External Peer Review of EPA's MS-COMBO Multi-tumor Model and Test Report

Prepared for:
Allen Davis, MSPH
U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment
109 T. W. Alexander Drive
Research Triangle Park, NC 27711

Prepared by:
Versar, Inc.
6850 Versar Center
Springfield, Virginia 22151

Contract No. EP-C-07-025
Task Order 97

Peer Reviewers:
Kenneth T. Bogen, Dr.PH., DABT
Kenny S. Crump, Ph.D.
Kerby A. Shedden, Ph.D.

March 3, 2011

Peer Reviewers:

Kenneth T. Bogen, Dr.PH., DABT
Exponent
Oakland, CA 94607

Kenny S. Crump, Ph.D.
Louisiana Tech University
Ruston, LA 71270

Kerby A. Shedden, Ph.D.
University of Michigan
Ann Arbor, MI 48109

Review by Kenneth T. Bogen, Dr.PH., DABT

Peer Review Comments on EPA's MS-COMBO Multi-tumor Model and Test Report

Kenneth T. Bogen, Dr.PH., DABT
Exponent*

February 27, 2011

I. GENERAL IMPRESSIONS

Documentation provided to users is clear enough to be adequate to allow users to run the program and obtain program output, but is not adequate to inform users concerning details about the context in which applying the model is appropriate or intended, nor does the documentation properly credit published sources concerning the origin of the multi-tumor modeling concept and related mathematical and biological considerations. The accuracy of information presented was assessed and confirmed using an independent, bootstrap method of parameter estimation. While the model thus appears to provide sound results, the format of result delivery appears to be arcane and inefficient, apparently offering no convenient ("Session-type" tabular) summary of model output as an alternative to a simple concatenation of tumor-specific outputs each in standard BMDS long-form ASCII format. Standard output for the multi-tumor model, as for other BMDS models, should provide the user with the entire estimation-error distribution for each estimated BMD, rather than just a MLE and single user-specified percentile.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: *Are the documentation and model output associated with the MS-COMBO model clear and transparent?*

Background Information Concerning Motivation and Origins of the Missing from Help Documentation

Documentation provided does not (but should) include explicit MLE equations that are solved to estimate the BMD and specified percentile(s) of its distribution characterizing estimation error. Specifically, the draft Help documentation states that "The calculation of the combined BMDL is a more complicated computation based on the profile-likelihood approach. As such, it gives the lowest value of the dose that satisfies the following conditions: there is a combination of parameters (across all models) for which the value of the BMDL gives a combined extra risk equal to the BMR and, using those parameter values, the combined log-likelihood is greater than or equal to a minimum log-likelihood defined by the maximum log-likelihood and the confidence level specified by the user (i.e., the parameters that give the desired extra risk when the dose is equal to the BMDL give a combined log-likelihood that is "close enough" to the maximum combined log-likelihood)." However, no explicit details are provided about how the computation is actually implemented, and no proof is provided that the implementation of the profile-likelihood method used guarantees that the results obtained reflect global rather than local maxima, insofar as the method must trace likelihoods over multiple (including competing)

* The attached review represents the personal opinion of Dr. Kenneth Bogen, an employee of Exponent. This review has not undergone QA/QC or corporate review by, and does not comprise a work product of, Exponent.

parameter-vector pathways, where deviations of each parameter in opposite directions from its MLE may yield equivalent decrements in log-likelihood from its global maximum value that occurs at the MLE values of all parameters. At a minimum, the explicit log-likelihood equations that are optimized should be specified, for the multi-tumor model as well as for all other BMDS models (e.g., in a technical appendix to the Help documentation).

Background Information Concerning Motivation and Origins of the Missing from Help Documentation

The draft Help documentation presently includes no references specific to the multi-tumor model, but rather includes only three general references on BMD methodology. Users of the multi-tumor model should be given a brief description of the origin, context, and implied assumptions of this model. Two such references (Bogen 1990; NRC 1994) are provided in supplemental material provided to model reviewers ("NCEA Statistics Workgroup Memo No. 1, January 2008'), but there is no indication of how or whether any of this supplemental material will be incorporated into Help documentation. The assumption of independence in tumor-type-specific tumor occurrence is particularly fundamental to the valid application of this model, as was emphasized in original descriptions and mathematical analyses concerning this model (Bogen 1986, 1990; Bogen and Spear 1987; NRC 1994). However, this critical assumption and conditions under which it is likely to be violated are not discussed in the Help documentation. Citation of publications in which the multi-tumor model was first presented, discussed, illustrated and recommended, will allow users to better understand its origin and purpose. To facilitate a summary of this background information, the following synopsis is offered.

A formula stating that, conditional on a multistage cancer risk model and assuming independent occurrence of different tumor types, the (e.g., Monte-Carlo) sum of estimated tumor-specific potencies equals the aggregate potency for increased risk of inducing one or more of the set of tumor types addressed first appeared in my own publications (Bogen 1986, 1990; Bogen and Spear 1987). A proof of this relationship first appeared in Bogen (1986, 1990), and a similar proof appeared in Appendix I-1 of NRC 1994), which I wrote. In Chapter 11 of Science and Judgment in Risk Assessment (which chapter I also wrote), the NRC (1994) specifically recommended to EPA that, to address multiple tumor types, the Agency should adopt an approach such as the Monte Carlo approach identified and illustrated by Bogen (1986, 1990), and by Bogen and Spear (1987), which was summarized in Appendix I-1 of the NRC (1994) report.

The publications mentioned (Bogen 1986, 1990; Bogen and Spear 1987; NRC 1994) all pointed out that the multistage potency-summation approach is valid only conditional on independent occurrence of different tumor types. The summation approach is not valid if elevations in the incidence rate of different tumor types occur in a correlated manner. Tumor-type-occurrence correlations can occur, e.g., when it is known that hormone-secreting tumors promote the occurrence of secondary tumors by enhancing cell proliferation in those secondary tumor sites. Although the null hypothesis of tumor-type independence can be tested statistically using individual animal data in case such data are available, this is generally labor-intensive. An examination and demonstration of the general validity of the tumor-type-independence assumption for most common tumor types that occur in NTP rodent bioassays appeared as

Appendix 1-2 of NRC (1994). Appendix I-2 of the NRC (1994) report essentially reprints an earlier report (Bogen and Seilkop 1993) I did for my NRC committee on this topic with Dr. Steve Sielkop of Analytical Sciences, Inc. (Alston Technical Park, 100 Capitola Drive, Suite 106, Durham, NC), who had access to the complete NTP rodent bioassay data base at that time, prior to when these data were made electronically accessible to the general public.

2. Adequacy of Testing Methods and Results: ***The testing process should ensure that the MS-COMBO model results are reliable, accurate and clear.***

(a) Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?

Yes, except to the extent that appropriate tests were not included in the set of tests documented, as explained below.

(b) Have appropriate aspects of the MS-COMBO model been tested?

Appropriate aspects of the MS-COMBO model appear to have been tested, except insofar as no test was performed addressing BMDL estimation for the simplest scenario involving k identical data sets for large k that allows comparison of MS-COMBO likelihood-based results with expected BMDL values at any specified confidence level as predicted by the Central Limit Theorem. An upper-bound q^* potency (i.e., the upper bound on the linear coefficient Q in dose) is related to BMDL by BMDL = –log(1-BMR)/q^*, so both bounds essentially provide redundant information for a wide variety of data sets (Bogen 2011). Because aggregate potency Q is just the sum of tumor-specific potencies Q_i, i = 1,...,k, for sufficiently large k and (for convenience) assuming $Q_i = Q_j$ = for all $\{i, j\}$, the Central Limit Theorem guarantees that aggregate potency Q is normally distributed as ~N(k E(Q_i), k Var(Q_i)). Under these conditions, the statistics of Q (and thus of –log(1-BMR)/Q) are known functions of just the first two moments of Q_i, and hence these statistics may be compared to those calculated for BMDL by MS-COMBO.

(c) Do the test results indicate that the MS-COMBO model provides reliable, accurate and clear results? (Note: Reviewers are encouraged, but not required, to apply alternative statistical methods and software to validate the MS_COMBO results.)

Test results provided appear to indicate that the MS-COMBO model results are reasonably reliable and accurate. However, an important missing test would involve the case of k identical data sets, as described above, insofar as in this case exact statistics are readily computed by independent methods. This test was performed using a Bootstrap Monte Carlo approach consisting of a ("linearized") modification of a "Generic Hockey-Stick" (GHS) model previously described (Bogen 2011), where the modification used was to constrain all multistage model parameters to be non-negative, and the degree of the multistage polynomial to be ≤3 (i.e., constraints identical to those that users may implement via the BMD Multistage Cancer model). In this test, doses were set to {0, 1, 2, 3, 5}/5, the corresponding number of animals per dose used to {50, 50, 50, 50, 52}, BMR was set to 0.10, and the number of animals with tumor type i (for all i) to {0, 2, 4, 6, 10}, respectively, and k was set to be 7. To simulate dose-response data,

binomial error was assumed about the observed data. In this test, no attempt was made to estimate or correct for bias associated with bootstrap potency estimation from simulated data sets.

The attached pdf file documents estimates of multi-tumor BMDL obtained using the modified GHS bootstrap approach, by three methods (an asymptotic method, and two bootstrap methods, the first being approximate and the second a more exact method), and compares these to the BMDL estimate produced by MS-COMBO. For the seven indicated data sets, MS-COMBO estimates the BMDL to be 0.0625. The linearized GSH method starts by simulating 4000 sets of 5-dose dose-response data assuming binomial error about the observed data as specified above. A total of 3,696 of these were estimated (analytically, as described by Bogen 2011) to have positive (as opposed to zero-valued) "potency" coefficients (i.e., linear coefficients in dose). Only these 3,696 positive-potency fits were included in further analysis (an arbitrary, conservative decision that reflects one of two plausible interpretations of how parameter estimation ought to be done for the multistage cancer model using a bootstrap procedure, the alternative being to include all fits including those with an estimated potency of zero). For each fit, the corresponding complete fitted model and associated numerically calculated BMD value were saved. The mean (±1 SD) of estimated potency and BMD were found to be 0.181 (± 0.063) and 0.612 (±0.327), respectively (pdf, page 2), with a corresponding upper-bound potency of 0.277 and BMDL of 0.380. For comparison MS-COMBO applied to the same single dose-response data set yields BMDL = 0.357 (pdf, page 4).

For multi-tumor BMDL involving seven such data sets, the asymptotic linearized GHS method thus estimates multi-tumor BMDL = ln(10/9)/[7*0.181 + 1.6448*Sqrt(7)*0.063] = 0.0682 (pdf, page 6). Bootstrap "Method 1" estimates multi-tumor BMDL = ln(10/9)/Sum(Q_i, i = 1,...7) = 0.0684 (pdf, page 9), where Q_i is the empirical bootstrap distribution of 3,696 positive-valued potencies obtained, and stochastic summation was implemented by Monte Carlo methods. Bootstrap "Method 2" estimates multi-tumor BMDL = 0.0686 (pdf, page 10), as the 5th (i.e., 1-tail lower 95th) percentile of the distribution of the numerical solution for BMD to the equation BMR = FIT_j = 1 – exp[Sum($X_{i,j}$, i = 1,...7)], where $X_{i,j}$ is the j^{th} realization of the sum over seven random permutations of the vector of (saved) fitted multistage-cancer-model polynomials referred to above. The slight (~1%) difference of the latter estimate from the corresponding asymptotic normal approximation is understandable, in view of the significant non-normality of the underlying aggregate potency distribution that dominates the calculation of BMDL (p = 0.0021 by Shapiro-Wilk test; pdf page 7).

The MS-COMBO estimate of multi-tumor BMDL based on the same data (again, with k = 7) is 0.0625 (pdf, page 11). The MS-COMBO estimates of multi-tumor BMDL is therefore within 10% of the estimate produced by the linearized-GHS Bootstrap "Method 2," and on this basis the results agree fairly well. In the context of estimating multi-tumor BMDL, this specific example emphasizes the importance of an accurate estimate of the expected value and variance of the distribution of aggregate potency. This central importance is created by the Central Limit theorem, which ensures that confidence bounds on aggregate multi-tumor potency must, in the limit, be governed by only these two moments. Unfortunately, bias concerning estimates of the mean and variance of aggregate potency, conditional on realistically small sample sizes and binomial sampling error in dichotomous dose-response data, cannot be evaluated by methods

used in material provided to MS-COMBO reviewers. In general, such potential bias can be evaluated only by Monte Carlo simulations, like those conducted by Bogen (2011).

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results that give you special cause for concern? If so, please describe your concerns and recommendations.***

MS-COMBO, and other BMDS models, should allow users, on request, access to each entire estimated (tumor-specific, and multi-tumor) BMD distribution, not just a single specified percentile of it, in addition to the MLE (see attached pdf). Multi-tumor model output should be, on request, output to the user in summary Sessions format, rather than only in the ASCII long form that seems now to be the default (or only?) mode of output.

III. SPECIFIC OBSERVATIONS

No specific comments or corrections, other than those provided above.

IV. REFERENCES

Bogen KT. Uncertainty in environmental health risk assessment: A framework for analysis and an application to a chronic exposure situation involving a chemical carcinogen. Doctoral Dissertation, University of California Berkeley, School of Public Health, Berkeley, CA, 1986.

Bogen KT, Spear RC. Integrating uncertainty and inter-individual variability in environmental risk assessment. Risk Anal 1987; 7:427-436.

Bogen KT. Uncertainty in environmental health risk assessment. Garland, New York, NY, 1990.

Bogen KT, Seilkop S. Investigation of Independence in Inter-Animal Tumor-Type Occurrences within the NTP Rodent-Bioassay Database: Report prepared for the National Research Council, Board on Environmental Studies and Toxicology, Committee on Risk Assessment of Hazardous Air Pollutants, 1993. http://www.osti.gov/bridge/product.biblio.jsp?osti_id=10121101

Bogen KT. Generic Hockey-Stick model for estimating benchmark dose and potency: performance relative to BMDS and application to anthraquinone. Dose Response 2011; 9 (in press).

National Research Council (NRC). Science and Judgment in Risk Assessment. Chapter 11 ("Aggregation"), Appendix I-1 ("Aggregate Risk of Nonthreshold, Quantal, Toxic End Points Caused by Exposure to Multiple Agents (Assuming Independent Actions)"), and Appendix I-2 ("Independence in Inter-Animal Tumor-Type Occurrence in the NTP Rodent-Bioassay Database"). National Academy Press, Washington DC, 1994.

Review by Kenny S. Crump, Ph.D.

Peer Review Comments on EPA's MS-COMBO Multi-tumor Model and Test Report

Kenny S. Crump, Ph.D.
Louisiana Tech University

February 17, 2011

I. GENERAL IMPRESSIONS

The program appears to be working properly based on comparisons between the provided output and independent calculations that I have made. However, there were a few discrepancies that, although not large, should be looked into. The evaluation of the program was limited in terms of the number of tumors and degree of the polynomial. Additional testing could be useful in determining the range of number of tumors and degree of polynomial over which the program provides accurate answers. The presentation was clear enough for persons thoroughly familiar with BMD analysis. However, it would not be adequate for persons who were not familiar with the process.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: *Are the documentation and model output associated with the MS-COMBO model clear and transparent?*

The test report is not in a form that would be suitable for general distribution. Although I had no trouble understanding the report, I think it would be difficult for some who is not familiar with the subject matter. Terms are used without being defined, descriptions are not complete, and references are limited. The parameterization described in the test report of the background response in the model is apparently different from that implemented in MS-COMBO. A number of specific comments are included in Section III below.

The model output seems adequate and I assume it is consistent with output generally provided by BMDS. However, the volume of the output was such that I tended to lose track of the model being reported on. It might help to repeat the name of the run at various places in the output. I did not implement the software.

2. Adequacy of Testing Methods and Results: *The testing process should ensure that the MS-COMBO model results are reliable, accurate and clear.*

(a) Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?

Yes, the record provided is sufficient for an understanding of the test methods and results.

(b) Have appropriate aspects of the MS-COMBO model been tested?

Test results are presented only for combinations of three tumor types, and then only using models of degree 2. Higher degree models are implemented only for two tumor types. Increasing the number of tumors and the polynomial degree increases the number of parameters to be estimated and places increasing strain on the optimizer. Additional testing is needed to determine how the program performs with more tumor types and higher degree polynomials. From that, some guidance would be useful on the numbers of tumors and polynomial degree combinations the program can reasonably handle. Is there an upper bound set by the program on the number of tumors? It may be that three will be sufficient for most applications, but guidance in this area would be useful.

It seemed odd that a fourth degree polynomial was implemented for the example that modeled two tumors, but only a second degree polynomial was used to model three tumors. The test report says that higher degree polynomials were applied and compared with Excel calculations, but these were not reported. It should be noted that the conditions checked using Excel (as I understand them) are necessary for the program to give the correct answer but not sufficient. If the optimizer used by the program provided a suboptimal answer, this would not be identified by the checks performed in Excel.

The ability of the program to use higher degree polynomials in combination with multiple tumors can be important. For example, I applied a model with a fourth degree polynomial to the data sets 1, 2, and 3 combined (details provided below). This model fit was highly statistically improved over the fit from the example provided that used only a second degree polynomial. Moreover, the fourth degree polynomial gave a substantially different MLE BMD and BMDL from that obtained using a second degree polynomial.

(c) Do the test results indicate that the MS-COMBO model provides reliable, accurate and clear results? (Note: Reviewers are encouraged, but not required, to apply alternative statistical methods and software to validate the MS_COMBO results.)

As detailed below, I independently implemented several of the analyses reported in the test report. The results I obtained are generally in good agreement with those obtained by MS-COMBO. The BMD and BMDL obtained from my quadratic fit to the combined Data Sets 1-3 differs enough from the MS-COMBO values to suggest looking into, although this could be a problem on my end. The corresponding calculations for combined Data Sets 4-6 agreed closely with those in reported by MS-COMBO. I was concerned at first by the fact that the MLE parameters estimates I obtained agreed very closely with those from MS-COMBO except for the background parameter. However, later I decided this was due to use of a different, but equivalent, model parameterization in MS-COMBO than was presented in the test report.

Independent calculations I made are summarized below and compared to results from MS-COMBO. Differences are minor in most cases. However, the differences in the MLE BMD and BMDL obtained in the three tumor fit should be looked into.

Fourth Degree Fit to Data Set 1

ML verified to 6 digits (number of digits reported in MS-COMBO).

Some small differences in MLE parameter estimates:

Table 1: Fourth Degree Fit to Data Set 1: differences in MLE parameter estimates.

MLE Parameters	MS-COMBO	My values
β 0	0	0
β 1	0	0
β 2	1.05563e-005	1.05644E-05
β 3	2.3908e-007	2.39011E-07
β4	0	8.42585E-15

Some small differences in BMD estimates:

Table 2: Fourth Degree Fit to Data Set 1: differences in BMD estimates.

BMD Estimates	MS-COMBO	My values
MLE	63.8712	63.82780376
BMDL	52.0372	52.25954381
BMDU	72.7481	72.74811282

Fourth Degree Fit to Data Set 2

ML verified to 6 digits (number reported in MS-COMBO).

Some small differences in BMD estimates.

Table 3: Fourth Degree Fit to Data Set 2: differences in BMD estimates.

BMD Estimates	MS-COMBO	My values
MLE	63.8279	63.82756387
BMDL	42.8487	42.82684083
BMDU	78.1258	78.1375515

MLE parameter estimates are substantially in agreement.

Table 4: Fourth Degree Fit to Data Set 2: differences in MLE estimates.

MLE Parameters	MS-COMBO	My values
β 0	0.0493062	0.050562943 <= difference explained by different parameterization
β 1	0	0
β 2	1.52649e-005	1.52662E-05
β 3	1.28069e-007	1.28047E-07
β 4	5.94628e-010	5.94703E-10

Multi-tumor Quadratic Run on Data Sets 1-2

Some small differences in BMD estimates.

Table 5: Multi-tumor Quadratic Run on Data Sets 1-2: differences in BMD estimates.

BMD Estimates	MS-COMBO	My values
MLE	48.3967	48.39443254
BMDL	32.4165	32.43760158

Multi-Tumor Run on Data Sets 1-3

Quadratic Fit to Data set 1.

ML is missing from MS-COMBO output.

MLE parameter estimates verified to 6 digits (number reported in MS-COMBO).

BMD MLE, BMDL and BMDU all verified to 6 digits (number reported in MS-COMBO).

Quadratic Fit to Data Set 2

ML and BMD MLE agree to 6 digits.

MLE parameter estimates are in agreement.

Table 6: Quadratic Fit to Data Set 2: differences in MLE estimates.

MLE Parameters	MS-COMBO	My values
β 0	0.034571	0.035182733 <= difference explained by different parameterization
β 1	0	0
β 2	4.90516e-05	4.90516E-05

Quadratic Fit to Data Set 3

ML and BMD MLE agree to 6 digits.

MLE parameter estimates are substantially in agreement.

Table 7: Quadratic Fit to Data Set 3: differences in MLE estimates.

MLE Parameters	MS-COMBO	My values
β 0	0.0420129	0.042920864 <= difference explained by different parameterization
β 1	0	0
β 2	5.04927e-05	5.049264362E-05

Quadratic Fit to Data Sets 1-3

I obtained different MLE BMD and BMDL.

Table 8: Quadratic Fit to Data Sets 1-3: differences in BMD estimates.

BMD Estimates	MS-COMBO	My values
MLE BMD	26.9335	25.7182794
BMDL	25.1391	24.79672285

Quadratic Fit to Each of Data Sets 4-6

Each of these analyses gave ML and ML parameter estimates that were in agreement with MS-COMBO (results not tabulated).

Quadratic Fit to Each of Data Sets 4-6

MLs were in agreement:

Table 9: Quadratic Fit to Each of Data Sets 4-6: differences in ML estimates.

ML Estimate	MS-COMBO	My value
ML	-170.1197623344506837384	-170.119762334501

MLE BMDs and BMDLs were also in agreement:

Table 10: Quadratic Fit to Each of Data Sets 4-6: differences in BMD estimates.

BMD Estimates	MS-COMBO	My value
MLE BMD	0.334355	0.334370616396791
BMDL	0.263129	0.263148213704089

Fourth Degree Polynomial Fit to Data Sets 1-3

(This analysis was not part of the testing provided.)

ML -589.6941311

This likelihood is a highly significant improvement over the value of -604.157 obtained using a quadratic polynomial.

BMD 40.52947009
BMDL 33.82020238

These values are substantially different from the values BMD =26.9335 and BMDL = 25.1391 obtained from MS-COMBO using a quadratic model. This suggests that it may be important at times to employ polynomials with degree higher than 2 to analyses of three or more tumors, and further testing in this direction may be warranted.

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results that give you special cause for concern? If so, please describe your concerns and recommendations.***

Additional testing using three or more tumors and higher degree polynomials could be helpful.

III. SPECIFIC OBSERVATIONS

The following are specific comments on the test report. It was not clear to me who the intended audience is for this report. If it was written only to assist in this review, then it served this purpose adequately. However, if it is intended for wider use, it needs to be revised with these comments taken into consideration.

Page 3, first paragraph
"one or more of" instead of "one or more or"

Page 3, second paragraph
The notation needs some attention. Although the symbol Ai was not defined, phrases like "the presence or absence of Ai in an animal" indicate that it refers to a tumor. However, it is used subsequently as if it refers to the event an animal has a tumor of this type. The symbol ∩ used in the report to refer to the union of event normally refers to the intersection of events. Why not use the conventional symbol for union of events?

Page 3, a few lines below eq. 1
"parameters of the multistage model fit (estimated) for tumor Ai" "Fit (estimated)" is not needed. They are the parameters whether or not they are fit.

Page 3 eq. 1
It should be stated here that the parameters are non-negative. The notation also should be changed to indicate a last coefficient e.g., β_{Ki}. Otherwise it appears that there are an infinite number of parameters. Also, background is parameterized differently in eq. 1than in MS-COMBO. This should be changed to agree with MS_COMBO or explained somewhere in the document.

Page 4, first paragraph
"maximum likelihood estimate of a function of random variables is the function applied to the maximum likelihood estimates of those random variables."
This phrase should refer to "parameters" instead of "random variables".

"observations of A and B"
What do A and B refer to?

Page 4, eq. 5
The BMD should be defined before using it in this equation.

Page 4, first paragraph below equation 5
Likewise, "BMDL" should be defined. Also, it would be better to say "the calculation" is based on the profile likelihood, rather than "the derivation" as derivation usually refers to manipulation of algebraic symbols, and calculation to manipulation of numbers.

Next sentence

The profile likelihood is more complicated that what?

Next to last line in first paragraph below equation 5
"extended" rather than "extend"

"Likelihood" should be explained.

One or more references to maximum likelihood estimation and to the profile likelihood method of computing statistical confidence intervals is needed in this section. Also, I think the discussion of one degree of freedom is not very useful. It would be better just to refer to some reference on this point. If a person knows enough statistics to follow this discussion, he/she will probably already know that one degree of freedom is appropriate and if he/she doesn't know that they are unlikely to follow this discussion anyway. If the report is going to get into this at all, it should indicate how the chi-square distribution with one degree of freedom is used in the calculation. It could also be worthwhile to at least mention the limitations of the chi-square assumption resulting from parameters estimates lying on the boundary of the parameter space (i.e., estimated as zero). It could also be worthwhile to discuss the advantages of the profile likelihood method over other methods, particularly those based on the asymptotic normality of ML estimators. A mathematical description of the profile likelihood method would be helpful to supplement the heuristic description.

Page 7, first paragraph in section 3.2
What is meant by a matter of algebra? In particular, doesn't the MLE BMD require a search routine?

Additional Comments

The NRC claim (referred to in the test report) that dependence among tumors is not likely to introduce significant errors may be true, but it would still be best to avoid this assumption whenever possible. When it is possible to identify whether an animal had at least one of the tumors of interest (e.g., when individual animal pathology data are available), it would be preferable simply to apply a one-tumor model to the tumor category defined by this combination. The analyst would then have the option of using the same model (multistage) as implemented in MS-COMBO or a different one, and the assumption of independent tumors would not be needed.

The fact that independence is assumed and the multistage model used for individual tumors, the probability of at least one tumor of the modeled types also has a multistage form is a nice simplification, but it is not necessary. The approach could easily be extended to other dose-response models and even situations in which different dose-response models are applied to different tumors.

USEPA generally prefers to utilize pharmacokinetic data on the dose to the target organ in its risk assessments. However, different tumor sites will have different internal doses and it will not be possible to take these differences into account properly with the current implementation of MS-COMBO. Conceptually, accounting for target organ doses would require incorporation of a quantitative physiologically-based pharmacokinetic (PBPK) model into the analysis. However,

if the BMD is low enough so that the PBPK model is linear (which may often be the case), the model can be adequately represented by constants (one for each tumor site) that convert a steady state external exposure into the corresponding internal dose at the tumor site. It would only require a minor extension of MS-COMBO to allow for inputting these constants and appropriately incorporating them into the calculations. Consistent with the manner in which EPA normally uses PBPK data to convert from animals to humans, the animal tumor data would be modeled using tumor site-specific internal doses estimated from the animal PBPK model, and the BMD calculation would use the human PBPK model (implemented using the simple linear approximation) to calculate the human external BMD corresponding to these internal doses.

Review by Kerby A. Shedden, Ph.D.

Peer Review Comments on EPA's MS-COMBO Multi-tumor Model and Test Report

Kerby A. Shedden, Ph.D.
University of Michigan

February 28, 2011

I. GENERAL IMPRESSIONS

The proposed MS-COMBO method is a reasonable way to approach the problem of obtaining point estimates and uncertainty bounds for the probability that any one of a number of adverse events will occur in an individual animal. Its main weakness, which it shares with much of the methodology used in this field, is that it uses convenient parametric models that may or may not hold in any given setting. In a single-endpoint analysis, this weakness can be addressed to some extent by the use of model assessment procedures. For multiple endpoints, this issue is heightened by the number of endpoints under consideration (all of which must fit into the same modeling framework), and the potential that multiple small errors will be compounded when calculating the combined-endpoint probabilities.

The presentation is overly brief. While the presentation is generally clear, some major issues are barely addressed. One such issue is the means by which constrained optimization is used in the profile likelihood analysis. Since the constraints are linear, it should be straightforward to constrain the numerical search to the subspace of feasible beta vectors, simply by projecting the gradients and making corresponding adjustments to the working scale matrix (assuming a quasi-Newton or conjugate-gradient approach is used, although this is not stated). It would have been helpful to learn more specifically how this was done. Solving for one of the beta values (e.g. the intercept), and resubstituting into the original equation (as done in section 1.2 to justify the sampling distribution of the profile likelihood), is not the best choice from an algorithmic point of view, partly because not all models have a free intercept (for example, some of the test models do not).

It would have been helpful to include more information about how the experimental data used to fit the model should be obtained. Presumably the same animals are assessed for all tumor types. Although a 16 year-old NRC report is cited to support modeling the tumor occurrences within an animal as independent events, it is reasonable to ask whether subsequent work has borne this out. I would also like to know whether the experiments are carried out in a way to minimize cluster effects between batches of animals and the dose levels (e.g. are the animals randomly assigned to the dose levels, are the experiments run in parallel, and are there non-negligible cluster effects related to litter, cage, etc.?). While a review of experimental design for animal toxicity studies is beyond the scope of this proposal, it's worth noting that these issues become relevant in a new way when considering the multiplicative model for outcome probabilities. Specifically, if there are cluster effects or cryptic dependencies in the data, point estimates would be wrong, whereas these types of problems would only affect confidence limits (but not point estimates) for conventional single-endpoint studies.

The profile likelihood approach is an elegant and rigorous way to address complex inference problems, such as the one that arises in this application. Conveniently, the exponential model specification of the dose/response relationship allows the profile analysis to be relatively easily implemented (subject to questions expressed above about the lack of detail regarding the numerical approach). There is at least one major challenge for inference in this setting, which is not addressed in the proposal. This is the fact that constrained estimates are employed (i.e. the coefficients must be non-negative to ensure a monotone dose/response relationship). When doing this, it will not be uncommon for coefficient estimates to lie on the boundary of the domain. It is well-known that likelihood ratio procedures (including profile-likelihood) do not follow the usual asymptotic law, even approximately, in these settings.

Examining test data set 1 from appendix A, it appears that all single-variable models have an interior MLE, but only 3 of the 6 two-variable models have an interior MLE (i.e. the other three models have a negative coefficient). The fitted model from appendix A, which somewhat bizarrely includes quadratic and cubic terms, but no intercept or linear term, is seen with some calculation to give the highest likelihood among all two-variable models. Thus there evidently was some pre-testing and model selection done to arrive at a model with an MLE that was interior to the domain. The profile likelihood approach is not guaranteed to give meaningful results when there is pre-testing and model selection, or when the point estimates lie on the boundary of the domain.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: ***Are the documentation and model output associated with the MS-COMBO model clear and transparent?***

The documentation is generally clear insofar as stating the form of the model, the calculations that are done to relate the BMR to the BMD, and setting up the profile likelihood procedure.

2. Adequacy of Testing Methods and Results: ***The testing process should ensure that the MS-COMBO model results are reliable, accurate and clear.***

(a) Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?

I clearly understand the testing methods that were employed, based on the description of this process given in the proposal. Thus, I view the documentation of these tests as adequate.

(b) Have appropriate aspects of the MS-COMBO model been tested?

All the tests described in the proposal are helpful and support the implementation as being correct. I have high confidence that the two-tumor results are correct. The results for the multi-tumor models are encouraging, but not iron-clad. As noted above, I would have liked to see more documentation as to how the constrained optimizations for the multi-tumor models were implemented. The constraint (equation 6) involves all beta values. Thus, in the profile likelihood

analysis the optimization over the tumor types must be done jointly (unlike in the point estimation, where it may be done separately for each tumor type). This substantially increases the computational burden, and raises greater risks for numerical errors, non-convergence, convergence to local modes, and convergence to boundary points. Thus it would have been helpful to see more extensive assessments of the performance of the MS-COMBO procedure for multiple tumor types.

(c) Do the test results indicate that the MS-COMBO model provides reliable, accurate and clear results? (Note: Reviewers are encouraged, but not required, to apply alternative statistical methods and software to validate the MS_COMBO results.)

I implemented the likelihood, score, and Hessian functions for the single-endpoint model. Using Fisher-scoring, I was able to obtain equivalent results as given in the proposal for the first test set.

A more thorough assessment of the profile likelihood procedure should use simulation studies, and consider how the boundary behavior is handled. A simulation study showing coverage probabilities for the population BMD in situations where some population parameters are nearly, or exactly zero would address this issue in part. In situations where the parameter estimate is near, or on the boundary, it often turns out that inference procedures based on likelihood ratios are conservative. Demonstrating this for the MS-COMBO model would reduce concerns about this issue.

Ambitious testing would also consider the sensitivity of the results to small deviations in the form of the model, including the assumption of independence, and the form of the dose/response curve.

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results that give you special cause for concern? If so, please describe your concerns and recommendations.***

All of my concerns are discussed above.

III. SPECIFIC OBSERVATIONS

I don't have any concerns about the text currently in the document. My only concerns relate to omitted material, as discussed above.