

QA Testing for the AOT425StatPgm Program

Prepared for

The US Environmental Protection Agency

Prepared by

Westat

Under Contract No. 68-W7-0025

Task Order No. 5-03

June 2001

QA Testing for the AOT425StatPgm Program

Prepared for the US Environmental Protection Agency
By Westat, June 2001

Table of Contents

1.	Introduction.....	1
2.	The Probit Model and the Profile Likelihood.....	2
3.	Overview of the Testing.....	3
4.	Data Set Construction.....	4
5.	Basis for Comparison.....	5
6.	Results for the Test Data Sets.....	8
6.1	Comparison of the Confidence Bounds.....	9
6.2	Comparison of the LD50 and Slope.....	9
7.	Comparison of the AOT425StatPgm Simulations with EPA Simulations.....	10
8.	Other Testing of the AOT25StatPgm Calculations.....	14
9.	Test Data Set Results.....	16
10.	Summary Statistics for the Test Simulations.....	20
	Appendix A: Test Data Sets.....	24
	Appendix B: Evaluation of Cumulative Normal Distributions..... and Calculation of the Chi-square Test	33

1. Introduction

Under contract to USEPA, Westat has prepared the AOT425StatPgm program to perform calculations required to implement the OECD Acute Oral Toxicity Test Guideline 425 procedure (OECD TG 425). The program calculates 1) the doses for the test animals, 2) when to stop dosing animals, and 3) the specified LD50 and a confidence interval for the LD50. All of the tests were performed using the draft version of the program and most were performed using the final program. Very few changes were made to the program between the draft and the final versions and those changes did not affect the calculations and thus are not expected to affect the test results.

This document presents the results of testing to assess and to document the quality of the calculations performed by the program. Tables that support the text are at the end of the document.

2. The Probit Model and the Profile Likelihood

The test data can be classified into the following five cases:

Case 1, No positive dose-response relationship.

Case 2, Lowest dose with a death is below the highest dose with survival.

Case 3, Lowest dose with a death is above the highest dose with survival.

Case 4, Lowest dose with a death is the highest dose with survival and there are both lower and higher doses.

Case 5, Lowest dose with a death is the highest dose with survival and there is either a lower dose with survival or a higher dose with death, but not both.

Case 2 data can be fit using an appropriate algorithm to find the parameters that maximize the likelihood. For Case 4 and 5 data, the slope that maximizes the likelihood is infinite. The profile likelihood confidence interval is calculated for Case 2, 4, and 5 data as described below.

For Case 3 data, there is a range of values for which the likelihood is maximized. For Case 3 data, the LD50 estimate is obtained by fixing the slope at the assumed slope (the inverse of the assumed sigma) and maximizing the likelihood over the LD50. Unless otherwise specified, the assumed sigma is 0.5 and the assumed slope is 2.0. A profile likelihood confidence interval is not fit for Case 3 data. Instead, an approximate confidence interval is used, going from the highest dose with survival to the lowest dose with a death.

Although maximum likelihood techniques may be used for Case 1 data, possibly resulting in a negative slope estimate, EPA has decided that the program will not calculate an LD50 estimate for Case 1 data.

The Probit model used in the AOT425StatPgm program assumes that the probability of an animal responding (dying) at a specified dose can be written as:

$$P(\text{response}) = \text{CumNorm}(\text{Intercept} + \text{Slope} * \text{Log}(\text{Dose})) \quad (1)$$

Where CumNorm is the cumulative normal distribution function and Log is the base 10 logarithm.¹ Equation (1) can also be written as:

$$P(\text{response}) = \text{CumNorm}(\text{Slope} * (\text{Log}(\text{Dose}) - \text{Log}(\text{LD50}))) \quad (2)$$

Equation (1) will be referred to as the linear parameterization. Equation (2) will be referred to as the LD50 parameterization. The maximum likelihood estimates of the parameters using one parameterization can be easily transformed to get the maximum likelihood estimates in the other parameterization. Using the linear parameterization the likelihood surface appears to have a roughly quadratic-shaped surface with one

¹ The BMDS program uses the natural logarithm. In this report, the slope reported for the BMDS program is the slope reported in the BMDS report multiplied by ln(10) to make the value comparable to the other programs.

maximum, making it easy to locate the maximum. Using the LD50 parameterization, the likelihood surface has one maximum but a contorted shape, making it more difficult to find the maximum. Nonetheless, both models generally converge when using zero for the starting parameter values. Let MaxLL be the value of the log-likelihood at the maximum.

The profile likelihood (PL) is the likelihood for the Probit model when the LD50 is held fixed and the likelihood is maximized over the slope. When fixing the LD50, the shape of the log likelihood surface as a function of the slope appears to have a roughly quadratic shape, making it easy to locate the maximum.

The profile likelihood function is a function of the LD50. The profile likelihood confidence region comprises the set of all LD50 values for which the profile likelihood is greater than a critical value. For a (1-a)% confidence interval the critical value is:

$$\text{CritLL} = \text{MaxLL} - (Z(1 - a/2)**2)/2 \quad (3)$$

where $Z(\cdot)$ is the inverse of the cumulative normal probability function.

The PL confidence region for the LD50 may be 1) a single interval bounded at both ends, 2) two disjoint intervals, each bounded on one end and infinite on the other (one going to positive infinity and one going to negative infinity), or 3) the entire range of LD50 values from minus infinity to positive infinity. When the region has two disjoint infinite intervals, one interval is associated with a positive slope and the other with a negative slope.

The OECD TG 425 procedure assumes that the test substance will have a positive dose-response relationship. It was therefore decided to report profile likelihood confidence intervals for which the slope was restricted to be positive. With this restriction, the confidence regions are either 1) a single finite bounded interval, 2) a single interval bounded on one side and infinite on the other, or 3) an infinite interval from negative to positive infinity.

The profile likelihood function may be one continuous curve or several disjoint continuous curves. Whether the function has disjoint sections depends on 1) the overall proportion of deaths in the data, 2) if there are one or more doses with partial response (i.e., case 4 or 5 data versus case 2 data), and 3) the response proportion at the partial dose (if any).

3. Overview of the Testing

A simulation module was created to evaluate the statistical performance of the TG 425 procedure. Given the true LD50 and true sigma used for the simulations, the simulation module calculates the probability of death given an animal's dose and calculates a random response, survival or death, based on that probability. Given the simulated animal's response, the program calculates the dose for the next animal. These steps continue until

one of the stopping criteria is met, at which time an LD50 estimate and a confidence interval are calculated.

The program's calculations were tested using two approaches:

- 1) For test data sets, the LD50 estimates and profile likelihood confidence intervals calculated by AOT425StatPgm were compared to estimates from other comparison programs.
- 2) The probability of different simulated outcomes were compared to simulations performed by EPA using the same simulation scenarios.

Each of these approaches to testing the program is discussed below. The use of test data sets is discussed first, followed by a discussion of the simulation tests.

4. Data Set Construction

The test data sets were constructed using three approaches:

- 1) Construct data sets to test all types of confidence intervals (finite, one-sided, and infinite) and all types of discontinuities in the profile likelihood functions;
- 2) Construct data sets that have selected parameter characteristics, and
- 3) Select a subset of simulated data sets.

Except when no data sets were found corresponding to a combination of characteristics, data sets were constructed for all combinations of the following characteristics:

- Case (2 to 5);
- Whether the confidence interval is finite, infinite, or one-sided
- Whether or not the overall response proportion was 0.5;
- For Case 4 and 5 data, whether or not the response proportion at the partial dose was 0.5.

These data sets were constructed to vary:

- The magnitude of the parameters;
- The side on which the confidence interval was finite, if the confidence interval was one sided; and
- Whether CritLL fell in the discontinuity (for profile likelihood functions with discontinuities).

Additional test data sets were constructed with very large or very small slopes, very wide confidence intervals, and large LD50 estimates.

Finally, simulated tests were sorted by stopping criteria, case, and number of animals. From each combination of stopping criteria and case, the simulated test with the smallest

and largest number of animals was selected. From this set, 15 tests were randomly selected for manual checking and comparison using other programs.

The final list had 48 data sets for which the AOT425StatPgm program calculations were compared to the results from the comparison programs. Those data sets are listed in appendix A.

5. Basis for Comparison

The statistical results from AOT425StatPgm were compared to similar values from BMDS, JMP, and Excel. The results that were compared are: Log(LD50), slope, Log-likelihood, Log(lower confidence bound) and Log(upper confidence bound).

For the Log(LD50), slope, Log(lower confidence bound) and Log(upper confidence bound), the absolute differences were compared. The absolute difference was calculated as:

Absolute value(AOT425StatPgm value – Comparison value)

For the log likelihood, the absolute relative difference was compared. The absolute relative difference is defined as:

Absolute value((AOT425StatPgm LL – Comparison LL)/AOT425StatPgm LL)

Differences greater than 0.0001 were investigated. Differences that could not be explained by deficiencies in the comparison programs were considered to indicate possible problems for the AOT425StatPgm program.

The comparison programs were:

- Probit regression fit using the JMP non-linear regression platform;
- BMDS, created by EPA, which calculates the lower profile likelihood confidence bound (labeled BMDL); and
- Excel, programmed to fit the probit model and profile likelihood. The confidence bounds were located through linear interpolation of the profile likelihood function.

These programs were selected because they all could be programmed to calculate the profile likelihood confidence intervals with relative ease. Although the programming in Excel is somewhat cumbersome, it has the advantage that intermediate results are easy to check for possible problems.

The correspondence between AOT425StatPgm results and results from other programs may be affected by differences in the 1) convergence criteria, 2) iteration method, 3) parameterization, 4) maximum number of iterations, 5) exact versus numerical derivatives, 6) calculation of the profile likelihood, and 7) precision of the output.

For all programs, an important convergence criterion is the relative change in the log likelihood from iteration to iteration. All programs have other convergence criteria. For the AOT425StatPgm program, iterations do not stop until the Log Likelihood convergence criteria is met and the slope is converging. The slope is judged as converging when the change in the slope on the last iteration is less than one half the change on the previous iteration. Excel was iterated by hand until the best convergence was achieved, corresponding roughly to a precise calculation of both the log likelihood and slope within the precision of the Excel functions. BMDS has two convergence criteria—one for the relative change in the log likelihood and the other for parameter convergence. BMDS acts as if convergence stops when the first criterion is met. JMP has three criteria and stops when the first criterion is met. In both JMP and BMDS, the convergence criteria can be changed. Only in JMP was the maximum number of iterations reached.

The primary iteration algorithms are the Newton-Raphson and Gauss-Newton. Although the Newton Raphson algorithm is better near the maximum likelihood solution, far from the solution the Gauss-Newton algorithm may be preferred. Newton-Raphson uses both first and second derivatives. Gauss-Newton uses only first derivatives. AOT425StatPgm and Excel used the Newton-Raphson algorithm. In JMP the user can easily select or change the algorithm used. In several cases, the algorithm or convergence criteria were modified to coax the program to converge. The BMDS algorithm apparently uses a variation of the Gauss-Newton algorithm (describes as quasi-Newton method that minimizes a general unconstrained objective function using an analytic gradient and a hessian [second derivative matrix] approximation from a secant update).

If the Newton-Raphson or Gauss-Newton parameter step (or change in parameters) from one iteration to the next does not increase the likelihood, step-halving can be used to find parameters for which the likelihood does increase. When the log-likelihood has a quadratic shape, step-halving is not required. Using the LD5 parameterization for which the log-likelihood surface does not have a quadratic shape, step-halving or some equivalent modification to the basic algorithm is usually needed. AOT425StatPgm, Excel, and JMP used step-halving. BMDS apparently used an equivalent procedure.

The probit model can be fit using either the linear or LD50 parameterization. The log likelihood surface using the linear parameterization has much better numerical properties than the LD50 parameterization and is likely to result in much faster convergence. When fitting the profile likelihood, the LD50 parameterization is efficient and easier to use than the linear parameterization. However, the BMDS software may use the linear parameterization to fit the profile likelihood. AOT425StatPgm and Excel used the linear parameterization to fit the LD50 and used the LD50 parameterization to fit the profile likelihood. JMP used only the LD50 parameterization, which may have contributed to convergence problems.

The precision of the calculation of the cumulative normal and inverse cumulative normal can affect the precision of the resulting estimates. The algorithms used in

AOT425StatPgm and BMDS were reported to be precise to one part in $1e15$. The built-in Excel function is accurate to about one part in $1e7$. The precision of the cumulative normal functions was evaluated by comparing different functions and comparing exact derivatives to numerical derivatives. The details of the evaluation are in appendix B. This approach provides a lower bound on the precision of the algorithms. For very extreme confidence bounds, the difference in the algorithms made a big difference in the estimated bounds.

AOT425StatPgm and Excel used exact derivatives and second derivatives. The BMDS and JMP programs used numerical derivatives (and second derivatives if needed). Exact derivatives are always preferred. Use of numerical derivatives apparently contributed to some convergence problems in JMP.

In effect, the AOT425StatPgm program calculates the profile likelihood restricted to positive slopes. Excel was used to calculate the confidence bounds for positive slopes only. There are no restrictions on the slope when JMP calculates the profile likelihood confidence bounds. This lack of a restriction might have contributed to difficulties converging when searching for the confidence bounds. Possible restrictions on the slope in BMDS are not documented. However, BMDS never failed to calculate a reasonable lower confidence bound.

Finally, the ability to compare results from difference programs is limited by the precision of the results reported by the program. In general, all values could be compared to at least 6 significant figures.

These characteristics of the different methods are summarized in Table 1.

Table 1 Characteristics of AOT425StatPgm and the Comparison Programs

	AOT425Stat Pgm	Excel	BMDS	JMP
Convergence criteria	Relative LL change = 1e-8 and slope converging	Relative LL change about 1e-15 and slope convergence	Relative LL change = 1e-8 or parameter convergence (1e-8)	Relative LL change = 1e-7 or parameter convergence
Iteration method	Newton Raphson	Newton Raphson	Variation on Gauss-Newton	Can select Newton-Raphson or Gauss Newton
Maximum number of iterations (Default)	40	Manually determined	250, Can be changed	50, Can be changed
Parameterization for calculating the LD50	Linear	Linear	Linear	LD50
Parameterization for calculating the profile likelihood	LD50	LD50	Not determined	LD50
Calculation of P from Z, precision	1 part in 1e15	1 part in 1e7	1 part in 1e15	Better than 1 part in 1e11
Exact versus numerical derivatives	Exact	Exact	Numerical	Numerical
Calculation of profile likelihood	Restricted to positive slope	Restricted to positive slope as used	Apparently restricted to positive slope	No restrictions
Typical output precision, Log LD50	8 digits	15 digits	6 digits	10 digits
Typical output precision, Confidence bound	8 digits	4 digits	6 digits	7 digits

6. Results for the Test Data Sets

The parameter estimates for the different data sets and programs are in Tables 2 through 5 at the end of this report. Based on the comparison of the parameter estimates, the conclusions are:

- The log likelihood at the maximum is very close in all programs (the absolute relative differences in the log likelihood values are less than 7e-6 and average less than 1e-6).
- If the comparison programs calculate the profile likelihood, the profile likelihood confidence bounds are very close except in a few cases where the differences can be explained by deficiencies in the comparison programs.
- The log LD50 and slope estimates agree well except in a few cases where the difference appears to be related to deficiencies in the comparison programs, or one case where the AOT425StatPgm slope is slightly different from the Excel slope and the excel slope is considered more accurate. In this case, the AOT425StatPgm slope was more accurate than the slope from the other programs.
- Thus, the tests provide no basis for concluding that the calculations in the AOT425StatPgm program need modification.

6.1 Comparison of the Confidence Bounds

JMP often does not calculate a profile likelihood confidence bound when other programs do. Also, JMP does not try to calculate a confidence bound if the LD50 convergence criteria are not met. JMP rarely considered the results converged if the slope was infinite. Therefore, JMP often did not provide estimates of the confidence bounds. When JMP did calculate a confidence bound, it agreed with the AOT425StatPgm bound, except in one case. The largest difference (on the log scale) is 0.0009 for one case, where JMP's convergence criteria were relaxed to coax JMP to converge.

BMDS only calculates the lower confidence bound (BMDL). BMDS also fits a background response. For a few cases, the background response estimate was non-zero. In these cases, the lower confidence bound was not used. In all but one case, the confidence bounds agreed. In one case, the confidence bound calculated by AOT425StatPgm was very large in magnitude (log of lower bound = -83.7). The log confidence bound calculated by BMDS was large but very different (-49.6). Far from the LD50 the confidence bound can be very sensitive to convergence criteria, and round-off error in MaxLL and the critical value from the inverse cumulative normal function. This difference was not judged to be important.

Except for three bounds whose absolute magnitude was greater than 40 on the log scale, the confidence bounds determined using Excel agree with those from the ATO425 program. In the three problem cases, the confidence bound determined in Excel was based on Excel's built-in function to calculate the inverse cumulative normal and CritLL. When the value corresponding to the more precise function in the AOT425StatPgm program was used, the agreement between the confidence bounds was greatly improved. The differences between Excel and the ATO425StatPgm program are attributed to either slightly different convergence, the differences in the inverse cumulative normal functions and round-off error. This difference was not judged to be important.

6.2 Comparison of the LD50 and Slope

The LD50 and slope estimates from AOT425StatPgm agreed with those from Excel with two exceptions. For data sets AI and AK the slope converges slowly. For these data sets, more iterations were performed in Excel to get what was hoped to be the best possible estimates of the parameters. The AOT425StatPgm program used fewer iterations. The differences in the slope estimates are 0.002 and 0.035 for slope estimates of 20.1 and 23.9. These differences are small, and much smaller than the differences between Excel and the other comparison programs (for which the minimum difference is 0.15). These differences are not judged to be important.

For data with an infinite slope, BMDS generally fit a slope of about 20 before stopping iterations. BMDS did not recognize that the true slope was infinite. Excluding data sets with infinite slopes and the AI and AK data sets discussed above, the AOT425StatPgm LD50 and slope estimates agreed with the BMDS with one exception. Data set AJ was

constructed to have a small slope. The slope estimate from Excel, JMP, and AOT425StatPgm is 0.0143. Using the default convergence criteria, the slope estimate from BMDS is 0.0077 (equal to BMDS's initial parameter estimate). When the BMDS convergence criteria were set to 1e-10, the parameter estimates agreed with those from the other programs.

The AOT425StatPgm and JMP estimates of LD50 and slope generally agreed to within 0.07 except for the AI and AK data sets discussed above. In some cases, better estimates could be achieved by changing the convergence criteria. In other cases, this was not true. The use of numerical derivatives may have affected the precision of the estimates in JMP.

Except for the two cases where the Excel slope is believed to be more precise than the AOT425StatPgm slope, the AOT425StatPgm program is judged to provide estimates as good or better than the comparison programs.

7. Comparison of the AOT425StatPgm Simulations with EPA Simulations

EPA reported simulation results for the OECD TG 425 procedure in "LD50 Confidence Bounds for Guideline 425: Statistical Approach and Performance Characteristics" by David Farrar, dated January 19, 2001. The summary tables presented:

- The percentage of simulated tests classified as Case 2, Case 3, Case 4, or Case 1 or Case 5;
- The percentage of Case 2, 3, and 4 data with finite confidence intervals (PM1);
- The coverage of the 95% confidence intervals (PM2) for Case 2, 3, and 4 data; and
- The median ratio of the upper to lower confidence bounds (PM3).

David Farrar also defined the PM4 measure as the width of the confidence interval as a percentage of the true LD50. Because PM3 and PM4 are similar and provide similar conclusions regard the comparison of the AOT425StatPgm and EPA simulations, only the values for PM3 are presented.

There is one difference between how the AOT425StatPgm program made calculations and how the simulations were performed by EPA. In the AOT425StatPgm program, doses that are close to the limit dose (or the smallest usable dose) relative to the assumed $\sigma/2$ are dropped from the dose progression. In the EPA simulations, no doses were dropped from the dose progression. The AOT425StatPgm program does not drop a dose from the dose progression for 52 of the 112 simulation scenarios used by EPA. The corresponding summary statistics were compared for these 52 scenarios.

Table 6, at the end of this report, shows the percentage of simulated tests by case. Table 6 is comparable to Table A.1 in David Farrar's paper (starting with Scenario 46). Table 6 shows values of PM1 and PM2 for all scenarios. Cases with an asterisk in the last column are cases for which the dose sequence used in the EPA simulations and the AOT425StatPgm simulations are the same. The statistical results were compared for these cases. Table 6 shows values of PM3 for only those cases for which the statistical results are compared and for which PM3 is finite. Figure 1 shows a plot of the percentages from the AOT425StatPgm simulations and from the EPA simulations. Although the percentages generally agree, a chi-square test suggests there may be differences between the simulations. The p-value comparing the percentage of simulations by case is 0.018 (Chi-square = 157 with 122 degrees of freedom). The calculation of the chi-square test is described in Appendix B.

Figure 1 Comparison of the Percentage of Simulations by Case

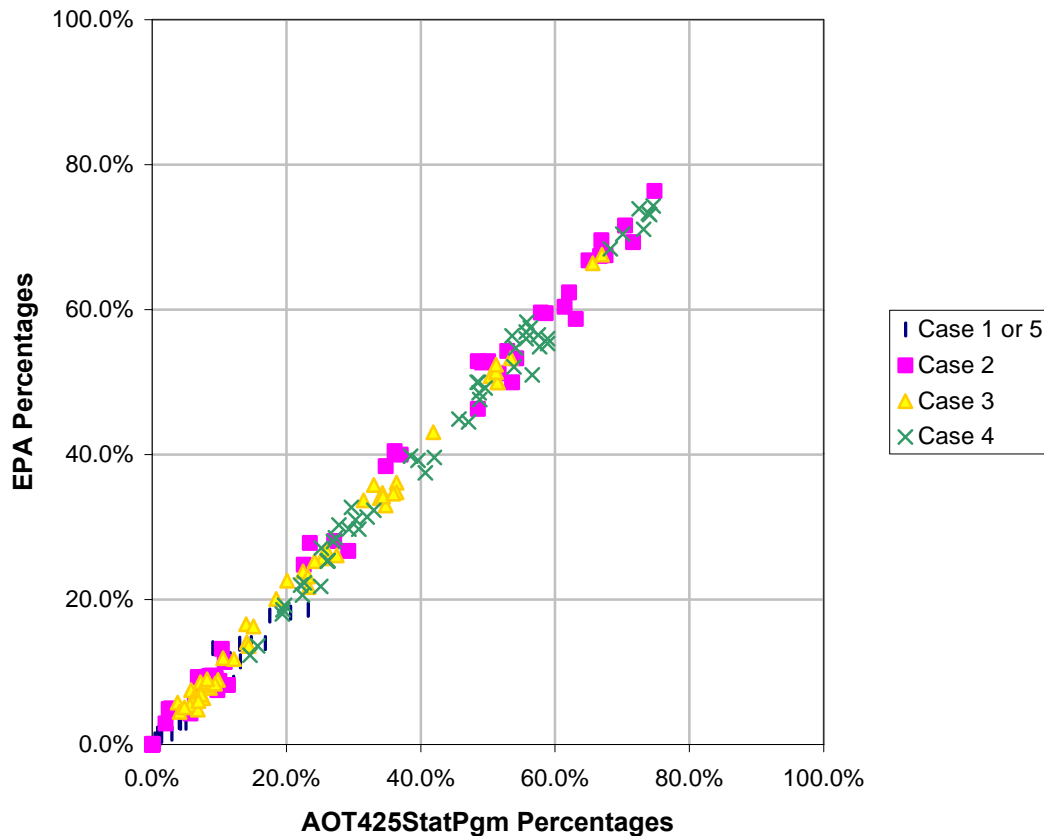


Table 6 shows the AOT425StatPgm values of PM1, PM2, and PM3. The percentage of finite confidence intervals (PM1) and the coverage (PM2) agree reasonably well between the AOT425StatPgm and EPA simulations based on the plot. No formal chi-square test was performed to compare the percentages. Figure 2 shows a plot of the PM1 and PM2 percentages.

For some scenarios the coverage of the 95% confidence intervals is well below 95%. These scenarios generally have small values for the true slope (i.e., the assumed slope is greater than the true slope).

Figure 2 Comparison of the PM1 and PM2

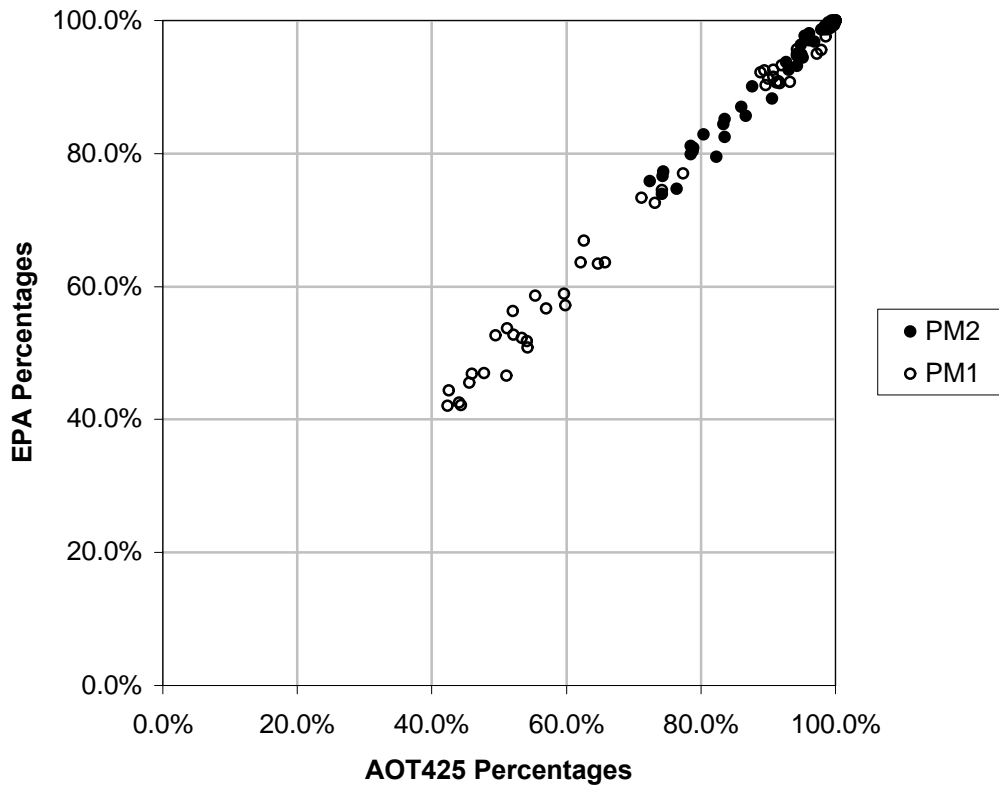
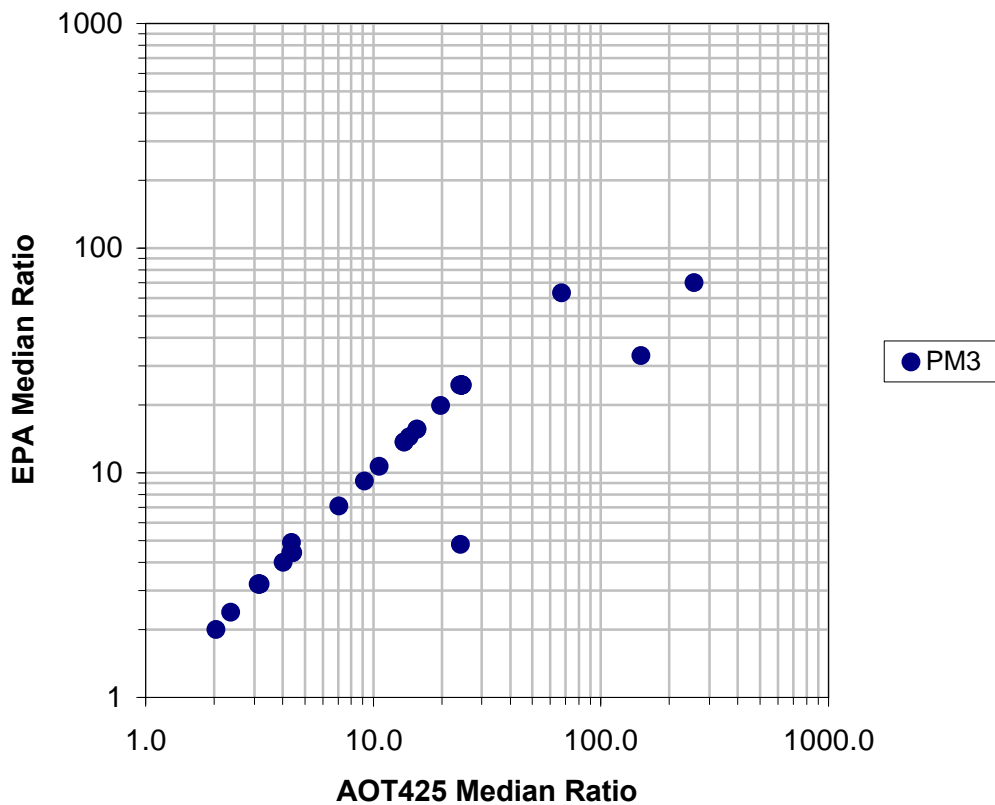


Figure 3 shows a plot of the median ratio of the upper to lower confidence bounds (infinite bounds are included in the median calculations as a large number). EPA apparently calculated the median using the bounds in dose units. For the AOT425StatPgm data, the median was based on the log of the ratio of the bounds. The median length was then transformed back to the dose units. The difference in how the median was calculated is not expected to affect the comparison. Except for four scenarios, there is good agreement between the median ratio of the upper to lower confidence bound from the AOT425StatPgm and EPA simulations.

Three of the four scenarios for which there is a large difference are shown in Figure 3. For the fourth scenario the median AOT425StatPgm ratio was infinite and the EPA ratio was very large (865). For the scenarios for which there are large differences, the simulated ratios bunch at discrete values. Differences between the discrete values in each scenario's distribution, as well as random variation in the location of the median relative to the change from one discrete value to another, may explain the differences. Based on approximate variance calculations, this explanation is possible, but appears unlikely.

Figure 3 Comparison of the PM3



Overall, the AOT425StatPgm and EPA simulations show similar patterns in the percentage of simulated tests in difference cases, the percentage of finite confidence intervals, the coverage of the 95% confidence intervals, and the median ratio of the upper to the lower confidence bound. However, a more detailed look at the differences suggests that there may be small differences between the simulations that have not been accounted for. David Farrar noted that his simulations did not restrict the slope to be positive when calculating the profile likelihood. There may be other, as yet unrecognized differences between the simulations that explain the slightly-more-than-expected differences.

8. Other Testing of the AOT425StatPgm Calculations

SAS programs were written to check the following characteristics of the simulated data:

- 1) The calculation of the next dose and when doses are dropped from the dose sequence
- 2) The stopping criteria used within the simulations to stop dosing;
- 3) The probability of survival versus dose (to check the simulation of survival or death).

No problems were found.

Simulations were performed using a very small value for the standard deviation of the log dose to see if there were any computational errors. Convergence problems were expected to be most acute when doses are very close. No problems were found.

The last modifications to the computational algorithms were made before completing the draft version of the program. Since then, millions of simulated tests using a variety of scenarios have been performed without computational errors.

Additional test data sets have been used to test error messages, dose recommendations, and stopping criteria.

Furthermore, the user interface has been tested by specifically testing various items such as:

- file saving,
- screen color settings,
- screen resolution,
- smallest usable dose,
- limit dose,
- warning messages,
- confidence interval selection,
- menus,
- files formats,
- header change warnings,

- extreme input values,
- entering unusual characters into each field (every character on the keyboard was checked),
- unusual patterns of switching between windows and window options,
- invalid input data,
- turning off the computer while the program was running,
- printing on different printers, and
- printing many files.

Finally, the program was tested by asking the tester to try to "break the program," i.e., try to get the program to provide an error message for an error that was not handled properly, to provide nonsensical results, or to fail and quit (as in, "Your program has performed an illegal operation and will be shut down").

9. Test Data Set Results

Table 2 AOT425StatPgm Parameter Estimates for Test Data

Data Set	LogLD50	Slope	Log Likelihood	Upper Bound	Lower Bound
B	0.150515	2.91515875	-3.21451799	-INF	INF
C	-0.10971942	4.56362805	-3.98460164	-0.37577585	0.17018444
D	2.89640427	2.31924327	-5.7493211	-INF	INF
E	3.01601776	3.1657783	-4.77727373	2.59239384	3.63009457
F	-0.07080445	4.0634154	-4.21998812	-INF	0.16661155
G1	0.22880048	2	-3.31724808	0	0.30103
G2	0.31851058	1	-4.00190346	0	0.30103
H	2.30103	INF	-1.38629436	-INF	INF
I	0.30103	INF	-1.38629436	-0.0159741	0.61803409
K	2.30103	INF	-2.77258872	2.1119837	2.54617563
N	1	INF	-1.9095425	0.57189215	1.59091781
O	0	INF	-2.87081423	-0.20558287	0
P	1	INF	-1.9095425	-INF	INF
Q	1	INF	-1.9095425	0.16995403	2.38384109
R	0	INF	-3.01416129	-0.24222278	0
S	1	INF	-1.9095425	0.78361461	INF
T	0	INF	-2.87081423	-INF	0
W	1	INF	-1.38629436	-INF	INF
X	1	INF	-1.38629436	0.85397294	5.89740333
Y	0.60205999	INF	-2.77258872	0.37301975	INF
Z	0.30103	INF	-2.24934058	-INF	INF
AA	0.30103	INF	-2.50201212	0.02168438	0.36743564
AB	0	INF	-3.01416129	0	0.20855028
AC	0.30103	INF	-1.38629436	-INF	INF
AD	0.30103	INF	-2.24934058	0.02588278	0.5198684
AE	0	INF	-5.58119929	0	0.25467288
AF	0	INF	-2.70336725	-INF	0.01074094
AG	0	INF	-3.01416129	-INF	0
AH	0.31700125	9.1343752	-1.75616521	0.14242791	0.46944941
AI	0.30103001	20.1043152	-1.38629472	0.13753475	0.45827717
AJ	0.89894818	0.01431379	-9.70372937	-INF	INF
AK	2.26582802	23.9083557	-2.50201214	2.06995071	2.35958034
AL	0.41982475	3.17902788	-5.01072116	-83.7106913	84.56797927
AM	0.35843352	3.17567935	-4.25724255	-41.94750703	1.25069194
Large LD50	4.74529574	0.42920992	-2.84161672	2.27804989	INF
Sim1	2.63765621	0.90259254	-8.7440285	-INF	INF
Sim2	2.31698626	4	-1.80218725	2.24303805	2.49136169
Sim3	0.35595341	4.06479506	-3.30199094	-0.55503385	0.68803894
Sim4	0.74036269	INF	-3.81908501	0.48440747	1.32321043
Sim5	0.81803024	2.99781782	-5.04509803	0.38844865	1.17669187
Sim7	2.49136169	INF	-1.9095425	-INF	INF
Sim8	0.74036269	INF	-1.9095425	0.40692993	0.92777653
Sim9	2.28931452	2.87174207	-2.72528117	-INF	INF
Sim10	0	INF	-2.50201212	-0.09789187	0.38453848
Sim11	0.1775768	2	-2.69483615	0.03554166	0.34618935
Sim13	-6.11107182	0.17611389	-3.11561195	-INF	0.35405199
Sim14	0	INF	-5.29250591	-7.48022608	0.20059392
Sim15	4.22895092	1.26708262	-2.63084068	3.34905575	INF

Table 3 Excel Parameter Estimates for Test Data

Data Set	LogLD50	Slope	Log Likelihood	Upper Bound	Lower Bound
B	0.150515	2.91515905	-3.214518046	-INF	INF
C	-0.10971944	4.56362778	-3.984601366	-0.3758	0.1702
D	2.89640435	2.31924314	-5.749321063	-INF	INF
E	3.01601778	3.16577764	-4.777273715	2.5924	3.6301
F	-0.07080447	4.06341574	-4.219988138	-INF	0.1666
G1	0.2288005	2	-3.317248056	0	0.301029996
G2	0.3185106	1	-4.001904098	0	0.301029996
H	2.30103	INF	-1.386294361	-INF	INF
I	0.30103	INF	-1.386294361	-0.01597	0.618
K	2.30103	INF	-2.772588722	2.112	2.5462
N	1	INF	-1.909542505	0.5719	1.5909
O	0	INF	-2.870814228	-0.2056	0
P	1	INF	-1.909542505	-INF	INF
Q	1	INF	-1.909542505	0.17	2.3838
R	0	INF	-3.01416129	-0.2422	0
S	1	INF	-1.909542505	0.7836	INF
T	0	INF	-2.870814228	-INF	0
W	1	INF	-1.386294361	-INF	INF
X	1	INF	-1.386294361	0.854	5.8972
Y	0.60205999	INF	-2.772588722	0.373	INF
Z	0.30103	INF	-2.249340578	-INF	INF
AA	0.30103	INF	-2.502012118	0.02169	0.3674
AB	0	INF	-3.01416129	0	0.2086
AC	0.30103	INF	-1.386294361	-INF	INF
AD	0.30103	INF	-2.249340578	0.02588	0.5199
AE	0	INF	-5.581199285	0	0.2547
AF	0	INF	-2.703367253	-INF	0.0107
AG	0	INF	-3.01416129	-INF	0
AH	0.31700124	9.13437519	-1.756165798	0.1424	0.4694
AI	0.30103001	20.1060543	-1.386294718	0.1375	0.4583
AJ	0.89894826	0.01431371	-9.703729373	-INF	INF
AK	2.26588052	23.9440595	-2.502012136	2.07	2.3596
AL	0.41982479	3.17902793	-5.010721311	-74.5133	75.3706
AM	0.35843356	3.17567904	-4.25724286	-41.8997	1.2507
Large LD50	4.74529532	0.42921003	-2.841616753	2.2781	INF
Sim1	2.63765628	0.90259231	-8.744028553	-INF	INF
Sim2	2.31698628	4	-1.802187163	2.243038049	2.491361694
Sim3	0.35595343	4.06479506	-3.30199111	-0.555	0.688
Sim4	0.74036269	INF	-3.81908501	0.4844	1.3232
Sim5	0.81803021	2.99781775	-5.045097551	0.3885	1.1767
Sim7	2.49136169	INF	-1.909542505	-INF	INF
Sim8	0.74036269	INF	-1.909542505	0.4069	0.9278
Sim9	2.28931453	2.87174177	-2.725291354	-INF	INF
Sim10	0	INF	-2.502012118	-0.0979	0.3845
Sim11	0.17757682	2	-2.694836578	0.035541659	0.346189352
Sim13	-6.11107437	0.17611387	-3.115611834	-INF	0.3541
Sim14	0	INF	-5.292505905	-7.4801	0.2006
Sim15	4.22895075	1.26708275	-2.630840578	3.3491	INF

- Shaded values in the slope column are probably more accurate than the corresponding value from AOT425StatPgm.
- Differences between the shaded values in the confidence bound columns and those from the AOT425StatPgm program appear to be due to use of the Excel's inverse cumulative normal function that is less precise than the AOT425StatPgm function.

Table 4 BMDS Parameter Estimates for Test Data

Data Set	LogLD50	Slope ²	Log Likelihood	Upper Bound
B	0.150516975	2.915141805	-3.21452	-INF
C	-0.109719856	4.563654577	-3.9846	-0.375775676
D	2.896407156	2.319255809	-5.74932	-INF
E	3.016017406	3.165778193	-4.77727	2.592393295
F	-0.07080441	4.063417965	-4.21999	-INF
G1				
G2				
H	2.301029996	19.08295027	-1.38629	-INF
I	0.301029996	19.48081395	-1.38629	-0.015974328
K	2.301029996	19.06300988	-2.77259	2.111984617
N	1.02820512	16.82614057	-1.90954	0.571891593
O	-0.046896351	22.76434634	-2.87081	-0.205583064
P	1.025285379	17.03417913	-1.90954	-INF
Q	1.025617135	16.81561775	-1.90954	0.169953492
R	-0.051049878	22.53376547	-3.01416	-0.242222992
S	1.024900061	17.29706527	-1.90954	0.783614906
T	-0.049021627	21.7775735	-2.87081	-INF
W	1	18.69722121	-1.38629	-INF
X	1	19.83124437	-1.38629	0.853973056
Y	0.602059991	18.53097457	-2.77259	0.373020563
Z	0.26860605	20.80222148	-2.24934	-INF
AA	0.262299179	21.72940342	-2.50201	0.021685352
AB				
AC	0.301029996	18.80739991	-1.38629	-INF
AD	0.269171913	21.17167126	-2.24934	0.025883175
AE				
AF				
AG				
AH	0.317001357	9.13437809	-1.75617	0.142427012
AI	0.301029996	19.95848522	-1.38629	0.137534502
AJ	0.898799751	0.007704427	-9.7038	-INF
AK	2.262213705	21.68215437	-2.50201	2.069949151
AL	0.419825275	3.179041083	-5.01072	-49.4558389
AM	0.358433618	3.175679309	-4.25724	-41.94752538
Large LD50	4.74529735	0.429208769	-2.84162	2.278049016
Sim1	2.63765581	0.902592633	-8.74403	-INF
Sim2				
Sim3	0.35595315	4.064799516	-3.30199	-0.555034477
Sim4	0.802743631	6.904761918	-3.81909	0.484408044
Sim5	0.818029823	2.997827636	-5.0451	0.388449351
Sim7	2.473206595	23.72422499	-1.90954	-INF
Sim8	0.722489134	24.10046739	-1.90954	0.406930019
Sim9	2.289314645	2.871738076	-2.72529	-INF
Sim10	0.031788053	26.47696547	-2.50201	-0.097892085
Sim11				
Sim13				
Sim14				
Sim15	4.228949297	1.267089551	-2.63084	3.349055952

- Data sets with no data are either Case 3 data or data for which BMDS fit a positive background response.
- "-INF" in the lower bound column corresponds to the BMDS message "Lower limit includes zero."
- For data sets AI and AK, BMDS apparently stopped converging before the ATO425 program stopped converging. Poor agreement is associated with very large slopes.
- For data set AJ, the only possible explanation for the poor agreement between BMDS and AOT425StatPgm is the use of numerical derivative in BMDS.
- For data set AL, agreement for the extreme confidence bound is sensitive to convergence criteria.

²The BMDS program uses the natural logarithm. In this report, the slope reported for the BMDS program is the slope reported in the BMDS report multiplied by ln(10) to make the value comparable to the other programs.

Table 5 JMP Parameter Estimates for Test Data

Data Set	LogLD50	Slope	Log Likelihood	Upper Bound	Lower Bound
B					
C	-0.109729005	4.564130246	-3.984601664	-0.3757779	0.17018487
D					
E	3.016018115	3.165760441	-4.777273733	2.5923923	3.63010357
F	-0.070606641	4.065588194	-4.219988571		0.1666115
G1					
G2					
H					
I					
K					
N					
O					
P					
Q					
R					
S					
T					
W					
X					
Y					
Z					
AA					
AB					
AC					
AD					
AE					
AF					
AG					
AH	0.316990936	9.120572155	-1.75616815	0.14243206	
AI	0.301004756	19.75124695	-1.386294883	0.13756266	
AJ	0.898948182	0.014313789	-9.703729373		
AK	2.263917736	22.67682695	-2.502012141	2.06995016	
AL	0.419824591	3.179031158	-5.010721156		
AM	0.358437413	3.175753734	-4.257242554		
Large LD50					
Sim1	2.637656206	0.902611718	-8.744028498		
Sim2					
Sim3	0.355952932	4.063682252	-3.301991035		0.68803979
Sim4					
Sim5	0.818186011	2.995360224	-5.045100077	0.38844506	1.17669216
Sim7					
Sim8					
Sim9					
Sim10					
Sim11					
Sim13	-6.176620444	0.174415939	-3.115614231		0.35491724
Sim14					
Sim15	4.230177377	1.265076912	-2.630841124		

- JMP did not converge for many data sets. Modeling using JMP was not attempted for some case 4 and case5 data sets (with infinite slopes). For these data sets convergence was generally not achieved.
- For Data sets AH, AI, and AK, the parameters in the shaded areas have relatively poor agreement with the AOT425StatPgm program, apparently due to poorer convergence.
- Poor agreement for data set Sim13 is due to relaxing convergence criteria to get JMP to fit the confidence bound.
- Poor agreement between JMP and AOT425StatPgm may be due to JMPs implementation of numerical derivatives.

10. Summary Statistics for the Test Simulations

Table 6 AOT425StatPgm Test Simulation Results, Percentage by Case

TrueLD50	TrueSlope	Dose0	Case1&5	Case1	Case2	Case3	Case4	
1.5	8.33	1.1	20.3%	0.1%	0.0%	79.6%	0.1%	
1.5	8.33	1.5	19.8%	0.0%	0.0%	80.2%	0.0%	
1.5	8.33	1.9	7.5%	0.0%	8.6%	41.9%	42.0%	*
1.5	4	1.5	55.0%	1.7%	4.6%	37.6%	2.8%	
1.5	4	2.4	23.2%	1.6%	23.5%	22.5%	30.8%	*
1.5	2	1.5	45.6%	6.3%	30.1%	14.7%	9.6%	
1.5	2	4.0	32.5%	5.1%	38.5%	14.0%	15.0%	
1.5	0.8	1.5	36.4%	14.8%	47.0%	6.5%	10.1%	
1.5	0.8	16.9	22.0%	9.0%	56.1%	6.3%	15.6%	
1.5	0.5	1.5	36.1%	18.2%	48.0%	4.5%	11.4%	
1.5	0.5	72.3	10.0%	7.3%	66.7%	3.8%	19.5%	*
2.5	8.33	1.8	0.0%	0.0%	0.0%	67.0%	33.0%	*
2.5	8.33	2.5	0.0%	0.0%	0.0%	27.5%	72.5%	*
2.5	8.33	3.2	0.0%	0.0%	0.0%	56.2%	43.8%	
2.5	4	1.2	9.4%	0.5%	6.1%	40.9%	43.6%	
2.5	4	2.5	0.4%	0.0%	11.3%	20.1%	68.2%	*
2.5	4	4.1	6.0%	0.0%	9.1%	47.9%	37.0%	
2.5	2	2.5	4.0%	1.1%	34.8%	14.1%	47.1%	*
2.5	2	6.6	2.9%	0.9%	37.0%	14.4%	45.7%	*
2.5	0.8	2.5	11.6%	6.0%	54.2%	7.2%	27.0%	*
2.5	0.8	28.2	6.7%	3.8%	61.4%	5.8%	26.1%	*
2.5	0.5	2.5	14.7%	9.2%	52.9%	6.2%	26.2%	*
2.5	0.5	120.6	7.8%	5.3%	66.0%	5.8%	20.4%	
20	8.33	14.0	0.0%	0.0%	0.3%	73.5%	26.2%	
20	8.33	20.0	0.0%	0.0%	0.0%	25.4%	74.6%	*
20	8.33	25.2	0.0%	0.0%	0.1%	51.5%	48.4%	*
20	4	9.6	0.0%	0.0%	8.6%	33.9%	57.5%	*
20	4	20.0	0.0%	0.0%	2.8%	23.3%	73.9%	*
20	4	32.5	0.0%	0.0%	9.0%	36.2%	54.8%	
20	2	4.6	0.0%	0.0%	41.2%	13.9%	44.9%	
20	2	20.0	0.0%	0.0%	27.1%	14.0%	58.9%	*
20	2	52.7	0.0%	0.0%	34.2%	12.4%	53.4%	
20	0.8	20.0	0.1%	0.1%	51.6%	9.8%	38.5%	*
20	0.8	225.5	0.0%	0.0%	62.1%	8.6%	29.3%	*
20	0.5	20.0	0.9%	0.6%	58.6%	8.5%	32.0%	*
20	0.5	964.4	0.8%	0.6%	70.4%	6.4%	22.4%	*
50	8.33	35.1	0.0%	0.0%	0.1%	72.3%	27.6%	
50	8.33	50.0	0.0%	0.0%	0.0%	24.4%	75.6%	
50	8.33	63.1	0.0%	0.0%	0.0%	50.4%	49.6%	*
50	4	23.9	0.0%	0.0%	9.5%	36.4%	54.1%	*
50	4	50.0	0.0%	0.0%	3.6%	22.2%	74.2%	
50	4	81.2	0.0%	0.0%	10.0%	36.4%	53.6%	*
50	2	11.4	0.0%	0.0%	36.2%	15.9%	47.9%	
50	2	50.0	0.0%	0.0%	26.1%	15.9%	58.0%	
50	2	131.8	0.0%	0.0%	33.5%	15.1%	51.4%	
50	0.8	1.3	1.9%	1.3%	66.1%	6.2%	25.8%	
50	0.8	50.0	0.1%	0.0%	48.7%	9.1%	42.1%	
50	0.8	563.6	0.0%	0.0%	63.1%	7.2%	29.7%	*
50	0.5	50.0	0.6%	0.6%	55.5%	8.6%	35.3%	
50	0.5	2411.1	0.8%	0.7%	66.9%	7.2%	25.1%	*
150	8.33	105.3	0.0%	0.0%	0.0%	75.6%	24.4%	
150	8.33	150.0	0.0%	0.0%	0.0%	23.9%	76.1%	
150	8.33	189.3	0.0%	0.0%	0.0%	51.4%	48.6%	*
150	4	71.7	0.0%	0.0%	8.9%	34.8%	56.3%	*
150	4	150.0	0.0%	0.0%	4.9%	24.2%	70.9%	
150	4	243.5	0.0%	0.0%	6.8%	34.3%	58.9%	*

Table 6 AOT425StatPgm Test Simulation Results, Percentage by Case (Cont.)

TrueLD50	TrueSlope	Dose0	Case1&5	Case1	Case2	Case3	Case4	
150	2	34.3	0.0%	0.0%	37.8%	14.2%	48.0%	
150	2	150.0	0.0%	0.0%	23.0%	18.1%	58.9%	
150	2	395.3	0.0%	0.0%	33.9%	13.5%	52.6%	
150	0.8	3.8	0.0%	0.0%	71.2%	6.8%	22.0%	
150	0.8	150.0	0.0%	0.0%	51.1%	8.2%	40.7%	
150	0.8	1690.9	0.7%	0.4%	60.6%	7.1%	31.6%	
150	0.5	150.0	0.6%	0.2%	57.8%	7.0%	34.6%	
600	8.33	421.0	0.0%	0.0%	0.0%	72.1%	27.9%	
600	8.33	600.0	0.0%	0.0%	0.0%	25.9%	74.1%	*
600	8.33	757.2	0.0%	0.0%	0.1%	51.2%	48.7%	*
600	4	286.9	0.0%	0.0%	10.8%	31.5%	57.7%	*
600	4	600.0	0.0%	0.0%	5.7%	24.2%	70.1%	*
600	4	974.0	0.0%	0.0%	10.4%	33.0%	56.6%	*
600	2	137.2	0.0%	0.0%	38.9%	13.8%	47.3%	
600	2	600.0	0.0%	0.0%	29.2%	15.1%	55.7%	*
600	2	1581.1	0.1%	0.1%	32.9%	13.5%	53.5%	
600	0.8	15.0	0.7%	0.2%	66.2%	6.9%	26.2%	
600	0.8	600.0	1.4%	0.9%	48.5%	9.4%	40.7%	*
600	0.5	1.6	2.5%	1.9%	79.3%	3.5%	14.7%	
600	0.5	600.0	4.2%	2.8%	57.9%	7.6%	30.3%	*
1500	8.33	1052.6	0.0%	0.0%	0.0%	71.4%	28.6%	
1500	8.33	1500.0	0.0%	0.0%	0.0%	25.7%	74.3%	
1500	8.33	1892.9	0.0%	0.0%	0.0%	51.2%	48.8%	*
1500	4	717.3	0.1%	0.0%	9.7%	34.4%	55.8%	*
1500	4	1500.0	0.2%	0.0%	4.8%	22.9%	72.1%	
1500	4	2435.0	0.0%	0.0%	8.4%	35.9%	55.7%	*
1500	2	343.0	0.6%	0.5%	38.8%	14.2%	46.4%	
1500	2	1500.0	3.5%	0.4%	28.0%	13.4%	55.1%	
1500	2	3952.8	8.5%	1.5%	35.8%	17.9%	37.8%	
1500	0.8	37.5	5.6%	2.4%	67.6%	7.3%	19.5%	
1500	0.8	1500.0	10.4%	6.1%	48.2%	9.3%	32.1%	
1500	0.5	4.1	4.4%	2.5%	76.7%	4.3%	14.6%	
1500	0.5	1500.0	13.7%	9.1%	54.2%	7.8%	24.3%	
3000	8.33	2105.1	6.3%	0.0%	2.0%	65.6%	26.1%	*
3000	8.33	3000.0	1.1%	0.1%	2.5%	23.2%	73.2%	*
3000	8.33	3785.8	8.2%	0.0%	0.0%	91.7%	0.1%	
3000	4	1434.6	30.5%	0.8%	13.9%	38.3%	17.3%	
3000	4	3000.0	5.0%	2.3%	22.6%	18.5%	53.9%	*
3000	4	4870.0	35.2%	1.8%	12.3%	35.5%	17.0%	
3000	2	686.0	13.1%	2.9%	48.5%	10.6%	27.8%	*
3000	2	3000.0	12.1%	5.1%	36.1%	12.2%	39.6%	*
3000	0.8	75.0	7.2%	3.2%	67.5%	5.6%	19.7%	*
3000	0.8	3000.0	16.8%	9.8%	49.1%	6.8%	27.3%	*
3000	0.5	8.2	5.4%	3.4%	74.8%	4.1%	15.7%	*
3000	0.5	3000.0	20.6%	13.6%	50.0%	7.3%	22.1%	*
3500	8.33	2456.0	17.5%	0.2%	6.4%	53.4%	22.7%	*
3500	8.33	3500.0	2.8%	0.5%	11.6%	20.9%	64.7%	
3500	8.33	4416.8	24.5%	0.0%	0.0%	75.3%	0.2%	
3500	4	1673.7	39.2%	1.7%	16.8%	29.6%	14.4%	
3500	4	3500.0	7.6%	3.3%	25.9%	17.0%	49.5%	
3500	2	800.4	13.0%	4.8%	53.6%	8.2%	25.2%	*
3500	2	3500.0	13.6%	7.0%	38.1%	11.2%	37.1%	
3500	0.8	87.5	8.7%	4.3%	65.0%	6.9%	19.4%	*
3500	0.8	3500.0	20.5%	12.0%	46.1%	5.4%	28.0%	
3500	0.5	9.6	9.0%	6.7%	71.6%	4.8%	14.6%	*
3500	0.5	3500.0	20.9%	13.3%	47.1%	6.6%	25.4%	

Table 7 AOT425StatPgm Test Simulation Results, PM1, PM2, PM3

TrueLD50	TrueSlope	Dose0	PM1	PM2	PM3	
1.5	8.33	1.1	100.0%	100.0%		
1.5	8.33	1.5	100.0%	100.0%		
1.5	8.33	1.9	92.0%	99.1%	9.1	*
1.5	4	1.5	93.1%	99.8%		
1.5	4	2.4	74.2%	97.9%	14.4	*
1.5	2	1.5	51.8%	94.1%		
1.5	2	4.0	54.2%	95.7%		
1.5	0.8	1.5	34.1%	86.6%		
1.5	0.8	16.9	41.8%	81.7%		
1.5	0.5	1.5	33.8%	81.1%		
1.5	0.5	72.3	46.0%	76.4%		*
2.5	8.33	1.8	100.0%	99.9%	3.2	*
2.5	8.33	2.5	100.0%	100.0%	15.6	*
2.5	8.33	3.2	100.0%	98.9%		
2.5	4	1.2	94.7%	95.8%		
2.5	4	2.5	88.9%	99.6%	15.6	*
2.5	4	4.1	94.5%	99.3%		
2.5	2	2.5	64.7%	94.9%	19.8	*
2.5	2	6.6	65.7%	95.2%	24.4	*
2.5	0.8	2.5	42.5%	90.6%		*
2.5	0.8	28.2	49.5%	82.3%		*
2.5	0.5	2.5	44.1%	86.0%		*
2.5	0.5	120.6	47.4%	76.9%		
20	8.33	14.0	99.7%	99.8%		
20	8.33	20.0	100.0%	100.0%	24.4	*
20	8.33	25.2	99.9%	99.4%	3.1	*
20	4	9.6	91.5%	95.4%	4.4	*
20	4	20.0	97.2%	99.3%	24.4	*
20	4	32.5	91.0%	99.0%		
20	2	4.6	78.9%	90.2%		
20	2	20.0	73.2%	92.7%	24.6	*
20	2	52.7	72.1%	92.3%		
20	0.8	20.0	52.2%	83.6%	67.2	*
20	0.8	225.5	62.1%	74.3%	257.4	*
20	0.5	20.0	47.7%	78.9%		*
20	0.5	964.4	57.0%	74.4%		*
50	8.33	35.1	99.9%	99.7%		
50	8.33	50.0	100.0%	100.0%		
50	8.33	63.1	100.0%	99.3%	3.2	*
50	4	23.9	90.8%	96.9%	4.4	*
50	4	50.0	96.4%	99.7%		
50	4	81.2	90.0%	98.3%	24.1	*
50	2	11.4	79.0%	90.1%		
50	2	50.0	74.1%	90.4%		
50	2	131.8	72.4%	90.2%		
50	0.8	1.3	70.4%	77.7%		
50	0.8	50.0	55.9%	83.4%		
50	0.8	563.6	62.6%	74.2%	150.5	*
50	0.5	50.0	51.8%	77.3%		
50	0.5	2411.1	59.7%	72.4%		*
150	8.33	105.3	100.0%	99.9%		
150	8.33	150.0	100.0%	100.0%		
150	8.33	189.3	100.0%	99.0%	3.1	*
150	4	71.7	91.2%	95.6%	4.4	*
150	4	150.0	95.1%	98.6%		
150	4	243.5	93.2%	98.4%	24.4	*
150	2	34.3	79.2%	89.6%		
150	2	150.0	77.4%	91.1%		
150	2	395.3	72.0%	91.3%		

Table 7: AOT425StatPgm Test Simulation Results, PM1, PM2, PM3 (Cont.)

TrueLD50	TrueSlope	Dose0	PM1	PM2	PM3
150	0.8	3.8	68.8%	78.8%	
150	0.8	150.0	54.0%	80.6%	
150	0.8	1690.9	61.1%	73.9%	
150	0.5	150.0	49.2%	79.5%	
600	8.33	421.0	100.0%	99.9%	
600	8.33	600.0	100.0%	100.0%	24.4
600	8.33	757.2	99.9%	99.8%	3.2
600	4	286.9	89.4%	96.1%	4.4
600	4	600.0	94.3%	98.8%	24.4
600	4	974.0	89.6%	98.5%	24.2
600	2	137.2	79.8%	89.2%	
600	2	600.0	71.2%	93.1%	24.5
600	2	1581.1	72.0%	92.9%	
600	0.8	15.0	68.1%	80.1%	
600	0.8	600.0	54.3%	83.4%	24.5
600	0.5	1.6	56.5%	80.9%	
600	0.5	600.0	45.6%	83.5%	
1500	8.33	1052.6	100.0%	99.9%	
1500	8.33	1500.0	100.0%	100.0%	
1500	8.33	1892.9	100.0%	98.9%	3.1
1500	4	717.3	90.8%	96.1%	4.0
1500	4	1500.0	95.4%	99.5%	
1500	4	2435.0	91.7%	98.9%	10.6
1500	2	343.0	74.8%	90.3%	
1500	2	1500.0	71.7%	95.6%	
1500	2	3952.8	70.8%	90.8%	
1500	0.8	37.5	58.8%	83.6%	
1500	0.8	1500.0	51.3%	88.1%	
1500	0.5	4.1	54.7%	81.0%	
1500	0.5	1500.0	42.5%	88.2%	
3000	8.33	2105.1	98.6%	99.8%	2.4
3000	8.33	3000.0	97.9%	100.0%	7.1
3000	8.33	3785.8	100.0%	100.0%	
3000	4	1434.6	87.6%	99.7%	
3000	4	3000.0	77.4%	99.5%	13.7
3000	4	4870.0	83.3%	99.4%	
3000	2	686.0	55.4%	94.2%	24.6
3000	2	3000.0	59.8%	94.3%	13.7
3000	0.8	75.0	54.1%	80.4%	
3000	0.8	3000.0	44.4%	87.6%	
3000	0.5	8.2	53.4%	78.5%	
3000	0.5	3000.0	42.3%	86.6%	
3500	8.33	2456.0	94.3%	99.8%	2.0
3500	8.33	3500.0	88.4%	100.0%	
3500	8.33	4416.8	100.0%	100.0%	
3500	4	1673.7	82.2%	99.5%	
3500	4	3500.0	72.4%	99.5%	
3500	2	800.4	51.1%	94.9%	
3500	2	3500.0	56.3%	94.8%	
3500	0.8	87.5	52.0%	78.8%	
3500	0.8	3500.0	44.9%	87.9%	
3500	0.5	9.6	51.1%	78.5%	
3500	0.5	3500.0	47.2%	83.8%	

PM1 = Percentage of finite confidence intervals

PM2 = Coverage of the 95% confidence intervals

PM3 = Ratio of the upper to lower 95% confidence bounds

* indicates scenarios for which the EPA and AOT425StatPgm simulations are comparable.

Appendix A: Test Data Sets

The following table lists the dose and outcome for the test data sets. For the tests, the short- and long-term outcomes were assumed to be equal. The doses in the tests derived from the simulations (Sim1 to Sim15) follow the OECD TG 425 procedures. The doses in the other tests were constructed to test the LD50 and confidence interval calculations and, in general, do not follow the OECD TG 425 procedures.

Test Data Sets		
DataSet	Dose	Result (X = died)
AA	1	O
AA	1	O
AA	2	O
AA	2	X
AA	2	X
AA	2	X
AA	1	O
AA	2	X
AB	1	O
AB	1	O
AB	1	O
AB	1	O
AB	1	O
AB	1	O
AB	1	X
AB	2	X
AB	2	X
AB	2	X
AB	2	X
AB	2	X
AB	2	X
AC	1	O
AC	1	O
AC	2	O
AC	2	X
AD	1	O
AD	1	O
AD	1	O
AD	2	O
AD	2	X
AD	2	X
AD	2	X
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	O
AE	1	X
AE	1	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
AE	1.5	X
AE	1.5	X
AF	1	O
AF	1	X
AF	1	X
AF	1	X
AF	1	X
AF	1	X
AF	1.5	X
AF	1.5	X
AG	1	O
AG	1	X
AG	1	X
AG	1	X
AG	1	X
AG	1	X
AG	1	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AG	1.5	X
AH	1	O
AH	1	O
AH	1	O
AH	1	O
AH	1	O
AH	1	O
AH	2	X
AH	2.2	O
AH	4	X
AH	4	X
AH	4	X
AH	4	X
AH	4	X
AH	4	X
AH	4	X
AI	1	O
AI	1	O
AI	1	O
AI	1	O
AI	1	O
AI	2	X
AI	2.0000001	O
AI	4	X
AI	4	X
AI	4	X
AI	4	X
AI	4	X
AI	4	X
AI	4	X
AI	4	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
AJ	1	X
AJ	1	O
AJ	2	X
AJ	2	O
AJ	4	X
AJ	4	O
AJ	8	X
AJ	7	O
AJ	16	X
AJ	16	O
AJ	32	X
AJ	32	O
AJ	64	X
AJ	64	O
AK	100	O
AK	100	O
AK	100	O
AK	100	O
AK	200	O
AK	199.999999	X
AK	200	X
AK	200	X
AK	200	X
AL	1	O
AL	1	O
AL	2	X
AL	2	X
AL	4	O
AL	4	X
AL	8	X
AL	8	X
AL	2.1729	O
AL	2.1729	O
AM	1	O
AM	1	O
AM	2	X
AM	2	X
AM	4	O
AM	4	X
AM	8	X
AM	8	X
AM	1.738	O
B	0.5	O
B	1	O
B	1	X
B	2	O
B	2	X
B	4	X
C	0.25	O
C	0.25	O
C	0.5	O
C	0.5	X
C	1	X
C	1	O
C	2	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
C	2	X
C	2	X
C	2	X
C	0.5	O
C	0.5	O
D	175	O
D	550	O
D	1750	X
D	550	O
D	1750	X
D	550	O
D	1750	O
D	5000	X
D	1750	X
D	550	X
D	550	X
D1	1517.8933	O
D1	4800	X
D1	4800	O
D1	4800	O
D1	4800	O
D1	5000	X
D1	5000	X
D1	5000	X
D1	5000	O
E	175	O
E	550	O
E	1750	X
E	550	O
E	1750	X
E	550	O
E	1750	O
E	5000	X
E	1750	X
E	550	X
E	550	O
Exempl 5	175	O
Exempl 5	550	O
Exempl 5	1750	X
Exempl 5	550	O
Exempl 5	1750	X
Exempl 5	550	O
Exempl 5	1750	O
Exempl 5	5000	X
Exempl 5	1750	X
F	0.5	O
F	1	X
F	2	O
F	2	X
F	2	X
F	2	X
F	2	X
F	2	X
F	2	X
F	2	X
F	2	X
F	2	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
F	2	X
F	2	X
F	2	X
F	2	X
G	1	O
G	2	X
G	1	O
G	2	X
G	1	O
G	2	X
G	1	O
H	200	O
H	400	X
H	200	X
H	100	O
I	1	O
I	1	O
I	2	O
I	2	X
I	4	X
I	4	X
K	100	O
K	200	O
K	400	X
K	200	X
K	100	O
K	200	O
K	400	X
K	200	X
K	100	O
N	10	X
N	10	O
N	10	O
N	5	O
N	25	X
N	25	X
O	0.5	O
O	0.5	O
O	0.5	O
O	0.5	O
O	0.5	O
O	0.5	O
O	1	O
O	1	X
O	1	X
O	1	X
O	1	X
O	1	X
O	1	X
O	2	X
P	10	X
P	10	O
P	10	O
P	5	O
P	25	X
Q	10	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
Q	10	O
Q	10	O
Q	5	O
Q	25	X
Q	9999	X
Q	9999	X
R	0.5	O
R	0.5	O
R	0.5	O
R	0.5	O
R	1	X
R	1	X
R	1	O
R	1	X
R	1	X
R	1	X
R	1	X
R	1	X
R	1	X
R	2	X
R	2	X
R	2	X
S	10	O
S	10	X
S	10	O
S	25	X
S	5	O
S	5	O
T	0.5	O
T	1	X
T	1	O
T	1	X
T	1	X
T	1	X
T	1	X
T	1	X
T	2	X
W	5	O
W	5	O
W	10	X
W	10	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	5	O
X	10	X
X	10	O
Y	1	O
Y	1	O
Y	2	O
Y	2	O

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
Y	4	O
Y	4	X
Y	4	O
Y	4	X
Z	1	O
Z	2	O
Z	2	X
Z	2	X
Z	1	O
Z	2	X
LargeLD50	100	O
LargeLD50	100	O
LargeLD50	100	O
LargeLD50	100	O
LargeLD50	200	O
LargeLD50	199.999999	X
LargeLD50	1000	O
Sim1	55	O
Sim1	550	X
Sim1	55	X
Sim1	5.5	O
Sim1	55	O
Sim1	550	O
Sim1	2000	X
Sim1	550	O
Sim1	2000	X
Sim1	550	X
Sim1	55	O
Sim1	550	O
Sim1	2000	O
Sim1	2000	X
Sim1	550	X
Sim2	310	X
Sim2	175	O
Sim2	310	X
Sim2	175	O
Sim2	310	X
Sim3	310	X
Sim3	175	X
Sim3	98	X
Sim3	55	X
Sim3	31	X
Sim3	17.5	X
Sim3	9.8	X
Sim3	5.5	X
Sim3	3.1	O
Sim3	5.5	X
Sim3	3.1	X
Sim3	1.75	X
Sim3	1	O
Sim3	1.75	O
Sim4	55	X
Sim4	5.5	X
Sim4	1	O
Sim4	5.5	O
Sim4	55	X

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
Sim4	5.5	X
Sim4	1	O
Sim4	5.5	O
Sim4	55	X
Sim4	5.5	O
Sim4	55	X
Sim4	5.5	O
Sim4	55	X
Sim5	331.0637328	X
Sim5	88.64637256	X
Sim5	29.52718394	X
Sim5	8.424332589	O
Sim5	31.73039581	X
Sim5	8.46723423	O
Sim5	37.14208397	X
Sim5	12.69530716	X
Sim5	3.203443823	X
Sim5	0.992084281	O
Sim5	2.846794701	O
Sim5	10.86162988	X
Sim5	3.001967031	O
Sim5	10.94615778	X
Sim5	2.735951204	O
Sim6	55	X
Sim6	5.5	X
Sim6	1	X
Sim6	1	X
Sim6	1	X
Sim7	310	X
Sim7	175	O
Sim7	310	O
Sim7	550	X
Sim7	310	X
Sim8	310	X
Sim8	175	X
Sim8	98	X
Sim8	55	X
Sim8	31	X
Sim8	17.5	X
Sim8	9.8	X
Sim8	5.5	X
Sim8	3.1	O
Sim8	5.5	X
Sim8	3.1	O
Sim8	5.5	O
Sim9	310	O
Sim9	550	X
Sim9	310	X
Sim9	175	X
Sim9	98	O
Sim10	320	X
Sim10	100	X
Sim10	32	X
Sim10	10	X
Sim10	3.2	X
Sim10	1	O

Test Data Sets (Cont.)

DataSet	Dose	Result (X = died)
Sim10	3.2	X
Sim10	1	X
Sim10	1	O
Sim10	3.2	X
Sim10	1	O
Sim10	3.2	X
Sim10	1	O
Sim11	77.4980487	X
Sim11	18.27562034	X
Sim11	8.647620955	X
Sim11	3.094556046	X
Sim11	1.085279649	O
Sim11	2.219163761	X
Sim11	0.995617398	O
Sim11	2.255869667	X
Sim11	1.018223623	O
Sim12	175	O
Sim12	550	O
Sim12	2000	O
Sim12	2000	O
Sim12	2000	O
Sim13	175	X
Sim13	55	X
Sim13	17.5	X
Sim13	5.5	O
Sim13	17.5	X
Sim13	5.5	X
Sim13	1	X
Sim13	1	X
Sim13	1	X
Sim14	175	X
Sim14	55	X
Sim14	17.5	X
Sim14	5.5	X
Sim14	1	O
Sim14	5.5	X
Sim14	1	X
Sim14	1	O
Sim14	5.5	X
Sim14	1	X
Sim14	1	O
Sim14	5.5	X
Sim14	1	X
Sim14	1	X
Sim14	1	X
Sim15	320	O
Sim15	1000	O
Sim15	3200	X
Sim15	1000	O
Sim15	3200	O
Sim15	5000	O
Sim15	5000	O

Appendix B: Evaluation of Cumulative Normal Distributions and Calculation of the Chi-square Test

The precision of the cumulative normal distribution function used in AOT425StatPgm was evaluated using two methods: 1) comparing numerical to exact derivatives, and 2) comparing different published functions.

Three different algorithms for calculating the cumulative normal distribution were evaluated: NORMP, NPROB, (Algorithm A66 from Statlib@stat.cmu.edu) and the built in Excel function. The cumulative normal distribution calculates a normal probability value P from a standard normal deviate Z. The NORMP function was reported as providing P values good to $1e-15$. There was no stated precision for the other algorithms. The relative accuracy of the algorithms was evaluated by comparing the exact derivative of the cumulative normal distribution to a numerical derivative calculated from the function. The numerical derivative at x is $(FN(x+\delta)-FN(x-\delta))/\delta$, where FN is the function for which the derivative is calculated. Delta was varied to find the value that gave the best agreement between the numerical and exact derivatives; Delta was set to 0.00001. The maximum difference between the numerical and exact derivative was used as a measure of the relative precision of the cumulative normal algorithm. The maximum was calculated over a range for Z from -14 to 0 (the P value for positive Z is just 1-P for negative Z). For Excel, the maximum difference in the derivatives was $4.1e-5$. For the NORMP, the maximum difference was $6e-12$. For NPROB the difference was $4e-11$. These values suggest that the NORMP function is the most precise of the three.

Assuming that NORMP is the most precise function, the difference between the NORMP and Excel was calculated for different Z values. The maximum difference was $7e-8$. Thus, the Excel cumulative normal function is good to about $1e-7$ or better. The maximum difference between NORMP and NPROB was $3e-12$. If NORMP is truly good to $1e-15$, then the precision of NPROB is about $3e-12$. Alternatively, if NORMP is not as precise as reported, the precision of NORMP made be as low as $3e-12$.

Another consideration for selecting the algorithm is the range of Z over which P is non-zero. This range was greatest for the NORMP function (-37 to 37), less for the NPROB function (-12.7 to 12.7), and least for the Excel function (-7 to 7). AOT425StatPgm uses the NORMP function. The evaluation of the different cumulative normal functions was performed in Excel.

The inverse cumulative normal function used by AOT425StatPgm is algorithm AS241 (PPND16) from the Statlib site mentioned above. The reported precision of Z from this algorithm is $1e-16$. A comparison of Z to $PPND16(NORMP(Z))$ suggested that the precision of either the NORMP or PPND16 function was not quite as precise as reported.

Chi-square Test Calculation

A chi-square test was used to evaluate the agreement between the proportion of simulated tests falling into different cases from the EPA and AOT425StatPgm simulations. The simulation results were compared for 52 scenarios for which the same dose progression was used. The simulated tests were divided into four classes, Case 1 or 5, Case 2, Case 3, and Case 4. For each scenario a two-by-four contingency table was constructed. When calculating the chi-square value, classes with fewer than five simulated tests were excluded from the calculations. Since the chi-square values are independent, the chi-square values and the degrees of freedom were summed across the 52 scenarios to get a more precise overall evaluation. The maximum possible value of degrees of freedom for each overall test was $52*3 = 156$, however, after removing classes with few simulated tests, the value of degrees of freedom for the overall chi-square was only 122.