Some New Aspects of Statistical Inference for Multistage Dose-Response Models with Applications

Bimal K. Sinha
Department of Mathematics & Statistics,
University of Maryland, Baltimore County, USA
sinha@math.umbc.edu

Leonid Kopylev
National Center for Environmental Assessment,[i]
Office of Research and Development,Environmental Protection Agency, USA
kopylev.leonid@epa.gov

John Fox
National Center for Environmental Assessment,[ii]
Office of Research and Development,Environmental Protection Agency, USA
fox.john@epa.gov

Abstract

Applications of the results to dose-response multistage models serve as illustrations. The paper considers various scenarios, and is expected to be of interest to practitioners of risk analysis.

Keywords: Benchmark dose, BMDL, Confidence limits, Dose-response models, Likelihood ratio tests, Maximum likelihood estimates, Multistage Weibull model, Parameters on the boundaries, Profile likelihood

1. Introduction

1.1 Dose-Response Models and Risk Estimates

Quantitative risk assessment for toxic and carcinogenic chemicals relies largely upon fitting dose-response models to data from animal bioassays. A variety of models are in use (Krewski and van Ryzin, 1980, Filipsson *et al*. 2003, U.S. EPA 2006) to represent both quantal (dichotomous) and continuous responses. Toxicological experience and principles indicate that the response will generally be bounded and non-decreasing (Eaton and Klaassen 2001).

Quantal response models represent the probability $P(d,\theta)$ of a quantal response, like presence or absence of a particular type of cancer, in relation to dose (d). The experimental data consist of counts of animals exposed to a chemical, the numbers exhibiting the response, and the dose levels (e.g., $n_i, x_i, d_i, i = 0,1,...,m$), from a bioassay

conducted with mice or rats, typically in 3--5 dose groups including the control, each having 10--50 animals.

The maximum likelihood estimation for the parameters θ employs the binomial likelihood

$$L(\theta \mid x) = \prod_{i=0}^{m} \binom{n_i}{x_i} P(d_i, \theta)^{x_i} \left(1 - P(d_i, \theta)^{n_i - x_i}\right) \tag{1}$$

The benchmark dose method (Crump 1984, Filipsson *et al*. 2003, Parham and Portier 2005) consists of estimating a lower confidence limit for the dose associated with a specified increase γ in adverse response (i.e., increased risk) above the background level. In practice, the specified increase is typically 1% to 10% for cancer quantal response models (Filipsson *et al*. 2003, Parham and Portier 2005).

We will use "absolute risk" to refer to the probability $P(d, \theta)$ modeled by a dose response model for a quantal response. The increase in risk above background for a quantal response is quantified as "extra risk" or "additional risk" (Filipsson *et al*. 2003). These quantities are defined below.

$$\text{Absolute Risk}: \quad AR = P(d, \theta) \tag{2}$$

$$\text{Additional Risk}: \quad ADR = P(d, \theta) - P(0, \theta) \tag{3}$$

$$\text{Extra Risk}: \quad ER = \frac{P(d, \theta) - P(0, \theta)}{1 - P(0, \theta)} \tag{4}$$

The benchmark dose may be determined for any of these risk types. For example, the benchmark dose for extra risk of γ is the solution of

$$\text{Benchmark Dose (BMD)}: \quad d^* : \gamma = \frac{P(d^*, \theta) - P(0, \theta)}{1 - P(0, \theta)} \tag{5}$$

1.2 Statistical Inference

Statistical inference for chemical risk assessment has mainly emphasized finding confidence limits for the dose associated with a specified risk and for the risk associated with a specified dose. In this context, the profile likelihood method as applied in Crump and Howe (1985) and in U.S. EPA's Benchmark Dose Software (U.S. EPA 2006) assumes that $-2log(L(\theta \mid \mathbf{x}))$ is distributed as χ_1^2. This is correct, asymptotically, under certain regularity conditions (Rao 1973, Cox and Hinkley 1974), one of which is that the true parameters are interior to the parameter space (for more details, see: Chernoff (1954), Feder (1968) and Self and Liang (1987); see also Molenberghs and Verbeke (2007) for a nice summary with applications). However, when one or more parameters (those of interest, or nuisance parameters, or both) lie on the boundary of the parameter space, the distribution of the likelihood ratio test statistic may not be χ_1^2 and may be difficult to derive (Self and Liang 1987). Moreover, some older software found upper confidence limits by the profile likelihood method using linear approximations to the risk and benchmark dose that are functions of a single parameter (aka 'q*' method). This method may be questionable if some parameters are on a boundary. Also, confidence

limits of model parameters are often approximated by Wald intervals, which are known to be inaccurate (Bailer and Smith 1994, Moerbeek *et al*. 2004, Nitcheva *et al*. 2007). These problems were acknowledged long ago (Crump *et al*. 1984), but have not been resolved clearly for the practitioners of dose-response modeling.

Research problems

Our primary goal in this paper is to develop appropriate asymptotic statistical methods in a very general multi-parameter framework when some parameters may lie on their boundaries. Our focus is mainly on the asymptotic properties of the MLEs and the LRTs. The problem of their actual computation, which involves use of sophisticated computer software and codes, is not discussed here. As an important application of the asymptotic theory, we discuss in detail inference about the parameters θ_j in the multistage model (6) and about the quantities of interest, namely, AR, ADR, ER and BMD, when some of the basic parameters may lie on their boundaries. This is a significant point that has not been properly addressed in the relevant literature dealing with such models.

In the sequel, we discuss one commonly used model for quantal responses, called the multistage model:

$$P(d,\theta) = 1 - exp(-\sum_{j=0}^{k} \theta_j d^j) \tag{6}$$

In applications, the coefficients θ's are often constrained to be non-negative so that the dose-response function will be non-decreasing.

While AR, ADR and ER are direct and simple functions of θ, the benchmark dose $d*$ (5) can obviously be a complicated function of θ, especially when k is large. To circumvent this potential difficulty, we proceed in an alternative fashion. Note that $d*(\theta) = \psi_\gamma(\theta)$ satisfies:

$$ln(1/(1-\gamma)) = \theta_1 d* + \theta_2 d*^2 + \cdots + \theta_k d*^k \tag{7}$$

We propose to test $H_{0\delta} : \psi_\gamma(\theta) = \delta$ versus $H_{1\delta} : \psi_\gamma(\theta) \neq \delta$ for a given $\delta > 0$. If $H_{0\delta}$ is accepted, we include δ in the confidence interval for $\psi_\gamma(\theta)$. Otherwise, we exclude it. We carry out this procedure for *all* possible values of δ, thus hopefully generating an interval of *accepted* δ-values, all belonging to the acceptance set, and thereby leading to the lower and upper bounds of δ, which is $\psi_\gamma(\theta)$. Since $\psi_\gamma(\theta)$ is linear in θ, a test for H_0 versus H_1 is easily carried out! In fact, since $ln(1/(1-\gamma)) = \theta_1\delta + \theta_2\delta^2 + \cdots + \theta_k\delta^k$ is equivalent to $ln(1/(1-\gamma))/\delta = \theta_1 + \theta_2\delta + \cdots + \theta_k\delta^{k-1}$, we can work with a re-parametrization of the parameters $(\theta_1, \theta_2, \cdots, \theta_k) \rightarrow (\theta_1* = \theta_1 + \theta_2\delta + \cdots + \theta_k\delta^{k-1}, \theta_2, \cdots, \theta_k)$ with θ_1* staying away from the boundary even if some of the θ's may lie on their boundaries.

The organization of the paper is as follows. We develop at length in Section 2, the core section of the paper, the necessary statistical inference in a very general multi-parameter framework with some parameters possibly lying on their boundaries, borrowing ideas

from a host of *key* papers on this topic, notably from Self and Liang (1987). We develop both profile likelihood based inference (LRT) as well as Wald-type inference for a variety of situations with respect to boundary parameters. These results can be applied to a wide variety of models used in dose-response analysis. Applications to the specific multistage models (6) are discussed in Section 3. Simulations for linear, quadratic and cubic multistage models are reported in Section 4. These simulations clearly reveal that the use of Self & Liang's procedure over the Wald procedure considerably improves the expected lengths of the confidence intervals for all the relevant parameters. Some concluding remarks and directions for future research are mentioned in Section 5. An Appendix at the end contains a proof of a basic result of the paper.

2. New results on boundary value problems

Let us recall the general discussion mentioned in the previous section about the asymptotic properties of the MLEs in multi-parameter problems and the asymptotic distribution of the likelihood ratio test of a function of such parameters. Let X denote a random data set which is typically a collection of N independent and identically or independent but non-identically distributed random variables and let $\theta = (\theta_1, \cdots, \theta_p)$ denote a p-dimensional real parameter vector which governs the distribution of X through a joint density $f(X | \theta)$ or equivalently the likelihood function $L(\theta | X)$. Let us assume that $\theta \in \Omega \in R^p$ and we write $\Omega = \Omega_1 \times \cdots \Omega_p$ where it is assumed that $\theta_i \in \Omega_i$, $i = 1, \cdots, p$. We also assume that the p parameters $\theta_1, \cdots, \theta_p$ are functionally independent. Keeping the specific dose-response multistage Weibull models in mind, we further assume that $\Omega_i = [\theta_{i0}, \infty)$, a half-closed interval or $\Omega_i = (\theta_{i0}, \infty)$, an open interval, where θ_{i0} is specified. In the former case, θ_{i0} is referred to as a boundary point of θ_i, and in the latter case, all points of θ_i are interior points. In most dose-response models, $\theta_{i0} = 0$ or 1 for all i. The statistical problem in such a set-up is to estimate the parameters θ and test suitable hypotheses about θ or some functions thereof based on the data set X. Since often the joint density or the likelihood function can be quite complicated as in the case of dose-response multistage Weibull models, both estimation of θ and tests about θ are carried out using suitable asymptotic theory under the assumption of a large sample size N.

It is well known that, under some very general conditions, the asymptotic theory of estimation based on the maximum likelihood estimates (MLEs) and the asymptotic theory of tests based on the likelihood ratio tests (LRTs) are valid and provide useful tools for meaningful statistical inference. Typically, we conclude the asymptotic normality of the MLE $\hat{\theta}_N = (\hat{\theta}_{1N}, \cdots, \hat{\theta}_{pN})$ of θ and the asymptotic chi-square distribution of $-2ln(LRT)$ with a suitable *df* under a null hypothesis. However, it should be noted that the validity of such results is based on a *crucial* assumption that the true state of nature of the unknown parameter vector θ is an interior point of Ω. In many applications, including multistage dose-response Weibull models, it can often hold that some parameter spaces are half-closed and, in fact, some parameters may actually lie on their respective boundaries, thus making the standard asymptotic results on MLEs and

LRTs unjustified and incorrect. Fortunately, this point has indeed been seriously addressed in the literature and a series of papers concerning this vital issue have appeared, most notably by Self and Liang (1987).

Based on Chernoff (1954), Feder (1968), Moran (1971), Chant (1974), Fahrmeir and Kaufman (1985), Self and Liang (1987), Geyer (1994), and Vu and Zhou (1997), the following two general results can be stated under fairly standard regularity conditions on the joint density of X and nature of θ. The conditions are typical Cramer-type and are satisfied in our applications to multistage dose-response models.

Let $\hat{\theta}_N$ denote the MLE of θ when the likelihood function $L(\theta \,|\, X)$ is maximized wrt $\theta \in \Omega$. Recalling the very general nature of Ω, the maximization of the likelihood with respect to (wrt) θ might mean unrestricted maximization in an open set (actually *product* of open intervals) or a restricted maximization in a product of open and *half-closed* intervals, depending on which parameters can assume their boundary values. We will call these as *natural* restrictions on θ and reserve the use of the term *restricted* maximization to the situation when the components of θ are restricted by a null hypothesis of the form $H_0 : \theta \in \Omega_0$. Assume $\theta^* = (\theta_1^*, \cdots, \theta_p^*)$ to be the true value of the vector parameter θ and let $I(\theta^*)$ be the Fisher information matrix evaluated at $\theta = \theta^*$, which is assumed to be positive definite with $\Sigma(\theta^*) = [I(\theta^*)]^{-1}$. For ready reference, we also mention that the (i, j)th element of the information matrix $I(\theta^*)$ is computed as $E[-\dfrac{\partial^2}{\partial \theta_i \partial \theta_j} lnL(\theta \,|\, X)]$ where the expectation is evaluated at θ^*. It is understood that in case of boundary points, the derivatives are computed from appropriate directions.

Our first result is concerned with the asymptotic distribution of the *MLE* $\hat{\theta}_N$ of θ under the true value θ^* where θ^* is either an interior point in Ω or a combination of both interior and boundary points of Ω.

Proposition 2.1. *The asymptotic distribution (as $N \to \infty$) of the p-vector $\sqrt{N}(\hat{\theta}_{1N} - \theta_1^*, \cdots, \hat{\theta}_{pN} - \theta_p^*)$ is the same as the (exact) distribution of the MLE $\hat{\theta}(Z)$ of θ based on a normal p-vector Z with mean θ and dispersion matrix $\Sigma(\theta^*)$ under $\theta = 0$.*

Here the MLE $\hat{\theta}(Z)$ of θ based on Z is computed under the assumption that the mean vector θ of Z lies either in an open set in R^p including the point 0, corresponding to the case when $(\theta_1^*, \cdots, \theta_p^*)$ is an interior point in Ω with reference to the distribution of X, or in a product of open and half-closed intervals of the type $(-\infty, \infty), [0, \infty)$, whose nature depends on which true parameters θ_i^*'s are the boundary points θ_{i0}'s of θ with reference to the distribution of X. Thus, if $\theta_i^* = \theta_{i0}$ for some i in the distribution of X, we take $\theta_i \in [0, \infty)$ in the distribution of Z while maximizing its *pdf* wrt θ_i. Otherwise, we take $-\infty < \theta_i < \infty$.

When in fact $\theta_i^* \geq \theta_{i0}$, the maximization of the normal *pdf* of Z is carried out wrt θ, taking $\theta_i \geq 0$. Under these conditions, it is no longer true that Z is the MLE of θ based on Z, and hence the asymptotic normality of $\hat{\theta}_N$ does not hold! Moran (1971), Chant (1974) and, most importantly, Self and Liang (1987) have clearly spelled out the MLEs of θ based on Z under various scenarios of boundary points. Once the MLEs of θ based on Z are identified and their joint distribution under normality of Z is derived, this would provide the asymptotic joint distribution of $\sqrt{N}(\hat{\theta}_{1N} - \theta_1^*, \cdots, \hat{\theta}_{pN} - \theta_p^*)$. Because of the nature of Ω assumed above, it will hold that $\hat{\theta}_{iN} \geq \theta_{i0}$, for all i. Thus, in the exact and asymptotic distributions of $(\hat{\theta}_{iN} - \theta_i^*)$, it will hold that this difference can assume both positive and negative values when θ_i^* is an interior point of $\Omega_i = (\theta_{i0}, \infty)$, and is nonnegative when $\theta_i^* = \theta_{i0}$, the boundary point of $\Omega_i = [\theta_{i0}, \infty)$.

One major goal of this paper is to further develop this part when we have one, two or three boundary points, clearly explaining the joint distribution of the resultant MLEs of θ based on Z which then readily yields the asymptotic joint distribution of the MLEs of θ under X. This knowledge is useful when one is interested in developing Wald-type inference about a smooth function of θ based on the MLEs $\hat{\theta}_N$ of θ.

To be specific, we establish the following results in Section 2 concerning the asymptotic distribution of the *MLE* $\hat{\theta}_N$ of θ, which is computed by maximizing the likelihood function $L(\theta \mid X)$ wrt $\theta \in \Omega$.

- The asymptotic joint distribution of $\sqrt{N}(\hat{\theta}_{1N} - \theta_{10}, \hat{\theta}_{2N} - \theta_2^*, \cdots, \hat{\theta}_{pN} - \theta_p^*)$ with one boundary point $\theta_1^* = \theta_{10}$, $\Omega_1 = [\theta_{10}, \infty)$, $\theta_i^* > \theta_{i0}$, $\Omega_i = (\theta_{i0}, \infty)$, $i \geq 2$.

- The asymptotic joint distribution of $\sqrt{N}(\hat{\theta}_{1N} - \theta_{10}, \hat{\theta}_{2N} - \theta_{20}, \hat{\theta}_{3N} - \theta_3^*, \cdots, \hat{\theta}_{pN} - \theta_p^*)$ with two boundary points $\theta_i^* = \theta_{i0}$, $\Omega_i = [\theta_{i0}, \infty)$, $i = 1, 2$, $\theta_j^* > \theta_{j0}$, $\Omega_j = (\theta_{j0}, \infty)$, $j \geq 3$.

- The asymptotic joint distribution of
$\sqrt{N}(\hat{\theta}_{1N} - \theta_{10}, \hat{\theta}_{2N} - \theta_{20}, \hat{\theta}_{3N} - \theta_{30}, \hat{\theta}_{4N} - \theta_4^*, \cdots, \hat{\theta}_{pN} - \theta_p^*)$ with three boundary points
$\theta_i^* = \theta_{i0}$, $\Omega_i = [\theta_{i0}, \infty)$, $i = 1, 2, 3$, $\theta_j^* > \theta_{j0}$, $\Omega_j = (\theta_{j0}, \infty)$, $j \geq 4$.

In each case, we also describe the asymptotic distribution of $\sum_i c_i \hat{\theta}_{iN}$ which can be used to draw suitable inference about $\sum_i c_i \theta_i$. In particular, these results can be directly used to derive Wald-type tests of hypotheses concerning linear functions of θ, without an appeal to the alternative *profile* likelihood method. We will illustrate this point later by some applications.

We now describe another important aspect of the papers by Moran (1971), Chant (1974), and Self and Liang (1987) in the context of the derivation and properties of the likelihood ratio tests for θ based on X when some parameter points may lie on the boundary. To make matters simple and easy to understand, let us consider the following two testing problems about just one component, say θ_1, of θ.

Problem 2.1. *Test the null hypothesis* $H_0 : \theta_1 = \theta_{10} + \delta$ *versus* $H_1 : \theta_1 \neq \theta_{10} + \delta$ *where* $\delta > 0$. *Here naturally* θ_1 *is an interior point of* Ω_1.

Problem 2.2. *Test the null hypothesis* $H_0 : \theta_1 = \theta_{10}$ *versus* $H_1 : \theta_1 > \theta_{10}$ *where* θ_{10} *is the boundary point of* $\Omega_1 = [\theta_{10}, \infty)$.

It is well known that when both Ω, the general parameter space, and Ω_0, the parameter space under a null hypothesis, are *smooth* in the sense of being open sets in R^p and in a lower dimensional subspace, respectively, the asymptotic distribution of $-2ln(LRT)$ under the null hypothesis is central chi-square with an appropriate *df*. However, this result is far from being true when some parameters may lie on the boundary either under the null hypothesis or even otherwise. It is precisely in the context that Ω and Ω_0 may not be open sets in R^p that we have the following general result primarily due to Self and Liang (1987). We remark that Self and Liang's results on the asymptotic distribution of the *LRT* are quite general, but here we state the results keeping in mind the two null hypotheses given above under Prob. 2.1 and Prob. 2.2.

Proposition 2.2.

(a) Consider the problem of testing the null hypothesis $H_0 : \theta_1 = \theta_{10} + \delta$ versus $H_1 : \theta_1 \neq \theta_{10} + \delta$, for some $\delta > 0$ when $\Omega_1 = (\theta_{10}, \infty)$, and suppose we compute the *LRT* by maximizing $L(\theta \mid X)$ *wrt* $\theta \in \Omega$ and also under H_0. Write $\Omega_0 = \{\theta : \theta_1 = \theta_{10} + \delta, \theta_2, \cdots, \theta_p \ unspecified\}$. Then the null distribution of the profile log likelihood based LRT using X, namely, the null distribution of

$$-2ln\left[\frac{max_{\theta \in \Omega_0} L(\theta \mid X)}{max_{\theta \in \Omega} L(\theta \mid X)}\right] \tag{8}$$

is asymptotically equivalent to the distribution of the profile log likelihood based LRT using Z under $N[\theta, \Sigma]$, namely, the distribution of

$$min_{\theta \in \Omega_0^*} Q(\theta \mid Z) - min_{\theta \in \Omega^*} Q(\theta \mid Z) \tag{9}$$

when $\theta = 0$. Here $Q(\theta \mid Z) = (Z - \theta)'[\Sigma]^{-1}(Z - \theta)$ is the exponent of the normal likelihood of Z with mean θ and dispersion Σ, minimum under $\theta \in \Omega_0^*$ is computed when $\theta_1 = \delta$ and $\theta_2, \cdots, \theta_p$ are unspecified in $(-\infty, \infty)$, and minimum under $\theta \in \Omega^*$ is computed when all the parameters θ are unspecified in $(-\infty, \infty)$. In other words, quite

generally, Ω^* is a translation of Ω by Ω_0 so that Ω^* includes the point 0 in the parameter space of Z.

(b) Consider the problem of testing the null hypothesis $H_0 : \theta_1 = \theta_{10}$ versus $H_1 : \theta_1 > \theta_{10}$, and suppose we compute the *LRT* by maximizing $L(\theta \mid X)$ *wrt* $\theta \in \Omega$ and also under H_0. Write $\Omega_0 = \{\theta : \theta_{10}, \theta_2, \cdots, \theta_p \ unspecified\}$. Then the null distribution of the profile log likelihood based LRT using X, namely, the null distribution of

$$-2ln\left[\frac{max_{\theta \in \Omega_0} L(\theta \mid X)}{max_{\theta \in \Omega} L(\theta \mid X)}\right] \tag{10}$$

is asymptotically equivalent to the distribution of the profile log likelihood based LRT using Z under $N[\theta, \Sigma]$, namely, the distribution of

$$min_{\theta \in \Omega_0^*} Q(\theta \mid Z) - min_{\theta \in \Omega^*} Q(\theta \mid Z) \tag{11}$$

when $\theta = 0$. Here, as before, $Q(\theta \mid Z) = (Z - \theta)'[\Sigma]^{-1}(Z - \theta)$ is the exponent of the normal likelihood of Z with mean θ and dispersion Σ, minimum under $\theta \in \Omega_0^*$ is computed when $\theta_1 = 0$ and $\theta_2, \cdots, \theta_p$ are unspecified in $(-\infty, \infty)$, and minimum under $\theta \in \Omega^*$ is computed when $\theta_1 \geq 0$ and all the other parameters θ are unspecified in $(-\infty, \infty)$. In other words, as before, Ω^* is a translation of Ω by Ω_0 so that Ω^* includes the point 0 in the parameter space of Z.

Remark 2.1. It should be noted that the null hypotheses mentioned above are concerned only with θ_1 and do not mention anything about the nuisance parameters $\theta_2, \cdots, \theta_p$. It is quite possible that some of the nuisance parameters may lie in a parameter space containing the boundary points. In other words, it is possible that $\theta_i \in \Omega_i = [\theta_{i0}, \infty)$ for $i > 1$. When this happens, it is implied that the minimization wrt θ in the quadratic form $Q(\theta \mid Z)$ is done under the restriction that $\theta_i \geq 0$, implying that 0 is a boundary point wrt θ_i rather than being an interior point.

Moran (1971), Chant (1974) and, most importantly, Self and Liang (1987) discussed at length computation of the LRT based on Z and its null distribution under various forms of the null hypothesis well beyond the two cases mentioned above. In general, as remarked earlier, it follows that the null distribution of the LRT based on Z is central chi-square when Ω_0^* consists of interior points of the mean vector θ in the normal distribution of Z. However, this distribution is quite often a mixture of chi-squares when Ω_0^* contains some boundary points of θ as in Prob. 2.2 above.

A second major objective of this paper is to discuss in detail the null distribution of the LRT based on Z for testing the two hypotheses about θ_1 mentioned above under Prob. 2.1 and Prob. 2.2 when it is likely that there are none or some parameters lying on the boundaries, thus supplementing the earlier works of Moran (1971), Chant (1974) and Self and Liang (1987).

Referring to Prop. 2.1, we should note that once the MLEs of θ based on Z under Ω^* and their joint distribution are derived, we can use Prop. 2.1 to approximate the joint as well as the marginal distributions of the actual MLEs of θ based on X. These distributions can then be effectively used to draw suitable Wald-type inference about a smooth function of θ such as tests for a single component or a linear function of the components of θ, thus providing an alternative to the profile likelihood approach.

Likewise, referring to Prop. 2.2, once the LRT of H_0 versus H_1 based on Z is derived and its null distribution is obtained, we can use it to get the approximate cut-off points of the LRT based on X. We remark that this reduction of the original inference problem based on X with an arbitrary distribution to a canonical form using Z which has a normal distribution, though only asymptotically valid, is a key feature of the asymptotic theory and the spirit of all the earlier works of these authors.

We are now in a position to describe the main results of this section. Based on the distributional assumption $Z \sim N_p[\theta, \Sigma]$, where θ is unknown and Σ is positive definite known, we develop some *new* results for *exact* inference on θ_1 in presence of a few nuisance parameters, allowing the possibility that some of the nuisance parameters may lie on the boundaries.

Towards deriving the LRT of $H_0 : \theta_1 = \delta$ versus $H_1 : \theta_1 \neq \delta$ for some $\delta > 0$ based on Z, which corresponds to testing $H_0 : \theta_1 = \theta_{10} + \delta$ versus $H_1 : \theta_1 \neq \theta_{10} + \delta$ for some $\delta > 0$ in the original distribution of X, we note that the likelihood function of Z, namely, $L(\theta \mid Z)$ can be written as

$$L(\theta \mid Z) = K \exp[-(Z - \theta)' \Sigma^{-1} (Z - \theta)]. \tag{12}$$

Since we assume that the p parameters are functionally independent, for testing $H_0 : \theta_1 = \delta$ versus $H_1 : \theta_1 \neq \delta$ based on Z, it is well known that if the parameter vector θ is an interior point in R^p, which means the true parameter values of the parameter θ_1 of interest as well as the nuisance parameters $\theta_2, \cdots, \theta_p$ are interior points, then the usual normal test based on Z_1 is the LRT and it provides a a valid test. Recall that the dispersion matrix of Z is assumed known which results in a normal test rather than a t-test.

As mentioned earlier, Self and Liang (1987), based on previous works of Chernoff (1954), Chant (1974), Moran (1971) and Shapiro (1985), developed appropriate solutions to this kind of problem for a wide variety of scenarios involving several parameters under H_0, and allowing some of them and also some nuisance parameters to be on the boundary. In the huge literature on mathematical statistics, Self and Liang (1987) paper is indeed a landmark paper with a novel contribution to this important problem particularly because it is quite common that some parameters of interest as well as some nuisance

parameters in a typical multi-normal set-up can indeed lie on the boundary and also because of the fact that such situations often arise in applications.

Following essentially Self and Liang's ideas, we derive below the likelihood ratio tests (LRTs) of H_0 versus H_1 based on Z, allowing one, two and three nuisance parameters to be on the boundary. It turns out that, as one can expect, the form of the LRT becomes quite complex with the increase in the number of nuisance parameters which lie on the boundary. In the sequel, we also derive the maximum likelihood estimates of all parameters θ based on Z under the condition that some of the parameters may lie on the boundary, and derive their joint and marginal distributions. As mentioned before, these results would be useful when one is interested in deriving Wald-type asymptotic inference based on X about a smooth function of θ, and in particular, about a linear function of these parameters which is the case for dose-response multistage Weibull models mentioned in Section 1.

Some standard results from classical multivariate analysis which are needed in the sequel are listed below. Write $Z = (Z_{(1)}, Z_{(2)})$, $\theta = (\theta_{(1)}, \theta_{(2)})$ where $Z_{(1)} : q \times 1$, $Z_{(2)} : (p-q) \times 1$, $\theta_{(1)} : q \times 1$, $\theta_{(2)} : (p-q) \times 1$, and $\Sigma_{(11)} = var(Z_{(1)})$, $\Sigma_{(22)} = var(Z_{(2)})$, and $\Sigma_{(12)} = cov(Z_{(1)}, Z_{(2)})$. Then:

distribution of a quadratic form
$$[Z - \theta]'(\Sigma)^{-1}[Z - \theta] \sim \chi_p^2 \tag{13}$$
marginal distribution
$$Z_1 \sim N_q[\theta_{(1)}, \Sigma_{(11)}] \tag{14}$$
conditional distribution of Z_1, given $Z_2 = z_2$
$$Z_1 \mid z_2 \sim N_q[\theta_1 + \Sigma_{(12)}(\Sigma_{(22)})^{-1}(z_2 - \theta_2), \Sigma_{(11)} - \Sigma_{(12)}(\Sigma_{(22)})^{-1}\Sigma_{(21)}] \tag{15}$$

2.1 One parameter on the boundary

In this subsection we assume that there is one parameter, say θ_i, which may lie on the boundary in the sense that $\theta_i \in \Omega_i = [\theta_{i0}, \infty)$, and carry out appropriate inference about θ. Before we develop inferential tools, let us make a remark about testing for the existence of a boundary parameter.

Consider a general multi-parameter model based on X involving p parameters $\theta = (\theta_1, \cdots, \theta_p) \in \Omega$ where $\Omega = \Omega_1 \times \cdots \Omega_p$ with $\Omega_i = [\theta_{i0}, \infty)$ or (θ_{i0}, ∞) for all i. Suppose it is suspected that one parameter lies on the boundary! How do we determine which one? Here is an ad hoc approach. Assuming that $\theta_i = \theta_{i0}$ is the point on the boundary, we can maximize the likelihood $L(\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_p \mid \theta_i = \theta_{i0}, X)$ wrt $\theta_j \in \Omega_j = (\theta_{j0}, \infty)$ for all $j \neq i$ and compute the maximum value of the likelihood, say L_i. Comparing $L_1, \cdots, L_p$ and selecting the index k such that $L_k = max(L_1, \cdots, L_p)$, we can conclude that θ_k lies on the boundary. Alternatively, we can test the hypothesis $H_{0i} : \theta_i = \theta_{i0}$ versus $H_{1i} : \theta_i > \theta_{i0}$ by computing the *LRT* statistic, say λ_i, and choose the

index k for which λ_k is the smallest with the conclusion that θ_k is likely to lie on the boundary. Details about such a test are given below.

Let us consider the two testing problems about θ_1 mentioned earlier, under the assumption that there is one boundary parameter. We will derive exact tests based on Z which would yield asymptotic tests based on X. Consider testing $H_0 : \theta_1 = \delta$ versus $H_1 : \theta_1 \neq \delta$ based on $Z \sim N[\theta, \Sigma]$. We distinguish between two cases depending on whether the parameter on the boundary is itself the parameter of interest (θ_1) or a nuisance parameter, say θ_2. In the former case, we take $\delta = 0$ as the boundary point of $\Omega_1^* = [0, \infty)$ and write $H_1 : \theta_1 > 0$ while the other parameters are free. In the latter case, we take $\delta > 0$ as an interior point of $\Omega_1^* = (0, \infty)$, and assume that one of the remaining unspecified nuisance parameters, say θ_2, may lie on the boundary in the sense that $\theta_2 \in [0, \infty)$. Note that, by definition, a nuisance parameter can never be known so that we cannot conclude that $\theta_2 = 0$!

Case 2.1.1 *θ_1 is a boundary point. This has been discussed at length in Self and Liang**Error! Reference source not found.** who derived the LRT of this problem. It turns out that the LRT is based on $Z_1^2 I[Z_1 > 0] / \sigma_{11}$ and its null distribution is a $50:50$ mixture of χ_0^2 and χ_1^2 distributions. Details are omitted.*

Case 2.1.2 *θ_1 is an interior point and θ_2 is a boundary point. To derive the LRT for $H_0 : \theta_1 = \delta + \theta_{10}$ versus $H_1 : \theta_1 \neq \delta + \theta_{10}$ for some $\delta > 0$ on the basis of X, note that, asymptotically, this is equivalent to testing $H_0 : \theta_1 = \delta$ versus $H_1 : \theta_1 \neq \delta$ for $\delta > 0$ based on Z. We proceed to apply Prop. 2.2. It is clear from the expression of the likelihood function $L(\theta \mid Z)$ given in (12) that what matters is the quadratic form $Q(\theta \mid Z)$ given by*

$$Q(\theta \mid Z) = (Z - \theta)' \Sigma^{-1} (Z - \theta). \tag{16}$$

Writing $Q(\theta \mid Z) = Q(\theta_1, \theta_2 \mid Z_1, Z_2) + Q(Z_3, \cdots, Z_p \mid Z_1, Z_2; \theta)$ where the first part is the marginal bivariate quadratic of (Z_1, Z_2) and the second part is the $(p-2)$-dimensional conditional quadratic of $(Z_3, \cdots, Z_p)$, given (Z_1, Z_2), it follows from Self and Liang (1987) that due to the interior nature of the parameters $\theta_3, \cdots, \theta_p$, the only part we need to study is the first part, and maximization of the likelihood corresponds to finding the two minimums of the first part, one under the union of null and alternative hypotheses, and the other under the null hypothesis.

To derive the maximum likelihood estimates (MLEs) of the parameters θ under the union of null and alternative parameter spaces, we can express $Q(\theta_1, \theta_2 \mid Z_1, Z_2)$ as

$$Q(\theta_1, \theta_2 \mid Z_1, Z_2) = Q(Z_2 \mid \theta_2) + Q(Z_1 \mid Z_2; \theta_1, \theta_2) \tag{17}$$

where $Q(Z_2 \mid \theta_2)$ is the marginal *pdf* of Z_2 and $Q(Z_1 \mid Z_2; \theta_1, \theta_2)$ is the conditional *pdf* of Z_1, given Z_2. It is clear from the discussion in Self and Liang (1987) that due to the

interior nature of θ_1, the minimum value of $Q(Z_1 | Z_2; \theta_1, \theta_2)$ wrt θ_1 is 0, irrespective of the value of θ_2. Now a minimization of $Q(Z_2 | \theta_2)$ wrt θ_2 subject to $\theta_2 \geq 0$ readily yields:

$$
\begin{aligned}
min_{\theta_2 \geq 0} Q(Z_2 | \theta_2) \quad &= \quad 0 \quad if \quad Z_2 > 0 \\
&= \quad Z_2^2 / \sigma_{22} \quad if \quad Z_2 \leq 0.
\end{aligned}
\tag{18}
$$

To minimize $Q(\theta_1, \theta_2 | Z_1, Z_2)$ wrt θ_2 under the null hypothesis when $\theta_1 = \delta$, we write

$$
Q(\theta_1, \theta_2 | Z_1, Z_2) = Q(Z_1 | \theta_1 = \delta) + Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2)
\tag{19}
$$

where $Q(Z_1 | \theta_1 = \delta)$ is the *pdf* of Z_1 under the null hypothesis and $Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2)$ is the *pdf* of Z_2, given Z_1 when $\theta_1 = \delta$. Since the first term is independent of θ_2, it is clear that the minimum value of $Q(\theta_1, \theta_2 | Z_1, Z_2)$ wrt θ_2 arises essentially from minimizing $Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2)$ wrt θ_2. Since

$$
Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2) = \frac{[Z_2 - \theta_2 - \rho \frac{\sigma_2}{\sigma_1}(Z_1 - \delta)]^2}{\sigma_{22}(1 - \rho^2)}
\tag{20}
$$

minimization of $Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2)$ wrt θ_2 under the condition $\theta_2 \geq 0$ readily gives

$$
\begin{aligned}
min_{\theta_2 \geq 0}[Q(Z_2 | Z_1; \theta_1 = \delta, \theta_2)] \quad &= \quad 0 \quad if \quad Z_{2.1} > 0 \\
&= \quad \frac{Z_{2.1}^2}{\sigma_{22}(1 - \rho^2)} \quad if \quad Z_{2.1} \leq 0
\end{aligned}
\tag{21}
$$

where $Z_{2.1} = Z_2 - \rho \frac{\sigma_2}{\sigma_1}(Z_1 - \delta)$. Hence we get

$$
\begin{aligned}
min_{\theta_1 = \delta; \theta_2 \geq 0} Q(\theta_1, \theta_2 | Z_1, Z_2) \quad &= \quad \frac{(Z_1 - \delta)^2}{\sigma_{11}}, \quad if \quad Z_{2.1} > 0 \\
&= \quad \frac{(Z_1 - \delta)^2}{\sigma_{11}} + \frac{Z_{2.1}^2}{\sigma_{22}(1 - \rho^2)}, if \quad Z_{2.1} \leq 0.
\end{aligned}
\tag{22}
$$

Combining (18) and (22), and taking the difference, we get $-2ln(LRT) = W(say)$ as

$$
\begin{aligned}
W \quad &= \quad (Z_1 - \delta)^2 / \sigma_{11}, \ if \quad Z_2 > 0, Z_{2.1} \geq 0 \\
&= \quad (Z_1 - \delta)^2 / \sigma_{11} - Z_2^2 / \sigma_{22}, \ if \quad Z_2 \leq 0, Z_{2.1} > 0 \\
&= \quad Z_{1.2}^2 / \sigma_{11}(1 - \rho^2), \ if \quad Z_2 \leq 0, Z_{2.1} \leq 0 \qquad (23) \\
&= \quad \frac{(Z_1 - \delta)^2}{\sigma_{11}} + \frac{Z_{2.1}^2}{\sigma_{22}(1 - \rho^2)}, \ if \quad Z_2 > 0, Z_{2.1} < 0.
\end{aligned}
$$

It is easy to verify that when $\rho = 0$, W reduces to $(Z_1 - \delta)^2 / \sigma_{11}$ which has a chi-square distribution with 1 d.f. under H_0, a familiar result.

To derive the null distribution of W for any given ρ, we assume without loss of generality that $\sigma_1 = \sigma_2 = 1$ and $\delta = 0$. It is proved in the Appendix that the cdf of W is given by the following.

Theorem 2.1 The cdf $G(w)$ of W, for $0 < w < \infty$, is given by the sum of four parts:

(i) First part $= \int_0^{r\sqrt{w}} \left[\int_{-\sqrt{\{(\omega - v^2)(1-\rho^2)\}}}^{v\rho + \sqrt{w}} N(0,1)dx \right] N(0,1)dv$ (24)

(ii) Second part $= \int_{\rho\sqrt{w}}^{\infty} \left[\int_{v\rho - \sqrt{w}}^{v\rho + \sqrt{w}} N(0,1)dx \right] N(0,1)dv$ (25)

(iii) Third part $= \int_{-\rho\sqrt{w/(1-\rho^2)}}^{0} \left[\int_{-\sqrt{\omega(1-\rho^2)}}^{v\rho + \sqrt{\omega + v^2}} N(0,1)dx \right] N(0,1)dv$ (26)

(iv) Fourth part $= \left[\int_{-\sqrt{\omega(1-\rho^2)}}^{\sqrt{\omega(1-\rho^2)}} N(0,1)dx \right] \cdot \left[\int_{-\infty}^{-\rho\sqrt{\omega/(1-\rho^2)}} N(0,1)dx \right]$ (27)

The above distribution can be used to get a cut-off point of the statistic $-2ln(LRT)$ which can then be used to carry out the LRT based on X. *Quite surprisingly*, our simulations indicate that the above distribution does *not* depend on ρ!

We now discuss the application of Wald-type test for testing H_0 versus H_1. Towards this end, we note from Self and Liang (1987) that the MLEs of the parameters θ_i when θ_2 may be on the boundary are given by the following. Define $Z_{i.2} = Z_i - \frac{\sigma_i \rho_{i2}}{\sigma_2} Z_2$.

Theorem 2.2 The MLE of θ_i, $i = 1, 3, \cdots, p$ when θ_2 is on the boundary is given by

$$\hat{\theta}_i = Z_i I[Z_2 > 0] + Z_{i.2} I[Z_2 \leq 0].$$ (28)

The marginal distribution of the MLE of θ_i, which depends on the correlation ρ_{i2} between Z_i and Z_2, is given below. This would be useful if one is interested in drawing suitable inference about just one parameter θ_i.

Theorem 2.3 The pdf of the MLE $\hat{\theta}_i = U_i$ of θ_i is given by

$$f(u_i) = \int_0^\infty \frac{e^{-\left[\frac{x^2 + \frac{u_i^2}{\sigma_{ii}} - \frac{2x_i \rho_{i2}}{\sigma_i}}{2(1-\rho_{i2}^2)} \right]}}{2\pi \sqrt{\sigma_{ii}\left(1-\rho_{i2}^2\right)}} dx + \frac{e^{-\frac{u_i^2}{2\sigma_{ii}\left(1-\rho_{i2}^2\right)}}}{2\sqrt{2\pi\sigma_{ii}\left(1-\rho_{i2}^2\right)}}.$$ (29)

Proof. The pdf of U_i is derived from the cdf of U_i which is readily obtained as follows. Since Z_2 and $Z_{i.2}$ are independent and $P(Z_2 \leq 0) = 0.5$, the second part is obvious. For

the first part, we first write down the bivariate normal pdf of Z_2 and Z_i, and then integrate out Z_2 over $(0, \infty)$. The pdf of U_i for the first part is then directly obtained from the cdf of the first part upon differentiation wrt u_i. This completes the proof.

Remark 2.2 Specializing to the case of testing $H_0 : \theta_1 = \delta + \theta_{10}$ versus $H_1 : \theta_1 > \delta + \theta_{10}$ based on X, we can easily derive a Wald-type test based on the MLE of θ_1, namely, $\hat{\theta}_{1N}$ which is the standard MLE of θ based on X under $\theta \in \Omega$ with $\Omega_2 = [\theta_{20}, \infty)$ and $\Omega_i = (\theta_{i0}, \infty)$ for $i \neq 2$. Because of the nature of H_0 and H_1, it makes sense to reject H_0 for large values of $\hat{\theta}_{1N}$, i.e., when $\hat{\theta}_{1N} > c_N$ for some c_N. To determine the value of c_N for a given level of significance α, it follows that, asymptotically, $\alpha = P[\hat{\theta}_{1N} > c_N \mid H_0] = P[\sqrt{N}(\hat{\theta}_{1N} - \delta - \theta_{10}) > \sqrt{N}(c_N - \delta - \theta_{10})] \sim P[Z_1 I(Z_2 > 0) + Z_{1.2} I(Z_2 \leq 0) > \sqrt{N}(c_N - \delta)]$. We can now use the result of Theorem 2.3 to claim: $\sqrt{N}(c_N - \delta) = z_{\alpha|\rho}$ where $z_{\alpha|\rho}$ is the upper α cut-off point of the distribution of $Z_1 I[Z_2 > 0] + Z_{1.2} I[Z_2 \leq 0]$ given in Theorem 2.3 for $i = 1$. Hence we reject H_0 when $\hat{\theta}_{1N} > \delta + \theta_{10} + z_{\alpha|\rho} / \sqrt{N}$. This is of course a very easy test to carry out without the need to compute a profile likelihood which is the basis of LRT. However, we should also note that we have taken a one-sided alternative as H_1. For a both-sided alternative, we may choose to reject H_0 for large values of $|\hat{\theta}_{1N} - \delta - \theta_{10}|$ and determine the cut-off point appropriately. Note that this would provide an alternative approach to the profile likelihood method (LRT).

Remark 2.3 If one is interested in making inference about a linear combination of θ_i's which *excludes* the boundary point θ_2, we note that the distribution of $U = \sum_{i \neq 2} c_i U_i = [\sum_{i \neq 2} c_i Z_i] I[Z_2 > 0] + [\sum_{i \neq 2} c_i Z_{i.2}] I[Z_2 \leq 0]$ is again readily obtained as the sum of two parts of which the second part (when $Z_2 \leq 0$) is normal with mean 0 and variance obtained from the variances and covariances of the residuals $Z_{i.2}$'s, and the other part is a convolution of two normals. The latter term is obtained by writing down the bivariate normal pdf of $(\sum_{i \neq 2} c_i Z_i)$ and Z_2, and then integrating out Z_2 over $(0, \infty)$. Thus, defining $c^* = (c_1, 0, c_3, .., c_p)$, $\sigma^{2*} = c^* \Sigma c^{*\prime}$, $\rho^* = \sum_{i \neq 2} c_i \rho_{i2} \sigma_i$ and $\Delta^2 = \sum_{i \neq 2} c_i^2 \sigma_{ii}(1 - \rho_{i2}^2) + \sum_{i \neq j \neq 2} c_i c_j \sigma_i \sigma_j (\rho_{ij} - \rho_{i2} \rho_{j2})$, the *pdf* of U is given by $f(u) = f_I(u) + f_{II}(u)$ where

$$f_I(u) \quad = \quad \frac{e^{-\frac{u^2}{2\Delta^2}}}{2\Delta\sqrt{2\pi}}$$

$$f_{II}(u) \;=\; \int_0^\infty e^{\left[-\dfrac{x^2+\dfrac{u^2}{\sigma^{2*}}-2\dfrac{xu\rho^*}{\sigma^*}}{2(1-\rho^{2*})}\right]} dx \, / \, [2\pi\sigma^*\sqrt{1-\rho^{2*}}] \qquad (30)$$

Let us recall that the distribution of U given here is precisely the asymptotic distribution of $\sqrt{N}[\sum_{i\neq 2}c_i\hat{\theta}_{iN} - \sum_{i\neq 2}c_i\theta_i]$ and hence can be readily used for drawing valid asymptotic inference about $\sum_{i\neq 2}c_i\theta_i$. Note that this so-called Wald-type approach avoids computing profile likelihoods for testing hypotheses about $[\sum_{i\neq 2}c_i\theta_i]$.

Remark 2.4 If, on the other hand, we are interested in making inference about a linear function of the θ_i's which also includes θ_2, naturally we need to derive the distribution of $V = \sum_{i\neq 2}c_i\hat{\theta}_i + c_2\hat{\theta}_2$. Since $\hat{\theta}_2 = Z_2 I[Z_2 > 0]$, it follows from (28) that the distribution of V will again consist of two parts: one part, corresponding to $Z_2 < 0$, is just normal and is independent of Z_2. This is in fact the distribution of $\sum_{i\neq 2}c_i Z_{i.2}$. The other part, for $Z_2 > 0$, is obtained by first deriving the conditional distribution of $\sum_{i\neq 2}c_i Z_i$ which is normal and then convoluting it with $c_2 Z_2$. This argument leads to the *pdf* of V as $f(v) = f_I(v) + f_{II}^*(v)$ where $f_I(v)$ is the same as $f_I(u)$, and $f_{II}^*(v)$ is given by

$$f_{II}^*(v) = \int_0^\infty \frac{e^{\left(-\dfrac{\left[(v-x)\left(\sum_{i=1}^p c_i \rho_{i2}\sigma_i\right)\right]^2}{2\Delta^2} - \dfrac{x^2}{2}\right)}}{\Delta 2\pi} dx. \qquad (31)$$

We recall that the distribution of V given here is precisely the asymptotic distribution of $\sqrt{N}[\sum_{i=1}^p c_i\hat{\theta}_{iN} - \sum_{i=1}^p c_i\theta_i]$ and hence can be readily used for drawing valid asymptotic inference about $\sum_{i=1}^p c_i\theta_i$. Again, this approach avoids the computation of a profile likelihood.

2.2 Two parameters on the boundaries

We now discuss the case of two boundary parameter points. Kopylev and Sinha (2011) discussed the case when the parameter of interest and a nuisance parameter lie on the boundary. Here we consider the case when two nuisance paramaters lie on the boundary and the parameter of interest is an interior point.

Assume without any loss of generality that θ_2 and θ_3 lie on the boundary and our primary interest lies in the parameter θ_1 which is an interior point. In the sequel, we

consider both LRT and Wald-type tests for $H_0 : \theta_1 = \theta_{10} + \delta$ versus $H_1 : \theta_1 \neq \theta_{10} + \delta$ for some $\delta > 0$ based on X.

As before, write $Q(Z \mid \theta) = Q(Z_1, Z_2, Z_3 \mid \theta_1, \theta_2, \theta_3) + Q(Z_4, \cdots, Z_p \mid Z_1, Z_2, Z_3; \theta)$ and recall that in the θ-space for Z, $\theta_2 \in [0, \infty)$, $\theta_3 \in [0, \infty)$ and $\theta_i \in (-\infty, \infty)$ for $i \neq 2, 3$. Hence, for LRT as well as for Wald, we need to concentrate only on the first part. For the unrestricted MLEs of θ_2 and θ_3, we get from Self and Liang (1987):

$$
\begin{aligned}
(\hat{\theta}_2, \hat{\theta}_3)' &= (Z_2, Z_3)', \, if \quad Z_2 > 0, Z_3 \geq 0 \\
&= (Z_{2.3}, 0)', \, if \quad Z_{2.3} \leq 0, Z_3 \leq 0 \\
&= (0, Z_{3.2})', \, if \quad Z_2 \leq 0, Z_{3.2} > 0 \\
&= (0, 0)', \, if \quad Z_{2.3} < 0, Z_{3.2} < 0.
\end{aligned}
\tag{32}
$$

Since the minimum of $Q(Z_1 \mid Z_2, Z_3; \theta_1, \theta_2, \theta_3)$ wrt θ_1 for any given (θ_2, θ_3) is 0 and also the minimum of $Q(Z_4, \cdots, Z_p \mid Z_1, Z_2, Z_3; \theta)$ wrt $\theta_4, \cdots, \theta_p$ for any given θ_1, θ_2 and θ_3 is 0, we get

$$
min_{\theta \in \Omega^*} Q(Z \mid \theta) = Q(Z_2, Z_3 \mid \hat{\theta}_2, \hat{\theta}_3)
\tag{33}
$$

It is easy to show that the above minimum simplifies to

$$
\begin{aligned}
min_{\theta \in \Omega^*} Q(Z \mid \theta) &= 0, \, if \quad Z_2 > 0, Z_3 \geq 0 \\
&= Z_3^2 / \sigma_{33}, \, if \quad Z_3 < 0, Z_{2.3} > 0 \\
&= Z_2^2 / \sigma_{22}, \, if \quad Z_2 \leq 0, Z_{3.2} > 0 \\
&= Q(Z_2, Z_3 \mid \theta_2 = \theta_3 = 0), \, if \quad Z_{2.3} \leq 0, Z_{3.2} \leq 0
\end{aligned}
\tag{34}
$$

Once $\hat{\theta}_2$ and $\hat{\theta}_3$ are derived, the MLEs of the rest of the (interior) parameters are readily obtained from $Q(Z_1, Z_4, \cdots, Z_p \mid Z_2, Z_3; \theta)$ as the *residuals* of Z_i, given (Z_2, Z_3), and are given by

$$
\hat{\theta}_i = Z_i - E[Z_i \mid Z_2 - \hat{\theta}_2, Z_3 - \hat{\theta}_3].
\tag{35}
$$

To derive the LRT of $H_0 : \theta_1 = \delta$ versus $H_1 : \theta_1 \neq \delta$ based on Z, we need to derive the restricted MLEs of θ's under the null hypothesis H_0. Without any loss of generality, let us assume $\delta = 0$, and write $Q(Z_1, Z_2, Z_3 \mid \theta_1 = 0, \theta_2, \theta_3) = Q(Z_1 \mid \theta = 0) + Q(Z_2, Z_3 \mid Z_1; \theta_1 = 0, \theta_2, \theta_3)$. Since the first part is independent of θ_2 and θ_3, writing $Y_2 = Z_{2.1}$ and $Y_3 = Z_{3.1}$, we get from Self and Liang (1987) the following restricted MLEs of θ_2 and θ_3:

$$
\begin{aligned}
(\hat{\theta}_2, \hat{\theta}_3)_{null}{}' \quad = \quad & (Y_2, Y_3)', \quad if \quad Y_2 > 0, Y_3 \geq 0 \\
= \quad & (Y_{2.3}, 0)', \quad if \quad Y_{2.3} \leq 0, Y_3 \leq 0 \\
= \quad & (0, Y_{3.2})', \quad if \quad Y_2 \leq 0, Y_{3.2} > 0 \\
= \quad & (0, 0)', \quad if \quad Y_{2.3} < 0, Y_{3.2} < 0.
\end{aligned}
\tag{36}
$$

Since as before the contribution from the minimization of $Q(Z_4 \cdots, Z_p \mid Z_1, Z_2, Z_3; \theta)$ with respect to $\theta_4, \cdots, \theta_p$ is 0 even under H_0 due to the interior nature of the parameters $\theta_4, \cdots, \theta_p$, it follows that

$$
\begin{aligned}
min_{\theta \in \Omega_0^*} Q(Z \mid \theta; H_0) \quad = \quad & Q(Z_1 \mid \theta_1 = 0) \\
& + min_{\theta_2, \theta_3 \geq 0} Q(Z_2, Z_3 \mid Z_1, \theta_1 = 0, \theta_2, \theta_3) \\
= \quad & Q(Z_1 \mid \theta_1 = 0) + Q(Z_2, Z_3 \mid Z_1, \hat{\theta}_2 \, null, \hat{\theta}_3 \, null)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & Z_1^2 / \sigma_{11}, if \, Z_{2.1} > 0, Z_{3.1} \geq 0 \tag{37} \\
= \quad & Z_3^2 / \sigma_{33} + \frac{Z_{1.3}^2}{\sigma_{11}(1 - \rho_{13}^2)}, if \, Z_{3.1} < 0, Z_{2.13} > 0 \\
= \quad & Z_2^2 / \sigma_{22} + \frac{Z_{1.2}^2}{\sigma_{11}(1 - \rho_{12}^2)}, if \, Z_{2.1} \leq 0, Z_{3.12} > 0 \\
= \quad & Z_1^2 / \sigma_{11} + (Z_{2.1} / \sigma_2, Z_{3.1} / \sigma_3) A^{-1} (Z_{2.1} / \sigma_2, Z_{3.1} / \sigma_3)', \\
& \quad if \, Z_{2.13} \leq 0, Z_{3.12} \leq 0
\end{aligned}
$$

where the matrix $A : 2 \times 2$ is defined as

$$
A = \begin{pmatrix} (1 - \rho_{12}^2) & \rho_{23} - \rho_{12}\rho_{13} \\ \rho_{23} - \rho_{12}\rho_{13} & 1 - \rho_{13}^2 \end{pmatrix}.
\tag{38}
$$

Combining (34) and (37), the LRT of $H_0 : \theta_1 = 0$ is obtained as W given by

$$
W = min_{\theta \in \Omega_0^*} Q(Z \mid \theta; H_0) - min_{\theta \in \Omega^*} Q(Z \mid \theta).
\tag{39}
$$

To simulate the null distribution of W, we can take without any loss of generality $(Z_1, Z_2, Z_3) \sim N[(0,0,0), \Sigma^*]$ where Σ^* has its diagonal elements as 1. Upon generating the j th iteration element as (Z_{1j}, Z_{2j}, Z_{3j}), we compute $(Z_{2.3j}, Z_{3.2j}, Z_{3.1j}, Z_{2.1j}, Z_{2.13j}, Z_{3.12j})$, leading to a value w_j of W.

Obviously, the LRT of H_0 versus H_1 rejects H_0 when W computed as above is large. Again, *quite surprisingly*, it turns out from our simulation studies that the null distribution of W does not depend on the correlations between Z_1 and (Z_2, Z_3).

For Wald-type inference about θ_1, we concentrate on the distribution of the MLE of θ_1 given in (35) under the natural restriction of two boundary points. Such a distribution can

be obtained from (34) by conditioning on four disjoint subsets in (z_2, z_3)-plane and deriving each component distribution and mixing them with suitable proportions. Details are omitted as this is similar to Theorem 2.1. This distribution can be used for testing $H_0 : \theta_1 = \delta + \theta_{10}$ based on X. Thus, if the alternative is $H_1 : \theta_1 > \delta + \theta_{10}$, a reasonable test is to reject H_0 for large values of the MLE $\hat{\theta}_{1N}$ of θ_1, namely, when $(\hat{\theta}_{1N} - \delta - \theta_{10}) > c_N$. To determine the value of c_N for a given significance level α, we can use the asymptotic null distribution of $\sqrt{N}(\hat{\theta}_{1N} - \delta - \theta_{10})$ in presence of two boundary parameters, which is precisely the distribution mentioned above.

On the other hand, for a both-sided alternative $H_1 : \theta_1 \neq \delta + \theta_{10}$, a reasonable test is to reject H_0 for large values of $|\hat{\theta}_{1N} - \delta - \theta_{10}|$. To determine the cut-off point for a given significance level α, we can again use the asymptotic null distribution of $\sqrt{N}(\hat{\theta}_{1N} - \delta - \theta_{10})$ in presence of two boundary parameters, which is precisely the distribution mentioned above. Details are omitted.

Remark 2.5 Suppose we are interested in making a Wald-type inference about a linear function of θ, say $\sum_{i \neq 2,3} c_i \theta_i$ which excludes the two boundary points θ_2 and θ_3. Naturally we would then consider the distribution of $\sum_{i \neq 2,3} c_i \hat{\theta}_{iN}$. From (32), such a distribution can be obtained by conditioning on values of (z_2, z_3) and then suitable unconditioning.

Remark 2.6 Suppose now that we are interested in making a Wald-type inference about a linear function of θ, say $\sum_{i \neq 2,3} c_i \theta_i + c_2 \theta_2$ which contains one of the boundary points, say θ_2. We would then consider the obvious statistic $\sum_{i \neq 2,3} c_i \hat{\theta}_{iN} + c_2 \hat{\theta}_{2N}$. Its asymptotic distribution can be derived from the results given above.

Remark 2.7 Let us also assume that we are interested in drawing suitable Wald-type inference about a linear function of θ, which includes both the boundary parameters. Obviously, such a linear function can be written as $\eta = \sum_i c_i \theta_i$ and inference about η is drawn by studying the asymptotic distribution of $W = \sum_i c_i \hat{\theta}_{iN}$. This distribution can also be derived from the above results.

2.3 Three points on the boundaries

In this section we deal with the case when three parameters lie on their boundaries. Kopylev and Sinha(2011) discussed the case when the parameter of interest and two nuisance parameters lie on the boundary. We consider here the case when all three nuisance parameters lie on the boundary and the parameter of interest is an interior point. Without any loss of generality, suppose θ_1, θ_2 and θ_3 are the three boundary parameters with $\Omega_i = [\theta_{i0}, \infty)$ for $i = 1, 2, 3$ and $\Omega_j = (\theta_{j0}, \infty)$ for $j > 3$, and θ_4 is the parameter of

interest. The relevant testing problem about θ_4 is then to decide between $H_0 : \theta_4 = \theta_{40} + \delta$ versus $H_1 : \theta_4 \neq \theta_{40} + \delta$ on the basis of X where $\delta > 0$ is specified. Equivalently (and asymptotically), this is the same as testing $H_0 : \theta_4 = \delta$ versus $H_1 : \theta_4 \neq \delta$ based on $Z \sim N[\theta, \Sigma]$, when $\theta_i \geq 0, \quad i = 1, 2, 3$ and $-\infty < \theta_j < \infty$ for $j > 3$. To derive the LRT for the above testing problem, we proceed in the usual fashion by deriving both unrestricted and restricted (under H_0) MLEs of θ based on X and computing the value of $W = -2ln(LRT(X))$, which is indeed the profile likelihood approach. To carry out the LRT test, naturally we need to determine the cut-off point of the null distribution of W. Following Self and Liang (1987) and Prop. 2.2, asymptotically, such a distribution is given by the distribution of the difference of two quadratic forms based on Z. This is discussed in Kopylev and Sinha (2011) and we have the following expressions for the unrestricted MLEs of the parameters θ in the distribution of Z.

$$
\begin{aligned}
(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)' &= (Z_1, Z_2, Z_3)', if \quad Z_1 \geq 0, Z_2 \geq 0, Z_3 \geq 0 \\
&= (0, Z_{2.1}, Z_{3.1})', if \quad Z_1 < 0, Z_{2.1} > 0, Z_{3.1} \geq 0 \\
&= (Z_{1.2}, 0, Z_{3.2})', if \quad Z_2 < 0, Z_{1.2} > 0, Z_{3.2} \geq 0 \\
&= (Z_{1.3}, Z_{2.3}, 0)', if \quad Z_3 < 0, Z_{1.3} \geq 0, Z_{2.3} \geq 0 \\
&= (0, 0, Z_{3.12})', if \quad Z_{1.2} \leq 0, Z_{2.1} \leq 0, Z_{3.12} > 0 \\
&= (0, Z_{2.13}, 0)', if \quad Z_{1.3} < 0, Z_{3.1} < 0, Z_{2.13} > 0 \\
&= (Z_{1.23}, 0, 0)', if \quad Z_{2.3} < 0, Z_{3.2} < 0, Z_{1.23} > 0 \\
&= (0, 0, 0)', if \quad Z_{1.23} \leq 0, Z_{2.13} \leq 0, Z_{3.12} \leq 0.
\end{aligned}
\tag{40}
$$

where $Z_{..}$ are the *usual* residual terms. Naturally we need to plug in these estimates of the MLEs in $Q(Z_1, Z_2, Z_3 \mid \theta_1, \theta_2, \theta_3)$ and simplify to get an expression of the unrestricted minimum, i.e., $Q_1(Z)$. Writing $Q(Z \mid \theta) = Q(Z_1, Z_2, Z_3 \mid \theta_1, \theta_2, \theta_3) + Q(Z_4, \cdots, Z_p \mid \theta)$, since the unrestricted minimum of the second term is 0, it follows that

$$
\begin{aligned}
min_{\theta \in \Omega^*} Q(Z \mid \theta) &= min_{\theta_1 \geq 0, \theta_2 \geq 0, \theta_3 \geq 0} Q(Z_1, Z_2, Z_3 \mid \theta_1, \theta_2, \theta_3) \\
&= Q(Z_1, Z_2, Z_3 \mid \hat{\theta}_{1N}, \hat{\theta}_{2N}, \hat{\theta}_{3N}).
\end{aligned}
\tag{41}
$$

Because of the nature of $(\hat{\theta}_{1N}, \hat{\theta}_{2N}, \hat{\theta}_{3N})$ given above in (40), it is obvious that the unrestricted minimum value of the relevant quadratic of the likelihood based on Z can be divided into eight disjoint sets, each set having a distinct value of the quadratic. This is explicitly presented in Kopylev and Sinha (2011).

On the other hand, to compute the restricted maximum likelihood under the null hypothesis $H_0 : \theta_4 = \theta_{40} + \delta$ in the X space which is equivalent to $\theta_4 = \delta$ in the Z space, note that all we need to compute are the restricted MLEs of θ_1, θ_2 and θ_3, subject to their being non-negative. This is derived as follows. We write

$$Q(Z_1, Z_2, Z_3, Z_4 \mid \theta_1, \theta_2, \theta_3, \theta_4 = \delta)$$
$$= Q(Z_4 \mid \delta) + Q(Z_1, Z_2, Z_3 \mid Z_4, \theta_1, \theta_2, \theta_3, \theta_4 = \delta) \qquad (42)$$

and note from (15) that the second quadratic can be expressed as

$$Q(Z_1, Z_2, Z_3 \mid Z_4, \theta_1, \theta_2, \theta_3, \theta_4 = \delta) = (Z_{1.4}, Z_{2.4}, Z_{3.4}) \tilde{\Sigma}^{-1} (Z_{1.4}, Z_{2.4}, Z_{3.4})' \qquad (43)$$

where the residuals $Z_{1.4}, Z_{2.4}, Z_{3.4}$ and $\tilde{\Sigma}$ are defined as

$$Z_{i.4} = Z_i - \theta_i - \rho_{i4}\sigma_i (Z_4 - \delta)/\sigma_4, \quad i = 1, 2, 3$$
$$\tilde{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{21}/\sigma_{44}$$
$$\Sigma_{11} : 3 \times 3 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33.} \end{pmatrix} \qquad (44)$$
$$\Sigma_{12} = (\sigma_{14}, \sigma_{24}, \sigma_{34})'$$

To determine the *restricted* MLEs of the parameters $\theta_1, \theta_2, \theta_3$ which must be non-negative, by minimizing $Q(Z_1, Z_2, Z_3 \mid Z_4, \theta_1, \theta_2, \theta_3, \theta_4 = \delta)$, under the null hypothesis: $\theta_4 = \delta$, we can readily apply the above results. Based on the *first step* residuals $(Z_{1.4} = Z_1^*, Z_{2.4} = Z_2^*, Z_{3.4} = Z_3^*)$ and their dispersion matrix $\tilde{\Sigma}$, we define the *second step* residuals $(Z_{2.1}^*, Z_{1.2}^*, Z_{2.3}^* Z_{3.2}^* Z_{1.3}^* Z_{3.1}^*, Z_{2.13}^* Z_{1.23}^* Z_{3.12}^*)$ in the usual fashion. Then, the restricted MLEs are given by:

$$
\begin{aligned}
(\hat{\theta}_{1N}, \hat{\theta}_{2N}, \hat{\theta}_{3N})_{null}' &= (Z_1^*, Z_2^*, Z_3^*)', if \quad Z_1^* \geq 0, Z_2^* \geq 0, Z_3^* \geq 0 \\
&= (0, Z_{2.1}^*, Z_{3.1}^*)', if \quad Z_1^* < 0, Z_{2.1}^* > 0, Z_{3.1}^* \geq 0 \\
&= (Z_{1.2}^*, 0, Z_{3.2}^*)', if \quad Z_2^* < 0, Z_{1.2}^* > 0, Z_{3.2}^* \geq 0 \\
&= (Z_{1.3}^*, Z_{2.3}^*, 0)', if \quad Z_3^* < 0, Z_{1.3}^* \geq 0, Z_{2.3}^* \geq 0 \\
&= (0, 0, Z_{3.12}^*)', if \quad Z_{1.2}^* \leq 0, Z_{2.1}^* \leq 0, Z_{3.12}^* > 0 \\
&= (0, Z_{2.13}^*, 0)', if \quad Z_{1.3}^* < 0, Z_{3.1}^* < 0, Z_{2.13}^* > 0 \\
&= (Z_{1.23}^*, 0, 0)', if \quad Z_{2.3}^* < 0, Z_{3.2}^* < 0, Z_{1.23}^* > 0 \\
&= (0, 0, 0)', if \quad Z_{1.23}^* \leq 0, Z_{2.13}^* \leq 0, Z_{3.12}^* \leq 0.
\end{aligned} \qquad (45)
$$

This then results in the restricted quadratic form

$$
\begin{aligned}
min_{\theta \in \Omega_0^*} Q(Z \mid \theta \in \Omega_0^*) &\\
&= min_{\theta \in \Omega_0^*} Q(Z_1, Z_2, Z_3, Z_4 \mid \theta_1, \theta_2, \theta_3, \theta_4 = \delta) \\
&= Q(Z_4 \mid \delta) + Q(Z_1, Z_2, Z_3 \mid Z_4, \hat{\theta}_{1;null}, \hat{\theta}_{2;null}, \hat{\theta}_{3;null}, \theta_4 = \delta)
\end{aligned} \qquad (46)
$$

Using (41) and (46), we obtain the difference $\Delta(Z)$ between the two quadratic forms, whose distribution under $Z \sim N[0, \Sigma]$ is essentially the asymptotic null distribution of $-2ln(LRT)$ based on X. Obviously, the distribution of $\Delta(Z)$ would depend only on the

nature of correlations among the estimates of the four parameters - one parameter of interest, namely, θ_4, and the three boundary parameters, $\theta_1, \theta_2, \theta_3$. It may be remarked that the null distribution of $\Delta(Z)$ may have to be obtained by simulation and it is obvious that we can take the dispersion matrix of Z as the correlation matrix, without any loss of generality. It would be interesting to check through simulations if this null distribution really depends on the correlations. Let us recall that this was not the case in the previous two instances.

A Wald-type test of $H_0 : \theta_4 = \theta_{40} + \delta$ versus the one-sided alternative $H_1 : \theta_4 > \theta_{40} + \delta$ based on X can be easily derived using the unrestricted MLE $\hat{\theta}_{4N}$ of θ_4. This readily follows from the conditional distribution of Z_4, given (Z_1, Z_2, Z_3), and plugging in the *restricted* MLEs of $\theta_1, \theta_2, \theta_3$ derived above. Such a test would reject H_0 for large values of $\hat{\theta}_{4N}$, or $\sqrt{N}(\hat{\theta}_{4N} - \theta_{40} - \delta)$ and the cut-off point of its null distribution is asymptotically computed from the distribution of the MLE of θ_4 based on Z in presence of three boundary parameters. Similarly, for testing H_0 versus the both-sided alternative $H_1 : \theta_4 \neq \theta_{40} + \delta$, one could reject H_0 for large values of $\sqrt{N} | \hat{\theta}_{4N} - \theta_{40} - \delta |$ and the cut-off point is determined analogously from its null distribution. Obviously, this distribution, which depends on the correlation structure between Z_4 and (Z_1, Z_2, Z_3), can be easily derived.

3. Applications of general theory to multistage Weibull models

In this section we consider some applications of the general theory developed in Section 2 in the context of multistage Weibull models and analyze some relevant data sets. In the sequel, we consider three cases: *linear* Weibull model, *quadratic* Weibull model and *cubic* Weibull model.

Quite generally, assume that there are $m+1$ dose groups with doses $d_0 = 0$ (control), $d_1, \cdots, d_m$ and n_i subjects in the *ith* dose group, yielding independent $X_i \sim B[n_i, \pi_i(\theta \mid d_i)]$, $i = 0, \cdots, m$ where

$$\pi_i(\theta \mid d_i) = 1 - e^{-\left[\theta_0 + d_i\theta_1 + \cdots + d_i^k\theta_k\right]}. \tag{47}$$

Here $\theta = (\theta_0, \cdots, \theta_k) \neq 0$ and $\theta_i \geq 0$ for all i, which are considered as the *natural* model restrictions! There are three kinds of inference problems of interest to us. First, inference about the absolute risk (AR) $\pi(\theta \mid d^*)$ at a given dose d^*; secondly, inference about the extra risk (ER) $\gamma(d^*)$ at a specified dose d^*, where $\gamma(d^*)$ is defined by

$$\gamma(d^*) = [\pi(\theta \mid d^*) - \pi(\theta \mid d_0)] / [1 - \pi(\theta \mid d_0)] = 1 - e^{-\left[\theta_1 d^* + \theta_2 d^{*2} + \cdots + \theta_k d^{*k}\right]}. \tag{48}$$

The third inference problem is about the *benchmark* dose (BMD) d^* which yields a specified relative risk γ. From the expression for $\pi(\theta \mid d^*)$, it is clear that the first two

inference problems are essentially concerned with linear functions of θ, while the third problem is about drawing inference for $d*$ which satisfies the polynomial equation:

$$ln(1/(1-\gamma)) = \theta_1 d* + \theta_2 d*^2 + \cdots + \theta_k d*^k .$$ (49)

Obviously, $d*$ is an implicit function of θ whose solution is easy to obtain in linear, quadratic and cubic cases. We mention in passing that sometimes we may be interested in what is called an *added risk* at a given dose d, defined by $ADR(d) = \pi(\theta \mid d) - \pi(\theta \mid d_0)$. Since the inference on ADR is very similar to that on AR, we do not pursue it here.

Case 3.0.1 Linear Weibull model. Taking $k = 1$ in (47), we get

$\pi_i(\theta \mid d_i) = 1 - e^{-\left[\theta_0 + d_i \theta_1\right]}$, which readily gives

$$AR(d*) = 1 - e^{-\left[\theta_0 + d* \theta_1\right]}$$ (50)

$$ER(d*) = 1 - e^{-d* \theta_1}$$ (51)

$$BMD = (1/\theta_1) ln(1/(1-\gamma)).$$ (52)

Appropriate statistical inference about the above quantities follows upon computing the *natural* MLEs of the two parameters θ_0 and θ_1 based on the data $(X_0, \cdots, X_m)$, namely $\hat{\theta}_{0N}$ and $\hat{\theta}_{1N}$, and applying their asymptotic distributional results derived in Section 2. It is easy to derive the Fisher information matrix $I(\theta_0, \theta_1) : 2 \times 2$ under the assumption of independent binomial distributions of the X_i's, and hence $\Sigma(\theta_0, \theta_1) = I(\theta_0, \theta_1)^{-1}$, where $I(\theta_0, \theta_1)$ is given by

$$I(\theta_0, \theta_1) = \begin{pmatrix} \sum_{i=0}^{m} \dfrac{1 - \pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i \dfrac{1 - \pi_i(\theta)}{\pi_i(\theta)} \\ \sum_{i=0}^{m} d_i \dfrac{1 - \pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^2 \dfrac{1 - \pi_i(\theta)}{\pi_i(\theta)} \end{pmatrix}.$$ (53)

When $\theta_0 > 0$, $\theta_1 > 0$, we clearly have a *regular* parametric scenario. By applying the standard asymptotic theory, we then get the following result.

$$\sqrt{N}(\hat{\theta}_{0N} - \theta_0, \hat{\theta}_{1N} - \theta_1) \to N_2[(0,0), \Sigma(\theta_0, \theta_1)]$$ (54)

$$\sqrt{N}[(\hat{\theta}_{0N} + d* \hat{\theta}_{1N}) - (\theta_0 + d* \theta_1)] \to N[0, (1, d*)\Sigma(\theta_0, \theta_1)(1, d*)'].$$ (55)

We remark that for most inference purposes, θ_0 and θ_1 in $\Sigma(\theta_0, \theta_1)$ are replaced by their MLEs or their null hypotheses values, if any. Drawing inference about AR, ER and BMD is fairly routine in this case.

When $\theta_0 = 0$, which may well happen in some situations, we have the case of one parameter point being on the boundary, and hence the asymptotic distributional and inferential results mentioned above are *not* true. Using Self and Liang (1987), we get

$$\sqrt{N}(\hat{\theta}_{0N}, \hat{\theta}_{1N} - \theta_1) \rightarrow (\hat{\theta}_0(Z_0, Z_1), \hat{\theta}_1(Z_0, Z_1)) \qquad (56)$$

where $Z = (Z_0, Z_1) \sim N[0, \hat{\Sigma}]$ and $\hat{\theta}_0(Z_0, Z_1) = 0$ if $Z_0 \leq 0$, $= Z_0$ if $Z_0 > 0$, and $\hat{\theta}_1(Z_0, Z_1) = Z_{1,0}$ if $Z_0 \leq 0$, and $= Z_1$ if $Z_0 > 0$. Here $Z_{1,0} = Z_1 - \rho \sigma_1 Z_0 / \sigma_0$ is the standard residual of Z_1 on Z_0.

A remark about $\hat{\Sigma}$ is in order here. Since Σ is obviously a function of $\theta = (\theta_0, \theta_1)$, we can use $\hat{\Sigma}(\hat{\theta}_N)$. In case it so happens that $X_0 = 0$ so that $\hat{\theta}_{0N} = 0$, computation of $\hat{\Sigma}$ may pose some difficulty (in terms of singularities!), and then we can take an arbitrary small value of $\hat{\theta}_{0N}$.

Returning to the inference problems, although we still have the same point estimates as before, namely, $\hat{AR}(d^*) = 1 - e^{-\left[\hat{\theta}_{0N} + d^* \hat{\theta}_{1N}\right]}$, $\hat{ER}(d^*) = 1 - e^{-d^* \hat{\theta}_{1N}}$ and $\hat{BMD} = (1 / \hat{\theta}_{1N}) ln(1 / (1 - \gamma))$, the asymptotic distributions of these estimates are not normal any more, and in fact are obtained by replacing $\hat{\theta}_{0N}$ and $\hat{\theta}_{1N}$ by $\hat{\theta}_0(Z_1, Z_2)$ and $\hat{\theta}_1(Z_1, Z_2)$, respectively, and deriving the resultant distributions by simulation.

We have pursued in Section 4 the inference methods for $ER(d^*) = 1 - e^{-d^* \theta_1}$ and $BMD = (1 / \theta_1) ln(1 / (1 - \gamma))$ on the basis of $\hat{ER} = 1 - e^{-d^* \hat{\theta}_{1N}}$ and $\hat{BMD} = (1 / \hat{\theta}_{1N}) ln(1 / (1 - \gamma))$.

What we have described so far is the Wald approach! For the derivation of the LRT for hypotheses about $ER(d^*)$ and BMD, since these quantities involve only θ_1, it follows from the general theory discussed in Section 2 that irrespective of the nature of θ_0, whether a boundary point or not, the asymptotic distribution of $-2ln(LRT)$ remains as χ^2 with 1 d.f. We have compared the Wald test with the LRT test in a few cases in Section 4.

Case 3.0.2 Quadratic Weibull model. Taking $k = 2$ in (47), we get $\pi_i(\theta | d_i) = 1 - e^{-\left[\theta_0 + d_i \theta_1 + d_i^2 \theta_2\right]}$, which readily gives

$$AR(d^*) = 1 - e^{-\left[\theta_0 + d^* \theta_1 + d^{*2} \theta_2\right]} \qquad (57)$$

$$ER(d^*) = 1 - e^{-(d^* \theta_1 + d^{*2} \theta_2)} \qquad (58)$$

$$BMD = \frac{\sqrt{[\theta_1^2 + 4\theta_2\gamma]} - \theta_1}{2\theta_2}.$$

(59)

Appropriate statistical inference about the above quantities follows upon computing the *natural* MLEs of the three parameters θ_0, θ_1 and θ_2 based on the data $(X_0, \cdots, X_m)$, namely $\hat{\theta}_{0N}$, $\hat{\theta}_{1N}$ and $\hat{\theta}_{2N}$, and applying their asymptotic distributional results derived in Section 2. It is easy to derive the Fisher information matrix $I(\theta_0, \theta_1, \theta_2) : 3 \times 3$ under the assumption of independent binomial distributions of the X_i's, and hence $\Sigma(\theta_0, \theta_1, \theta_2) = I(\theta_0, \theta_1, \theta_2)^{-1}$, where $I(\theta_0, \theta_1, \theta_2)$ is given by

$$I(\theta_0, \theta_1, \theta_2) = \begin{pmatrix} \sum_{i=0}^{m} \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^2 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} \\ \sum_{i=0}^{m} d_i \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^2 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^3 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} \\ \sum_{i=0}^{m} d_i^2 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^3 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m} d_i^4 \frac{1-\pi_i(\theta)}{\pi_i(\theta)} \end{pmatrix}.$$

(60)

When $\theta_0 > 0$, $\theta_1 > 0$, $\theta_2 > 0$, we clearly have a *regular* parametric scenario situation. By standard asymptotic theory, we then get the following result:

$$\sqrt{N}(\hat{\theta}_{0N} - \theta_0, \hat{\theta}_{1N} - \theta_1, \hat{\theta}_{2N} - \theta_2) \rightarrow N_3[(0,0,0), \Sigma(\theta)]$$

(61)

$$\sqrt{N}[(\hat{\theta}_{0N} + d * \hat{\theta}_{1N} + d^2 * \hat{\theta}_{2N}) - (\theta_0 + d * \theta_1 + d^2 * \theta_2)] \rightarrow$$
$$N[0, (1, d*, d^2*)\Sigma(\theta)(1, d*, d^2*)']$$

(62)

$$\sqrt{N}[(\hat{\theta}_{1N} + d * \hat{\theta}_{2N}) - (\theta_1 + d * \theta_2)] \rightarrow N[0, (1, d*)\tilde{\Sigma}(\theta)(1, d*)'].$$

(63)

Here $\tilde{\Sigma}(\theta)$ is the 2×2 lower sub matrix of $\Sigma(\theta)$ obtained by deleting the first row and the first column. Again, drawing inference about AR, RR and BMD is very routine here. Details are omitted.

We now consider the case when some parameters may lie on the boundaries. Since we have assumed a quadratic Weibull model, there are three possible scenarios: (*i*) $\theta_0 = 0$, (*ii*) $\theta_1 = 0$, and (*iii*) $\theta_0 = \theta_1 = 0$.

When (*i*) $\theta_0 = 0$, which may well happen in some situations, we have the case of one parameter point being on the boundary, and hence the asymptotic distributional and inferential results mentioned above are not true. We should note that for statistical inference about $ER(d*)$ and BMD which involve the parameters θ_1 and θ_2, all we would

need is the asymptotic joint distribution of the relevant MLEs. Using Prop. 2.1, quite generally we get

$$\sqrt{N}(\hat{\theta}_{0N},\hat{\theta}_{1N}-\theta_1,\hat{\theta}_{2N}-\theta_2) \rightarrow (\hat{\theta}_0(Z_0,Z_1,Z_2),\hat{\theta}_1(Z_0,Z_1,Z_2),\hat{\theta}_2(Z_0,Z_1,Z_2)) \quad (64)$$

where $Z=(Z_0,Z_1,Z_2) \sim N[0,\hat{\Sigma}]$ and from the discussion in Section 2, Theorem 2.2, we get

$$\hat{\theta}_0(Z_0,Z_1,Z_2)=Z_0 \quad if \quad Z_0>0, \quad and \quad =0 \quad if \quad Z_0 \le 0 \qquad (65)$$

$$\hat{\theta}_1(Z_0,Z_1,Z_2)=Z_1 \quad if \quad Z_0>0, \quad and \quad =Z_{1.0} \quad if \quad Z_0 \le 0. \qquad (66)$$

$$\hat{\theta}_2(Z_0,Z_1,Z_2)=Z_2 \quad if \quad Z_0>0, \quad and \quad =Z_{2.0} \quad if \quad Z_0 \le 0. \qquad (67)$$

where $Z_{1.0}$ and $Z_{2.0}$ are the usual *residual* variables. A remark about $\hat{\Sigma}$ is in order here. Since Σ is obviously a function of θ, we can use $\hat{\Sigma}(\hat{\theta}_N)$. In case it so happens that $X_0=0$ so that $\hat{\theta}_{0N}=0$, computation of $\hat{\Sigma}$ may pose some difficulty (in terms of singularities!), and then we can take an arbitrary small value of $\hat{\theta}_{0N}$.

Returning to the inference problems, although we still have the same point estimates as before, namely,

$$\hat{AR}(d^*)=1-e^{-[\hat{\theta}_{0N}+d^*\hat{\theta}_{1N}+d^{*2}\hat{\theta}_{2N}]}, \qquad (68)$$

$$\hat{ER}(d^*)=1-e^{-(d^*\hat{\theta}_{1N}+d^{*2}\hat{\theta}_{2N})}, \qquad (69)$$

$$\hat{BMD}=\left[\sqrt{[\hat{\theta}_{1N}^2+4\hat{\theta}_{2N}\gamma]}-\hat{\theta}_{1N}\right]/2\hat{\theta}_{2N}, \qquad (70)$$

the asymptotic distributions of these estimates are not normal any more, and, in fact, these are obtained by replacing $\hat{\theta}_{0N}$, $\hat{\theta}_{1N}$ and $\hat{\theta}_{2N}$ by $\hat{\theta}_0(Z_0,Z_1,Z_2)$, $\hat{\theta}_1(Z_0,Z_1,Z_2)$ and $\hat{\theta}_2(Z_0,Z_1,Z_2)$, respectively, and deriving the resultant distributions by simulation.

We have explicitly pursued below the inference methods for $ER(d^*)=1-e^{-(d^*\theta_1+d^{*2}\theta_2)}$ and $BMD=[\sqrt{[\theta_1^2+4\theta_2\gamma]}-\theta_1]/2\theta_2$ on the basis of $\hat{ER}(d^*)=1-e^{-(d^*\hat{\theta}_{1N}+d^{*2}\hat{\theta}_{2N})}$ and $\hat{BMD}=[\sqrt{[\hat{\theta}_{1N}^2+4\hat{\theta}_{2N}\gamma]}-\hat{\theta}_{1N}]/2\hat{\theta}_{2N}$. Here are the results.

Inference about ER. Since $ER(d^*)=1-e^{-(d^*\theta_1+d^{*2}\theta_2)}$, without any loss of generality, we consider equivalently the problem of drawing suitable inference about $\tilde{ER}(d^*)=\theta_1+d^*\theta_2$ which is based on $\widehat{\tilde{ER}(d^*)}=\hat{\theta}_{1N}+d^*\hat{\theta}_{2N}$. Applying Theorem 2.2, we conclude that

$$\sqrt{N}[(\hat{\theta}_{1N}+d^*\hat{\theta}_{2N})-(\theta_1+d^*\theta_2)] \quad \sim \quad Z_1+d^*Z_2 \quad if \quad Z_0>0$$
$$\sim \quad Z_{1.0}+d^*Z_{2.0} \quad if \quad Z_0 \le 0. \qquad (71)$$

Since $\hat{\Sigma}(\hat{\theta})$ is known, it is rather easy to simulate the asymptotic distribution of $\sqrt{N}[(\hat{\theta}_{1N} + d * \hat{\theta}_{2N}) - (\theta_1 + d * \theta_2)]$, from which the cut-off points of this distribution can be obtained. Tests and confidence intervals about $\tilde{ER}(d*)$ are then routinely derived.

Inference about BMD. Using (64) and (70), it is easy to generate the asymptotic distribution of $BMD(d*)$ from which its cut-off points can be derived and used for tests and confidence intervals for $BMD(d*)$.

What we have described so far is the Wald approach! For the derivation of the LRT for hypotheses about $ER(d*)$ and BMD, since these quantities involve only θ_1 and θ_2 which are *regular* parameters, it follows from the general theory discussed in Section 2 that irrespective of the nature of θ_0, whether a boundary point or not, the asymptotic distribution of $-2ln(LRT)$ remains as χ^2 with an appropriate d.f. We have compared the Wald test with the LRT test in a few cases in Section 4.

We next consider the case *(ii)* $\theta_1 = 0$ which is still the case of one boundary parameter. Using Prop. 2.1, the asymptotic joint distribution of the MLEs $\hat{\theta}_N$ is given by

$$\sqrt{N}(\hat{\theta}_{0N} - \theta_0, \hat{\theta}_{1N}, \hat{\theta}_{2N} - \theta_2) \rightarrow$$
$$\left(\hat{\theta}_0(Z_0, Z_1, Z_2), \hat{\theta}_1(Z_0, Z_1, Z_2), \hat{\theta}_2(Z_0, Z_1, Z_2) \right) \tag{72}$$

where, as before, $Z = (Z_0, Z_1, Z_2) \sim N[0, \hat{\Sigma}]$, and

$$\hat{\theta}_0(Z_0, Z_1, Z_2) = Z_0 \quad if \quad Z_1 > 0, \quad and \quad = Z_{0.1} \quad if \quad Z_1 \leq 0 \tag{73}$$
$$\hat{\theta}_1(Z_0, Z_1, Z_2) = Z_1 \quad if \quad Z_1 > 0, \quad and \quad = 0 \quad if \quad Z_1 \leq 0 \tag{74}$$
$$\hat{\theta}_2(Z_0, Z_1, Z_2) = Z_2 \quad if \quad Z_1 > 0, \quad and \quad = Z_{2.1} \quad if \quad Z_1 \leq 0 \tag{75}$$

Using the above asymptotic distribution of the MLEs of θ, we can approximate the distributions of $\hat{ER}(d*) = 1 - e^{-\left(d*\hat{\theta}_{1N} + d*^2\hat{\theta}_{2N} \right)}$ and $\hat{BMD} = [\sqrt{[\hat{\theta}_{1N}^2 + 4\hat{\theta}_{2N}\gamma]} - \hat{\theta}_{1N}] / 2\hat{\theta}_{2N}$ by simulations and use them to carry out tests and confidence intervals of ER and BMD. We remark that the LRT for $ER(d*)$ which is equivalent to the linear function $\theta_1 + d * \theta_2$ can be carried out as follows. From the discussion in Section 2, we first consider a reparametrization from $(\theta_0, \theta_1, \theta_2)$ to $(\theta_0, \theta_1, \theta_1 + d * \theta_2 = \theta_2*)$ and express the joint distribution in terms of these *new* parameters. The problem then boils down to inference about $\theta_2 *$ when θ_1 lies on the boundary! This is precisely the set up discussed in Section 2.

Finally, we consider the case when *(iii)* $\theta_0 = \theta_1 = 0$, which involves two points on the boundary, a case discussed in Section 2. Following Prop. 2.1, the asymptotic distribution of the MLEs $\hat{\theta}_N$ of θ is given by

$$\sqrt{N}(\hat{\theta}_{0N},\hat{\theta}_{1N},\hat{\theta}_{2N}-\theta_2) \rightarrow \left(\hat{\theta}_0(Z_0,Z_1,Z_2),\hat{\theta}_1(Z_0,Z_1,Z_2),\hat{\theta}_2(Z_0,Z_1,Z_2)\right) \qquad (76)$$

where again $Z = (Z_0,Z_1,Z_2) \sim N[0,\hat{\Sigma}]$ and $\hat{\theta}_0(Z_0,Z_1,Z_2)$, $\hat{\theta}_1(Z_0,Z_1,Z_2)$ and $\hat{\theta}_2(Z_0,Z_1,Z_2)$ are as defined in (73) - (75) with obvious changes.

As remarked earlier, while using $\hat{\Sigma}$, we can use either the natural MLEs $\hat{\theta}_N$ or take $\theta_0 = \theta_1 = 0$ and replace θ_2 by $\hat{\theta}_{2N}$.

From the above asymptotic distribution, it is indeed possible to derive the asymptotic distribution of any function of the MLEs such as $\hat{AR}(d^*)$, $\hat{ER}(d^*)$ and $\hat{BMD}$. Details are omitted.

Case 3.0.3 Cubic Weibull Model. Taking $k = 3$ in (47), we get

$\pi_i(\theta \mid d_i) = 1 - e^{-[\theta_0 + d_i\theta_1 + d_i^2\theta_2 + d_i^3\theta_3]}$, which readily gives

$$AR(d^*) = e^{-\left[\theta_0 + d^*\theta_1 + d^{*2}\theta_2 + d^{*3}\theta_3\right]} \qquad (77)$$

$$ER(d^*) = 1 - e^{-(d^*\theta_1 + d^{*2}\theta_2 + d^{*3}\theta_3)}. \qquad (78)$$

Regarding BMD d^*, we note from (49) that d^* is the solution of a cubic equation. Following Cardano, an explicit formula for d^* is given as follows. Let $ln(1/(1-\gamma)) = -u$. Define

$$A(\theta) = \frac{3\theta_3\theta_1 - \theta_2^2}{3\theta_3^2} \qquad B(\theta) = \frac{9\theta_3\theta_2\theta_1 - 2\theta_2^3 - 27u\theta_3^2}{27\theta_3^3} \qquad (79)$$

$$v(\theta) = \frac{-\sqrt{[B^2(\theta) + \frac{4A^3(\theta)}{27}]} - B(\theta)}{2}, \quad t(\theta) = v(\theta)^{1/3} \qquad (80)$$

$$s(\theta) = \frac{A(\theta)}{3t(\theta)}, \quad y(\theta) = s(\theta) - t(\theta). \qquad (81)$$

Then BMD $d^*(\theta)$ is given by

$$d^*(\theta) = y(\theta) - \frac{\theta_2}{3\theta_3}. \qquad (82)$$

Appropriate statistical inference about the above quantities follows upon computing the *natural* MLEs of the three parameters θ_0, θ_1, θ_2 and θ_3 based on the data $(X_0,\cdots,X_m)$, namely $\hat{\theta}_{0N}$, $\hat{\theta}_{1N}$, $\hat{\theta}_{2N}$ and $\hat{\theta}_{3N}$, and applying their asymptotic distributional results derived in Section 2. It is easy to derive the Fisher information matrix $I(\theta_0,\theta_1,\theta_2,\theta_3):4\times4$ under the assumption of independent binomial distributions of the X_i's, and hence $\Sigma(\theta_0,\theta_1,\theta_2,\theta_3) = I(\theta_0,\theta_1,\theta_2,\theta_3)^{-1}$, where $I(\theta_0,\theta_1,\theta_2,\theta_3)$ is given by

$$I(\theta_0,\theta_1,\theta_2,\theta_3) = \tag{83}$$

$$\begin{pmatrix} \sum_{i=0}^{m}\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^2\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^3\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} \\[2ex] \sum_{i=0}^{m}d_i\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^2\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^3\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^4\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} \\[2ex] \sum_{i=0}^{m}d_i^2\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^3\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^4\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^5\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} \\[2ex] \sum_{i=0}^{m}d_i^3\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^4\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^5\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} & \sum_{i=0}^{m}d_i^6\dfrac{1-\pi_i(\theta)}{\pi_i(\theta)} \end{pmatrix}.$$

When $\theta_0 > 0$, $\theta_1 > 0$, $\theta_2 > 0$, $\theta_3 > 0$, we clearly have a *regular* parametric scenario situation. By Prop. 2.1, we then get the following result.

$$\sqrt{N}(\hat{\theta}_{0N}-\theta_0,\hat{\theta}_{1N}-\theta_1,\hat{\theta}_{2N}-\theta_2,\hat{\theta}_{3N}-\theta_3) \to N_4[(0,0,0,0),\Sigma(\theta)] \tag{84}$$

$$\sqrt{N}[(\hat{\theta}_{0N}+d*\hat{\theta}_{1N}+d^2*\hat{\theta}_{2N}+d^3*\hat{\theta}_{3N})-(\theta_0+d*\theta_1+d^2*\theta_2+d^3*\theta_3)] \tag{85}$$

$$\to N[0,(1,d*,d^2*,d^3*)\Sigma(\theta)(1,d*,d^2*,d^3*)']$$

$$\sqrt{N}[(\hat{\theta}_{1N}+d*\hat{\theta}_{2N}+d^2*\hat{\theta}_{3N})-(\theta_1+d*\theta_2+d^2*\theta_3)]$$

$$\to N[0,(1,d*,d^2*)\tilde{\Sigma}(\theta)(1,d*,d^2*)']. \tag{86}$$

Here $\tilde{\Sigma}(\theta)$ is the 3×3 lower sub matrix of $\Sigma(\theta)$ obtained by deleting the first row and the first column. Statistical inference about AR, RR and BMD is very routine here. Details are omitted.

We now consider the case when some parameters may lie on the boundaries. Since we have assumed a cubic Weibull model, there are several possible scenarios: (*i*) $\theta_0 = 0$, (*ii*) $\theta_1 = 0$, (*iii*) $\theta_2 = 0$, (*iv*) $\theta_0 = \theta_1 = 0$, (*v*) $\theta_0 = \theta_2 = 0$, (*vi*) $\theta_1 = \theta_2 = 0$, and (*vii*) $\theta_0 = \theta_1 = \theta_2 = 0$.

When (*i*) $\theta_0 = 0$, which may well happen in some situations, we have the case of one parameter point being on the boundary, and hence the asymptotic distributional and inferential results mentioned above are not true. We should note that for statistical inference about $ER(d*)$ and BMD which involve the parameters θ_1, θ_2 and θ_3, all we would need is the asymptotic joint distribution of the relevant MLEs. Using Prop. 2.1, quite generally we get

$$\sqrt{N}(\hat{\theta}_{0N},\hat{\theta}_{1N}-\theta_1,\hat{\theta}_{2N}-\theta_2,\hat{\theta}_{3N}-\theta_3) \to$$

$$(\hat{\theta}_0(Z_0,Z_1,Z_2,Z_3),\hat{\theta}_1(Z_0,Z_1,Z_2,Z_3),\hat{\theta}_2(Z_0,Z_1,Z_2,Z_3),\hat{\theta}_3(Z_0,Z_1,Z_2,Z_3)) \tag{87}$$

where $Z = (Z_0,Z_1,Z_2,Z_3) \sim N[0,\hat{\Sigma}]$ and from the discussion in Section 2, Theorem 2.2, we get

$$\hat{\theta}_0(Z_0, Z_1, Z_2, Z_3) = Z_0 \quad if \quad Z_0 > 0, \quad and \quad = 0 \quad if \quad Z_0 \le 0 \tag{88}$$

$$\hat{\theta}_1(Z_0, Z_1, Z_2, Z_3) = Z_1 \quad if \quad Z_0 > 0, \quad and \quad = Z_{1.0} \quad if \quad Z_0 \le 0. \tag{89}$$

$$\hat{\theta}_2(Z_0, Z_1, Z_2, Z_3) = Z_2 \quad if \quad Z_0 > 0, \quad and \quad = Z_{2.0} \quad if \quad Z_0 \le 0. \tag{90}$$

$$\hat{\theta}_3(Z_0, Z_1, Z_2, Z_3) = Z_3 \quad if \quad Z_0 > 0, \quad and \quad = Z_{3.0} \quad if \quad Z_0 \le 0. \tag{91}$$

where $Z_{1.0}$, $Z_{2.0}$ and $Z_{3.0}$ are the usual *residual* variables. A remark about $\hat{\Sigma}$ is in order here. Since Σ is obviously a function of θ, we can use $\hat{\Sigma}(\hat{\theta}_N)$. In case it so happens that $X_0 = 0$ so that $\hat{\theta}_{0N} = 0$, computation of $\hat{\Sigma}$ may pose some difficulty (in terms of singularities!), and then we can take an arbitrary small value of $\hat{\theta}_{0N}$.

Returning to the inference problems, although we still have the same point estimates as before, namely,

$$\hat{AR}(d^*) = 1 - e^{-\left[\hat{\theta}_{0N} + d^* \hat{\theta}_{1N} + d^{*2} \hat{\theta}_{2N} + d^{3*} \hat{\theta}_{3N}\right]}, \tag{92}$$

$$\hat{ER}(d^*) = 1 - e^{-(d^* \hat{\theta}_{1N} + d^{*2} \hat{\theta}_{2N} + d^{3*} \hat{\theta}_{3N})}, \tag{93}$$

and a similar expression for $\hat{BMD}$, the asymptotic distributions of these estimates are not normal any more, and, in fact, these are obtained by replacing $\hat{\theta}_{0N}$, $\hat{\theta}_{1N}$, $\hat{\theta}_{2N}$ and $\hat{\theta}_{3N}$ by $\hat{\theta}_0(Z_0, Z_1, Z_2, Z_3)$, $\hat{\theta}_1(Z_0, Z_1, Z_2, Z_3)$, $\hat{\theta}_2(Z_0, Z_1, Z_2, Z_3)$ and $\hat{\theta}_3(Z_0, Z_1, Z_2, Z_3)$, respectively, and deriving the resultant distributions by simulation.

We have explicitly pursued below the inference method for $ER(d^*) = 1 - e^{-(d^* \theta_1 + d^{*2} \theta_2 + d^{3*} \theta_3)}$ on the basis of $\hat{ER}(d^*) = 1 - e^{-\left(d^* \hat{\theta}_{1N} + d^{*2} \hat{\theta}_{2N} + d^{3*} \hat{\theta}_{3N}\right)}$. The same for *BMD* can be pursued with some extra effort! Here are the results.

Inference about ER. Since $ER(d^*) = 1 - e^{-(d^* \theta_1 + d^{*2} \theta_2 + d^{3*} \theta_3)}$, without any loss of generality, we consider equivalently the problem of drawing suitable inference about $\tilde{ER}(d^*) = \theta_1 + d^* \theta_2 + d^2 * \theta_3$ which is based on $\widetilde{ER}(d^*) = \hat{\theta}_{1N} + d^* \hat{\theta}_{2N} + d^2 * \hat{\theta}_{3N}$. Applying Theorem 2.2, we conclude that

$$\sqrt{N}\left[(\hat{\theta}_{1N}+d*\hat{\theta}_{2N}+d^2*\hat{\theta}_{3N})-(\theta_1+d*\theta_2+d^2*\theta_3)\right]$$

$$\sim Z_1+d*Z_2+d^2*Z_3 \quad if \quad Z_0 > 0 \qquad\qquad (94)$$

$$\sim Z_{1.0}+d*Z_{2.0}+d^2*Z_{3.0} \quad if \quad Z_0 \leq 0.$$

Since $\hat{\Sigma}(\hat{\theta})$ is known, it is rather easy to simulate the asymptotic distribution of $\sqrt{N}[(\hat{\theta}_{1N}+d*\hat{\theta}_{2N}+d^2*\hat{\theta}_{3N})-(\theta_1+d*\theta_2+d^2*\theta_3)]$, from which the cut-off points of this distribution can be obtained. Tests and confidence intervals about $\tilde{E}R(d^*)$ are then routinely derived.

What we have described so far is the Wald approach! For the derivation of the LRT for hypotheses about $ER(d^*)$ and BMD, since these quantities involve only θ_1 and θ_2 which are *regular* parameters, it follows from the general theory discussed in Section 2 that irrespective of the nature of θ_0, whether a boundary point or not, the asymptotic distribution of $-2ln(LRT)$ remains as χ^2 with an appropriate d.f. We have compared the Wald test with the LRT test in the case of ER below.

The cases (*ii*) $\theta_1 = 0$ and (*iii*) $\theta_2 = 0$ which are still concerned with one boundary parameter can be handled similarly. When (*iv*) $\theta_0 = \theta_1 = 0$, which involves two points on the boundary, following Prop. 2.1, the asymptotic distribution of the MLEs $\hat{\theta}_N$ of θ is given by

$$\sqrt{N}(\hat{\theta}_{0N}, \hat{\theta}_{1N}, \hat{\theta}_{2N}-\theta_2, \hat{\theta}_{3N}-\theta_3) \rightarrow \qquad\qquad (95)$$

$$\left(\hat{\theta}_0(Z_0,Z_1,Z_2,Z_3), \hat{\theta}_1(Z_0,Z_1,Z_2,Z_3), \hat{\theta}_2(Z_0,Z_1,Z_2,Z_3), \hat{\theta}_3(Z_0,Z_1,Z_2,Z_3)\right)$$

where again $Z = (Z_0,Z_1,Z_2,Z_3) \sim N[0,\hat{\Sigma}]$ and $\hat{\theta}_0(Z_0,Z_1,Z_2,Z_3)$, $\hat{\theta}_1(Z_0,Z_1,Z_2,Z_3)$, $\hat{\theta}_2(Z_0,Z_1,Z_2,Z_3)$ and $\hat{\theta}_3(Z_0,Z_1,Z_2,Z_3)$ are defined in Section 2.

As remarked earlier, while using $\hat{\Sigma}$, we can use either the natural MLEs $\hat{\theta}_N$ or take $\theta_0 = \theta_1 = 0$ and replace θ_2 and θ_3 by $\hat{\theta}_{2N}$ and $\hat{\theta}_{3N}$, respectively.

From the above asymptotic distribution, it is indeed possible to derive the asymptotic distribution of any function of the MLEs such as $\hat{A}R(d^*)$, $\hat{E}R(d^*)$ and $\hat{B}MD$. Details are omitted.

Finally, the case (*vii*) $\theta_0 = \theta_1 = \theta_2 = 0$ which involves three parameters on their boundaries can be dealt on the basis of the theory developed earlier. Again, details are omitted.

4. Applications

4.1 Simulations

Monte-Carlo simulations were performed to compare coverage of the confidence intervals for the multistage model parameters θ, extra risk (a linear function of parameters) as well as benchmark dose (linear or nonlinear, depending on the order of the model). We considered linear, quadratic and cubic multistage models (k=1,2,3). In each case, the true value of one or two of the parameters (but not the leading coefficient) was on the boundary (equal to zero) or the true value of one or two parameters was not on the boundary, but their estimates often were. All simulations were performed for 4 groups (1 control group and 3 dose groups), each of 50 animals. The chosen doses were 0, 1/4, 1/2 and 1. This imitates a standard rodent bioassay setup. The nonzero parameters were chosen to provide a range of incidences for each dose group.

Monte-Carlo simulations were performed as follows. The model parameters define the binomial response probabilities $P(d,\theta)$ at each of the four doses. For each of 5000 realizations, these probabilities were used to generate tumor incidence data $X \sim Binom(N = 50, P(d,\theta))$. The resulting experiment data was used to estimate parameters of the model, using BMDS programs (U.S. EPA, 2006) for extra risk equal to 0.1, a standard EPA practice when analyzing quantal response bioassay data. The information matrix was estimated. To avoid singularities, if the background rate was estimated to be 0, θ_0 was taken to be 5E-4, as discussed in Section 3. In such cases, the confidence interval for θ_0 depends on the chosen substitution value and therefore confidence intervals for θ_0 were not considered. It should be possible to derive a confidence interval for θ_0 via a re-sampling procedure, but this is beyond the scope of this report.

Using the information matrix, Wald confidence intervals were computed for all parameters described above. If a parameter estimate was on the boundary, the confidence interval was computed using the asymptotic distribution derived according to Self and Liang (1987) (See Section 3). The confidence interval values were retained and procedure repeated 5000 times. For estimating coverage of confidence intervals, only those realizations were retained for which the χ^2 goodness-of-fit P-value exceeded 0.1, conforming to the practice recommended by EPA (U.S. EPA 2012), and for which the estimate of the leading coefficient was positive, so that models of the same order are compared to each other.

In the following tables, the "Wald" column shows coverage for Wald confidence intervals, while the "Self--Liang" column shows coverage of the procedure such that for each simulation, depending on the estimated value of the corresponding parameter being on the boundary or not, the interval was computed according to the asymptotic

distribution derived in Self and Liang (1987) or as a Wald interval, respectively. The "Fit" column shows the percent of the simulations when the model fit well ($P > 0.1$) and the estimate of the leading coefficient was non-zero. The desired coverage is 90%. Only results for Benchmark Dose (BMD) are shown in the tables (results were similar for Extra Risk (ER) and non-background parameters).

For the linear model (Table 1), both Wald and Self--Liang confidence intervals have practically the same coverage and average length for a wide range of θ_1. The coverage is very close to the desired level.

Table 1: **BMD coverage for linear model under several scenarios, when some of the parameters or their estimates are on the boundary.**

Parameter	Wald		Self-Liang		LRT		Fit
True Values	CI	Length	CI	Length	CI	Length	%
$\theta_0 = 0$; $\theta_1 = 0.5$	90.2	0.121	89.1	0.132	90.2	0.125	95.8
$\theta_0 = 0$; $\theta_1 = 1$	90.1	0.046	90.2	0.048	90.5	0.046	95.4
$\theta_0 = 0$; $\theta_1 = 2$	90.3	0.020	90.4	0.020	90.6	0.020	95.7
$\theta_0 = 0$; $\theta_1 = 3$	90.2	0.013	90.8	0.013	90.6	0.013	96.2
$\theta_0 = 0.02$; $\theta_1 = 0.5$	90.3	0.138	92.1	0.141	89.3	0.154	91.9
$\theta_0 = 0.02$; $\theta_1 = 1$	90.1	0.049	91.6	0.050	89.7	0.051	91.8
$\theta_0 = 0.02$; $\theta_1 = 2$	90.2	0.020	91.0	0.021	90.2	0.021	92.7
$\theta_0 = 0.02$; $\theta_1 = 3$	90.3	0.013	91.1	0.013	90.4	0.013	92.9

For quadratic and cubic models (Table 2 and Table 3) with one or two parameters whose true or estimated value was on the boundary, the Wald confidence interval generally exceeded nominal coverage and Self-Liang confidence intervals are generally closer to the desired confidence level of 90%. Also, in almost all cases, Self--Liang intervals improve on average length compared to Wald intervals.

Table 2: **BMD coverage for quadratic model under several scenarios, when some of the parameters or their estimates are on the boundary.**

Parameter	Wald		Self-Liang		LRT		Fit
True Values	CI	Length	CI	Length	CI	Length	%
$\theta_0 = 0$; $\theta_1 = 0.05$; $\theta_2 = 0.5$	96.2	0.239	96.0	0.224	94.4	0.226	96.4
$\theta_0 = 0$; $\theta_1 = 0.05$; $\theta_2 = 1$	95.6	0.182	97.2	0.158	93.8	0.151	97.8
$\theta_0 = 0$; $\theta_1 = 0.05$; $\theta_2 = 2$	95.7	0.142	96.7	0.117	93.9	0.105	97.4
$\theta_0 = 0$; $\theta_1 = 0.05$; $\theta_2 = 3$	95.0	0.125	96.3	0.100	93.5	0.088	97.6
$\theta_0 = 0$; $\theta_1 = 0.8$; $\theta_2 = 0.5$	93.5	0.131	90.7	0.136	92.9	0.122	87.7
$\theta_0 = 0$; $\theta_1 = 0.8$; $\theta_2 = 1$	89.1	0.118	91.0	0.121	92.0	0.110	96.8
$\theta_0 = 0$; $\theta_1 = 0.8$; $\theta_2 = 2$	88.9	0.112	92.5	0.110	91.3	0.097	99.6
$\theta_0 = 0$; $\theta_1 = 0.8$; $\theta_2 = 3$	91.5	0.115	95.2	0.105	90.7	0.089	99.9
$\theta_0 = 0.02$; $\theta_1 = 0$; $\theta_2 = 0.5$	94.9	0.315	92.9	0.285	90.0	0.259	92.0
$\theta_0 = 0.02$; $\theta_1 = 0$; $\theta_2 = 1$	94.2	0.229	91.5	0.189	89.1	0.164	92.9
$\theta_0 = 0.02$; $\theta_1 = 0$; $\theta_2 = 2$	94.2	0.174	91.4	0.135	89.9	0.112	93.0
$\theta_0 = 0.02$; $\theta_1 = 0$; $\theta_2 = 3$	94.2	0.156	91.4	0.117	89.9	0.094	93.0
$\theta_0 = \theta_1 = 0$; $\theta_2 = 0.5$	96.3	0.238	93.8	0.226	92.0	0.229	97.9
$\theta_0 = \theta_1 = 0$; $\theta_2 = 1$	96.7	0.182	94.9	0.154	91.7	0.148	98.1
$\theta_0 = \theta_1 = 0$; $\theta_2 = 2$	96.3	0.143	94.5	0.114	91.1	0.102	97.5
$\theta_0 = \theta_1 = 0$; $\theta_2 = 3$	95.6	0.126	94.2	0.098	90.6	0.085	97.4
$\theta_0 = \theta_1 = 0.02$; $\theta_2 = 0.5$	93.8	0.333	94.0	0.296	91.3	0.265	91.8
$\theta_0 = \theta_1 = 0.02$; $\theta_2 = 1$	93.4	0.234	93.2	0.192	90.5	0.167	93.7
$\theta_0 = \theta_1 = 0.02$; $\theta_2 = 2$	93.0	0.169	92.8	0.134	90.8	0.113	93.4
$\theta_0 = \theta_1 = 0.02$; $\theta_2 = 3$	93.5	0.154	91.5	0.117	90.8	0.095	93.7

Table 3: **BMD coverage for cubic model under several scenarios, when some of the parameters or their estimates are on the boundary.**

Parameter	Wald		Self-Liang		LRT		Fit
True Values	CI	Length	CI	Length	CI	Length	%
$\theta_0 = 0.1$; $\theta_1 = 0$ $\theta_2 = 0$; $\theta_3 = 2$	98.8	0.569	92.7	0.240	90.3	0.184	79.9
$\theta_0 = 0$; $\theta_1 = 1$ $\theta_2 = 0$; $\theta_3 = 3$	94.9	0.920	89.6	0.122	89.9	0.120	83.9
$\theta_0 = 0$; $\theta_1 = 2$ $\theta_2 = 0$; $\theta_3 = 0.5$	100	0.731	86.6	0.055	91.9	0.074	43.6
$\theta_0 = 0.2$; $\theta_1 = 0.01$ $\theta_2 = 0.01$; $\theta_3 = 1$	99.9	1.8368	93.9	0.387	92.1	0.303	79.2

4.2. Example

In this section we consider an example from National Toxicological Program (NTP 1993). In this study, B6C3F1 mice were exposed to 1,3-butadiene. The outcome was heart hemangiosarcomas. This data was also analyzed by Bailer and Smith (1994). The data is given in the Table 1 (only doses up to 200ppm are given, following Bailer and Smith 1994):

Table 4: Bailer and Smith (1994) data

Dose (ppm)	Number at risk	Number of tumor-bearing animals
0	50	0
6.25	50	0
20	50	0
62.5	49	1
200	50	21

A 3-stage Weibull model (see Case 3.0.3 in Section 3) fits well to this data and two parameters, θ_0 and θ_1, are estimated to be on the boundary. Bailer and Smith (1994) considered several approaches to constructing an upper 95th confidence limit on Extra Risk. Their results for the confidence limit for extra risk at three doses are summarized in Table 5:

Table 5: Bailer and Smith (1994) results

Dose ppm	Likelihood	Nonparametric bootstrap	Parametric bootstrap
2.00	1.404E-3	5.167E-5	5.168E-5
0.20	1.405E-4	5.154E-7	5.156E-7
0.02	1.405E-5	5.153E-9	5.155E-9

The likelihood upper bound in the Table 5 is calculated according to a methodology (cited in Bailer and Smith 1994) based on Crump *et al.* (1977): it is the profile likelihood bound on a linear term of the multistage model using χ_1^2 as an asymptotic distribution. That is the situation considered in Self and Liang (1987), Case 6: one parameter of interest and one nuisance parameter are on their boundaries. In this case, the correct asymptotic distribution is a 50:50 mixture of χ_1^2 and χ_2^2.

We calculated 90 % confidence intervals for the extra risk using asymptotic normality (Wald) and also using asymptotic distribution derived according to Self and Liang (1987). The confidence intervals for Extra Risk at three doses are shown in Table 6:

Table 6: Wald and Self and Liang (1987) confidence intervals

Dose ppm	Wald Confidence Interval	Self-Liang Confidence Interval
2.00	(0; 4.476E-5)	(2.143E-6; 4.471E-5)
0.20	(0; 5.027E-7)	(1.804E-8; 4.982E-7)
0.02	(0; 1.060E-8)	(1.770E-10; 1.063E-8)

It is interesting to note that the upper bounds of both Wald and Self--Liang confidence intervals agree well with each other and with both nonparametric and parametric bootstrap confidence bounds in Table 5. However, for all 3 doses, the Wald 90% confidence interval contains 0, but confidence interval calculated according to Self and Liang (1987) does not. Also, the Self-Liang interval is shorter than Wald's.

5. Concluding Remarks

There have been long-standing questions about the accuracy and validity of various asymptotic confidence interval methods applied to risk estimates from dose response models. Very recently, Molenberghs and Verbeke (2007) sketched a framework for testing in constrained parameter spaces and Nitcheva *et al.* (2007) explored Wald tests for the multistage model using simulation when some of the parameters are on the boundary. Our report develops a theoretical approach, based on Self and Liang (1987), to resolve some of the remaining issues when parameters of the dose-response model are on the boundary but the parameter of interest is not. It also provides theoretical results, implementing Self and Liang's methodology, for the case of parameters, risk estimates, and the benchmark dose for the multistage dose response model, often employed in cancer risk assessments.

Numerical results for various scenarios involving the multistage model demonstrate substantial differences between several confidence interval approaches. The Self-Liang method improves upon Wald intervals for all the parameters and performs comparably with the likelihood ratio test intervals, the latter being slightly shorter. However, programming the profile likelihood method requires complicated routines for nonlinear optimization with inequality constraints. In contrast, Self and Liang's improvement on the Wald method can be programmed very easily, as it involves only inversion of the information matrix and simulation of a multivariate normal distribution, both operations being part of many standard software packages and not requiring much computer time.

This report investigated coverage of two-sided intervals only. A next step would be to examine one-sided intervals when some parameters are on the boundary. Initial investigations show that, unlike two-sided intervals, coverage of one-sided intervals can be far from nominal when some parameters of the multistage model are on the boundary. Another useful applied project would be to evaluate coverage of profile likelihood, Self-Liang and Wald intervals for the multistage model using various sample sizes.

Acknowledgment

Appendix A.

Here we provide a proof of Theorem 2.1. From (23), by a direct computation, we get

$$
\begin{aligned}
G(w) \;=\;& P[W \le w] \\
=\;& P[(Z_1^2 \le w)]I_{(Z_2>0,\, Z_{2.1}>0)} \\
+\;& P[(Z_1^2 - Z_2^2 \le w)]I_{(Z_2<0,\, Z_{2.1}>0)} \qquad\text{(A.1)}\\
+\;& P[(Z_{1.2}^2 \le w(1-\rho^2))]I_{(Z_2<0,\, Z_{2.1}<0)} \\
+\;& P[(Z_1^2 + \frac{Z_{2.1}^2}{1-\rho^2} \le w)]I_{(Z_2>0,\, Z_{2.1}<0)}.
\end{aligned}
$$

We now apply the method of transformation from $(Z_1, Z_2) \to (Z_{2.1} = U, Z_2 = V)$ with $(U,V) \sim N[(0,0), var(U) = 1-\rho^2, var(V) = 1, cov(U,V) = 1-\rho^2]$. Obviously, $Z_1 = (V-U)/\rho$ and $Z_{1.2} = (V-U)/\rho - \rho V$. Moreover, $V \mid U = u \sim N[u, \rho^2]$, implying $(V-U)/\rho \sim N[0,1]$. This yields the various terms in (A.1) as follows.

$$
\begin{aligned}
Term\,1 \;=\;& P[\{(V-U)^2 \le w\rho^2\}I(U>0,V>0)] \\
=\;& P\Big[\{U - \rho\sqrt{w} \le V \le U + \rho\sqrt{w}\}I(U>0,V>0)\Big] \\
=\;& \int_0^{\rho\sqrt{w}}\Big[\int_0^{u+\rho\sqrt{w}} f(v\mid u)\,dv\Big]f(u)\,du \\
+\;& \int_{\rho\sqrt{w}}^{\infty}\Big[\int_{u-\sqrt{w}}^{u+\sqrt{w}} f(v\mid u)\,dv\Big]f(u)\,du \qquad\text{(A.2)}\\
=\;& \int_0^{\frac{\rho\sqrt{w}}{\sqrt{1-\rho^2}}}\Big[\int_{-\frac{u\sqrt{1-\rho^2}}{\rho}}^{\sqrt{w}} N(0,1)\,dx\Big]N(0,1)\,du \\
+\;& \Big[\int_{-\sqrt{w}}^{\sqrt{w}} N(0,1)\,dx\Big]\Big[\int_{\frac{\rho\sqrt{w}}{\sqrt{(1-\rho^2)}}}^{\infty} N(0,1)\,dx\Big].
\end{aligned}
$$

$$
\begin{aligned}
Term\,2 \;=\;& P\Big[\{(V-U)^2 \le \rho^2(w+V^2)\}\cdot I(U>0,V<0)\Big] \\
=\;& P\Big[(U \le V + \rho\sqrt{w+V^2})\cdot I(V<0, V+\rho\sqrt{w+V^2}>0)\Big] \qquad\text{(A.3)}\\
=\;& \int_{-\sqrt{\frac{w\rho^2}{1-\rho^2}}}^{0} P\Big[-\frac{v(1-\rho^2)}{\rho} \le N(0,1) \le v\rho + \sqrt{w+v^2}\Big]N(0,1)\,dv.
\end{aligned}
$$

$$Term\ 3 = P\left[\{(V - U - \rho^2 V)^2 \le w\rho^2(1-\rho^2)\} \cdot I(U < 0, V < 0)\right] \tag{A.4}$$

$$= P\left[\left\{V(1-\rho^2) - \sqrt{w\rho^2(1-\rho^2)} \le U \le V(1-\rho^2) + \sqrt{w\rho^2(1-\rho^2)}\right\} \cdot I(U < 0, V < 0)\right]$$

$$= P\left[\left\{V(1-\rho^2) - \sqrt{w\rho^2(1-\rho^2)} \le U \le V(1-\rho^2) + \sqrt{w\rho^2(1-\rho^2)}\right\}, V < -\sqrt{\frac{w\rho^2}{(1-\rho^2)}}\right]$$

$$+ P\left[V(1-\rho^2) - \sqrt{w\rho^2(1-\rho^2)} \le U \le 0, -\sqrt{\frac{w\rho^2}{(1-\rho^2)}} < V < 0\right]$$

$$= P\left[-\sqrt{w(1-\rho^2)} < N(0,1) < \sqrt{w(1-\rho^2)}\right] \cdot P\left[N(0,1) < -\sqrt{\frac{w\rho^2}{(1-\rho^2)}}\right]$$

$$+ P\left[\left\{-\sqrt{w(1-\rho^2)} < N(0,1) < -V(1-\rho^2)/\rho\right\}, -\sqrt{\frac{w\rho^2}{(1-\rho^2)}} < V < 0\right].$$

$$
\begin{aligned}
Term\ 4 \ &= \ P\left[\{(V-U)^2(1-\rho^2) + U^2\rho^2 \le w\rho^2(1-\rho^2)\} \cdot I(U < 0, V > 0)\right] \\
&= \ P\left[\{(U - V(1-\rho^2))^2 \le (w - V^2)\rho^2(1-\rho^2)\}, I(U < 0, V > 0)\right] \\
&= \ P\left[V(1-\rho^2) - \sqrt{(w-V^2)\rho^2(1-\rho^2)} \le U < 0, 0 < V < \rho\sqrt{w}\right] \\
&= \ \int_0^{\rho\sqrt{w}} P\left[-\sqrt{(w-v^2)\rho^2(1-\rho^2)} \le N(0,1) \le -v(1-\rho^2)/\rho\right] \cdot N(0,1) dv.
\end{aligned}
\tag{A.5}
$$

The theorem follows upon combining and simplifying these four terms.

References

1. D. Krewski and J. van Ryzin, *Dose-response models for quantal response toxicity data*, in *Statistics and Related Topics: Proc. Inter. Symp. on statistics and related topics (Ottawa, Ontario, May 5--7, 1980)* (North--Holland, Amsterdam, 1981).

2. A.F. Filipsson, S. Sand, J. Nilsson and K. Victorin, *Critical Reviews in Toxicology*, **33**, 505 (2003).

3. U.S. EPA (U.S. Environmental Protection Agency), *Help Manual for Benchmark Dose Software. Version 1.4*, EPA 600/R-00/014F, (2006).

4. D.L. Eaton and C.D. Klaassen, *Principles of Toxicology*, in *Casarett and Doull's Toxicology: The Basic Science of Poisons (6th edn.)* (McGraw--Hill, New York, 2001).

5. K.S. Crump, *Fundamental and Applied Toxicology*, **4**, 845 (1984).

6. F. Parham and C. Portier, *Benchmark Dose Approach*, in *Recent Advances in Quantitative Methods In Cancer and Human Health Risk Assessment* eds. L. Edler and C.P. Kitsos, (John Wiley & Sons, Chichester, 2005).

7. K.S. Crump and R. Howe, *A review of methods for calculating statistical confidence limits in low dose extrapolation*, in *Toxicological Risk Assessment* eds. D.B. Clayson, D. Krewski and I. Munro, (CRC Press, Boca Raton, 1985).

8. C.R. Rao, *Linear Statistical Inference and Its Applications*, 2nd edn. (John Wiley and Sons, New York, 1973).

9. D.R. Cox and D.V. Hinkley, *Theoretical Statistics*, (Chapman and Hall, London, 1974).

10. H. Chernoff, *Annals of Mathematical Statistics*, **25**, 573 (1954).

11. P.I. Feder, *Annals of Mathematical Statistics*, **39**, 2044 (1968).

12. S.G. Self and K.--Y. Liang, *Journal of the American Statistical Association*, **82**, 605 (1987).

13. G. Molenberghs and G. Verbeke, *American Statistician*, **61**, 22 (2007).

14. A.J. Bailer and R.J. Smith, *Risk Analysis*, **14**, 1001 (1994).

15. M. Moerbeek, A.H. Piersma and W. Slob, *Risk Analysis*, **24**, 31 (2004).

16. D.K. Nitcheva, W.W. Piegorsch and R.W. West, *Regulatory Toxicology and Pharmacology*, **48**, 135 (2007).

17. K.S. Crump, H.A. Guess and K.L. Deal, *Biometrics*, **33**, 437 (1977).

18. P.A.P. Moran, *Proceedings of the Cambridge Philosophical Society*, **70**, 441 (1971).

19. D. Chant, *Biometrika*, **61**, 291 (1974).

20. L. Fahrmeir and H. Kaufmann, *Annals of Statistics*, **13**, 342 (1985).

21. C.J. Geyer, *Annals of Statistics*, **22**, 1993 (1994).

22. H.T.V. Vu and S. Zhou, *Annals of Statistics*, **25**, 897 (1997).

23. A. Shapiro, *Biometrika*, **72**, 133 (1985).

24. G. Cardano, *Ars Magna*, (1501--1576).

25. U.S. EPA (U.S. Environmental Protection Agency), *Benchmark Dose Technical Guidance Document. External Review Draft*, EPA/630/R-00/001, (2000).

26. NTP (National Toxicology Program), *Toxicology and Carcinogenesis Studies of 1,3-Butadiene (CAS No. 106-99-0) in B6C3F1 Mice (Inhalation Studies)*, NTP TR 434, NIH Publ. No. 93-3165 (1993).

27. Leonid Kopylev and Bimal Sinha. Sankhya, Series B, **73**, 20-41 (2011).

---

[i]The views expressed in this report are solely those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency

[ii]The views expressed in this report are solely those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency