples is a critical decision. Find out if your lab has established a good quality assurance program and is accustomed to working with veterinary samples. Quality assurance programs have become a major priority to most clinical laboratories in recent years, resulting in increasingly more accurate and reproducible laboratory results.

Quality assurance is definitely not restricted to the clinical lab. The most accurate analysis in the world is useless unless you can be confident that your sample is handled according to strict guidelines before it reaches the laboratory. This may not seem like a major point, but in reality sample handling is often the single biggest source of error in sample processing and analysis. Generally, laboratory personnel are quite accustomed to the influences of sample transport and storage. To them, taking the proper precautions becomes second nature. Common sources of error in collection and handling of serum samples are hemolysis from poor venipuncture technique, letting the serum sit on the clot, exposing the sample to air and light, and storing the sample at room temperature. Common sources of error in hematologic samples are delayed or inadequate distribution of the anticoagulant, transporting cells for morphologic examination in the collection tube and prolonged transportation times for platelet counts.

Most of the large clinical pathology laboratories have strict sample requirements and well established quality assurance programs in order to ensure accuracy in reporting results. As a brief overview, all quality assurance programs include regular instrument calibration and daily use of control fluids. Clinical laboratory instruments are calibrated in order to make test results closely match the so called "real values." Of course, real values are theoretical and are never completely certain. In reality, calibration causes the instrument's results to closely approximate values obtained from a method accepted as the analytical gold standard. Why don't we just skip all of this nonsense and run all clinical tests by the gold standard method? Often, this method is extremely accurate, but is impractical for routine testing in a clinical setting due to cost per test, or the time it takes to produce results. Some clinical instruments typically remain calibrated for about 6 months, while others require daily calibration. The retention of proper calibration is checked daily on each analyte by running commercially available control fluids. These controls are assayed repeatedly by the manufacturer using a gold standard method, yielding a mean and standard deviation. When daily controls are analyzed and reviewed in the clinical lab, results are generally accepted as "in control" if they fall within ±2 standard deviations from this assayed mean. Once all analytes are determined to be "in control", patient samples are analyzed. Records of the daily control results are kept to anticipate problems with the various analytes. These results are charted by a computer program and inspected visually to detect shifts and trends.

The majority of large human-oriented labs participate in formal reviews and accreditation programs by external evaluators. These laboratories also participate in interlaboratory comparison surveys, whereby samples with unknown quantities of various analytes are prepared, divided and sent to all of the participating labs in the country. The samples are analyzed and individual results are tabulated and compared with all of the other participating labs. These programs are in their infancy in exclusively veterinary-oriented labs, but it will not be long until these are routine procedures in most veterinary labs.

References


