Paper 2563-2018

LinAcc: A SAS® Macro for Assay Linearity and Accuracy Determination

Jesse A. Canchola, Ed Marins, Ellen Paxinos

Roche Molecular Diagnostics, Pleasanton, California

ABSTRACT

Linearity and accuracy are essential metrics for a linear response assay. Linearity can be defined in various ways while the accuracy is defined as the log observed difference from the "perfect accuracy" line. The ***LinAcc*** SAS Macro packages both linearity and accuracy in one easy to use implementation on the SAS System.

INTRODUCTION

In the context of molecular assay diagnostics, what are "linearity and accuracy" and why do we need them? First, molecular assays (MA) are more commonly known as medical/biological tests (e.g., using a blood or culture sample). However, MAs are much more than that and generate data about the presence (qualitative), amount (quantitative), function, structure, or activity of biological molecules or processes that take place at the molecular level. Molecular assays can be categorized as either qualitative, or quantitative. One category of molecular tests measure pathogens through detection of specific DNA or and RNA sequence targets. Most clinical molecular assays today use the polymerase chain reaction (PCR) technique to exponentially amplify a single or several copies of a DNA fragment, and can span several orders of magnitude (e.g., from 20 copies per milliliter to 10 million copies per milliliter). Because of the exponential nature of PCR technique, the analysis is typically performed on the logarithm base 10 ($\log_{10}$) scale, which, in turn, makes the assay "linear in the log scale." Further discussion will center on the linearity and accuracy performance metrics as applied to the quantitative assay.

The linear range of the assay is defined to be the range of the quantitative assay that has been determined (or verified) in the development stage and validated in clinical trials to be the measurable portion of the assay with acceptable (or best) precision, linearity, accuracy. How are linearity and accuracy measured across the linear range of the assay? The answer lies in creating discrete levels or panels that span the linear range. An example of such a linearity and accuracy experiment, its methodology and analysis using the SAS LinAcc macro, follows.

ASIDE: The reader should note that there is a key performance metric in assay development/verification not discussed here, viz., the sensitivity of the assay. The sensitivity of the quantitative assay is the ability of the assay to detect the lowest possible concentration of target with a certain percentage (typically 95%). Additionally, there are at least two more studies that must be carried out in the validation phase (also can be performed in the verification phase), namely, method comparison/correlation and clinical utility. Method comparison/correlation compares the current assay with an approved reference assay to show similar or better performance. Finally, the clinical utility study is performed to demonstrate the ability or usefulness of the assay to impact care at the patient level.

Example of Linearity and Accuracy Study.

The use of data from an actual study may help us better understand linearity and accuracy and further motivate this discussion. In this example, the cytomegalovirus (CMV) target is used. CMV is a common virus infecting people in every age range. Though most people with CMV infections are asymptomatic, patients with weakened immune systems (e.g., organ or hematopoietic stem cell transplant recipients on immunosuppressive therapies) are at high risk of health complications due to CMV not being held in check by the body's natural immune response. Therefore, it is of clinical importance to have a good assay to detect and quantitate CMV for monitoring and therapy.

The CMV assay under discussion is tested using twelve contrived panels across a portion of the linear range of the assay from 250 International Units per milliliter (IU/mL) to 10000 IU/mL [Note: "International Unit is a unit of measurement for the amount of a substance"[1]]. On the logarithmic scale, this range is from 2.3979 $\log_{10}$ IU/mL to 4.0 $\log_{10}$ IU/mL. Figure 1 shows the 210 individual observed CMV sample observations with 30 samples the each of the expected (i.e., nominal) concentration levels in $\log_{10}$ IU/mL with the ordinary least squares (OLS) regression line overlay.

Figure 2 shows the mean $\log_{10}$ IU/mL version of Figure 1 with an OLS regression line overlay. Incidentally, the OLS regression line will also include the 95% confidence interval for both the slope and intercept. If the 95% confidence interval for the slope contains 1.0, this is evidence of linearity and if the 95% confidence interval for the intercept contains 0, this is further evidence for accuracy as the regression line would be parallel (slope≅1) and go through the origin (intercept≅0), thus approximately matching the (dashed) unity line equation. This is one of several methods to show evidence of linearity, amongst others such as the lack of fit test, cumulative sum of residuals test, quadratic equation term testing or a simple pass/fail range interval from the development process which is then compared to the observed linearity metrics. Since many of the tests described for statistically testing linearity are sensitive to small departures from linearity, we use the latter method described prior as the most non-parametric method for declaring linearity and accuracy. For this CMV assay, let us use prior specifications from the assay capability as derived in the development process to declare

linearity if the log difference between the observed and linearized value (to be described later below) or, alternatively, the OLS value is within ± 0.2 $\log_{10}$ IU/mL, representing from 60% to 160% acceptable "recovery" of the original input material from the predicted value.

Now, the regression line in Figure 2 appears to be slightly non-parallel to the unity line. From the data, this would be because the bottom level mean is "pulling" the regression line towards the dashed unity line. Though technically still within the linear range, a quantitative assay can sometimes have issues of precision at the lower levels which may unfairly bias the accuracy estimate. For a better accuracy, estimate, if we can assume a perfect dilution series from the top level to the lower levels, we can fit a line that more accurately reflects the dilution process that was used for creating the panels themselves. This process is called linearization. Figure 3 shows the results of this linearization process (using the top 6 levels) that allows for a better estimate of linearity and accuracy of the measurement. The reader should note that this process will remove the error variation contribution of the dilution itself. Figure 4 overlays the regression line and the line from the linearization process.

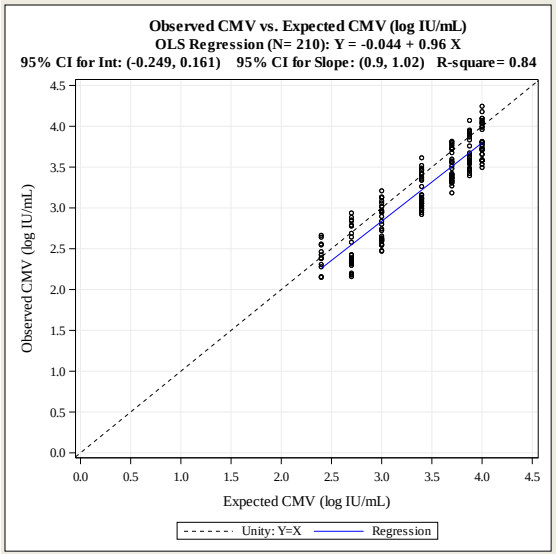

Figure 1. Observed CMV vs. Expected CMV ($\log_{10}$ IU/mL) with Regression Overlay.

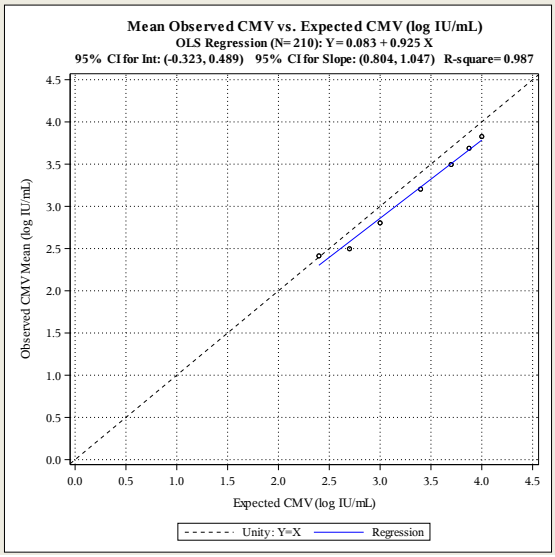

Figure 2. Mean Observed CMV vs. Expected ($\log_{10}$ IU/mL) with Regression Overlay.

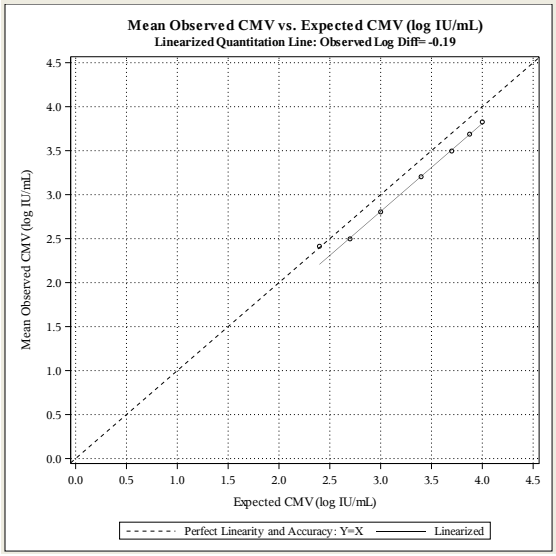

Figure 3. Observed CMV vs. Expected CMV ($\log_{10}$ IU/mL) with Linearization Overlay.

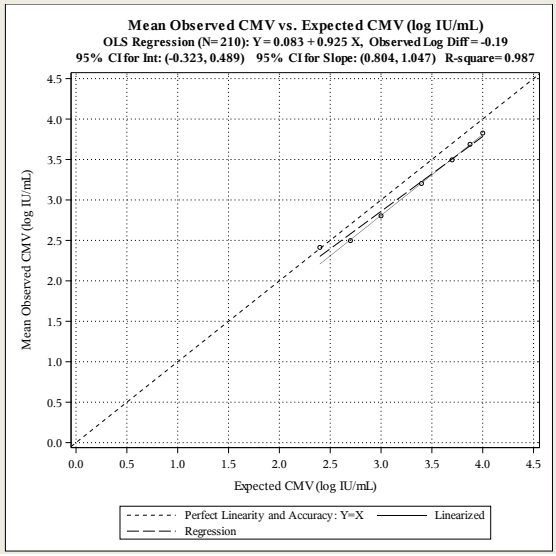

Figure 4. Observed CMV vs. Expected CMV ($\log_{10}$ IU/mL) with Regression and Linearization Overlays.

Each of the Figures 1 to 4 are useful for gauging linearity and accuracy. They will produce the metrics in **Error! Reference source not found.**, below. Each of these metrics is defined in Table 2.

Target Concentration (IU/mL) A	Target Conc (log IU/mL) B	Level C	Total Valid Sample D	Observed Mean (IU/mL) E	Observed Mean (log IU/mL) F	Linearized Quant (IU/mL) G	Linearized Log Quant (log IU/mL) H	Accuracy: Log Recovery (log IU/mL) I	Linearity: Log Diff (log IU/mL) J	Avg. Acc: Observed Log Diff (log IU/mL) K	% Recovery L
10000	4.0000	1	30	7530.7	3.8267	6444.8	3.8092	-0.1733	0.0175	-0.1908	67.1
7500	3.8751	2	30	5392.3	3.6879	4833.6	3.6843	-0.1872	0.0036	-0.1908	65.0
5000	3.6990	3	30	3425.0	3.4958	3222.4	3.5082	-0.2032	-0.0124	-0.1908	62.6
2500	3.3979	4	30	1780.6	3.2035	1611.2	3.2071	-0.1944	-0.0036	-0.1908	63.9
1000	3.0000	5	30	729.3	2.8041	644.5	2.8092	-0.1959	-0.0051	-0.1908	63.7
500	2.6990	6	25	371.6	2.4973	322.2	2.5082	-0.2017	-0.0109	-0.1908	62.9
250	2.3979	7	12	277.7	2.4134	161.1	2.2071	0.0154	0.2062	-0.1908	103.6

Table 1. Linearity and Accuracy of the CMV Assay.

Column	Description / Definition
A	CMV Target Concentration in international units per milliliter (IU/mL).
B	CMV Target Concentration in logarithm base 10 units ($\log_{10}$ IU/mL)
C	Concentration level. Typically, this can go from highest to lowest concentration but can go the opposite direction.
D	For this study, each level starts out with 30 observations. Note that not all observations were detected at the bottom two levels.
E	Observed mean of each level in IU/mL.
F	Observed mean of each level in $\log_{10}$ IU/mL.
G	Linearized quantitation of each level in IU/mL. See process of linearization in the section below.
H	Linearized quantitation of each level in $\log_{10}$ IU/mL. See process of linearization in the section below.
I	Log recovery (aka **Accuracy**) = F – B = $\log_{10}$ Observed Mean – Log_{10} Target Concentration
J	Log difference (aka **Linearity**) = E – H = $\log_{10}$ Observed Mean - Linearized Log_{10} Quantitation
K	Observed log difference (aka Average Accuracy) = H – B = Linearized Log_{10} Quantitation - Log_{10} Target Concentration
L	Percent (%) recovery = Accuracy in percentage from 100% = 10^I • 100% = $10^{(\text{Log recovery})}$ • 100%

Table 2. Linearity and Accuracy of the CMV Assay.

Linearity Assessment

For this example, Table 1 shows that linearity (column J) ranges from -0.0036 log10 IU/mL at the 2500 IU/mL level to 0.2062 $\log_{10}$ IU/mL at the 250 IU/mL level. . Since all of the log difference values are within our *a priori* set cutoff of ± 0.2 log10 IU/mL, we can declare linearity within the assay range measured.

Accuracy Assessment

Similarly, Table 1 shows the accuracy (column I) ranges from -0.2032 at 5000 IU/mL to 0.0154 at 250 IU/mL with the overall accuracy is -0.1908 (column K). Coupled with the fact that these accuracy metrics are within of ± 0.2 log10 IU/mL (rounded to the specification precision level) and the 95% confidence interval for the intercept of the regression lines in Figures 1 and 2 include zero, we can declare accuracy for the CMV assay.

The following sections will expand on the definitions and motivation for linearity and accuracy then the authors will describe a simple method for the linearization process. Finally, the authors will show how the LinAcc SAS Macro can be used for assessing linearity and accuracy of your assay.

DEFINITION FOR LINEARITY

Linearity answers the question, "is the assay response linear in the log scale?" The CLSI guideline EP6-A roughly states that "the linearity of an analytical procedure is its ability (within a given range) to obtain test results which are directly proportional to the concentration of analyte in the sample."

One may than ask the subsequent question, "But, what does this mean operationally?", to which the authors reply with a simpler definition:

Operational Definition of Linearity

"An Assay is defined to be perfectly linear if the average assay output is proportional to the predicted input concentration." That is, if the input concentration is doubled, then the average predicted output quantitation should double. In general, this translates directly to the following equation:

$$\text{Quantitation} = K \bullet (\text{Predicted Concentration}) \qquad (1)$$

On the logarithmic scale, this linearity requirement can be restated as:

$$\text{Log}_{10}(\text{Quantitation}) = \text{Log}_{10}(K) + \text{Log}_{10}(\text{Predicted Concentration}) \qquad (2)$$

Or

$$\text{Log}_{10}(\text{Quantitation}) = \text{Log}_{10}(K) + 1 \bullet \text{Log}_{10}(\text{Predicted Concentration}) \qquad (3)$$

with equation (3) exemplifying that the requirement translates to a fixed slope of 1.0 on the $\log_{10}$ scale.

Graphically, linearity is demonstrated in Figure 5 as the distance between the lower bracket edge (i.e., the mean $\log_{10}$ observed value) and upper bracket edge or the $\log_{10}$ linearized value (i.e., on the solid line). This brings up an easy check on your results. If the mean $\log_{10}$ observed value (i.e., represented by the open circles in Figure 5) is below the line in your graph, the difference result in your table will be negative and so the opposite is true. From Figure 5, the top two values and third value are above the solid line (so positive differences will appear in your table) whilst the third and bottom three values at the lower left bottom appear below the line (so negative differences will appear in your table).

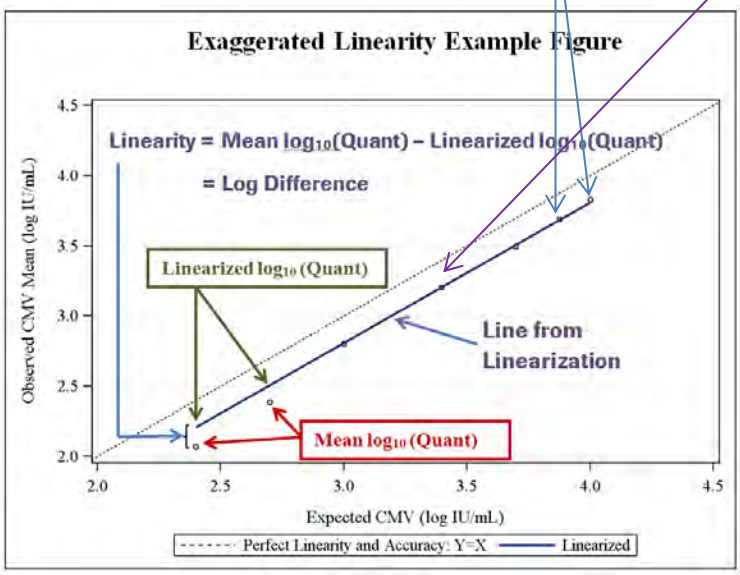

Figure 5. Example of Linearity using the Observed CMV vs. Expected CMV (log IU/mL).

DEFINITION FOR ACCURACY

Accuracy answers the question, "is the assay accurate with respect to a reference standard or expected value?" The CLSI guideline MM6-P roughly states that "the accuracy of an analytical process is the difference *(column I, Table 1)* between the average result obtainable by a method under specified conditions *(column F, Table 1)* and the value accepted as true *(column B, Table 1)."*

An Assay is defined to be accurate if the average observed log quantitation is equal (or close) to the log of the actual or expected concentration. The latter is also known as the nominal concentration.

That is, if the assay is "accurate", then if the input concentration is doubled, then the average observed output quantitation should also be doubled. In general, this translates directly to the following equation:

$$\text{Quantitation} = K \bullet (\text{Expected Concentration}) \qquad (4)$$

On the logarithmic scale, this accuracy requirement can be restated as:

$$\text{Log}_{10}(\text{Quantitation}) = \text{Log}_{10}K + \text{Log}_{10}(\text{Expected Concentration}) \qquad (5)$$

So that when K=1, we have perfect accuracy since Log10(1) = 0:

$$\text{Log}_{10}(\text{Quantitation}) = \text{Log}_{10}(\text{Expected Concentration}) \qquad (6)$$

If $k \neq 1$ along with the fit linearized line (with slope=1), this implies that there is constant bias = K or constant bias $\log_{10}K$, on the log scale. If an OLS line is fit rather than the linearized line, then and the OLS line that might be fit is parallel to the unity line (i.e., the 95% CI for the slope includes 1.0), then this will also imply that there is constant bias = K (or $\log_{10}K$ on the log scale).

Graphically, accuracy is demonstrated in Figure 6 as the distance between the lower bracket edge (i.e., the mean $\log_{10}$ observed value) and upper bracket edge or the $\log_{10}$ target value (i.e., on the dashed unity line). Again, this brings up an easy check on your results. If the mean $\log_{10}$ observed value (i.e., represented by the open circles in Figure 6) is below the dashed unity line in your graph, the difference result in your table will be negative and so the opposite is true. In this example, all the mean $\log_{10}$ observed values are below the dashed line so the differences in your table will all be negative.

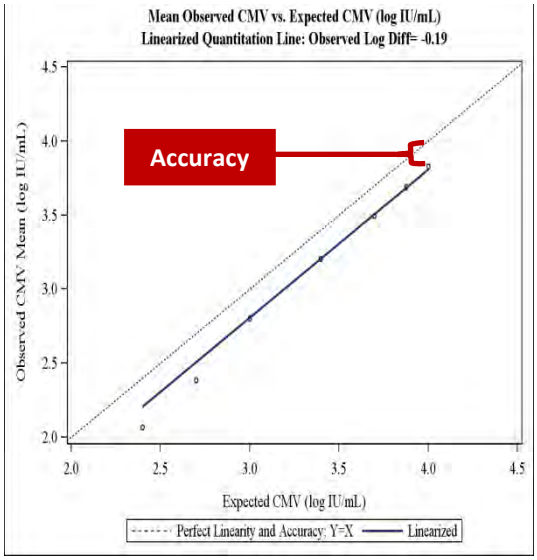

Figure 6. Example of Accuracy using the Observed CMV vs. Expected CMV (log IU/mL).

One can fit an OLS regression equation to the data point means and determined the log difference (i.e., linearity) from the predicted values of the regression equation. Assuming one has a perfect dilution series,

this regression line should be parallel (i.e., 95% CI for the slope contains 1.0) to the expected (or nominal) values of the assay panel members. Unfortunately, this is not always the case. However, not all is lost as, if one can ASSUME a perfect dilution series starting at the top or near top of the assay range, one can force a line with slope of 1.0 and have more reasonable estimates of linearity AND average accuracy. To do this one can choose the mean points that have the best relationship with a perfect dilution series process and use a method called linearization to draw the line with slope 1.

LINEARIZATION PROCESS

Linearization is the process of fitting a regression line with a slope of 1.0 assuming a perfect dilution series. The line can be constructed using the best sequence of mean $\log_{10}$ observations in the dilution series that best describes the line parallel to the nominal concentrations. This line can then be used to describe the average accuracy constant and obtain a linearized $\log_{10}$ quantitation value.

Recall the equation of a line:

$$Y = b_0 + b_1 X \quad (7)$$

where b_0 is the intercept and b_1 is the slope of the line. Next, recall that the least squares solution to the intercept is [2]:

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (8)$$

Then if we are to linearize, this implies that the slope is 1.0 so that (7) becomes:

$$\hat{Y}_{(\text{log linearized value})} = b_0 + X_{(\text{log target concentration})} \quad (9)$$

and (8) becomes:

$$b_0 = \bar{Y} - \bar{X} \quad (10)$$

so that our final equation for obtaining the linearized value combining (9) and (10) is:

$$\hat{Y}_{(\text{log linearized value})} = (\bar{Y} - \bar{X}) + X_{(\text{log target concentration})} \quad (11)$$

Note that equation (11) holds as long as the user does not have any fixed effects that need to be controlled for. In those cases, one must fit a multivariable fixed effects regression model.

INTRODUCING THE LINACC SAS MACRO

LinAcc is a SAS macro for linearity and accuracy evaluation of molecular diagnostic (or similar) assays. The macro uses the SAS/IML facility (i.e., PROC IML) to store regression parameter estimates and create other estimates for use in PROC SGPLOT. Figure 7 shows the macro call for the ***LinAcc*** SAS Macro. The macro code itself is included in Appendix A.

LINACC SAS MACRO CALL

```
%include "C:\SAS\macros\LinAcc.SAS" ;
%LinAcc(DSIN    = Main1   ,  * SAS Data Set ; ❶
        Path    = C:\SAS  ,  * Output path  ; ❷
        Target  = CMV     ,  * Short Target Descriptor ; ❸
        Units   = IU/mL   ,  * Target Units ; ❹
        TargetConc = TargetConc     , * Target Concentrations ; ❺
        X          = Log10TargetConc, * Log10 Target Concentrations ; ❻
        Y          = Log10Observed  , * Log10 Observed Concentrations;❼
        GraphMin   = 0   ,  * Minimum X-Y value for SGPLOT graph ; ❽
```

```
GraphMax    = 8   ,  * Maximum X-Y value for square SGPLOT graph;❾
GraphIncrement = 0.5 , * SGPLOT X-Y graph increment ; ❿
LinMin      = 1   ,  * Starting top level to linearize on ; ⓫
LinMax      = 5   ,  * Ending bottom level to linearize on ;⓬
LinAccTableLevel = 6) ; * How much of the Linearity and Accuracy
                          table to show: eg, "where Level < 6" ; ⓭
```

Figure 7. SAS Code for *LinAcc* SAS Macro call.

Table 3, below, shows a more formal set of specifications and defaults with examples of inputs.

Parameter #	Macro Parameter	General Macro Parameter Description	Specifications / Defaults	Example
❶	DSIN	SAS input data set (permanent or temporary)	< previous data set >	data.Main1
❷	Path	Output Path	Current Directory	c:\sas
❸	Target	Short Target Descriptor	[Required]	CMV
❹	Units	Target Units	[Required]	IU/mL
❺	TargetConc	Target Concentrations	[Required]	TargetConc
❻	X	Log10 Target Concentrations	[Required]	Log10TargetConc
❼	Y	Log10 Observed Concentrations	[Required]	Log10Observed
❽	GraphMin	Minimum X-Y value for SGPLOT graph	[Required] / 0	0
❾	GraphMax	Maximum X-Y value for SGPLOT graph	[Required] / 8	8
❿	GraphIncrement	SGPLOT X-Y graph increment	[Not Required] / 0.5	0.5
⓫	LinMin	Starting top level to linearize on	[Required] / 1	1
⓬	LinMax	Starting bottom level to linearized on	[Required] / 4	5
⓭	LinAccTableLevel	How much of linearity and accuracy table to show: eg, “where Level < 6”	[Not Required] / 50	6

Table 3. *LinAcc* SAS Macro Parameter Specifications and defaults

PRACTICAL EXAMPLE

A practical example was exhibited in the Introduction section using the CMV target and its expected values for each level. That example used the following ***LinAcc*** SAS Macro code:

```
* directory where your SAS data reside ;
libname "c:\SASGF2018\Data";

* SAS data step to define TargetConc, Level and log10 of TargetConc and
  Observed ;
data Main1 ;
set data.valid ;

* Defining expected/target concentrations in IU/mL ;
* PanelMember were randomized in protocol ;
if PanelMember = 1 then do ; TargetConc = 1.0E+04 ; Level = 1 ; end ;
if PanelMember = 5 then do ; TargetConc = 7.5E+03 ; Level = 2 ; end ;
if PanelMember = 6 then do ; TargetConc = 5.0E+03 ; Level = 3 ; end ;
if PanelMember = 7 then do ; TargetConc = 2.5E+03 ; Level = 4 ; end ;
if PanelMember = 3 then do ; TargetConc = 1.0E+03 ; Level = 5 ; end ;
if PanelMember = 2 then do ; TargetConc = 5.0E+02 ; Level = 6 ; end ;
if PanelMember = 4 then do ; TargetConc = 2.5E+02 ; Level = 7 ; end ;

* log10 of concentration and extracted/observed result ;
Log10TargetConc = log10(TargetConc) ;
Log10Observed   = log10(Observed)   ;
run ;


* Drop the LinAcc.SAS macro into this or your own directory ;
%include c:\SASGF2018\Macros\LinAcc.sas" ;
%LinAcc(DSIN    = Main1                ,
        Path    = C:\SASGF2018\RTFOUT ,
        Target  = CMV                  ,
        Units   = IU/mL                ,
        TargetConc = TargetConc        ,
        X          = Log10TargetConc   ,
        Y          = Log10Observed     ,
        GraphMin   = 0                 ,
        GraphMax   = 8                 ,
        GraphIncrement = 0.5           ,
        LinMin     = 1                 ,
        LinMax     = 5                 ,
        LinAccTableLevel = 6) ;
```

RESULTS

Figures 1 to 4 and Table 1 are the output from the *LinAcc* SAS Macro. One can "tweak" the macro code to produce custom tables. For example, if one desires grid-free graphs, one can simply search for and remove the "GRID" option in the ***LinAcc*** SAS Macro code for PROC SGPLOT.

CONCLUSION

Assay performance metrics include linearity and accuracy. Whenever linearity or accuracy assessments are needed, it is important to have easy to use tools to produce camera-ready tables and graphs. The ***LinAcc*** SAS Macro is recommended as a good tool for the analyst to have for this purpose.

REFERENCES

1 International Unit. https://en.wikipedia.org/wiki/International_unit accessed on 05FEB2018.

accessed on 05MAY2016.

2 Myers R. 1986. *Classical and Modern Regression with Applications*. Boston, MA: Duxbury Press.

ACKNOWLEDGMENTS

The authors thank the internal Roche reviewers, Christian Simon and Michael Santulli, for their help in making the manuscript more readable. Any remaining errors or omissions are the authors' alone.

CONTACT INFORMATION

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

Jesse A. Canchola, MS, PStat®
Roche Molecular Diagnostics
4300 Hacienda Blvd.
Pleasanton, CA 94588
eMail: Jesse.Canchola@Roche.Com or JCanch@Gmail.Com

APPENDIX A. *LinAcc* SAS Macro Code

```
/* ************************************************************************************
* Macro LinAcc
* Calculates the Linearity and Accuracy Table and Graphs
*
* Macro Inputs (with examples):
*
* DSIN          = Main1      ; * enter data set name ;
* Path          = B:\SASout  ; * enter your path ;
* Target        = CMV        ; * enter target name, e.g., HIV, HBV, HCV, Zika ;
* Units         = IU/mL      ; * enter units of target measurement, e.g., copies/mL or cp/mL ;
* TargetConc = TargetConc ; * enter target concentration variable ;
* X             = Log10TargetConc ; * enter log target concentration variable ;
* Y             = Log10Observed ; * enter log 10 of the observed variable ;
* GraphMin   = 0 ; * enter minimum value for SGPLOT graph square X- and Y-axis in log10 units ;
* GraphMax   = 8 ; * enter maximum value for SGPLOT graph square X- and Y-axis in log10 units ;
* GraphIncrement = 0.5 ; * enter graph increment value for SGPLOT graph square X- and Y-axis ;
* LinMin        = 1 ; * Minimum Point to Fit Linearization Curve - typically 1;
* LinMax        = 5 ; * Maximum Point to Fit Linearization Curve - typically one or two less
*                      than the total or can be total number of levels ;
* LinAccTableLevel = 6 ; * Enter  how much of the Linearity and Accuracy table to show
*                         "where Level < 6" ;
* ODStype    = RTF ; * enter ODS output file type ;
* ODStype     = RTF      ; * enter ODS output type: RTF, PDF, HTML, etc. ;
* RTFout      = \\RPBMSSASP1\Biometrics\Global\Temp\LoD_&target._&study._&sysdate..&ODStype ;
*                         * output &ODStype file ;
*
* Created by : Jesse A. Canchola (JAC)
* Creation Date:  26-Feb-2018
*
* Modified by:
* Modification Date:
* Modification:
* ------------------------------------------------------------------------ ;
* Linearization Process Ingredients
*  E = Expected (Exp)
*  O = Observed (Obs)
*  L10E = Log10(Expected)
*  L10O = Log10(Observed)
*  Linearization Range (line fit in this range with good precision):
*         = Level 1 to 5 (out of 7 levels)
*  IntEst = Intercept Estimate =
*         = Average of (Exp Hi to Exp Lo) - Average of (Obs Hi to Obs Lo)
*  LinMin = Minimum Point to Fit Linearization Curve
*  LinMax = Maximum Point to Fit Linearization Curve
* ------------------------------------------------------------------------ ;
* ************************************************************************************ */
%MACRO LinAcc(DSIN=,
              Path=,
              Target=,
              Units=,
              TargetConc=,
              X=,
              Y=,
              GraphMin=0,
              GraphMax=8,
              GraphIncrement=0.5,
              LinMin=1,
              LinMax=4,
              LinAccTableLevel=50) ;

ods output ParameterEstimates = Parms1 FitStatistics = FitStats1 ;
proc reg data = &DSIN. ;
 model &Y. = &X. / clm cli clb ;
 output out=OLSresids1 predicted=pred1 residual=resids1 press=press1 ;
run ; quit ;
ods output close ;

proc iml ;
```

```
use Parms1 ;
read all VAR{estimate lowercl uppercl} into X ;
close Parms1 ;

use FitStats1 ;
read all VAR{NVALUE2} into Y ;
close Fitstats1 ;

ols1_b0 = X[1,1] ;
ols1_b1 = X[2,1] ;

CI95 LL1 b0 = X[1,2] ;
CI95_UL1_b0 = X[1,3] ;

CI95_LL1_b1 = X[2,2] ;
CI95_UL1_b1 = X[2,3] ;

Rsquare1 = Y[1,1] ;

call symputx("ols1_b0",round(ols1_b0,0.001)) ;
call symputx("ols1_b1",round(ols1_b1,0.001)) ;

call symputx("CI95_LL1_b0",round(CI95_LL1_b0,0.001)) ;
call symputx("CI95_UL1_b0",round(CI95_UL1_b0,0.001)) ;

call symputx("CI95_LL1_b1",round(CI95_LL1_b1,0.001)) ;
call symputx("CI95_UL1_b1",round(CI95_UL1_b1,0.001)) ;

call symputx("Rsquare1",round(Rsquare1,0.01)) ;

quit ;
run ;


* Forces a square plot (no proc template necessary) ;
ods graphics / width = 550px height = 550px ;


ods rtf body = "&path.\Graphs1_&sysdate..rtf" ;
* Linear Fit on Log10 Scale ;

proc sgpanel  data = &DSIN.  ;
  panelby RunOp / columns = 2 rows = 2 ;

  scatter x = &X. y = &Y. / markerattrs=(size=5px) ;
  reg     x = &X. y = &Y. / markerattrs=(size=5px)
                                          LINEATTRS = (THICKNESS=0.2 COLOR=Blue PATTERN=1)
name="OLS" ;

  LINEPARM x = 0 y = 0 slope = 1 / LEGENDLABEL = "Unity: Y=X"
                                   LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ;
* works in SAS >= 9.3 ;

  ROWAXIS LABEL = "Observed &Target. (log &Units.)" VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  COLAXIS LABEL = "Expected &Target. (log &Units.)" VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  title1 "Observed &Target. vs. Expected &Target. (log &Units.) by Run/Operator" ;
run ;

* Linear Fit on Log10 Scale ;
proc sgplot data = &DSIN.  ;
  scatter x = &X. y = &Y. / markerattrs=(size=5px) ;
  reg     x = &X. y = &Y. / markerattrs=(size=5px)
                                          LINEATTRS = (THICKNESS=0.2 COLOR=Blue PATTERN=1)
name="OLS" ;

  LINEPARM x = 0 y = 0 slope = 1 / LEGENDLABEL = "Unity: Y=X"
                                   LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ;
* works in SAS v9.3 ;
```

```
  YAXIS LABEL = "Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "OLS" / LOCATION=outside POSITION=bottom ;
  title1 "Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "OLS Regression (N= &N_Tot): Y = &ols1_b0 + &ols1_b1 X" ;
  title3 "95% CI for Int: (&CI95_LL1_b0, &CI95_UL1_b0)    95% CI for Slope: (&CI95_LL1_b1,
&CI95_UL1_b1)   R-square= &Rsquare1" ;
run ;


* Linear Fit on Log10 Scale with Box Plots ;
proc sgplot data = &DSIN.  ;
  *reg     x = &X. y = &Y. / markerattrs=(size=5px)
                                          LINEATTRS = (THICKNESS=0.9 COLOR=Blue PATTERN=1)
name="OLS" ;
  vbox &Y. / category = &X. ;

  LINEPARM x = 0 y = 0 slope = 1 / LEGENDLABEL = "Unity: Y=X"
                                   LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ;
* works in SAS v9.3 ;

  YAXIS LABEL = "Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "OLS" / LOCATION=outside POSITION=bottom ;
  title1 "Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "Box Plots by Expected Concentration Level" ;
run ;
ods rtf close ;

* Obtaining means of replicates ;
proc sort data = &DSIN. ; by Sample ; run ;
proc means data = &DSIN. noprint  ;
by Sample ;
id &TargetConc. &X. Level ;
var Observed &Y. ;
output out =  MeanObserved1 mean = ObservedMean &Y.Mean N=N_obs ;
run ;
proc sort data = MeanObserved1  ; by Level ; run ;

* Creating Mean Observed Difference for bias plots later below ;
data MeanObserved ;
set MeanObserved1 ;
LDiff = &Y.Mean - &X. ;
run ;
proc print data = MeanObserved ; run ;


ods output ParameterEstimates=Parms2 FitStatistics=FitStats2 ;
proc reg data= MeanObserved  ;
 model &Y.Mean = &X. / clm cli clb ;
 output out=OLSresids2 predicted=pred2 residual=resids2 press=press2 ;
run ; quit ;
ods output close ;
proc iml ;
use Parms2 ;
read all VAR{estimate lowercl uppercl} into X ;
close Parms2 ;

use FitStats2 ;
read all VAR{NVALUE2} into Y ;
close Fitstats2 ;

ols2_b0 = X[1,1] ;
ols2_b1 = X[2,1] ;
```

```
CI95_LL2_b0 = X[1,2] ;
CI95_UL2_b0 = X[1,3] ;

CI95_LL2_b1 = X[2,2] ;
CI95_UL2_b1 = X[2,3] ;

Rsquare2 = Y[1,1] ;

call symputx("ols2_b0",round(ols2_b0,0.001)) ;
call symputx("ols2_b1",round(ols2_b1,0.001)) ;

call symputx("CI95_LL2_b0",round(CI95_LL2_b0,0.001)) ;
call symputx("CI95_UL2_b0",round(CI95_UL2_b0,0.001)) ;

call symputx("CI95_LL2_b1",round(CI95_LL2_b1,0.001)) ;
call symputx("CI95_UL2_b1",round(CI95_UL2_b1,0.001)) ;

call symputx("Rsquare2",round(Rsquare2,0.001)) ;

quit ;
run ;


ods rtf body = "&path.\Graphs2_&sysdate..rtf" style=monochromeprinter ;
* Linear Fit on Log10 Scale ;

proc sgplot data = MeanObserved  ;
  scatter x=&X. y=&Y.Mean / markerattrs=(size=5px) ;
  reg     x=&X. y=&Y.Mean / markerattrs=(size=5px)
                                              LINEATTRS = (THICKNESS=0.2 COLOR=Blue PATTERN=1)
name="OLS" ;

  LINEPARM x=0 y=0 slope=1 / LEGENDLABEL = "Unity: Y=X"
                             LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ; *
works in SAS v9.3 ;

  YAXIS LABEL = "Mean Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)"      GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "OLS" / LOCATION=outside POSITION=bottom ;
  title1 "Mean Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "OLS Regression (N= &N_Tot): Y = &ols2_b0 + &ols2_b1 X" ;
  title3 "95% CI for Int: (&CI95_LL2_b0, &CI95_UL2_b0)    95% CI for Slope: (&CI95_LL2_b1,
&CI95_UL2_b1)   R-square= &Rsquare2" ;
run ;

* ------------------------------------------------------------------- ;
* Linearization Process Ingredients
*  E = Expected (Exp)
*  O = Observed (Obs)
*  L10E = Log10(Expected)
*  L10O = Log10(Observed)
*  Linearization Range (line fit in this range with good precision):
*         = Level 1 to 5 (out of 7 levels)
*  IntEst = Intercept Estimate =
*         = Average of (Exp Hi to Exp Lo) - Average of (Obs Hi to Obs Lo)
*  LinMin = Minimum Point to Fit Linearization Curve
*  LinMax = Maximum Point to Fit Linearization Curve
* ------------------------------------------------------------------- ;
proc iml ;
use MeanObserved ;
read all var{&TargetConc. &X. Level ObservedMean &Y.Mean _FREQ_ N_obs} into X ;

LinearizeMin = &LinMin ;
LinearizeMax = &LinMax ;

* Total Sample Size for each Level ;
N_Reps = X[,7] ;
```

```
* Number of rows of observations ;
N = nrow(X[LinearizeMin:LinearizeMax,]) ;

* obtain sum only for Exp Hi to Exp Lo ;
SumLog10ObsMean = sum(X[LinearizeMin:LinearizeMax,5]) ; *sum of Log10 observed mean from low to hi
                                                          the 5th variable in X is &Y.Mean ;
SumLog10ExpMean = sum(X[LinearizeMin:LinearizeMax,2]) ; *sum of Log10 expected mean from low to hi
                                                          the 2nd variable in X is &X. ;

* create mean for Exp Hi to Exp Lo ;
MeanLog10ObsMean = SumLog10ObsMean / N ; * N_Reps ;
MeanLog10ExpMean = SumLog10ExpMean / N ; * N_Reps ;

* Obtain Intercept Estimate = Averge Observed - Average Expected ;
IntEst = MeanLog10ObsMean - MeanLog10ExpMean ;

* Obtain Linearized Log Quant = IntEst(a constant) + Expected Value(a vector) ;
LinLogQuant = IntEst + X[,2] ;

* Obtain log diff (mean log observed - log linearized) ;
LogDiff = X[,5] - LinLogQuant ;

* Obtain Log Recovery = log observed - log VA ;
LogRecovery = X[,5] - X[,2] ;

* Obtain Observed Log Difference = lin log quant - log VA ;
ObsLogDiff = LinLogQuant - X[,2] ;

* Obtain Linearized Quantitation = 10^LinLogQuant ;
LinQuant = 10**LinLogQuant ;

* Calculate percent recovery = 10^(LogRecovery) * 100 percent ;
PercentRecovery = ( 10**LogRecovery ) * 100 ;

* Bring back original variables to save in one file;
&TargetConc.      = X[,1] ;
&X.               = X[,2] ;
Level             = X[,3] ;
ObservedMean      = X[,4] ;
&Y.Mean           = X[,5] ;
N_Obs             = X[,7] ;

* Since the ObsLogDiff is the same all the way down, take the first instance then send it to a macro
variable for the graph ;
ObsLogDiff1 = obsLogDiff[1,1] ;
call symputx("OLD1",round(ObsLogDiff1,0.01)) ;

print X SumLog10ObsMean SumLog10ExpMean MeanLog10ObsMean MeanLog10ExpMean N_Reps LinLogQuant
LogRecovery LogDiff ObsLogDiff LinQuant ObsLogDiff1 PercentRecovery ;

create LinDat var { &TargetConc. &X. Level N_Reps ObservedMean &Y.Mean LinQuant LinLogQuant
LogRecovery LogDiff ObsLogDiff PercentRecovery } ;
append ;
close LinDat ;
run ;
quit ;

* ------------------------------------------------------------------------------------------------ ;
* Linear Fit on Log10 Scale
* All Lines on Graph
* ------------------------------------------------------------------------------------------------ ;
proc sgplot data = LinDat  ;
  * plotting the points ;
  scatter x=&X. y=&Y.Mean / MARKERATTRS  = (size=5px) ;
  * Graphing the linearized quant line ;
  reg     x=&X. y=LinLogQuant       / NOMARKERS
                                                LEGENDLABEL  = "Linearized"  name = "Linearized" ;
  * Graphing the regression equation of the mean points ;
  reg     x=&X. y=&Y.Mean / NOMARKERS
                                                LEGENDLABEL  = "Regression"  name = "Regression" ;
```

```
  LINEPARM x=0 y=0 slope=1 / LEGENDLABEL = "Perfect Linearity and Accuracy: Y=X"
                             LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ; *
works in SAS v9.3 ;

  YAXIS LABEL = "Mean Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)"      GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "Linearized" "Regression" / LOCATION=outside POSITION=bottom ;
  title1 "Mean Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "OLS Regression (N= &N_Tot): Y = &ols2_b0 + &ols2_b1 X,  Observed Log Diff = &OLD1" ;
  title3 "95% CI for Int: (&CI95_LL2_b0, &CI95_UL2_b0)    95% CI for Slope: (&CI95_LL2_b1,
&CI95_UL2_b1)   R-square= &Rsquare2" ;
run ;


* ------------------------------------------------------------------------------------------ ;
* Linear Fit on Log10 Scale
* Just Linearization Quant Line on Graph
* ------------------------------------------------------------------------------------------ ;
proc sgplot data = LinDat  ;
  * plotting the points ;
  scatter x=&X. y=&Y.Mean / MARKERATTRS  = (size=5px) ;
  * Graphing the linearized quant line ;
  reg     x=&X. y=LinLogQuant       / NOMARKERS
                                              LEGENDLABEL  = "Linearized"  name = "Linearized" ;

  LINEPARM x=0 y=0 slope=1 / LEGENDLABEL = "Perfect Linearity and Accuracy: Y=X"
                             LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ; *
works in SAS v9.3 ;

  YAXIS LABEL = "Mean Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)"      GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "Linearized" / LOCATION=outside POSITION=bottom ;
  title1 "Mean Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "Linearized Quantitation Line: Observed Log Diff= &OLD1" ;
run ;


* ------------------------------------------------------------------------------------------ ;
* Linear Fit on Log10 Scale
* Just Regression Line on Graph
* ------------------------------------------------------------------------------------------ ;
proc sgplot data = LinDat  ;
  * plotting the points ;
  scatter x=&X. y=&Y.Mean / MARKERATTRS  = (size=5px) ;
  * Graphing the regression equation of the mean points ;
  reg     x=&X. y=&Y.Mean / NOMARKERS
                                              LEGENDLABEL  = "Regression"
                                              LINEATTRS = (THICKNESS=0.2 COLOR=red PATTERN=1)
name="Regression" ;

  LINEPARM x=0 y=0 slope=1 / LEGENDLABEL = "Perfect Linearity and Accuracy: Y=X"
                             LINEATTRS = (THICKNESS=0.2 COLOR=Black PATTERN=2) name="Unity" ; *
works in SAS v9.3 ;

  YAXIS LABEL = "Mean Observed &Target. (log &Units.)" GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;
  XAXIS LABEL = "Expected &Target. (log &Units.)"      GRID VALUES = (&GraphMin TO &GraphMax BY
&GraphIncrement) ;

  KEYLEGEND "Unity" "Regression" / LOCATION=outside POSITION=bottom ;
  title1 "Mean Observed &Target. vs. Expected &Target. (log &Units.)" ;
  title2 "OLS Regression (N= &N_Tot): Y = &ols2_b0 + &ols2_b1,  Observed Log Diff= &OLD1" ;
```

```
  title3 "95% CI for Int: (&CI95_LL2_b0, &CI95_UL2_b0)    95% CI for Slope: (&CI95_LL2_b1,
&CI95_UL2_b1)   R-square= &Rsquare2" ;
run ;
ods rtf close ;

options orientation = landscape ;
ods rtf body = "&path.\LinAccTable_&sysdate..rtf" style=Monochromeprinter ;
proc print data = LinDat label noobs style(data header obs)={just=c} ;
  var &TargetConc. &X. Level N_Reps ObservedMean &Y.Mean LinQuant LinLogQuant LogRecovery LogDiff
ObsLogDiff PercentRecovery ;
  format &X.               8.4
         ObservedMean      8.1
         &Y.Mean           8.4
         LinQuant          8.1
         LinLogQuant       8.4
         LogRecovery       8.4
         LogDiff           8.4
         ObsLogDiff        8.4
         PercentRecovery   8.1 ;

  label &TargetConc.       = "Target Concentration (&Units.)"
        &X.                = "Target Conc (log &Units.)"
        Level              = "Level"
        N_Reps             = "Total Valid Sample"
        ObservedMean       = "Observed Mean (&Units.)"
        &Y.Mean            = "Observed Mean (log &Units.)"
        LinQuant           = "Linearized Quant (&Units.)"
        LinLogQuant        = "Linearized Log Quant (log &Units.)"
        LogRecovery        = "Accuracy: Log Recovery (log &Units.)"
        LogDiff            = "Linearity: Log Diff (log &Units.)"
        ObsLogDiff         = "Avg. Acc: Observed Log Diff (log &Units.)"
        PercentRecovery    = "% Recovery" ;

  title1 "Linearity and Accuracy" ;
  title2 ; title3 ;
run ;
ods rtf close ;
%MEND LinAcc ;

* ************************************** END OF PROGRAM ************************************** ;
```