

Paper 326-2012

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10

D. W. McCormack, Jr., SAS Institute Inc., Cary, NC

ABSTRACT

Nonlinear models are frequently encountered in applied technological areas such as the semiconductor and pharmaceutical industries. Defect density and pharmacokinetic models are two examples where fitting a straight line to data is inadequate or inappropriate. Exploration of these models is often tedious and time consuming, and model comparison is difficult. Through its Nonlinear platform, JMP 10 introduces a quick and simple way to fit multiple curves and explore competing models, streamlining the discovery process. Two detailed examples are given using data similar to that found in practice.

INTRODUCTION

Exploration of nonlinear models can be a tedious and daunting task. An initial step, as with most analyses, is outlier screening. The data should be representative of the population in question before conclusions can be derived from the fitted model. Unfortunately, this step might require multiple iterations of outlier removal, refitting to the remaining data, and then assumption checking. This process might be further complicated when the investigator is interested in examining several competing models. This is often the case when there is interest in finding a model that both fits the data well and is parsimonious.

JMP 10 has added several features to its Nonlinear platform to ameliorate the issues mentioned above. At the heart of these improvements is the addition of approximately 20 commonly used predefined models (Table 1). These models free the user from having to create (and re-create or reset) the parameterized formula needed for nonlinear fitting in earlier versions. This speeds up the iterative process. As in the past, it is still possible to input custom models by creating the appropriate parameterized formula. A number of model comparison and parameter comparison methods are available for the predefined models, all of which have an attendant graph. The exploration process becomes even more efficient when these new features are combined with the visual and interactive functionality of JMP.

Group	Models
Polynomial	To degree 5
Sigmoidal	Gompertz (3P, 4P), Logistic (3P, 4P, 5P)
Exponential Growth & Decay	Exponential (2P, 3P), Biexponential (4P, 5P), Mechanistic Growth
Peak	Gaussian, Lorentzian
Pharmacokinetic	One Compartment, Two Compartment
Michaelis-Menten	Michaelis-Menten

Table 1. Predefined Models in JMP 10.

OUTLIER SCREENING: AN EXAMPLE

Pharmacokinetic models are commonly used to model how a drug, drug metabolite, or other drug by-product accumulates and is eliminated from the bloodstream. Samples are taken over multiple time points and measured for the substance of interest. This is often done on multiple subjects. For oral dosing, a one-compartment model usually fits the data well. Figure 1 shows the typical shape for this type of model. The steep peak in the early part of the curve is associated with accumulation in the bloodstream of the measured substance. Over time, accumulation slows and stops, and blood concentrations begin to fall. This is indicated by the negative sloping curve to the right of the peak. It is not unusual for elimination to occur much more slowly than accumulation. Given the unusual shape of the curve (steep peak and asymptotic right tail) and lower bounds at $y = 0$ and $x = 0$, it is unlikely that polynomial models and data transformations (that is, techniques used to transform the input and/or response so that linear modeling methods can be used) can consistently produce good-fitting models.

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

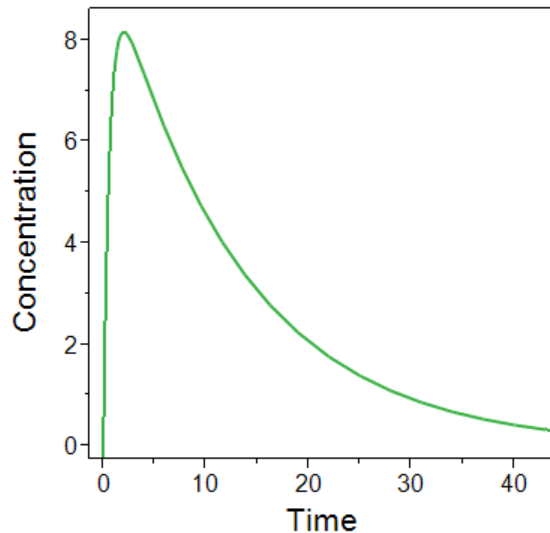


Figure 1. One-Compartment Oral Dose Curve

The functional form of the one-compartment model is given by

$$y = \frac{abc}{c-b}(e^{-bx} - e^{-cx})$$

where

- y is blood concentration.
- x is time.
- a is clearance rate.
- b is elimination rate.
- c is absorption rate.

To access the predefined models, specify the predictor variable as a column of data rather than a parameterized formula. The model groups are located under the report window hotspot in the top-most outline box as seen in Figure 2.

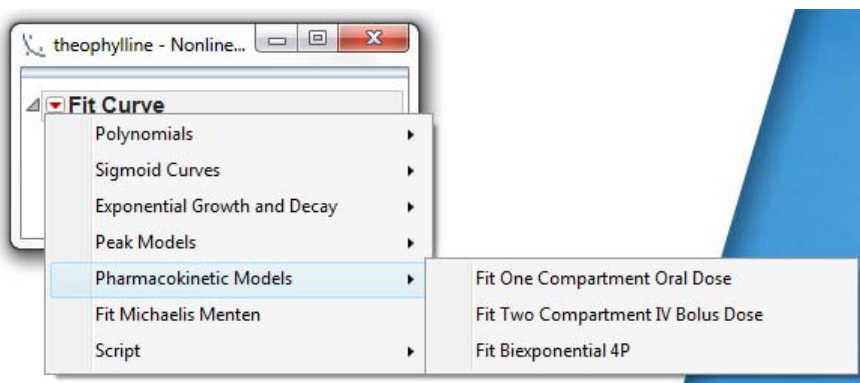


Figure 2. Predefined Pharmacokinetic Models

Data for this first example comes from a study in Boeckmann, Sheiner, and Beal (1994) examining theophylline blood concentration over time after oral administration. The data set is freely available from the R package (<http://cran.r-project.org/index.html>). The study consisted of 12 subjects, each of whom received a single dose of theophylline. Theophylline blood concentration readings were taken 11 times for each subject over a 25-hour period.

For this analysis, Subject is specified as the grouping variable so that separate models are fitted for each person. Selection of the one-compartment menu choice results in the graph shown in Figure 3, which can be used as a visual tool for identifying outliers. Two possible outliers have been marked manually with a red asterisk. To complete the outlier screening iteration, the two points are selected using one of the interactive selection tools and are excluded and hidden (Figure 4). The model is then refitted (Figure 5). This process is easily repeated if necessary.

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

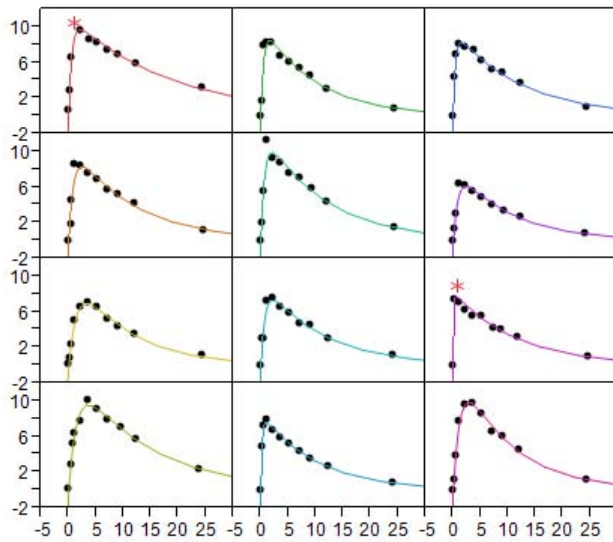


Figure 3. Curve Fits for Each Subject. Possible Outliers Marked by a Red Asterisk.

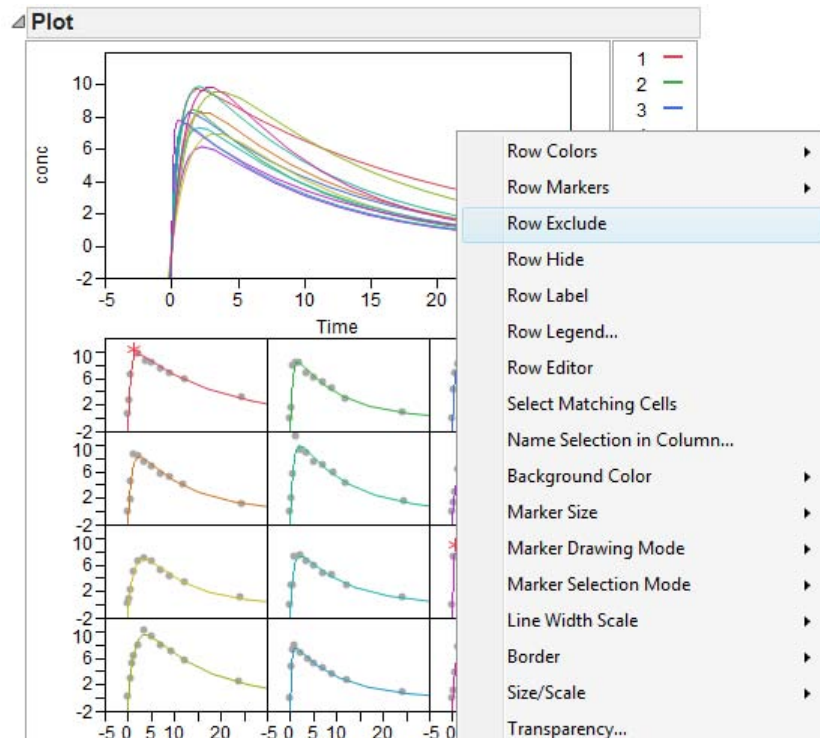


Figure 4. Right-Clicking with Selected Points Allows for Quick Row Exclusion and Hiding.

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

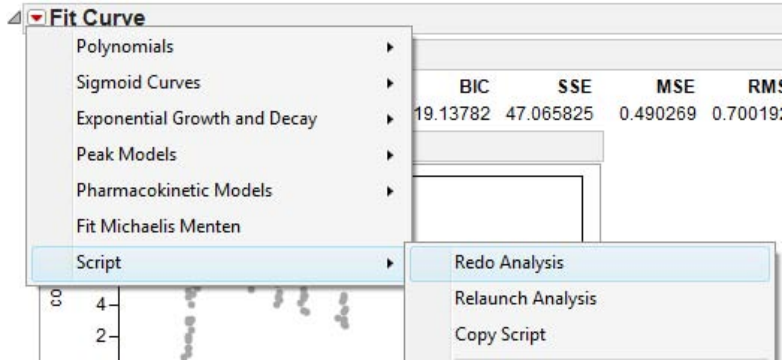


Figure 5. Model Refitting after Outlier Removal.

IDENTIFYING DIFFERENCES IN SUBJECTS

Examining the subjects for atypical results is another useful step in insuring sample homogeneity. One approach is to see how a model parameter differs for each subject as compared to the overall parameter mean (Figure 6). This test is implemented using analysis of means (ANOM). ANOM was introduced in the linear modeling platform Fit Y by X in JMP 9 and extended to nonlinear models in JMP 10. The three ANOM graphs for the example data set are shown in Figure 7. Observations in red are considered statistically different from the mean. For example, the graphs indicate that subject 1 has a greater clearance rate and is eliminating theophylline from his system more slowly than the average subject.

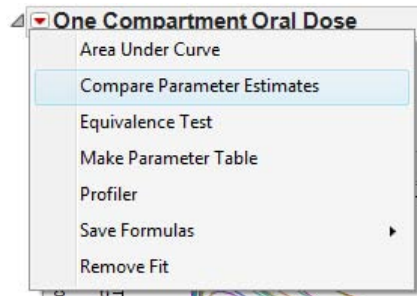
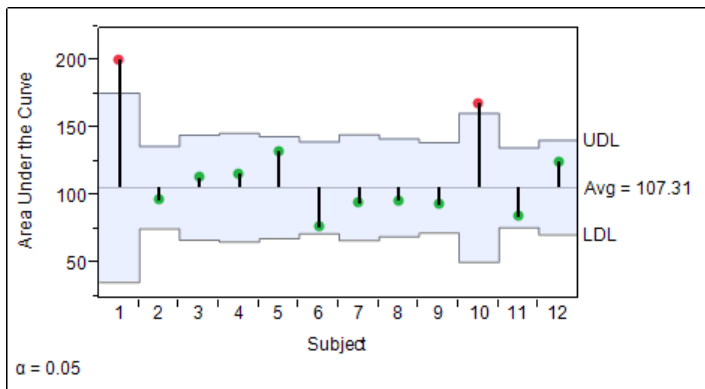


Figure 6. Comparing Parameters across Subjects. This Option Is Found under the Fitted Model's Hotspot.



Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

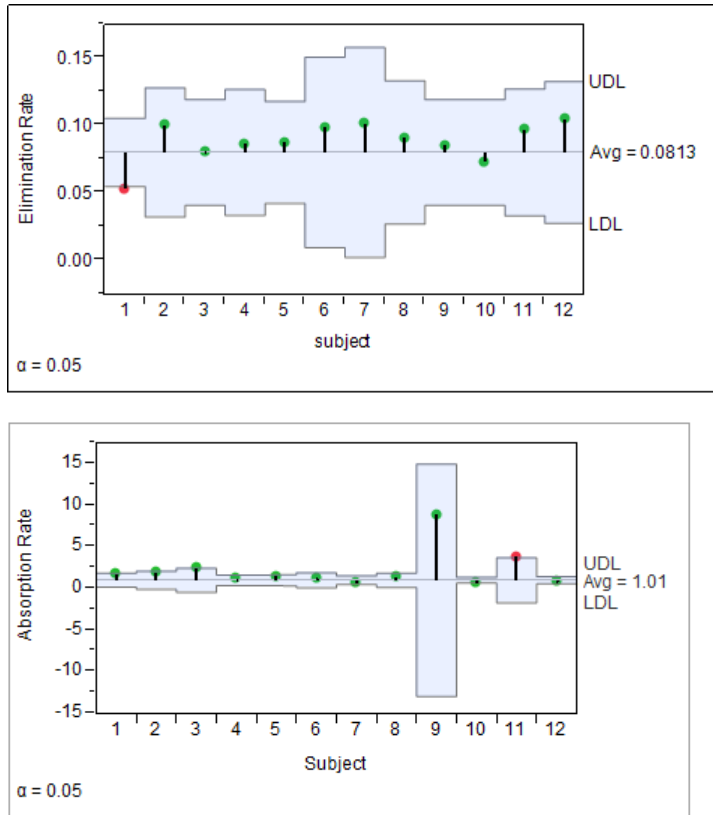


Figure 7. ANOM Graphs for the Three Parameters.

COMPARING MODELS

Dose response curve fitting is another situation where nonlinear models are commonly used. Dose response curves have a shape similar to that shown in Figure 8. As with pharmacokinetic models, attempts at linearization, traditionally by taking the log of both dose and response, usually produce substandard results. The models of choice for this application come from the family of Logistic functions. A commonly used five-parameter version is

$$y = c + \frac{d - c}{(1 + \exp\{-a[x - b]\})^k}$$

where

- y is response.
- x is dose.
- a is growth rate.
- b is inflection point (when $k = 1$).
- c is lower asymptote.
- d is upper asymptote.
- f is asymmetry parameter.

When $f = 1$, the function simplifies to a four-parameter model. If $f = 1$ and $c = 0$, a three-parameter model results.

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

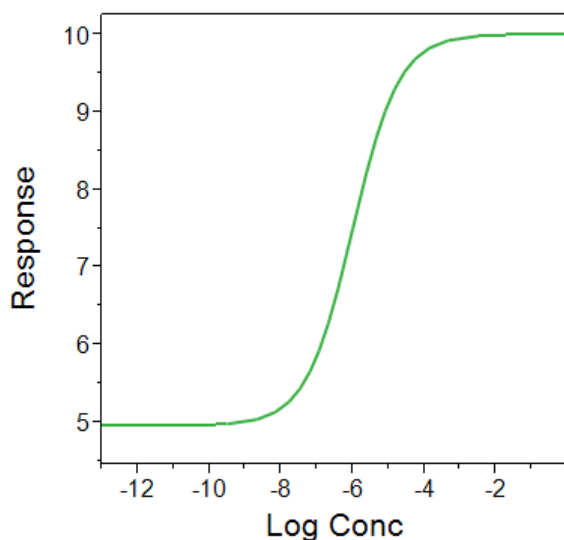


Figure 8. A Commonly Seen Relationship between Dose and Response Is an S-shaped Curve.

In the area of assay analysis, two types of model comparisons frequently occur. The four- and five-parameter models are compared to determine the better of the two, when both fit and model parsimony are considered. In cases where calculation of relative potency is of interest, a test for parallelism is conducted to determine if the curves fitted for the separate assay groups (reference versus one or more other samples) have the same slopes and upper asymptotes.

Using the data set Bioassay.jmp from the JMP 10 sample data directory to illustrate these concepts, we start by fitting the three-, four-, and five- parameter models to the data with Toxicity as the Y, Log Conc as the X, and Formulation as the Group variable.

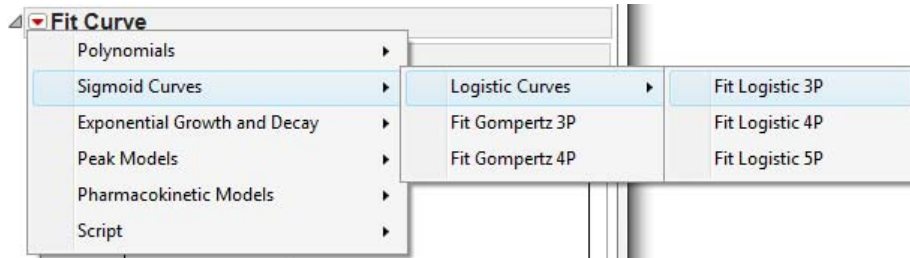


Figure 9. Location of the Logistic Curves in the Predetermined Models.

The Model Comparison outline is generated when more than one model is fitted. In it a number of comparison statistics are calculated, allowing the user to select among competing alternatives. Figure 10 indicates that the four-parameter model provides the best choice based on Akaike's Information Criteria (AIC), which considers both fit and parsimony.

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Logistic 4P	-372.3021	0.8764336					-348.9054	0.005325	0.0001109	0.0105327	0.9980584
Logistic 5P	-368.3839	0.1235664					-345.0473	0.0043613	9.912e-5	0.0099559	0.9984098
Logistic 3P	-169.6644	8.72e-45					-148.8789	0.1572372	0.0030238	0.054989	0.942668

Figure 10. Comparison Criteria for the Three Models.

Suitability for calculating relative potency can be determined using an equivalency test on the growth rate and asymptote parameters. A two-sided t-test of the reference against each sample is used. Graphs of the results in Figure 12 through Figure 14 indicate all three samples can be used. In the example data set, groups A and B are compared to the standard group.

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

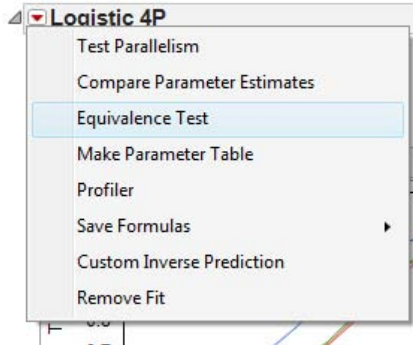


Figure 11. Equivalence Testing Can Be Found under the Hotspot for the Fitted Model.

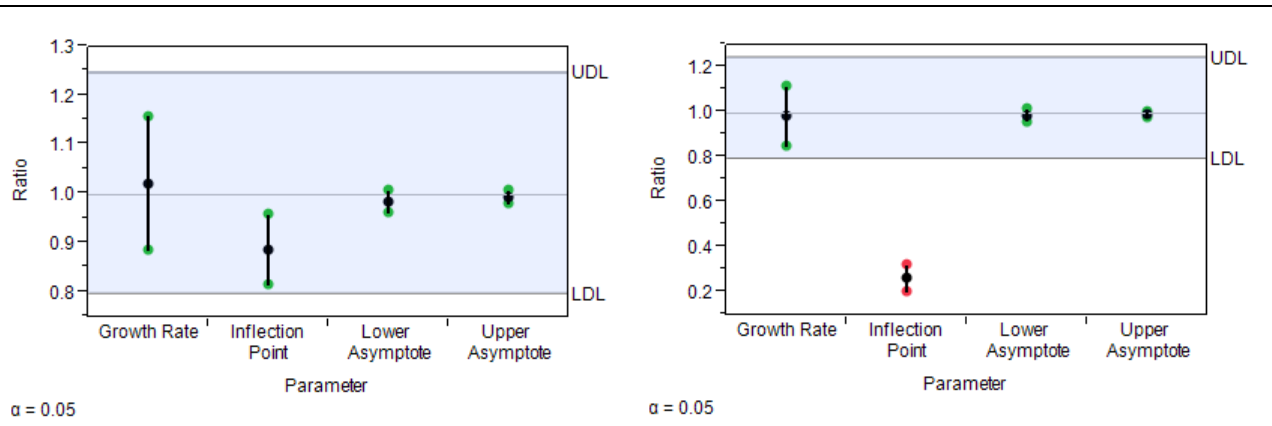


Figure 12. Equivalence between Test A and Standard. Figure 13. Equivalence between Test B and Standard.

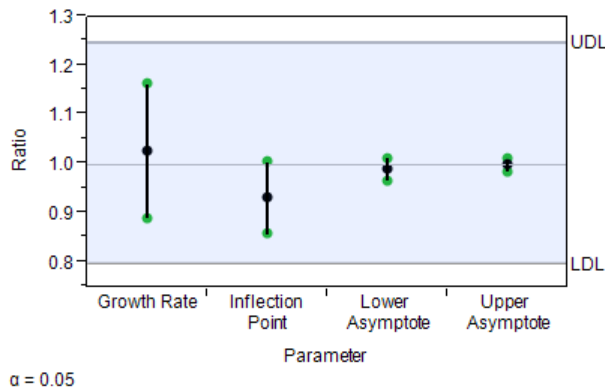


Figure 14. Equivalence between Test C and Standard.

Alternatively, the Parallelism test (also shown in Figure 11) can be used to decide whether the curves differ by their inflection point only. In this test, the error sum of squares for the full model (all parameters are allowed to vary) is compared to the error sum of squares for the reduced model (only the inflection points are allowed to vary). The relative potency calculation is directly available from this option ().

Making Exploration of Nonlinear Models Fast and Easy Using JMP® 10, continued

Parallelism Test						
Test Results						
Parallelism F Test						
Parallel						
Fit SSE	Full SSE	NDF	DDF	F Ratio	Prob > F	
0.0060088	0.005325	9	48	0.685	0.7187	
Parallelism Chi-Square Test						
ChiSquare	DF	Prob>ChiSq				
0.001	9	1.0000				
Parallel Fit Parameter Estimates						
Relative Potencies						
Relative Potency versus standard						
Group	Potency	Relative Potency		Std Error		
standard	1.3646989	1		0		
test A	1.3348492	0.9781273		0.0087756		
test B	1.0951576	0.8024903		0.0071991		
test C	1.3445131	0.9852086		0.0088392		

Figure 15. Calculation of Relative Potency Using the Test Parallelism Option.

EXTENSION TO OTHER MODELS

Although the two examples given in this paper were from the pharmaceutical industry, the techniques used can be generalized to any analysis using one of the predefined models. Exponential growth models, for example, can be used to understand how a population increases over time. Lorentzian peak models have been used to extrapolate spectral data (Rusling & Kumosinski 1996).

CONCLUSION

The new features added to JMP 10 have made nonlinear model exploration faster and easier. With the addition of approximately 20 commonly used predefined models, users no longer have to worry about building and updating parametric functions. Combined with a number of model and parameter comparison tests and the interactive capabilities of JMP, these new features can make the time gap between data collection and data discovery shorter.

REFERENCES

- Boeckmann, A. J., L. B. Sheiner, and S. L. Beal. 1994. *NONMEM Users Guide: Part V*. San Francisco, CA: NONMEM Project Group, University of California.
- Rusling, J. F. and T. F. Kumosinski. 1996. *Nonlinear Computer Modeling of Chemical and Biochemical Data*. San Diego, CA: Academic Press.

CONTACT INFORMATION

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

Donald McCormack
 SAS Institute Inc.
 11920 Wilson Parke Ave.
 Austin, TX 78726
 512.916.8060
 don.mccormack@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.