Abstract

In everyday bioassay conversations, the word “sensitivity” is usually interpreted as having something to do with how small a concentration of analyte can be detected – as in, “The greater the sensitivity, the smaller the concentration the assay can detect.” The statistical meaning of “sensitivity” (and the meaning accepted by most international organizations such as the IUPAC) is quite different: it is the slope of the calibration relationship ($\Delta y/\Delta x$) at a given point, the change in response $y$ (e.g., $ED_{50}$) for a given change in analyte concentration $x$. For our purposes, the bioassayist’s “sensitivity” can be given another name, the “minimum consistently detectable concentration” (MCDC). The slope of the calibration relationship does influence the MCDC – in general, the steeper the slope, the lower the MCDC – but the measurement variability also plays a large role. Relationships among $\Delta y/\Delta x$, MCDC, limit of detection ($L_D$), and limit of quantification ($L_Q$) will be discussed. Complications often arise in the statistical treatment of this type of data because of censoring that commonly arises from the bioassay’s SOP.
The Problem (well stated):

“A fundamental principle behind immunogenicity assessment methods is that a cut point or threshold determines whether a sample is considered positive or negative for the presence of antibodies capable of binding to the therapeutic protein. The cut point is established statistically by observing the binding of samples that do not contain anti-therapeutic protein antibodies and determining what magnitude of signal in a test sample is sufficient to have it stand apart from negative samples. Any test sample that is greater than the established cut point is then considered to be positive for the presence of antibodies. In some assays, there is a reason why a sample might exhibit binding above the threshold other than the presence of specific antibodies. For these situations, a confirmatory assay can be utilized that provides additional confidence that a sample that tests positive in an assay contains specific anti-drug antibodies.”

Lloyd A. Currie


He made sense of the swamp.

“Until Lloyd Currie’s paper *Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry* ... was published, there was enough inconsistency in the definition of "detection limit" to conceal a great deal of disagreement. In just over seven pages, this tightly written communication established a high level of uniformity in answering these questions. The paper contains fundamental information that has made it influential far beyond its size, and it is rich enough to be discussed actively in e-mail newsgroups over 30 years later. This is surely one of the most often cited publications in analytical chemistry.”


Sensitivity
Sensitivity: Calibration

When talking with the rest of the world about “sensitivity”, bioassayists need to be careful:

The observed relationship between the size of the measured signal (activity y) and the magnitude of the property (concentration x) is called the calibration relationship, or the calibration curve or the calibration line.

With many measurement processes, the measured signal is directly proportional to the magnitude of the property, especially at low concentrations, as shown in the calibration line at the right.

(The fundamental ideas in this talk can be developed equivalently using “amount” instead of “concentration” for the horizontal axis.)

Sensitivity: Definition

The rest of the world defines the sensitivity (S) of a measurement process as the change in the measured signal (Δy) divided by the change in the property being measured (Δx):

\[ S = \frac{\Delta y}{\Delta x} \]

The sensitivity of a measurement process is thus the slope of the calibration line at a given point.

This is the definition of sensitivity adopted by most international organizations.

THIS DEFINITION DOES NOT MENTION THE SMALLEST CONCENTRATION THAT CAN BE DETECTED
**Sensitivity:** Non-Constant Sensitivity

The sensitivity of a measurement process might not be constant.

As an example, a fixed amount of a critical reagent (e.g., cells) in a bioassay might prevent the true activity from being realized at high concentrations.

In this figure, the sensitivity (slope) is greater at smaller concentrations than it is at larger concentrations.

When bioassayists use the word "sensitivity", they usually mean either

(a) how large the measured activity must be before they're willing to say that something is present (limit of detection)

(b) the minimum concentration that can be detected reliably (minimum consistently detectable concentration)

**Sensitivity:** What To Use On The x-Axis

As an aside, consider the simple binding equation:

\[ A + B \rightleftharpoons C \]

\[ K_{eq} = \frac{[C]}{[A][B]} \]

Assume that the equilibrium constant \( K_{eq} = 10.0 \) and the initial concentration of \( B = 2.0 \times 10^{-5} \). The panel at the right shows how the concentration of C increases as the initial concentration of A is increased.

Note that B is the limiting reagent. At low concentrations of A, the relationship between C and A is approximately a straight line.

When \([C]\) is plotted against \(\log_{10}[A]_{init}\) the resulting "calibration curve" is a sigmoid.
**Sensitivity**: What To Use On The x-Axis

The curve in the lower panel on the previous slide can be approximated very closely with a “three-parameter logistic” (3PL) model:

\[
y = \frac{D}{1 + \left(\frac{x}{C}\right)^B}
\]

The following 3PL model mimics the sigmoid almost exactly:

\[
[C] = \frac{2.0 \times 10^{-5}}{1 + \left(\frac{x}{1.0 \times 10^{-7}}\right)^{-1}}
\]

So what’s the point? When considering the limit of detection and minimum consistently detectable concentration, a non-log concentration (or amount) axis is usually more useful.

---

**Limit Of Detection**
**Limit Of Detection: Introduction And Statistical Concepts**

A discussion of the limit of detection focuses on the risk $\alpha$ of a false positive (Type I Error).

The concept requires a good estimate of the true mean signal of the blank ($\mu_b$), usually obtained from repeated measurement of a true blank (a sample containing no analyte – antibody in this case), and a good estimate of the standard deviation of the blank ($\sigma_b$), again usually obtained from repeated measurements of a true blank.

"The cut point should be statistically determined by using negative control samples (e.g., samples from patients not exposed to product)." [5-10 samples for development; 50-100 samples for validation]


**Limit Of Detection: True Mean Of The Blank**

The mean of all these $n$ measurements is labeled $\mu_b$, the "true mean of the blank." It is approximated by the calculation

$$\mu_b \approx \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

The variation in the measurements can be described by the standard deviation of the blank or near blank, $\sigma_b$, approximated by the calculation

$$\sigma_b \approx s_b = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}}$$

The true mean of the blank, $\mu_b$, and the standard deviation of the blank, $\sigma_b$, describe a Gaussian distribution showing the frequency of occurrence of the measured values.
**Limit Of Detection:** If $L_D = \mu_b$

Given a limit of detection $L_D$ (the "cut point"), the rules are:

- if $y \geq L_D$, analyte is said to be **detected**
- if $y < L_D$, analyte is said to be **not detected**

Consider a sample that contains no analyte. Suppose the limit of detection $L_D$ is set at $\mu_b$.

Half of the time, this sample that contains no analyte will produce a measured signal $y$ that is less than $\mu_b$, and the analyte will be said to be not detected or absent. This is correct.

Half of the time, this sample that contains no analyte will produce a measured signal $y$ that is equal to or greater than $\mu_b$, and the analyte will be said to be detected or present. This is incorrect. It is called a false positive.

The fractional risk of detecting the analyte when, in fact, it is absent as $\alpha = 0.5$.

“FDA recommends that the cut point have an upper negative limit of approximately 95 percent. While this value yields a 5 percent false positive rate, it improves the probability that the assay will identify all [sic] patients who developed antibodies. This sensitivity [sic] is particularly important in the initial screening assay as the results dictate the further analysis of the sample for NAB.

Several approaches can be used. For example, parametric approaches using the mean plus 1.645 standard deviation (SD), where 1.645 is the 95th percentile of the normal distribution may be appropriate.”


**Limit Of Detection:** If $L_D = \mu_b + 1.645\sigma_b$

If $L_D$ is set higher at $\mu_b + 1.645\sigma_b$, then the fractional risk of detecting analyte when, in fact, it is absent is only $\alpha = 0.05$. 

"FDA recommends that the cut point have an upper negative limit of approximately 95 percent. While this value yields a 5 percent false positive rate, it improves the probability that the assay will identify all [sic] patients who developed antibodies. This sensitivity [sic] is particularly important in the initial screening assay as the results dictate the further analysis of the sample for NAB.

Several approaches can be used. For example, parametric approaches using the mean plus 1.645 standard deviation (SD), where 1.645 is the 95th percentile of the normal distribution may be appropriate.”
Limit Of Detection: $L_D$ Is On The Measurement Axis

Returning to the calibration relationship, note that $L_D$ is on the signal or measurement axis.

$L_D$ does not appear on the horizontal axis. It has nothing to do with the concentration of analyte. (The next section will discuss MCDC, which does have something to do with the horizontal axis, the concentration of analyte.)

Anticipating the next module, note that setting $L_D$ greater than $\mu_b + 1.645\sigma_b$ greater in this example) does reduce the number of false positive results when the concentration of analyte is zero, but the price to be paid for this is that false negative results occur for small but real concentrations of analyte.

“Not detected” doesn’t mean “absent”. Thus, the FDA didn’t really mean “it improves the probability that the assay will identify all patients who developed antibodies.” There will always be some false negatives.

Minimum Consistently Detectable Concentration
**MCDC: LDC**

The limit of detection, $L_D$, is on the vertical measurement axis.

$L_D$ represents the minimum "signal strength" (measured activity $y$) above which it can be said with confidence that analyte is present (95% confidence if $L_D = \mu_b + 1.645\sigma_b$).

By working back through the calibration line, it can be seen that there is a concentration of analyte that corresponds to this limit of detection.

This concentration of analyte will be called the "limit-of-detection concentration" or the LDC.

**MCDC: Detected concentrations**

The limit of detection concentration (LDC) is on the concentration axis, the horizontal axis.

The limit of detection ($L_D$) is on the signal axis, the vertical axis.

Signals less than $L_D$ will not lead to confidently detected concentrations of analyte.

Reported values of detected analyte will be at or above the LDC.

Thus, there exists a sort of "blackout region" below $L_D$ and to the left of the LDC where the client doesn't get any quantitative information — just "Not Detected" or its equivalent.
**MCDC: Consistent Detection**

What will happen if we repeatedly measure a sample carefully prepared to contain the LDC of analyte? The results are shown in the figure at the right.

The variation of individual measurements about this mean is assumed to be the same as the standard deviation of the blank, \( \sigma_b \) (homoscedastic noise at low concentrations).

It is clear that repeated measurements of a sample carefully prepared to contain the LDC of analyte do not always result in the analyte being detected!

The results are not consistent

- sometimes the analyte is detected
- sometimes the analyte is not detected.

**MCDC: False Negatives**

Measurements that fail to detect the presence of analyte when it is known to be present are called "false negatives." In this example, the sample is known to contain the LDC of analyte — analyte is present and should be detected. But half of the time, because of statistical variation, the measured signal will be less than \( L_D \) and the analyte will not be detected.

The fractional risk of not detecting the analyte when it should be detected is \( \beta = 0.5 \) for a sample that contains the LDC of analyte.

Again, many clients (and analysts) wrongly interpret “not detected” to mean “analyte is absent”. “Not detected” should be interpreted as “don't know”: analyte might or might not be present — the evidence just isn't strong enough to say with enough certainty that it is present.
**MCDC:** MCDC

The risk $\beta$ is not a constant but depends on the concentration of analyte. Thus, the false negative rate can be reduced if more analyte is added to the sample.

If enough analyte is added to raise the mean signal to, say, $1.645\sigma_b$ above $L_D$ (equivalent to $3.29\sigma_b$ above $\mu_b$), then the false negative rate can be reduced to $\beta = 0.05$. The analyte will be detected more consistently.

The concentration of analyte that gives sufficiently consistent detection will be called the "minimum consistently detectable concentration," or MCDC.

The concept of "consistently detectable" is important. It means that if a sample containing sufficient analyte is measured again and again and again, most of the time (95% of the time, in this example) it will be concluded that analyte is present.

---

**MCDC:** MCDC

For a straight-line calibration relationship, if $\alpha = \beta$, the MCDC will always be twice the LDC, as seen in this example.

However ... the consequences of false positives are often different from the consequences of false negatives. As a result, the corresponding risks $\alpha$ and $\beta$ will often be set to different values.

Once $\alpha$ and $\beta$ are set, the corresponding multipliers (e.g., 1.645, 3.29) can be determined from a table of critical values of $z$.

The lower the risk $\alpha$ of a false positive (the higher the cut point), the more likely it will be that small but non-zero concentrations will not be detected. Thus, in the FDA document referenced earlier, it was recommended that $\alpha$ be set at a relatively high risk (a lower cut point) so that small concentrations would have a greater probability of being detected.
**MCDC**: Resolving A Conundrum

How can an analyst detect a concentration of analyte that is less than the MCDC?

The LDC and the MCDC are different because of the different concepts involved:

If a value is reported above the LDC, there is at least 95% confidence that the sample does contain analyte. But if the concentration of analyte in a sample is less than the MCDC, there is not at least 95% confidence that it will be detected.

LDC is the "minimum detected concentration"

MCDC is the "minimum detectable concentration"

**Consistently detectable**, that is. It's OK (it's logical) to report with confidence that we have detected analyte, even though we can't be confident that its detection will be consistent.

---

**Limit Of Quantitation**
**Limit Of Quantitation: Noise At The LDC**

Because noise tends to be homoscedastic at low levels of analyte, the standard deviation of the blank $\sigma_b$ is a good estimate of the noise above and below the limit of detection $L_D$.

When measured signals are converted to concentrations through the calibration curve, the standard deviation of the resulting concentrations ($\sigma_c$) on the horizontal axis is calculated as

$$\sigma_c \text{ (conc.)} = \frac{\sigma_b \text{ (activity)}}{\text{sensitivity (activity/conc.)}}$$

For a given uncertainty of measurement ($\sigma_b$), a steeper slope (greater sensitivity) gives a smaller uncertainty of reported concentration ($\sigma_c$), and LDC and MCDC move to smaller concentrations. *This is why high sensitivity (slope) is so important.*

---

**Limit Of Quantitation: %RSD vs. q**

If $q$ is the analyte concentration expressed as a number of standard deviations ($\sigma_c$),

- $q = 1.645$ (the LDC) = ~60% RSD
- $q = 3.29$ (the MCDC) = ~30% RSD
- $q = 10.0$ (my LQC) = ~10% RSD

Recall that these three quantities are based on three separate statistical concepts:

- the *limit of detection concentration*, LDC, is based on the fractional risk $\alpha$ of stating that the property is present when, in fact, it is not present (the risk of a false positive)

- the *minimum detectable concentration*, MCDC, is based on the fractional risk $\beta$ of stating that the property is not present when, in fact, it is present (the risk of a false negative)

- the *limit of quantification concentration*, LQC, is based on the maximal acceptable percent relative standard deviation (%RSD) of a reported value (it's an ego thing)
Censoring: “Non-Reportable” Values

The standard operating procedure (SOP) of some bioassays often censors the data available for determining the limit of detection and related quantities.

For example, “activity” is sometimes reported as a so-called “NF$_{50}$” value (a relative concentration measured against a reference standard, a ratio of two so-called “ED$_{50}$” values). However, the SOP might arbitrarily insist that an ED$_{50}$ value is not reliable if it is based on “less than half a curve” and therefore the NF$_{50}$ value calculated from it cannot be reported. This censors the data.

Because the noise structure of the data is usually homoscedastic at the low end of the calibration relationship, the use of a “near blank” can often overcome this potential limitation.
ROC Curves: Opinion

Receiver Operating Characteristic curves show “the performance of a binary classifier as its discrimination threshold is varied. It plots the fraction of true positives out of the positives (true positive rate) vs. the fraction of false positives out of the negatives (false positive rate), at various threshold settings.” The curve was developed during World War II to increase the prediction accuracy of identifying Japanese aircraft.

The decision of Japanese aircraft or not is not a clear binary distinction (either completely positive or completely negative).

Opinion: The decision of antibody or not is a fuzzy binary distinction (either completely negative or various shades of positive). Need independent assessment of categorization before ROC curves are useful. OR don’t use antibody as a surrogate – use “taken drug” or “not taken drug” as the direct categorization.
Some Additional References


Laureen Little (well read).

---

Thanks Danke Merci Grazie Engraziel

Statistical Definition of Sensitivity and How it Pertains to Cut-Off Assays

BEBPA

Biopharmaceutical Emerging Best Practices Association

Basel, Switzerland
25 September 2013

Stanley N. Deming, Ph.D.
Statistical Designs
9941 Rowlett Road, Suite 10
Houston, TX 77075-3404 U.S.A.
713.947.1551
standeming@statisticaldesigns.com