Analysis of Experiments for Reliability Improvement and Robust Reliability

M.S. Hamada
University of Waterloo

RR-94-07

May 1994

ANALYSIS OF EXPERIMENTS FOR RELIABILITY IMPROVEMENT AND ROBUST RELIABILITY

Michael S. Hamada[1]
The Institute for Improvement in Quality and Productivity
University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

ABSTRACT

Statistically designed experiments have been used extensively for estimating or demonstrating existing reliability but have seldom been used for improving reliability. Genichi Taguchi has advocated their use not only for improving reliability but also for achieving robust reliability. Robust reliability is part of his robust design philosophy whose aim is to make processes/products insensitive to "noise" factors which are hard or impossible to control such as manufacturing variables that cannot easily be controlled or environmental conditions in which the product is operated. This paper first discusses experimental designs for reliability improvement and robust reliability. Then analysis methods for reliability data are considered. These methods need to be able to handle censored data which commonly occur because all units tested have not failed by the end of the experiment. In analyzing censored data from these experimental designs, some difficulties with standard methods have been encountered and have provided the motivation for recent work. This paper presents an overview of analysis methods which include standard methods, an iterative imputation-based model selection procedure and one which takes a Bayesian approach. Examples of fluorescent lamps, heat exchangers and drill bits are given to illustrate the use of experimental design for improving reliability. The data from these experiments are reanalyzed to show how some of the analysis methods can be applied.

Key words: Bayesian, Censoring, Complex aliasing patterns, Control and noise factors, Estimability, Fractional factorial, Maximum likelihood estimation, Mixed-level orthogonal array, Plackett-Burman design, Robust design, Product array, Response surface design.

---

[1]This research was supported by General Motors of Canada Limited, the Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada.

1. INTRODUCTION

While statistically designed experiments have been used extensively for estimating or demonstrating existing reliability (Nelson 1982), they have seldom been used proactively to improve reliability as advocated by Genichi Taguchi (Taguchi 1986, 1987), i.e., to identify factors that affect reliability and to recommend factor levels that lead to improved reliability. Taguchi is better known for robust design, whose aim is to make processes/products insensitive to "noise" factors which are hard or impossible to control. Such products/processes are said to be *robust* to the noise factors. Examples of noise factors include manufacturing variables that cannot easily be controlled and environmental conditions in which the product is used. This important paradigm for improving products/processes has attracted much attention in recent years (Kackar 1985) and can be applied to reliability. In order to ensure good stability and adequate life, Taguchi (1986, page 149) recommended that noise factors be considered in any experiment to improve reliability whenever practical to do so.

It is somewhat surprising that statistically designed experiments have not received more attention as a means for improving reliability. Methods for analyzing lifetime data which handle censored data (e.g., arising from units which have not failed by the end of the experiment) existed as early as 1959. See Sampford and Taylor (1959) and Zelen (1959) which discussed, respectively, how maximum likelihood estimation could handle right-censoring (or Type I censoring) and Type II censoring in factorial experiments. Zelen (1959) looked at the effect of two accelerating factors, temperature and voltage, on capacitor lifetime, however, so that the focus was not improvement. Hitzelberger (1967) in an article, "Improve Your Reliability", presented factorial experiments as a way to establish cause-and-effect relationships between factors and a product/process characteristic. Unfortunately, the example he gave dealt with quality rather than reliability. While there has been isolated use of such experiments in the years following, the industrial statistical literature has been rather silent on this matter.

It is in the 1980's with North American industry's introduction to Taguchi's quality engineering philosophy and methodology (first Taguchi and Wu 1980 and later Taguchi 1986, 1987) that we find a clear message to use designed experiments to improve reliability. In his books, Taguchi provides examples of improving clutch spring and fluorescent lamp reliability. Furthermore, there is documented evidence in the *Symposia on Taguchi Methods* (1984-1993) that industry has heard the message. Specht (1985) reported on the improvement of heat-exchanger reliability in a commercial heating system. Montmarquet (1988) discussed the improvement of drill bit reliability in a multilayer printed circuit board drilling operation. Phadke (1986) also reported an early application of Taguchi's methodology at AT&T which improved router-bit reliability in a printed circuit board cutting operation. The message has recently made its way into textbooks on reliability. O'Connor (1991) in his third edition has a chapter on designed experiments which is apparently influenced by Taguchi. No reliability improvement examples are given, however. In Condra's (1993) *Reliability Improvement of Design of Experiments* we find the first textbook which is entirely devoted to this subject and in which Taguchi's robust design philosophy figures prominently. In assessing the use of designed experiments for reliability improvement, Condra (1993 page 127) concludes that they have not been used widely by reliability engineers in the past few decades. As for robust design, he states that it is a potentially powerful tool whose exploitation for reliability improvement is only beginning.

As North American industry began applying Taguchi's methods, his philosophy and methods also attracted the attention of researchers. Various studies were undertaken that generated new lines of research which included improved experimental designs and analysis methods for implementing his philosophy. Regarding experiments to improve reliability, analysis of censored data from them has presented new challenges. Besides right-censored data from units not failing by the end of the experiment, other types of censored data arise when units are inspected periodically for failure. In such situations in which units cannot be monitored continuously, units produce left-censored data if they fail before the first inspection and interval-censored data, otherwise. It is the censored data coupled with the moderately to highly fractionated designs commonly used in industry to study a large number of factors in a small number of runs that causes problems when standard methods are used. In order to overcome these problems, Hamada and Wu (1991) proposed an iterative procedure based on building up a model. Problems with standard methods were further explored by Hamada and Tse (1992), which provided support for the strategy used in the iterative procedure. Hamada (1993) also explored the analysis of robust reliability experiments with censored data. Recently, Hamada and Wu (1992b) proposed a Bayesian approach which overcomes limitations of the iterative procedure and also provides a natural framework for analyzing robust reliability experiments.

This article focuses on the analysis methods for experiments to improve reliability and to achieve robust reliability and is organized as follows. First two classes of experimental designs, fractional factorials and product arrays, are discussed with examples in Sections 2 and 3, respectively. In Section 4, an overview of analysis methods is given. Some of these methods are demonstrated in Sections 5 through 7 which presents respective analyses of experiments to improve fluorescent lamp and heat exchanger reliability and to achieve robust reliability of drill bits in printed circuit board fabrication. Section 8 concludes with a discussion and comments on other types of reliability improvement experiments.

2. EXPERIMENTS FOR IMPROVING RELIABILITY

Experiments for improving reliability have the following goals: 1) identify the important factors that affect the reliability of a product/process and 2) choose levels of these factors that lead to improved reliability. Like other quality characteristics, the relationship between various factors and reliability can be studied using an appropriately chosen experimental design. Typically in industry, a large number of factors may need to be studied in a relatively small number of runs. Thus, highly fractionated 2^{k-p} designs (Box, Hunter and Hunter 1978) or non-geometric Plackett-Burman (1946) designs such as their 12-run design are often used. In a sequential approach to experimentation, these designs would initially be used for screening out the unimportant factors before conducting a follow-up experiment with more levels so that the response-factor relationship can modeled in more detail. In reality, the initial experiment may be the only one performed so that a proper assignment of factors for the 2^{k-p} designs can allow some potential interactions to be studied. Hamada and Wu (1992a) has also shown that some information on interactions may also be obtained from the Plackett-Burman designs. Taguchi often initially uses designs with more levels. These include the 3^{k-p} designs and mixed-level designs such as the L18(2×3^7), which can be used to study one two-level factor and up to seven three-level factors. Dey (1985) and Wang and Wu (1991) catalogue other mixed-level designs. In a sequential experimental approach, however, a design with more levels such as a response surface design (Box and Draper 1987) which usually is not a fractional factorial would be used in a follow-up experiment.

The applications given in the Introduction illustrate the use of some of these designs. Specht (1985) used a 12-run Plackett-Burman design to study ten factors (denoted by A-H,J,K) chosen from many possible product design, material selection and manufacturing factors to improve heat exchanger reliability. See Table 1 which gives the design and interval-censored lifetime data from eight inspections. The interval (93.5,105) indicates that the unit failed between 9350 and 10500 cycles, i.e., between the fifth and sixth inspections. $(128,\infty)$ indicates that the unit was still working at 12800 cycles, the last inspection. Note that one unit was still functioning at the last inspection.

Taguchi (1987 page 930) presented an experiment to improve the lifetime of fluorescent lamps which used a 2^{k-p} design. Five two-level factors denoted by A-E were studied using a twice replicated 8-run experiment with defining relations D=AC and E=BC which was conducted over 20 days with inspections every two days. The design and lifetime data appear in Table 2. Besides the main effects, the experimenter also thought that the AB(=DE) interaction might be potentially important. Note that there are right-censored data because seven of the 16 lamps had not failed by the 20 day inspection

Examples of other designs discussed above include the clutch spring experiment in Taguchi (1986, chapter 9) which used a 3^{10-7} to study seven factors. Phadke used a mixed-level 32-

Table 1: Design and Lifetime Data for the Heat Exchanger Experiment

Factor											
F	B	A	C	D	E	G	H	J	K	U	Lifetime
1	1	1	1	1	1	1	1	1	1	1	(93.5, 105)
1	1	1	1	1	2	2	2	2	2	2	(42, 56.5)
1	1	2	2	2	1	1	2	2	2	1	(128, ∞)
1	2	1	2	2	2	2	1	1	2	1	(56.5, 71)
1	2	2	1	2	1	2	1	2	1	2	(56.5, 71)
1	2	2	2	1	2	1	2	1	1	2	(0, 42)
2	1	2	2	1	2	2	1	2	1	1	(56.5, 71)
2	1	2	1	2	2	1	1	1	2	2	(42, 56.5)
2	1	1	2	2	1	2	2	1	1	2	(82, 93.5)
2	2	2	1	1	1	2	2	1	2	1	(82, 93.5)
2	2	1	2	1	1	1	1	2	2	2	(82, 93.5)
2	2	1	1	2	2	1	2	2	1	1	(42, 56.5)

run design to study two four-level factors and seven two-level factors. We also know of one study that used a Box-Behnken response surface design (Box and Draper 1987) as a follow-up experiment to investigate four factors previously identified in a 2^{k-p} screening design.

The data from these designs can be analyzed using a parametric model such as a lognormal or Weibull regression model. These models and their analyses are presented in Section 4. Analyses of the fluorescent lamp and heat exchanger experiments will be discussed in Sections 5 and 6, respectively.

Table 2: Design and Lifetime Data for the Fluorescent Lamp Experiment

Factor					Lifetime	
A	B	C	D	E		
1	1	1	1	1	(14,16)	(20,∞)
1	1	2	2	2	(18,20)	(20,∞)
1	2	1	1	2	(08,10)	(10,12)
1	2	2	2	1	(18,20)	(20,∞)
2	1	1	2	1	(20,∞)	(20,∞)
2	1	2	1	2	(12,14)	(20,∞)
2	2	1	2	2	(16,18)	(20,∞)
2	2	2	1	1	(12,14)	(14,16)

3. EXPERIMENTS FOR ACHIEVING ROBUST RELIABILITY

Taguchi's robust design is also referred to as parameter design because its objective is to find levels of engineering parameters (called control factors here) that yield a robust product/process, i.e., that make the product/process insensitive to the variation of hard or impossible to control noise factors. Taguchi's tactics for carrying out robust design are to specify a criterion for assessing noise factor effects and then use experimentation to estimate the criterion. Note that while noise factors are difficult or impractical to control in production or in use, for purposes of the experiment (i.e., to learn about the effect of the noise factors), the noise factors need to be controlled during the experiment. Following the notation used in Welch, Yu, Kang and Sacks (1990), a criterion for assessing the effect of the noise factors (termed the loss and denoted by $L(\cdot)$) at a particular combination of control factor levels $\mathbf{x}_{control}$ can be defined for a general loss function $l(\cdot)$ as:

$$L(\mathbf{x}_{control}) = \int l(Y(\mathbf{x}_{control}, \mathbf{x}_{noise}))f(\mathbf{x}_{noise})d\mathbf{x}_{noise} \ , \tag{1}$$

where $Y(\mathbf{x}_{control}, \mathbf{x}_{noise})$ is the random quality characteristic observed at a particular combination of control and noise factor levels $(\mathbf{x}_{control}, \mathbf{x}_{noise})$ and $f(\cdot)$ is the joint probability density function of the noise factors. In this formulation, the objective of robust design is to find a product/process design $\mathbf{x}_{control}$ with minimum loss. In applying robust design to reliability, Y is the lifetime random variable; some appropriate loss functions will be discussed later in the example given in Section 7.

Taguchi (1987) proposed using experimentation to estimate the loss (1) and modelling the estimated losses in terms of the control factors. Taguchi recommends using specialized experimental plans referred to as product (or crossed) arrays. A product array consists of two plans or arrays, one for the control factors called the "control array" and the other for the noise factors called the "noise array". The product or crossed array design is so named because all the noise factor combinations specified by the noise array are experimented at every combination of the control factors specified by the control array.

As an example, consider an experiment for improving the lifetime of drill bits (i.e., number of holes drilled before breakage) used in fabricating multilayer printed circuit boards (Montmarquet 1988). In designing multilayer circuit boards, small diameter holes are desired because they allow more room for the circuitry. The strength of small diameter drill bits is greatly reduced, however, so that breakage becomes a serious problem; broken bits cannot be removed from the boards requiring the boards to be scrapped. A product array consisting of a 16-run control array and an eight run noise array was used to study 11 control factors (A at four levels and B–J and L at two levels) and five noise factors (M–Q at two levels) as displayed in Table 3. The control factors were selected material composition and geometric characteristics of drill bits such as the carbide cobalt percentage in a drill bit (factor A) and radial rake (factor F). The noise factors dealt with characteristics of different types of multilayer circuit boards that would be drilled such as board material (factor O) and number of layers in a board (factor P). Thus, 16 different drill bit designs specified by the control factor array were used in the eight different production conditions specified by the noise factor array. Note that testing was stopped after 3,000 holes were drilled and 11% of the tested drill bits had not failed by that time.

Taguchi (1987) originally proposed estimating the loss $L(\mathbf{x}_{control})$ for each $\mathbf{x}_{control}$ specified

Table 3: Product Array Design and Lifetime Data for the Drill Bit Experiment

(experiment ended at 3000 cycles)

											Noise Array								
											1	1	1	1	2	2	2	2	M
											1	1	2	2	1	1	2	2	N
											1	1	2	2	2	2	1	1	O
											1	2	1	2	1	2	1	2	P
											1	2	2	1	2	1	1	2	Q
Control Array																			
A	D	B	C	F	G	H	I	E	J	L	Lifetime								
1	1	1	1	1	1	1	1	1	1	1	1280	44	150	20	60	2	65	25	
1	1	1	1	1	2	2	2	2	2	2	2680	125	120	2	165	100	795	307	
1	2	2	2	2	1	1	1	1	2	2	2670	480	762	130	1422	280	670	130	
1	2	2	2	2	2	2	2	2	1	1	2655	90	7	27	3	15	90	480	
2	1	1	2	2	1	1	2	2	1	2	3000	440	480	10	1260	5	1720	3000	
2	1	1	2	2	2	2	1	1	2	1	2586	6	370	45	2190	36	1030	16	
2	2	2	1	1	1	1	2	2	2	1	3000	2580	20	320	425	85	950	3000	
2	2	2	1	1	2	2	1	1	1	2	800	45	260	250	1650	470	1250	70	
3	1	2	1	2	1	2	1	2	1	1	3000	190	140	2	100	3	450	840	
3	1	2	1	2	2	1	2	1	2	2	3000	638	440	145	690	140	1180	1080	
3	2	1	2	1	2	1	2	1	1	1	3000	180	870	310	2820	240	2190	1100	
4	1	2	2	1	1	2	2	1	1	2	3000	612	1611	625	1720	195	1881	2780	
4	1	2	2	1	2	1	1	2	2	1	3000	1145	1060	198	1340	95	2509	345	
3	2	1	2	1	1	2	1	2	2	2	3000	970	180	220	415	70	2630	3000	
4	2	1	1	2	1	2	2	1	2	1	3000	3000	794	40	160	50	495	3000	
4	2	1	1	2	2	1	1	2	1	2	680	140	809	275	1130	145	2025	125	

by the control array from the data obtained by varying the noise factors according to the noise array and then modeling it as a function of the control factors. Alternatively, Welch et al. (1990) proposed modeling the response Y directly as a function of both the control and noise factors and then evaluating the loss using the estimated response model. Their rationale for the latter approach, termed the response-model approach by Shoemaker, Tsui and Wu (1991), was that it would be more likely to find a simple model for the response than one for the much more complicated estimated loss. Examples in Welch et al. (1990) and Shoemaker et al. (1991) provide evidence for preferring the response-model approach. Shoemaker et al. (1991) showed that the approach also provides more information.

For achieving robust reliability, the response-model approach is a natural one because the same parametric regression models mentioned in Section 2 can be used. The product array data allows a model to be fit consisting of all C main effects (possibly some $C \times C$ interactions), all $C \times N$ interactions and all N main effects (possibly some $N \times N$ interactions), where C and N denotes control and noise factors, respectively. The $C \times N$ interactions play an important role because the fact that the loss (1) changes for different control factor combinations means that these interactions must exist. Figure 1a displays a simplified relationship between a response Y and one control factor (at two levels) and one noise factor (over an interval) and shows that the effect of the noise factor is substantially smaller at control factor level 1 (C1). Thus, robust design exploits the existence of interactions between control and noise factors. Note that having a $C \times N$ interaction is not sufficient for an opportunity for robustness as is shown in Figure 1b where the magnitude of the change over the noise factor interval at both levels of the control factor is the same. Consequently, an N main effect is also needed which explains the inclusion of both $C \times N$ interactions and N main effects in the model. The C main effects and $C \times C$ interactions indicate the general response value about which the response varies as the noise factors vary according to their distribution; the amount of variation depends on the magnitudes of the N main effects and $C \times N$ interactions.

By taking the response-model approach, alternate designs to a product array have also been suggested. For example, Welch et al. (1990) proposed using a single plan or array for both the control and noise factors. Shoemaker et al. (1991) explored the economic advantages of single arrays over product arrays.

4. ANALYSIS METHODS FOR CENSORED DATA

For analyzing the experiments discussed in the previous two sections, we consider the following parametric regression model (Lawless 1982):

$$y_i = \log(t_i) = \mathbf{x}_i{}^T\beta + \sigma\epsilon_i,\ i = 1,\ \ldots,\ n, \tag{2}$$

where the $\{t_i\}$ are the lifetimes, the $\{\mathbf{x}_i\}$ are the corresponding vectors of covariates values, β is the vector of location parameters and σ is the scale parameter. The errors $\{\epsilon_i\}$ are i.i.d. standard extreme-value r.v.'s if the lifetimes have a Weibull distribution and are i.i.d. standard normal r.v.'s if the lifetimes follow a lognormal distribution. The models with these two error distributions are called the lognormal and Weibull regression models, respectively. For the reliability improvement experiments given in Section 2, the covariates consist of an intercept, the factor main effects and possibly some interactions. For the robust reliability experiments presented in Section 3, the covariates consist of an intercept, the C main effects, possibly some $C \times C$ interactions, the $C \times N$ interactions, the N main effects and possibly some $N \times N$ interactions, where C and N denote control and noise factors, respectively. Because of the typically small amount of data collected in these experiments, there appear to be little qualitative differences between the use of either model; there is not enough data to differentiate between the two error distributions. The lognormal regression model has some advantages, however, because of its connection with the normal regression model and has been exploited in recent work (Hamada and Wu 1991, 1992b). For complete data, i.e., when all lifetimes observed, the analysis is straightforward using maximum likelihood estimation (Lawless 1982). As mentioned in the Introduction, the censored data present new challenges in the context of analyzing reliability experiments. Next we give an overview of methods for handling the censored data which highlights these challenges.

4.1 A Quick and Dirty Method

A quick and dirty (QD) method which continues to be used in practice treats the right-censoring times as actual failure times and then analyzes them by standard methods for complete data. (For interval-censored data, an interval endpoint or midpoint is used.) Although simple, ignoring the censoring can lead to wrong decisions because the unobserved failure times and right-censoring times may differ greatly depending on the particular factor level combination. A simulation study in Hamada and Wu (1991) showed that the QD method can perform quite poorly. Hamada (1992) also pointed out that Taguchi's (1987) minute accumulating analysis treats the right-censored data similarly.

4.2 Fitting Saturated Models and Their Submodels

One obvious approach for handling censored data is to specify a saturated model and fit it using maximum likelihood estimation (MLE). The approach has several drawbacks, however. First, the MLEs need not exist, i.e., at least one is infinite, so that testing cannot be done by comparing the MLEs with their standard errors. Silvapulle and Burridge (1981) gave necessary and sufficient conditions for the existence of MLEs for model (2). In the reliability context, Hamada and Tse (1992) concluded that estimability problems will tend to occur for saturated and nearly saturated submodels.

Krall, Uthoff and Harley (1975) proposed using an MLE-based forward selection procedure. Building the model up tends to mitigate estimability problems, but still requires a certain amount of computation; an iterative algorithm is required to obtain the MLEs for each model fit. If a stepwise selection procedure is used instead, the amount of computation required increases substantially. Lawless and Singhal (1980) proposed an efficient algorithm for an all subsets procedure which finds good submodels of a saturated model. While the

MLEs of saturated and nearly saturated models are not likely to exist, the likelihood is still well defined so that sequences of submodels could be fit and compared using appropriate likelihood ratio tests. There are computational difficulties associated with the estimability problems, however, which the software needs to handle as shown in Clarkson and Jennrich (1991). Nevertheless, the computational cost can be quite high because many possible models may be fit.

For the Plackett-Burman 12-run, mixed-level fractional factorial and 3^{k-p} designs, the computational cost is even higher and may be prohibitive because of the enormous number of possible models. If we consider a comprehensive model for these designs, say containing all main effects and two-factor interactions, the number of effects exceeds the number of runs and therefore cannot be fit. As Hamada and Wu (1992b) pointed out, there is no saturated comprehensive model for these designs because of their complex aliasing patterns; e.g., a main effect is partially aliased with several if not many two-factor interactions rather than being completely aliased with a one or a few two-factor interactions as is the case for 2^{k-p} designs. Take for example the 12-run Plackett-Burman design where each factor main-effect is partially aliased with all two-factor interactions not involving the factor. For the heat exchanger experiment with ten factors in Section 2, a second-order model would have 55 effects plus an intercept. Consequently, the number of possible models to be fit can be enormous, even when restrictions are made on the class of models such as including at least one of the corresponding main effects for any interaction appearing in the model.

As a final comment, the likelihood ratio testing approach which handles the estimability problem may not be entirely useful for robust reliability experiments unless the MLEs of the final model selected exist. In the context of analyzing ordinal data (which applies here as well), Chipman and Hamada (1993) pointed out that in choosing control factor levels, finite estimates are needed to evaluate the loss (1).

4.3 Iterative Imputation-Based Model Selection

The computational problems and requirements of the standard MLE based methods presented in the previous section motivated Hamada and Wu (1991). They proposed an iterative model selection procedure based on imputation. By imputing the censored data to obtain "complete normal" or pseudo-complete data, any of the standard model selection techniques can be used at much less computational cost. Note that this procedure can be used for the lognormal regression model (1) since the log lifetimes are normally distributed.

The iterative model selection procedure consists of a three-step loop:

1. Imputation
2. Model Selection
3. Model Fitting

An initial model chosen by the experimenter which may consist only of main effects for highly fractionated designs is fit using maximum likelihood estimation. In the imputation step, the censored data are replaced by conditional mean lifetimes based on the current model. The next model is then chosen based on these pseudo-complete or pseudo-uncensored data using any standard method such as stepwise selection. The chosen model is then fit using maximum likelihood estimation, the censored data are imputed, and so forth until the next chosen model stops changing.

The procedure exploits the simplicity of a complete data problem to solve an incomplete data problem. While the use of standard methods on pseudo-complete data lacks theoretical justification, the simulation study given in Hamada and Wu (1991) showed that the method performs well and is far superior to the QD method. The procedure also relies on maximum likelihood estimation so that there are potential estimability problems. These tend to be avoided since the procedure builds up the model rather than starting with a saturated model.

In earlier work, Hahn, Morgan and Schmee (1981) used imputation to analyze left-censored yield data from an experiment on a chemical process. Their method starts out by imputing the censored data based on the model obtained by the QD method, i.e., a least-squares fit treating the censoring values as actual observations. Imputing the censored data by their conditional means, the next model is chosen based on the least-squares estimates (LSEs) from the pseudo-complete data. The LSEs of the chosen model, which do not require an additional fitting, are used in the next imputation. A key difference with that proposed by Hamada and Wu (1991) is the reliance on LSEs rather than MLEs; the latter directly incorporates the censoring information in fitting the current chosen model. A rationale for the LSEs is that for a fixed model, iterative LSEs at convergence are nearly equal to MLEs (Aitken 1981) and therefore can be viewed as one-step MLEs. Despite its simplicity, the simulation study in Hamada and Wu (1991) indicates that the LSE based procedure performs worse especially for heavier censoring.

4.4 A Bayesian Approach

The estimability problem of the MLE based methods presented in the previous sections motivated the Bayesian approach proposed in Hamada and Wu (1992b). The Bayesian approach is a natural one because important factor effects might be expected to be large but not infinite. By using proper prior distributions, posterior distributions with finite modes result and can be used to obtain finite estimates. Also, the posterior distributions allow the importance of factorial effects to be assessed without using the asymptotic approximations employed by the MLE based methods.

This approach is relatively simple to implement using the recent advances in Bayesian computing. That is, the resampling methodology makes calculating the entire posterior distribution or some characteristic pretty straightforward to do. Wei and Tanner (1990b) showed how posteriors for censored data regression models could be calculated using data augmentation (Tanner and Wong 1987). Wei and Tanner (1990a) also showed how the posterior maximizer could be calculated without computing the entire posterior by what they called the Monte Carlo EM. Wei and Tanner (1990 a,b) used an improper prior, however, which Hamada and Wu (1992b) showed could be extended to the proper conjugate prior given by Raiffa and Schlaifer (1961). Besides producing well-behaved posteriors, the conjugate prior allows the robustness of the results to be investigated. See Hamada and Wu (1992b) for more details.

In analyzing 2^{k-p} designs, a saturated comprehensive can be entertained using a standard Bayesian analysis. Thus, for these designs, the iterative model selection procedure of Hamada and Wu (1991) is not needed. For designs with complex aliasing patterns such as the 12-run Plackett-Burman design, however, Hamada and Wu (1992b) proposed adapting this iterative model selection procedure. That is, the imputation is based on the posterior maximizer instead of the MLEs. For the same reasons given previously, this adapted iterative model selection procedure requires much less computation than an entirely Bayesian approach.

For analyzing robust design experiments, the Bayesian approach has an additional advantage. Chipman and Hamada (1993) showed how the Bayesian approach can easily combine the uncertainty of the model estimates with the variation of the noise factors in choosing control factor levels; previous work had not accounted for the model estimates' uncertainty. This idea can be applied to robust reliability experiments and is currently being implemented.

5. ANALYSIS OF FLUORESCENT LAMP EXPERIMENT

For the fluorescent lamp experiment presented in Section 2, first consider a standard MLE based analysis using the lognormal regression model. Table 4 gives the MLEs and standard errors for the five main effects (A-E) and the AB interaction which the experimenters thought might be important. The intercept is denoted by Int. Based on these results, the main effects D, B, E and A are important in the order given.

An additional effect (BD=AE) can be entertained if a saturated model is fit, but the MLEs do not exist for this model since both replicates at the fifth run are right-censored. The Bayesian approach of Section 4.4, which circumvents the estimability problem, can be taken using a relatively diffuse prior distribution with mean zero for the effects. The quantiles corresponding to central 0.95 and 0.99 intervals of the posterior distribution are displayed in Table 5. These results show that BD is not important and confirm the importance of main-effects, D, B, E and A. Hamada and Wu (1992b) also showed that these results hold for a

Table 4: MLEs and Standard Errors of Lognormal Regression Model for the Fluorescent Lamp Experiment

Effect	MLE	Std Err
Int	2.939	0.064
A	-0.117	0.062
B	0.201	0.060
AB	-0.049	0.062
C	0.051	0.062
D	-0.273	0.062
E	0.153	0.063
σ	1.590	0.043

much sharper prior distribution. The sign of these effects suggests that reliability will improve at recommended settings $A_1B_2D_1E_2$, where the subscript indicates the recommended level.

6. ANALYSIS OF THE HEAT EXCHANGER EXPERIMENT

Hamada and Wu (1991) used the iterative model selection procedure to analyze the heat exchanger experiment presented in Section 2. The main effects model for the ten factors whose MLEs exist was fit and used to impute the censored data. A stepwise selection procedure applied to the pseudo-complete data identified the model (E, EG, EH). The same model was chosen in the next iteration.

Now consider the saturated model consisting of the ten main effects and the effect associated with the unassigned design column U in Table 1. Since the MLEs do not exist for this model, the Bayesian approach can be taken. Table 6 gives the quantiles corresponding to central 0.95 and 0.99 intervals of the posterior using a relatively diffuse prior. While B, C, E, K and U appear important, recall that all main effects are aliased with the 36 two-factor interactions not involving the factor; U is aliased with all 45 two-factor interactions between the ten factors. The U effect's importance means that some two-factor interactions are important and can explain the main effects B, C and K. See Hamada and Wu (1991) for details. E is much larger so that E is clearly important. Thus, the Bayesian analysis of the saturated

Table 5: Posterior Quantiles Using Diffuse Prior
for the Fluorescent Lamp Experiment

	Quantile			
Effect	.005	.025	.975	.995
Int	2.83	2.84	3.02	3.07
A	-0.24	-0.19	-0.02	-0.00
B	0.07	0.09	0.26	0.31
C	-0.06	-0.05	0.12	0.17
AB	-0.15	-0.10	0.07	0.08
E	0.02	0.04	0.21	0.26
D	-0.40	-0.35	-0.18	-0.17
BD	-0.10	-0.05	0.11	0.13
σ	0.10	0.10	0.20	0.24

model confirms the importance of E and identifies the presence of some interactions.

7. ANALYSIS OF THE DRILL BIT EXPERIMENT

Taking the response-model approach, a Weibull regression model (1) consisting of an intercept, C main effects, one $C \times C$ interaction ($D \times E$), N main effects, two $N \times N$ interactions ($M \times P$, $M \times Q$) and all the $C \times N$ interactions, was fit using the product array data from Table 3. Table 7 presents only the MLEs and respective standard errors for the important effects. Since factor A has four levels, the main effect is modeled by linear, quadratic and cubic components which are denoted by A_l, A_q and A_c, respectively.

Once the response has been modeled, recommendations for the important control factors settings need to be made. For a simple model with few noise factors, they may be apparent from inspection of the model directly; i.e., by observing what the significant effects are and their magnitudes. Shoemaker et al. (1991) gave an example, but for complicated models, this approach may be tedious if not impossible.

An alternative is to specify some meaningful criterion or loss and use the identified model to evaluate them. The loss (1) can then be estimated using some distribution for the noise factors. In practice, because it may be difficult to specify the distribution, the criterion can

Table 6: Posterior Quantiles Using Diffuse Prior for the Heat Exchanger Experiment

	Quantile			
Effect	.005	.025	.975	.995
Int	4.180	4.190	4.262	4.273
F	-0.050	-0.040	0.031	0.041
B	-0.145	-0.134	-0.059	-0.049
A	-0.068	-0.054	0.019	0.029
C	0.034	0.044	0.117	0.123
D	-0.026	-0.016	0.057	0.069
E	-0.340	-0.329	-0.257	-0.247
G	-0.061	-0.049	0.023	0.035
H	-0.058	-0.045	0.029	0.040
J	-0.022	-0.012	0.058	0.068
K	0.028	0.039	0.111	0.121
U	-0.191	-0.176	-0.104	-0.094
σ	0.028	0.028	0.028	0.028

be evaluated over the noise combinations given by the noise array. The same noise array from the experiment does not have to be used. For example, instead of a fractional factorial design, the loss could be evaluated using a full factorial design. The noise combinations can also be weighted appropriately to reflect their probabilities of occurrence. Similarly, the loss can be estimated for all possible settings of the control factors.

For achieving robust reliability, besides requiring high reliability on average, as little dependence as possible on the noise factors is desired. This suggests estimating the mean and standard deviation of the mixture of lifetime distributions given by (2) with the mixture defined by the noise factor distribution. Ideally, there is a control factor setting that simultaneously maximizes both the mean and minimizes the standard deviation; otherwise, tradeoffs between the two need to be made. By taking a worst case approach, the minimum mean lifetime over the noise factor distribution provides another criterion which can be maximized as a basis for choosing the control factor settings. Finally, Taguchi's (1987) larger-the-better

Table 7: MLEs and Standard Errors of Weibull Regression Model for the Drill Bit Experiment

Effect	MLE	Std Err	Effect	MLE	Std Err
Int	6.182	0.047	$A_l N$	-0.047	0.021
A_l	0.279	0.021	BN	-0.094	0.042
A_q	-0.268	0.043	IN	-0.111	0.044
A_c	0.071	0.018	DO	0.181	0.058
C	-0.194	0.043	BO	-0.136	0.047
D	-0.265	0.043	GO	0.277	0.054
F	0.154	0.042	IO	-0.429	0.054
G	0.132	0.048	EO	-0.376	0.061
H	0.218	0.048	LO	0.294	0.059
I	-0.272	0.044	$A_q P$	-0.123	0.061
J	-0.231	0.043	DP	0.269	0.059
L	-0.272	0.043	BP	0.213	0.048
DE	-0.225	0.041	CP	-0.119	0.059
M	0.179	0.058	IP	0.195	0.054
N	0.136	0.047	EP	0.143	0.061
O	0.898	0.059	JP	0.156	0.060
P	0.862	0.057	LP	-0.194	0.060
Q	0.548	0.057	MP	0.237	0.057
$A_q Q$	-0.174	0.061	GM	0.236	0.054
HQ	0.079	0.061	JM	0.149	0.059
IQ	-0.117	0.054	LM	0.123	0.059
LQ	0.202	0.060	σ	0.350	0.030

signal-to-noise ratio (LTB S/N ratio) for assessing the effect of noise factors is applicable here and is based on the loss in (1) using the loss function $l(Y) = 1/Y^2$. Control factor settings with large LTB S/N ratios, defined as $-10log_{10}L(\mathbf{x}_{control})$, can then be identified. Other criteria are possible such as that based on the probability of exceeding a specified time such as a warranty period, but are not considered further.

For the drill bit experiment, the relationship between the response and the control and noise factors as seen from Table 7 is too complicated to make control factor level recommendations simply by inspecting the model. Consequently, the criteria discussed above can be estimated, namely the mean, standard deviation, minimum mean and LTB S/N ratio using all possible combinations of noise factors ($32 = 2^5$) and evaluated at all possible combinations of control factor ($4096 = 4 \times 2^{10}$) and then ranked appropriately (out of 4096, with 1 being the best). Table 8 presents the best five control factor combinations for each criterion along with the other criteria and their ranks. Several observations can be made: (i) the combination least sensitive to the noise factors ranks rather poorly on other criteria, especially the mean; (ii) the other three criteria identify many of the same combinations; (iii) there is little difference between the top few combinations. Based on Table 8, a good choice of factor levels would be $A_4D_2B_2C_2F_1G_2H_1I_2E_1J_2L_2$. Note that this combination is also rather robust to the noise factors.

8. DISCUSSION

The paper has presented the use of designed experiments for reliability improvement and for achieving robust reliability. While the experimental designs are the same ones used for improving any quality characteristic, it is the analysis of censored lifetime data from these designs that has provided new challenges. In this context, standard methods need to deal with estimability problems because of the censoring or require a large amount of computation because many possible models may be fitted, especially for designs with complex aliasing patterns. Standard MLE based methods may be used if the censoring is not too severe as was demonstrated in the drill bit experiment, however. A Bayesian approach (Hamada and

Table 8: Best Factor Settings for the Drill Bit Experiment

settings											mean		std dev		LTB S/N		min mean	
five largest means																		
A	D	B	C	F	G	H	I	E	J	L	value	rank	value	rank	value	rank	value	rank
4	2	1	2	1	1	1	2	1	2	2	8.877	1	1.084	497	18.776	2	7.122	4
4	2	2	2	1	1	1	2	1	2	2	8.877	2	1.064	469	18.781	1	7.088	7
4	2	1	2	1	2	1	2	1	2	2	8.613	3	0.903	205	18.552	4	7.150	3
4	2	2	2	1	2	1	2	1	2	2	8.613	4	0.681	66	18.618	3	7.116	5
3	2	1	2	1	1	1	2	1	2	2	8.571	5	1.396	1246	18.312	10	6.128	71
five smallest standard deviations																		
A	D	B	C	F	G	H	I	E	J	L	value	rank	value	rank	value	rank	value	rank
4	2	2	1	2	2	2	1	1	2	1	6.209	2171	0.325	1	15.825	1071	5.540	199
4	2	2	1	1	2	2	1	1	2	1	6.517	1629	0.325	2	16.249	735	5.848	117
4	2	2	1	2	2	1	1	1	2	1	6.829	1102	0.325	3	16.658	463	6.160	66
4	2	2	1	1	2	1	1	1	2	1	7.137	674	0.325	4	17.043	271	6.468	35
1	2	2	1	1	2	1	1	1	2	1	5.321	3306	0.418	5	14.439	2233	4.414	835
five largest LTB S/N ratios																		
A	D	B	C	F	G	H	I	E	J	L	value	rank	value	rank	value	rank	value	rank
4	2	2	2	1	1	1	2	1	2	2	8.877	2	1.064	469	18.781	1	7.088	7
4	2	1	2	1	1	1	2	1	2	2	8.877	1	1.084	497	18.776	2	7.122	4
4	2	2	2	1	2	1	2	1	2	2	8.613	4	0.681	66	18.618	3	7.116	5
4	2	1	2	1	2	1	2	1	2	2	8.613	3	0.903	205	18.552	4	7.150	3
4	2	2	2	2	1	1	2	1	2	2	8.569	8	1.064	470	18.461	5	6.780	17
five largest minimum means																		
A	D	B	C	F	G	H	I	E	J	L	value	rank	value	rank	value	rank	value	rank
4	2	2	2	1	1	1	1	1	2	2	8.333	20	0.815	120	18.299	11	7.424	1
4	2	2	1	1	1	1	1	1	2	2	7.945	88	0.594	32	17.934	41	7.274	2
4	2	1	2	1	2	1	2	1	2	2	8.613	3	0.903	205	18.552	4	7.150	3
4	2	1	2	1	1	1	2	1	2	2	8.877	1	1.084	497	18.776	2	7.122	4
4	2	2	2	1	2	1	2	1	2	2	8.613	4	0.681	66	18.618	3	7.116	5

Wu 1992b) summarized in the paper addresses both drawbacks. Its current implementation is for the lognormal regression model or for the normal regression model after transforming the lifetime response. The Bayesian approach could be extended to the Weibull regression model.

Other types of reliability improvement experiments need to be explored for highly reliable product where lifetime based experiments discussed in this paper are not feasible. Two possibilities are the use of acceleration factors to speed-up failures (Nelson 1990 and Meeker and Escobar 1993) and the collection of degradation data, i.e., monitoring the degradation of surrogate characteristics for reliability (Lu and Meeker 1993).

There are indications that statistically designed experiments are now being used more often for improving reliability than in the past. It is hoped that the trend continues with more experiments for achieving robust reliability being performed. As other types of reliability improvement experiments are considered, new challenges for analyzing them will arise. Finally, the use of alternate experimental designs needs to be investigated.

REFERENCES

Aitken, M. (1981), "A Note on the Regression Analysis of Censored Data," *Technometrics*, 23, 161-163.

Box, G.E.P. and Draper, N.R. (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978), *Statistics for Experimenters*, New York: John Wiley.

Chipman, H. and Hamada, M. (1993), "A Bayesian Approach for Analyzing Ordinal Data from Industrial Experiments," University of Waterloo Institute for Improvement in Quality and Productivity Research Report RR-93-06.

Clarkson, D.B. and Jennrich, R.I. (1991), "Computing Extended Maximum Likelihood Estimates for Linear Parameter Models," *Journal of the Royal Statistical Society*, Ser. B, 53, 417-426.

Condra, L.W. (1993), *Reliability Improvement with Design of Experiments* New York: Marcel Dekker.

Dey, A. (1985), *Orthogonal Fractional Factorial Designs*, New York: Halsted Press.

Hahn, G.J., Morgan, C.B. and Schmee, J. (1981), "The Analysis of a Fractional Factorial Experiment with Censored Data Using Iterative Least Squares," *Technometrics*, 23, 33-36.

Hamada, M. (1992), "An Explanation and Criticism of Minute Accumulating Analysis," *Journal of Quality Technology*, 24, 70-77.

Hamada, M. (1993), "Reliability Improvement Via Taguchi's Robust Design," *Quality and Reliability Engineering International*, 9, 7-13.

Hamada, M. and Tse, S.K. (1992), "On Estimability Problems in Industrial Experiments with Censored Data," *Statistica Sinica*, 2, 381-391.

Hamada, M. and Wu, C.F.J. (1991), "Analysis of Censored Data from Highly Fractionated Experiments," *Technometrics*, 33, 25-38.

Hamada, M. and Wu, C.F.J. (1992a), "Analysis of Designed Experiments with Complex Aliasing," *Journal of Quality Technology*, 24, 130-137.

Hamada, M. and Wu, C.F.J. (1992b), "Analysis of Censored Data from Fractionated Experiments: a Bayesian Approach," University of Waterloo Institute for Improvement in Quality and Productivity Research Report RR-92-11.

Hitzelberger, A.J. (1967), "Improve Your Reliability," *Industrial Quality Control*, 24, 313-316.

Kackar, R.N. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Method (with discussion)," *Journal of Quality Technology*, 17, 176-209.

Krall, J.M., Uthoff, V.A. and Harley, J.B. (1975), "A Step-Up Procedure for Selecting Variables Associated with Survival," *Biometrics*, 31, 49-57.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley.

Lawless, J.F. and Singhal, K. (1978), "Efficient Screening of Nonnormal Regression Models," *Biometrics*, 34, 318-327.

Lu, C.J. and Meeker, W.Q. (1993), "Using Degradation Measures to Estimate a Time-to-Failure Distribution," *Technometrics*, 35, 161-174.

Meeker, W.Q., and Escobar, L.A. (1993), "A Review of Recent Research and Current Issues in Accelerated Testing," *International Statistical Review*, 61, 147-168.

Montmarquet, F. (1988), "Printed Circuit Drill Bit Design Optimization Using Taguchi's Methods – .013″ Diameter Bits," *Sixth Symposium on Taguchi Methods,* Romulus, MI: American Supplier Institute, 70-77.

Nelson, W. (1982), *Applied Life Data Analysis*, New York: John Wiley.

Nelson, W. (1990), *Accelerated Testing-Statistical Models, Test Plans, and Data Analysis*, New York: John Wiley.

O'Connor, P.D.T. (1991), *Practical Reliability Engineering*, 3rd Edition, Chichester, England: John Wiley.

Phadke, M.S. (1986), "Design Optimization Case Studies," *AT&T Technical Journal*, 65, 51-68.

Plackett, R.L. and Burman, J.P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika* 33, 305-325.

Raiffa, H. and Schlaifer, R. (1961), *Introduction to Statistical Decision Theory*, New York: McGraw-Hill.

Sampford, M.R. and Taylor, J. (1959), "Censored Observations in Randomized Block Experiments," *Journal of the Royal Statistical Society*, Ser. B, 21, 214-237.

Shoemaker, A.C., Tsui, K.L. and Wu, C.F.J. (1991), "Economical Experimentation Methods for Robust Design," *Technometrics*, 33, 415-427.

Silvapulle, M.J. and Burridge, J. (1986), "Existence of Maximum Likelihood Estimates in Regression Models for Grouped and Ungrouped Data," *Journal of the Royal Statistical Society*, Ser. B, 48, 100-106.

Specht, N. (1985), "Heat Exchanger Product Design via Taguchi Methods," in *Third Symposium on Taguchi Methods*, Romulus, MI: American Supplier Institute, 302-318.

Taguchi, G. (1986), *Introduction to Quality Engineering*, Tokyo, Japan: Asian Productivity Organisation.

Taguchi, G. (1987), *System of Experimental Design,* White Plains NY: Unipub/Kraus International Publications.

Taguchi, G. and Wu, Y.I. (1980), *Introduction to Off-Line Quality Control*, Nagoya, Japan: Japan Quality Control Association.

Tanner, M.A. and Wong, W.H. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 82, 528-540.

Wang, J.C. and Wu, C.F.J. (1991), "An Approach to the Construction of Asymmetrical Orthogonal Arrays," *Journal of the American Statistical Association*, 86, 450-456.

Wei, G.C.G. and Tanner, M.A. (1990a), "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms," *Journal of the American Statistical Association*, 85, 699-704.

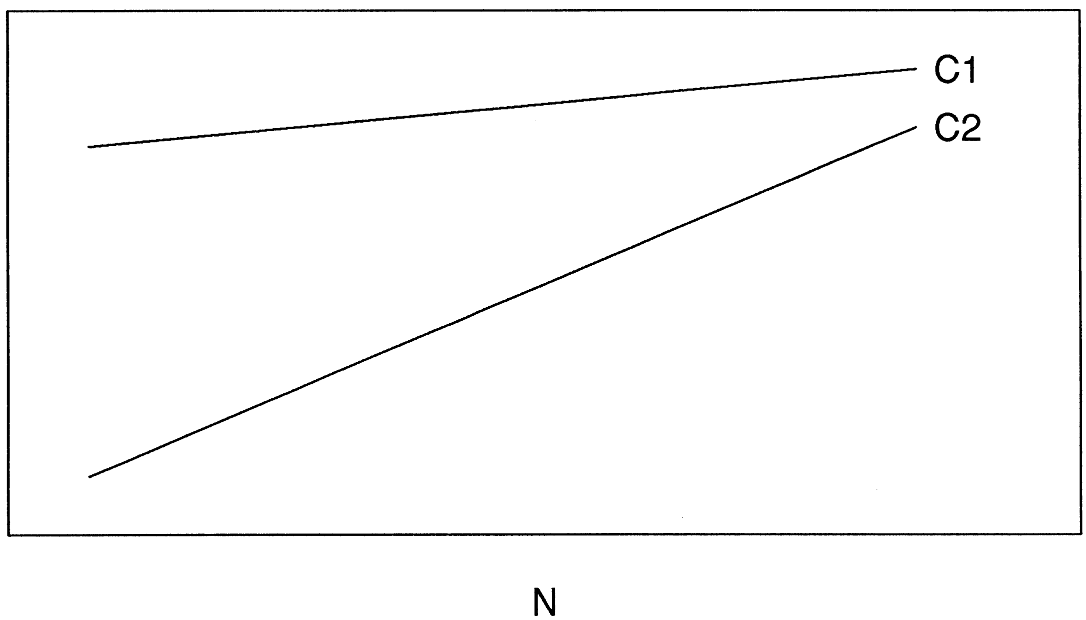
Wei, G.C.G. and Tanner, M.A. (1990b), "Posterior Computations for Censored Regression Data," *Journal of the American Statistical Association*, 85, 829-839.

Welch, W.J., Yu, T.K., Kang, S.M. and Sacks, J. (1990), "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, 15-22.

Zelen, M. (1959), "Factorial Experiments in Life Testing," *Technometrics*, 1, 269-288.

Figure 1: Example Response Functions and Opportunity for Robustness

(a) opportunity for robustness



(b) no opportunity for robustness