CONFIDENCE REGIONS FOR THE EFFECTIVE DOSE WITH COMBINATIONS OF AGENTS

HANNA K. JANKOWSKI, XIANG JI AND LARISSA I. STANBERRY

ABSTRACT. The effective dose is the dose or amount of drug required to produce a therapeutic response in a fixed proportion of the subjects. When only one drug is considered, the problem is a univariate one and has been well-studied. However, in the multidimensional setting, i.e., in the presence of combinations of agents, estimation of the effective dose becomes more difficult. This article considers the case where the multidimensional effective dose is estimated using a plug-in logistic regression. We discuss consistency of such estimators, and focus on the problem of simultaneous confidence regions. Existing methodology for 95% simultaneous confidence regions yields empirical coverage probabilities with values often larger than 99%. We introduce an empirical set quantile algorithm which outperforms existing methodology. Through simulation, we show that our algorithm gives 95% confidence regions which have better empirical coverage then the previous method for moderate to large sample sizes. The new algorithm is illustrated on a cytotoxicity study on the effect of two toxins in the leukaemia cell line HL-60 and a decompression sickness study of the effects of the duration and depth of the dive.

1. INTRODUCTION

In the analysis of biological assays, one is often interested in the covariate, or combination of covariates, which yields a specific response. For example, consider the situation when the efficacy of a drug is being investigated. The effective dose (ED) is then the dose or amount of drug required to produce a therapeutic response in a desired proportion of the population under study. When this response is binary, a logistic linear regression model is a popular choice. That is, let p denote the probability of a positive response, and let x denote the drug dosage. Then the model is

$$\log(p/(1-p)) \;=\; \beta_0 + \beta_1 x,$$

and the dosage associated with a 50% response rate (ED_{50}) is the point $-\beta_0/\beta_1$. When a combination of drugs and/or covariates is being studied, the model becomes

$$\log(p/(1-p)) \;=\; \beta_0 + \sum_{i=1}^{k} \beta_i x_i.$$

If only a single x_i, say x_1, represents a drug and the remaining x_i represent covariates such as age or weight, then the median effective dose is $\mathrm{ED}_{50} = -(\beta_1)^{-1}(\beta_0 + \sum_{i=2}^{k} \beta_i x_i)$ when the patient-specific covariates are held fixed. Alternatively, multiple x_i can represent drug levels, and one is interested in the combinations of drugs or other agents required to yield a

response. Such instances arise, for example, in Skarin et al. (1983) where lymphoma treatments are considered or in Lang et al. (1980). Another example, considered extensively in Li et al. (2008b,a, 2010); Li and Wong (2011), is that of analysing the risk of decompression sickness among deep sea divers based on both the duration and pressure of the dive. In this setting, the median effective dose is the set

$$\mathrm{ED}_{50} = \left\{ (x_1, x_2, \ldots, x_p) : \beta_0 + \textstyle\sum_{i=1}^{k} \beta_i x_i = 0 \right\}.$$

Li et al. (2008b) refer to this as the *multidimensional* effective dose. In the single drug setting (with or without covariates), the quantity of interest is real-valued, and therefore much easier to handle. However, for combinations of drugs, the object under study is considerably more complicated.

Many interesting statistical problems are associated with the study of the multidimensional effective dose. Here, we focus on one specific aspect of the extensive problem at hand. We assume that the effective dose is estimated using a plug-in estimator from a logisitic regression. We focus on this plug-in estimator as it is likely to be used in practice. We provide conditions for consistency of the estimator and we study *simultaneous* confidence regions, or supersets, for the true effective dose. In the univariate setting, the variability of the estimator can be quantified by estimation of its standard deviation. However, when combinations of agents are being considered, we note that a confidence region provides a natural and visually appealing way of quantifying the variability of the estimator.

To our best knowledge, the first study of such confidence regions was done in Carter et al. (1986). The proposed solution is to create large sample confidence regions by inverting Scheffé's bounds for simultaneous confidence intervals (Scheffé, 1953). As noted by Carter et al. (1986), this method is quite conservative. Further study of this method was carried out in Li et al. (2008b), who consider the estimation of both (1) a conditional single-dimensional effective dose in the presence of covariates and (2) an unconditional multidimensional effective dose in the linear logistic model. The conditional single-dimensional effective dose was also used by Chen (2007) in the analysis of a dose-time response model. The extensive simulations of Li et al. (2008b) show how conservative the Scheffé-inversion method is: Empirical coverage probabilities of 95% confidence regions vary from 98% to 100% in the two-dimensional linear model.

In this work we consider several different existing approaches to the problem of finding a confidence superset. We also introduce a novel empirical set quantile algorithm, and use it to compute a confidence region for the effective dose. The various methods are described in Section 4. We study the empirical coverage of the confidence regions through extensive simulations (see Section 5 as well as the Appendix). Our choice of designs for the simulations is similar to that of Li et al. (2008b), in particular to allow for direct comparison. We go beyond the cases studied in Li et al. (2008b) in that we study more than just the linear regression setting. Notably, the over coverage of the Carter method is only increased for these settings, and therefore it is important to consider such cases.

The methodology is illustrated on both theoretical (Section 4.6) and real-world examples (Section 6).

2. Notation and Assumptions

Let $\mathcal{D} \subset \mathbb{R}^d$ denote the possible domain of the covariates, which we assume to be compact. We assume the a binary response Y is observed and let $f(x) = E[Y|X = x]$. We are interested in sets of the form

$$\mathrm{ED}^{+}_{100p} = \{x \in \mathcal{D} : f(x) \geq p\} \qquad \text{and} \qquad \mathrm{ED}_{100p} = \{x \in \mathcal{D} : f(x) = p\}$$

where $x = \{x_1, \ldots, x_d\}$ denotes the observed covariates. One may also be interested in $\mathrm{ED}^{-}_{100p} = \{x \in \mathcal{D} : f(x) \leq p\}$, but as this may be studied by considering $g(x) = -f(x)$, we consider only ED^{+}_{100p} and ED_{100p} in what follows. We assume that the data can be modelled as a logistic regression with

$$\log\left(\frac{f(x)}{1-f(x)}\right) \;=\; \beta_0 x_0^* + \beta_1 x_1^* + \ldots + \beta_k x_k^* = \beta^T x^*. \tag{2.1}$$

We use the notation $x^* = (x_0^*, x_1^*, \ldots, x_k^*)$ to denote the function of the covariates $x^*(x) : \mathbb{R}^d \mapsto \mathbb{R}^{k+1}$, where $k + 1 \geq d$. For example, if $d = 2$ and the right-hand side of (2.1) is $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2$ then $x^*(x) = (1, x_1, x_2, x_2^2)$. We require that the function $x^*(x)$ is continuous on the domain $\mathcal{D}$.

Let X denote the design matrix of size $\kappa \times (k + 1)$. In our simulations, we will use a balanced design with m replicates of each combination of covariates, for an overall sample size of $n = m\kappa$. Under this notation, the estimator $\widehat{\beta}_n$ is asymptotically normal with mean β and variance given by Σ/m where $\Sigma = \left(X^T \mathrm{diag}\{p_i(1-p_i)\}X\right)^{-1}$, and $p_i = E[Y|X = x_i]$ for $i = 1, \ldots, \kappa$.

Next, let $\eta(u) = \log(u/(1-u))$, and note that this is an increasing function of u. Therefore, we may re-write

$$\mathrm{ED}^{+}_{100p} = \left\{x \in \mathcal{D} : \beta^T x^* \geq \eta(p)\right\} \quad \text{and} \quad \mathrm{ED}_{100p} = \left\{x \in \mathcal{D} : \beta^T x^* = \eta(p)\right\}.$$

We estimate these using the plug-in estimators

$$\widehat{\mathrm{ED}}^{+}_{100p} = \left\{x \in \mathcal{D} : \widehat{\beta}_n^T x^* \geq \eta(p)\right\} \quad \text{and} \quad \widehat{\mathrm{ED}}_{100p} = \left\{x \in \mathcal{D} : \widehat{\beta}_n^T x^* = \eta(p)\right\},$$

where $\widehat{\beta}_n$ is the maximum likelihood estimator of β. As long as $x^*(x)$ is a continuous function of x, all of the sets considered above are closed. However, note also that it is not necessarily the case that $\partial \mathrm{ED}^{+}_{p} = \mathrm{ED}_p$, where ∂A denotes the boundary of the set A.

We denote Euclidean distance for $x, y \in \mathbb{R}^d$ as $|x - y|$. Then, for a set $A \subset \mathbb{R}^d$ and $\varepsilon > 0$ we define the dilation of A as $A^{\varepsilon} = \{x \in \mathbb{R}^d : |x - y| \leq \varepsilon, \text{ for some } y \in A\}$. Finally, we define the Hausdorff distance between two sets

$$\rho(A, B) \;=\; \inf\{\varepsilon > 0 : A \subset B^{\varepsilon},\, B \subset A^{\varepsilon}\}.$$

3. CONSISTENCY

To our best knowledge, the works studying multidimensional effective dose estimation focus on confidence regions and do not discuss consistency (Carter et al., 1986; Li et al., 2008b,a; Li and Wong, 2011). Here, we fill this gap, by extending the work of Molchanov (1998) and Cuevas et al. (2006) on level sets to this setting. A very similar problem was recently studied in Jankowski and Stanberry (2011), and the following result follows directly from the proof of Jankowski and Stanberry (2011, Theorem 3.1) as well as consistency of the maximum likelihood estimator $\widehat{\beta}_n$.

Theorem 3.1. *Let β denote the value of the parameters in model* (2.1)*, and assume that $x^*(x)$ is continuous. If $\widehat{\beta}_n \to \beta$ almost surely, then, for all compact K, $\rho(\widehat{\mathrm{ED}}^+_{100p} \cap K, \mathrm{ED}^+_{100p} \cap K) \to 0$ almost surely if*

$$\overline{\{x \in \mathcal{D} : \beta^T x^* > \eta(p)\}} \;=\; \mathrm{ED}^+_{100p} \tag{3.2}$$

holds. Furthermore, $\rho(\widehat{\mathrm{ED}}_{100p}, \mathrm{ED}_{100p}) \to 0$ almost surely if both (3.2) *and*

$$\overline{\{x \in \mathcal{D} : \beta^T x^* < \eta(p)\}} \;=\; \{x \in \mathcal{D} : \beta^T x^* \leq \eta(p)\} \tag{3.3}$$

hold.

The two conditions (3.2) and (3.3) can be re-stated mathematically in different ways; for example, (3.2) is equivalent to ED^+_{100p} being regularly closed. In terms of the response surface, condition (3.2) requires that $f(x) = \beta^T x^*$ have no local maxima on ED_{100p} while condition (3.3) requires that $f(x) = \beta^T x^*$ have no local minima on ED_{100p}. Thus, in the linear model, we are guaranteed to have consistency if at least one β_i, $i = 2, \ldots, k$ is non-zero. For other models, one could perform a heuristic check of critical points of $f(x)$, if x^* is differentiable as a function of x. As an example, consider the model $f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$. The determinant of the Hessian matrix is $-\beta_3^2 < 0$. Thus, if β_3 is non-zero, any critical points of $f(x)$ are saddle points and therefore no local minima/maxima exist on ED_{100p}.

4. CONFIDENCE REGIONS: DESCRIPTION OF METHODS

We now turn to the problem of calculating confidence regions for ED_{100p} and ED^+_{100p}, and begin by describing various approaches. A discussion and initial comparison of the methods follows the definitions.

For each method, we describe how to calculate a $100(1-\alpha)\%$ confidence region for ED^+_{100p}. Confidence regions for ED^-_{100p} are formed in an analogous fashion. A $100(1-\alpha)\%$ confidence region for ED_{100p} is taken as the intersection of the $100(1-\alpha/2)\%$ confidence region for ED^+_{100p} and the $100(1-\alpha/2)\%$ confidence region for ED^-_{100p}. In what follows, we assume that the domain $\mathcal{D}$ is compact. If it is not compact, then the methods can be modified to consider the domain $\mathcal{D} \cap K$, for any compact set K.

4.1. **Supremum inversion.** By definition, $\widehat{\mathrm{ED}}_{100p}$ is the level set of the random function $\widehat{f}_n(x) = \widehat{\beta}_n^T x^*$. The variability of $\widehat{f}_n(x)$ away from its expectation is described by the fluctation field $\mathbb{Z}_n(x) = \sqrt{n}(\widehat{f}_n(x) - E[\widehat{f}_n(x)]) = \sqrt{n}(\widehat{\beta}_n - \beta)^T x^*$, or its Gaussian limit $\mathbb{Z} = \lim_n \mathbb{Z}_n$. Uniform bounds on the fluctuation field can be inverted to find uniform bounds on level sets of $E[\widehat{f}_n(x)]$, in an approach we call "supremum inversion." Although conservative by definition, it is quite natural for plug-in estimators.

The existing methodology for simultaneous confidence regions is based on Scheffé's uniform contrast bounds. As shown in Carter et al. (1986), $\sup_x \mathbb{Z}_n^2(x)/x^{*T}\Sigma x^* \leq Y_n \Rightarrow Y$, where Y is a $\chi^2(k+1)$ random variable. Let q_α denote the upper quantile of Y, a value such that $P(Y > q_\alpha) = 2\alpha$. The confidence region for ED_{100p}^+ is then defined as

$$\mathrm{CR}_{\mathrm{SCH},1-\alpha}^+ \;=\; \left\{x : \widehat{f}_n(x) \geq \eta(p) - \sqrt{q_\alpha\, x^{*T}\Sigma x^*/m}\right\}.$$

In practice, the covariance matrix Σ is estimated using standard methods. It is possible that the confidence region as defined above is empty, but this occurs rarely for larger sample sizes. As shown in Carter et al. (1986), the probability that $\mathrm{CR}_{\mathrm{SCH},1-\alpha}^+$ covers ED_p^+ is at least $100(1-\alpha)\%$.

4.2. **Central limit theorem for level sets.** Another possible approach is to consider a central limit theorem for the sets $\widehat{\mathrm{ED}}_{100p}$ themselves. Such a result, for general level sets, was given in Molchanov (1998). We note that the result is valid for more general settings with an appropriately modified definition for $L(x)$ (Molchanov, 1998).

Theorem 4.1. *(Molchanov (1998)) Let K denote a compact subset of $\mathbb{R}^d$ with C^1 boundary, and let $n(x)$ denote the unit outer normal vector to K at $x \in \partial K$. Then, for $\widehat{f}_n(x) = \widehat{\beta}_n^T x^*$,*

$$\sqrt{n}\;\rho\left(\widehat{\mathrm{ED}}_{100p}^+ \cap K, \mathrm{ED}_{100p}^+ \cap K\right) \Rightarrow \sup_{x \in \mathrm{ED}_{100p} \cap K} |\mathbb{Z}(x)/L(x)|\,,$$

where $L(x)$ is given by

$$|L(x)| \;=\; \begin{array}{ll} |\beta^T \nabla x^*| & x \in K^0 \\ |\beta^T \nabla x^*| & x \in \partial K \text{ and } \theta(x) \geq \pi/2 \\ |\beta^T \nabla x^*| \sin(\theta) & x \in \partial K \text{ and } 0 \leq \theta(x) < \pi/2, \end{array}$$

for $\theta(x)$ equal to the angle between $\beta^T \nabla x^*$ and $n(x)$. Here, ∇x^* denotes the gradient of $x^*(x)$.

If $f(x) = \beta_0 + \beta_1 x$, the result becomes

$$\sqrt{n}\left(-\frac{\widehat{\beta}_{n,0}}{\widehat{\beta}_{n,1}} + \frac{\beta_0}{\beta_1}\right) \;\Rightarrow\; \frac{Z}{\beta_0},$$

where Z is a Gaussian random variable, at $p = 0.50$, as expected. The more general result is still not surprising. The term $L(x)$ in the denominator tell us that the less steep the slope of the function $\beta^T x^*$ at the boundary, the more variable $\widehat{\mathrm{ED}}_{100p}$ will be.

Let q_α denote the α–level upper quantile of the random variable $\sup_{x \in \mathrm{ED}_{100p}} |\mathbb{Z}(x)/L(x)|$. Then a natural definition for the the $100(1-\alpha)\%$ confidence region for ED^+_{100p} is

$$\mathrm{CR}^+_{\mathrm{CLT},1-\alpha} \;=\; \left(\widehat{\mathrm{ED}}^+_{100p}\right)^{q_\alpha},$$

noting that $\widehat{\mathrm{ED}}^+_{100p}$ is closed.

4.3. **Quantiles of random sets.** An appealing method of forming confidence intervals in the real-valued setting is to use the quantiles of a statistic or pivotal quantity, and it seems natural to develop this notion for the effective dose, whose estimator is a random closed set. The space of closed sets is non-linear, and therefore there is no single natural way to define expectation or quantile. The following definition for the quantile of a random set was introduced in Molchanov (1990).

The distribution of a random closed set (RCS), $\boldsymbol{A}$, is determined by the functional $T(K) = P(\boldsymbol{A} \cap K \neq \emptyset)$ for every compact set $K \subset \mathbb{R}^d$ (see, for example, Molchanov (2005)). Let $\mathcal{K}$ denote the class of all compact sets on $\mathbb{R}^d$, and let $\mathcal{M} \subset \mathcal{K}$ be a subcollection of sets which is closed in the Hausdorff metric. In Molchanov (1990), the q–quantile of the RCS $\boldsymbol{A}$ is defined as the set $M_q = \cup\{K \in \mathcal{M} : T(K) < q\}$. As a concrete example let Θ be a real-valued random variable with distribution function F, and let $\boldsymbol{A} = [\Theta, \infty)$. Then, taking the class $\mathcal{M} = \{\{x\}, x \in \mathbb{R}\}$ yields $M_q = (-\infty, \theta_q)$, where $\theta_q = \sup\{x : F(x) < q\}$. Therefore, in some sense, the RCS quantile definition coincides with the quantile definition of a real-valued random variable. We consider two different choices of the class $\mathcal{M}$. Our selection is motivated, in large part, by ease of computability.

(1) $\mathcal{M}_1 = \left\{\{x\}; x \in \mathbb{R}^d\right\}$.
(2) $\mathcal{M}_2 = \left\{\{b^T x = a\}; a \in \mathbb{R}, b, x \in \mathbb{R}^d\right\}$

Some care needs to be taken before applying this definition to create confidence regions, as the above example illustrates that the choice of covering set M^c_{1-q} may be more appropriate for our purposes (essentially, we seek a *lower* confidence bound).

Let us first consider the quantile $M^c_\alpha(\mathcal{M}_1)$ of the set $\widehat{\mathrm{ED}}^+_{100p}$. We can calculate this exactly, assuming that $\widehat{\beta}_n$ behaves according to its asymptotic normal distribution. Let z_α denote a value such that $P(Z > z_\alpha) = \alpha$, where Z denotes a standard normal random variable. Then

$$\begin{aligned} M^c_\alpha(\mathcal{M}_1) &= \left\{x \in \mathcal{D} : p\left(\widehat{\mathrm{ED}}^+_{100p} \cap \{x\} \neq \emptyset\right) < \alpha\right\}^c \\ &= \left\{x \in \mathcal{D} : p\left(\widehat{\beta}_n^T x^* \leq \eta(p)\right) \leq 1-\alpha\right\} \\ &= \left\{x \in \mathcal{D} : \beta^T x^* \geq \eta(p) - z_\alpha \sqrt{x^{*T} \Sigma x^*/m}\right\}. \end{aligned}$$

This quantile depends on β, which is not surprising as $\widehat{\beta}_n$ is centred around β. To define the confidence region, we therefore, "re-centre" the distribution around $\widehat{\beta}_n$. That is, we define

the confidence region based on the class $\mathcal{M}_1$ to be

$$\mathrm{CR}^{+}_{\mathrm{Q1},1-\alpha} \;=\; \left\{x \in \mathcal{D} : \widehat{\beta}_n^T x^* \geq \eta(p) - z_\alpha \sqrt{x^{*T}\Sigma x^*/m}\right\}. \tag{4.4}$$

The difference between this definition and $\mathrm{CR}^{+}_{\mathrm{SCH},1-\alpha}$ lies entirely in the quantile z_α vs $\sqrt{q_\alpha}$. Clearly, $\mathrm{CR}^{+}_{\mathrm{Q1},1-\alpha}$ is a *local* (i.e. pointwise) confidence region and not a global one. We therefore expect that simultaneous coverage rates for these quantiles will be smaller than $1-\alpha$. We also note that these regions bear much resemblance to the conditional confidence regions considered in Li et al. (2008b). By definition, this region is the collection of all points such that each point is in $\widehat{\mathrm{ED}}^{+}_{100p}$ with at least a probability of α, and as such, it coincides with the definition of an α excursion set used in the definition of the Vorob'ev quantile (Molchanov, 1990, Section 2.2).

It seems natural to next consider a different class for $\mathcal{M}$, and the class of all lines is one simple alternative. Write $\ell_{a,b} = \{x \in \mathcal{D} : b^T x = a\}$ for some $a \in \mathbb{R}, b \in \mathbb{R}^d$, so that $\mathcal{M}_1 = \{\ell_{a,b}\}$. Then

$$\begin{aligned} M^c_\alpha(\mathcal{M}_2) &= \left\{\cup\left\{\ell_{a,b} : p\left(\widehat{\mathrm{ED}}^{+}_{100p} \cap \ell_{a,b} \neq \emptyset\right) < \alpha\right\}\right\}^c \\ &= \left\{\cup\left\{\ell_{a,b} : p\left(\widehat{\beta}_n^T x^*(x) \geq \eta(p)\ \ \exists\, x \in \ell_{a,b}\right) < \alpha\right\}\right\}^c. \end{aligned}$$

In a manner similar to that for the class $\mathcal{M}_1$, we define the confidence region as

$$\mathrm{CR}^{+}_{\mathrm{Q2},1-\alpha} \;=\; \left\{\cup\left\{\ell_{a,b} : p\left(\widehat{\beta}_n^T x^* + Z^T x^* \leq \eta(p)\ \ \exists\, x \in \ell_{a,b}\right) < \alpha\right\}\right\}^c,$$

where Z is a multivariate normal random variable mean zero and variance Σ/m, and $\widehat{\beta}_n$ is *held fixed at its observed value*.

4.4. **New method.** Suppose that we observe $A_1, \ldots, A_B$ sample sets. Then a sample quantile for these sets can be calculated as follows.

Empirical set quantile algorithm

Suppose that we observe $A_1, \ldots, A_B$ closed sets.

(1) Let $K_B = \cap_{i=1}^B A_i$.
(2) Let $\xi_{i,B} = \rho(A_i, K_B)$, and let $\gamma_{\alpha,B}$ denote the empirical α quantile of the $\xi_{i,B}$ values. That is, let $\xi_{(i),B}$ denote the ordered values of $\xi_{i,B}$. Then $\gamma_{\alpha,B}$ is the value such that

$$\gamma_{\alpha,B} \;=\; \max\{\xi_{(i),B} : \#\{j : \xi_{(j),B} \leq \xi_{(i),B}\} < \alpha B\}.$$

(3) Calculate the quantile of the collection $\{A_1, \ldots, A_B\}$ to be $\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) \leq \gamma_{\alpha,B}\}$.

As an example, consider the simple setting with $A_i = [\theta_i, \infty)$, $i = 1, \ldots, B$. Let $\theta_{(B)} = \max(\theta_1, \ldots, \theta_b)$. Then $K_B = [\theta_{(B)}, \infty)$, $\xi_{i,B} = \theta_{(B)} - \theta_i$, and $\gamma_{\alpha,B} = \theta_{(B)} - \widetilde{\gamma}_{\alpha,B}$, where $\widetilde{\gamma}_{\alpha,B}$ denotes the empirical α quantile of θ_i, i.e., $|\{i : \theta_i < \widetilde{\gamma}_{\alpha,B}\}|/B \approx \alpha$. Finally, the empirical quantile becomes $\cup\{A_i : \theta_{(B)} - \theta_i < \theta_{(B)} - \widetilde{\gamma}_{\alpha,B}\} = \cup\{A_i : \theta_i > \widetilde{\gamma}_{\alpha,B}\} = [\widetilde{\gamma}_{\alpha,B}, \infty)$. For ease of presentation, we have assumed here that $\widetilde{\gamma}_{\alpha,B}$ is one of the observed values of $\theta_1, \ldots, \theta_B$. Therefore, choosing α=95% would yield an approximate 95% covering set for the sets $A_i = [\theta_i, \infty)$. Clearly, the quantiles depend on the centre of the observed data, $K_B = \cap_{i=1}^{B} A_i$, and the relative location of K_B greatly influences the resulting quantile superset.

Remark 4.2. *In the algorithm, if $K_B = \emptyset$, then the resulting quantile is the empty set, and therefore uninformative. For the effective dose application, this becomes an issue for smaller sample sizes. In most cases, this happens because one of the sampled sets A_i is empty. We considered several solutions, and the optimal of these was to simply remove all empty A_i from the algorithm apriori.*

Proposition 4.3. *Let $A_1, \ldots, A_B$ be a collection of nonempty subsets of $\mathbb{R}^d$.*

(1) *The proportion of sets $A_1, \ldots, A_B$ which is contained in $\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma_\alpha\}$ is at least $(\lfloor \alpha\, B \rfloor)/B$.*
(2) *Consider $\gamma_1 \leq \gamma_2$. Then $\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma_1\} \subseteq \cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma_2\}$.*
(3) *Fix a rigid motion $g \in E^+(d)$ and let $C_i = g(A_i)$. Then*

$$\begin{aligned} g\left(\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_n) < \gamma\}\right) \\ = \ \cup\{C_i, i = 1, \ldots, B : \rho(C_i, g(K_B)) < \gamma\}. \end{aligned}$$

(4) *Fix $\alpha > 0$ and let $C_i = \alpha A_i$. Then $\cup\{C_i, i = 1, \ldots, B : \rho(C_i, \alpha K_B) < \alpha\gamma\} = \alpha\left\{\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma\}\right\}$.*

It follows therefore that the new definition has some desirable properties. It is invariant under rotations, and satisfies a natural scaling property. Most importantly, the quantiles create a covering set which contains a desirable proportion of the observed data. The proof of Proposition 4.3 appears in the Appendix.

Using the new method, we estimate a confidence region for ED^+_{100p} by using a parametric bootstrap approach. That is, let $(\widehat{\mathrm{ED}}^+_{100p})^*_i,\ i = 1, \ldots, n$ denote observed values of

$$(\widehat{\mathrm{ED}}^+_{100p})^* = \{x \in \mathcal{D} : \widehat{\beta}_n^T x^* + Z^T x^* \geq \eta(p)\},$$

where (as in the definition for $\mathrm{CR}^+_{\mathrm{Q2},1-\alpha}$) $\widehat{\beta}_n$ is held fixed while Z is a normal random variable with mean zero and variance Σ/m. In practice, if there is much variability in the response surface $\widehat{\beta}_n^T x^*$, then some of the observed sets $\widehat{\mathrm{ED}}^+_{100p,i}$ may be empty. In this case, we modify the algorithm by considering only those sets which are nonempty (cf. Remark 4.2). The $1 - \alpha$ quantile of the observed (nonempty) sets is then the estimated confidence region $\mathrm{CR}^+_{1-\alpha}$.

4.5. **Computation.** The regions $\mathrm{CR}_{\mathrm{SCH},1-\alpha}$ and $\mathrm{CR}_{\mathrm{Q1},1-\alpha}$ can be computed exactly if Σ is known. Similarly, if Σ is known, then the region $\mathrm{CR}_{1-\alpha}$ can be calculated using the algorithm described above. In practice, however, only its estimate, denoted here as $\widehat{\Sigma}_n$, is available and is provided in all standard statistical packages. The three regions can then be approximated by replacing Σ with its consistent estimate $\widehat{\Sigma}_n$ throughout.

To calculate the confidence region $\mathrm{CR}_{\mathrm{B},1-\alpha}$, we would need to know the function $L(x) = \beta^T \nabla x^*$ and the set ED_{100p}, both of which we would then use to obtain the quantile of $\sup_{x \in \mathrm{ED}_{100p}} |\mathbb{Z}(x)/L(x)|$. Only the distribution of $\mathbb{Z}(x)$ is known, it is equivalent to the distribution of $Z^T x^*$ where Z is multivariate normal with mean zero and variance Σ/κ. In some simulations which follow, we estimated β and Σ with $\widehat{\beta}_n$ and $\widehat{\Sigma}_n$, ED_{100p} with $\widehat{\mathrm{ED}}_{100p}$, and based on these values, we estimate quantiles of $\sup_{x \in \mathrm{ED}_{100p}} |\mathbb{Z}(x)/L(x)|$ via re-sampling.

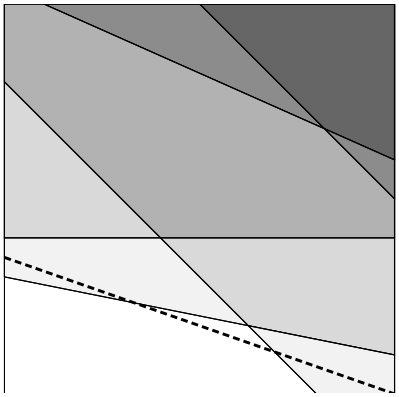

FIGURE 1. A sample of five planar sets. The darker a region the more sets intersect with this region.

In general, the distribution of a random set is difficult to obtain explicitly, and in practice, Molchanov's quantiles may also be estimated from observed data. Let $\boldsymbol{A}$ denote a random closed set and let $A_1, \ldots, A_n$ denote n IID observations of $\boldsymbol{A}$. Then, the capacity functional of $\boldsymbol{A}$, $T(K)$, may be estimated by the empirical functional $\mathbb{T}_n(K) = \sum_{i=1}^n 1_{A_i \cap K \neq \emptyset}/n$, which in turn gives the empirical quantile $\widehat{M}_{q,n}(\mathcal{M}) = \cup\{K \in \mathcal{M} : \mathbb{T}_n(K) < q\}$. Asymptotic properties of this procedure were studied in Molchanov (1990), who provides necessary and sufficient conditions for consistency, among other things. (We note that the condition $M_{p+} \subset M_p$ in Molchanov (1990, Theorem 1, English version) should actually read $M_{p+} \subset \overline{M_p}$, which can be confirmed by reading the proof or the original Russian text.) In Molchanov (1990, Corollary 2), the result is given for the case $\mathcal{M}_1$. It follows that, when $\boldsymbol{A} = \widehat{\mathrm{ED}}^+_{100p}$, the empirical quantiles with $\mathcal{M}_1$ are consistent for the true quantiles if the distribution of $\widehat{\beta}_n$ is continuous. As we have an explicit formula for $\mathrm{CR}_{\mathrm{Q1},1-\alpha}$, it is not necessary to estimate it by re-sampling. On the other hand, an explicit formula

for $\mathrm{CR}_{\mathrm{Q2},1-\alpha}$ is not as immediate. Therefore, to estimate the region $\mathrm{CR}^{+}_{\mathrm{Q2},1-\alpha}$, we have devised the following method.

First, we estimate $\mathrm{CR}^{+}_{\mathrm{Q1},1-\alpha}$ by some re-sampling method. Then $\mathrm{CR}^{+}_{\mathrm{Q2},1-\alpha} = \text{chull}\left(\mathrm{CR}^{+}_{\mathrm{Q1},1-\alpha}\right)$, where chull$(A)$ denotes convex hull of the set A. To see why this should hold consider a simple setting with $n = 5$ and suppose that we wish to calculate $\mathrm{CR}^{+}_{\mathrm{Q2},0.60}$ for these five sets. By definition, we are interested in the region so that any line contained in this region intersects no more than one of these five observed sets. Finally, the collection of lines which intersect two or more of these sets is the convex hull of $\mathrm{CR}^{+}_{\mathrm{Q1},1-\alpha}$.

It was noted in Li et al. (2008b, page 113) that their conditional univariate approach is simpler to implement than a multivariate approach ($\mathrm{CR}_{\mathrm{SCH},1-\alpha}$ in this paper). However, we do not find this to be the case, and $\mathrm{CR}_{\mathrm{SCH},1-\alpha}$ is quite easy to calculate in, for example R. All computations and plots shown in this paper were done using R and/or Matlab. Matlab's imaging toolbox contains the function bwdist, which allows for easy computation of the Hausdorff distance, required for calculation of $\mathrm{CR}_{1-\alpha}$. This function is available both for $d = 2$ and $d = 3$, but not for higher dimensions. Sample programs are available from `www.math.yorku.ca/~hkj/Software/`. In all computations, including the simulations which follow, it was necessary to discretize the underlying domain. For $d = 2$, we used 401^2 pixels, and when $d = 3$, we used 101^3 voxels.

4.6. **Comparison of Methods.** To compare the various methods, we first consider the linear parametric model $\widehat{f}_n(x) = \widehat{\beta}_n^T x^*$ as given in Table 1. We assume that data is collected according to a design with $\kappa = 36$ points uniformly spaced over the domain in a grid-like pattern (this is design one in Figure 9).

TABLE 1. True model parameters for $x = (x_1, x_2) \in \mathbb{R}^2$

	true model	domain
linear	$-6 + 6x_1 + 6x_2$	$[0,1]^2$
interaction	$-6 + 6x_1 + 6x_2 - 3x_1x_2$	$[0,1]^2$
quadratic	$-6 + 6x_1 + 6x_2 + 10x_1^2 + 3x_1x_2 + x_2^2$	$[0,1]^2$
log term	$-10 + 6\log x_1 + 6x_2$	$[1,2]^2$

For our linear model, the median effective dose is the straight line $\mathrm{ED}_{50} = \{(x_1, x_2) \in \mathcal{D} : 6 + 6x_1 + 6x_2 = 0\}$. We first assume that the estimator $\widehat{\beta}_n$ behaves according to its asymptotic distribution with $\kappa = 36$ and $m = 10$. Under this assumption, $B = 25$ samples of $\widehat{\mathrm{ED}}_{50}$ are shown in Figure 2 (top left). We then drew one sample of $\widehat{\mathrm{ED}}_{50}$ and calculated the various 95% confidence regions, as described in Section 4.5. The results are also shown in Figure 2. In each case, the confidence region is shown in grey, with ED_{50} shown as the bold line and the sample $\widehat{\mathrm{ED}}_{50}$ shown as the thin line. The different regions are

(1) $\mathrm{CR}_{\mathrm{SCH},0.95}$ based on inverting Scheffé's bounds

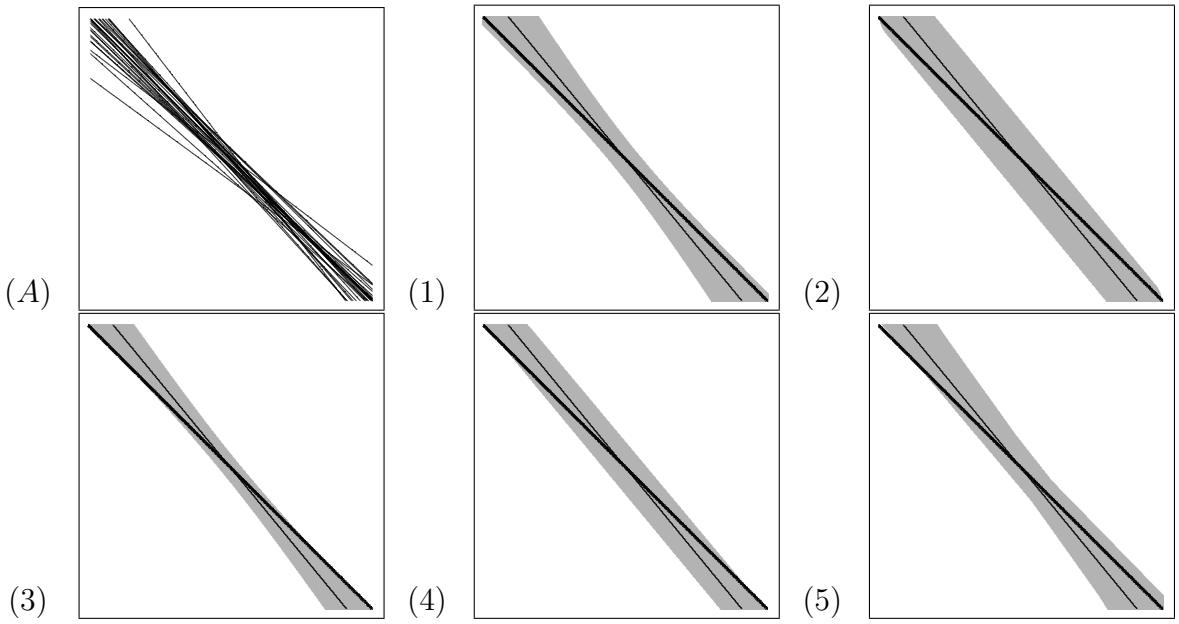

FIGURE 2. $(1)-(5)$: 95% confidence regions in gray for $\mathrm{ED}_{0.5}$ (bold line) in the linear model with $n = 360$ as indicated in the text; $\widehat{\mathrm{ED}}_{50}$ is shown as the thin line. Plot (A) shows 25 random samples of $\widehat{\mathrm{ED}}_{50}$.

(2) $\mathrm{CR}_{\mathrm{CLT},0.95}$ based on the central limit theorem for level sets
(3) $\mathrm{CR}_{\mathrm{Q1},0.95}$ based on random set quantiles with $\mathcal{M}_1$
(4) $\mathrm{CR}_{\mathrm{Q2},0.95}$ based on random set quantiles with $\mathcal{M}_2$
(5) $\mathrm{CR}_{0.95}$ based on the new empirical quantile algorithm

A comparison of the sampled $\widehat{\mathrm{ED}}_{50}$ values and the various confidence regions in Figure 2 reveals that $\mathrm{CR}_{\mathrm{CLT},0.95}$ and $\mathrm{CR}_{\mathrm{Q2},0.95}$ do not capture the local variability of $\widehat{\mathrm{ED}}_{50}$. On the other hand, $\mathrm{CR}_{\mathrm{SCH},0.95}$, $\mathrm{CR}_{\mathrm{Q1},0.95}$ and $\mathrm{CR}_{0.95}$ have shapes which describe well the behaviour of $\widehat{\mathrm{ED}}_{50}$. For this reason, we now focus only on these three methods.

We next consider the other three models of Table 1 in Figure 4, again with $\kappa = 36, m = 10$ and under design one. Contour plots of the true models are shown in Figure 3. Note that the domain for the model with a logarithmic term is different than the others. This is because the function $x^*(x)$ is not a continuous function on the domain $\mathcal{D} = [0, 1]^2$. Again, we now study only the three confidence regions: $\mathrm{CR}_{\mathrm{SCH},0.95}$, $\mathrm{CR}_{\mathrm{Q1},0.95}$ and $\mathrm{CR}_{0.95}$. Table 2 gives the proportions of the domain $\mathcal{D}$ covered by each confidence region. Figure 4 along with Table 2 reveal that $\mathrm{CR}_{\mathrm{Q1},0.95}$ is smaller than the other two regions, while $\mathrm{CR}_{0.95}$ and $\mathrm{CR}_{\mathrm{SCH},0.95}$ are similar in size, except for the quadratic model. That $\mathrm{CR}_{\mathrm{Q1},0.95}$ is smaller than the other two regions is not surprising. The Vorob'ev quantile takes a pointwise approach instead of a global one, and we expect it to be the smallest region. The fact that $\mathrm{CR}_{0.95}$ is smaller than $\mathrm{CR}_{\mathrm{SCH},0.95}$ for the quadratic model is probably caused by the fact

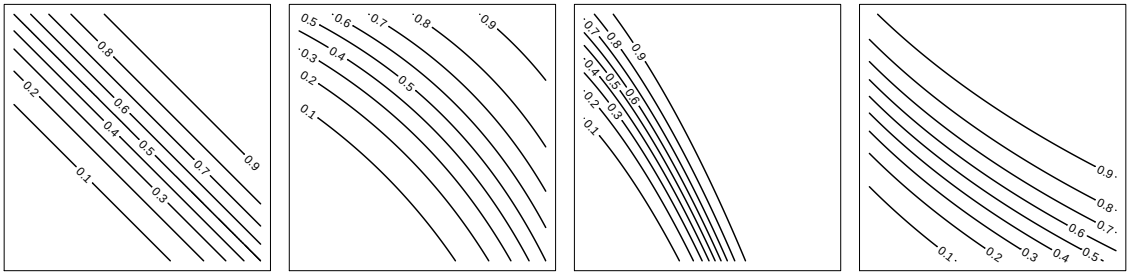

FIGURE 3. Contour plots for the true function f in the parametric setting. From left the right the models from Table 1 are linear, interaction, quadratic, and log-term.

that Scheffé's bounds are conservative, and this is particularly emphasized with an increase in the difference between $k+1$ (the number of parameters in the model) and d (the dimension of the covariates). Indeed, the region $\mathrm{CR}_{0.95}$ is nearly 11% smaller than the region $\mathrm{CR}_{\mathrm{SCH},0.95}$.

All three confidence regions show that there is considerable variability in the quadratic model, in that the function $\widehat{\beta}_n^T x^*$ dips down with some frequency in the upper right corner of the domain. Although not shown, this is consistent with samples of $\widehat{\mathrm{ED}}_{50}$. Note how much additional information is revealed about the variability of $\widehat{\mathrm{ED}}_{50}$ through viewing the confidence region in this case. In particular the true function $f(x) = \beta^T x^*$ does not exhibit such dips on $\mathcal{D}$ (Figure 3, second from left), and neither does the function $\widehat{\beta}_n^T x^*$ (not shown).

TABLE 2. Proportion of the domain covered by the confidence regions shown in Figures 2 and 4.

	linear	interaction	quadratic	log-term
$\mathrm{CR}_{\mathrm{SCH},0.95}$	0.161	0.230	0.803	0.159
$\mathrm{CR}_{0.95}$	0.168	0.245	0.718	0.172
$\mathrm{CR}_{\mathrm{Q1},0.95}$	0.116	0.147	0.474	0.109

5. SIMULATION RESULTS

We next study the behaviour of the three confidence regions, $\mathrm{CR}_{\mathrm{SCH},0.95}$, $\mathrm{CR}_{\mathrm{Q1},0.95}$ and $\mathrm{CR}_{0.95}$, through simulations. Each simulation is the result of 1000 samples, and we also used $B = 1000$ in the quantile set algorithm throughout. Whenever possible, the data in each simulation was the same for the different confidence regions.

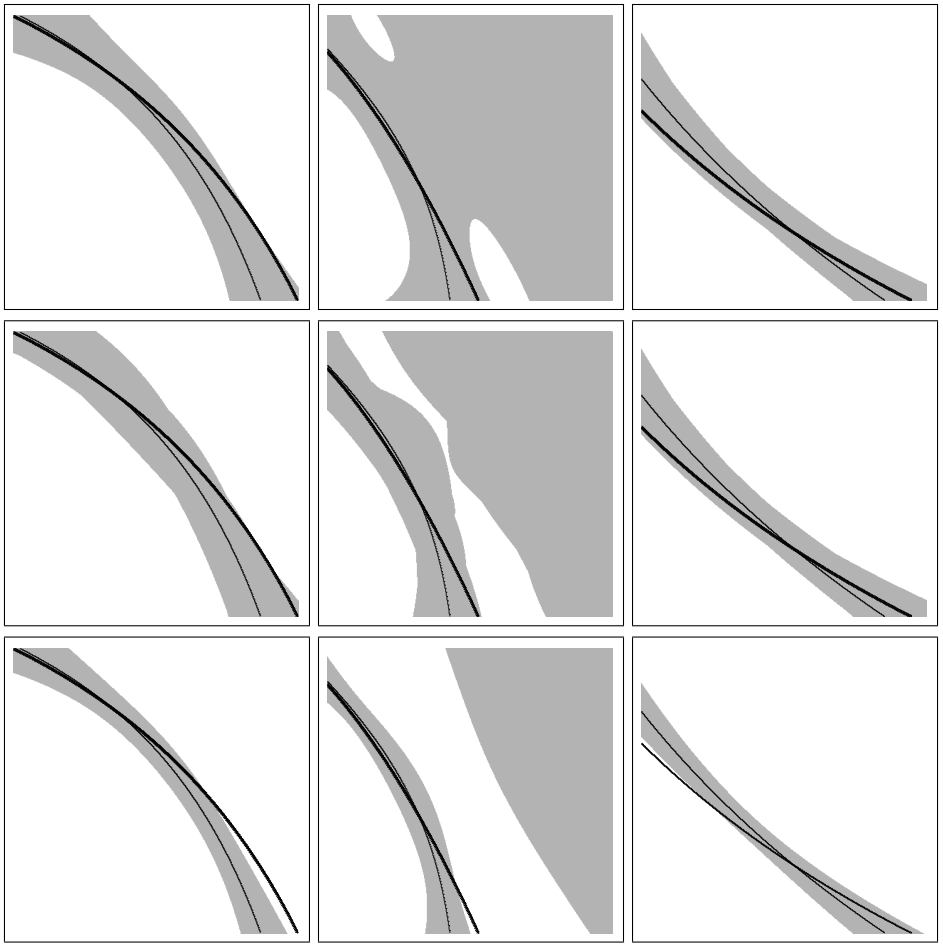

FIGURE 4. Confidence regions (gray) for ED_{50} (bold); $\widehat{\mathrm{ED}}_{50}$ is shown as the thin line. From top to bottom, the regions are $\mathrm{CR}_{\mathrm{SCH},0.95}$, $\mathrm{CR}_{\mathrm{Q1},0.95}$, and $\mathrm{CR}_{0.95}$. From left to right, the models are interaction, quadratic, and log term for $m = 10$.

We first look at the behaviour of the confidence regions when sampling from the asymptotic distribution. The results for design one are shown in Table 3, where $\kappa = 36$ and $m = 1, 10, 100$ for all four models given in Table 1. The region $\mathrm{CR}_{\mathrm{Q1},0.95}$ undercovers consistently, as expected. For the linear, interaction and log term models, the coverage of $\mathrm{CR}_{0.95}$ is either better or similar to that of $\mathrm{CR}_{\mathrm{SCH},0.95}$, while the sizes of the confidence regions are similar. For the quadratic model, the coverage of $\mathrm{CR}_{0.95}$ is smaller than that of $\mathrm{CR}_{\mathrm{SCH},0.95}$, as is the mean size of the region. $\mathrm{CR}_{0.95}$ is the only region which achieves its nominal coverage, if only in a few instances. Since we are sampling from the asymptotic

TABLE 3. Empirical coverage probabilities of 95% confidence regions for ED_{100p} for the parametric models in Table 1 for design 1 (see Figure 9) for simulations from the asymptotic normal distribution. Results *not* statistically different from 0.95 are shown in bold. The corresponding sizes of the confidence regions, measured as the mean proportion of the domain covered by the region, are given in brackets.

		linear			interaction			quadratic			log term		
n	p	SCH	Q1	CR	SCH	Q1	CR	SCH	Q1	CR	SCH	Q1	CR
36	.1	.993	.897	.976	.995	.889	.972	1.00	.865	.985	.994	.928	.983
		(.46)	(.34)	(.50)	(.59)	(.42)	(.60)	(.95)	(.66)	(.91)	(.26)	(.18)	(.32)
	.5	.984	.874	.969	.995	.808	.980	.997	.792	.980	.985	.884	.974
		(.75)	(.51)	(.80)	(.90)	(.63)	(.90)	(.99)	(.81)	(.98)	(.72)	(.49)	(.79)
	.9	.990	.898	.996	1.00	.921	1.00	.998	.796	.987	.982	.872	.992
		(.47)	(.35)	(.51)	(.42)	(.24)	(.45)	(.96)	(.74)	(.93)	(.64)	(.50)	.(70)
360	.1	.993	.900	.973	.995	.891	.968	1.00	.880	.989	.994	.930	.981
		(.18)	(.12)	(.17)	(.26)	(.16)	(.21)	(.37)	(.15)	(.32)	(.10)	(.07)	(.10)
	.5	.984	.874	**.947**	.995	.810	**.942**	.998	.805	.982	.985	.886	**.963**
		(.19)	(.13)	(.20)	(.27)	(.18)	(.27)	(.57)	(.18)	(.47)	(.18)	(.13)	(.19)
	.9	.991	.903	.990	1.00	.922	.997	.998	.801	.972	.982	.873	**.962**
		(.18)	(.12)	(.19)	(.12)	(.08)	(.10)	(.64)	(.33)	(.59)	(.26)	(.18)	(.28)
3600	.1	.993	.911	.975	.995	.893	.968	1.00	.880	.987	.994	.931	.981
		(.06)	(.04)	(.05)	(.08)	(.05)	(.07)	(.06)	(.03)	(.05)	(.03)	(.02)	(.03)
	.5	.984	.874	**.948**	.996	.814	**.942**	.998	.814	.970	.985	.887	.970
		(.06)	(.04)	(.06)	(.09)	(.06)	(.09)	(.06)	(.03)	(.05)	(.06)	(.04)	(.06)
	.9	.993	.917	.985	1.00	.923	.976	.998	.805	**.959**	.982	.876	.967
		(.06)	(.04)	(.06)	(.05)	(.03)	(.05)	(.09)	(.04)	(.08)	(.08)	(.06)	(.09)

TABLE 4. Empirical coverage probabilities of 95% confidence regions for ED_{100p} for the parametric models in Table 1 for design one (see Figure 9) using the maximum likelihood estimate of Σ. Results *not* statistically different from 0.95 are shown in bold. The corresponding sizes of the confidence regions, measured as the mean proportion of the domain covered by the region, are given in brackets.

		linear			interaction			quadratic			log term		
n	p	SCH	Q1	CR	SCH	Q1	CR	SCH	Q1	CR	SCH	Q1	CR
36	.1	.989	**.948**	.995	.995	**.948**	.992	*	*	*	.990	**.956**	.997
		(.47)	(.39)	(.54)	(.61)	(.48)	(.64)	*	*	*	(.28)	(.20)	(.36)
	.5	1.00	**.943**	.998	1.00	**.942**	1.00	*	*	*	1.00	.930	.993
		(.93)	(.44)	(.95)	(1.0)	(.58)	(1.0)	*	*	*	(.85)	(.43)	(.96)
	.9	.991	.936	.993	.997	**.957**	.997	*	*	*	.994	.922	.996
		(.47)	(.39)	(.53)	(.43)	(.27)	(.49)	*	*	*	(.67)	(.57)	.(72)
360	.1	.988	.908	.977	.992	.899	.966	1.00	.884	.994	.988	.932	.980
		(.18)	(.12)	(.17)	(.26)	(.16)	(.21)	(.34)	(.14)	(.29)	(.10)	(.07)	(.10)
	.5	.988	.861	**.946**	.994	.802	**.945**	1.00	.844	.995	.983	.867	.965
		(.19)	(.13)	(.20)	(.27)	(.17)	(.27)	(.51)	(.12)	(.40)	(.18)	(.12)	(.19)
	.9	.990	.896	.985	.997	.928	.989	.996	.864	.973	.980	.891	.971
		(.18)	(.12)	(.19)	(.12)	(.09)	(.11)	(.69)	(.26)	(.66)	(.26)	(.18)	(.28)
3600	.1	.995	.917	.980	.999	.902	.980	1.00	.880	.971	.988	.921	.970
		(.06)	(.04)	(.05)	(.08)	(.05)	(.07)	(.06)	(.03)	(.05)	(.03)	(.02)	(.03)
	.5	.986	.867	**.944**	.989	.820	**.937**	.999	.828	.982	.991	.865	.968
		(.06)	(.04)	(.07)	(.09)	(.06)	(.09)	(.06)	(.03)	(.05)	(.06)	(.04)	(.06)
	.9	.991	.905	.973	.996	.920	.969	.999	.836	.975	.989	.898	.973
		(.06)	(.04)	(.06)	(.05)	(.03)	(.05)	(.08)	(.04)	(.07)	(.08)	(.06)	(.09)

distribution, the behaviour of the confidence regions is relatively similar as the sample size varies. The mean sizes of the regions do decrease with the sample size, which is again expected.

Next, we repeated the experiment, but in the more realistic situation where the true Σ is unknown. In this case, we use its estimate $\widehat{\Sigma}_n$ either directly (in $\mathrm{CR}_{\mathrm{SCH},0.95}$ and $\mathrm{CR}_{\mathrm{Q1},0.95}$) or in the parametric bootstrap (in $\mathrm{CR}_{0.95}$). The results are shown in Table 4. Note that we did not report the results for the quadratic model when $n = 36$. This is because there was too large a proportion of observations in "complete separation" (Albert and Anderson, 1984) in this case. When the sample size is $n = 360$ and $n = 3600$, the results in Table 4 and Table 3 are similar. When $n = 36$, however, $\mathrm{CR}_{\mathrm{SCH},0.95}$ and $\mathrm{CR}_{0.95}$ behave comparably. Surprisingly, the additional variability of the problem due to the estimation of the variance matrix also increases the empirical coverage of $\mathrm{CR}_{\mathrm{Q1},0.95}$. For the linear, interaction, and log term models, $\mathrm{CR}_{\mathrm{Q1},0.95}$ reaches nominal levels more often than not in the simulations. Notably, we also tried estimating the z_α quantile in $\mathrm{CR}_{\mathrm{Q1},0.95}$ through both a parametric and non-parametric bootstrap approach for $n = 36$. In the parametric bootstrap the results are similar to those in Table 4, while for the nonparametric bootstrap the results again undercover in a manner similar to that in Table 3. Additional simulations for $d = 2$ are provided in the Appendix.

We also considered two simple models for the higher dimensional $d = 3$ case: the true linear model was $\eta(p) = -6 + 3x_1 + 3x_2 + 3x_3$ and the true quadratic model was $\eta(p) = -6 + 3x_1 + 3x_2 + 3x_3 + x_1^2 + x_2^2 + x_3^2$. The design was similar to design one, in that data were observed on a uniform grid inside $\mathcal{D} = [0, 1]^3$ with $\kappa = 6^3$ and $m = 10$. The domain was the same for both models, and we simulated only from the asymptotic distribution. The results are shown in Table 5.

TABLE 5. $d = 3$

		linear			quadratic		
n	p	S	Q1	CR	S	Q1	CR
2160	.1	.991	.811	.977	.997	.664	**.937**
		(.17)	(.11)	(.16)	(.16)	(.08)	(.13)
	.5	.983	.769	.921	.999	.569	**.938**
		(.10)	(.06)	(.09)	(.17)	(.09)	(.14)
	.9	.997	.932	.988	1.00	.692	**.961**
		(.01)	(.01)	(.01)	(.10)	(.05)	(.08)

6. EXAMPLES

6.1. **A decompression sickness study.** Our first example comes from a decompression sickness (DCS) study from the University of Wisconsin, Madison. DCS is most often associated with diving, but can be experienced in other depressurisation events such as caisson

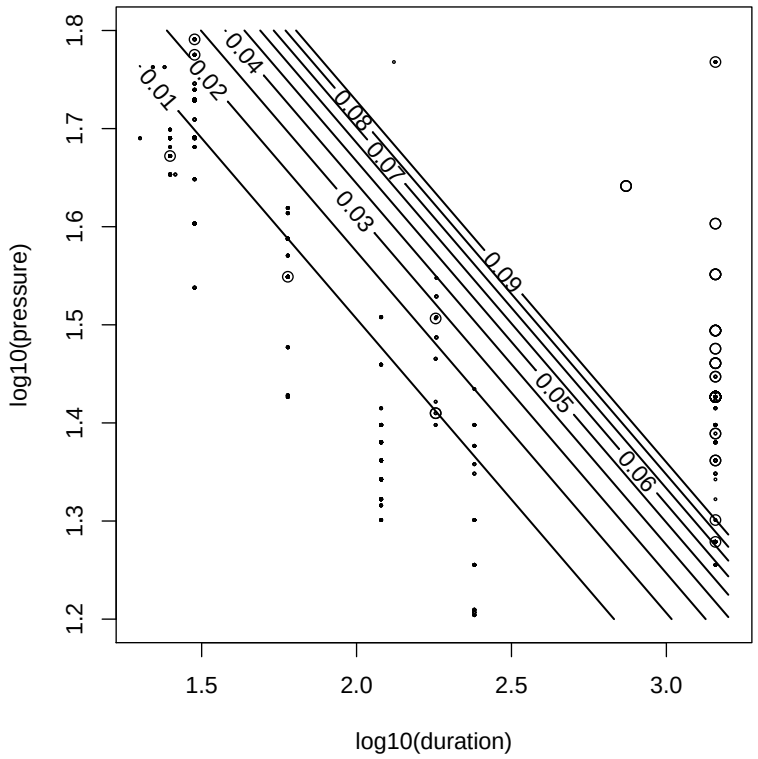

FIGURE 5. Estimates of ED_{100p} for different values of p for the cytotoxicity data. Data points with response equal to 0/1 are shown as closed/open circles.

working (i.e. during the building of dams or tunnels) or unpressurized flight. In this study, sheep were exposed to a variety of exposure pressures and durations in a pressure chamber. Sheep have a similar body mass to humans, and were therefore used as an approximate model for human response. DCS is caused by dissolved gasses vaporising on depressurization forming dangerous bubbles of gas throughout the body. The bubbles can form in different locations in the body, and therefore lead to a variety of symptoms as well as severities. In the study, central nervous system and respiratory DCS as well as limb bends and mortality were included in the recorded outcomes. The sample size is $n = 1108$. The data set has been considered extensively in Li et al. (2008b,a, 2010); Li and Wong (2011) and we refer to these papers for further details on data collection and analysis. Here, we compare our methods to those of Li et al. (2008b) where a linear model was fit for the mortality response. The data has been updated several times in the past few years, and therefore our results differ slightly from those found in Li et al. (2008b).

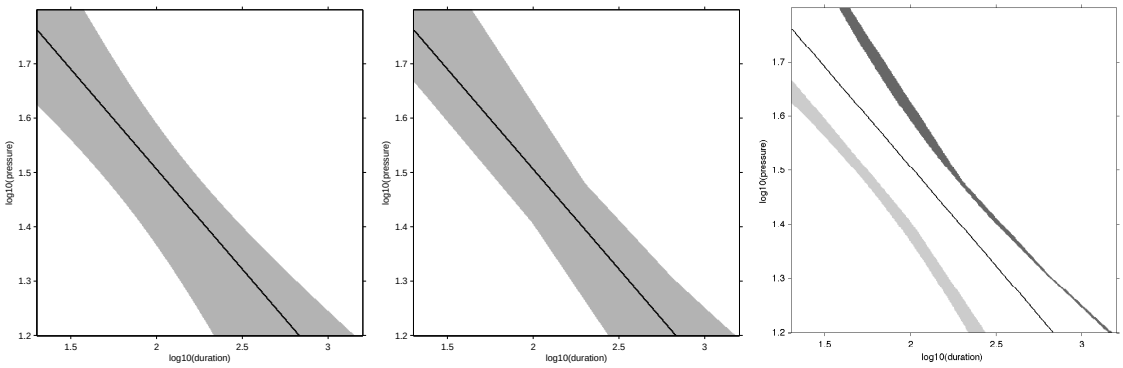

FIGURE 6. 95% confidence regions for ED_{50} based on Scheffé's upper bound (left) and the new CRS quantiles (centre). The difference between the sets is shown on the left ($\mathrm{CR}_{\mathrm{SCH},0.95}/\mathrm{CR}_{0.95}$ in light and $\mathrm{CR}_{0.95}/\mathrm{CR}_{\mathrm{SCH},0.95}$ in dark gray). $\widehat{\mathrm{ED}}_1$ is superimposed throughout for reference.

When the survival response is considered, a simple linear model gives an adequate fit to the data. The effective dose (with death denoted as "success") is thus estimated as

$$\widehat{\mathrm{ED}}_{100p} \;=\; \{-27.64 + 3.78x_1 + 10.28x_2 = \eta(p)\},$$

where we take $\mathcal{D} = [1.3, 3.6] \times [1.2, 1.8]$ with x_1 and x_2 denoting the base 10 logarithm of duration and pressure of the dive. The data and results for $\widehat{\mathrm{ED}}_{100p}$ are shown in Figure 5. Fix $p = 0.01$. From Theorem 3.1 we know that the plug-in estimator $\widehat{\mathrm{ED}}_1$ is consistent for ED_1, assuming a linear model. We are also interested in the variability of $\widehat{\mathrm{ED}}_1$, and therefore we calculate simultaneous confidence regions for ED_1. Results for $\mathrm{CR}_{\mathrm{SCH},0.95}$ and $\mathrm{CR}_{0.95}$ are shown in Figure 6. The regions are similar in size, covering 30% and 29% of the domain respectively. However, $\mathrm{CR}_{0.95}$ appears more centred around $\widehat{\mathrm{ED}}_1$, particularly for large values of x_1.

6.2. **Cytotoxicity in the leukemia cell line HL-60.** Carter et al. (1986) consider a cytotoxicity data set where the effect of two toxins, methylmethanesulfonate (MMS) and phorbol 12-myristate 13-acetate (PMA), on the human promyeloctic leukemia cell line HL-60 was evaluated. Both MMS and PMA have demonstrated carcinogenic properties, and it was of interest to understand their interactive properties. In the study, 16 treatments were considered with 83 to 98 observations per treatment, for a total sample size of 1436. The data and a detailed analysis is available in Carter et al. (1986). There, a logistic regression model was fit resulting in the plug-in estimate

$$\widehat{\mathrm{ED}}_{100p} \;=\; \{-1.330 - 0.084x_1 + 0.159x_2 + 0.003883x_1^2 - 0.001308x_2^2 = \eta(p)\},$$

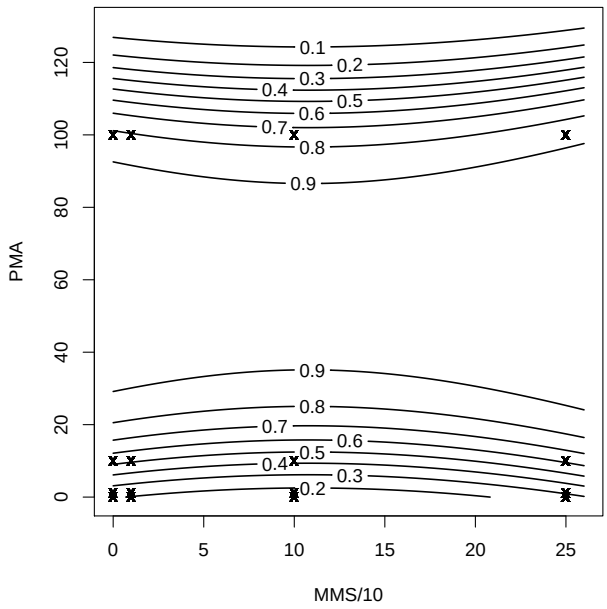

FIGURE 7. Estimates of ED_{100p} for different values of p for the DCS data (left). Locations where data was observed are marked with an "x" (the treatments had 83 to 98 observations each).

with x_1, x_2 corresponding to MMS/10 and PMA, respectively. Several estimates for different values of p are shown in Figure 7 (left). The appropriateness of the quadratic model is at first counterintuitive, however, due to lysis, certain treatments can become so toxic that cells become uncountable (Carter et al., 1986). Thus, an increase in perceived survival at higher toxicities is an appropriate result in this experiment.

The fitted model is non-linear, however, we can preform a simple check for consistency as follows. The determinant of the Hessian matrix for the model $f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2$ is equal to $4\beta_3\beta_4$. If this quantity is negative, then all critical points of $f(x)$ are saddle points, and we would therefore have consistency. Now, if $\beta_3 > 0$ and $\beta_4 < 0$, then $4\beta_3\beta_4 < 0$. From Carter et al. (1986), the p-values for each of these tests are smaller than 0.0002 and 0.0001 respectively. Therefore, as long as the underlying model is correct, we can be fairly confident that the estimators $\widehat{\mathrm{ED}}_p$ are consistent.

With such a large sample size, we would expect little variability in the values of $\widehat{\beta}_n$. However, it is not immediately clear how this translates to the variability of $\widehat{\mathrm{ED}}_{100p}$. In Figure 7 (right), we show several (parametric) bootstrap re-samples of $\widehat{\mathrm{ED}}_{50}$. Next, we calculated 95% confidence regions for ED_{50} using Scheffé's upper bound (Figure 8, left) and the new CRS quantiles (Figure 8, centre). The new method yields a tighter confidence band.

Indeed, here, the percentage of the domain covered is 14.92%, whereas using Scheffé's upper bound 18.31% of the domain was covered. Figure 8 (right) shows that $\mathrm{CR}_{0.95}$ is almost entirely contained in $\mathrm{CR}_{\mathrm{SCH},0.95}$.

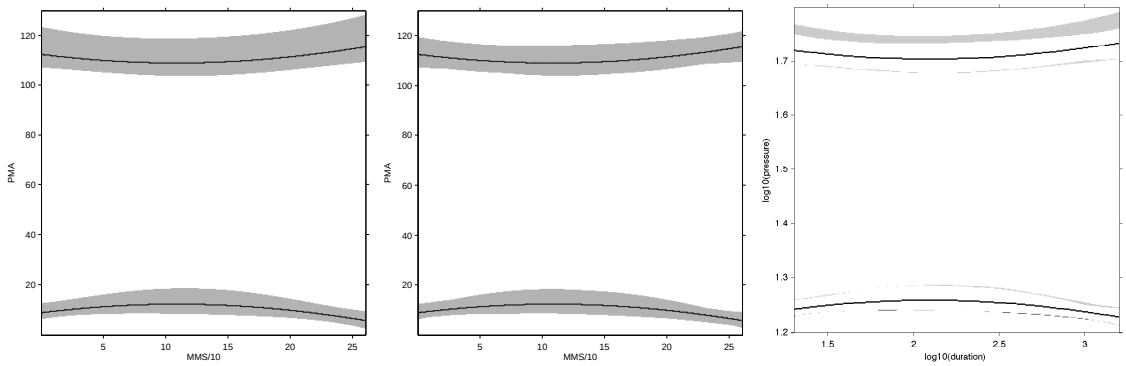

FIGURE 8. 95% confidence regions for ED_{50} based on Scheffé's upper bound (left) and the new CRS quantiles (centre). The difference between the sets is shown on the left ($\mathrm{CR}_{\mathrm{SCH},0.95}/\mathrm{CR}_{0.95}$ in light and $\mathrm{CR}_{0.95}/\mathrm{CR}_{\mathrm{SCH},0.95}$ in dark gray). $\widehat{\mathrm{ED}}_{50}$ is superimposed throughout for reference.

Carter et al. (1986) also calculate confidence regions using Scheffé's method for $p = 0.4, 0.5,$ and 0.6. These regions overlap, and hence Carter et al. (1986) conclude that we "cannot confidently distinguish among the respective ED_{100p} sets." Although not shown, the regions in this case using the new quantile method would also overlap. However, the estimates $\widehat{\mathrm{ED}}_{100p_1}$ and $\widehat{\mathrm{ED}}_{100p_2}$ are highly positively correlated, and therefore such an ad-hoc comparison is probably overly conservative. The question of how to correctly account for this correlation is also interesting and important, but is beyond the scope of this work.

7. DISCUSSION

When studying confidence intervals in the univariate setting, one is typically looking for two properties: the confidence interval should be (1) as small as possible while (2) reaching the nominal level, but without under-coverage. In our context, the observed data comes from a non-Euclidean space, so some additional properties are desirable. Confidence regions provide visual information about the variability of our estimates, and therefore the shape of the confidence regions should reflect the behaviour of the estimates and have the ability to detect local variability. For practitioners, it would also be desirable that the confidence region be relatively easy to calculate.

In this work we have studied various methods of calculating a confidence region for a multi-dimensional effective dose. The confidence regions $\mathrm{CR}_{\mathrm{SCH},0.95}$, $\mathrm{CR}_{\mathrm{Q1},0.95}$ and $\mathrm{CR}_{0.95}$ all nicely show the local variability of the effective dose estimators. However, for moderate

and large sample sizes, $\mathrm{CR}_{\mathrm{SCH},0.95}$ systematically exhibits over-coverage while $\mathrm{CR}_{\mathrm{Q1},0.95}$ systematically under covers in simulation studies. The new set quantile algorithm introduced here yields the region $\mathrm{CR}_{0.95}$ which outperforms the others in terms of empirical coverage and size of region for moderate and large samples sizes. For smaller sample sizes the behaviour of $\mathrm{CR}_{0.95}$ and $\mathrm{CR}_{\mathrm{SCH},0.95}$ is, however, equally disappointing.

The confidence regions $\mathrm{CR}_{\mathrm{SCH},0.95}$ and $\mathrm{CR}_{\mathrm{Q1},0.95}$ are the asymptotic confidence regions in that they use the asymptotic distribution quantiles. For smaller sample sizes, one could potentially either bootstrap these quantiles or replace them with their non-asymptotic counterparts. Thus, in $\mathrm{CR}_{\mathrm{SCH},0.95}$ one could replace the asymptotic χ^2 quantile with it's appropriate F distribution version (Scheffé, 1953). However, these quantiles would be larger, and $\mathrm{CR}_{\mathrm{SCH},0.95}$ already exhibits over-coverage. Another potential avenue here would be to consider the method of Piegorsch and Casella (1988) (see also Casella and Strawderman (1980)), but we do not explore this here. Notably, this modification would greatly increase the computational complexity of the regions $\mathrm{CR}_{\mathrm{SCH},0.95}$. Recall that when the sample size is small, the region $\mathrm{CR}_{\mathrm{Q1},0.95}$ occasionally achieves nominal levels, however, this not a systematic property. We believe that this behaviour is coincidental only, and not indicative of any theoretical properties. Additional simulations not shown here revealed that when the asymptotic quantiles in $\mathrm{CR}_{\mathrm{Q1},0.95}$ were replaced with empirically estimated ones the empirical coverage probabilities for $n = 36$ resembled those of $n = 360$, and systematic under-coverage was again observed.

The new algorithm introduced here is essentially a bootstrapping procedure. Our original motivation for this approach came from the setting $d = 1$, where we found that a simple bootstrap approach had the best small sample size performance. Indeed, when $d = 1$, our method is equivalent to univariate bootstrap methods. This new bootstrapped quantile does have some appealing properties, and these are summarized in Proposition 4.3. In particular, the empirical quantile is invariant under rotations and equivariant under re-scaling.

Of the confidence regions consider here, the new confidence region $\mathrm{CR}_{1-\alpha}$ is also the most computationally intensive and requires some simulations. On the other hand, the regions $\mathrm{CR}_{\mathrm{SCH},1-\alpha}$, and $\mathrm{CR}_{\mathrm{Q1},1-\alpha}$ may be calculated directly. However, in all of the two-dimensional examples considered here, $\mathrm{CR}_{1-\alpha}$ took less then one minute to compute on a 2.4 GHz dual core Macintosh laptop. This could probably be reduced even further by studying more efficient programming techniques. Thus, the time requirements to calculate these regions do not carry great practical constraints. Matlab script calculating all three regions for the example of Section 6.2 is available online at `www.math.yorku.ca/~hkj/Software/`.

Previous work on the problem of confidence regions for the effective dose is, to our best knowledge, not extensive. The region $\mathrm{CR}_{\mathrm{SCH},1-\alpha}$ was introduced in Carter et al. (1986) and studied extensively in Li et al. (2008b) for the parametric model. Li et al. (2008b) also consider the problem of the conditional effective dose (i.e. the effective dose obtained after holding certain covariates fixed). Although our simulations do not explore this, the new method presented here can also be applied in this setting. We conjecture that empirical coverage will be similar, particularly for models without interactions, as fixing a covariate

essentially changes the behaviour of the intercept in such cases. This conditional approach would be most appealing in higher dimensions, since relationships are more difficult to visualize when $d > 2$.

Li et al. (2010) also study an extension of this approach to the case where $\widehat{f}_n$ is estimated using a semi-parametric model. They consider two methods: one is based on a theoretical bound of $\sup \mathbb{Z}_n^2$ similar to the parametric case, and the second is based on empirical estimation of the theoretical quantiles of $\sup \mathbb{Z}_n^2$ via a parametric bootstrap procedure. Unfortunately, their theoretical bounds in this setting are incorrect (this is the reason for the discrepancy between the bootstrap and theoretical procedure noted in Li et al. (2010, Section 4.1 and Figure 1)). The bootstrap approach, although quite conservative, is more promising. However, although the semi-parametric estimates are insensitive to the choice of bandwidth, the parametric bootstrap confidence regions increase in size and coverage as h decreases. This is not surprising as the covariance depends on $\sqrt{nh}$ in the denominator. The question of optimal choice of h was not considered in Li et al. (2008b). One could also try the region $\mathrm{CR}_{1-\alpha}$ in the semi-parametric setting, and we intend to study this problem thoroughly in a future work.

8. Acknowledgements

The first author thanks Georges Monette from York University and Ruxandra Pinto from Sunnybrook Hospital for helpful discussions. All authors are grateful to Jialiang Li for sharing the DCS data set. Parts of this work were made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada.

9. Appendix

9.1. **Additional simulations.** Following Li et al. (2008b) we consider several different experimental designs in our simulations. The various designs are given in Figure 9, and recall that contours of the true models are given in Figure 3. In these simulations we consider only the case where Σ must be estimated from the data. The results for $n = 36$ are given in Table 6 and those for $n = 360$ in Table 7. We note that for $n = 36$ and the linear and interaction models, under designs 1,3, and 5 one expects at least 10 failures and 10 successes, but not so for designs 2,4, and 6. This is also holds for $n = 36$ and $n = 360$ for all designs in the log term model. As in Section 5, we do not report the simulations which had a large proportion (more than $\sim 20\%$) of data in complete separation.

9.2. **Technical work.**

Proof of Proposition 4.3. (1) By definition, γ_α is such that $\lfloor \alpha B \rfloor$ of the observed sets satisfy $\rho(A_i, K_B) \leq \gamma_\alpha$.

(2) This again follows from the definition.

TABLE 6. Empirical coverage probabilities of 95% confidence regions for ED_{100p} for the parametric models give in Table 1 using $\widehat{\Sigma}_n$ throughout. The designs are shown in Figure 9. Results not statistically different from 0.95 are shown in bold. The mean proportion of the domain covered by the region is given in brackets.

			linear			interaction			log term		
n	design	p	SCH	Q1	CR	SCH	Q1	CR	SCH	Q1	CR
36	2	.1	.989	**.956**	.993	.992	.917	.992	.984	.930	.993
			(.70)	(.47)	(.83)	(.98)	(.74)	(.98)	(.23)	(.16)	(.32)
		.5	1.00	**.938**	1.00	1.00	**.943**	1.00	1.00	.888	.997
			(.93)	(.57)	(.98)	(.99)	(.67)	(.98)	(.92)	(.46)	(.97)
		.9	1.00	**.957**	1.00	1.00	.987	1.00	.993	.937	.997
			(.51)	(.43)	(.57)	(.59)	(.44)	(.62)	(.65)	(.55)	(.68)
	3	.1	.989	.921	.993	.996	.925	.995	.992	.911	.998
			(.74)	(.61)	(.81)	(.85)	(.68)	(.87)	(.59)	(.42)	(.68)
		.5	1.00	.968	1.00	1.00	.916	1.00	1.00	.930	.999
			(.98)	(.68)	(1.0)	(1.0)	(.82)	(1.0)	(.95)	(.65)	(.98)
		.9	.979	.912	.986	.995	.908	.992	.981	.905	.989
			(.74)	(.61)	(.81)	(.80)	(.53)	(.82)	(.84)	(.73)	(.88)
	4	.1	.985	.915	.932	*	*	*	.989	.906	.993
			(.85)	(.62)	(.93)	*	*	*	(.49)	(.33)	(.61)
		.5	1.00	.963	1.00	*	*	*	.998	.902	.997
			(.97)	(.73)	(.99)	*	*	*	(.93)	(.59)	(.98)
		.9	1.00	**.955**	1.00	*	*	*	.994	.905	.997
			(.71)	(.58)	(.77)	*	*	*	(.78)	(.66)	(.81)
	5	.1	.993	.930	.995	.999	.931	.997	.989	.933	.994
			(.65)	(.51)	(.74)	(.81)	(.68)	(.83)	(.51)	(.34)	(.62)
		.5	1.00	.910	.998	1.00	.899	1.00	1.00	.912	.999
			(.99)	(.63)	(1.0)	(1.0)	(.79)	(1.0)	(.98)	(.63)	(.99)
		.9	.987	.916	.989	.994	.920	.994	.986	.911	.992
			(.65)	(.51)	(.74)	(.73)	(.53)	(.76)	(.78)	(.65)	(.84)
	6	.1	.989	.919	.996	.993	.908	.993	.990	**.945**	.996
			(.85)	(.61)	(.93)	(1.0)	(.98)	(.1.0)	(.59)	(.33)	(.75)
		.5	1.00	**.957**	1.00	1.00	**.957**	1.00	.998	.894	.998
			(.97)	(.70)	(.99)	(1.0)	(.97)	(1.0)	(1.0)	(.70)	(1.0)
		.9	.999	.979	1.00	1.00	.987	1.00	.993	.931	.997
			(.65)	(.54)	(.72)	(.96)	(.94)	(.97)	(.74)	(.65)	(.78)

TABLE 7. Empirical coverage probabilities of 95% confidence regions for ED_{100p} for the parametric models give in Table 1 using $\widehat{\Sigma}_n$ throughout. The designs are shown in Figure 9. Results not statistically different from 0.95 are shown in bold. The mean proportion of the domain covered by the region is given in brackets.

			linear		interaction		quadratic		log term	
n	design	p	SCH	CR	SCH	CR	SCH	CR	SCH	CR
360	2	.1	.983	.972	.987	.970	.998	.995	.991	.980
			(.16)	(.15)	(.30)	(.29)	(.33)	(.29)	(.08)	(.08)
		.5	.983	**.947**	.993	.979	1.00	.987	.986	.965
			(.23)	(.26)	(.53)	(.54)	(.46)	(.38)	(.16)	(.17)
		.9	.994	.983	1.00	.999	.997	.968	.990	.971
			(.27)	(.28)	(.27)	(.24)	(.63)	(.57)	(.30)	(.32)
	3	.1	.985	.969	.992	.986	*	*	.987	.984
			(.34)	(.36)	(.49)	(.45)	*	*	(.18)	(.21)
		.5	.986	**.944**	.992	**.947**	*	*	.985	.970
			(.39)	(.40)	(.47)	(.46)	*	*	(.37)	(.40)
		.9	.987	.991	.995	.993	*	*	.982	.968
			(.34)	(.33)	(.30)	(.31)	*	*	(.42)	(.43)
	4	.1	.985	**.961**	.986	.988	.996	.981	.993	.978
			(.29)	(.31)	(.77)	(.78)	(.69)	(.64)	(.12)	(.14)
		.5	.987	**.956**	.992	**.963**	1.00	.985	.973	**.960**
			(.37)	(.40)	(.71)	(.72)	(.74)	(.69)	(.30)	(.32)
		.9	.991	.995	.995	.992	.998	.982	.988	.970
			(.36)	(.37)	(.51)	(.52)	(.78)	(.74)	(.38)	(.40)
	5	.1	.985	.970	.991	.989	*	*	.993	.975
			(.27)	(.27)	(.48)	(.48)	*	*	(.14)	(.16)
		.5	.987	.971	.992	**.963**	*	*	.977	.968
			(.31)	(.31)	(.49)	(.46)	*	*	(.30)	(.31)
		.9	.990	.972	.995	.992	*	*	.983	**.963**
			(.27)	(.26)	(.35)	(.35)	*	*	(.37)	(.38)
	6	.1	.994	.977	.991	.992	.999	.997	.991	.976
			(.16)	(.26)	(.85)	(.85)	(.91)	(.90)	(.11)	(.12)
		.5	.994	.976	.994	.983	1.00	.996	.982	**.966**
			(.36)	(.38)	(.83)	(.84)	(.94)	(.92)	(.28)	(.31)
		.9	.996	.994	.998	.998	.997	.990	.988	**.967**
			(.35)	(.34)	(.69)	(.70)	(.96)	(.95)	(.43)	(.44)

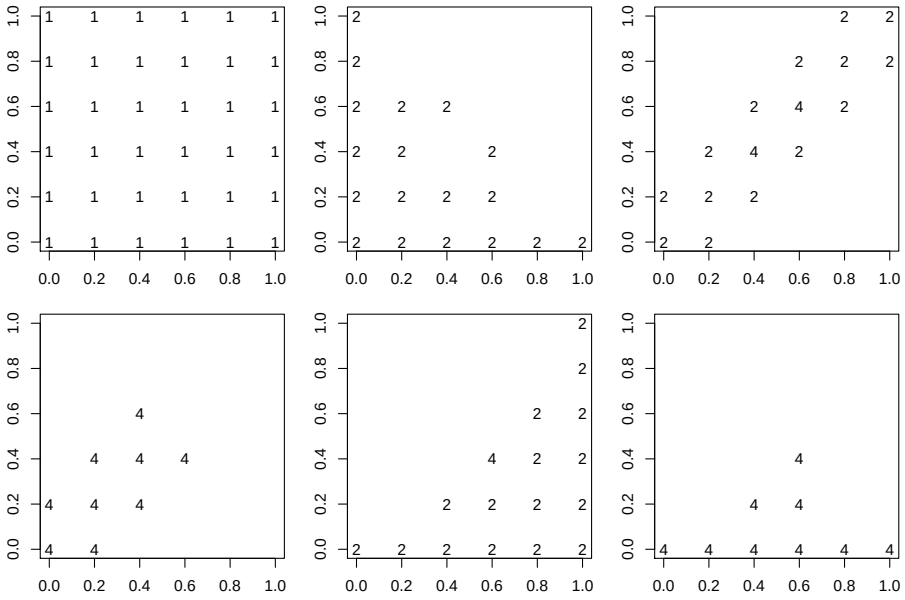

FIGURE 9. Different design matrices considered for the parametric model. The designs are shown from left to right: 1–3 in the top row and 4–6 in the bottow row.

(3) Recall that $\rho_H(A,B) = \sup_{x\in\mathcal{D}} |d_A(x) - d_B(x)|$, where $d_A(\cdot)$ denotes the distance transform of A (Delfour and Zolésio, 1994), and also that $d_{g(A)}(x) = d_A(g^{-1}(x))$ (Jankowski and Stanberry, 2010, see e.g.). It follows that

$$\rho_H(g(A_i), g(K_B)) \quad = \quad \rho_H(A_i, K_B),$$

which implies the result.

(4) The same argument as for rigid motions works for dilations by $\alpha > 0$.

□

REFERENCES

ALBERT, A. and ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10.

CARTER, W., CHINCHILLI, V., WILSON, J., CAMPBELL, E., KESSLER, F. and CARCHMAN, R. (1986). An asymptotic confidence region for the ED100p from the logistic response surface for a combination of agents. *The American Statistician* **40** 124–128.

CASELLA, G. and STRAWDERMAN, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *J. Amer. Statist. Assoc.* **75** 862–868.

CHEN, D. G. (2007). Dose-time response cumulative multinomial generalized linear model. *J. Biopharm. Statist.* **17** 173–185.

CUEVAS, A., GONZÁLEZ-MANTEIGA, W. and RODRÍGUEZ-CASAL, A. (2006). Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* **48** 7–19.

DELFOUR, M. C. and ZOLÉSIO, J.-P. (1994). Shape analysis via oriented distance functions. *J. Funct. Anal.* **123** 129–201.

JANKOWSKI, H. and STANBERRY, L. (2010). Expectations of random sets and their boundaries using oriented distance functions. *Journal of Mathematical Imaging and Vision* **36** 291–303.

JANKOWSKI, H. and STANBERRY, L. (2011). Confidence regions for means of random sets using oriented distance functions. *Scandinavian Journal of Statistics* To appear.

LANG, D. R., KURZEPA, H., COLE, M. S. and LOPER, J. C. (1980). Malignant transformation of BALB/3T3 cells by residue organic mixtures from drinking water. *Journal of Environmental Pathology & Toxicology* **4** 41–54.

LI, J. and WONG, W. K. (2011). Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics* **12** 143155.

LI, J., ZHANG, C. M., DOKSUM, K. A. and NORDHEIM, E. V. (2010). Simultaneous confidence intervals for semiparametric logistic regression and confidence regions for the multi-dimensional effective dose. *Statistica Sinica* **20** 637659.

LI, J., ZHANG, C. M., NORDHEIM, E. V. and LEHNER, C. E. (2008a). Estimation and confidence regions of multi-dimensional effective dose. *Biometrical Journal* **50** 110–112.

LI, J., ZHANG, C. M., NORDHEIM, E. V. and LEHNER, C. E. (2008b). On the multivariate predictive distribution of multi-dimensional effective dose: a Bayesian approach. *Journal of Statistical Computation and Simulation* **78** 429–442.

MOLCHANOV, I. (1990). Empirical estimation of distribution quantiles of random closed sets (Russian). *Teor. Veroyatnost. i Primen.* **35** 586–592. English translation in: *Theory of Probability and Its Applications*, 1990, **35**, pp.594-600.

MOLCHANOV, I. (2005). *Theory of random sets*. Probability and its Applications (New York), Springer-Verlag London Ltd., London.

MOLCHANOV, I. S. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics* **25** 235–242.

PIEGORSCH, W. W. and CASELLA, G. (1988). Confidence bands for logistic regression with restricted predictor variables. *Biometrics* **44** 739–750.

SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40** 87–104.

SKARIN, A. T., CANELLOS, G. P., ROSENTHAL, D. S., CASE, D. C., MACINTYRE, J. M., PINKUS, G. S., MOLONEY, W. C. and FREI, E. (1983). Improved prognosis of diffuse histiocytic and undifferentiated lymphoma by use of high dose methotrexate alternating with standard agents (M-BACOD). *Journal of Clinical Oncology* **1** 91–98.

HANNA JANKOWSKI
HKJ@MATHSTAT.YORKU.CA
DEPARTMENT OF MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, CANADA