Radioimmunoassay Kit Evaluation and Selection

Mohini R. Pimputkar and Carolyn S. Feldkamp

Henry Ford Hospital, Detroit, Michigan

This is the second continuing education article in a four-part series on radioimmunoassay. After reading and studying this article, the reader should be able to: 1) describe the procedures used for kit evaluation, including the utility of each experiment, 2) describe the requirements for each experiment, and 3) discuss acceptable results and pitfalls.

The goal of an RIA kit evaluation is to obtain objective data in order to support manufacturer claims of precision, accuracy, sensitivity, and specificity. These data support a laboratory's selection of a kit by clarifying which performance features justify choosing one method over another. In addition, these data provide a yardstick for future kit performance—essential for long-term quality control and trouble-shooting.

The performance tests described in the following protocol attempt to establish or verify these desirable characteristics. In the routine clinical laboratory setting, it is understood that the laboratory personnel are often limited by the amount of reagents available for this purpose, by predilution or premixing, and often by the physical format of a given kit.

GETTING STARTED (DRY LAB)

Specific objectives established for a kit evaluation direct which tests are done and how extensive the evaluation is to be. Preliminary considerations include: (a) Clinical need for the test. Will the test be done frequently enough to provide the service required in terms of timing, turnaround, and cost? (b) Facilities and resources available. Is new equipment needed? (c) If a kit currently in use is unsatisfactory, then specific performance criteria (e.g., better precision) may have been defined based on the problem-solving experiments.

The College of American Pathologists' (CAP) survey or others provide information on the popularity of various kits (number of users), and precision between laboratories. However, consensus alone does not prove the accuracy or reliability.

Obtain product information from 2-4 kits selected from preliminary considerations and review of the protocols. Compare claims of precision between kits, assay set-up times, cost per tube, and format of reagents. Sensitivity, standard range, and curve shape characteristics are compared by plotting data from different kits on the same scale. This simple procedure may uncover the strength or weaknesses of one kit or another that may then be investigated further.

For reprints contact: Carolyn S. Feldkamp, Department of Pathology, Henry Ford Hospital, 2799 West Grand Boulevard, Detroit, MI 48202.

PROTOCOL OUTLINE: LABORATORY EVALUATION

First Experiment
1. Sources of variation
2. Standard curve
3. Precision
4. Sensitivity (least detectable dose, LDD)

Second Experiment: Accuracy
1. Review of standards
2. Recovery, test of matrix effects
3. Parallelism (dilution of samples)
4. Cross-reaction
5. Clinical validation and normal range

Third Experiment
1. Specific activity (self displacement)
2. Tracer immunoreactivity
3. Scatchard plot
4. Data reduction

First Experiment

Read the product brochure carefully and use the recommended protocol. Careful control of details such as which tubes to use, incubation times, centrifugation, and counting time will optimize assay performance. The plan for the first assay allows assessment of: (a) sources of variation; (b) standard curve; (c) precision; and (d) sensitivity.

Sources of Variation. Systematic error, or bias, is often inherent in assay components or in the experimental design. In addition, a variety of experimental and instrumental factors influence assay reliability by contributing to both systematic and random error. Among these are pipette and thermometer calibration, careful timing, equilibrium conditions, and optimal choice of separation techniques.

Scintillation counters should be carefully calibrated and enough counts for each sample should be accumulated to ensure that counting errors do not contribute significantly to the total error (counting error = $\sqrt{\text{total counts}}$). The common recommendation is to count 10,000 counts. Care should be taken to select and calibrate pipettes to rule out carryover. Allocate enough quality control materials to complete the entire evaluation. Vial-to-vial variation is thus avoided. Plan to test more than one kit lot. The best sensitivity and maximum

precision should be obtained when the radioactive tracer is fresh. Save some tracer until near the expiration date to repeat selected tests.

Control Tubes and Standards. Set up total count tubes (tracer only). The assumption is implicit that all of the label is immunoreactive, that is, the tracer has not been damaged and could bind in the presence of an excess antibody.

Set up nonspecific binding tubes (NSB, buffer control) by adding tracer, a protein-containing buffer to replace the antiserum, and the sample. NSB as a percent of total counts should be minimal. Typical NSB tubes correct for tracer bound to the test tube, trapped in the precipitate, but do not account for specific or nonspecific binding to components (proteins) introduced from the patient's serum. Patient blanks (sample, tracer, buffer, and separation medium) run during the evaluation appropriately account for the true nonspecific binding present in the patient samples. Current kits rarely require patient blanks, but occasionally a variable blank will be observed.

Maximum binding tubes (B_0, zero standard) include tracer, antibody, and analyte free sample. B_0 may be followed as the tracer ages as evidence of tracer stability and immunoreactivity.

If there are large gaps between any two kit standards, ascertain the true shape of the curve by filling these gaps with intermediate standards. Occasionally, this procedure reveals an inflection point between two standards, which, combined with point-to-point or spline curve fitting methods, might result in significant inaccuracy in patient results. Once a kit and data reduction method are selected for routine use, the number of standards can be reduced to those recommended by the manufacturer.

Precision. Precision indices such as coefficient of variation (CV) between run and within run, reflect the ability of an assay to reproduce a result on the replicates of the same sample, or the confidence interval associated with a single result. They are assumed to reflect random errors.

Select samples for precision studies in at least three concentrations: low, midrange (50% B_0), and high doses on the standard curve. For some tests, additional concentrations may be selected to evaluate parts of the curve range that are of particular clinical interest such as an abnormal-normal cutoff or a portion of range for which commercial controls are available.

Assay each of the selected test samples in two or more replicates in every run. Calculate the mean and standard deviation for each sample. Practically speaking, $n = 10$ is often used to estimate the within-run precision for the initial evaluation. The calculated standard deviation will be a better estimate of the true value if at least 20 observations are included between run and within run. An estimate of standard deviation can also be made by grouping the variance of duplicates of individual samples. With kits, a CV—between run and within run—of $< 10\%$ in most ranges is obtainable and acceptable for clinical applications.

Precision is affected by all components of the immunoassay such as the equilibrium constant, the rate of reaction, completeness of separation, and interference. Kit users assume that the assays are robust to normal variations in the environment such as differences in room temperature. Because reaction times are often shortened to improve turnaround time, true equilibrium cannot be assumed and reaction times must be kept within specified limits.

Sensitivity. Sensitivity is commonly understood to mean the smallest amount of ligand detectable or, in some contexts, the greatest change in response for a given increase in dose. The concept encompasses two related assay characteristics: precision and slope of the standard curve (Fig. 1). A value of ligand that can be reliably distinguished from zero (95% confidence) is defined as the least detectable dose (LDD). LDD for RIA is the dose calculated as the mean -2 SD (expressed in cpm or other response variable) of replicates of the zero standard. Clearly, both the precision and the slope of the curve contribute to a small LDD or sensitivity. Because some error is associated with every point of the standard curve, a similar calculation ($\pm$ 2 SD) at points other than zero define resolution.

Least Detectable Dose for RIA Kits. Set up 10 maximum binding tubes (B_0, zero standard) distributed throughout the assay, as well as one or two dilutions of the lowest standard. These "low standards" verify the shape of the curve (linearity) and precision near zero. As with the specific activity measurement described later, a bound counts/free counts (B/F) plot with its steep slope in the low end is recommended for this experiment.

Least Detectable Dose for IRMA Kits. Set up NSB tubes distributed throughout the assay and one or more dilutions of the lowest standard. Since in IRMA methods the counts bound are proportional to concentration, LDD is the mass

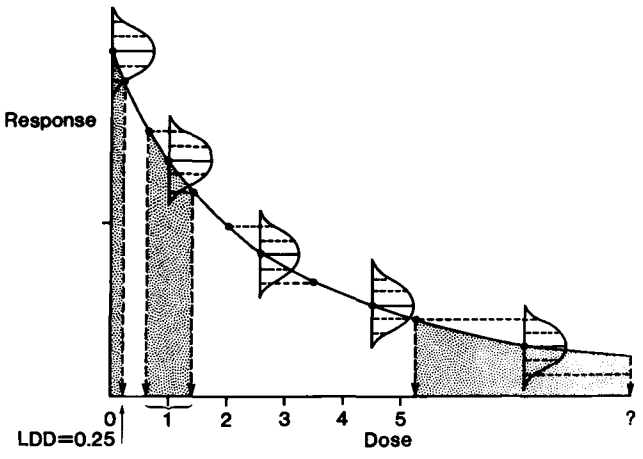

FIG. 1. Effect of precision on sensitivity. LDD (95%) is defined as the dose represented by the mean response -2 SD of replicates of B_0 ($n=10$). Resolution (95% confidence) is the smallest detectable difference between two dose values ($\pm$ 2 SD).

calculated from the mean + 2 SD (in cpm) of replicates of the NSB (zero standard).

It is a very practical problem to decide whether to report results between the lowest standard and the LDD. Common sense and a careful evaluation of the shape of the curve as well as the LDD all contribute to this assessment.

Second Experiment

The second assay of the evaluation addresses accuracy. Tests for accuracy include: (a) review of standards; (b) recovery, test of matrix effects; (c) parallelism (dilution of samples); (d) test for cross-reactions; and (e) clinical validation and normal range.

Standards: The Ultimate Accuracy. When a dose estimate is calculated from a standard curve, the analytical assumption is made that the compound being measured is well characterized, chemically pure, and has been measured by an independent, valid measurement (i.e., weighing). For an accurate measurement, we must also assume that the assay method is specific, precise, and free from bias or interference introduced by the matrix in which the standard is placed.

Curve Shapes. Standard curves plotted using different variables, B/F, bound counts/total counts (B/T), or bound/zero standard (B/B_0) often reveal high- or low-dose regions in which dose response is too flat to use. Also, the true shape of a dose response curve may be obscured by the choice of the data reduction model. An appreciation of the impact of the curve shape on other derived data can help the kit user understand such observations as poor parallelism or recoveries. "Low-dose hook," occasionally observed in RIA kits, has been attributed to positive cooperativity (Fig. 2A). This phenomenon may be seen as a poor fit in the low end of the curve or $B > B_0$. The analogous "high-dose hook" in the IRMA assays may cause considerable trouble by attributing a low or normal value to a high sample (Fig. 2B).

Matrix. The matrix of the standard curve is defined as everything in the standard (or sample) that is not the analyte (proteins, lipids, ionic strength, pH). Matrix may affect the separation steps as well as the binding reaction. For example, polyethylene glycol separations are often found to be relatively sensitive to variation in protein composition.

In a kit, the matrix should be identical in all standards, including the zero standard. The matrix does not actually have to be human serum, but the manufacturer should have tested the antibody to demonstrate that no shift in the standard curve occurs between standards in human serum and standards in the matrix chosen. Thyroid-stimulating hormone (TSH) is an example of an assay that frequently shows significant differences depending on the matrix used.

Recovery. The most popular test for accuracy is the recovery study. Briefly, a known amount of the analyte is added to a base medium (sample-like matrix that contains little of the analyte) and is then tested in the assay being evaluated. The difference between concentration observed (base + standard added) and the base assayed alone is "recovered."

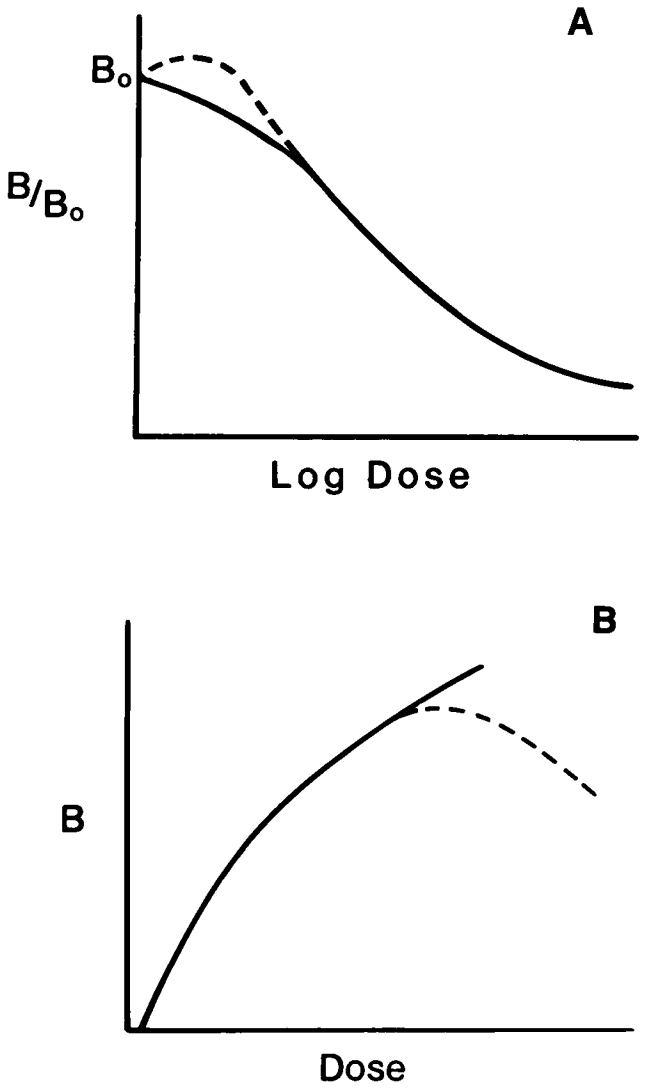

FIG. 2. "Hook" effects in RIA and IRMA. Anomolous "hooks" (---) observed at extremely low or high concentrations may result in falsely low values. The expected dose-response curve is denoted by (——).

Ideally, recovery is tested in different bases since the test is a measure of the influence of matrix on the assay. Popular base media include the zero standard, a low concentration patient sample, or a low pool of samples. At least three concentrations of added standard are selected to fall within the most precise and linear range of the standard curve.

The success of this test depends on the availability of a known amount of pure ligand, the precision of the assay, and the choice of the base medium. The base medium, which is also measured in the assay being evaluated, is implicitly assumed to be accurate. This contradiction creates certain problems in the design and interpretation of recovery studies.

The apparent recovery of an unknown analyte depends on the method of calculation. Method II (Fig. 3) always gives higher values, but cannot reveal a proportional error. Method I (recommended) may give lower values, but can reveal a systematic bias that should be attributed to the base medium as well.

RECOVERY CALCULATION

Base + Recovery std. $\stackrel{?}{=}$ Analyte Observed

B A C

Method I

$$\%R = \frac{\text{Observed-Base}}{\text{Std. added}} = \frac{C-B}{A}$$

Method II

$$\%R = \frac{\text{Observed}}{\text{Expected}} = \frac{C}{A+B}$$

FIG. 3. Recovery calculations. Both methods give equivalent values if assay is accurate. If there is a matrix effect on recovery, Method I will give lower values. To minimize errors in the estimate of R, B is usally selected to have very low concentration of the analyte.

The precision and fit of the standard curve as well as the precision of the recovery study contribute to the outcome (and satisfaction) of this test. Poor recoveries may reveal an inaccurate test, systematic bias, or a matrix effect. A 100% recovery does not in itself prove accuracy.

Parallelism. A necessary, but not sufficient, criterion for accuracy is parallelism of response between dilutions of sample and standards. Lack of parallelism may reflect antigen heterogeneity, antibody heterogeneity, or cross-reactions. In addition, nonparallel behavior may be an artifact of an inappropriate choice of diluent (i.e., matrix effects, inappropriate blanking, dilution error, or measuring on an imprecise part of the standard curve). Currently, many kit standards are prepared in analyte-free human serum and kit protocols specify that the zero standard should be used to dilute high samples.

Demonstrate parallelism by diluting a high patient sample with an appropriate diluent and run the assay. Test several different patient samples and at least three different dilutions selected to fall in a precise portion of the standard curve. Dilutions of the high standard should also be parallel to the standard curve if the diluent is compatible.

One of the simplest ways to demonstrate linearity and parallelism is to calculate the final concentration for each sample (multiplying by the appropriate dilution and plotting the result versus dilution or sample size). The graph should be a horizontal line within the error of measurement. Observed value plotted versus expected (based on the undiluted value) should be a 45° line. In this graph, different samples can be displayed together and should be parallel to each other. If the intercept is not zero, the assay may be parallel but biased. Although it is possible to apply statistical tests of the differences from ideal behavior, in our experience visual inspection of such a graph has usually been adequate to have a good sense of the performance of a kit.

Cross-reactivity. An important aspect of accuracy is the specificity of the assay. Although cross-reactivity data are published in kit brochures, it may be desirable to selectively test some compounds in the laboratory to verify accuracy with current lots of reagent.

Cross-reactivity is usually expressed as the relative dose required for 50% displacement of the maximum tracer binding (Fig. 4). The experiment is done by running an entire standard curve of the cross-reacting substance and at the same time the usual standard curve. Cross-reactivity is conventionally defined as (mass of standard at 50% B_0/mass of competitor at 50% B_0) × 100. Even when the apparent cross-reactivity is acceptably low, it is important to consider how high the concentration of interfering substances in samples being measured might be.

Clinical Validation and Normal Range. Tests for accuracy do include normal range study. Expense and the difficulty in obtaining specimens are important reasons for performing normal study at the end of an evaluation.

Third Experiment

A fundamental concept underlying competitive displacement assays is that both unlabeled and labeled species have strong affinity for the same binding site, but can be conveniently detected in the presence of the other. However, the tracer reagent is not ideal; it is rarely homogeneous and has some finite mass. Thus, sensitivity and precision are not only influ-

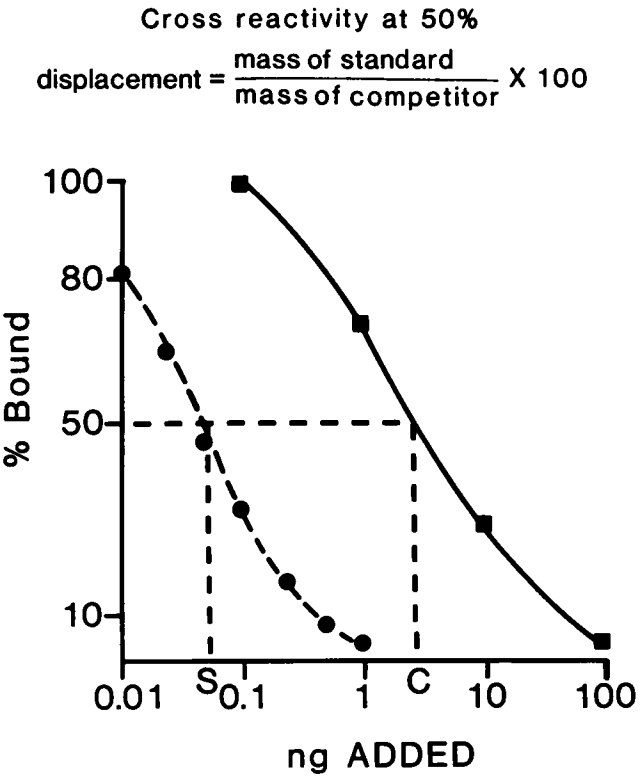

FIG. 4. Cross-reactivity. Displacement of tracer by a competing (cross-reacting) antigen (C) is compared with the standard analyte (S).

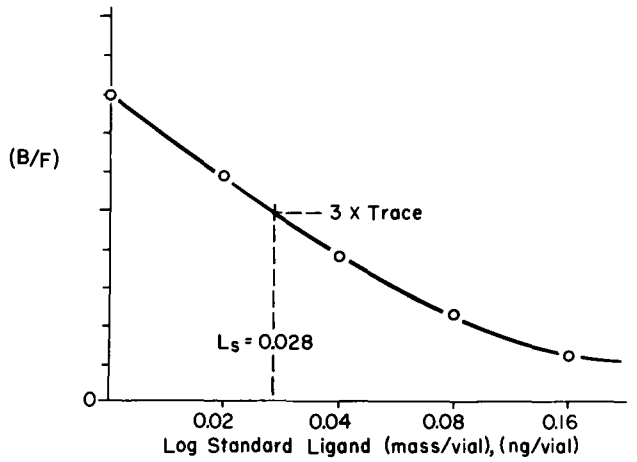

FIG. 5. Tracer self-displacement. The mass of an aliquot of tracer is read directly from the standard curve. In the example shown, B/F is calculated individually since increased tracer mass increases the total counts. Since the normal curve contains 1× tracer, the sample shown (3×) contains two extra aliquots. Thus, mass/aliquot is Ls/2.

enced by statistics and antibody affinity, but also by the amount of immunoreactive mass in the tracer.

The third experiment includes tests for: (a) specific activity from self-displacement data; (b) tracer immunoreactivity; (c) Scatchard plot; and (d) data reduction.

Specific Activity. Specific activity (activity/mass) of kit tracers can be estimated by a technique called self-displacement. This is accomplished by treating the tracer as a sample (additional tracer mass). Several different amounts of additional tracer are assayed. The exact experimental protocol varies with the format in which kit tracer is provided by the manufacturer.

Mass of the samples (concentrated tracer) is read from standard curve (B/F vs log mass/vial). Each sample response is individually corrected for NSB. The response of the tracer is compared to the slope of unlabeled standard (i.e., parallelism is observed) (Fig. 5).

TABLE 1. Specific Activity Experiment*

Tube Number	Concentrated Tracer (ml)	Buffer (ml)	Antibody (ml)	Regular Tracer (ml)
1, 2	0.010	0.090	0.1	0.1
3 (NSB)	0.010	0.090	—	0.1
4, 5	0.025	0.075	0.1	0.1
6 (NSB)	0.025	0.075	—	0.1
7, 8	0.050	0.050	0.1	0.1
9 (NSB)	0.050	0.050	—	0.1
10, 11	0.100	0.0	0.1	0.1
12 (NSB)	0.100	0.0	—	0.1

* Protocol for an experiment is as follows: (a) run several levels of concentrated tracer in triplicate; (b) precount the tubes; and (c) incubate, separate, centrifuge, decant, and count each tube.

For this laboratory exercise, maximum slope in the low end of the standard curve is desirable, so the B/F vs dose plot is recommended. If the tracer has a high specific activity, the lowest standard may be diluted to define the curve for the range of this experiment. This test will yield the specific activity of immunoreactive tracer because it is measured by its reaction with the specific antibody. Aside from being able to compare this parameter in different kits of the tracer, the measured specific activity is needed to calculate the total mass of tracer for the Scatchard plot.

The kit tracer may be lyophilized, or the manufacturer may be asked to supply some concentrated tracer. If the tracer is prediluted, the measurement of the specific activity may be impossible with some kit formats. Occasionally a conjugate-labeled hapten will have such a high specific activity that no displacement in the range of standard curve is observed.

The experiment shown in Table 1 illustrates an example of self-displacement. Assuming the recommended tracer volume is 7 ml and 0.1 ml is the usual volume in the assay, reconstitute the lyophilized tracer with 2 ml of water (instead of 7 ml). Remove 1 ml of concentrated tracer for self-displacement study (concentrated tracer). To the remaining tracer, add 2.5 ml of water (final volume 3.5 ml). This will now be the usual tracer dilution. Run recommended protocol for the standard curve and quality control samples. Table 1 shows the set up in more detail. For this example the specific activity is calculated as follows:

$$\underset{\text{(corrected)}}{\text{*Bound}} = \text{Bound} - \left(\underset{\text{(precount)}}{\text{Total}} \times \frac{\text{NSB after assay}}{\text{NSB precount}} \right)$$

$$\text{Free} = \underset{\text{(precount)}}{\text{Total}} - \underset{\text{(corrected)}}{\text{*Bound}}$$

$$\text{Slope} = \text{Specific activity (Fig. 5)}$$

$$\text{Mass bound per tube} = \text{Mass/tube} \times \frac{\text{*B}}{\text{T}}$$

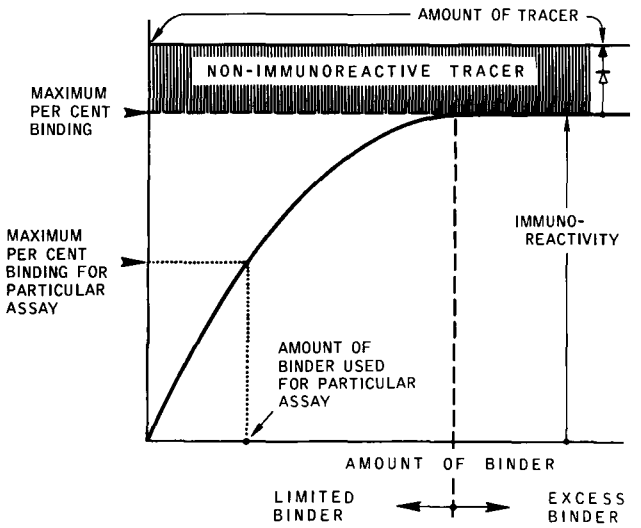

FIG. 6. Immunoreactivity. The amount of tracer bound (% of total) by excess antibody is immunoreactivity. For a given assay system, a fixed amount of binder is selected which then defines the maximum binding for the assay (B_0).

Immunoreactivity. The ability of a tracer to bind to the antibody is called immunoreactivity. This property is measured by reacting a constant amount of tracer, usually the amount used in the assay, with increasing amounts of antibody until a plateau is reached. The maximum percentage bound under these conditions is called the immunoreactive fraction (IF) (Fig. 6). Most calculations and theoretical models assume that all of the tracer is immunoreactive. If the immunoreactivity is low (< 80%, for example), then the kit may be expected to have poor sensitivity and a nonlinear Scatchard plot. Better standard curves and Scatchard plots may be obtained by multiplying the observed total counts by the immunoreactive fraction (TC × IF) in all calculations.

Scatchard Plot. The Scatchard plot is constructed by plotting the ratio of bound to free counts against the mass of ligand bound (Fig. 7). For RIA data, mass bound = B/T × (mass in standard + mass in tracer). Although criteria must be fulfilled before Scatchard's analysis can be strictly applied to competitive displacement assays, the Scatchard plot may be used qualitatively on kit curve data to assess assay conditions and reagent stability (assay-conditional Scatchard plot, ACSP) (Table 2).

Problems associated with calculating the data to be plotted are nonspecific binding correction and the estimation of tracer mass. One correction for NSB assumes that it is a nonsaturable binding that binds ligand linearly as ligand mass increases (e.g., is proportional to the free fraction). The immunoreactive mass of the tracer, evaluated as described above, may be used in the calculation. If one has no knowledge of the tracer mass, one can assume zero mass, or some other low value, in order to make the plot. When immunoassay data are plotted, one of three basic shapes will be generated that reflects the general adherence to the fundamental assumptions (Fig. 7). The utility

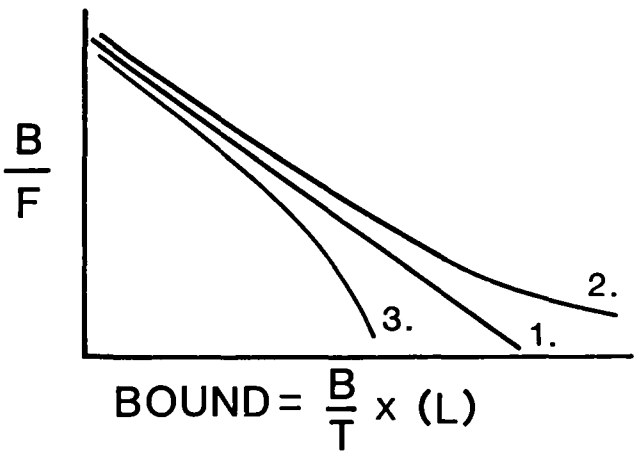

FIG. 7. Assay-conditional Scatchard plot (ACSP). Curve 1: apparently meets assumptions. Curve 2: multiple binding sites, K* > K. Curve 3: K* < K, poor separation, non-equilibrium. The ACSP is prepared by plotting bound/free counts (B/F) against mass of ligand bound (B/T × L), where L = (mass of standard + mass of tracer). Total counts may be corrected for immunoreactivity, Tcorr = T × IF. B may be corrected for nonspecific binding.

TABLE 2. Functional Requirements for the Assay-Conditional Scatchard Plot

Scatchard's Assumptions	Functional Requirements
First-order mass-action kinetics	The antibody is not divalent; no strong cooperativity
Single type of binding site	The antiserum has functionally homogeneous binding sites at the assay dilution
Thermodynamic equilibrium	The assay reaction is indistinguishable from equilibrium
Unbiased measure of bound and free ligand in the presence of the binder	The binding affinities of labeled and unlabeled ligands are similar
	The separation of bound and free ligand is approximately 100%
	The separation procedure does not disrupt the primary ligand-antibody complex

of the assay-conditional Scatchard plots for a routine laboratory derives from the fact that the data are easily available and, for a given assay system, the general shape remains the same from run to run. If reagents deteriorate, the plot will shift in a manner that correlates which reagent undergoes deterioration and thus will help in troubleshooting.

Data Reduction. Selection of the data reduction method is one of the critical steps in kit evaluation. The most important consideration is how well the mathematical model underlying a particular plot corresponds to the reality of the biochemical reaction being monitored. "Lack of fit" can be accounted for by poor precision and corrected by suitable weighing functions. Other curves necessarily require alternative models for an accurate assay. An initial recommended evaluation is to test several models for the one that best fits the individual kit. Parameters of fit are noted and are followed as quality control tools. Loss of fit may indicate that there has been some deterioration in a kit component, or that the individual standard concentrations particularly in low or high end are inappropriate for the assay.

NORMAL RANGE

The establishment of a normal, or reference, range, or even the definition of "normal" remains a controversial subject in the clinical literature. The discussions range from the statistical to the philosophical. Still, the laboratory must eventually select some estimates of expected values for healthy and diseased individuals in order for the test to be useful. Because many different approaches can be taken, the laboratory should at least know what assumptions were made and the limitations of any method selected.

Most kits are for tests that are relatively well established

in the research literature. The clinical application does not have to be reevaluated. In addition, the current regulatory and commercial climate has mandated that normal range studies be available from the manufacturer. The individual laboratory is usually limited in attempting to verify the limits of the normal range using a small sample size, and observing that the lab values are consistent with the published values.

The most frequently asked questions are "What subjects may I use as normals?" and "How many samples do I need?" Many factors affect the normal range: age, sex, race, weight, geography, diet, menstrual cycle, etc. Dealing with all these variables is difficult and expensive. For most assays, the important variables have been age and sex.

For all practical purposes, one may collect the samples from employees, pre-employment or yearly physical. Other sources are blood donors or the hospitalized population with no disease known to affect the test of interest.

There are two different statistical approaches to normal range estimation: parametric and nonparametric.

Parametric Methods. The most popular method to calculate normal range is the mean $\pm$ 2 SD calculated from observed values. This represents 95% of the normal population. The method assumes Gaussian or log-Gaussian distribution and depends on the adequacy of that assumption. A severely skewed distribution may be seen with a limited number of observations. Tests of the Gaussian assumption such as Chi-square, Kolmogorov-Smirnov, or plotting on proportionality paper may or may not be sensitive to some nonnormal distributions.

Nonparametric Methods. These methods have been highly recommended because there are no prior assumptions about the data distribution. The method is also valid for Gaussian or log-Gaussian distributions. A minimum of 120 samples are required to estimate 95% of the range with 90% confidence, and 75 samples for 95% of the range with 70% confidence. Because extreme values play a greater role than intermediate ones in these calculations, different tests are available to identify and reject outliers.

The laboratory can make simplifying assumptions to establish or verify reference ranges which are clinically useful, but the limitations should be kept in mind.

In summary, the selection of an assay kit that effectively meets the needs of the laboratory pays off in long-term assay stability, and provides accurate results and user satisfaction. With careful planning, the protocol described can be accomplished with approximately 400 tubes, excluding normal range study. When a kit does not perform well in early experiments, the evaluation can be discontinued so that when clinical samples and normal ranges are evaluated, the effort is directed only toward the final kit selection. When a kit performs well in the tests of precision and accuracy, the performance under expected laboratory running conditions should be tested. Full-sized runs should not reveal any front-to-back drift. The assay should be "robust" to environmental changes.

Although the complete protocol is elaborate, selected individual tests may be run to answer specific questions. A thoughtful plan allows the selection of the "best" kit, increases the understanding of how individual components contribute to good performance, and provides objective data which can be the foundation for ongoing trouble-shooting and quality control.

ADDITIONAL READINGS

1. Ekins R. Misunderstanding about sensitivity and precision, *Ligand Quarterly* 1981;4:60.
2. Ekins R. The "precision profile": Its use in RIA assessment and design, *Ligand Quarterly* 1981;4:33-44.
3. Yalow RS, Berson SA. Parallelism and cross reactivity. In *Principles of Competitive Protein-binding Assays*, Odell and Daughaday, eds, Philadelphia: Lippincott, 1971:374-400.
4. Scatchard G. The attractions of proteins for small molecules and ions. *Ann NY Acad Sci* 1949;51:660-72.
5. Chamness GC, McGuire WL. Scatchard plots: Common errors in correction and interpretation. *Steroids* 1975;26:538-42.
6. Hollemans HJG, Bertina RM. Scatchard plot and heterogeneity in binding affinity of labelled and unlabelled ligand. *Clin Chem* 1975;21:1769-73.
7. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. *Clin Chem* 1971;17:275-84.