FIELD RELIABILITY ASSESSMENT USING THE PROPORTIONAL HAZARDS MODEL

Raymond V. Spring, *U.S. Army Natick Research, Development and Engineering Center*
Stephen A. Freitas, *U.S. Army Natick Research, Development and Engineering Center*

1. INTRODUCTION

The regression model for survival analysis introduced in Cox (1972) was developed with applications for "industrial reliability studies and medical studies" in mind. While this model has had a significant impact on the biomedical field, it has received little attention in the reliability literature. Only recently has the model been used for the analysis of hardware reliability [see Bendell and Wightman (1985) and Bendell et al (1985)], software reliability [see Bendell and Wightman (1985), Nagel and Skrivan (1985), and Font (1985)] and repairable systems [see Ascher (1983)].

The Cox model, referred to in the literature as the Proportional Hazard Model (PHM), is most useful as an exploratory technique which sheds light on the underlying structure of the data. As such, it may be used to analyze data that may prove insufficient or even inappropriate for more conventional analysis procedures. This type of data is all too common in real-world reliability applications. We illustrate the use of the PHM in such a situation, where there are many demands on the data and the controls on experimental data gathering are loose. We will discuss the use of both Bayesian and non-Bayesian estimation techniques for the PHM.

Due to the expository nature of this paper, the underlying theory of the PHM model and the associated estimation procedure are not discussed in detail. For further theoretical development, the reader is referred to Cox (1972, 1975), Kalbfleisch and Prentice (1980), Tsiatis (1981), or Lawless (1982). Bryson and Johnson (1981) give procedures for resolving problems that arise during the implementation of these models, and Schoenfeld (1982) has developed a test for the appropriateness of the model. The Bayesian estimation procedure is discussed in Kalbfleisch (1978).

The current SAS/STAT* procedure LIFEREG estimates a parametric type Cox regression while this reliability estimate uses the non-parametric approach. This type of option or procedure for the Cox regression or Proportional Hazards Model (PHM) would be very useful if included in LIFEREG or be the basis for a new procedure.

In Section 2 we present an overview of the PHM and associated estimation procedures. In Section 3 we discuss the estimation requirements, design of experiment, and proposed analysis of our specific problem.

2. OVERVIEW OF THE PROPORTIONAL HAZARD MODEL

In the more traditional life testing model, it is assumed that a certain number of identical items are tested under identical conditions. This is due to statistical, rather than practical, considerations. Under these assumptions, the times to failure of these items are independent and identically distributed with some density f(t) and corresponding reliability function R(t). Most often the failure characteristics of the test item are best modeled via the failure rate (or hazard) function

$$h(t) = \frac{f(t)}{R(t)} \tag{2.1}$$

where h(t)dt is approximately the conditional probability of failure in (t, t+dt] given survival to time t. Choices for the form or restrictions on h(t) are usually done with the physics of failure, aging characteristics or simple statistical convenience in mind.

The problem with the above approach is that often times actual field data may violate the underlying assumptions in that [see Ascher (1983)].

a. Items tested may not be totally indistinguisable;
b. The test conditions may vary from item to item;
c. A specific parametric model for the underlying hazard function cannot be specified.

In contrast to the above, the PHM overcomes these difficulties by relying on only one simple assumption. The only assumption of the PHM is that the hazard function of any item tested has the form

$$h(t) = h_0(t)\, g(\underset{\sim}{\beta},\underset{\sim}{z}) \tag{2.2}$$

where $h_0(t)$ is the baseline hazard function for all items tested, and $g(\cdot,\cdot)$ is a function that measures the hazard contribution for an individual test item based on the covariates $\underset{\sim}{z}$ describing the item. The regression coefficients $\underset{\sim}{\beta}$ are a measure of the importance or weight of each covariate. The PHM gets it's name from the fact that due to (2.2), the ratio of any two individual hazard functions is time invariant (i.e., any two hazards are proportional). The usual form considered for $g(\underset{\sim}{\beta},\underset{\sim}{z})$ is

$$g(\underset{\sim}{\beta},\underset{\sim}{z}) = \exp\{\sum_{i=1}^{m} \beta_i z_i\} \tag{2.3}$$

due to its nonnegativity property. The above formulation facilitates the estimation of the regression coefficients $\beta_1, \ldots, \beta_m$.

2.1 Estimation Procedures

Due to the data gathering structure of our problem, we deal with a special case of the PHM used for analyzing grouped data as in life-table analysis. The model discussed here was introduced by Kalbfleisch and Prentice (1973) and further discussed in Prentice and Gloeckler (1978).

In the sequel we consider the situation where n individual items are tested. Associated with the i^{th} test item is a 1 x m vector of covariates $\underset{\sim}{z}_i$. The value of the covariate vector $\underset{\sim}{z}_i$ is known for each individual. The items are observed over a fixed period $(0,T^*]$ which is partitioned into k+1 intervals $(t_{i-1}, t_i]$, i=1, ... ,k+1 where

$$0 \equiv t_0 < t_1 < t_2 < \cdots < t_k \equiv T^* \tag{2.4}$$

and $t_{k+1} = \infty$. Items are observed at times $t_1, \ldots, t_k$ and may be observed to have failed or been censored (removed from test *without failure*) at these items. Thus, for failed items the interval in which failure occurs is known and for convenience we assume that all censored items are censored just prior to observation (i.e., at the end of the interval).

We define the following values [see Lawless (1982) p. 372].

$$P_i(\underline{z}) \equiv \Pr\{\text{an item survives past } t_i \mid \underline{z}\},$$

$$p_i = \frac{P_i(\underline{z})}{P_{i-1}(\underline{z})}$$

$$\equiv \Pr\{\text{an item survives past } t_i \mid \text{it survives past } t_{i-1}, \underline{z}\},$$

$$P_0(\underline{z}) \equiv 1,$$

$$R_i \equiv \text{the set of items at risk of failure at time } t_{i-1},$$

$$D_i \equiv \text{the set of items observed to fail in } (t_{i-1}, t_i].$$

The likelihood function under these assumptions is given by

$$L = \prod_{i=1}^{k} \left\{ \prod_{j \in D_i} \left(1 - p_i(\underline{z}_j)\right) \right\} \left\{ \prod_{j \in R_i - D_i} p_i(\underline{z}_j) \right\} \tag{2.5}$$

2.1.1 The Grouped Proportional Hazard Model

Based on (2.5) a parametric analysis may be performed by assuming a parametric form for $P_i(\underline{z})$ in conjunction with the proportional hazard assumption. Here we choose a nonparametric approach because of no apparent justification for the use of any particular parametric family.

The grouped PHM is obtained by specifying

$$P_i(\underline{z}) = P_i^{\exp\{\underline{z}'\underline{\beta}\}} \tag{2.6}$$

where $P_i \equiv P_i(0)$, i=1, ... ,k is the baseline reliability function value at time t_i and $P_0 \equiv 1$. This gives

$$p_i(\underline{z}) = p_i^{\exp(\underline{z}'\underline{\beta})} \tag{2.7}$$

and

$$p_i = \frac{P_i}{P_{i-1}} \tag{2.8}$$

The model may be used with (2.5) to estimate $\underline{\beta}$ and $p_1, \ldots, p_k$.

Prentice and Gloeckler (1978) suggest a maximum likelihood estimation (MLE) procedure using the reparameterization

$$\gamma_i = \text{Log}\left\{-\text{Log}(p_i)\right\} \tag{2.9}$$

The Newton-Raphson technique may be used to solve the likelihood equations

$$\frac{\partial \text{Log } L}{\partial \beta_r} = \sum_{i=1}^{k} \left\{ \left(\sum_{j \in D_i} \frac{z_{j,r} X_{i,j}}{1 - e^{-X_{i,j}}} \right) - \left(\sum_{j \in R_i} z_{j,r} X_{i,j} \right) \right\}$$

$$= 0 \qquad r = 1, \ldots, m \tag{2.10}$$

$$\frac{\partial \text{Log } L}{\partial \gamma_i} = \sum_{j \in D_i} \frac{z_{j,r} X_{i,j}}{1 - e^{-X_{i,j}}} - \sum_{j \in R_i} X_{i,j} = 0$$

$$i = 1, \ldots, k \tag{2.11}$$

where $z_{j,r}$ is the r^{th} covariate value of the j^{th} covariate vector and

$$X_{i,j} = e^{\gamma_i + \underline{z}_i'\underline{\beta}} \tag{2.12}$$

Both tests and interval estimates for the coefficients $\beta_1, \ldots, \beta_m$ and values $p_1, \ldots, p_k$ may be obtained via the usual large sample theory for MLE's or the use of likelihood ratio techniques.

2.1.2 Bayesian Estimations

The Bayesian estimation procedure centers around obtaining a prior distribution for the baseline cumulative hazard rate

$$\Lambda_0(t) = \int_0^t h_0(u)\, du \tag{2.13}$$

The cumulative hazard rate may be written in terms of the sum of random variables [see Kalbfleisch (1978)]

$$\Lambda_0(t_i) = \sum_{j=1}^{i} r_j \tag{2.14}$$

$$= \sum_{j=1}^{i} -\text{Log}\{1-q_j\}$$

where

$$q_i = \frac{R(t_{i-1}) - R(t_i)}{R(t_{i-1})} \tag{2.15}$$

(assuming $R(t_i) > 0$) is the hazard contribution in the i^{th} interval. Assuming prior distributions for the variables r_i (or q_i) indirectly defines the prior distribution of $\Lambda_0(t)$. For mathematial convenience it is assumed that these variables are independent.

The gamma distribution is choosen as the prior distribution for r_i. Letting $X \sim G(\alpha, \beta)$ denote the fact that X is distributed according to a gamma density of the form

$$f(x) = \beta^{\alpha}\Gamma(\alpha)\, x^{\alpha-1}\, e^{-\beta x} \tag{2.16}$$

for $\alpha, \beta, x > 0$, then $r_i \sim G(c[\Lambda_0^*(t_i) - \Lambda_0^*(t_{i-1})], c)$. The prior parameters $\Lambda_0^*(t)$ and c are chosen as the prior best guess of the baseline cumulative hazard function and the measure of strength of conviction in the prior best guess.

This can be demonstrated by the fact that

$$E[r_i] = \Lambda_0^*(t_i) - \Lambda_0^*(t_{i-1}) \tag{2.17a}$$

$$VAR[r_i] = \frac{\Lambda_0^*(t_i) - \Lambda_0^*(t_{i-1})}{c} \tag{2.17b}$$

To date the Bayesian analysis of the PHM proceeds by estimating the coefficients $\beta_1, \ldots, \beta_m$ using MLE techniques and obtaining the posterior estimates of the cummulative hazard function conditioned on these estimates. This is clearly not a fully Bayesian procedure. For a fully Bayesian procedure we

would proceed as before but also assume that the β_i values are random variables independent of the r_i values. As such, prior distributions must be obtained for these values as well. Bayesian estimates may then be obtained from the posterior distributions of *both* the βi and r_i values.

3. OVERVIEW OF THE PROBLEM

3.1 Description and Test Goals

A test is designed to assess the ability of a certain permeable (uniform) material to repel contaminating agents under various test scenarios. A test scenario consists of a specific dose level, contaminant type, and test type. Several external factors (i.e., those not related to the contaminant) are also considered. Specifically, samples of the material are tested against two contaminating agents, at four different dose levels, using either single or multiple or multiple challenge test. The test outcome of the agent through the material constitutes a "failure" of the material sample, and thus interest centers around the "life length" of the material after the initial agent challenge (i.e, contamination). The statistical goal then is to determine the life length distribution of the material after contamination. The determination of the effects (if any) of agent type, dose level, and external factors on the life length distribution are also an important consideration.

After the initial agent challenge, the state of the material sample is observed over a fixed period of time at fixed time intervals.

3.2 Experimental Design

The experimental designs necessary using proportional hazard models require a very simple design when conducting actual wear testing and a second more complex design for the agent testing.

3.2.1 Wear Testing Design

The actual wear testing is conducted using a single level design with only the different wear times being addressed. The remaining factors must be handled by trying to achieve balance through randomization of the assignments of the different subjects to the five basic categories of possible wear:7, 14, 22, and 30 days of wear.

The only variable to be considered in the wear testing is the so-called prior wear. This is the amount of time that the uniform is worn prior to pulling it from wear testing and subjecting it to agent testing. The days of wear consist of the number of total hours worn converted to days of wear by equating 24 hours of wear to one wear day (continuous wear). It should also be noted that for one wear period, all uniforms may not be worn for the exact same number of hours due to the inability to change uniforms at the exact time necessary.

These types of wear tests are usually marked by participation, but not by control on the part of the experimenter. Thus, the experimenter can structure the experiment, but not completely control it. As for structuring the experiment, we may setup the groups for testing and randomize which uniforms become part of which wear groups. Thus, we will randomize the assignment of the individual soldiers to each wear group by carefully balancing and randomizing the soldiers unit type and uniform sizes that will become part of each wear group. It is easier to wait and select individual soldiers at the end of each wear period and avoid the problem of attrition. This does, however require more work.

3.2.2 Experimental Design for Agent Testing

The experimental designs required for chemical agent testing are more complex but still straight forward. The first part of the overall design is to make two separate but equal designs for each agent. Each of these designs will have the same factors and will have additional data collection. If there is no difference between the two agent test results, then collapsing of the two data sets will be available to increase the sample for estimation purposes.

The experimental design set will be done at the start of the agent testing phase. This is the time at which the uniforms are withdrawn from wear testing and cut into material samples for the agent testing. This phase is completely under the control of the experimenter and barring any disastrous loss of uniforms from the wear portions of the test (and thus lost samples from agent testing), the design will have no missing data. Thus, the ability to incorporate the necessary randomization into the design is easily accomplished..

Each of the designs will be a two-way factorial ANOVA with further randomization of additional factors. The factors to be considered are:

a. dose level (contamination level) at 0.5, 1.0, 5,0, and 10.0 g/m^2;
b. test type at single or multiple contamination.

Besides these factors, there will be a randomization of the soldiers' unit and uniform sizes from the wear phase of the testing. Also, there will be a randomization of the actual uniform used for each phase in each cell. Each wear phase (7, 14, 21, and 30 days) will be represented once in each cell and thus we have four data points per cell. For each data point, the sample swatch will also be randomized to ensure complete coverage of all locations across the testing.

The agent test results will consist of a measure of agent penetration at each fixed sampling point. Thus, the testing must continue for a set length of time even if the usual breakthrough or stopping criteria have been exceeded. This will insure that the same data is available for all samples tested and there will be no need for estimation of missing data points.

This type of design is similar to that used for our usual analysis procedures, except that our emphasis is on the estimated time when a breakthrough criteria has been exceeded. In addition, we use many more data points because we are not using all the available data. Thus, this type of analysis offers a significant benefit in the reduction of the required sample size and the utilization of all the available data. Hopefully, the two agent test designs will not produce different data and thus the collapsing of the two designs together would be possible.

The reduction of the sample sizes has a significant cost implication in that each of the required agent tests will cost at least \$170 for a single contamination, and at least \$350 for a multiple contamination test. These costs could be inflated if special conditions are required such as more frequent sampling.

3.3 Proposed Approach

A PHM is assumed for describing the life test results. Due to the life table structure that is imposed on the data, we will use a special application of the PHM first discussed in Kalbfleisch and Prentice (1973).

The baseline hazard function will be determined for each agent type, while dose level, test type, and external factors are treated as covariates affecting the life length of the material. For each agent type i, we will define the affects of the covariates through the function,

$$g(\underline{\beta},(\underline{x},\underline{y})) = \exp\{\beta_1 x_1+\beta_2 x_2+\gamma_1 y_1+\gamma_2 y_2+\gamma_3 y_3+\gamma_4 y_4\} \qquad (3.1)$$

where,

$x_1 = 0.5, 1.0, 5.0, 10.0$ g/m^2, is the dose level,
$x_2 = 0, 1$ is the test type (0 for single challange),
$y_1 = 7, 14, 21,$ and 30 is the prior wear time,
$y_2 = 1, \ldots, 11$ indicates swatch location,
$y_3 = 1, 2$ indicates troop unit,
$y_4 = 1, \ldots, 5$ indicates uniform size.

and $\underline{x}$ is the covariate vector for contaminant related factors, and $\underline{y}$ is a covariate for external factors. The estimation procedure is conducted as discussed in Section 2.

4. ANTICIPATED BENEFITS OF THE USE OF THE PROPORTIONAL HAZARDS MODE

The use of the proportional hazards model (PHM) has several benefits including the following:

a. More accurate reliability estimates
b. Ability to determine important conditions affecting reliability estimates
c. Possibility of reducing sample sizes
d. Better utilization of the data collected
e. Possibly extendable to a wide range of reliability estimation situations

These benefits will be more fully detailed in the following discussion. In addition there are some minor disadvantages to the use of the PHM which will also be discussed, although they should not bar the use of the technique.

The first benefit is that the technique should provide more accurate reliability estimates than the current use of a binomial estimate. The greater accuracy will result from the basic principle that a fuller utilization of the data that has been collected in the test will yield more accurate reliability estimates. The binomial estimator simply utilizes the numbers of tests that exceed a predetermined criterion level (failures). This does not take into account any conditions that may relate to this failure. The estimator is a very simple one and makes no assumptions about the data and thus is a very conservative estimator. In addition, this simplicity means that greater sample sizes are required to achieve respectable reliability levels. This has been the usual estimator because the nature of testing chemical protective equipment has required destruction of the item for testing purposes and makes other estimators unusuable. The PHM can easily utilize the type of data that results from chemical agent testing and because it does not require the use of a pass-fail criterion uses the data as interval type measures rather than ordinal measurements as the binomial estimator requires. In addition, the conditions of field use and agent testing can be included as covariates in the equation.

The second anticipated benefit is the ability to determine the important conditions that affect the PHM reliability estimate. This ability to determine the important conditions is accomplished by including the condition as a covariate in the PHM and then conducting hypothesis tests on the resulting coefficient. This testing of the coefficients would be relatively easy to accomplish and would provide a whole new tool for use in reliability estimation. Currently the determination of important conditions requires separate testing and comparison of the resulting estimates. This is time consuming and expensive method and thus usually not done. This usual type of comparison would also require that the comparison conditions be carefully monitored and placed into predetermined categories.

The third benefit of reducing sample size has already been mentioned and the key idea in this benefit is that a fuller utilization of the data collected means that the current level of reliability estimates may be achieved with smaller sample sizes. This estimation would require that an adequate number of covariates must be included and be relevant to the reliability to be estimated. But when the current costs of wear and agent testing are considered it may be desirable to reduce the sample sizes to save both time and money.

The fourth benefit of better data utilization has also been previously mentioned but still needs to be emphasized. The most expensive part of many development programs is the testing of the items developed and the better utilization of data is simply getting the full value from testing done.

The final benefit is the possible extension of the technique to other reliability testing situations. While the technique has obvious advantages for field testing with the naturally occuring variability in the testing conditions from day to day there are many other testing situations where this technique may be highly effective. One of these of particular note is the use in on-line reliability testing and estimation where the results can be immediately utilized to determine reliability rather than the collecting of data for later analysis. Such a use is viewed by many as the direction that reliability will take in its future development.

Finally, the disadvantages of use of the PHM are not major and can be easily handled. The first of these disadvantages is that the PHM would require computer support and adequate software that is only now starting to be developed for microcomputers. This problem is coupled with the second disadvantage of PHM use which is that this technique requires greater involvement of the reliability analyst in the testing. In addition, the PHM requires a greater statistical expertise on the part of the reliability analyst to correctly utilize and interpret the output of the model. But it appears that the potential for better estimates outweighs these minor difficulties.

REFERENCES

Ascher, H. (1983). Regression analysis of repairable systems reliability, Electronic Systems Effectiveness and Life Cycle Costing, J. K. Skwirzynski (ed.), Springer–Verlag, Berlin, pp 119–133.

Bendell, A. et al (1985). Proportional hazard modeling in reliability analysis – an application to break discs on high speed trains, to appear in *Quality and Reliability International*..

Bendell, A. and Wightman, D. W. (1985). The practical application of proportional hazard modeling, Proceedings of the 5th National Reliability Conference, Birmingham, England pp. 2B/3/1–16.

Bryson, M. C. and Johnson, M. E. (1981). The incidence of monotone likelihood in the Cox model, *Technometrics, Vol.* 23, pp. 381–383.

Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B* 34, pp. 187–220.

Cox, D. R. (1975). Partial likelihood, *Biometrika, Vol.* 62, pp. 269–276.

Font, V. (1985). Une approche de fiabilité des logiciels; modèles classiques et modèle lineaire generalise, *Thesis L'Universitè Paul Sabatier de Toulouse*, France.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data, *Journal of the Royal Statistical Society, Series B* 40, pp. 212–220.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life table model, *Biometrics, Vol.* 60, pp. 267–279.

Kalbfleisch, J. D. and Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data, John Wiley, New York.

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

Nagel, P. M. and Skrivan, J. A. (1985). Software reliability; repetitive run experimentation and modeling, Technical Report BCS–40336, Boeing Computer Company.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with applications to breast cancer data, *Biometrics, Vol.* 34, pp. 57–67.

Schoenfeld, D. (1982). Partial residuals for the proportional hazard regression model, *Biometrica, Vol.* 69, pp. 239–241.

Tsiates, A. A. ((1981). A large sample study of Cox's regression model, *Annals of Statistics, Vol.* 9, pp.93–108.

★ SAS/STAT is the registered trademark of SAS Institute Inc., Cary, NC, USA

For further information please contact:

Raymond Spring
US Army Natick R, D, & E Center
ATTN: STRNC-AC
Natick, MA 01760-5015
(508) 651 - 4340