

Additional BMDS Dose-Response Models for Risk Assessment:

Quantal Models With Background Additive to Dose

*Quantal Models Reporting Dose-Response Function Slope With
Confidence Interval*

Multistage Weibull Time to Tumor Model

Disposition of Comments by External Reviewers

This document was drafted in July, 2007 and revised in mid-September, 2007, after EPA and Battelle conducted more testing and made necessary changes to some of the programs.

Background

Additional dose-response models for risk assessment constitute an FY07 APM (APM 493, MYP: HHRA LTG2).

NCEA-W, Quantitative Risk Management Group, undertook development of additional dose-response models of three types:

(1) versions of existing BMDS quantal models which have the background additive to dose, rather than additive to the response;

(2) versions of some frequently used BMDS quantal models (multistage, logistic, and log-logistic) that can report the slope of the dose-response function at a user-specified dose, with a confidence interval for the slope;

(3) a multistage Weibull time to tumor model.

Models with background additive to dose have been presented, discussed and analyzed mathematically in risk analysis literature. Adding these to the BMDS suite of models is thus a logical extension of BMDS capabilities, increasing analysts' ability to evaluate model uncertainty and to evaluate a wider range of models that may fit data well.

Adding the capability of quantifying the dose-response slope with a confidence interval is, at this stage, exploratory, and is expected to assist analysts in understanding the uncertainty attending 'extrapolation' from observed doses to lower doses.

The multistage Weibull time to tumor model requires data on time and context of tumor observations for individual animals. When such data are available, the model could provide more precise estimates than the BMDS 'cancer' (multistage) model. A time to tumor model accounts for the time course of tumor appearance and other sources of mortality over time, while the simpler quantal models use a count of tumors at the experiment's end (sometimes adjusted approximately for mortality differences among dose groups). Such data are occasionally available and have been used in some EPA risk assessments (e.g., those for 1,3-butadiene and dibromomethane). This software will estimate benchmark doses. A MSW time-to-tumor model of similar form was available commercially within "ToxRisk" software¹, which is now no longer commercially distributed or supported. The MSW software prepared by EPA differs substantially from ToxRisk.² It is simpler to use, but does not make any dose conversions or adjustments. It uses the profile likelihood method to find lower and upper confidence limits for a benchmark dose. (ToxRisk used an approximation which has some deficiencies under certain conditions).

¹ Toxicological Risk Assessment Program (1985) Developed by K. Crump, R. Howe, C. Van Landingham, and W. Fulton. Clement International Corporation, Ruston, LA, under contract to Electric Power Research Institute, Palo Alto, CA.

² EPA chose to develop software independently for a number of reasons, including the desire to thoroughly document software quality assurance, testing, and the associated theory and numerical methods.

Major Comments and EPA Responses

Comments are summarized by model type and then by major comment topics. Reviewers are cited below by initials: RK for Ralph Kodell, LR for Louise Ryan, WW for Webster West.

A. Dose-Response Slope and its Confidence Interval for Selected Quantal Models

1. Needs more testing to verify correctness of numerical results (WW).

Response: We believe the testing for this module was as thorough as that conducted for the background-dose modules (below). However, the documentation was unclear and will revise it to better demonstrate the thoroughness of our verification testing.

2. Need to evaluate coverage of confidence interval (WW), because this is a new implementation of profile confidence interval methods, rather than a simpler modification of existing BMDS code.

Response: We have done this for the multistage model, using a Monte Carlo method, documented in a separate report. The coverage is accurate.

3. Provide more detail about how SAS and Mathematica were used to verify results (LR).

Response: This will be done in the testing report. EPA will also archive the verification programs with the testing report.

4. Confusing terminology in output file report from the models (RK).

Response: This will be changed to read (changes are in red-line):

```
-----
Slope Confidence-Interval Calculations
-----
Unconstrained Likelihood:   -372.7778112 (donlp3)
Chi-Square Boundary:        -374.6985416
Alpha:                       0.05000
Restrict Gamma [0,1]:       Yes

User-selected dose (d0):    150.0000000
Slope at d0:                0.0073267
Confidence Interval:        [0.0021002, 0.0139010]
```

and the User Guide will better explain and interpret this part of the output report.

B. Background Additive to Dose for BMDS Quantal Dose-Response Models

1. Complete the testing using some extreme cases (RK, LR, WW).

Response: We will do this to verify numerical accuracy and 'graceful' handling of convergence failures and unusual results.

2. Restrictions on model parameters seem not always to be stated. Clarify distinction between cancer model and multistage in documents (RK). Why does the log-probit model have a constraint on the slope parameter to exceed one? (See Table 9 of background document. (LR)

Response: The model output reports do state restrictions, near the top. This comment may refer to the testing document or the user help files. In those documents, we have insured that parameter 'natural' restrictions are specified in a table and that they are discussed in the user help file. The user is referred to the Benchmark Dose Technical Guidance for advice on user-specified restrictions (e.g., coefficients non-negative and powers ≥ 1). We have added a brief note on consequences of parameters on boundaries, citing the Molenberghe and Verbeke paper (esp. in view of Dr. Ryan's observations about parameters on boundaries, see **D.1** below). We have clarified the distinction between the multistage and cancer models in our reports and in the user help file.

3. Change help-file language on dose-additive parameter to reflect the mechanistic motivation for this parameter (RK).

Response: This has been done to reflect Dr. Kodell's recommendations.

4. Logistic & probit models with new background response parameter - add discussion and guidance to user regarding the two-parameter (reduced) model, etc. (RK).

Response: Guidance to users will reflect Dr. Kodell's recommendations and the points mentioned in his comments. The user help file now states that a 2-parameter model (with parameters gamma and beta) can be fitted and compared to the current model with parameters alpha and beta, and to the 3-parameter model. [However, the scope of application is limited to response probabilities $> 50\%$ (including the control).]

5. Provide more detail about how SAS and Mathematica were used to verify results (LR).

Response: This will be done in the testing report. There was more in the Testing Document than in the background report cited by Dr. Ryan, but we will insure that the testing report provides more detail. EPA will also archive the verification programs with the testing report.

6. Convergence and identifiability issues (LR).

Response: We are creating a new report on convergence and identifiability issues, detailing the results of investigations by Battelle and EPA. This document will be cited in the testing report. The user help file will be changed to reflect our current understanding of the practical limits on applying some of these models (log-logistic, log-probit, and Weibull with background dose parameter) perhaps with a heuristic explanation of the reasons. EPA or Battelle will conduct some more testing to better identify the types of data associated with identifiability problems for these models.

7. Clarify explanations of new vs. current models. (LR).

Response: This applies to the User Help file and any background documents meant to provide an overview, comparing & contrasting old & new quantal models. We will work on this description and have it reviewed internally to be sure it is clear.

8. Awkward and nonsensical discussion in Testing document. (LR)

Response: This will be corrected. We will have Battelle review and edit the Testing and Methodology documents again.

9. More detail about how parameter constraints are applied in maximum likelihood and confidence limit computation (discuss statistical implications as well as numerical methods). (LR)

Response: This will be addressed. We will have Battelle revise the Methodology and Testing documents accordingly. (We believe the MSW documents already do this well.)

10. More detail in User Help file about constraints (multistage, esp.) and confidence limit computation. (LR)

Response: This is now addressed in a section titled Parameter Constraints. We cite key published papers, noting that Wald intervals are not asymptotically correct when a parameter is on a boundary of parameter space, but that likelihood-based confidence intervals for BMD and Extra Risk are asymptotically correct in this case.

11. Testing - compare new models to corresponding existing models by setting background parameter to zero for both (outcomes should then be identical). (WW)

Response: This will be done. Some such testing was done, but a more systematic description is needed.

12. Initialize parameters at different starting points in the MLE and profile likelihood optimizations to be sure that the optimization result is unique and/or global. (LR, WW)

Response. This is done in the MSW time to tumor model software. It is not done in the BMDS quantal models. Doing so has its own liabilities, esp. in the profile likelihood optimizations. We shall address this issue indirectly, by exploring the shape of the likelihood, and by evaluating coverage of the BMD confidence interval (see below).

13. Conduct further independent review of the software itself. (WW)

Response. We will look into the Journal of Statistical Software. At this point, the software has undergone internal and external review. The testing has been thorough and is well documented. We feel that further testing will not add value. However, for those modules that are released (and also those used internally for research), users will be advised that the modules are "beta" versions and they will be asked to comment on the model performance and especially to report any apparent errors or problems.

C. Multistage Weibull Time to Tumor Model

1. Use and Interpretation of parameter t_0 . (all) There were comments on the meaning and interpretation of t_0 , the lack of realism of making t_0 a constant, the fact that t_0 often occurs at a boundary of its parameter space, and estimability issues.

Response: EPA and Battelle staff had begun to question this model set-up and we expressed our reservations in the review documents. The reviewers' comments confirmed the need for a careful re-examination of the 'classical' MSW model. Battelle is drafting a report on these issues, due later in 2007.

We note that EPA's use of the MSW time to tumor model in risk assessment has been to estimate tumor onset for Incidental tumors. For that purpose, t_0 was fixed at zero (i.e., it is not present in the model) and then the likelihood contribution for observed tumors is correct.

While the time lag represented by t_0 is in reality almost certainly a random variable, it is not clear that treating it as a constant sacrifices much accuracy or precision. The MSW model was developed by experts who probably thought carefully about this choice. The assumption seems reasonable as an approximation if the distribution of t_0 is narrow compared with the distribution of times being modeled (times of onset or of death).

The reviewer reservations about the interpretation of t_0 were made in the context of the then-erroneous programming of the likelihood and incorrect constraint on t_0 , resulting in unreasonable estimates for t_0 (which we called to the attention of the reviewers). Since the review, EPA and Battelle identified and corrected these errors. The C code was revised to correctly estimate parameter t_0 , and the Methodology document was revised to

correctly state the constraint for t_0 and to correctly represent the likelihood. Now, t_0 is constrained to be less than the smallest observed time for a Fatal tumor observation (not an Incidental observation as previously stated), and the likelihood contribution for an Incidental observation made at $t < t_0$ has been revised accordingly (the gradient and hessian calculations also had to be revised).

Since these corrections were made, the MSW model gives plausible estimates (reasonably close to the true parameters) for large datasets produced by simulation, for Fatal tumors (generating a mixture of Incidental, Fatal, and Censored observations). For non-fatal tumors (i.e., pure current-status data), the model has been providing plausible estimates from the start, and providing a parametric estimate that agrees closely with the nonparametric estimate given by `survfit(Surv(x, ...))` in S-Plus.

Estimates of t_0 in the MSW model for Fatal tumors are now less often at a boundary (i.e., zero or the upper constraint value). The “MSW Time to Tumor model description for users” provides a warning about confidence limits when a parameter estimate takes on a boundary value.

Thus, we now believe that the software is functioning well and giving correct estimates for the MSW model for fatal tumors and the MSW model for non-fatal tumors.

Technical Comments from Battelle. In current versions of the software documentation, the interpretation of t_0 as the time between tumor onset and death from tumor is correct. Based on the derivation of the distribution function for tumor onset in item **C4.** below, the cumulative distribution function $F(t)$ for death from tumor at time t is equal to the probability from tumor onset by time $t - t_0$, i.e.,

$$F(t) = G(t - t_0) = 1 - \exp\{-(t - t_0)^c p(d)\}$$

The only way to interpret $F(t)$ as a time-to-onset distribution is to change, as Dr. Kodell suggests, the interpretation of t_0 to the one used traditionally for the location parameter in the 3-parameter Weibull distribution, i.e., the earliest possible time (after time 0) when tumor onset can occur. For non-fatal tumors, this change introduces an additional parameter which creates the same types of problems with estimation of t_0 as that seen in the multistage Weibull fatal tumor model. For the fatal tumors, death from tumor will no longer be a modeled tumor response, and the analysis will therefore be the same as for the non-fatal tumor model, except subjects categorized as Fatal (F) tumor context would be re-categorized as Incidental (I) tumor context. This nullifies the assumption that the censoring distribution is conditionally independent from the tumor onset distribution, and requires the censoring distribution to be specified in the model likelihood. In addition, by ignoring the additional information provided by the category for Fatal (F) and Unknown (U) tumor contexts, the statistical analysis is less powerful because the data are being “diluted”. We recommend that the original interpretation of $F(t)$ as a time-to-death (from tumor) distribution be retained within the software until sufficient time has been spent to properly investigate and understand the implications of such a change.

2. Parameter Constraints. (RK, LR) Not clearly stated.

Response: The parameter constraints are now clearly stated in the User Help File.

3. Testing. (a) Has MSW been tested for BMD calculation at times other than 104 weeks? (RK) (b) Test with more datasets and data configurations (WW). (c) It is also interesting that t_0 is estimated to be zero in every single case. Clearly, more diversity in this regard is desirable. (WW)

Response: (a) MSW has been tested for BMD calculation at times other than 104 weeks; this will be described in the testing document. (b) This will be done. (c) In all examples provided, " t_0 " was fixed at zero, not estimated. We are adding test results showing estimation of $t_0 > 0$.

4. Constraint on parameter "c". Dr. Kodell suggests that it may be appropriate to restrict " c " \geq number of dose-related stages (" k ", the highest power of dose in the model). (RK)

Response. We will need time to evaluate this idea. To adopt it, we will need to justify it carefully in the Methodology document. We have asked Battelle to give their expert opinion (below), and we will need to review the literature on this subject. Currently, users could fix " c " at several values to evaluate sensitivity of estimates to " c ," so we might provide such a suggestion to users.

Technical Comments from Battelle. Dr. Kodell's interpretation of the shape parameter c as the total number of stages appears to be based on the derivation of the statistical multistage Weibull model from the physical multistage model for describing a particular mechanism in carcinogenesis. (This speculation needs to be verified either by Dr. Kodell or some other subject matter expert.) In the physical model, cancer develops by sequential stages of genetic cell mutation, and the hazard rate of transitioning from one stage to the next is assumed to be constant in time. If tumor onset occurs at stage c of the cell mutation process, and r_i is the hazard rate of transition from stage $i-1$ to stage i , then Armitage (1953) derives the hazard $h(t)$ of tumor onset at time t to be approximated by

$$h(t) = \prod_{i=1}^c r_i t^{c-1} / (c-1)!$$

provided that values of $r_i t$ are small. By definition of the hazard function, the probability $G(t)$ of tumor onset by time t (i.e., the cumulative distribution function for tumor onset) is therefore

$$G(t) = 1 - \exp\left\{-\int_0^t h(\tau) d\tau\right\} = 1 - \exp\left\{-\prod_{i=1}^c r_i t^c / c!\right\}$$

For the multistage Weibull, the stagewise hazard rates r_i are assumed to be polynomials of dose d , so that $p(d) = \prod_{i=1}^c r_i(d) / c!$ is also a polynomial in d of some degree k . This explains Dr. Kodell's assertion of the shape parameter c (rather than k) as the actual number of stages. Under the additional restriction that the rates $r_i(d)$ are linear in d , the degree k is less than or equal to c , because some of the hazard rates could be constants. In light of this derivation, Dr. Kodell's recommendation of limiting $c \geq k$ and c as a positive integer makes sense. Nevertheless, before changing the software to account for these restrictions, some major preliminary steps are necessary. In particular:

- Certain performance features of the software (e.g., stability, efficiency) in its current capability needs to be improved, and
- The statistical and numerical implications of these new restrictions need to be properly understood.

5. Initialize parameters at different starting points in the MLE and profile likelihood optimizations. (LR, WW)

Response. This is done in the MSW time to tumor model software.

6. The example output shown for the MSW model does not appear as well laid out as it might be. (RK)

Response. We will review this, but note that it is very similar to current BMDS output formats.

7. Concerns related to the report "MSW Time to Tumor model description for users" (LR)

Response. Using the same letters as in Dr. Ryan's comments:

(a) The paragraph has been changed to be, we hope, unambiguous now. Non-fatal tumors require a different model (MSW model for non-fatal tumors), explained in the document. Dr. Ryan probably meant "how are Incidental tumors handled?", and that is now explained in the Methodology document. We have added to the section on Maximum Likelihood Estimation a passage explaining the likelihood contributions of observations with each context (C,I,F, and U), which should answer that question.

(b) See response under (e).

(c) This will be corrected

(d) "Subject group" is now defined explicitly. The logic of the likelihood

contribution for Incidental tumors is explained in the Methodology document and now in a new section added to the User Description (see (a) above). The User Description document warns about the assumption of fixed t_0 . However, the model was developed and used by experts and the assumption seems reasonable as an approximation if the distribution of t_0 is narrow compared with the distribution of times being modeled (times of onset or of death). The likelihood contribution suggested by Dr. Ryan for Incidental observations is indeed used in the model for non-fatal tumors, but is not appropriate for modeling time to death from fatal tumors. In the latter case, an Incidental observation could be treated as left-censored (like a "C" observation), or treated as we did, as interval censored, with death from tumor predicted to occur between time t and $t+t_0$. Regarding the first term in the log-likelihood: $F(tjs, d)$ is increasing in tjs , so the suggested change would produce a negative likelihood component for Incidental tumors, i.e.,

$$F(tjs-t_0, d) - F(tjs, d) < 0.$$

To explain, for an incidental tumor context for fatal tumors at time t , tumor onset must have occurred some time between $t-t_0$ and t ; therefore, death from tumor must occur some time between t and $t+t_0$. The cdf $F(t, f)$ represents death from tumor by time t , the probability of death from tumor between t and $t+t_0$ is $F(t+t_0, d) - F(t, d)$

(e) The description of BMDL computation is indeed terse. The language is modeled after that in the BMDS help files. However, this computation is explained better in the Methodology document. Typical users would gain nothing from a detailed explanation. The current explanation should make sense to a user who has read another part of the BMDS user help files which explains two methods of solving for BMDL. The impact of parameter restrictions on inference is now explained briefly. However, for the BMD confidence interval, we expect parameters on the boundary to have no effect on the 2-sided confidence interval coverage for the BMD, and we are still investigating the 1-sided (BMDL) coverage, so any further advice to users would be premature.

8. Concerns related to the “MSW Time to Tumor model description for users” (WW). A bit heavy on mathematical detail, a bit light on implementation tips.

Response. We have added more on implementation that should help the user. For the data format, the user is directed (at the beginning of this document) to another document titled "User Guide", which gives instructions for running the program and assembling data files. We have combined the two documents into one "User Guide". We will continue to improve this document based on feedback from users.

9. “Documentation on issues that are more closely aligned to statistical concepts is often poorer” (LR).

Response. Dr. Ryan’s statement appears to criticize an apparent lack of exposition on statistical issues. Dealing with statistical issues is complicated by the fact that

- The multistage Weibull model (for fatal tumor) fails to satisfy the standard regularity conditions for asymptotic normality of maximum likelihood estimators, because the support of the likelihood over the t_0 parameter is dependent on the data (and the Fisher information is not finite if shape parameter $c \leq 2$), and
- Time-to-tumor experiments are subject to a complex censoring scheme based on tumor contexts.

In a broad sense, Chapter 7 of the multistage Weibull Time-to-Tumor Methodology Description document covers some of the issues, but a significant effort of theoretical research would be necessary to provide more detailed insight. Nevertheless, Dr. Ryan's concerns about parameter estimates that end up on the boundary of the parameter space (although not specifically directed at just the multistage Weibull model) are important and should be addressed in some fashion within the documentation.

10. Specification of $(t-t_0)$ term in the model. (LR).

Response. Dr. Ryan suggests that the definition of the distribution function associated with the multistage Weibull should be simplified by replacing $(t-t_0)^c$ with $(t-t_0)_+^c$, and states, "*I don't think it is correct to simply say that there is a restriction that $t > t_0$* ". Our restriction was based on the statistical literature for the 3-parameter Weibull distribution, where the location parameter is commonly restricted in this way. Dr. Ryan's objection may be based on a semantic difference in the interpretation of the location parameter between the multistage and 3-parameter Weibull models. After examining the two options, we have concluded that Dr. Ryan's definition of the distribution function is correct and we have adopted it.

A particular benefit can be seen with the parameter t_0 . In our original definition, the parameter space for t_0 was bounded above by the minimum observation time in the data (except for observations with tumor contexts C). This was puzzling, because a parameter describing a feature of the tumor response (time between onset and death from tumor) was being constrained by the censoring mechanism (sacrifice or premature death); however, the censoring mechanism was supposed to be independent from the tumor response. Dr. Ryan's definition removes this constraint, and limits the MLE for t_0 to lie somewhere between 0, and the minimum observation time for fatal (F) tumor contexts. (This restriction, in turn, is consistent with the assumption that subjects are tumor-free at time 0).

Both the software and documentation will be changed to implement the revised definition for the distribution function.

D. Comments that Apply Generally to BMDS Models (as well as the models under review).

1. Parameter constraints; parameters at boundaries. There is a need to deal with statistical issues related to parameters on boundary of parameter space. (e.g., coefficients equal to zero), which complicates statistical inference for confidence intervals (LR).

Response:

Fortuitously, NCEA-W staff were aware of this issue and developing ways to deal with it before this review began. We have read the two cited papers and all others we could find that pertain to this issue. This issue pertains to dose-response modeling generally, not merely to BMDS models (it applies very generally to hypothesis testing for linear and nonlinear models). How we are dealing with this matter is discussed in two parts below.

(Part i) The theory of Self and Liang (1987) requires specialization to our dose-response models. After that mathematical foundation is laid, numerical evaluations of coverage are needed for profile-likelihood confidence intervals and for the Self-Liang adjustment to Wald intervals. This will allow us to quantify the coverage when one or more parameters are on a boundary, and will provide evidence to support practical measures or advice applicable to BMDS.

We have accomplished much of this objective already. Paul White (QRMG Chief) arranged for Dr. Bimal Sinha (University of Maryland - UMBC) to hold a fellowship at NCEA-W and work on this matter with NCEA staff. A first report, laying out the application of methods to the multistage model, is published as a technical report (http://www.math.umbc.edu/~kogan/technical_papers/2007/Sinha_Kopylev_Fox.pdf) and a revised version will later appear as a publication of Indian Statistical Institute. We are now working on numerical evaluations of coverage under various conditions and expect that to be available late in 2007.

All of this work applies, at present, only to the multistage model, which is important because it is used so frequently in cancer dose-response modeling. We will need to extend this work to other the models. We are hopeful that this extension will be fairly straightforward, but until it is made we will provide suitable caveats in the User Help file that was part of this external review.

(Part ii) We will need to provide suitable information and advice to typical users of dose-response models and BMDS. As Dr. Ryan suggested, we will provide some suitable text in either the BMDS help files or output reports or both. This will probably have to be done by introducing this material in the Help file for the new models under review here and then 'catching up' the existing models later. We will also seek to publish an article in a widely read journal, in language much less technical than the report noted above, in order to make these issues and changes known to the wider community of risk assessment professionals who use dose-response models.

The bottom line is that, although confidence intervals for model coefficients are adversely effected when a parameter is on a boundary, the BMD and Extra Risk (or Added Risk) are always interior to their parameter space and their likelihood ratio test statistics are asymptotically distributed as Chi-square(1). This conclusion is developed rigorously in Dr. Sinha's paper. Thus the theory is now on a solid footing; it remains to be sure that it is implemented well in the BMDS software. Therefore, we are conducting Monte Carlo evaluations of BMD confidence interval coverage to verify that the BMDS quantal models in fact achieve the nominal coverage with huge sample sizes (n=100,000), and to see how the coverages (1 and 2 sided) behave for typical sample sizes.

2. Monte Carlo (simulation) studies in which empirical data are repeatedly generated from a known true model and then the results of fitting various models compared to the expected true values (LR, WW).

Response:

This would be a major undertaking, which we do not have the resources for currently. However, the studies described just above (for the multistage model) accomplish this objective in part, by checking on the distribution of the benchmark dose and extra risk and parameter estimates. We will evaluate our contract resources available for having Battelle conduct a limited number of evaluations for 1-3 background-dose models, using R code already developed for the studies described just above.

Dr. Ryan also suggested that the Figures in the User Help file (for the quantal models) would be more informative if based on simulations (Monte Carlo). In our experience, this would be appreciated by statisticians but would confuse the typical user. We believe it best to create a separate technical report on Monte Carlo studies (with suitable illustrations as suggested).

We are evaluating confidence interval coverage for the BMD for selected BMDS ver. 1.4.1 quantal models, using the suggested Monte Carlo approach. Once the programs had been developed for the evaluation of slope CI coverage, in response to Dr. Wheeler's review, they were easily modified for this purpose. It will take months to complete this work, because of the need to evaluate various data configurations, and the urgency of other work. Results for the multistage (for "nice" data) show coverages close to nominal.

We expect that measures described in the three paragraphs above will take place over late 2007 - early 2008 on a schedule separate from release of the new models.

3. Explore likelihood surface for multiple local maxima, esp. at boundaries of parameter space (WW). [this is related to reviewer (LR, WW) suggestions to incorporate multiple starting values for optimizations in maximum likelihood and confidence interval computations]

Dr. West observed: *The likelihood surfaces for the dose response models to be fit are extremely bumpy with sometimes peculiar boundary behavior for the common designs used. For example, with the multistage model, there are frequently values along the β_1 axis (where $\beta_2=0$) and values along the β_2 axis (where $\beta_1 = 0$) that provide roughly the same value of the likelihood function.* If it is correct, this is a novel observation that deserves thorough documentation and publication.

Response: Battelle has conducted limited evaluations of this sort, and will document those. Regarding further work, Battelle has been asked to evaluate Dr. West's comments and suggestions about exploring the likelihood surface using a grid and their response appears just below). We will evaluate the feasibility of such an exploration but make no commitment to a thorough investigation at this time, until we have evaluated the effort and cost required. This does not mean that the issue should be ignored. Instead, we intend to devise an incremental approach, selecting one model (one that does not have identifiability issues; probably the multistage) for a limited investigation. We have done such analyses for the log-logistic model with background dose parameter, as discussed in the review documents, and the findings will be reported in a new document on convergence issues. Revisions and corrections to the MSW model for time to tumor (noted above) appear have eliminated the irregular, multi-modal likelihood profile curve we observed for an earlier version of the model for fatal tumors

This exploration of the likelihood is also pertinent to Dr. Ryan's comments about identifiability. For some data configurations, the log-logistic model with background additive to dose shows a very shallow 'ridge' in the log-likelihood surface, and the maximum likelihood search stops short of the MLE. In such cases, the parameter standard errors (i.e., information matrix diagonal elements) may be very large, and parameter correlations may be essentially +/- 1. A warning about such indications of near-non-identifiability has been added to the User Help File (although the log-logistic model with background additive to dose is not being released, the other models can exhibit similar behavior for some datasets).

Response from Battelle: Dr. West expresses some concern on the reliability of model estimates from TOXRISK and proposes that multistage Weibull parameter estimates from BMDS be “*compared to the true optima in a number of test cases where the truth is determined by a complete grid search of the parameter space*” for purposes of testing to a better standard. In fact, the BMDS module for the multistage Weibull model selects starting values for estimating the model parameters by carrying out a complete grid search on two of the parameters and maximizing the likelihood over the remaining parameters. (Appendix C in the multistage Weibull Time to Tumor Methodology Description proves that the likelihood is concave over those remaining parameters.) Therefore, the BMDS module “globally” maximizes the multistage Weibull likelihood, under the assumptions that the search grid is sufficiently fine and the *donlp3* optimization is functioning properly. Therefore, a complete grid search seems unnecessary and is not recommended. While setting up a complete grid search may require only a couple of

days to program, a single execution of a complete grid search may take a few hours to a day of execution time, depending on the number of stages in the model.