ABSTRACT

The Limit of Detection (LoD) is defined as the lowest concentration or amount of material, target or analyte that is consistently detectable (for PCR quantitative studies, in at least 95% of the samples tested)\(^1\). In practice, the estimation of the LoD uses a parametric curve fit to a set of panel member (PM1, PM2, PM3, etc.) data where the responses are binary. Typically, the parametric curve fit to the percent detection levels takes on the form of a probit or logistic distribution. The LoD_Est SAS Macro (Canchola & Hemyari, SAS Global Forum 2016), using the SAS PROBIT procedure as the main engine, is used to fit such a parametric curves (Vaks, 2017 and 2018). The rarely used but preferred method uses the method of maximum likelihood (ML) to estimate the LoD assuming one detectable copy of template. We introduce the LOD_MLE SAS macro that maximizes the log likelihood function and returns the ML estimate (MLE) of the LoD along with its 95% confidence interval (CI). In addition, the macro returns the percent detection table with associated 95% exact (Clopper-Pearson) confidence intervals for the hit rates at each level.

INTRODUCTION AND BACKGROUND

In a PCR assay, analytical sensitivity is defined as the lowest amount of analyte one can reliably detect in, typically, 95% of samples tested. This is important in the diagnosis and monitoring of target analytes, for example, Hepatitis B and C (HBV and HCV, resp.) or the Human Immunodeficiency Virus (HIV), or other analytes of interest. For example, the analytical sensitivity of the assay test for the HCV virus in the blood stream can be used to diagnose disease in a patient who is then placed on a drug treatment regimen. With the progression of time, the viral load of the patient is monitored regularly to ensure the treatment is efficacious.

For quantitative assays, the analytical sensitivity is measured by the limit of detection (LoD) - sometimes called the lower limit of detection (LLoD). However, from here on, we use the more common LoD nomenclature and “sensitivity” to mean “analytical sensitivity”, as it may relate to a quantitative assay, throughout this document. This to differentiate this analysis from the qualitative analysis involving a 2x2 table where clinical sensitivity is well defined using concordance measures.

A typical experiment for finding the LoD is given in Canchola and Hemyari (2016) and is reviewed here. Basically, an experiment is performed in order to collect information about the lower end of the quantitative assay. Several levels are targeted at the lower portion of the assay range, near where the investigator believes the LoD may be, but where it may not be linear on the log\(_{10}\) response, in order to fit a parametric model curve to obtain an estimate of the LoD where the curve crosses 95% detection or hit rate (Figure 1).

Canchola and Hemyari (2016) give an operational description for the process of finding the LoD. That is, the levels or panels in the experiment are chosen to include at least one panel member at 100% detection, another to anchor the parametric curve at the bottom end (not including zero) with the remaining three or more levels targeting the region where one believes the LoD might be located. Table 1, below, shows an example of six panel levels (not including zero) with the top end anchored by an HCV RNA assay level at 50 IU/mL and one anchor at the bottom at 2.5 IU/mL. There are four levels in between both the top and bottom ends: 25, 15, 10 and 5 IU/mL.

\(^{1}\) This definition is used in the context of PCR quantitative studies, but similar definitions apply in other diagnostic assays.
THEORY

Maximum likelihood (ML) estimation, due primarily to Fisher (1950), is a statistical method or framework for maximizing the information in the data for estimating the parameters of a statistical model. Vaks (2017, 2018) first developed the theory as applied to PCR assays. In our case, assume that $\mu_i$ is the nominal or observed viral concentration for a particular $i$th level in a LoD study. Next, assume that for each $\mu_i$, we can model the hit rate using the following expression:

$$p_i = 1 - e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20)}$$

(1)

where LoD is the theoretical Limit of Detection parameter and $\ln(*)$ is the natural logarithm. It follows that the probability of observing $n_i$ “positives” from $N_i$ replicates at each level $i$ is:

$$P(n_i; N_i, p_i) = \left(\frac{N_i}{n_i}\right) \cdot \left(1 - e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20)}\right)^{n_i} \cdot e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20) \cdot (N_i-n_i)}$$

(2)

Maximum Likelihood theory then defines the univariate likelihood function for the LoD as formula (3) and (4).

$$L(\text{LoD}; \{\mu_i, n_i, N_i \}_{i=1}^l) = \prod_{i=1}^l P(n_i; N_i, p_i)$$

(3)

$$= \prod_{i=1}^l \left(\frac{N_i}{n_i}\right) \cdot \left(1 - e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20)}\right)^{n_i} \cdot e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20) \cdot (N_i-n_i)}$$

(4)

Now, typically, the likelihood is easier to work with on the logarithmic scale so that applying the logarithm to (4) produces the so-called "log-likelihood" function:

$$\ell(\text{LoD}; \{\mu_i, n_i, N_i \}_{i=1}^l) \propto \sum_{i=1}^l \ln[P(n_i; N_i, p_i)]$$

$$\propto \sum_{i=1}^l \left[n_i \cdot \ln \left(1 - e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20)}\right) - \frac{(N_i-n_i)\cdot \mu_i \cdot \ln(20)}{\text{LoD}}\right]$$

(5)

(6)

Then the Maximum Likelihood Estimate (MLE), denoted $\hat{\text{LoD}}$, of the LoD parameter is obtained by setting the first derivative of (6) with respect to LoD to zero as in equation (7).

$$\frac{d}{d(\text{LoD})} \ell(\text{LoD}; \{\mu_i, n_i, N_i \}_{i=1}^l) \bigg|_{\text{LoD}=\hat{\text{LoD}}} = 0$$

(7)

so that $\hat{\text{LoD}}$ is a unique positive root of the expression:

$$\sum_{i=1}^l \left[\frac{n_i \cdot \mu_i}{1 - e^{-\left(\frac{\mu_i}{\text{LoD}}\right)\cdot \ln(20)}}\right] = \sum_{i=1}^l [N_i \cdot \mu_i]$$

(8)

which can only be solved numerically.
We can then use the Fisher information [i.e., the second derivative of the log-likelihood function as shown in formula (9)] to obtain the asymptotic two-sided 95% confidence interval for LoD. Denote \( I^{(2)}(\widetilde{\text{LoD}}) \) as the second derivative of the log-likelihood function in (6) with respect to LoD and evaluated at \( \text{LoD} = \widetilde{\text{LoD}} \). The Fisher information is then derived as shown in (9).

\[
I^{(2)}(\text{LoD}) = -\left( \frac{\ln(20)}{\text{LoD}^2} \right)^2 \sum_{i=1}^{n} \left[ \frac{n_i \mu_i^2 e^{-\left( \frac{\mu_i}{\text{LoD}} \right) \ln(20)}}{\left( 1 - e^{-\left( \frac{\mu_i}{\text{LoD}} \right) \ln(20)} \right)^2} \right] 
\]

(9)

Then it follows that a 95% confidence interval (CI) for LoD is:

\[
\text{LoD} = \widetilde{\text{LoD}} \pm 1.96 \cdot \frac{1}{\sqrt{I^{(2)}(\text{LoD})}}
\]

(10)

**PROGRAMMING**

Figure 1 shows how we use the SAS Interactive Matrix Language (SAS/IML) to find the MLE of the LoD (i.e., \( \widetilde{\text{LoD}} \)) by first defining the log-likelihood function expression in (8) and then using the SAS “froot()” function for finding the root of (8) within the stated “bounds”. Finally, we use the Fisher Information in (9) to find the 95% confidence interval in (10).

```sas
proc iml;
  use DataIn;
  read all var {Titer Hits TotalTests} into Z;
  close DataIn;
  mu = Z[,1]; n = Z[,2]; NN = Z[,3];
  * Define log-likelihood function;
  start Fun(x) global(mu,n,NN);
    y1 = ( n # mu ) / (1 - exp( -mu * log(20) / x ) );
    y2 = NN * mu ;
    f = sum(y1) - sum(y2);
    return(f);
  finish Fun;
  * Define vector and bounds to contain LoD estimate;
  LoD_MLE = j(nrow(Z),1,1);
  bounds  = {0.0001 & upperBound};
  * Maximize the likelihood function;
  LoD_MLE = froot("Fun", bounds);
  * Define Fisher information (i.e., 2nd derivative) for confidence interval estimation;
  leftfish = ( log(20) / LoD_MLE**2 )**2;
  ritefish = sum( (n # mu##2 # exp(-mu * log(20) / LoD_MLE)) / (1 - exp(-mu * log(20) / LoD_MLE))##2 )
              / (1 - exp(-mu * log(20) / LoD_MLE))##2;
  fisher = leftfish # ritefish;
  LCL95  = LoD_MLE - 1.96 / sqrt(fisher);
  UCL95  = LoD_MLE + 1.96 / sqrt(fisher);
  call symputx("LoD_MLE", round(LoD_MLE,0.1) );
  call symputx("LCL95", round(LCL95,0.1) );
  call symputx("UCL95", round(UCL95,0.1) );
run;
quit;
```

**Figure 1.** SAS/IML code for LoD_MLE SAS Macro.
METHODS

INTRODUCING THE LOD_MLE SAS MACRO

LoD_MLE is a SAS macro for obtaining the maximum likelihood estimate (MLE) from hit-rate (or percent detection) data. The macro uses the SAS/IML as the main computation engine.

There are four main steps to successfully run the LoD_MLE SAS macro as follows.

**Step 1:** Enter the following (from Table 1):

a. Target descriptive (1)
b. Level concentrations (2)
c. Number of Positive Hit Counts (3)
d. Total Test Counts (4)

```sas
data LoD ;
input Target $3. Titer Hits TotalTests @ ;
Log10_Titer = log10(Titer) ;
Obs_HitRate = Hits / TotalTests ;
Obs_HitPcnt = Obs_HitRate * 100 ;
datalines ;
HCV 50  252 252
HCV 25  251 252
HCV 15  246 251
HCV 10  233 252
HCV 5   183 252
HCV 2.5 124 251 ;
run ;
```

**Step 2:** Load the LoD_MLE.sas SAS macro into your session.

```sas
%include "C:sas\macros\LoD\LoD_MLE.sas" ;
```

**Step 3:** Enter LoD_MLE SAS macro inputs.

```sas
%LOD_MLE (dataIn = LoD , study = CTM_1 , Target = HCV , subset = Target = "HCV" , Units = IU/mL , alpha = 0.05 , sided = 2 , upperBound = 200 , TableNo = 1 , ODStype = RTF , RTFout = C:\HCVstudy\RTFout\HCV_LoD.RTF ) ;
```

**DISCUSSION**

We have seen that the LoD_MLE SAS macro is simple to use when a Limit of Detection table, with information-rich details, is desired. The reader should note that there are some occasions where the MLE will not be defined and so the algorithm may not converge.
CONCLUSION

The LoD_MLE SAS macro can be used to produce a percent detection (or hit-rate) table with 95% exact (Clopper-Pearson) confidence intervals for the hit rate at each level, including the LoD ML estimate (MLE) and associated 95% confidence interval.
REFERENCES


ACKNOWLEDGMENTS

The authors thank Alison J. Canchola for their valuable input.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the first author at:

Jesse A. Canchola
Roche Molecular Systems, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588
E-mail: Jesse.Canchola@Roche.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.
APPENDIX A: LOD_MLE SAS MACRO CODE

* BEGIN MACRO CODE ;
/* ************************************************* ;
* Macro LoD_MLE
* Calculates the LoD MLE based on hit-rate data inputs
* Includes Percent Detection Table
*
* Macro Inputs (with examples):
*
* datain     = LoD ; * enter data set name ;
* study      = Investig ; * enter study name or investigator name ;
* target     = CMV ; * enter target/virus ;
* subset     = target="CMV" ; * enter "where" statement language ;
* Units      = IU/mL ; * enter units of measurement: IU/mL, copies/mL, cp/mL, etc;
* alpha      = 0.05 ; * significance leve in decimal ;
* sided      = 2 ; * sidedness of confidence interval for percent detection ;
* upperBound = 200 ; * define upper bound of concentration (any high enough number in units of assay) ;
* TableNo    = Z1 ; * Table Number ;
* ODStype    = RTF ; * enter ODS output type: RTF, PDF, HTML, etc. ;
* RTFout     = C:\LoD_&target._&study._&sysdate..&ODStype ; * output &ODStype file ;
*
* Created by: Jesse A. Canchola (JAC)
* Creation Date: 04Sep2015
*
* Updated on: 08-Sep-2015 (JAC)
* Update: 1. Added upperBound option.
* 2. Completed validation.
*
* ************************************************** USAGE ***************
%include "C:\SAS\Macros\LoD\LoD_MLE.sas" ;

* Enter data columns:
* Target : analyte such as CMV, HBV, HCV, HIV
* Titer : numerical concentration of analyte
* Hits : number of detected samples out of TotalTests experiment
* TotalTests : total number of tests for each concentration level of experiment
data LoDdat1 ;
input Target $3. Titer Hits TotalTests @ ; * MUST BE IN THIS ORDER - DO NOT MODIFY ;
Log10_Titer = log10(Titer) ; * DO NOT MODIFY ;
Obs_HitRate = Hits / TotalTests ; * DO NOT MODIFY ;
Obs_HitPcnt = Obs_HitRate * 100 ; * DO NOT MODIFY ;
datalines ;
HCV 50 252 252
HCV 25 251 252
HCV 15 246 251
HCV 10 233 252
HCV 5 183 252
HCV 2.5 124 251
;
run ;

%LET Target  = HCV   ;
%LET Study   = TEST1 ;
%LET ODStype = RTF   ;

options mprint mlogic symbolgen ;
%LOD_MLE(
datain      = LoDdat1 ,
study      = CTM_1 ,
target     = HCV ,
subset     = target = "HCV" ,
alpha = 0.05 ,
sided = 2 ,
ODStype = RTF ,
upperBound = 200 ,
TableNo = Z1 ,
RTFout = C:\LoD_HCV.TEST1.sysdate.RTF
);
Units       = IU/mL    ,
alpha       = 0.05     ,
sided       = 2        ,
upperBound  = 1E9     ,
TableNo     = 1        ,
ODStype     = RTF     ,
RTFout      = C:\LoD\LoD_MLE\Results\LoD&_target._&study._&sysdate..&ODStype
) ;
* ************************************************** */ ;
%MACRO LoD_MLE ( datain      = ,
study       = ,
target      = ,
subset      = ,
Units       = ,
alpha       = ,
sided       = ,
upperBound  = ,
TableNo     = ,
ODStype     = ,
RTFout      =
) ;
options orientation = landscape ;
proc format ;
  picture mltdecaf (multilabel)
    0-9   = '9.9'
    10-99 = '99'
    100-999 = '999'
    1000-9999 = '9999' ;
  picture mltdecbf (multilabel)
    0-9.9  = '9.99'
    10-99.9 = '99.9'
    100-999.9 = '999'
    1000-9999.9 = '9999' ;
  picture mltdeccf (multilabel)
    0-9.99  = '9.999'
    10-99.99 = '99.99'
    100-999.9 = '999'
    1000-9999.9 = '9999' ;
run ;
proc sort data = &DataIn ; by Target descending Titer ; run ;
data DataIn ;
set &DataIn ;
where &subset ;
run ;
ods &ODStype file = "&RTFout" ;
proc sort data = DataIn ; by Target ; run ;
* Nomenclature mapping from S. Tang -R- equivalent function: 
  Titer = mu   Hits  = n   TotalTests = NN ;
proc iml ;
use Datain ;
read all var {Titer Hits TotalTests} into Z ;
close Datain ;
mu = Z[,1] ; n = Z[,2] ; NN = Z[,3] ;
* Define likelihood function (Shaowu Tang, 2015) ;
start Fun(x) global(mu,n,NN) ;
y1 = ( n # mu ) / ( 1 - exp( -mu # log(20) / x ) ) ;
y2 = NN # mu ;
f = sum(y1) - sum(y2) ;
return(f) ;
finish Fun ;

* Define vector and bounds to contain LoD estimate ;
LoD_MLE = J(nrow(Z),1,1) ;
bounds = {0.0001 &upperBound} ;

* Maximize the likelihood function ;
LoD_MLE = froot( "Fun", bounds ) ;

* Define Fisher information (i.e., 2nd derivative) for confidence interval estimation
(Shaowu Tang, 2015) ;
leftfish = ( log(20) / LoD_MLE**2 )**2 ;
rightfish = sum( (n # mu##2 # exp(-mu # log(20) / LoD_MLE) ) / (1 - exp(-mu # log(20) / LoD_MLE))##2 ) ;
fisher = leftfish # rightfish ;
LCL95 = LoD_MLE - 1.96 / sqrt(fisher) ;
UCL95 = LoD_MLE + 1.96 / sqrt(fisher) ;
call symputx("LoD_MLE" , round(LoD_MLE,0.1) ) ;
call symputx("LCL95"   , round(LCL95,0.1)   ) ;
call symputx("UCL95"   , round(UCL95,0.1)   ) ;
run ;
quit ;

* ****************************************** *
* Hit Rate Table                             *
* ****************************************** *
data DataIn1 ;
set DataIn ;

* X-sided 95% Clopper-Pearson CI for hit rates ;
if &sided = 1 then do ;
zalpha = probit( 1 - ( &alpha ) ) ;
end ;
else if &sided = 2 then do ;
zalpha = probit( 1 - ( &alpha / 2 ) ) ;
end ;

n = TotalTests ;
p = Hits / TotalTests ;

** Exact Clopper Pearson ;
x = round( n * p, 0.1 ) ;

* Calculate the lower limit. ;
v1 = 2 * ( n - x + 1 ) ;
v2 = 2 * x ;
if &sided = 1 then do ;
a = 1 - ( &alpha ) ;
end ;
else if &sided = 2 then do ;
a = 1 - ( &alpha / 2 ) ;
end ;

c = ( n - x + 1 ) / x ;
fscore = finv( a, v1, v2 ) ;
exact_lcl = 1 / ( 1 + c * fscore ) ;
* Calculate the upper limit. 
\[ v_{11} = 2 \times (x + 1) ; \]
\[ v_{22} = 2 \times (n - x) ; \]
\[ \text{fscore2} = \text{finv}(a, v_{11}, v_{22}) ; \]
\[ \text{coef2} = \frac{x + 1}{(n-x)} ; \]
\[ \text{numer} = \text{coef2} \times \text{fscore2} ; \]
\[ \text{exact_ucl} = \frac{\text{numer}}{1 + \text{numer}} ; \]
\[ \text{if } \text{exact_ucl} = . \& p = 1 \text{ then } \text{exact_ucl} = 1.0 ; \]
\[ \text{if } \text{exact_ucl} = . \text{ then } \text{exact_ucl} = 9999 ; \]

* converting ci to percent AND HANDLING MISSING DATA IF ANY ;
\[ \text{if } \text{exact_lcl} \neq 9999 \text{ then } \text{exact_lclp} = \text{exact_lcl} \times 100 ; \]
\[ \text{if } \text{exact_ucl} \neq 9999 \text{ then } \text{exact_uclp} = \text{exact_ucl} \times 100 ; \]

* ****************** end Clopper-Pearson CI calculation ******************* ;

drop N p x V1 V2 A Coef Fscore V11 V22 Fscore2 Coef2 Numer ;
run ;

* Proc report ;
options orientation = landscape nodate nonumber ;
ods listing close ;
ods escapechar='~' ;
proc report data = DataIn1 split='*' missing headline headskip spanrows
   style(report)={just=center outputwidth=10 in}
   style(lines) =header{font_size=9pt font_face="Arial" font_weight=medium
   background=transparent just=left}
   style(header)=header{font_size=9pt font_face="Arial" font_weight=bold
   background=transparent just=center}
   style(column)=header{font_size=9pt font_face="Arial" font_weight=medium
   background=transparent just=center} ;
columns Target Titer Log10_Titer Hits TotalTests Obs_HitPcnt Exact_LCLP Exact_UCLP ;
define Target / display "Target" group order=freq
   style(header)={just=center cellwidth=.6 in}
   style(column)={just=center cellwidth=.6 in vjust=center} ;
* if low titers below 1.0, replace the mltdecaf with mltdeccf picture format defined above ;
define Titer / display "Titer (&units)" f=mltdecaf.
   style(header)={just=center cellwidth=.55 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx600 "*/
   cellwidth=.55 in} ;
define Log10_Titer / display "Titer*(log10 &units)" f=7.4
   style(header)={just=center cellwidth=.55 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx600 "*/
   cellwidth=.55 in} ;
define Hits / display "No. Detected" f=6.0
   style(header)={just=center cellwidth=.55 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx600 "*/
   cellwidth=.55 in} ;
define TotalTests / display "Total Tests" f=5.0
   style(header)={just=center cellwidth=.5 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx500 "*/
   cellwidth=.5 in} ;
define Obs_HitPcnt / display "Percent*Detection" f=mltdecbf.
   style(header)={just=center cellwidth=.5 in}
Limit of Detection (LoD) Estimation Using Maximum Likelihood from (Hit) Rate Data: The LoD_MLE SAS Macro

```sas
style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx500 "*/
cellwidth=.5 in} ;
define Exact_LCLP / display "95% Exact LL" f=mltdecbf.
   style(header)={just=center cellwidth=.5 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx500 "*/
cellwidth=.5 in} ;
define Exact_UCLP / display "95% Exact UL" f=mltdecbf.
   style(header)={just=center cellwidth=.5 in}
   style(column)={protectspecialchars=off just=center /*pretext="\qj\tqdec\tx500 "*/
cellwidth=.5 in} ;
compute after _page_ / center style=[font_weight=bold] ;
   line "LoD ML Estimate = &LoD_MLE &Units, 95% CI = ( &LCL95, &UCL95 )" ;
endcomp ;
title1 bold j=center height=12pt f='Times' "Table &TableNo.. Percent Detection" ;
run ;
ods &ODStype close ;
options orientation = portrait ;
%mend LoD_MLE ;
* END OF MACRO CODE ;
```