New Method of Evaluation of Limit of Detection in Molecular Diagnostics

Jeffrey E. Vaks
Roche Molecular Diagnostics, 4300 Hacienda Dr., Pleasanton, CA 94588

Abstract
With state-of-the art polymerase chain reaction based in-vitro diagnostics tests, the limit of blank (LoB) is zero and the limit of detection (LoD) is concentration of target nucleic acid corresponding to 95% probability of detection. It is an important characteristic of performance of in-vitro diagnostics tests that is inversely proportional to the analytical sensitivity. The current method of estimation of limit of detection (LoD) in molecular diagnostics, recommended in CLSI guideline EP17-A2 on evaluation of detection capability[3], is probit analysis. It consists of describing the relationship between the probability of detection and concentration using a cumulative probability curve, which is an empirical model for describing the relationship. The new method uses theoretical model to describe the relationship derived and validated with data and published by the author in 2017 JSM Proceedings[5] along with examples of applications. Using the theoretical model, maximum likelihood method of estimation of LoD has been developed. It has substantial advantages over the probit analysis: (1) both LoD and the minimum number of copies of nucleic acid required for detection are estimated, (2) the estimates of LoD have tighter confidence intervals and are not biased, (3) while substantial proportion of probit estimates suffer from lack of fit, there is no lack of fit of theoretical model to data problem, (4) smaller number of concentration levels and smaller number of replicates need to be tested with the new method. Also, efficient method of sequentially determining concentration levels to be tested in LoD studies is proposed, LoD in case of detection of more than one DNA/RNA region, and LoD for tests on microorganisms having large number of DNA/RNA copies per organism are discussed. Software for the calculations with the new method has been developed and copyrighted by Maplesoft, Inc., Waterloo, Canada.

Key Words: Molecular diagnostics, Limit of detection, Math Model for Probability of Detection, Maximum Likelihood Estimation.

1. Introduction

Molecular diagnostics tests are done on DNA and RNA biomarkers to identify infectious viruses, bacteria, parasites and human DNA mutations. Test results are used to screen for infections such as various genotypes of the human papilloma virus, HPV; screen donor blood for infections with human immunodeficiency virus, HIV, and other viruses; diagnose infectious disease (AIDS, Hepatitis B, STD's, etc.) and prescribe medications (personalized heath care); monitor therapy. Tests are done using polymerase chain reaction, PCR, to amplify and detect the target nucleic acids present in the samples in low counts of the molecules. In molecular diagnostics, test result is positive when the number of target nucleic acid copies per PCR is sufficient for detection, and the probability of detection is the probability that the number of copies extracted from randomly drawn sample is sufficient for detection.

Usually the number of amplification cycles available on commercial instruments is sufficient for amplification and detection of a single copy per PCR, and with no restriction

on the number of amplification cycles, *Ct*, a single copy per PCR is detected. In special cases, the number of amplification cycles and or fluorescent light intensity threshold are set so as more than 1 copy of target is required for detection. Examples of such cases are (1) intentional reduction of the assay analytical sensitivity with corresponding increase of LoD to better match the PCR results obtained on bacterial genetic material, that includes non-growing bacteria, to the bacterial growth test, (2) reduction of the rate of false positives due to cross-contamination of negative samples with the positive samples tested in the same batch.

Limit of detection (LoD) in in-vitro diagnostics is one of the most important assay performance characteristics. It is inversely proportional to the analytical sensitivity. The recommended methods of validation by the manufacturers and verification by the users of the LoD in in-vitro diagnostics were first described in the first edition of the Clinical and Laboratory Standards Institute (CLSI) guideline EP17-A[2] published in 2004. The statistical method used in EP17-A was published by K. Linnet and M. Kondratovich[1]. The method is applicable to quantitative or qualitative general clinical chemistry and immunoassays that have continuous internal signal with blank and containing the analyte samples. The LoD is defined in such case as the analyte concentration with specified β-quantile (typically, 0.05) of the distribution of replicated test results equal to the specified α-quantile (typically, 0.95) of the noise distribution of replicated tests with the blank samples. The latter is the LoB.

State-of-the-art PCR-based assays do not have a signal with blank samples, leading to zero LoB, with LoD = 1 - β. For PCR-based assays, the LoD is typically defined as the analyte concentration corresponding to 95% detection rate. The LoD evaluation study is done as recommended in CLSI Guideline[3] EP17-A2, 2nd Ed., published in 2012, using samples with several concentration levels having probabilities of detection between 0 (blank samples to confirm that LoB =0) and 100%, the latter samples having analyte concentrations above LoD. The rather low concentration samples, with low probabilities of detection, have some proportions of replicated tests done on samples with no copies of the target nucleic acid at the PCR input. Because of that, some of the test results have no internal signal, and the distribution of the internal signal, *Ct*, is truncated. Because of that, the LoD of PCR assays cannot be estimated as the 95th percentile of concentrations corresponding to the 95th percentile of the internal signal. This is the reason for the methods of LoD estimation for the PCR assays requiring use of the observed detection rates at the concentration levels tested for the estimation.

Nonparametric method of estimation of the PCR assay LoD consists of assigning to the LoD the lowest concentration of those tested that has at least 95% observed detection rate with all higher concentration levels having at least 95% observed detection rates. The advantages of such method are in its computational simplicity and absence of any assumptions about probability of detection vs. concentration. The flaws of the method are positive bias of the LoD estimate and very wide, virtually useless, limits of the confidence interval of the estimate with reasonable number of concentration levels tested. Though we have derived the equations for calculations of the bias and the confidence limits of the LoD estimate with this method, we do not include those in this paper because of the availability of better parametric methods.

The parametric method, recommended in the 2nd Edition of the CLSI Guideline EP17-A2 for the molecular diagnostics PCR assays, is based on fitting an empirical model to the observed detection rates vs. concentrations and calculating from the model the LoD as concentration corresponding to 95% detection rate, along with the confidence limits. E.g., in SAS® PROC PROBIT procedure, three cumulative distribution curves are available for use as empirical models for the probability of detection vs. concentration: normal, logistic and extreme values. The first two distributions are mentioned in EP17-A2. Our simulation studies have shown that the extreme values distribution has the lowest frequency of the lack of fit events and smallest deviations from the LoD estimates obtained using the theoretical math model. The frequency of lack of fit events is lower when the fit of the empirical models with PROC PROBIT is done to the observed detection rate vs. log10(Concentration). The flaws of the parametric method using empirical model for probability of detection vs. concentration are biased estimation, noticeable proportion of lack of fit events, often wide and sometimes non-existent confidence intervals for the LoD, and inability to estimate the minimum number of copies of target nucleic acid required for detection.

In this work, a new method, free of the flaws of the method based on empirical model, is described. The method is maximum likelihood estimation (MLE) of LoD based on using theoretical math model of the probability of detection vs. concentration. Also practical considerations of planning and conducting LoD evaluation studies are discussed. A special case of LoD, when detection is done with more than one region of DNA sequence, is also considered. The probability to pass verification by the user of the manufacturer claimed LoD was discussed in a paper by J. Vaks, et al[4].

1. The New Method of Estimation of LoD

2.1. Theoretical Math Model for Probability of Detection vs. Concentration

The probability of detection, p, vs. concentration and the minimum number of copies required for detection can be modelled with cumulative Poisson distribution (Vaks, JE, 2017 JSM Proceedings[5]):

$$p = 1 - \sum_{u=0}^{v-1} \frac{\xi^u e^{-\xi}}{u!} \tag{1}$$

In (1), v is the minimum number of copies of target nucleic acid required for detection; ξ is the mean number of copies per PCR, extracted and, if necessary, reverse transcribed from randomly drawn samples. ξ can be expressed as a product of the mean number of copies per test sample volume, μ, and the probability (efficiency) of extraction/reverse transcription, θ:

$$\xi = \mu\theta \tag{2}$$

Finally, the extraction/reverse transcription efficiency can be expressed as:

$$\theta = \frac{r_v \ln(20)}{LoD_v} \tag{3}$$

$$r_v = LoD_v / LoD_1 \tag{4}$$

LoD_v and LoD_1 are the limits of detection with v and 1 copies required for detection, respectively. Combining equations (1) to (4) yields for the probability of detection:

$$p = 1 - \sum_{u=0}^{v-1} \frac{\left(\frac{\mu \ln(20) r_v}{LoD_v} \right)^u e^{-\left(\frac{\mu \ln(20) r_v}{LoD_v} \right)}}{u!} \qquad (5)$$

The math model for the probability of detection based on Poisson approximation of the cumulative binomial distribution, with test samples are randomly drawn from a pool of a volume containing > 10 test sample volumes, is acceptable in practice[5]. That requirement is well satisfied when drawing a random sample of 1 mL from an adult patient blood volume of ~5000 mL. Since the probability of detection in math model (5) is a function of v and LoD it can be used for estimation of both.

It is easy to show that r_v is a function of v only, and its value can be calculated by solving numerically the following equation[5]:

$$\frac{e^{-r_v}}{\ln(20)} \sum_{u=0}^{v-1} \frac{r_v^u}{u!} = 0.05 \qquad (6)$$

The r_v values for v from 1 to 100 are summarized in Table 1 borrowed from the mentioned author's paper[5].

Table 1: r_v values for v from 1 to 100

v	r_v	v	r_v	v	r_v	v	r_v	v	r_v
1	1	21	9.701	41	17.381	61	24.832	81	32.162
2	1.584	22	10.095	42	17.758	62	25.201	82	32.527
3	2.102	23	10.487	43	18.134	63	25.569	83	32.891
4	2.588	24	10.877	44	18.509	64	25.938	84	33.255
5	3.056	25	11.267	45	18.884	65	26.306	85	33.618
6	3.509	26	11.655	46	19.259	66	26.674	86	33.982
7	3.953	27	12.043	47	19.633	67	27.041	87	34.345
8	4.389	28	12.429	48	20.007	68	27.408	88	34.708
9	4.818	29	12.815	49	20.380	69	27.775	89	35.072
10	5.243	30	13.199	50	20.753	70	28.142	90	35.434
11	5.662	31	13.583	51	21.126	71	28.509	91	35.797
12	6.078	32	13.966	52	21.498	72	28.875	92	36.160
13	6.490	33	14.348	53	21.870	73	29.241	93	36.522
14	6.899	34	14.729	54	22.241	74	29.607	94	36.884
15	7.306	35	15.110	55	22.612	75	29.973	95	37.246
16	7.710	36	15.490	56	22.983	76	30.338	96	37.608
17	8.112	37	15.869	57	23.353	77	30.703	97	37.970
18	8.512	38	16.248	58	23.723	78	31.068	98	38.332
19	8.910	39	16.626	59	24.093	79	31.433	99	38.693
20	9.306	40	17.004	60	24.463	80	31.798	100	39.055

2.2. MLE LoD Method Using Theoretical Model for Probability of Detection

Following CLSI EP17-A2 guideline, data for LoD estimation are collected in the form of Table 2 with minimum $I = 5$ concentration levels.

Table 2: Data Format for LoD Estimation

Target Concentration	Number of Tests	Number Positive
μ_1	n_1	x_1
…	…	…
μ_i	n_i	x_i
…	…	…
μ_I	n_I	x_I

The likelihood to observe a set of the numbers of positive results, x_i, in n_i tests of the samples with respective concentrations μ_i (expressed as the mean numbers of copies per sample volume) is the product of the probabilities to observe the respective numbers of positive results:

$$L\left(x_1,...,x_I\right)=\prod_{i=1}^{I}P\left(x_i\right) \tag{7}$$

The probability, $P(x_i)$, is binomial:

$$P\left(x_i\right)=\binom{n_i}{x_i}p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \tag{8}$$

In (8), p_i is calculated using (5) for respective concentration, μ_i. Pair (v, LoD_v) maximizing likelihood (7) is the maximum likelihood estimate (MLE) of LoD and v. Log-likelihood, l, is often used for numerical calculations.

From equations (3) and (4) it is easy to see that the limit of detection is inversely proportional to the target nucleic acid extraction efficiency and proportional to the ratio (values summarized in Table 1) of the LoDs corresponding to the minimum numbers of target copies required for detection, v and 1, respectively. The state-of-the art PCR tests detect a single copy per PCR[5]. Such PCR test performance provides for the lowest LoD and highest analytical sensitivity, which is important for donor blood screening for infectious agents and evaluating ability of new pharmaceuticals to cure such diseases as AIDS, Hepatitis B or C, etc. With $v = 1$ in (5) and (3), (5) simplifies:

$$p=1-e^{-\left(\frac{\mu\ln(20)}{LoD}\right)} \tag{5,a}$$

In other cases, such as maintenance of AIDS patients using a set of medications and monitoring progress of treatment of Hepatitis B or C, minimizing the rate of false positive results due to cross-contamination of low concentration samples with high concentration samples is more important. Also, better matching PCR microbiology test results, with detection of both growing and non-growing bacteria, to bacterial growth tests, by reducing the clinical sensitivity and increasing the LoD, is achieved by designing the PCR test so as a certain minimum number of copies $v > 1$ for target detection is required. The latter is accomplished by restricting the number of amplification cycles and/or increasing the fluorescent light threshold for detection.

When $v > 1$ improves the PCR test performance, the *Ct* cutoff and/or fluorescent light threshold can be optimized directly. The exact v in such case is not known. It is estimated, along with the LoD, with the new method, to provide additional information that is useful to the PCR test developer. Such approach also provides for unbiased LoD estimate and the confidence bounds for both v and LoD.

There is a special case of PCR tests – digital PCR – with specified and known exactly minimum number of positive droplets required for detection since they are counted directly. It should be noted though that the number of target nucleic acid copies per droplet varies, with vast majority of positive droplets having a single copy of target nucleic acid. The possibility of more than 1 copy per droplet is ignored here. In such case, only the LoD needs to be estimated with the new MLE method, and maximization of the likelihood (7) is done for fixed v. In both cases of known $v = 1$ with PCR tests, and specified v with digital PCR tests, the maximization of the likelihood is 1-dimensional. In such case, the bounds of the confidence interval for the LoD are found as two roots of the equation[6]:

$$l = \max_{LoD,v}(l) - \chi^2_{1-\alpha,df=1}/2 \qquad (9)$$

In (9), l = log of likelihood given in (7), $1 - \alpha$ = confidence level. For 95% confidence, $\chi^2_{0.95,df=1}/2 = 1.92$. So, from (9), the 95% confidence bounds for LoD correspond to the log likelihood that is 1.92 below the maximum log likelihood.

In case of estimation of both v and LoD, the bounds of 2D $(1 - \alpha)\cdot 100\%$ confidence region are calculated in a range of v solving (9) numerically for LoD at each v. For 95% confidence, the graph in Figure 1 shows (*LoD*, *v*) pairs corresponding to log likelihoods that are 1.92 below the maximum log likelihood – the points for integer v values on the plane crossing 3D log likelihood surface at the level 1.92 below the maximum log likelihood.

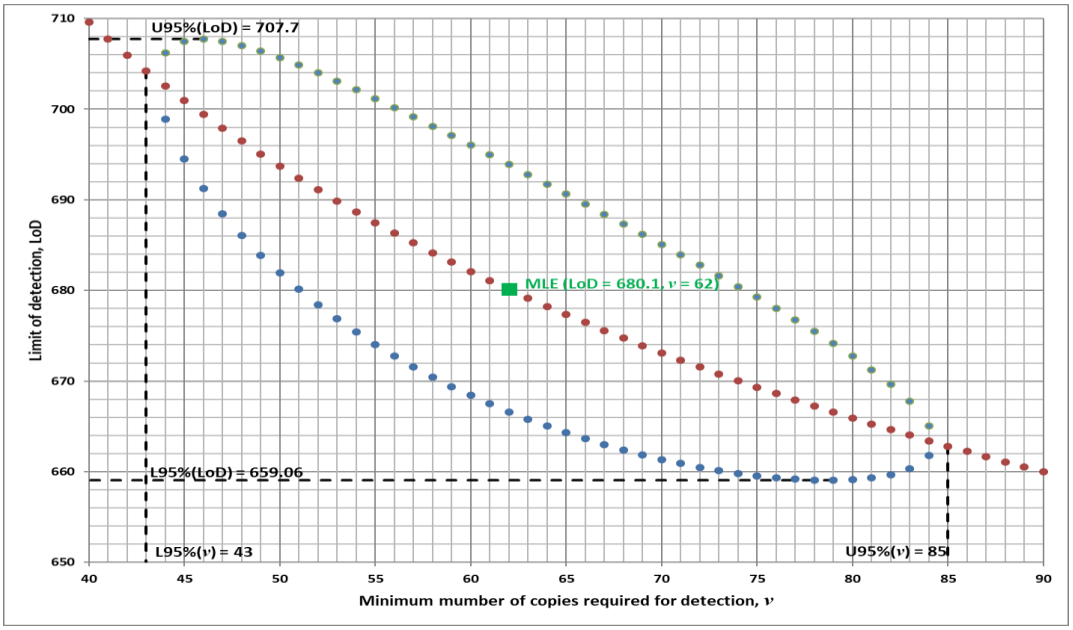

Figure 1. Example of a 2D 95% Confidence Region for (*LoD*, *v*)

The graph also shows the 1D 95% confidence bounds for v and LoD obtained from a 2D confidence region as the respective pairs of horizontal and vertical lines enclosing the 2D confidence region. The v and LoD estimates vary because of random variation of the observed detection rates at concentrations tested. It is possible to observe any number of positive results from 0 to n_i with binomial probability (8). Random variation of the observed detection rates in a particular study, the numbers of tests done at the concentration levels tested and their number and distribution affect the widths of the confidence intervals for v and LoD. If the widths of the confidence intervals are acceptable, the estimates of v and LoD are accepted along with their respective confidence intervals.

2.3. Real data examples

2.3.1. Example 1, HIV Blood Screening Test

An example of data collected at LoD evaluation of a blood screening test for HIV virus are given in Table 3. Graph with LoD estimation with probit analysis is shown in Figure 2.

Table 3: Data of LoD evaluation for HIV blood screening test

Concentration	Number of Tests	Number Positive
30	63	62
15	63	54
7.5	63	36
4.5	63	30
1.5	63	18

The analysis was done using SAS® PROC PROBIT procedure using log(concentration), cumulative normal curve and LACKFIT options. With LoD estimated at 34.6 IU/mL, lack of fit resulted in a very wide 95% confidence interval from 15.7 to 520.7 IU/mL. The LoD estimate as concentration corresponding to 95% detection rate is higher than the concentration 30 IU/mL corresponding to probability of detection 62 / 63 = 98.4% (see Table 3 and Figure 2). Since more often no lack of fit takes place, it must be a result of large variation of the observed detection rates. Looking at the graph in Figure 2, it is impossible to tell which detection rate vs. concentration observation may be an outlier that needs to be investigated and retested or to have additional replicates collected.

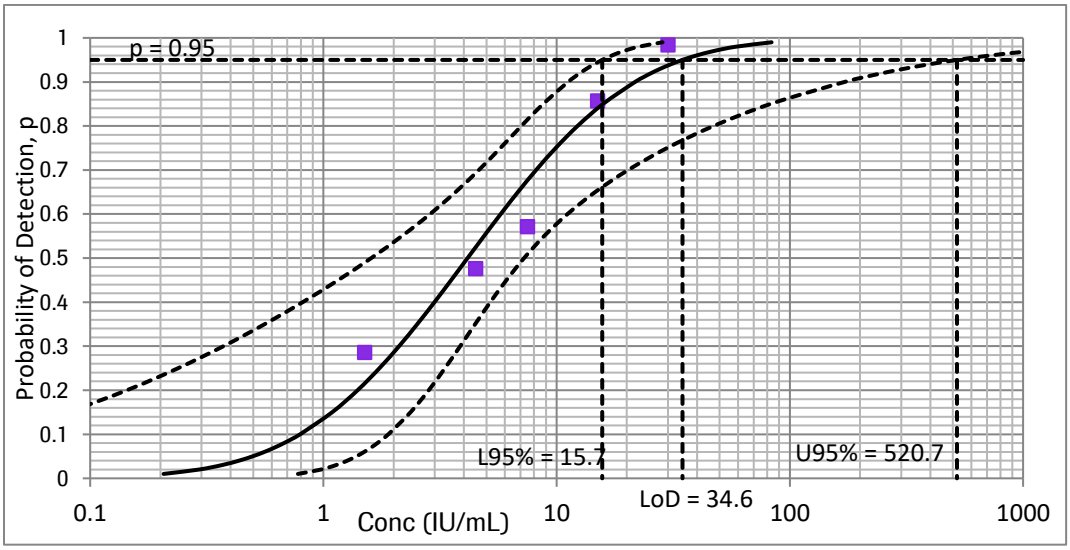

Figure 2. Probit Analysis of LoD Study Data for HIV Blood Screening Test.

Table 3 data analysis results using the new theoretical model based MLE method are shown in Figure 3.

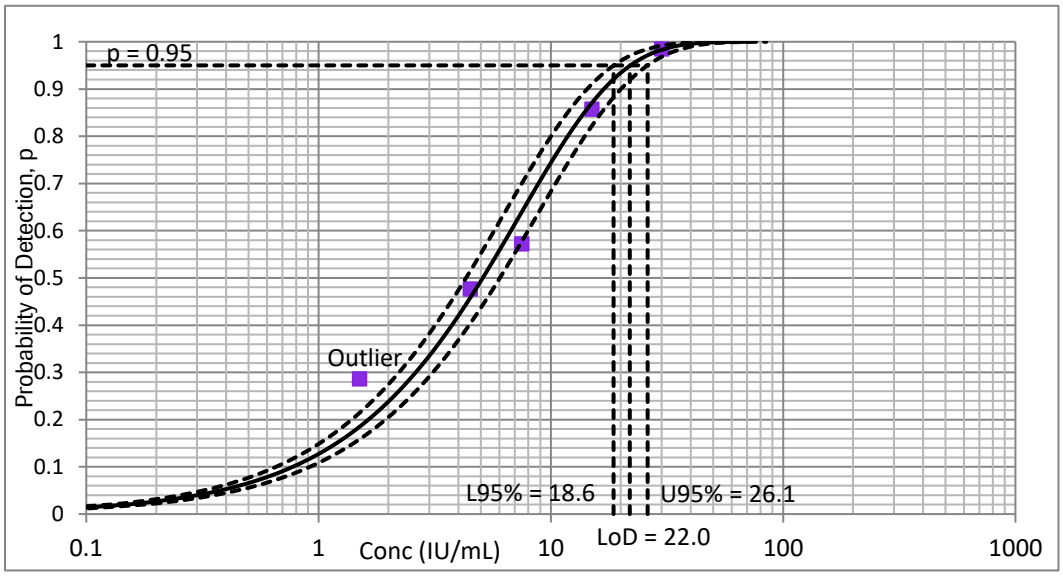

Figure 3. Analysis of LoD Study Data for HIV Blood Screening Test with New MLE Method

The results obtained with the new method have several favorable features compared to the results obtained with the probit analysis:

- As expected, the LoD estimate of 22.0 IU/mL at 95% detection rate is lower than the tested concentration of 30 IU/mL that has observed detection rate of 98.4%
- The 95% confidence interval from 18.6 to 26.1 IU/mL is ~67 times shorter than the confidence interval from 15.7 to 520.7 IU/mL obtained with the probit analysis
- From the graph in Figure 3, obtained with the new method, the problematic observation is clearly seen: it is the observed detection rate 18 / 63 = 28.6% at concentration 1.5 IU/mL.

From the graph, it is not clear what was the cause of the aberrant observation. The actual concentration might be higher than the nominal concentration. Assuming that this was the cause, and changing the concentration for a computational experiment from 1.5 to 2.5 IU/mL brings the data point on the theoretical curve describing the probability of detection as a function of concentration. The graph with results obtained with new method and outlier fixed are shown in Figure 4. Fixing the outlier changed the LoD estimate with the new method from 22 to 23.3 IU/mL by just 6%. The confidence interval shifted by about the same amount that the LoD estimate did.

Using such fix, the data were reanalyzed also with the probit analysis. The results are shown in graph in Figure 5. The estimate of LoD has been reduced by 22% from 34.6 to 27.1 IU/mL. As expected, it is now below 30 IU/mL that has observed detection rate of 98.4%. There is no lack of fit. The 95% confidence interval changed from 15.7 to 520.7 IU/mL to 20 to 42.7 IU/mL, 22 times tighter. It is still 3 times wider than the 95% confidence interval obtained with the new method.

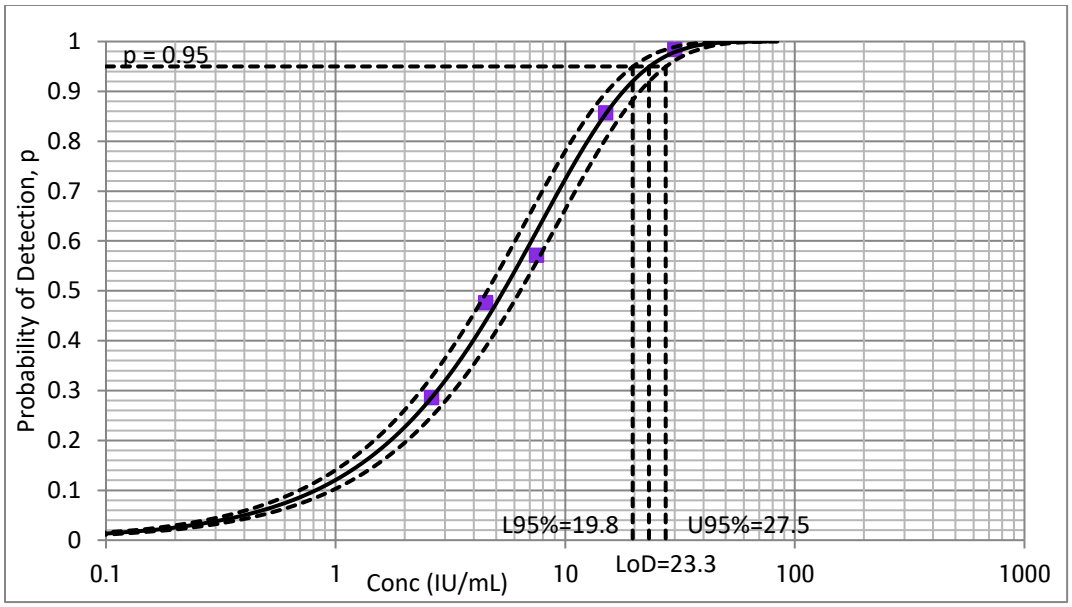

Figure 4. Analysis of LoD Study Data for HIV Blood Screening Test with New MLE Method; the Outlier is Fixed in Computational Experiment.

Comparison of the behaviors of the new method based on theoretical math model for probability of detection and the probit analysis based on empirical model shows one more advantage of the new method: it is substantially less sensitive to outliers.

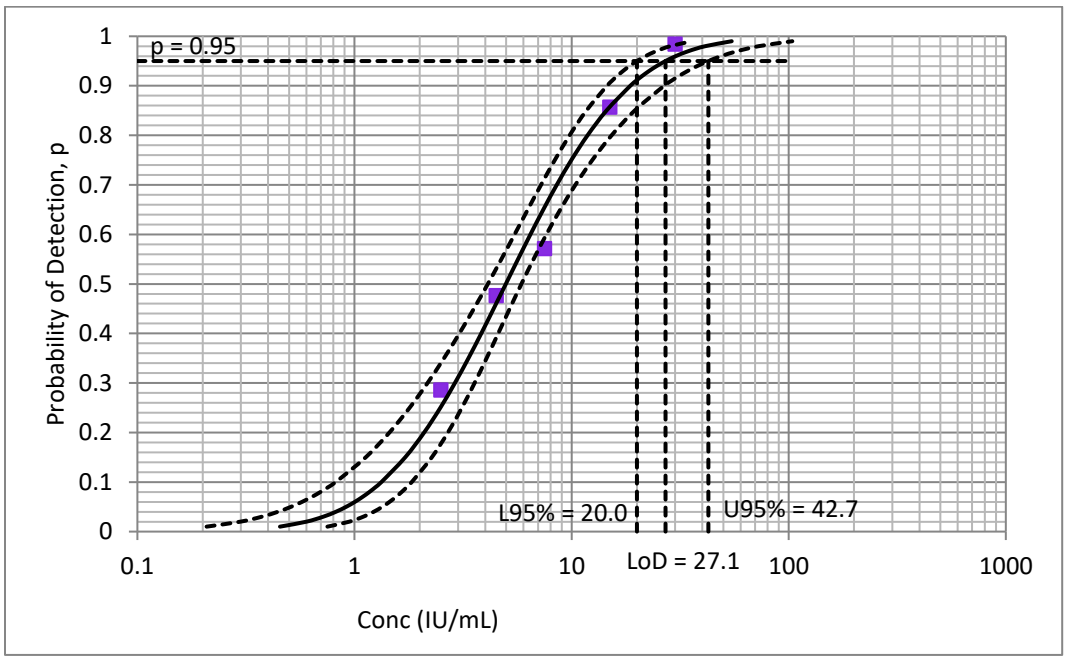

Figure 5. Probit Analysis of LoD Study Data for HIV Blood Screening Test – Computational Experiment with the Outlier Fixed

2.3.2. Example 2

Data collected in LoD study for Influenza B PCR test on POCT instrument are summarized in Table 4.

Table 4: LoD study for Influenza B PCR test on POCT instrument

Concentration	Number of Tests	Number Positive	Detection Rate
0.000125	10	2	0.2
0.000250	10	1	0.1
0.000500	10	6	0.6
0.001000	10	8	0.8
0.002000	10	7	0.7
0.004000	23	23	1.0

Such instruments provide fast test results in doctor's office or at bedside, but instrument has low throughput. For this reason, the reasonable number of tests with such instrument is substantially smaller than that with high throughput systems during performance evaluations. While the minimum number of tests per concentration level in LoD studies recommended in CLSI EP17-A2 guideline for primarily high and medium throughput systems is 20, in this study with the POCT instrument the number of replicates tested at 5 concentration levels was 10, and at 1 concentration level it was 23. Data in this example presented in Table 4 is for one of the reagent lots tested. The concentration units used with this PCR test are TCID50/mL, where TCID50 is the 50% lethality dose. The results of probit analysis are shown in the graph in Figure 6.

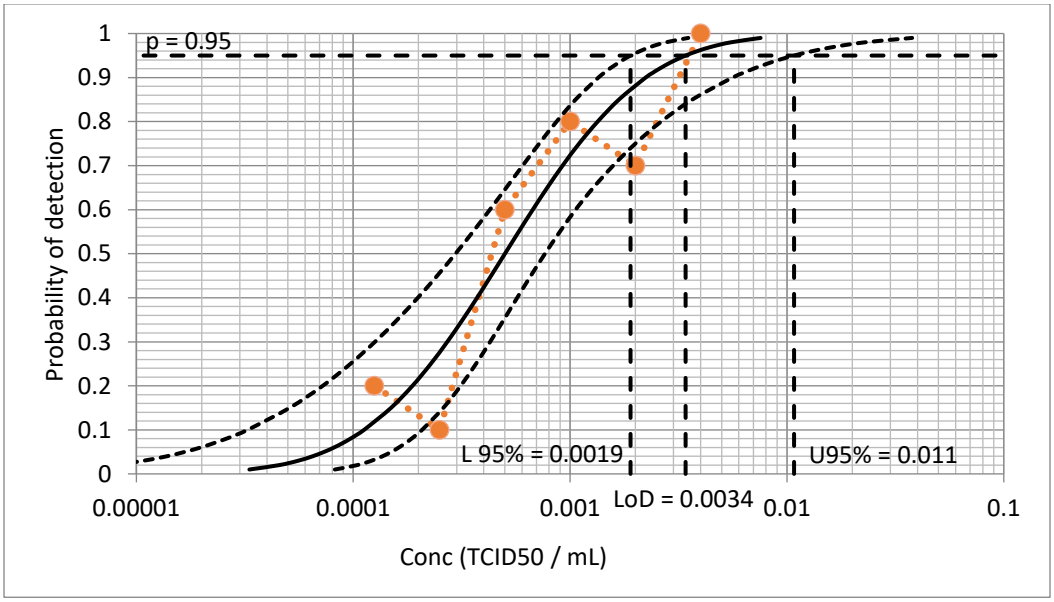

Figure 6. Results of probit analysis of Influenza B LoD evaluation data.

Despite the large variation of the observed detection rate, with small number of tests per concentration level, that even reduces observed detection rates with increases of concentration in two cases, no lack of fit of the probit model to data was reported by SAS® PROC PROBIT procedure because of rather small amount of data collected. The results of LoD estimation with the new MLE method, based on the theoretical model for probability of detection vs. concentration, are shown in the graph in Figure 7.

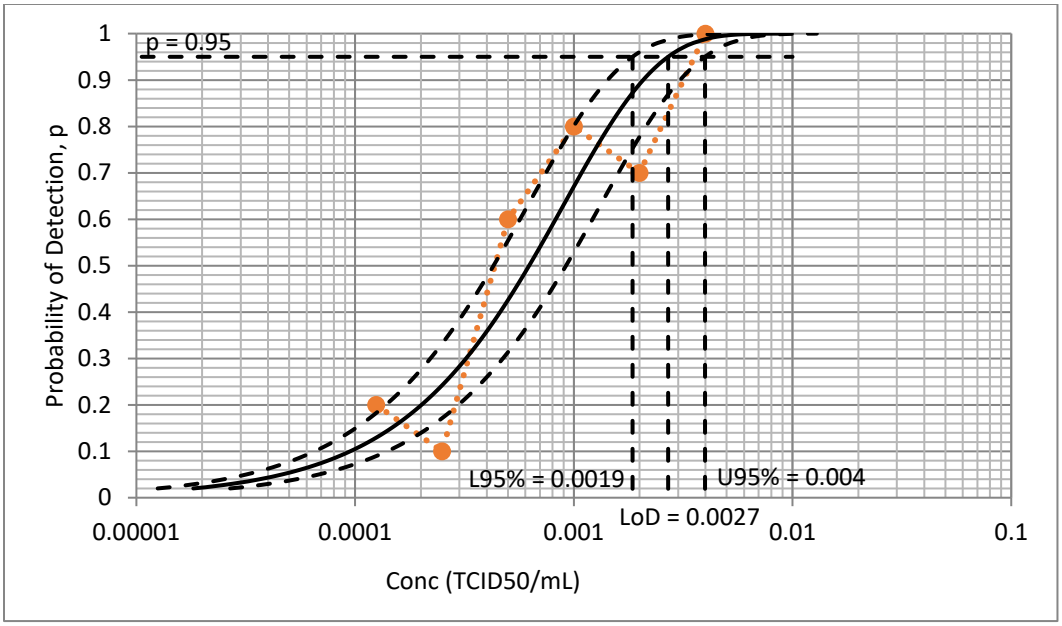

Figure 7. Results of LoD estimation with the new MLE method based on the theoretical model

With the theoretical model validated with data[5], the results of MLE are unbiased. Comparison of the probit LoD estimate with the theoretical model based LoD estimate shows that the probit LoD estimate is biased by 0.0034 – 0.0027 = 0.0007 TCID50/mL, which is 26% of the unbiased LoD estimate. The width of the probit 95% confidence interval is 0.0089 TCID50/mL, or 2.6 times the unbiased LoD estimate. The width of the 95% confidence interval obtained with the theoretical model based MLE method is 0.0021 TCID50/mL or 0.8 of the unbiased estimate of the LoD. The latter is 3.2 times tighter than the former.

2.4. Preliminary estimation of LoD

In many technical performance evaluations of PCR assays samples with concentrations equal to certain fractions and multiples of LoD are used. Those studies may be conducted before the LoD evaluation. Preliminary estimation of LoD can be done with a single concentration level using the following formula[5]:

$$LoD = \frac{\mu \ln(20)}{\ln(1-\hat{p})} \qquad (10)$$

In (10), μ is concentration expressed as mean number of copies per sample volume, $\hat{p}$ is observed detection rate (proportion of positive results). Confidence interval for the LoD can be calculated using confidence limits for the proportion of positive results in (10).

2.5. LoD Claim and Confidence Interval Considerations

CLSI EP17-A2 guideline recommends claiming the largest of LoD estimates obtained with 2 or 3 lots of reagent or LoD estimate obtained with combined data of 4 or more lots of reagent. The recommendation is not substantiated by scientific considerations. It is not clear what is the confidence interval for so claimed LoD. Knowledge of the confidence interval for the claimed LoD is important for the user validation of the claims and for

comparison of the PCR assays offered by different manufacturers. Treating LoD estimate as a normally distributed random variable, mean of the LoD estimates obtained with several lots of reagent could be claimed as the PCR assay performance characteristic. The limits [*L*, *U*] of two-sided (1 - α) confidence interval are then calculated as:

$$\left[L,U\right]=\overline{LoD}\mp t\left(1-\alpha/2,df\right)\cdot s \qquad (11)$$

In (11), *s* is the standard deviation estimate of the LoD estimates; *t* (1 – α/2, *df*) is 1 – α quantile of *t* distribution; *df* is the number of degrees of freedom of *s*. The most reliable estimates of *s* and *df* would be obtained from data that include two LoD estimates for each lot of reagent and at least 3 lots of reagent. Then *s* would be calculated as total standard deviation equal to the square root of the sum of the between lots and within lots variance components, and *df* as the number of degrees of freedom of the total variance.

3. Considerations on Planning and Conducting LoD Studies

For the most efficient use of the LoD study data, there should be one panel member with zero concentration and 0% detection rate for confirmation of *LoB* = 0, and one panel member with concentration close to 100% detection rate. Testing at least 3 panel members with observed detection rates greater than 0% and smaller than 100% is recommended in CLSI EP17-A2 guideline. In practice, there are cases when because of lack of knowledge of the new PCR assay being developed there are more than 1 concentration level with observed detection rates 0% and 100% each. Testing such samples increases costs of the LoD evaluations without adding useful information. At the same time, the number of concentration levels with observed detection rates greater 0% and smaller than 100% can be less than 3, making estimation, particularly when *LoD* and *v* have to be estimated, impossible and requiring testing additional concentration levels.

To do the LoD evaluation in the most economical way, the concentrations of the panel members can be established with sequential testing approach. The first concentration to be tested then is set to a level that is expected to produce detection rate closer to the middle of 0% - 100% interval. This is much easier to achieve than to determine all the concentrations of the positive samples within that interval before the beginning of the study. From the so tested concentration level, rough preliminary LoD estimate is calculated with formula (10). Depending on what fraction of preliminary LoD estimate the first concentration level tested is, the second concentration level is set to smaller or larger fraction of that LoD estimate. With detection rates observed at two concentration levels, an improved estimate of LoD is obtained with the new method, and one more concentration level is planned, and so on. The testing can be stopped when at least 2 concentration levels tested have detection rates below 95% corresponding to LoD, one concentration level has detection rate close to 95%, and one concentration level has detection rate above 95%. Such sequential planning and conducting LoD evaluation study provides for efficient use of data for estimation the LoD.

4. Special Cases of LoD Evaluation

4.1 Detection with More Than One DNA/RNA Region

Viruses mutate. Some regions of their genetic material are more stable, and those are often chosen for detection. Even this does not guarantee that the chosen region will not mutate in each copy of the virus under consideration. To increase the reliability of detection of

HIV-1M and HCV viruses, two regions are used for detection in Roche PCR tests. In case of HIV, two sets of primers and probes binding two most stable regions of the gene are used for detection. In case of HCV, the primers are designed for a longer sequence, and two probes, binding to two regions on that sequence are used for detection.

With such PCR tests, three types of the virus, (1) having one region not mutated, (2) having the other region not mutated, and (3) having both regions not mutated, are present in test sample in various proportions. The LoDs for each of two regions can be either equal or unequal. So, sample-specific LoD of such PCR test is dependent on the LoDs of the virus regions that are not mutated and the proportions of the above three types of virus in a sample being tested. We assume that there are no undetectable viruses with both regions mutated present in test samples. It is desirable to have equal LoDs for two regions used for detection, and efforts are made during development of such PCR test to have these LoDs equal. In case of equal LoDs of the two regions, the LoD for both regions not mutated is half the LoD for any one region not mutated present in test sample. With various proportions of the three types of the virus present in the sample, the sample-specific LoD is between the smallest when all viruses have two regions not mutated and the largest when all the viruses present in the sample have one region not mutated of the type with the largest LoD. So, viral test increased reliability of detection comes at the price of increased uncertainty of the LoD. It is obvious that higher reliability of detection of infectious virus fully justifies such increase of uncertainty of the LoD. It is suggested here that manufacturer claims the largest LoD across the LoDs for individual regions of the viral target nucleic acid sequence. For that, LoD has to be evaluated using samples with each type of the virus having one of the regions used for detection not mutated, the other mutated, and vice versa.

4.2 Detection of Microorganisms having Multiple Copies of DNA/RNA

An example of microorganism having multiple copies of DNA/RHA is Babesia parasite. Different types of individual Babesia parasite have from several hundred to several thousand copies of its genome. The desired number of copies of the organism can be picked up under the microscope and inserted in the plasma samples to be tested in LoD study. So, the number of copies of the organism in the samples tested is exact. PCR testing is done after lysing the organisms in the samples, and the number of copies per organism varies. With one copy of target nucleic acid detectable and multiple copies present in each organism, when expressing concentration as number of organisms per sample volume, the LoD can be a small fraction of the number of organisms required for 95% detection rate. This result is obtained with artificially prepared samples.

In reality, only whole organisms are present in clinical samples. This means that the real LoD in such case equals to the mean number of organisms per sample volume that provides for at least 1 copy of the organism in a randomly drawn clinical sample with probability 95%. This makes the LoD equal to $\ln(20) \approx 3$ copies, on average, per test sample volume. This follows from solving the equation:

$$p = 1 - e^{-\mu\theta} \qquad (5,b)$$

with $p = 0.95$ and $\theta = 1$, for $\mu = LoD$. The solution is $\mu = LoD = \ln(20) \approx 3$. The extraction efficiency, θ, is always less than 1, but with several hundred to several thousand copies of DNA/RNA per organism, there is always at least 1 copy extracted, amplified and detected per organism. Therefore, a single organism is always detectable even with not all DNA/RNA copies extracted from it. For this reason, the result is the same as with $\theta = 1$.

The above means that with microorganisms having large numbers of DNA/RNA copies, the LoD is the concentration corresponding to mean number of organisms per sample volume equal 3.

5. Conclusions

Limit of detection, defined as the target nucleic acid concentration corresponding to 95% probability of detection, is one of the most important performance characteristics of in-vitro diagnostics PCR assays. It is estimated with probit analysis of observed detection rate vs. concentration data following recommendations of CLSI EP17-A2 guideline. The probit analysis fits an empirical model to data and provides an estimate of LoD along with the confidence limits. Use of empirical model leads to biased estimation, lack of fit problems, wide or no confidence intervals and inability to estimate the minimum number of nucleic acid copies required for detection.

The new method described in this paper provides for maximum likelihood estimate of the LoD and the minimum number of copies required for detection along with their confidence intervals based on theoretical math model for probability of detection vs. concentration. With the math model validated with large amount of data, the new method provides for unbiased estimates of LoD and several times tighter confidence intervals compared to the probit analysis. This allows for acceptable LoD estimates with smaller amount of data and practically no retesting.

Also, consideration was given to (i) planning and conducting LoD studies more efficiently with sequential determination of the concentrations to be tested, (ii) increased LoD uncertainty in case of using more than one DNA/RNA region for detection, and (iii) in case the PCR assay is for a microorganism with large number of copies of DNA/RNA, the LoD is 3 microorganisms per sample volume, on average.

Acknowledgements

The author is grateful to his Roche Molecular Solutions, Pleasanton, CA, USA and Rotkreuz, Switzerland, colleagues, Jesse Canchola, John Duncan, Dmitriy Kosarikov, Kyle Lu, John Niemiec, Yoshi Ohhashi, Matthias Rullkoetter, Michael Santulli, Abha Sharma, Shaowu Tang and many others for their help in deeper understanding of PCR, encouragement and support in development of the new method of LoD evaluation in molecular diagnostics, and to the Maplesoft, Inc., Waterloo, Canada, Staff for their help in implementation of the new method in user-friendly software. Many thanks also go to the colleagues on the CLSI EP17-A2 Document Development Committee, and particularly to James Pierson-Perry, Paul Durham and Marina Kondratovich, for many discussions of the LoD concepts and methods of its evaluation and verification.

References

1. Linnet, K., Kondratovich, M., Partly Nonparametric Approach to Determining the Limit of Detection, Clinical Chemistry 50:4, pp. 732-740 (2004).
2. Protocols for Determination of Limits of Detection and Limits of Quantitation, Approved Guideline. CLSI Document EP17-A, Wayne, PA, USA (2004).
3. Evaluation of Detection Capability for Clinical Laboratory Procedures, Approved Guideline, Second Edition. CLSI Document EP17-A2, Wayne, PA, USA (2012).
4. Vaks, J.E., Hemyari, P., Rullkoetter, M., Santulli, M.J., Schoenbrunner, N., Verification of Claimed Limit of Detection in Molecular Diagnostics, Journal of Applied Laboratory Medicine, Nov. 2016, pp. 260-270.
5. Vaks, J.E., Probability of Detection of Target Nucleic Acid Sequence with PCR Assay in Molecular Diagnostics: Math Model Derivation, Validation and Applications, Proceedings of Joint Statistical Meetings, pp. 2460-2476 (2017).
6. Pawitan, Y., In All Likelihood, Statistical Modelling and Inference Using Likelihood, Clarence Press, Oxford, 2013.
7. Maple® 2016, Mathematical software by Maplesoft, Inc., Waterloo, Canada.