Robust Design for Censored Exponential Data

M.T.Mirnazari and William.J.Welch
University of Waterloo

I.I.Q.P. Research Report

RR-96-07

November 11, 1996

Robust Design for Censored Exponential Data

M. T. Mirnazari and William. J. Welch
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

November 11, 1996

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Robust Design for Censored Exponential Data

ABSTRACT

We study the design of experiments for censored life tests in which failure times depend on several designable explanatory variables. When censored observations are present, information (expected, observed, etc.) depends on the unknown values of the model parameters. Also, different realizations of the same experiment might, just by chance, lead to quite different censoring patterns and hence different information for the parameters of interest. We study the design of experiments in this context by "simulated analysis", where data are generated by simulation and analyzed. By repeating this process, one can estimate the distribution of observed information that would arise from a given experimental plan. This forms the basis for a criterion to choose the design: i.e., look for a design with good information for the majority of the simulated data sets. The criterion can also be used to augment an existing design with extra runs. Parameter estimates from the existing data can be used to form a prior for generating the augmented runs.

Key words and phrases: Accelerated testing, failure time, lifetime data, optimal design, reliability improvement.

1 INTRODUCTION

In this article we study criteria for designs involving censored data in reliability experiments.

Denote the vector of explanatory variables by $\mathbf{x}$ and the experimental region by $\boldsymbol{\chi}$. We observe the random variable $Y = \min(T, L)$, where T is the lifetime and L is the censoring limit. The distribution of T depends on the design variables $\mathbf{x}$ and on a vector of unknown parameters, $\boldsymbol{\beta}$. The design problem is to choose n vectors $\mathbf{x} \in \boldsymbol{\chi}$ (not necessarily distinct) in order to efficiently estimate functions of the model parameters, $\boldsymbol{\beta}$, of interest.

For linear regression models without censoring, design criteria include A, D, E, and G optimality, (e.g., Silvey 1980). Observed and expected information coincide and the information matrix is the basis for defining a design criterion. D optimality, i.e., maximizing the determinant of the information matrix, has frequently been used as a criterion to construct optimal designs. In the absence of censoring, the information matrix does not depend on the unknown model parameters, facilitating design construction.

For censored lifetime experiments, however, the information matrix depends on the parameters, $\boldsymbol{\beta}$. A similar problem arises in nonlinear regression, where a common approach is to obtain an efficient design for the best guess of the parameter values. This is called locally optimal design, introduced by Chernoff (1953). Another method is to assign a (Bayesian) prior distribution to the unknown parameters. Ford et al. (1989) provided a comprehensive review of nonlinear design methodology. Atkinson, Demetrio, and Zocchi (1995), Chaloner and Larntz (1989), Fedorov (1972, Chapters 2.8 and 4.4), Silvey (1980, Chapters 6 and 7), and St. John and

Draper (1975) also considered designs for nonlinear models without censoring.

For lifetime experiments, both with and without censoring, the expected Fisher information matrix has been the basis for obtaining optimal plans. Chernoff (1962) constructed locally-optimal accelerated test plans in one stress factor for the exponential distribution. Nelson and Kielpinski (1976) compared optimal, standard, and compromise plans for accelerated testing with one stress factor and censored normal or lognormal data. Escobar and Meeker (1986a) gave an algorithm to compute the expected information matrix for the smallest extreme value distribution. (For exponential or Weibull data the logarithm of life time can be expressed in terms of an extreme value distribution.) Escobar and Meeker (1986b, 1995) obtained test plans for accelerated life tests with one or more experimental factors. Chaloner and Larntz (1992) used a prior distribution for the model parameters.

The above papers deal largely with design criteria based on expected information. With censored lifetimes, particularly if the number of runs is small, observed information may vary considerably with the censoring pattern. Expected information might average data realizations containing little or no information with more informative realizations. For this reason we propose the use of observed information. Efron and Hinkley (1978) argued in general for using the inverse of the observed information matrix rather than the expected information matrix to obtain a covariance matrix for parameter estimates.

At the design stage, though, the data, the parameter estimates, and hence the observed information are unavailable. Thus, to evaluate a given design we propose repeatedly generating realizations of the data and numerically determining the distribution of observed information. The realizations have variability due to uncertainty about $\boldsymbol{\beta}$ and due to the lifetime distribution. This procedure, which we call "simulated analysis", is repeated for each design considered by a design-optimization algorithm. The design with the best distribution of observed information is chosen; it will tend to avoid poor censoring patterns with low information. Nelson (1990, Chapter 6.3) advocated simulating data for a tentative test plan to check its suitability. We are taking this idea one step further and using simulation to choose the design. Generating realizations requires the true values of the unknown parameters, $\boldsymbol{\beta}$. We sample $\boldsymbol{\beta}$ from a prior distribution. When augmenting a design the prior for $\boldsymbol{\beta}$ can be based on the analysis of the available data.

The methods described here could be aimed at two types of experiment. First, the explanatory variables $\mathbf{x}$ could be designable engineering factors. The design criteria proposed in Section 3 lead to designs that are efficient for estimating $\boldsymbol{\beta}$; good estimation of these parameters leads straightforwardly to optimization of the engineering design. There seems to be little work on experiments for this objective. Secondly, the explanatory variables could be stress factors in accelerated testing (e.g., Nelson 1990), where the objective is to estimate percentiles of the lifetime distribution. While we do not directly address accelerated testing, our design criteria for efficient estimation of $\boldsymbol{\beta}$ probably also lead to good plans for percentile estimation. An example in Section 4 illustrates. The design criteria proposed could be readily adapted to address accelerated testing directly.

In Section 2 we introduce notation and give the observed and expected information matrices. We concentrate on the exponential lifetime distribution. Section 3 proposes design criteria and describes their numerical implementation. Section 4 provides two examples to demonstrate the methodology, including design augmentation. Finally, Section 5 makes some concluding remarks.

2 OBSERVED AND EXPECTED INFORMATION

For this paper we assume that each unit is tested for a limited period of time, L, i.e., we may have Type-I censoring of the observations. Unequal censoring times, or other types of censoring, could be handled with minor adaptation.

Suppose that the failure time, T, is a non-negative, continuous random variable with survival function $\mathcal{F}(t) = \Pr(T \geq t)$ for $0 \leq t \leq \infty$. The density and hazard functions are written as

$$f(t) = -\frac{d}{dt}\mathcal{F}(t) = \lim_{\Delta \to 0} \frac{\Pr(t \leq T \leq t + \Delta)}{\Delta}$$

and

$$h(t) = \lim_{\Delta \to 0} \Pr(t \leq T \leq t + \Delta \mid T \geq t) = \frac{f(t)}{\mathcal{F}(t)}.$$

Unit i in a sample of size n would have failure time T_i. Define the random variables $Y_i = \min(T_i, L)$ and δ_i, an indicator taking the value 1 for an uncensored (complete) observation and zero otherwise. The dependence of failure times on explanatory variables $\mathbf{x}$ involves unknown parameters $\boldsymbol{\beta}$. Thus, $\mathcal{F}$, f, and h are functions of $\mathbf{x}$ and $\boldsymbol{\beta}$. For data $y_1, \ldots, y_n$, the likelihood function can be written (e.g., Lawless 1982, Chapter 1) as

$$\mathcal{L} = \prod_{i=1}^{n} [f(y_i, \mathbf{x}_i, \boldsymbol{\beta})]^{\delta_i} \prod_{i=1}^{n} [\mathcal{F}(y_i, \mathbf{x}_i, \boldsymbol{\beta})]^{1-\delta_i}.$$

After simplification, the log likelihood can be written

$$\ell = \sum_{u} \log h(y_i, \mathbf{x}_i, \boldsymbol{\beta}) \ + \ \sum_{i=1}^{n} \log \mathcal{F}(y_i, \mathbf{x}_i, \boldsymbol{\beta}), \qquad (1)$$

where the first summation is over the uncensored observations.

Let $\mathbf{y}$ denote the vector $y_1, \ldots, y_n$ and let $\mathbf{X}$ similarly denote the experimental plan, i.e., row i of $\mathbf{X}$ contains $\mathbf{x}_i^T$. Thus, the log likelihood can be written as $\ell(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y})$. The information matrix is $I(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}) = -\partial^2 \ell / \partial \boldsymbol{\beta}^2$. The expected or Fisher information is obtained by taking expectation with respect to the distribution of Y. Alternatively, evaluating I at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$, gives the observed information matrix. For details on calculating MLE's for censored experiments see, for example, Lawless (1982, Appendix F).

The exponential distribution with rate λ has $f(t; \lambda) = \lambda e^{-\lambda t}$. To ensure a positive rate the dependence on the explanatory variables is expressed through

$$\lambda(\mathbf{x}) = \exp(\boldsymbol{g}(\mathbf{x})^T \boldsymbol{\beta}), \tag{2}$$

where $\boldsymbol{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_k(\mathbf{x}))^T$ contains k known functions of the explanatory variables, e.g., linear, quadratic, and interaction terms. The log likelihood becomes

$$\ell(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}) = \sum_u \boldsymbol{g}(\mathbf{x}_i)^T \boldsymbol{\beta} - \sum_{i=1}^n y_i e^{\boldsymbol{g}(\mathbf{x}_i)^T \boldsymbol{\beta}},$$

and element (r, s) of the information matrix is

$$[I(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y})]_{rs} = -\frac{\partial^2 \ell(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n y_i g_r(\mathbf{x}_i) g_s(\mathbf{x}_i) e^{\boldsymbol{g}(\mathbf{x}_i)^T \boldsymbol{\beta}}.$$

Thus, the information matrix can be written as

$$I(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}) = \boldsymbol{G}^T \mathbf{W} \boldsymbol{G},$$

where $\boldsymbol{G}$ is an $n \times k$ matrix with $\boldsymbol{g}(\mathbf{x}_i)^T$ in row i, and $\mathbf{W} = \mathrm{diag}[y_1 \lambda(\mathbf{x}_1), \ldots, y_n \lambda(\mathbf{x}_n)]$. We will refer to $\mathbf{W}$ as the weight matrix. The observed and expected information matrices are

$$I(\hat{\boldsymbol{\beta}}, \mathbf{X}, \mathbf{y}) = \boldsymbol{G}^T \hat{\mathbf{W}} \boldsymbol{G}, \tag{3}$$

and

$$\mathcal{I}(\boldsymbol{\beta}, \mathbf{X}) = \mathrm{E}_Y\left[I(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y})\right] = \boldsymbol{G}^T \Pi \boldsymbol{G}, \tag{4}$$

where the estimated weights $\hat{w}_i = y_i \hat{\lambda}(\mathbf{x}_i)$ are based on the MLE's of $\boldsymbol{\beta}$, and $\Pi = \mathrm{diag}(\pi_1, \ldots, \pi_n)$ with π_i being the probability that unit i fails before time L.

3 DESIGN CRITERIA AND NUMERICAL IMPLEMENTATION

As noted in Section 1, expected information has often been used as a design criterion. For a design $\mathbf{X}$, the D-optimality criterion based on the expected information matrix is defined as

$$\Phi(\mathbf{X}) = E_{\boldsymbol{\beta}} \left| E_Y \left[I(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}_{\boldsymbol{\beta}}) \right] \right|^{1/k} \tag{5}$$

and called D_{exp}-optimality below. Normalizing the determinant by raising it to the power $1/k$ means that the criterion doubles if a design $\mathbf{X}$ is replicated to give $2n$ runs and the same data occur again. Thus, relative efficiencies can be interpreted in terms of equivalent numbers of observations. Note that the inner expectation with respect to the lifetime distribution can be obtained from (4). In general, the outer expectation with respect to a prior or weight function for $\boldsymbol{\beta}$ is computed numerically. Chaloner and Larntz (1989) used a similar criterion (on a log scale) for obtaining optimal approximate designs for logistic regression models with no censored observations. (In this article, we will be concerned only with exact, i.e., integer replication, designs including censored observations.)

We propose a criterion based on simulated analysis that takes account of uncertainty in $\boldsymbol{\beta}$ *and* in the data, rather than taking expectations with respect to Y. We generate m samples, $(\boldsymbol{\beta}_1, \mathbf{y}_1), \ldots, (\boldsymbol{\beta}_m, \mathbf{y}_m)$, where the dependence of $\mathbf{y}_j$ on $\mathbf{X}$ and on $\boldsymbol{\beta}_j$ is omitted in this notation. Here, $\mathbf{y}$ is a vector of length n, but if we are augmenting n_0 previous observations we need simulate only the $n - n_0$ new observations. For simulated sample or realization j, we compute the observed information, $I(\hat{\boldsymbol{\beta}}_j, \mathbf{X}, \mathbf{y}_j)$, given in (3). The D-optimality criterion for this sample (and design $\mathbf{X}$) is

$$D_j = |I(\hat{\boldsymbol{\beta}}_j, \mathbf{X}, \mathbf{y}_j)|^{1/k}, \tag{6}$$

where k is the number of parameters in the vector, $\boldsymbol{\beta}$.

The distribution of D_j values over the m samples needs to be reduced to a scalar quantity for design optimization. One possibility, called D_{ave} by Mirnazari (1995), is to take the average value,

$$\bar{D} = \frac{1}{m} \sum_j D_j. \tag{7}$$

This estimates the criterion

$$\Phi(\mathbf{X}) = E_{\boldsymbol{\beta}} E_Y |I(\hat{\boldsymbol{\beta}}_j, \mathbf{X}, \mathbf{y}_j)|^{1/k}.$$

Compared with D_{exp} in (5), D_{ave} reverses the order of taking the determinant and the expectation with respect to Y. Moreover, it uses observed information.

In practice, only one experiment is to be performed, and an experimenter would like good information for all, or at least most, realizations. Borrowing Taguchi's (1986, Chapter 8) larger-the-better loss function for robust product and process design, we suggest minimizing

$$\frac{1}{m} \sum_j \frac{1}{D_j^2}$$

over possible designs. This criterion will tend to give a distribution of D_j values with large average and small standard deviation. It heavily penalizes designs with any very poor realizations, i.e., with $D_j \simeq 0$.

Some numerical experimentation showed, however, that it may be impossible to find a design that avoids $D_j \simeq 0$ for all m realizations. To avoid domination of the criterion by a few very poor realizations, we sort $D_1, \ldots, D_m$ to give $D_{(1)} \geq \cdots \geq D_{(m)}$ and minimize

$$\frac{1}{m'} \sum_{j=1}^{m'} \frac{1}{D_{(j)}^2}, \tag{8}$$

where $m' < m$. In the examples of Section 4 we take $m' = 0.95m$. Thus, we aim to find a design such that the best 95% of realizations have good average information with small variability around the average. We call this criterion D_{rob}, as it seeks a design robust to variability from uncertainty in $\boldsymbol{\beta}$, from the lifetime distribution, and hence from the censoring pattern.

We now summarize the steps in computing the D_{rob}-optimal designs:

1. Simulate m random samples. (Numerical trial and error suggests that $m = 100$ is adequate.) Sample j has k parameters, $\boldsymbol{\beta}$, sampled from a possibly correlated prior distribution or weight function and n independent Uniform[0, 1] random variables, $\boldsymbol{u}$. Although the generated data will change with the design during design optimization, the same random $\boldsymbol{\beta}$ and $\boldsymbol{u}$ vectors will be used to generate the exponential lifetimes. This avoids problems that would occur with the design optimization algorithm if different random perturbations affect the criterion each time it is computed. The m samples $(\boldsymbol{\beta}_1, \boldsymbol{u}_1), \ldots, (\boldsymbol{\beta}_m, \boldsymbol{u}_m)$ are computed as a Latin hypercube sample (McKay, Conover, and Beckman 1979). Compared with pure Monte Carlo this gives a more systematic coverage of the prior distribution for $\boldsymbol{\beta}$ and of the uniform distribution for data generation.

2. Simulate the data. For each sample j we use the design $\mathbf{X}$ and the vectors $\boldsymbol{\beta}_j$ and $\boldsymbol{u}_j$ to generate n observations. Observation i has rate $\lambda(\mathbf{x}_i) = \exp(\boldsymbol{g}(\mathbf{x}_i)^T \boldsymbol{\beta}_j)$ as in (2). Lifetime t_i is given by $-\log(u_{ij})/\lambda(\mathbf{x}_i)$, where u_{ij} is element i of $\boldsymbol{u}_j$, and $y_i = \min(t_i, L)$. If n_0 observations are already available from a previous experiment, only the $n - n_0$ new observations are simulated.

3. Fit the parameters by maximum likelihood. For simulated data set j we find the maximum likelihood estimates, $\hat{\boldsymbol{\beta}}_j$. Currently, we use a conjugate gradient algorithm (Press et al. 1992, Chapter 10.6).

4. Compute the criterion. For sample j we compute the determinant of the observed information matrix, D_j, from (6). The D_{rob} criterion is then computed using (8).

5. Optimize the design. Steps 2–4 are carried out for every design, $\mathbf{X}$, tried during minimization of the design criterion. We currently use a modified Fedorov exchange algorithm (Cook and Nachtsheim 1980). This numerical optimizer is not guaranteed to find the best design; we typically make three optimization tries from different random starts and choose the best.

At Step 3, due to the particular censoring pattern, there may be some problems with maximum likelihood estimation of $\boldsymbol{\beta}$ for some designs and data realizations. Often the likelihood is maximized and flat as some of the $\boldsymbol{\beta}$ parameters go to infinity. This causes no practical problems. The numerical optimizer meets the convergence tolerance for large, finite values of the particular $\boldsymbol{\beta}$ parameters, where the determinant of the observed information matrix is close to zero. The D_{rob} criterion penalizes designs with more than a few realizations having this behaviour, and the optimizer will seek better designs.

At Step 4, other criteria based on a simulated analysis could be substituted. For example, instead of D_j in (6) we could compute the estimated variance of a percentile of the lifetime distribution. To obtain consistently good variances across realizations, the m variances would be combined in a way similar to D_{rob} in (8).

For comparison, in Section 4 we also present designs based on expected information. To compute D_{exp} in (5), the expectation over $\boldsymbol{\beta}$ is estimated by averaging m samples from the prior as in Step 1. The same m samples are used for computing D_{exp} and D_{rob} to make the comparisons fairer. The $\boldsymbol{u}$ samples are not required, nor is maximum likelihood estimation carried out in Step 3.

The computations presented here were made using software of Welch (1996).

4 EXAMPLES

Two examples are provided here. The first illustrates the differences between properties of D_{rob}- and D_{exp}-optimal designs. The second demonstrates augmentation of a design for an accelerated life test. The reader is also referred to Mirnazari (1995, Chapter 3) for examples using the D_{ave} criterion in (7).

4.1 Comparison of the D_{rob} and D_{exp} Criteria

Consider an experiment with two explanatory variables from $[-1, +1]^2$ and $n = 8$ runs. The rate function (2) is assumed to be $\lambda(\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$. To investigate the effect of censoring, we try $L = 50$, 75, 100, and 200.

We take independent normal priors for the three parameters with means and standard deviations as follows: β_0 has a $N(-4\frac{1}{2}, \frac{1}{6^2})$ distribution, while β_1 and β_2

are $N(1, \frac{1}{3^2})$. These choices, in conjunction with the L values tried, give fairly low to fairly high censoring probabilities. At the prior means, for $L = 50$ the censoring probability ranges from 0.017 (at $x_1 = x_2 = 1$) to 0.928 at ($x_1 = x_2 = -1$), whereas for $L = 200$ it ranges from 0.000 to 0.740.

The designs produced by the D_{rob} and D_{exp} criteria are shown in Figure 1. The design points have been "jittered", i.e., moved a small random distance, so that replicates are not overplotted. For both criteria, $m = 100$ samples were taken. For D_{rob}, we accumulated the $m' = 95$ best realizations when computing (8). It is seen that both criteria avoid the $x_1 = x_2 = -1$ corner where there is the largest probability of censoring. For all values of L, the D_{exp}-optimal designs are supported only at the other three corners. In contrast, for $L = 75$ and $L = 50$ the D_{rob}-optimal designs have support at the same three corners *and* at further locations with less censoring. Evaluation of alternative designs shows that the D_{exp}-optimal designs are not unique. For example in Figure 1, the design for $L = 50$ has only two points at (-1, +1). Putting three points at each of (-1, +1) and (+1, -1) with the other two points at (+1, +1) gives the same D_{exp} criterion value. We have not experienced similar nonuniqueness for D_{ave} or D_{rob}-optimal designs.

For each value of L, Figure 2 compares the properties of the two designs. Their distributions over all $m = 100$ prior-data realizations for the determinant of the observed information matrix in (6) are compared via a smoothed density plot and via a Q-Q plot. Note that the distribution of the normalized determinant, D, is bimodal for the D_{exp}-optimal designs when $L = 100, 75$, or 50. There is a proportion of very uninformative realizations, presumably because of poor censoring patterns. The D_{rob}-optimal design avoids many of these poor realizations. The two designs

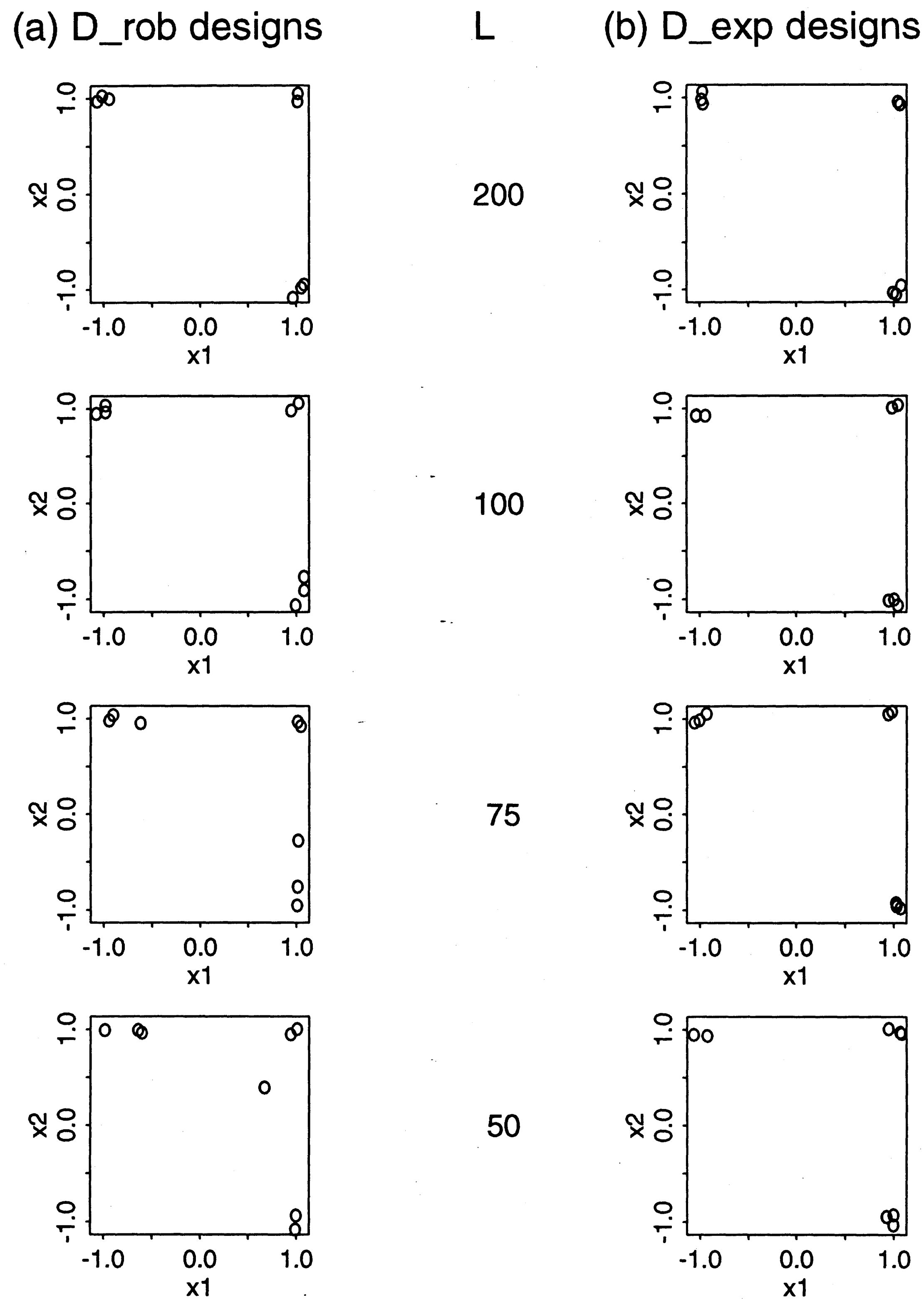

Figure 1: Optimal designs for two explanatory variables with eight runs and censoring limit $L = 200$, 100, 75, or 50: (a) D_{rob}-optimal designs and (b) D_{exp}-optimal designs.

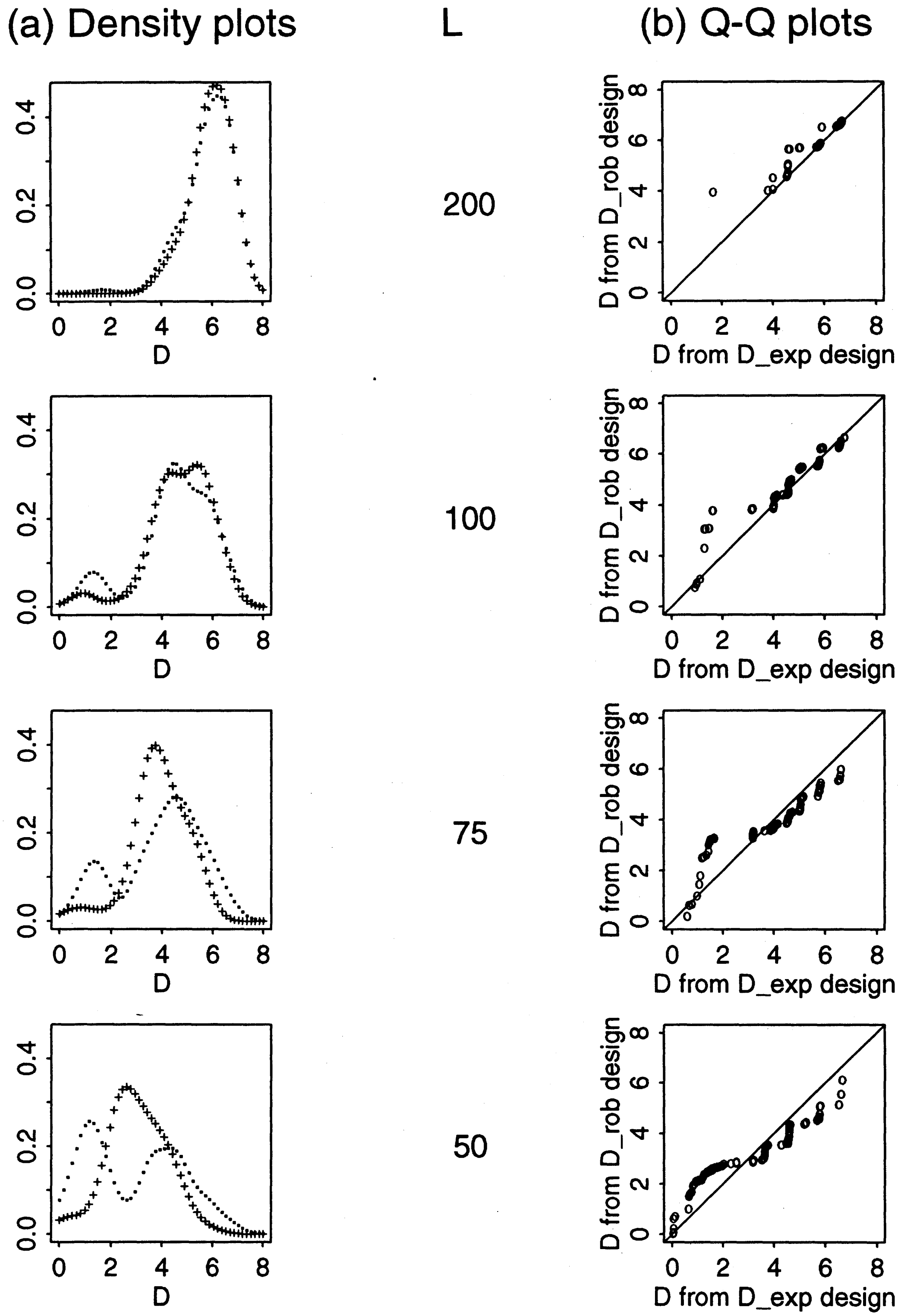

Figure 2: Comparison of $D_j = \left|I(\hat{\boldsymbol{\beta}}, \mathbf{X}, \mathbf{y}_j)\right|^{1/k}$ for 100 realizations from the D_{rob}- and D_{exp}-optimal designs: (a) Smoothed density plots (plotting symbol "+" for D_{rob} and "." for D_{exp}) and (b) Q-Q plots.

have similar *average* determinants, but the larger-the-better criterion motivating D_{rob} produces less variability around the average.

While this is just one example, it illustrates some typical qualitative features. In general our experience (see also Mirnazari 1995) shows that in comparison with D_{exp}, the D_{rob}-optimal designs tend to have:

- Support points favouring subregions of χ with the least probability of censoring. When many observations are censored a very poor determinant results; D_{rob} optimality penalizes designs with more than a few poor determinants. Thus, D_{rob}-optimal designs tend to have realizations with fewer censored observations.

- More support points.

- A unique design contrary to the lack of uniqueness for the D_{exp}-optimal designs.

- Greater sensitivity to the censoring limit, L. (Often the same D_{exp}-optimal design is generated for a wide range of L values.)

- High efficiency when evaluated under the D_{exp} criterion. Conversely, D_{exp}-optimal designs perform poorly when evaluated under D_{rob} unless there is little censoring.

4.2 Electrical Insulation Example

Nelson (1972, 1975) gave a set of data from an experiment on the breakdown of an insulating fluid subjected to various elevated test voltages, v. The data are also reported in Lawless (1982, Chapter 4.3). The main purpose of the accelerated life test was to estimate the distribution of the breakdown time at $v = 20KV$. Nelson (1975) used a power law model, i.e., a Weibull distribution with shape parameter σ and scale parameter $\mu(v) = cv^p$. Taking the explanatory variable $x = \log(v)$, his analysis of the data showed that the Weibull shape parameter is close to 1 and hence an exponential model is appropriate.

In the original experiment all 76 items on test failed, i.e., there was no censoring. To introduce censoring, the topic of this article, we consider a limit of $L = 50$. This would have given data with 17 units censored out of 76. Fitting an exponential model to the censored data and plotting residuals (adjusted for censoring) versus the exponential ordered statistics (Lawless 1982, Chapter 6) gives a roughly straight line, confirming the exponential model suggested for the original data.

We now consider augmenting the 76 (censored) observations with a second-stage experiment putting a further 24 items on test. The rate function (2) is assumed to be $\log(\lambda(v)) = \beta_0 + \beta_1 \log(v)$. The experimental region is the same as that for the

Parameter	Mean	Standard deviation
β_0	-80.85	6.04
β_1	22.17	1.71

Table 1: Prior means and standard deviations for the second-stage experiment.

first-stage experiment, i.e., $v = 26, 28, \ldots, 38$. Table 1 gives the means and standard deviations of the normal prior for the second stage based on the parameter estimates and standard errors from the first-stage analysis. When generating the $m = 100$ Latin hypercube samples from the prior distribution we also took account of the correlation of -0.999768 between $\hat{\beta}_0$ and $\hat{\beta}_1$. Iman and Conover (1982) described a method for obtaining a target covariance structure; iterating their procedure gives a sample correlation of -0.999843. For the 100 data realizations, the lifetimes for only the 24 new runs were simulated.

The design found using the D_{rob} criterion has 22 new runs at $v = 38$, one run at 30, and one run at 32. This is in contrast to the first-stage design which had fewer runs at $v = 38$ than at 30–36. The distribution of the simulated D_j values [see (6)] is shown in Figure 3(a). The corresponding value for the first 76 (censored) runs is 3.85, shown by an asterisk. The 24 new runs make more improvement in the determinant than would be expected from the change in sample size, presumably because the designs for the new and old runs have very different patterns.

Using least squares estimates of the parameters, Nelson (1975) estimated the first percentile and the mean life time at $x_{20} = \log(v = 20)$. Standard errors are calculated from normal approximations. From the model, the log of the first percentile of the life distribution is

$$\log(t_{0.01}) = \log(\mu) + \log[-\log(0.99)]$$

while the log of the mean at x_{20} is

$$\log(\mu) = \log(1/\lambda) = -\beta_0 - x_{20}\beta_1.$$

These quantities can be estimated using the maximum likelihood estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, for each of the 100 data sets simulated for the D_{rob}-optimal design. The standard error of the estimated log mean lifetime can be obtained from

$$\text{Var}((\widehat{\log(\mu)}) = \text{Var}(\boldsymbol{g}(x)^T\hat{\boldsymbol{\beta}}) = \boldsymbol{g}(x)^T V(\hat{\boldsymbol{\beta}})\boldsymbol{g}(x),$$

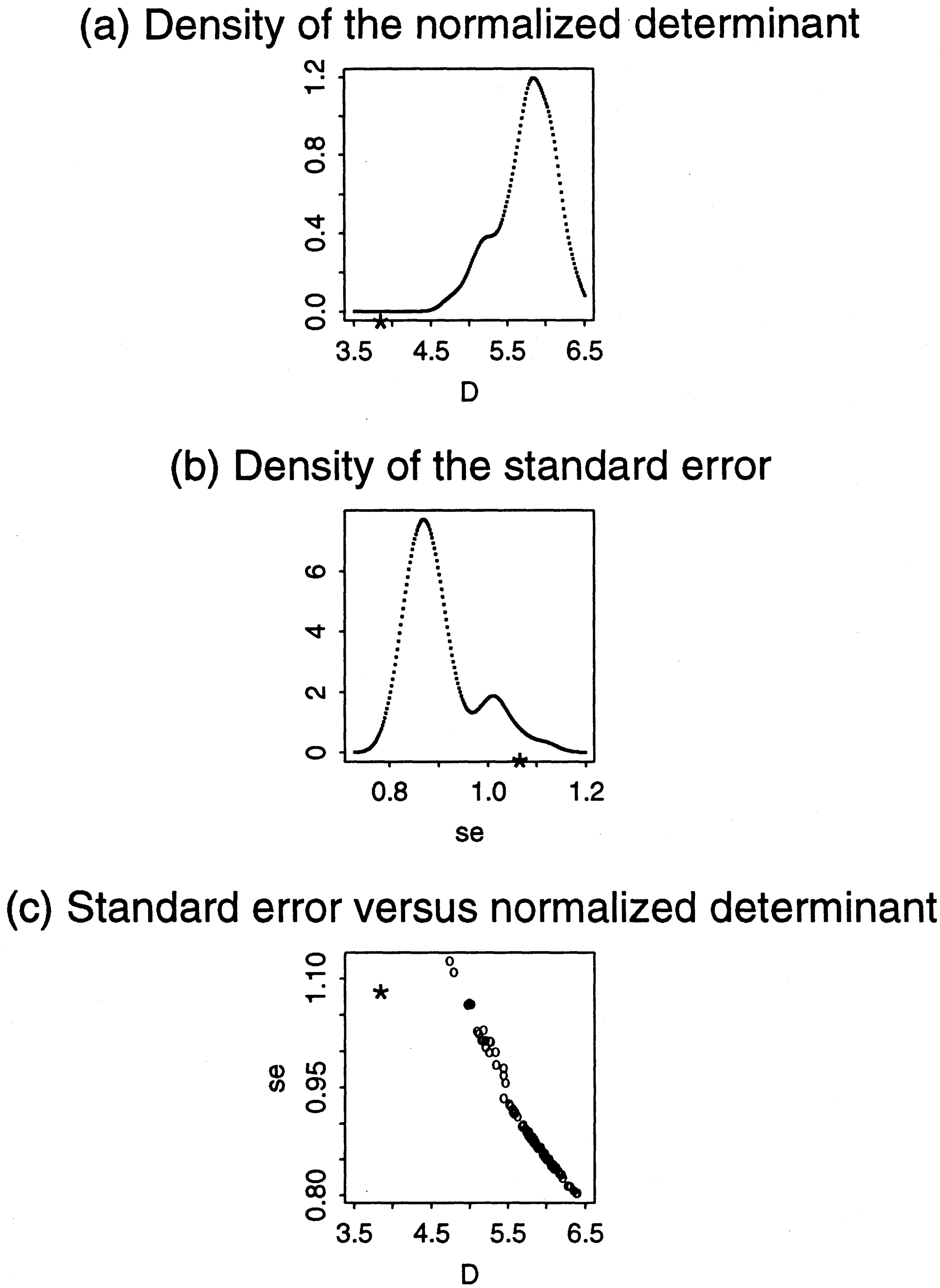

Figure 3: (a) Smoothed density of $D = \left|I(\hat{\boldsymbol{\beta}}, \mathbf{X}, \mathbf{y})\right|^{1/k}$, (b) Smoothed density of the standard error of $\widehat{\log(\mu)}$, and (c) The standard error of $\widehat{\log(\mu)}$ versus D. Asterisks indicate values for the first 76 runs only.

where $\boldsymbol{g}(x) = (1, x_{20})^T$ and $V(\hat{\boldsymbol{\beta}})$ is the asymptotic covariance matrix from the inverse of the observed information matrix. A similar method gives a standard error for $\log(\widehat{t_{0.01}})$. Figure 3(b) shows the estimated (smoothed) density plot for the standard error of $\widehat{\log(\mu)}$ from the $m = 100$ samples. Analysis of the first-stage 76-run design (with censoring) gives a standard error of 1.07. Only two of the $m = 100$ simulated 100-run data sets give a standard error slightly larger than 1.07 (due to changes in the parameter estimates). Most realizations substantially reduce the standard error. Finally, as shown in Figure 3(c), the standard error of $\widehat{\log(\mu)}$ is nearly linearly related to the normalized determinant, D, for the 100 realizations. Thus, using the standard error in the design criterion instead of D would make little difference here.

5 CONCLUSIONS

For small experimental plans, censoring introduces substantial variability. By chance, a censoring pattern might arise that has little or even no information for parameter estimation. This uncertainty is in addition to the dependence of the information matrix on the parameters being estimated. We have proposed simulating data sets that might arise and carrying out analyses based on observed information. By searching for a design that does well for most realizations, the experimental plan can obviate some of the uncertainty at the design stage.

Specifying a prior for the unknown parameters is easier in a follow-up experiment, where an empirical prior from the first-stage can be used. In design augmentation, expected-information criteria would mix observed information from the first-stage with expected information from the second. An approach based on observed information for all runs seems even more compelling here.

Further work on design for censored Weibull lifetimes will be reported elsewhere.

RERERENCES

Atkinson, A. C., Demetrio, C.G.B., and Zocchi, S.S. (1995). Optimum dose levels when males and females differ in response. *Appl. Statist.,* **44**, 213–226.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference,* **21**, 191–208.

Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. *J. Statist. Plann. Inference,* **33**, 245–259.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.,* **24**, 586–602.

Chernoff, H. (1962). Optimal accelerated life designs for estimation. *Technometrics,* **4**, 381–408.

Cook, R.D. and Nachtsheim, C.J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics,* **22**, 315–324.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika,* **65**, 457–487.

Escobar, L.A. and Meeker, W.Q. (1986a). Elements of the Fisher information matrix for the smallest extreme value distribution and censored data. *Appl. Statist.,* **35**, 80–86.

Escobar, L.A. and Meeker, W.Q. (1986b). Planning accelerated life tests with Type II censored data. *J. Statist. Comput. Simul.,* **23**, 273–297.

Escobar, L.A. and Meeker, W.Q. (1995). Planning accelerated life tests with two or more experimental factors. *Technometrics,* **37**, 411–427.

Fedorov, V.V. (1972). *Theory of optimal experiments.* Translated and edited by W. J. Studden and E. M. Klimko. Academic Press, New York.

Ford, I., Kistos, C.P., and Titterington, D.M. (1989). Recent advances in nonlinear experimental design. *Technometrics,* **31**, 49–60.

Iman, R.L. and Conover, W.J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Comm. Statist. Simul. Computation,* **11**, 311–334.

Lawless, J.F. (1982). *Statistical models and methods for lifetime data.* Wiley, New York.

McKay, M.D., Conover, W.J., and Beckman, R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics,* **21**, 239–245.

Mirnazari, M.T. (1995). *Optimal design of experiments: Criterion-robustness and censored data.* PhD thesis, University of Waterloo, Waterloo, Ontario.

Nelson, W. (1972). Graphical analysis of accelerated life test data with the inverse power law model. *IEEE Trans. Reliabil,* **R-21**, 2–11.

Nelson, W. (1975). Analysis of accelerated life test data—least squares methods for the inverse power law model. *IEEE Trans. Reliabil,* **R-24**, 103–107.

Nelson, W. (1990). *Accelerated testing: Statistical models, test plans, and data analyses.* Wiley, New York.

Nelson, W. and Kielpinski, T. J. (1976). Theory for optimum censored accelerated life tests for normal and lognormal life distributions. *Technometrics,* **18**, 105–114.

Press, W. H., Teukolsky, S.A., Vetterling, W.T., and Flanney, B.P. (1992). *Numerical recipes in C : The art of scientific computing.* Second Edition. Cambridge University Press, New York.

Silvey, S.D. (1980). *Optimal design: an introduction to theory for parameter estimation.* Chapman and Hall, New York.

St. John, R. C. and Draper, N. R. (1975). D-optimality for regression designs: A review. *Technometrics,* **17**, 15–23.

Taguchi, G. (1986). *Introduction to Quality Engineering.* Asian Productivity Organization, Tokyo.

Welch, W.J. (1996). *ACED: Algorithms for the Construction of Experimental Designs.* unpublished software.