

Monocyte Activation Test

Logistic regression model of the endotoxin standard curve and statistical tests

1. Introduction

The Monocyte Activation Test (MAT) is used to detect or quantify pyrogenic substances that activate human monocytes.

Pyrogenic substances in pharmaceutical products can induce life-threatening fever reactions after injection into the human body. Therefore, it is a regulatory requirement to test such products for pyrogens to ensure product quality and patient safety.

The MAT has been described as a compendial method for pyrogen detection in the European Pharmacopeia since 2010 (Chapter 2.6.30). It is a test that mimics the human reaction to pyrogens, by using a source of monocytes, which release some interleukins upon activation by pyrogenic substances:

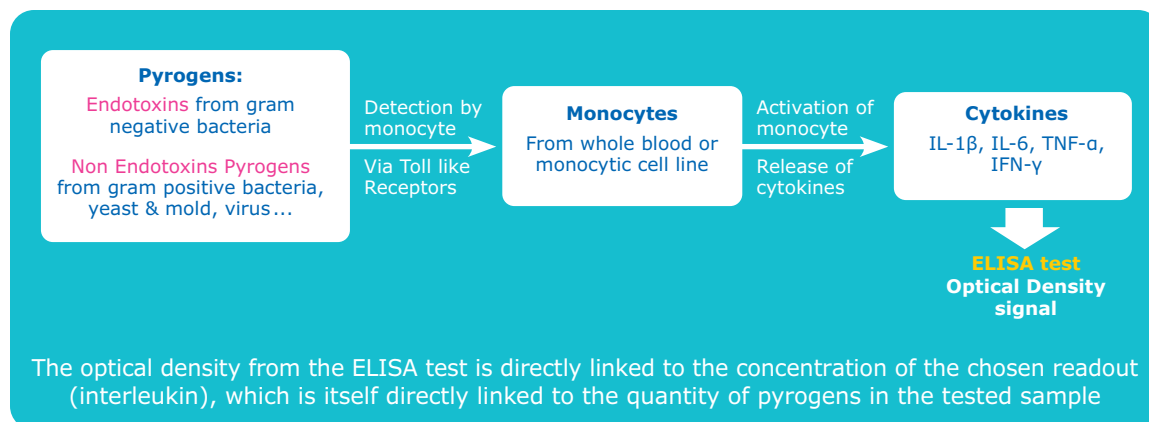


Figure 1: principle of the Monocyte Activation Test

The test requires the use of an endotoxin standard curve for Product Specific Validation (PSV) and for the Method A (quantitative test) described in the European Pharmacopeia. The Optical Density (OD) signal is the result of the biological reaction of monocytes to pyrogens and endotoxins. It is represented in function of the standard endotoxin concentration.

The endotoxin standard curve is then used to convert an OD signal into a pyrogen concentration and allows the quantification of pyrogens or endotoxin equivalent units in the tested sample.

A regression model is needed to modelize properly the response of monocytes to different Reference Standard Endotoxin (RSE) concentrations (Standard curve) in a continuous manner on the range of interest.

We currently provide two MAT solutions:

- The PyroDetect System uses cryopreserved whole blood as a source of monocytes, and the interleukin-1 β as a readout for the ELISA;
- The PyroMAT[®] System uses the Mono-Mac-6 (MM6) cell line as a source of monocytes, and the interleukin-6 as a source of monocytes for the ELISA.

The use of different sources of monocytes (blood vs cell line) leads to a different reactivity profile that requires the use of different regression models for each of these solutions.

To analyze the data generated by PyroDetect or PyroMAT[®] System, the Gen5[™] software¹ (Biotek, 2017) is used to carry out all the modelling and statistical analysis. Gen5[™] offers extensive curve fit methods for quantitative and qualitative analysis including 4- and 5-parameter logistic curve fits. We have developed specific protocols to perform automatically all the calculations and conclusions, according to Method A, B or C described in the European Pharmacopeia 07/2017:20630 (MAT Chapter).

The purpose of this white paper is to describe how the response of monocytes (either from cryoblood or MM6 cell line) to endotoxins has been modeled in order to fulfill the EP requirements. The statistical analysis performed in our data analysis tool supported by Gen5[™] Software is also described. The handling of outliers is described at the end of the document. This white paper was written with the support of a statistician from the Institute of Advanced Mathematical Research (UMR 7501) of the University of Strasbourg and the Centre National de la Recherche Scientifique (CNRS), who also provided support along the development of our data analysis tool.

Design of experiments

In order to get an efficient dose-response modelling, the following design of experiments is recommended by the EP chapter 07/2017:20630.

“Using the standard endotoxin solution, prepare at least 4 endotoxin concentrations to generate the standard curve. Perform the test using at least 4 replicates of each concentration of standard endotoxin.”

Both the design of experiments for PyroDetect and PyroMAT[®] Systems satisfy that criterion:

- For PyroDetect System, the experiment is made of 4 replicates of 6 RSE concentrations, including a blank.
- For PyroMAT[®] System, the experiment is made of 4 replicates of 8 RSE concentrations, including a blank.

What is a good curve model?

A good curve model should possess three properties²:

1. Curve model must do a good job at approximating the true curve. If the curve model does not do this, then whatever the number of replicates may be, the model will still be differing from the mean observed values because the model is wrongly specified: for instance, using a straight line to deal with nonlinear relationships.
2. A good curve model must be able to average out as much of the random variation in the assay data as possible to produce concentration estimates that are distorted by pure error as little as possible.
3. A good curve model must be able to predict concentrations well at points between the standard points and not just at the fitted data points.

¹ <https://www.biotek.com/products/software-robotics-software/gen5-microplate-reader-and-imager-software/features/#1>

² The five-parameter logistic: A characterization / P.G. Gottschalk, J.R. Dunn / Anal. Biochem. 343 (2005) 54–65

What is curve fitting and how can it be performed?

Once a curve model is chosen, one still needs to select the “best” curve among all the possible curves. This is called the fitting process and it aims to adjust the free parameters of the function until these parameters approximate the assay’s true curve better than any other parameter set. From a statistical point of view, this is called maximum likelihood estimation.

Once we assume that the error of the model, i.e. the residual discrepancies (or just residuals) between the predicted values (\check{Y}_i) using the fitted curve and the actual observed values (Y_i), is –approximately– Gaussian, then fitting a curve using maximum likelihood is equivalent to finding the curve whose parameters generate the smallest weighted sum of squared errors (SSE).

The weighted sum of squared errors (SSE) is the sum of all of the squares of the residuals ($Y_i - \check{Y}_i$) –the differences between the observed standard responses and the response predicted by the curve model–, weighted by the inverse variance ($w_i = 1/\text{Var}(Y_i)$) of the responses at that concentration³:

$$\text{SSE} = \sum_{i=1}^n w_i (Y_i - \check{Y}_i)^2$$

Why a curve would not perfectly fit the data?

In any regression, regardless of what curve model is used, there are two reasons why the curve will not fit the data perfectly:

1. Pure error: the presence of random variation in the data. It can be reduced by increasing the number of replicates of each standard concentration.
2. Lack-of-fit error: the curve model may not approximate the true curve very well. It cannot be reduced by increasing the number of standard replicates.

Two criteria for assessing dose-response modelling

According to the EP chapter 07/2017:20630 there are 2 acceptance criteria for the endotoxin standard curve: fit and effect.

1. To use a model, it is necessary to check if it correctly fits the data. For instance, for linear regression, the EP recommends that:

“The regression of responses on log dose must not deviate significantly from linearity ($p > 0.05$).”

As our model differs from linear regression, we have to adapt our goodness of fit procedure while still using a 5% level of significance.

2. Then, if the model fits the data, one wants to assess whether there is or not a significant effect of the doses on the optical densities. The EP recommends using a 1% level of significance:

“The regression of responses (appropriately transformed if necessary) on log dose shall be statistically significant ($p < 0.01$)”

³ According to regression theory, the weights should be set equal to the inverse variance of the responses at that concentration: weighting the squared errors in this way causes the fitting procedure to adjust the curve to be tighter around those standard responses with the smallest variance (error).

Data modelling using sigmoid dose-response curves

Figure 2 represents data generated with the PyroDetect System. The endotoxin concentrations are represented on the horizontal axis. The responses are indicated on the vertical axis. The individual responses (OD values) to each endotoxin concentrations are indicated with black dots.

As can be seen on **Figure 3**, the goodness of fit of a linear model is very poor. Monocytes do not respond in a linear way to the increasing pyrogens concentrations, that's why other models must be investigated.

According to The European Pharmacopoeia chapter 07/2016:50300 corrected 9.2, a sigmoid⁴ dose-response curve is suitable for some immunoassays. Such a model, the four-parameter logistic one, is fitted to the data on **Figure 4** and seems relevant.

In the chapter 07/2016:50300 corrected 9.2 the EP provides some guidance on selected statistical questions that are raised by the use of non-linear dose-response curves and their extensions.

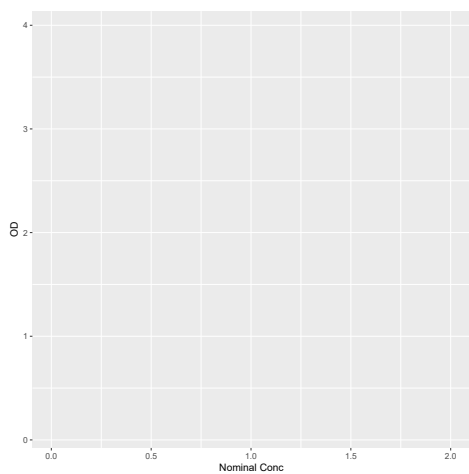


Figure 2: Raw data

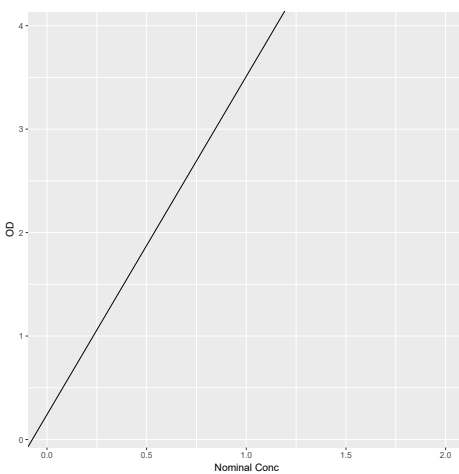


Figure 3: Linear fit

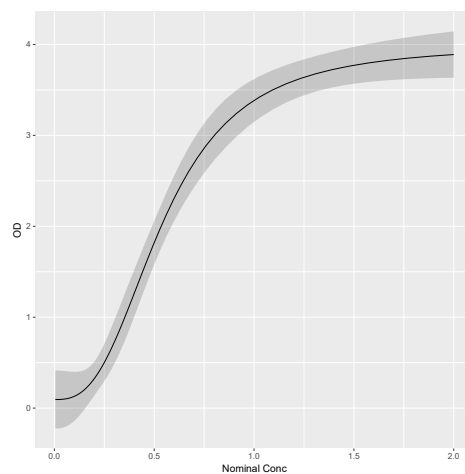


Figure 4: Sigmoid fit

- 1. [...] "However, models based on functions giving other sigmoid curves may also be used. Models incorporating additional asymmetry parameters have been suggested.**
- 2. Heterogeneity of variance is common when responses cover a wide range. If the analysis ignores the heterogeneity, interpretation of results may not be correct and estimates may be biased. Use of the reciprocal of the error variances as weights is unlikely to be reliable with limited numbers of replicates. It may be appropriate to estimate a function which relates variance to mean response.**
- 3. The statistical curve-fitting procedures may give different estimates depending on assumptions made about the homogeneity of the variance and on the range of responses used.**
- 4. In principle, equality of upper and lower response limits for the different preparations included in an assay can be directly tested in each assay. However, interpretation of the results of these tests may not be straightforward." [...]**

As we will quickly review now, the PyroDetect and PyroMAT[®] datasets feature some of the specificities highlighted by the EP. To provide the users with an efficient statistical methodology, a special care was devoted to address all these specificities:

1. For the PyroDetect datasets, the logistic four-parameter model will fit to most of the data. Yet, as we will show later, PyroMAT[®] datasets feature some additional asymmetry and will require the use of extended non-linear dose-response curves such as the five-parameter logistic function that incorporate an additional asymmetry parameter.
2. Heterogeneity of variances is observed for PyroMAT[®] dataset.
3. A comprehensive modelling of the variances leads to an improvement of the curve fitting procedures.
4. A dedicated statistical test to detect the effect of the dose on the optical densities is needed.

⁴ A non-linear model with an S-shape curve

2. Four-parameter logistic model to describe PyroDetect dataset

PyroDetect dataset

Figure 5 plots a dataset generated with the PyroDetect System. It exhibits non-linearity without asymmetry: monocytes from whole blood reach a reactivity plateau naturally before reaching the saturation of the absorbance reader. As a consequence, for PyroDetect data, the 4-parameters logistics regression model is the most suitable.

We will use this nonlinear model for the PyroDetect datasets and assume homogeneous variances in the fitting process and significance analysis.

Design of the assay

In order to fit the nonlinear logistic 4 parameters model, one needs to get observed optical density (OD) values before and after the turning point of the S-curve. It is recommended that:

- at least one standard is not far from each asymptote
- at least 2 standards fall within the linear area of the curve, on either side of the inflection point

The total number of standards was chosen in order to get enough measurements to estimate the four parameters of the model (4 standards required and 5 recommended for the calculation of the goodness of fit statistics).

Yet, the actual values for which the turning point appears depend on each assay. As a consequence a wider range composed of a blank and five concentrations was used to design experiments with the PyroDetect System:

0 EU/mL, 0.125 EU/mL, 0.25 EU/mL, 0.5 EU/mL, 1 EU/mL, 2 EU/mL.

The aim is to successfully find the turning point of the curve at each assay.

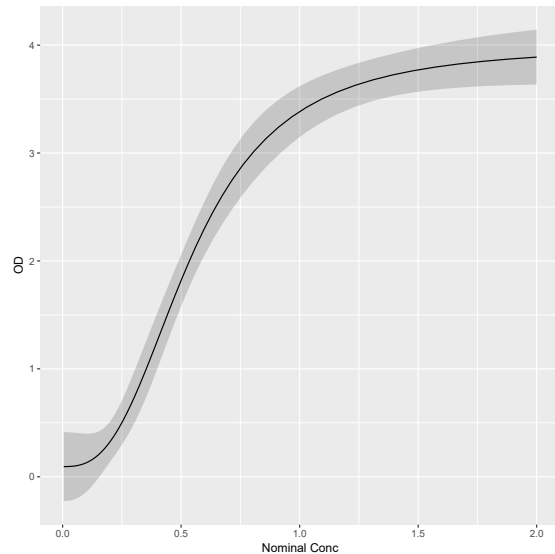


Figure 5: PyroDetect dataset with a 4P logistic fit

Features of the 4-parameters regression model

Definition

Here is the equation of the log-logistic 4 parameters model:

$$f(x) = \frac{a-d}{1 + (\frac{x}{c})^b} + d$$

where a, b, d can be any number and c is a positive number.

Symmetric model

The model is symmetric, which means that if you swap the a and d values and reverse the sign of b, the curve remains unchanged. Hence, one can assume that the slope b is positive without loss of generality or fitting performance. In the following of this white paper, we assume b positive ($b > 0$).

Four parameters

Even though the logistic 4 parameters model seems complicated, it can be easily grasped since its parameters have plain interpretation (remember that $b > 0$), see:

- a is the lower intercept (plateau) at the left of the curve (Y unit: OD)
- b is the unitless slope factor or Hill slope.
- c is the ED₅₀ (is the concentration that give half-maximal effects), in the same units as X. It also corresponds to the inflexion point (point of greatest slope) and to the point of symmetry.
- d is the upper intercept (plateau) at the right of the curve (Y unit: OD)

The effect of the four parameters on the shape of the curve are plotted on **Figure 6**, **Figure 7**, **Figure 8** and **Figure 9**.

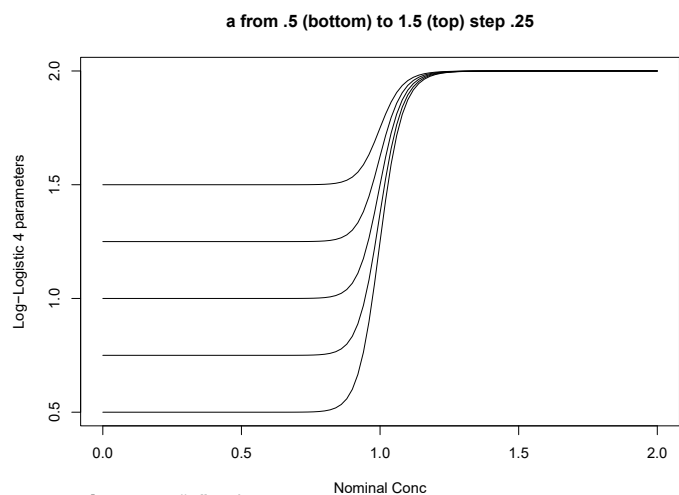


Figure 6: "a" values

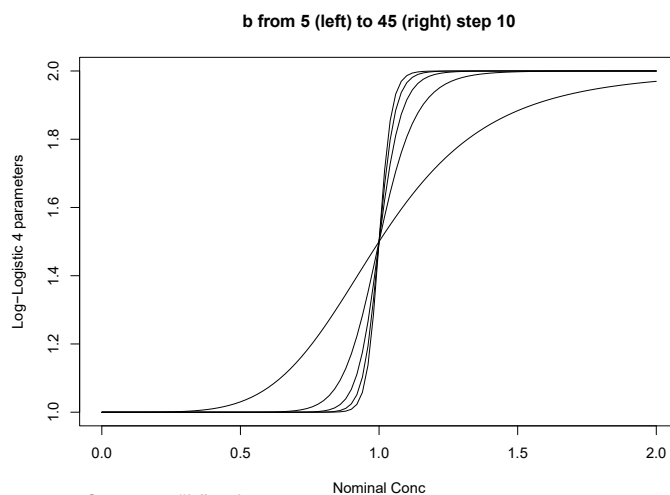


Figure 7: "b" values

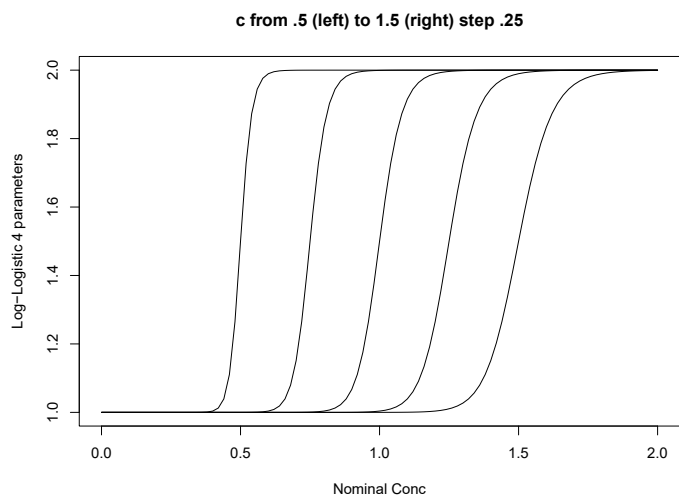


Figure 8: "c" values

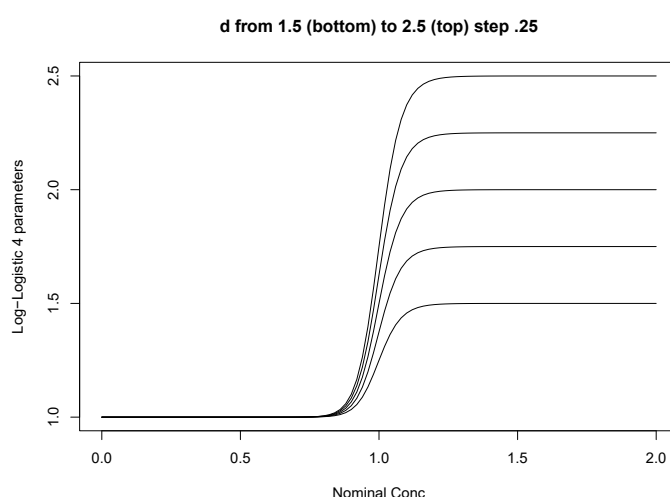


Figure 9: "d" values

Statistical significance testing to fulfill EP requirement

According to the EP, we need to assess the goodness of fit test at a 5% level, then search for a significant dose effect at a 1% level.

Goodness of fit

The classical lack-of-fit test is applied (Bates and Watts, 1988⁵). This test compares the dose-response model to a more general analysis of variance (ANOVA) model using an approximate F-test. It assumes that the responses of the individual standard concentrations are approximately normally distributed and this distribution allows us to determine the probability that curve fits having a particular value of lack of fit, or worse, will occur. The fit F-Prob test (Fisher Snedecor) is the probability of goodness of fit. It requires replicates but no weighting.

A significant test result is alarming and means that the model does not fit the data and that it must be changed.

Effect of dose

The dose effect is assessed using three different criteria:

1. Parameter a must be lower than d (the effect is increasing)
2. The curve must be monotonic: mean OD value for 0.125 EU/mL < mean OD value for 0.250 EU/mL < mean OD value for 0.500 EU/mL < mean OD value for 1.000 EU/mL,
3. The upper intercept (parameter d) must be significantly different from 0 (OD) at the 1% level.

The third criteria was chosen among several ones that are sensible for logistic four parameters models: significant dose effect is often tantamount to high b values (b is significantly > 0), **Figure 7**, or $d \gg 0$ (d is significantly greater than a null value), **Figure 9**.

Those significance criteria are based on Student's t-test. A p-value smaller than or equal to the alpha level 0.01=1% is equivalent to a t-value larger than, or equal to, the critical value $t_{crit}(Df, 1-\alpha)^{6,7}$, with Df the degree of freedom of the test ($=24-4=20$).

For our experimental design (4 replicates for 6 doses) and where a 4P model yields: $t_{crit}(20, 0.99)=2.528$.

The Statistical Table^{7a} for T-Student Critical values at the alpha level 0.01 is used as in some cases masking data in order to remove outliers is changing the degree of freedom:

Statistical Table for T-Student Critical values

df	12	13	14	15	16	17	18	19	20	21	22	23	24
p0.01	2.681	2.650	2.624	2.602	2.583	2.567	2.552	2.539	2.528	2.518	2.508	2.500	2.492

We now detail the test for $d=0$. Based on the previous data set and using Gen5 software, the d parameter value is 3.983 and its standard error is 0.148 (see **figure 10**).

One can use a student statistic to test if $d \gg 0$. The Student Ratio is equal to

$$SRatio = \frac{d_{upper} - 0}{Std\ error} = \frac{3.983}{0.148} = 26.912$$

The critical value for a one-sided test of significance at the 1% level is $t_{crit}(20, 0.99)=2.528$ and hence the absolute value of the Student Ratio must be equal or greater than 2.528 for the test to be significant at the 1% level. Since $26.912 > 2.528$, the test is significant at the 1% level.

⁵ Bates, D. M. and Watts, D. G. (1988) Nonlinear Regression Analysis and Its Applications, New York: Wiley & Sons (pp. 103-104).

⁶ Aide-mémoire pratique des techniques statistiques. Pour ingénieurs et techniciens supérieurs. / CERESTA - Centre d'enseignement et de recherche de statistique appliquée. Paris in Revue de Statistique Appliquée, vol. XXXIV - n° spécial (1986)

⁷ Pearson, E. S., and H. O. Hartley: Biometrika Tables for Statisticians, Vol. 1, 2. Aufl. Cambridge University Press, Cambridge 1962.

^{7a} Revue de Statistique Appliquée, 1988 - Tables Statistiques - Tables N°3 - Fractiles de la loi de Student

R statistical software used to confirm implementation in Gen5™

The results obtained with the Gen5™ software were confirmed using the R software, a well-known statistical language⁸. This process of verification was carried out by creating scripts and dedicated packages in order to reproduce all the statistical analyses done by Gen5. Both are equal if rounded at 3 decimal places.

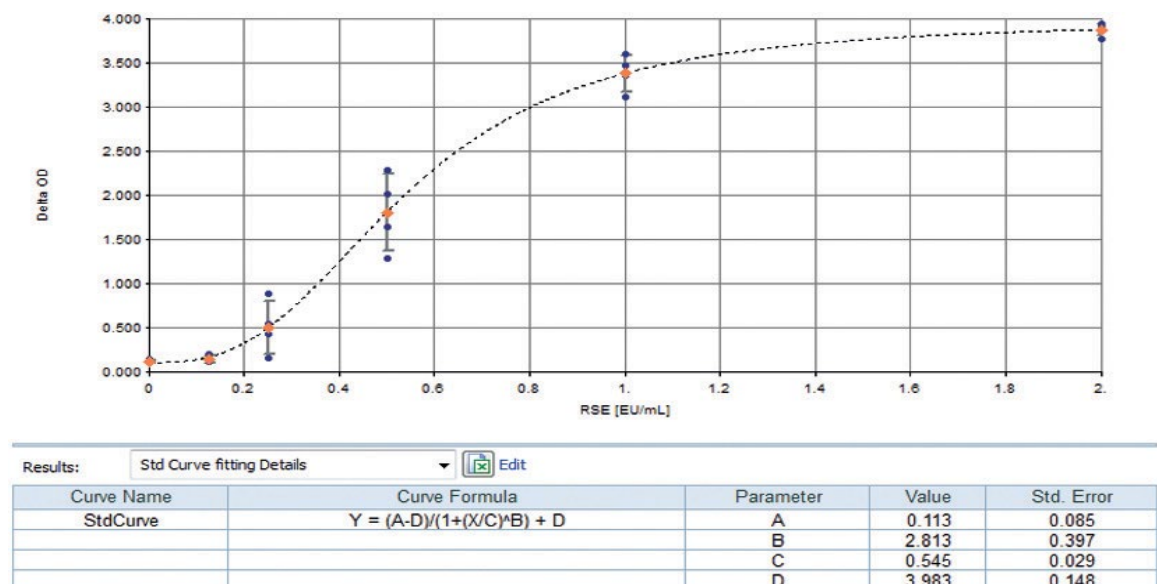


Figure 10: PyroDetect data. Regression and Values of the 4 parameters generated by Gen5

4P Parameters given by Gen5

Curve Name	Curve Formula	Parameter	Value	Std. Error	SRatio
StdCurve	$Y = (A-D)/(1+(X/C)^B) + D$	A	0.113	0.085	
		B	2.813	0.397	
		C	0.545	0.029	
		D	3.983	0.148	26.912

4P Parameters given by 'R'

Curve Name	Curve Formula	Parameter	Value	Std. Error	SRatio
StdCurve	$Y = (A-D)/(1+(X/C)^B) + D$	A	0.112842	0.084915	
		B	2.813270	0.396769	
		C	0.544765	0.029152	
		D	3.983084	0.148003	26.912

4P Parameters given by 'R' rounded at 3 decimals

Curve Name	Curve Formula	Parameter	Value	Std. Error	SRatio
StdCurve	$Y = (A-D)/(1+(X/C)^B) + D$	A	0.113	0.085	
		B	2.813	0.397	
		C	0.545	0.029	
		D	3.983	0.148	26.912

Error (in %) between parameters given by Gen5 and R

Curve Name	Curve Formula	Parameter	Value	Std. Error	SRatio
StdCurve	$Y = (A-D)/(1+(X/C)^B) + D$	A	0.000	0.000	
		B	0.000	0.000	
		C	0.000	0.000	
		D	0.000	0.000	0.000

Figure 11: Comparison of the regression generated by Gen5 and R.

⁸ R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2018, <https://www.R-project.org>

3. Five-parameter logistic model to describe PyroMAT® data

PyroMAT® data set

Figure 12 plots a dataset generated with PyroMAT® System.

As seen previously, monocytes from whole human blood react to pyrogens in dose-response manner until reaching a plateau, in the dynamic range of usual absorbance reader (OD between 0 – 4).

Mono-Mac-6 cells, used in the PyroMAT® System are more sensitive, react strongly to pyrogens and do not reach the plateau in the dynamic range of readers. The observed plateau is artificial and is due to the saturation of the signal. Therefore, the actual dose-response curve shows asymmetry and requires an adequate regression model.

To control asymmetry of the curve, the 4PL model can be extended by adding a fifth parameter e . This model is called the five-parameter logistic (5PL) model⁹. Asymmetry means that the curve does not near the lower or higher intercept at the same pace. **Figure 13** displays curve with no asymmetry (bold curve) and two curves with asymmetry values greater than 1 that nears the higher asymptote faster than they near the slower one.

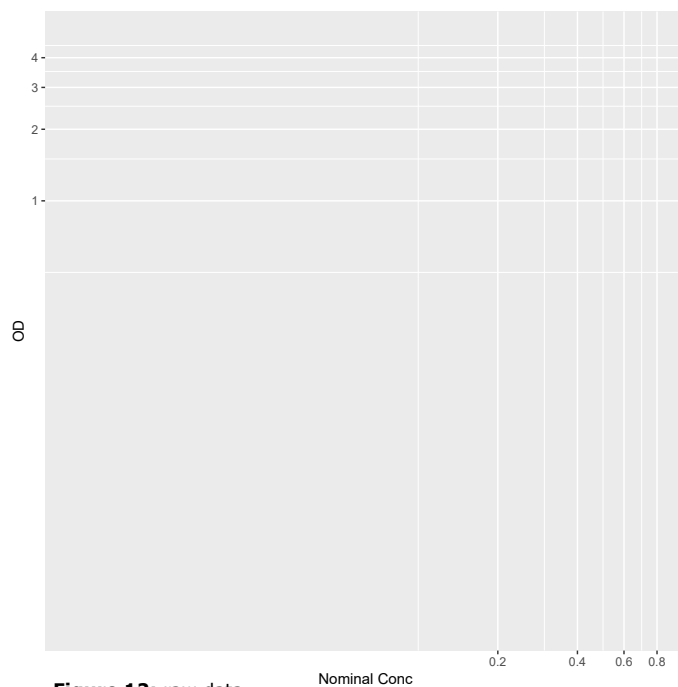


Figure 12: raw data

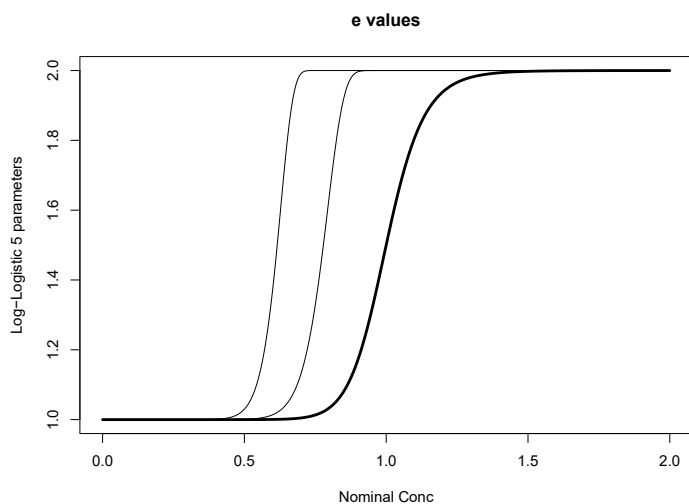


Figure 13: e values (from left to right 1000; 30; 1)

As shown on the **figure 14**, the 5-parameter logistic model fits PyroMAT® data.

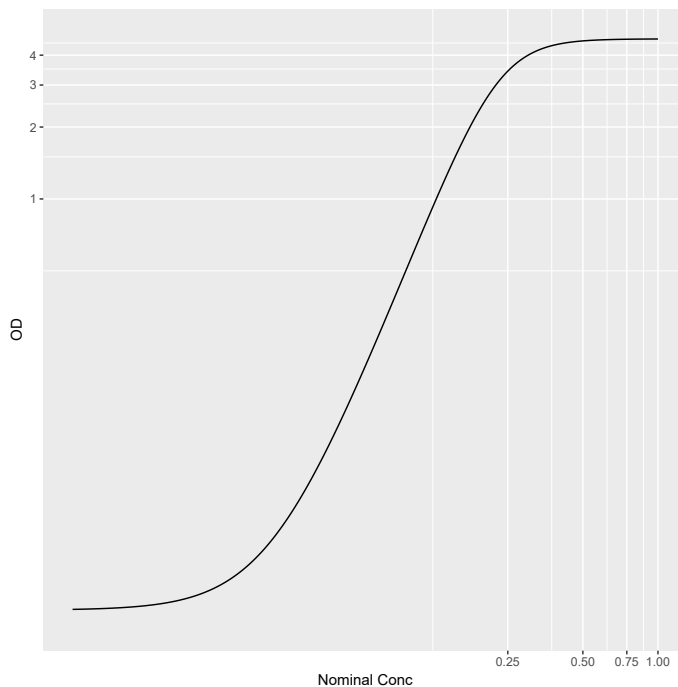


Figure 14: 5P logistic fit

⁹ In The five-parameter logistic: A characterization / P.G. Gottschalk, J.R. Dunn / Anal. Biochem. 343 (2005) 54–65, the authors have compared the 5PL to the 4PL in many thousands of dose-response curves from a wide variety of immunoassay and bioassay technologies. With the extra flexibility afforded by its asymmetry parameter, the 5PL model is able to virtually eliminate the lack-of-fit error that occurs when the 4PL is fitted to asymmetric dose-response data.

Design of the assay

In order to fit the nonlinear logistic 5-parameter model, one needs to get observed optical densities (OD) values before and after the turning point of the S-curve. It is recommended that:

- at least one standard is not far from each asymptote
- at least 3 standards fall within the linear area of the curve, one at the inflection point and one on either side of it.

The design of experiment for the PyroMAT® System satisfies these criteria.

Yet, the actual values for which the turning point appears depend on each assay and as a consequence a wider range of seven concentrations was used to design experiments in order to fit the 5P model with the PyroMAT® system:

0.0125 EU/mL, 0.025 EU/mL, 0.05 EU/mL, 0.1 EU/mL, 0.2 EU/mL, 0.4 EU/ml, 0.8 EU/mL.

The aim is to successfully find the turning point of the curve at each assay.

One can spot that additional low concentrations values were selected in order to improve the fit in the region used for back fitting and deriving concentrations of unknown samples.

The dynamic range of usual absorbance readers starts from 0 up to 4 OD. Hence, we will need specific care to deal with values over that threshold.

Features of the 5-parameters regression model

Definition

Here is the equation of the logistic 5 parameters model:

$$f(x) = \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} + d$$

Asymmetric model

The 5P logistic model is asymmetric, see Figure 13, unlike the 4P logistic model, introduced in Section 2. Five parameters are required because it is the fewest number of parameters that any general asymmetric sigmoidal function can possess.

Five parameters

Without hindering the fitting flexibility of the 5P model, we assume in the following of this white paper that b is positive ($b > 0$). Its parameters have plain interpretation:

- a is the lower intercept (plateau) at the left of the curve (Y unit: OD).
- b is unitless and sets the overall length of the function's transition region between asymptotes. It controls alone the rate of approach to the a asymptote and jointly with e controls the approach to the d asymptote.
- c sets the location of the transition region, in the same units as X .
- d is the upper intercept (plateau) at the right of the curve (Y unit: OD).
- e is the degree of asymmetry. It controls, with b , the rate of approach to the upper intercept d .

The a and d parameters have the same interpretation as for the four parameters logistic model. Yet, in a five parameters logistic model, other analogies with the four parameter logistic model are misleading.

- The b parameter of the model is no longer proportional to the slope at the inflexion point. Moreover, the slope depends more on the e parameter than on the b one.
- The location of the curve ED_{50} point does not correspond anymore to the c parameter: it can be larger or smaller than c , depending on the values of b and e .
- The inflection point of the 5PL curve is not the same as the ED_{50} . It can be larger or smaller than c depending on the values of b and e .

Weighing of the 5-parameters model: Mean-Variance relationship

Unfortunately, it is a common feature of dose-response experiments to show that the variance of the response (i.e. the optical density values) of each point (i.e. RSE concentration) depends on the mean level of this response.

This phenomenon can easily be seen on the PyroMAT® data set presented in **Figure 15**: the spread of the values increases then decreases with the values of concentrations. The highest value of standard deviation is about 250 times bigger than the smallest.

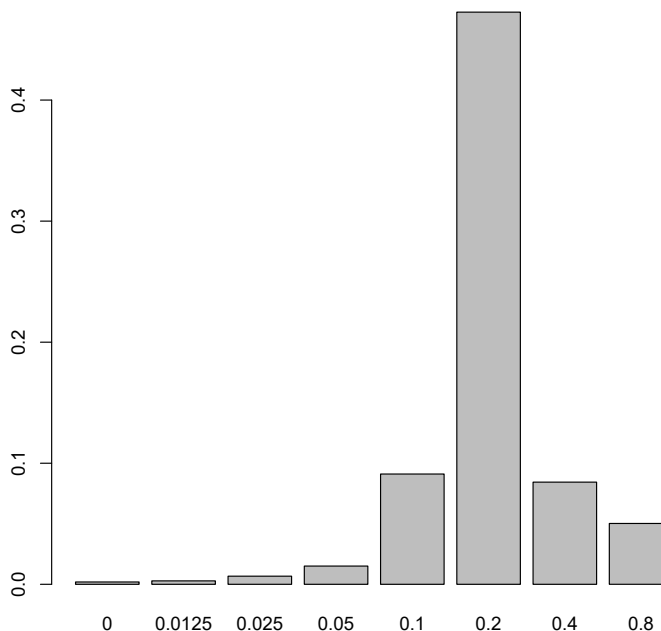


Figure 15: PyroMAT® dataset - standard deviations (Y axes) in function of RSE concentrations (X axis)

As a consequence, each point of the curve is not determined with the same accuracy. This is taken into account by weighting the curve fit with the variance of the responses.

To fulfill this requirement, a specific variance analysis has been performed on a large pool of historical data.

The **figure 16** shows a pool of observed data from RSE standard curves generated with PyroMAT® System:

- Variance increases with the Delta OD signal, on the OD range from 0 to 4
- Variance starts to decrease for the highest values of OD, greater than 4, due to the saturation of the reader
- There are less values of observed variances for OD between 2 and 4.

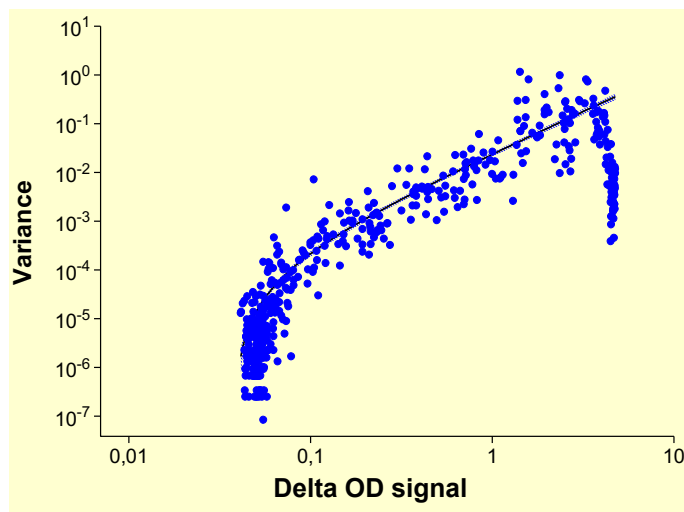


Figure 16: Observed variance versus Delta OD values

In order to determine the relationship between the mean OD and the variance of the OD, on the whole OD range, the following approach has been used:

- **Using a variance function model (see below) on the reduced OD range from 0 to 4.** In particular, this model allows to find accurate variance values for OD values in the 2 to 4 range.

To model the mean-variance relationship, we approximated the variance of the standards by the standard 3-parameter variance function^{10a} of the response:

$$\sigma^2 (Y) = (\beta_1 + \beta_2 * Y)^J$$

where σ^2 denotes variance, Y denotes the mean and β_1 , β_2 and J are the parameters. This function is always monotone over the range of the mean values; either an increasing function of the mean over the entire range, or a decreasing function of the mean.

^{10a} According to regression theory, the weights should be set equal to the inverse variance of the responses at that concentration: weighting the squared errors in this way causes the fitting procedure to adjust the curve to be tighter around those standard responses with the smallest variance (error).

Accurately estimating the true variance function from a single assay would require an impractical number of replicates. As a consequence, a large pool of historical assay data was collected during the development of the PyroMAT® System and then used to compute this variance function, once and for all, using a dedicated software, Variance function program 14^{10,11}.

- **using empirical variances for the OD values greater than 4.**

As the variance modelling is an increasing function, it was not fitting well the mean-variance relationship for OD values greater than 4, for which the variance decreases due to the saturation of the reader. As a consequence, empiric values from a large pool of historical assay data were considered.

This mixed approach allows to determine the parameters to weigh the 5-parameters regression model.

This regression model is robust to various cells reactivity profiles that could be observed even if not reaching the plateau at the RSE concentration of 0.4 EU/mL.

Statistical significance testing to fulfill EP requirement

According to the EP, the endotoxin standard curve has to fulfill several acceptance criteria:

- the goodness of fit test at a 5% level, then
- a significant dose effect at a 1% level.

Goodness of fit: a χ^2 test

Description of the χ^2 goodness of fit test.

Nonlinear regression minimizes the sum of the squared residuals (i.e. vertical distances between the data point and the curve). The value of that sum-of-squares (SS) depends on the number of data points collected and the units used to express the response. Hence, we need some reference values to carry out relevant interpretation.

The χ^2 (chi-square) statistic aims to compare the observed dispersion of the points (scatter) around the curve with the expected (based on theory or a large number of observations) amount of experimental scatter.

From a numerical point of view, this is done by computing the chi-square statistics:

$$\chi^2 = \sum_{i=1}^n \left(\frac{y_i - y_{model}}{\sigma_i} \right)^2$$

The standard deviation values must be computed from lots of data to yield accurate estimates (Subsection on Mean-Variance relationship).

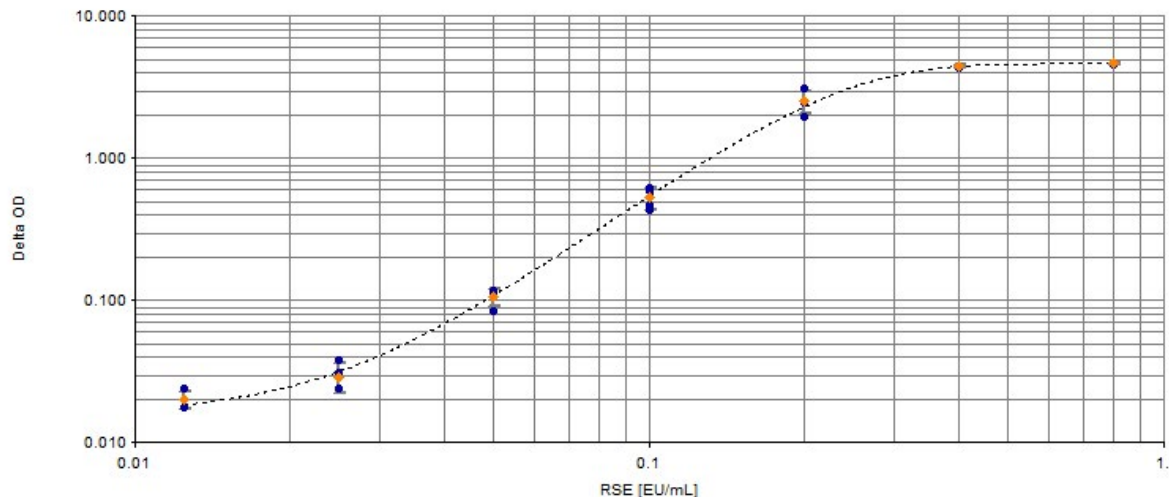
¹⁰ Sadler WA, Smith MH. A reliable method of estimating the variance function in immunoassay. *Comp Stat & Data Anal*, 3:227-239 (1986).
Sadler WA, Smith MH. A computer program for estimating imprecision characteristics of immunoassays. *Comput Biomed Res*, 23:105-114 (1990)
Sadler WA. A new Win32 computer program for estimating immunoassay variance functions. *Computer Meth & Prog in Biomed*, 67:195-199 (2002)
Sadler WA. An upgraded Win32 computer program for estimating immunoassay variance functions. *Clin Chem Lab Med*, 39(Suppl):S133 (2001)

¹¹ Available from the website of the Australasian Association of Clinical Biochemists at the http address <https://www.aacb.asn.au/professionaldevelopment/useful-tools/variance-function-program-version-140>.

Use of the χ^2 goodness of fit test.

If the model is correctly specified, the errors ($Y_i - Y_{\text{model}}$) are Gaussian and the σ_i values are accurately known, then the χ^2 statistics is distributed as a χ^2 distribution with a number of degrees of freedom, which equals the number of data points minus the number of parameters. A p -value can be computed accordingly and the fit of the model can be rejected if that p -value is lower than a given level of risk. A $\alpha = 5\%$ risk level is recommended by the EP.

With the PyroMAT[®] data set (**figure 12**), the following results are obtained with Gen5[™]:



Results: Std Curve fitting Details Edit

Curve Name	Curve Formula	Parameter	Value	Std. Error	Fit Chi2 Prob
StdCurve	$Y = (A-D)/(1+(X/C)^B)E + D$	A	0.016	0.004	0.878
		B	2.572	0.081	
		C	0.441	0.151	
		D	4.653	0.028	
		E	5.528	3.841	

Figure 17: PyroMAT[™] data. Regression and Values of the 4 parameters generated by Gen5

The Fit-Chi2-Prob value of the test is equal to 0.878 which is greater than $0.05=5\%$. Hence the test is not significant at the 5% level and the model does not missfit the data. The goodness of fit criteria is valid for this endotoxin standard curve.

Note: the main difference between the χ^2 and the Fit F-Prob test, is that the variance of the errors must be estimated using the sample for the Fit F-Prob test whereas it is estimated using accurate previous knowledge of the process (a large amount of sample is required) for the χ^2 test, see Subsection on Mean-Variance relationship.

Effect of dose

Why a global test of effect?

As stated in the 4P model section, if a dose-response experiment is modelled using a four parameters logistic model, different criteria can be used to detect and evaluate the significance of a dose effect, either based on the b parameter or d parameter (upper asymptote).

Yet, it is not convenient or even misleading to use those criteria with a five parameters logistic model and the design of experiments set up for the PyroMAT[®] System because:

- the b parameter of the model is no longer proportional to the slope at the inflexion point. Moreover, the slope depends more on the e parameter than on the b one. Hence, testing if the b parameter is equal to 0 is not testing for a 0 slope.
- the uncertainty on the d parameter is increased because of:
 - the presence of the asymmetry parameter e
 - the choice of the concentrations that favors lower values with OD before the inflection point
 - even if the uncertainty on the d parameter was low, it is most often estimated using OD values lying beyond the dynamic range of the reader.

On the contrary, the no effect test assesses, using the whole dose-response curve, whether there is a dose effect or not and does not focus only on some of the parameters.

In addition to the no effect test, additional criteria were implemented to assess the effect of dose and ensure a strictly monotonous increase of the curve:

Value of the parameter a should be below the value of the parameter d

Value of each standard should increase progressively: STD (0.05 EU/mL) < STD (0.1 EU/mL) < STD (0.2 EU/mL) < STD (0.4 EU/mL)

Description of the no effect test

A significance test¹² is provided for the comparison of the dose-response model considered and the simple linear regression model with slope 0 (a horizontal regression line corresponding to no dose effect).

Hypotheses tested by the no effect test

The aim of the no effect test is to decide between two hypotheses, named null hypothesis H_0 and alternative hypothesis H_1 , one of the two being the more likely to be true. Those two hypotheses are mutually exclusive (they cannot hold simultaneously):

- the H_0 hypothesis is that there is no dose effect at all,
- the H_1 hypothesis is that there is a dose effect.

The model used to describe the H_0 hypothesis is a model with a constant OD value whatever the concentration value, which is tantamount to no concentration effect on the OD. That constant value is the best one possible¹³ and is equal to the mean of all observed OD values.

The model used to describe the H_1 hypothesis is a model with dose-response relationship modelled by a five parameters logistic curve.

The test aims to compare those two models by evaluating whether there is a significant improvement when fitting a five parameters logistic model instead of the simpler constant one. The EP recommends a $\alpha = 1\%$ level of significance for that no effect test. By having a test with a small value of α , we feel that we have actually “proved” something when we reject the null hypothesis.

A p -value smaller than or equal to $\alpha = 0.01 = 1\%$ - p -value < 0.01- is equivalent to a Chi-square test value larger than, or equal to, the critical value $chisqcrit(Df, 1-\alpha)$ ^{14,15}, with Df the degrees of freedom of the test.

With our experimental design and 5P model yields to: $chisqcrit(Df, 1-\alpha) = 13.277$.

As a result, if the test is significant at the $\alpha=1\%$, then we reject the null hypothesis H_0 and decide that there is a significant dose effect, the risk of wrong decision being limited to 1%.

With the PyroMAT® dataset (**figure 12**), the following results are obtained with Gen5:

Chi-square test	Df
176.426	4.000

Since the Chi-square test is equal to 176.426 which is larger than 13.277, the test is significant at the 1% level: there is a significant dose effect.

¹² From a statistical point of view, this no effect test is a likelihood ratio test and is based on a chi-square distribution whose number of degrees of freedom is 4 -the number of free parameters 5 minus 1-. For more statistical details on the no effect test, see appendix 1.

¹³ For the least squares criterion

¹⁴ *Aide-mémoire pratique des techniques statistiques. Pour ingénieurs et techniciens supérieurs.* / CERESTA – Centre d’enseignement et de recherche de statistique appliquée. Paris in Revue de Statistique Appliquée, vol. XXXIV - n° spécial (1986)

¹⁵ Pearson, E. S., und H. O. Hartley: *Biometrika Tables for Statisticians*, Vol. 1, 2. Aufl. Cambridge University Press, Cambridge 1962.

4. Handling of outliers

When performing a MAT test, it happens that the endotoxin standard curve does not fulfill all the defined acceptance criteria. Therefore, a deeper analysis of the raw data is required and the handling of outlier can be a solution. Outliers can have several effects, not only on the validity of the standard curve but also on the test conclusion on pyrogenicity of the sample.

What is an outlier?

The MAT test requires to perform 4 replicates of each condition (e.g. endotoxin standard, samples, positive controls...). An outlier is a measurement that strongly deviates from the other replicates (extreme value).

The MAT test is based on a biological reaction of living cells. Therefore, it is expected that the test may demonstrate some variability and a broader distribution of the measured values than chemical reactions.

As a manually performed method, the test can also be strongly affected by handling issues such as pipetting errors or carry over contaminations from well to well.

What is the effect of an outlier ?

When computing a mean value, the smaller the sample size, the more influence an outlier will exert.

Depending on its location on the plate, outliers can have several impacts on the assay:

- Outlier in the blank can lead to invalid Limit of Detection
- Outlier in the standard curve can lead to a poorer curve fit or invalid criteria
- Outlier in the sample spiked with RSE can lead to invalid spike recovery
- Outlier in the sample could lead to inaccurate conclusions on pyrogenicity of the samples.

In a word, using or not these extreme values can change the result of the statistical analysis. As it may lead to wrong conclusion or results, outliers should be examined carefully.

Dealing with outliers

As a consequence, one should use dedicated statistical methods to deal with such a measurement and, from a data integrity point of view, cannot bluntly suppress it from the data. In order to temper the influence that such a single outlying measurement may have on the dose-response modelling or test conclusion, three solutions are commonly used:

- significance testing,
- median-based statistics,
- robust statistics.

In this white paper we focus on the first approach. The Gen5 software allows users to mask any value and it is the way to exclude any detected outlier from further analysis. To justify the presence of an outlier, 2 statistical tests can be used: Grubb's test and Dixon's test.

Grubbs' test

Grubbs' test is recommended by ISO. It computes the ratio of the deviation between the suspect value and the mean value by the standard deviation of the sample. Only one outlier can be detected by this test: the value the farther from the mean value of the sample.

Grubbs' test aims to select between two competing hypotheses, the null hypothesis H_0 and the alternative hypothesis H_1 :

- H_0 : Keep the tested value
- H_1 : Reject the tested value

For this, the T statistic is calculated:

$$T = \frac{|\text{suspect value} - \text{sample mean value}|}{\text{sample standard deviation}}$$

where both the sample mean value and the sample standard deviation are computed using the complete sample (hence including the suspect value). This test assumes that we are dealing with an approximately normal population¹⁶.

Grubbs recommends to use a smaller risk level than the one used to perform the foregoing statistical analysis of the data (typically the 1% level if the statistical analysis will be performed at the 5% level). Hence, in our case we will use the 1% level.

There are three types of critical values –see **figure 18**– for the T statistic of the Grubbs' test.

1. Testing a suspect value among the large values of the sample (T_n case).
2. Testing a suspect value among the small values of the sample (T_1 case).
3. Testing a suspect value among both the large and the small values of the sample (T case).
It is generally the case, if we do not know beforehand (before carrying out the experiment) if we will be looking for a high valued or a low valued outlier.

Number of observation n	Upper 0.1% significance level	Upper 0.5% significance level	Upper 1% significance level	Upper 2.5% significance level	Upper 5% significance level	Upper 10% significance level
4	1.499	1.496	1.492	1.481	1.463	1.425

Figure 18: Table of critical value of T (and one-sided tests T_1 or T_n)¹⁷

For small sample sizes, the outlying value must highly differ from the other values to be detected by Grubbs' test at a common risk level, namely the 1% level in our case. One of the major issue of Grubbs' test is masking effects due to more than one outlier being in the sample.

The critical values for Grubbs' test with a sample with 4 replicates at the 1% level are:

$$T_{1,crit} = T_{n,crit} = 1.492 \text{ and } T_{crit} = 1.496.$$

Note that $T_{1,crit}$ and $T_{n,crit}$ are unilateral tests at 1% level, therefore correspond to one-sided tests at 1% level, in the table. T_{crit} is a bilateral test at 1% level, therefore corresponds to a one-sided test at 0.5% level, in the table.

¹⁶ <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm#Normality>

¹⁷ Frank E. Grubbs and Glenn Beck, "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations", *Technometrics*, 14(4), 847-854 (1972).

Example

Below an example of an endotoxin standard curve obtained with the PyroMAT® System for which the goodness of fit criteria is INVALID.

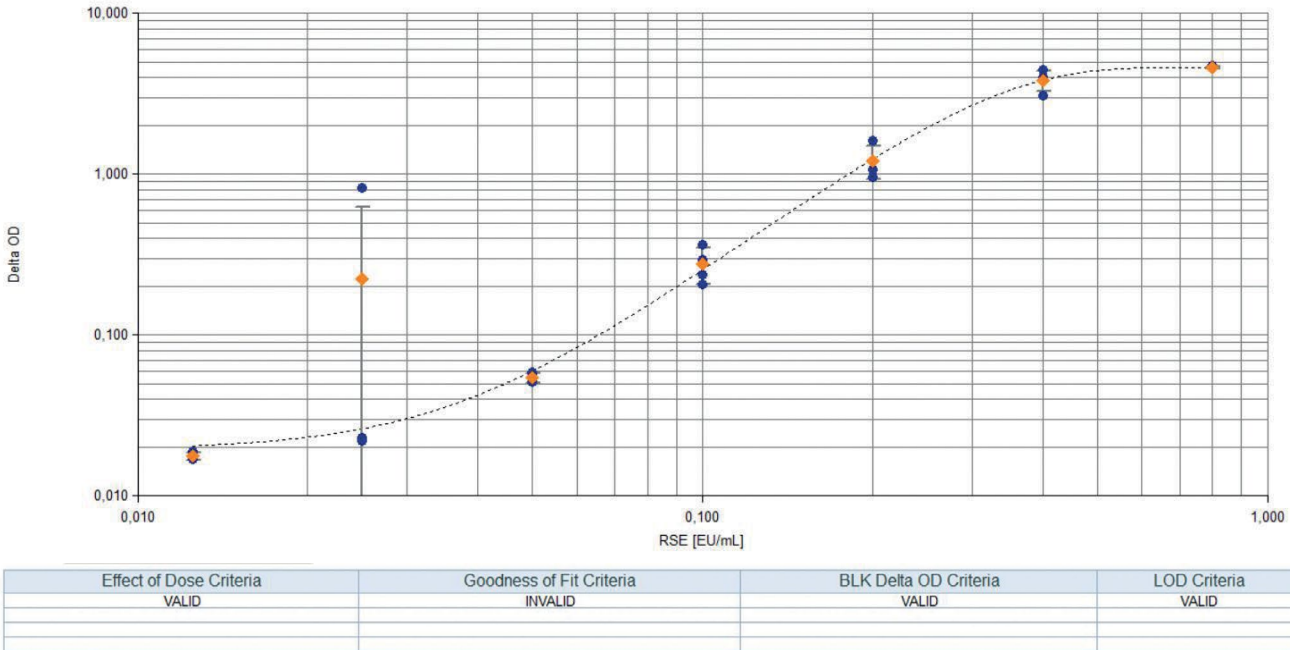


Figure 19: Example of pyroMAT data (RSE standard curve) showing a potential outlier and impact on acceptance criteria.

One OD replicates (“0.824”) of the standard at 0.025 EU/mL seems to be suspicious:

Delta OD at 0.025 EU/mL				Mean	Standard Deviation
0.022	0.022	0.824	0.023	0.223	0.401

Applying Grubbs’ test, we get:

$$T = \frac{|0.824 - 0.223|}{0.401} = 1.499$$

The T is higher than the critical T value for $\alpha=1\%$ (1.496). The test is significant and we can reject the value “0.824” as an outlier, leading to a valid endotoxin standard curve.

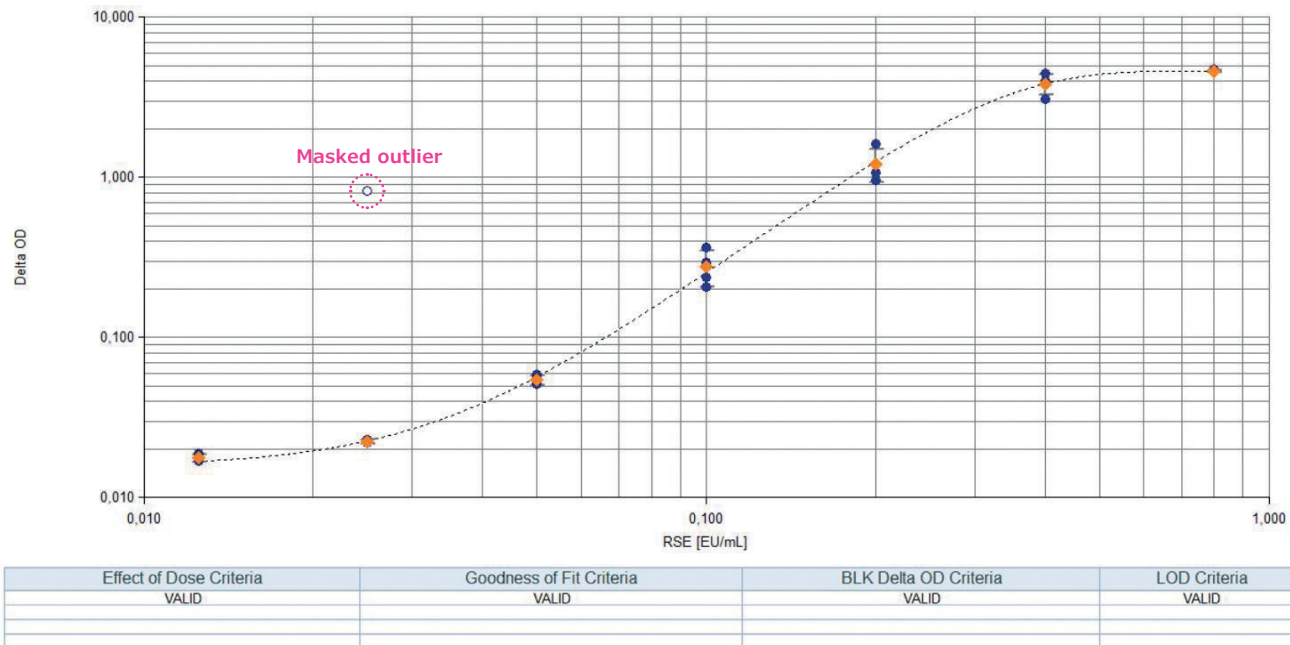


Figure 20: Example of PyroMAT® data (RSE standard curve) after masking of an outlier and impact on acceptance criteria.

Dixon's test

The scope of Dixon's test is to detect only one outlier (the same as Grubbs' test), the value the farther from the mean value of the sample. This test assumes that we are dealing with an approximately normal population.

Dixon's test aims to select between two competing hypotheses, the null hypothesis H_0 and the alternative hypothesis H_1 :

H_0 : Keep the tested value
 H_1 : Reject the tested value

Dixon's $r_{1,0}$ statistic is the ratio of the difference of the two largest values of the sample by the range (maximum values minus minimum value) of the sample. It aims to detect an outlier among the large values of the sample.

$$r_{1,0} = \frac{\text{largest value} - \text{second largest value}}{\text{range (maximum - minimum)}}$$

Dixon's $r'_{1,0}$ statistic is the ratio of the difference of the two smallest values of the sample by the range (maximum values minus minimum value) of the sample. It aims to detect an outlier among the small values of the sample.

$$r'_{1,0} = \frac{\text{second smallest value} - \text{smallest value}}{\text{range (maximum - minimum)}}$$

$r_{1,0}$ and $r'_{1,0}$ are combined and define:

$$r = \max(r_{1,0}, r'_{1,0})$$

Dixon's r search for an outlier among both the large and small values of the sample. To find if a Dixon's test is significant at the α level, we use critical values from the table given in **figure 21**. $r_{1,0}$ and $r'_{1,0}$ correspond to a one-sided test and r to a two-sided test.

Number of observation n	One sided critical values			Two-sided critical value		
	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
4	0.88940	0.76548	0.67867	0.92063	0.82973	0.76550

Figure 21: Critical value for Dixon's outlier test based on MC simulation¹⁸

The critical values for Dixon's test with a sample with size four at the 1% level are:

$$r_{1,0,crit} = r'_{1,0,crit} = 0.8894 \text{ and } r_{crit} = 0.9206.$$

It is the easiest test to carry out, even easier than Grubbs' test, since it requires very few calculations.

Example

Below an example of 4 replicates of Blank Delta OD obtained with PyroMAT® System during an assay and for which the LOD criteria is INVALID.

	1	2	3	4
A	BLK	BLK	BLK	BLK
	0,017	0,016	0,116	0,018

LOD Criteria
INVALID

Figure 22: Example of pyroMAT data (blank delta OD) showing a potential outlier and impact on LOD criteria.

The value "0.116" seems to be suspicious and a Dixon's test can be performed to determined if it is a true outlier.

¹⁸ Armin Böhner, Economic Quality Control, Vol 23 (2008), No. 1, 5 - 13
 One-sided and Two-sided Critical Values for Dixon's Outlier Test for
 Sample Sizes up to $n = 30$

Applying Dixon’s Test, we get:

$r_{1,0} = \frac{0.116 - 0.018}{0.116 - 0.016} = 0.98$	$r'_{1,0} = \frac{0.017 - 0.016}{0.116 - 0.016} = 0.1$	$r = \max(r_{1,0}, r'_{1,0}) = 0.98$
--	--	--------------------------------------

For $\alpha = 1\%$ and $n = 4$, the critical value is equal to 0.8886 for Dixon’s $r_{1,0}$ and Dixon’s $r'_{1,0}$ and even arger for Dixon’s r . As $r = 0.98$ is above the $r_{crit} = 0.9206$, we can reject the value “0.116”, which lead to a valid LOD criteria.

	1	2	3	4	
	BLK	BLK	BLK	BLK	LOD Criteria
A	0,017	0,016	*0,116*	0,018	VALID

Figure 23: Example of PyroMAT® data (blank delta OD) with masked outlier and impact on LOD criteria.

5. Conclusion

The Gen5™ protocol developed to support data analysis with PyroDetect System and PyroMAT® System allows to model the response of monocyte to endotoxin concentrations taking into account the specific features of each method and by using robust regression models that fulfill the EP requirements.

APPENDIX

The no effect test

Scope of the test

This test assesses a dose effect from a global point of view and not only through one parameter of the model. It is especially useful for the logistic 5-parameter model, since for this model it is difficult to single out one parameter that quantifies the dose effect.

Sample output

Chi-square test	Df	p-value
176.426	4.000	0.000

Output description

Chi-square test

Value of the likelihood ratio test statistic.

Df

It is the number of degrees of freedom for the test = number of independent parameters of the full model minus one (because the reduced model is limited to a single parameter: the intercept). As a result, the Df value is equal to $5-1=4$ for a five parameters logistic model.

p-value

Value of the likelihood ratio test statistic.

Hypotheses tested

The aim of the test is to decide between the two following hypotheses which one is the more likely to be true on the ground of the observed values.

- H_0 , there is no dose effect
- H_1 , there is a dose effect

How is the test statistic derived?

We first need to compute two likelihood values: the likelihood of the reduced model and the likelihood of the full model.

The log likelihood of the reduced model is equal to:

$$l_{Null} = -\frac{1}{2} (28 * (\log(2\pi) + 1 - \log(28) + \log(\text{RSS model is a single intercept}))),$$

where "RSS" is the residual sum of squares and 28 is the number of values used for the fit (7 doses without the blank with four replicates).

The log likelihood of the 5P logistic model (the full dose-response model) is equal to:

$$l_{5P} = -\frac{28}{2} (\log(2\pi) + \log(\text{RSS full 5P model}) - \log(28) + 1),$$

where "RSS" is the residual sum of squares and 5 is the number of free parameters of the model.

The likelihood ratio test statistic is equal to:

$$l_{rt} = -2(l_{Null} - l_{5P}).$$

When is the test significant at the 1% level?

- If the p-value is smaller than, or equal to, the level value $0.01=1\%$.
- If the observed l_{rt} value (=Chi-square test value in the output) is larger than, or equal to, the 99% quantile (=critical value) of the Chi-square distribution with four degrees of freedom that is equal to 13.277.

MilliporeSigma
400 Summit Drive
Burlington, MA 01803

