

3-23-2001

Early Detection of Reliability Problems Using Information From Warranty Databases

Huaiqing Wu
Iowa State University

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_preprints



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Wu, Huaiqing and Meeker, William Q., "Early Detection of Reliability Problems Using Information From Warranty Databases" (2001). *Statistics Preprints*. 27.
http://lib.dr.iastate.edu/stat_las_preprints/27

This Article is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Preprints by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Early Detection of Reliability Problems Using Information From Warranty Databases

Abstract

Most companies maintain warranty databases for purposes of financial reporting and warranty expense forecasting. In some cases, there are attempts to extract engineering information (e.g., on the reliability of components) from such databases. Another important application is to use warranty data to detect potentially serious field reliability problems as early as possible. When a serious problem arises, the existence of the problem will eventually be obvious. Early detection of serious problems through the use of sensitive statistical methods, allowing early action to mitigate potential reliability problems, could save large amounts of money and product good will.

This paper describes a detection procedure that has been designed for this purpose. In addition to the statistical decision rules, we suggest graphical tools for illustrating and describing the particular information in the data that caused the potential problem to be flagged. The methods are illustrated using data from an automobile warranty database.

Keywords

Average run length, Control chart, Poisson distribution, Sequential test, Statistical process monitoring, Warranty report

Disciplines

Statistics and Probability

Comments

This preprint has been published in *Technometrics* 44 (2001): 120–133, doi:[10.1198/004017002317375073](https://doi.org/10.1198/004017002317375073).

Early Detection of Reliability Problems Using Information From Warranty Databases

Huaiqing Wu
William Q. Meeker
Department of Statistics
Iowa State University
Ames, IA 50011

March 23, 2001

Abstract

Most companies maintain warranty databases for purposes of financial reporting and warranty expense forecasting. In some cases, there are attempts to extract engineering information (e.g., on the reliability of components) from such databases. Another important application is to use warranty data to detect potentially serious field reliability problems as early as possible. When a serious problem arises, the existence of the problem will eventually be obvious. Early detection of serious problems through the use of sensitive statistical methods, allowing early action to mitigate potential reliability problems, could save large amounts of money and product good will.

This paper describes a detection procedure that has been designed for this purpose. In addition to the statistical decision rules, we suggest graphical tools for illustrating and describing the particular information in the data that caused the potential problem to be flagged. The methods are illustrated using data from an automobile warranty database.

Key words: Average run length; Control chart; Poisson distribution; Sequential test; Statistical process monitoring; Warranty report.

1 Introduction

1.1 Motivation for early detection

The modern philosophy of manufacturing quality and reliability is to design reliability into a product and do up-front accelerated testing before manufacturing begins in an effort to avoid serious reliability/warranty problems for the product in the field. For example, programs such as “Reliability by Design” and “Design for Six Sigma” have become popular in manufacturing industries.

This focus on quality and reliability and the effective use of statistical methods have done much to improve the reliability of manufactured products in certain industries, notably the automobile industry. Automobile warranty periods have increased dramatically and 100 thousand miles is no longer considered to be the end of the useful life of an automobile.

Nevertheless, most manufacturing companies have, from time to time, faced and will continue to face serious reliability problems, most often caused by one or some combination of the following: an unanticipated failure mode, harsher than expected operating environment, an unknown change in raw material properties or supplier quality, an improperly verified design change, etc. In order to meet financial reporting requirements and to assure adequate financial reserves, manufacturing companies maintain warranty databases. The use of appropriate statistical detection rules in a warranty database has the potential to identify or warn of serious reliability problems long before they would otherwise be discovered. Detecting such a problem months or even weeks earlier than it would otherwise be detected can importantly reduce both tangible and intangible costs of poor reliability. Use of appropriate statistical tools can provide a framework to reliably and convincingly separate signals in the data from noise. Because warranty databases already exist, there is little extra cost in doing this monitoring.

1.2 Related work

A number of papers and books have been written to describe the use of warranty data. Blischke and Murthy (1994, 1996) cover a wide range of topics related to warranty issues. General reviews of statistical methods for warranty data are provided by Robinson and McDonald (1991), Lawless and Kalbfleisch (1992), and Lawless (1998). Specific technical methods for dealing with problems arising in field and warranty data (reporting delays, censoring, truncation, and missingness) are provided, for example, in Suzuki (1985a, b), Kalbfleisch and Lawless (1988), Lawless, Hu, and Cao (1995), Hu, Lawless, and Suzuki (1998), Karim, Yamamoto, and Suzuki (2001a), and Wang and Suzuki (2001a, b). Kalbfleisch, Lawless, and Robinson (1991) describe prediction methods. Karim, Yamamoto, and Suzuki (2001b) provide methods for detecting a change point from marginal count warranty data which arise when one cannot identify the date of manufacture of units that are serviced under warranty. In this paper we build on analysis and adjustment methods presented in some of these papers to develop specific methods for early detection of changes from a historical standard for product reliability.

1.3 Overview

The remainder of this paper is organized as follows. Section 2 describes the structure of warranty data as it applies to early detection monitoring and introduces the example that we use to illustrate the methods. Section 3 outlines the statistical formulation of the detection problem, including a discussion of the relationship between warranty monitoring and traditional Shewhart process monitoring schemes. Section 4 describes the analytical methods used to allocate false alarm probability and to balance with power for detection. Section 5 shows how to compute critical values for the early detection monitoring procedure. Section 6 gives methods for assessing the average run length (time until a detection signal is given), both under the baseline conditions and under out-of-control alternatives. Section 7 summarizes the behavior of our monitoring procedure over a number of “la-

bor codes” in our example warranty database. Section 8 describes some additional implementation issues and areas for further research.

2 Background and Example

2.1 Warranty data

Although it is common to speak of “warranty data” and the information in a “warranty database,” in most applications, inferences about field reliability from warranty data actually require information from two different databases. One database contains production information giving the unit-identification number (e.g., the vehicle identification number or VIN for automobiles), time and date of manufacture, assembly line, and other production-related information. For some products (e.g., automobiles) this database may also contain the date of sale.

A separate database contains records for each warranty report. A report may result in a repair, an adjustment, or a “no trouble found.” Each record contains the unit-identification number, date of the report, geographical information, a code indicating what action was taken, by whom, and, generally, the cost of the action. For an automobile warranty database, the record also provides the number of miles on the odometer at the time of the report.

Modern warranty reporting systems for field repairs use computer technology (e.g., bar code scanning and direct entry of information into a hand-held wireless computer terminals for field repairs). Shop repair facilities also use direct computer entry of information. These systems virtually eliminate reporting delays that used to be a serious difficulty in the analysis of warranty data.

2.2 Warranty report categorization in the warranty database

Warranty databases typically categorize reports to indicate the kind of repair or other action that was taken. The categorization variable is typically called something like “failure type,” “labor code,” “job code,” or something similar. We will use the term “labor code” in this paper because it is the term that is used in the automobile database that we use for our examples. Typically detection rules such as those described here would be used, more or less independently, for a chosen set of (possibly all) labor codes in the database.

One common difficulty with warranty data is that the report classification is often not specific enough to determine the actual cause of a report that is generally needed for engineering evaluation. Although there can be human-factors variability in the way that the repair person reports the labor code, this is generally the lowest level of information about the cause of the report and we will take this code to be indicative of a particular failure mode. Nevertheless, when a serious problem arises, the existing labor code definitions, if not too coarse, should be sufficient to identify and track the problem. Sometimes there may be special focus on a particular labor code if there has been a change in a corresponding component’s design and there is concern that the change might have a negative effect on reliability.

2.3 Distribution of time to a warranty report

Extraction of information relating to the distribution of time to a report for a particular labor code requires merging of information from the production and the report databases. In particular, the report database provides information on the calendar time of the report and other information (like cost) for all occurrences of the particular labor code. The production database provides the beginning of life definition so that the service time up to the time of the report can be determined and also so that the amount of service time can be found for those units with no reports.

Definition of the distribution of time to a report requires a definition of the beginning of life of the product. For most warranty analysis purposes, the beginning of life is defined as the time when the unit was sold. Because it is important for warranty costing, we will use this definition, although other definitions could be used for other specific purposes (e.g., in an engineering analysis, some failure mechanisms become active at the time of manufacture and operate on a calendar-time time scale).

The time that an individual unit is sold is part of the automobile industry production database. In some other industries, however, this information is missing. The date of sale generally becomes known for units that are returned for warranty repair. For units without a warranty report, this information can generally be supplemented by statistical information on the distribution of the time between manufacture and sale.

2.4 Production stratification and monitoring increment

For early detection of reliability problems for a given product, monitoring commences as soon as warranty data become available and monitoring actions should be repeated periodically as more data are accumulated. For purposes of detection, production is stratified into intervals or “periods” of time. Monthly, weekly, or even daily intervals might be appropriate, depending on the situation (number of labor codes or failure modes that can be reported, costs of not detecting an important problem, costs of false alarms, as well as cost and availability of computer processing power, etc.). Following the existing practices of the automobile company that provided the part of the database used in our examples, we will use one-month intervals in our examples. The methodology and notation in our development are, however, general enough to be easily adjusted to other products and time intervals. In our presentation we refer to a generic “period” for the time interval.

2.5 Example

To illustrate the methods presented in this paper, we will use warranty data that have been provided to us by an automobile manufacturer. Automobile warranties in the North America are two-dimensional. Typically the bumper-to-bumper warranties are for 36 thousand miles or 36 months, whichever ever comes first (although a number of possible variations exist, especially for parts of the automobile like the drive train). Two-dimensional warranties can complicate the analysis of warranty data as automobiles can leave the warranty region through “mileage-out.” Analysts do not have information on the number of miles driven for specific automobiles that have not returned for service. Which and when individual automobiles have exceeded the warranty mileage limit will, in general, not be known. For purposes of detection of changes in report rates over time, mileage-out

is not a problem.

The VIN (production) database used in our examples contains information on 566,406 different automobiles. These are automobiles, of a particular model, that were manufactured between January 1995 and August 1998. The report database provides information on all warranty reports received between January 1995 and November 1998 (the “data freeze” date). The report database contains a total of 1,350,675 records with 1,908 unique labor codes. To protect proprietary information, the make and model of the automobile and the precise meaning of the labor codes cannot be disclosed. Some labor codes have a large number of reports (in the thousands). Others have few or none. All of the unique labor codes were investigated individually as part of our exploratory data analyses. Most labor codes had patterns that were relatively stable over the 3.5 years of production. Many others showed dramatic changes.

Figure 1 provides a retrospective view of the fraction of automobiles with warranty reports for 12 selected labor codes. The particular labor codes in Figure 1 were chosen to give a sense of the different patterns and report rate levels that were observed in our exploratory analysis, with emphasis on those that exhibited rate changes over time. For example, the report rates for labor codes A5580 and N3350 are approximately constant over time, but the rate for labor code A5580 was more than an order of magnitude larger. The indications in the plot for N3350 correspond to a single report in the few production months for which there was a report.

The rate for E0432 suddenly jumped up and then continued to increase for a long period of time. The rate for labor code E9995 had a steady downward trend until it became negligible. Labor codes B0608, D3088, and N2015 indicate other warranty problems that were, apparently, solved.

Labor code C3301 jumped up for only in production month July 1997. In the first four months of service, there were 19 reports that caused the July 1997 production month spike. There were, however, only 3856 cars produced in July 1997; the average number for the other production months was about 13,000. The other production months had on the order of 8 reports in the first four months of service with approximately three times the production. The 4-month C3301 report rate for production month July 1997 was about .00505. For the other production months, the rate ranged from .00012 to .00156, with an average rate of .00071.

The focus in this paper is on early detection of changes for patterns like those seen in C0176, D4450, E0432, J4640 and T2020 where the report rate was relatively stable for a period of time, but then suddenly increased for a number of production months. For more detailed study we identified 48 labor codes for which there were at least 18 months with a stable report rate (so that we could establish a base line report rate from the data) followed by a sudden increase in the rate that persisted for at least two production months. We will focus on one of these labor codes for our detailed examples. In Section 7 we will report on some general findings from this larger selected group of labor codes.

For our detailed examples we will use the labor code designated C0140. The letter C indicates a particular subsystem in the automobile (e.g., the power train) and the number indicates the particular action that was taken. Figure 2 provides a retrospective view giving the fraction of automobiles with labor code C0140 warranty reports for different “production months.” The plot shows that there was serious deterioration of reliability starting in April 1997. In the following sections we will illustrate methods for early detection of problems like this.

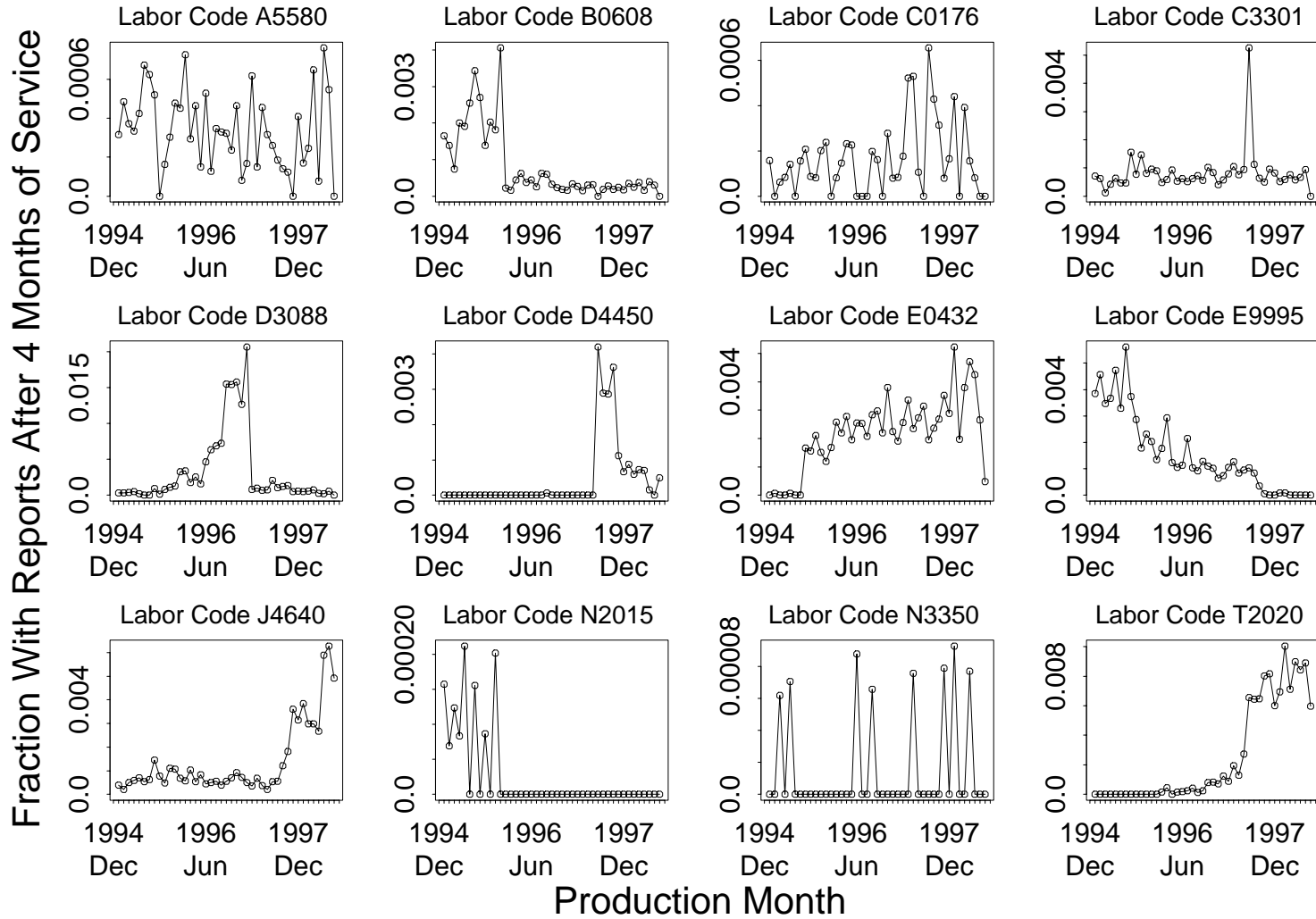


Figure 1: Retrospective View Giving the Fraction of Automobiles With Warranty Reports in the First Four Months of Service as a Function of Production Month for 12 Different Labor Codes.

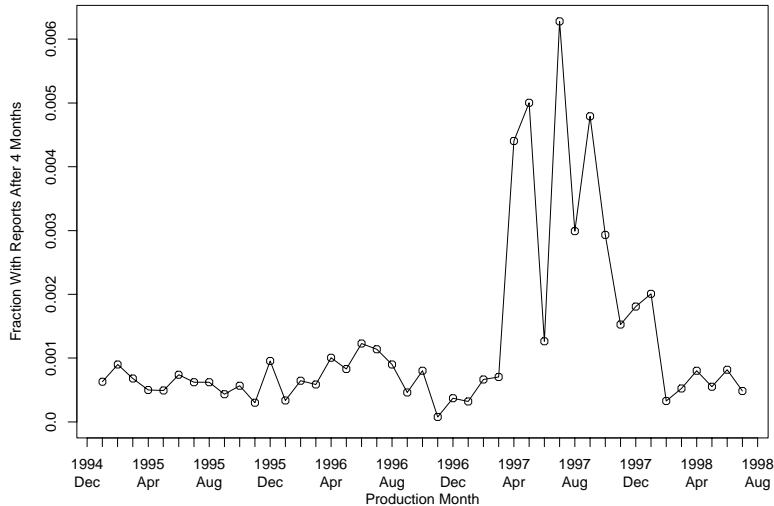


Figure 2: *Retrospective View Giving the Fraction of Automobiles With Labor Code C0140 Warranty Reports in the First Four Months of Service as a Function of Production Month.*

3 Formulation of the Detection Problem

This section outlines the statistical formulation of the early-warning detection problem, including discussion of the information available and the development of the detection rules.

3.1 Notation

The following notation will be used to describe the data from the warranty database used in the detection procedure. Let n_i denote the number of units produced in period i , and let n_{ij} denote the number of units produced in period i and sold in period $i + (j - 1)$, $j = 1, 2, \dots$ (i.e., sold in the j th period after they were manufactured). Also for a particular labor code under consideration, let R_{ijk} denote the number of warranty reports during the k th period in service for units that are manufactured in period i and sold in period $i + (j - 1)$. Note that R_{ijk} first becomes available in period $i + (j - 1) + k$.

3.2 Information for detection

The early detection problem can be viewed as an inference on product reliability. It might be suggested that it would be appropriate to fit a standard parametric distribution such as a Weibull or a lognormal distribution to provide structure for the needed inference. Instead, we recommend a nonparametric approach based on warranty report counts modeled with a Poisson distribution with report intensities that depend on production period and number of periods in service. This is equivalent to fitting a piece-wise exponential distribution to the available data and does not require specification or use of a particular distributional form for the time-to-report distribution. We use

this approach for the following reasons.

1. The implied underlying time-to-report model is flexible, allowing for different behaviors without having to make time-consuming modeling choices for each labor code. (There are 1,908 unique labor codes in the database used in our examples.)
2. There is, in most commercial applications, a sufficient amount of data to support the estimation of the (potentially) large number of report intensity parameters.
3. The detection decisions are made on the basis of report counts, which, along with historical report rates, contain all of the available information for the detection procedure.
4. The detection procedure proposed here allows flexibility in allocating power to detect problems in different parts of a unit's service life and to provide, if desired, special focus on a suspected labor code or codes (e.g., corresponding to a particular component or labor code that has had a recent design change).

Following the statistical model used in Kalbfleisch, Lawless, and Robinson (1991), we assume that R_{ijk} can be described as independently distributed Poisson ($n_{ij}\lambda_k$) random variables, where λ_k represents the report intensity for units during their k th service period (for the particular labor code under consideration). The use of this probability model is strongly supported by most warranty applications where there is a large number of units in the field, but the occurrence of any given failure mode, when reliability is as expected, should be rare and statistically independent from unit to unit with no underlying seasonality. (The detection rules developed in this paper can also be used when there is seasonality in warranty reports; see discussions in Section 8.) The reference value for λ_k , denoted by λ_k^0 , can be obtained based on historical records of report intensities for this type of unit over previous production periods (several years). In the absence of such historical data, λ_k^0 can be obtained from previous experience with similar products or design specifications.

3.3 Detection rule framework

The formal problem of detection of reliability deterioration can be formulated as a test of the multiple-parameter hypothesis

$$\begin{aligned}
 H_0: \lambda_1 \leq \lambda_1^0, \lambda_2 \leq \lambda_2^0, \dots, \lambda_M \leq \lambda_M^0 \\
 \text{versus } H_a: \lambda_1 > \lambda_1^0 \text{ or } \lambda_2 > \lambda_2^0 \text{ or } \dots \text{ or } \lambda_M > \lambda_M^0,
 \end{aligned}
 \tag{1}$$

where M is the prespecified number of future periods for which the report intensities will be monitored for units manufactured in any given production period. For a given overall false alarm rate, increasing M will require a reduction in power in order to spread protection over a larger number of monitoring periods.

Consider production period i . In this period, n_i units were manufactured, and n_{i1} of these units were sold. Among these, there were R_{i11} warranty reports during their first period of service, and these R_{i11} reports first became available in period $i + 1$. Note that $R_{i11} \sim \text{Poisson}(n_{i1}\lambda_1)$, and in

period $i + 1$, we can test only $\lambda_1 \leq \lambda_1^0$ versus $\lambda_1 > \lambda_1^0$; no information is available on $\lambda_2, \dots, \lambda_M$. In general, in period $i + k$, k periods after the units in the i th production period were produced and one period after $n_{i1} + \dots + n_{ik}$ of these were sold, we can test the joint hypothesis of whether $\lambda_1 \leq \lambda_1^0, \dots, \lambda_k \leq \lambda_k^0$ or not.

For testing λ_k , only the $R_{ijk}, j = 1, 2, \dots$ are relevant; the other $R_{ij\ell} (\ell \neq k)$ contain no information about λ_k . Because R_{ijk} and $R_{ij\ell} (\ell \neq k)$ are independent, testing $H_0: \lambda_1 \leq \lambda_1^0, \dots, \lambda_M \leq \lambda_M^0$ versus $H_a: \lambda_k > \lambda_k^0$ for some $k, k = 1, \dots, M$ can be done by testing, individually, $H_0^k: \lambda_k \leq \lambda_k^0$ versus $H_a^k: \lambda_k > \lambda_k^0$ for $k = 1, \dots, M$.

Consider first testing $H_0^1: \lambda_1 \leq \lambda_1^0$ versus $H_a^1: \lambda_1 > \lambda_1^0$, the Poisson report intensity for the first period in service. In period $i + 1$, we conclude that $\lambda_1 > \lambda_1^0$ if $R_{i11} \geq C_{i11}$ for some critical value C_{i11} (to be determined). In subsequent periods, additional information on the first period in service for production period i will accumulate due to units that were sold some number of periods after they were produced. In general, in period $i + j$, we will conclude that $\lambda_1 > \lambda_1^0$ if $S_{ij1} \geq C_{ij1}$, where $S_{ij1} = \sum_{\ell=1}^j R_{i\ell 1}$ is the cumulative number of reports during the first period in service for the units manufactured in period i . The optimality of such a rule arises from the theory of group sequential tests (e.g., Pocock 1977; Jennison and Turnbull 2000). The false alarm (Type I error) probability of this particular sub-test is $\alpha_1^* = 1 - \Pr(S_{i11} < C_{i11}, \dots, S_{iM1} < C_{iM1})$, which is less than or equal to α_1 , the nominal false alarm probability for this sub-test. Because of the discreteness of the Poisson distribution, it is typically impossible to have $\alpha_1^* = \alpha_1$. Throughout this paper, $\Pr(\cdot)$ is with respect to H_0 unless stated otherwise. The choices of α_1 and C_{i11}, \dots, C_{iM1} are important aspects of this test and will be described in Sections 4 and 5.1, respectively.

Similarly, for testing $H_0^k: \lambda_k \leq \lambda_k^0$ versus $H_a^k: \lambda_k > \lambda_k^0$ (the report intensity for the k th period in service), in period $i + (j - 1) + k$ (i.e., the j th period after information on service period k first becomes available), we conclude that $\lambda_k > \lambda_k^0$ if $S_{ijk} \geq C_{ijk}$, where $S_{ijk} = \sum_{\ell=1}^j R_{i\ell k}$. The false alarm probability of this sub-test is

$$\alpha_k^* = 1 - \Pr(S_{i1k} < C_{i1k}, \dots, S_{i, M-k+1, k} < C_{i, M-k+1, k}) \leq \alpha_k, \quad (2)$$

where α_k is the nominal false alarm probability. The choice of α_k and the determination of $C_{i1k}, \dots, C_{i, M-k+1, k}$ will be discussed in Sections 4 and 5.1, respectively.

Due to the independence of the tests for the different number of periods in service, the *overall* false alarm probability for testing the hypothesis H_0 versus H_a in (1) is then

$$\alpha^* = 1 - \prod_{k=1}^M (1 - \alpha_k^*) \leq 1 - \prod_{k=1}^M (1 - \alpha_k) = \alpha, \quad (3)$$

where α_k^* and α_k are given by (2) and α is the nominal overall false alarm probability.

3.4 Example

Figure 3 shows sequential test monitoring charts that would be available in August 1997 for labor code C0140 for the production periods from April to July 1997. Here for illustrative purposes, we have chosen $M = 4$ and $\alpha = .1\%$ and taken January 1995 to be the first production period. Computation of the critical limits will be described in Section 5.1. Figure 3 shows the detection alarm in the first service month of production month May 1997. (An alarm would be signaled,



Figure 3: Sequential Test Monitoring Charts for Labor Code C0140 Warranty Reports With + Indicating the Cumulative Number of Reports S_{ijk} and - Indicating the Corresponding Critical Limit C_{ijk} .

respectively, in both July and August 1997.) Section 5.2 gives details describing the reason for this detection.

In each subsequent monitoring period (month in this example), additional report information becomes available. As more production periods pass, additional sequential test charts become available, and some expire. In production period i ($i \leq M + 1$), there will be $[i(i - 1)]/2$ charts. At any point in time after there have been $M + 1$ production periods, there will be the maximum of $[M(M + 1)]/2$ such sequential test charts under consideration.

3.5 Relationship to Shewhart process monitoring schemes

The detection procedure presented here can be viewed as a generalization of a Shewhart process monitoring scheme (Shewhart 1931), where after each monitoring increment the available data are used to determine whether the process is out of control or not. In the warranty monitoring problem, however,

- Multivariate statistical inferences are being made over time about a report process as a function of periods in service parameterized by $(\lambda_1, \lambda_2, \dots, \lambda_M)$ and,
- Data on a particular period in service within a production period accumulate over time, due to sales that are staggered over time.

Instead of a single set of control limits, the set of critical values C_{ijk} is obtained in a way such that the nominal overall false alarm probability is small, say on the order of $\alpha = .1\%$ to 1% . When an R_{ijk} reaches its corresponding critical value within a particular sequential test for a production period/period in service combination, an out-of-control signal is triggered. In this case, when the manufacturing process is in control, the average run length (ARL) will be at least $1/\alpha$, as shown in Section 6.1. This relationship is similar to the traditional Shewhart chart.

Of course, different choices of C_{ijk} can be used to provide control charts with different performance properties, as described in the following sections. Relatedly, warning limits W_{ijk} can be obtained in a manner that is similar to that used to compute the C_{ijk} by setting α at 1% to 5% , and reaching W_{ijk} would suggest an action different from that of reaching C_{ijk} . Such warning limits might also be used as a basis for runs rules like those used in Shewhart monitoring schemes.

4 Allocation of False Alarm Probability and Power for Detection

4.1 Allocation of false alarm probability

Section 3.3 defined α to be the nominal overall false alarm probability for testing the hypothesis in (1), computed as $\alpha = 1 - \prod_{k=1}^M (1 - \alpha_k)$, where α_k is the nominal false alarm probability for testing the sub-hypothesis H_0^k versus H_a^k about λ_k , corresponding to the k th period in service. Given α , the choice or allocation of $\alpha_1, \dots, \alpha_M$ depends on specific considerations in the application. For example, if early detection of a potential reliability problem is critical and if it is believed that problems could arise in the early periods of service life, then one might choose the first several α_k 's

(e.g., $\alpha_1, \alpha_2, \alpha_3$, etc.) to be larger than those at the end of the monitoring period (i.e., $\alpha_{M-2}, \alpha_{M-1}, \alpha_M$, etc.).

To balance between quick detection and the overall probability of detection (power) over potential reliability problems over the first M periods of a unit's life, we propose a simple rule that chooses α_k to be proportional to the information available for testing H_0^k versus H_a^k . This information is proportional to the expected number of reports during the k th period in service. Thus we suggest

$$\alpha_k = \mathcal{C}(f_1^0 + \cdots + f_{M-k+1}^0)\lambda_k^0, \quad (4)$$

where the $f_1^0 + \cdots + f_{M-k+1}^0$, based on historical records, is the expected fraction of units sold during the first $M - k + 1$ periods after production and \mathcal{C} is a constant such that $\alpha = 1 - \prod_{k=1}^M (1 - \alpha_k)$. That is,

$$1 - \prod_{k=1}^M [1 - \mathcal{C}(f_1^0 + \cdots + f_{M-k+1}^0)\lambda_k^0] = \alpha. \quad (5)$$

Note that $\sum_{j=1}^M f_j^0 \leq 1$ and that the f_j^0 are likely to depend on the production period (e.g., inventory of some products may sell much more rapidly during the period that a new model is released or during the gift-giving season), although this dependency has been suppressed in our notation. Because $\mathcal{C}(f_1^0 + \cdots + f_{M-k+1}^0)\lambda_k^0 = \alpha_k$ is generally small, we can use the approximation

$$\mathcal{C} \approx \frac{\alpha}{\sum_{k=1}^M (f_1^0 + \cdots + f_{M-k+1}^0)\lambda_k^0}.$$

Alternatively, solving (5) for $\log(1 - \alpha)$ gives a function of \mathcal{C} that is increasing, making it easy to use a numerical root-finding algorithm to obtain \mathcal{C} and this is the approach that we used in our computations.

4.2 Error allocation for the sequential test region

Once $\alpha_1, \dots, \alpha_M$ (the nominal false alarm probabilities for the M service periods in the detection procedure) have been determined for a given production period, it is necessary to choose error probabilities corresponding to the sequential test region for testing λ_k for that production period. The sequential test for a particular period in service within a given production period assesses information as it accumulates on units that were produced in one production period but sold over a number of periods (i.e., accounting for sales delay). In particular, for production period i ,

$$1 - \Pr(S_{i1k} < C_{i1k}, \dots, S_{ijk} < C_{ijk})$$

for $k = 1, \dots, M$ and $j = 1, \dots, M - k + 1$ is the *cumulative* probability of a false alarm at the monitoring point that comes the j th period after information on service period k first becomes available. Note that

$$1 - \Pr(S_{i1k} < C_{i1k}, \dots, S_{ijk} < C_{ijk}) \leq \alpha_{jk}^{(i)},$$

where $\alpha_{jk}^{(i)}$ is the corresponding nominal false alarm probability and $\alpha_{M-k+1,k}^{(i)} \equiv \alpha_k$.

We will use the *error spending* approach for the choice of these $\alpha_{jk}^{(i)}$. This approach was originally developed for sequential clinical trials by Slud and Wei (1982), Lan and DeMets (1983), among others. It is explained in detail, for example, in Chapter 7 of Jennison and Turnbull (2000). The basic idea is as follows.

For the first period in service, consider testing $H_0^1: \lambda_1 \leq \lambda_1^0$ versus $H_a^1: \lambda_1 > \lambda_1^0$. The information available for testing λ_1 is proportional to the number of units sold during the first j periods after production (i.e., $n_{i1} + \dots + n_{ij}$), and the proportion of information accumulated during these periods is

$$\frac{f_{i1} + \dots + f_{ij}}{f_{i1} + \dots + f_{iM}},$$

where $f_{ij} = n_{ij}/n_i$ is the fraction of units produced in production period i and sold in the j th period after having been produced. The error spending approach uses a nondecreasing function $g(t)$ such that $g(0) = 0$ and $g(t) = \alpha_1$ for $t \geq 1$. It then assigns an amount

$$\alpha_{j1}^{(i)} = g\left(\frac{f_{i1} + \dots + f_{ij}}{f_{i1} + \dots + f_{iM}}\right).$$

A simple, flexible choice of $g(t)$ is

$$g(t) = \begin{cases} \alpha_1 t^\rho & 0 \leq t \leq 1 \\ \alpha_1 & t > 1, \end{cases} \quad (6)$$

as suggested in Jennison and Turnbull (2000, chap. 7). Kim and DeMets (1987) studied the cases $\rho = 1, 1.5$, and 2.

The error spending method is widely used in clinical trial applications where sequential decision rules are used and the emphasis is often placed on overall power. Different choices of ρ (e.g., $\rho = .5, 1, 2$) provide different amounts of emphasis on early versus later detection, in terms of power. In our applications, early detection is important. This suggests choosing a smaller value of ρ such as $\rho = .5$ or 1, allocating somewhat more power to earlier detection opportunities.

For testing other λ_k values (corresponding to other numbers of periods in service), we can choose different error spending functions [e.g., by changing the value of ρ in (6)]. Allowing for the differing amounts of information available for different numbers of periods in service,

$$\alpha_{jk}^{(i)} = \alpha_k \times \left(\frac{f_{i1} + \dots + f_{ij}}{f_{i1} + \dots + f_{i,j+k} + f_{i,j+k+1} + \dots + f_{i,M-k+1}} \right)^\rho. \quad (7)$$

In actual use it will be necessary to replace the $f_{i,j+k+1}, \dots, f_{i,M-k+1}$ in (7) by estimates of frequencies based on past data, $f_{j+k+1}^0, \dots, f_{M-k+1}^0$ because the actual fractions sold in these periods are not available at the time the $\alpha_{jk}^{(i)}$ values need to be computed (i.e., at the end of period $i + (j-1) + k$). After the $\alpha_{jk}^{(i)}$ have been computed, the computation of the critical values C_{ijk} can be done based on a recursive formula, as shown in Section 5.1.

4.3 Example

To illustrate the method of allocation of false alarm probabilities in the sequential test we use $\alpha = .1\%$ and $M = 4$ with labor code C0140 for production month May 1997. The historical sale patterns f_1^0, \dots, f_4^0 and report intensities (for this labor code) $\lambda_1^0, \dots, \lambda_4^0$ are computed from information available up to January 1997 based on the 24 production months from January 1995 (production month one) to December 1996 (production month 24). That is,

$$f_j^0 = \frac{\sum_{i=1}^{25-j} n_{ij}}{\sum_{i=1}^{25-j} n_i}, \quad \text{for } j = 1, 2, 3, 4, \quad (8)$$

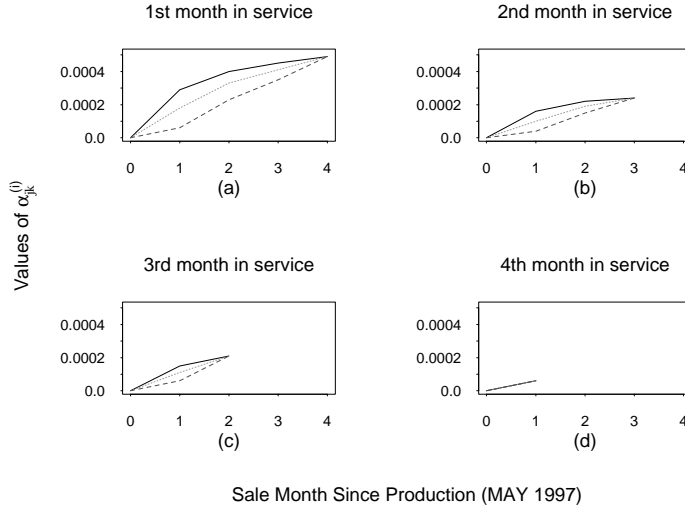


Figure 4: *Typical Spending Functions for Production Month May 1997 for Labor Code C0140 With $\alpha = .1\%$ and $M = 4$, and $\rho = .5$, —; $\rho = 1$, \cdots ; and $\rho = 2$, - - -.*

and

$$\lambda_k^0 = \frac{\sum_{i=1}^{25-k} \sum_{j=1}^{\min(4, 26-i-k)} R_{ijk}}{\sum_{i=1}^{25-k} \sum_{j=1}^{\min(4, 26-i-k)} n_{ij}}, \quad \text{for } k = 1, 2, 3, 4. \quad (9)$$

Using (8) and (9) and the data for labor code C0140 gives

$$f_1^0 = .133, f_2^0 = .241, f_3^0 = .165, f_4^0 = .123 \text{ and} \\ \lambda_1^0 = .00022, \lambda_2^0 = .00013, \lambda_3^0 = .00016, \lambda_4^0 = .00014.$$

Based on these results, the nominal false alarm probability α_k for testing H_0^k versus H_a^k can be obtained using the method outlined in Section 4.1 and (4). We have $\alpha_1 = .00049$, $\alpha_2 = .00024$, $\alpha_3 = .00021$, and $\alpha_4 = .00006$.

Note that, for purposes of illustration, we focus only on a single production month. The actual monitoring process will monitor all of the most recent M production months (not including the current production month) simultaneously after the monitoring process has been in operation for M months. (For $M = 4$ here, the production months from April 1997 to July 1997 would be monitored simultaneously in August 1997.) For production month May 1997, $i = 29$ and $n_{29} = 13203$, with sales in the first four months after production being $n_{29,1} = 4198$, $n_{29,2} = 3659$, $n_{29,3} = 1991$, and $n_{29,4} = 1791$. The $\alpha_{jk}^{(29)}$ values are computed using (7). For the detection rule outlined in Section 3.3, $\alpha_{11}^{(29)}$ would be computed in June 1997, $\alpha_{21}^{(29)}$ and $\alpha_{12}^{(29)}$ in July, $\alpha_{31}^{(29)}$, $\alpha_{22}^{(29)}$, and $\alpha_{13}^{(29)}$ in August, and $\alpha_{41}^{(29)}$, $\alpha_{32}^{(29)}$, $\alpha_{23}^{(29)}$, and $\alpha_{14}^{(29)}$ in September 1997.

Figure 4 illustrates the error spending functions in (7) for values $\rho = .5, 1$, and 2 . For example, Figure 4(a) gives the error allocations $\alpha_{11}^{(29)}$, $\alpha_{21}^{(29)}$, $\alpha_{31}^{(29)}$, and $\alpha_{41}^{(29)} = \alpha_1$ for the first month in service for $\rho = .5, 1$, and 2 . Clearly, a smaller value of ρ such as $\rho = .5$ gives more power to earlier detection

opportunities. Figures 4(b), (c), and (d) give error allocations for the second, third, and fourth service month, respectively. Note that, for the fourth service month (here $M = 4$), only data for the first sale month is available, and thus $\alpha_{14}^{(29)} = \alpha_4$ does not depend on the choice of ρ .

5 Computing the Critical Values

This section describes and illustrates the use of an algorithm to compute critical values for the warranty report monitoring procedure.

5.1 Recursive algorithm

First note that $S_{ijk} \sim \text{Poisson}(n_i(f_{i1} + \dots + f_{ij})\lambda_k^0)$. For a given production period i , service period k , and false alarm allocations $\alpha_{jk}^{(i)}$, $j = 1, \dots, M - k + 1$, the following recursive algorithm can be used to compute the sequential test critical values $C_{i1k}, \dots, C_{i, M-k+1, k}$.

1. Let r denote the smallest value of s such that

$$\sum_{y=0}^s \Pr(S_{i1k} = y) \geq 1 - \alpha_{1k}^{(i)}.$$

Then $C_{i1k} = r + 1$.

2. For $\ell = 2, \dots, M - k + 1$, to obtain $C_{i\ell k}$, given $C_{i1k} \dots C_{i, \ell-1, k}$, compute

$$\begin{aligned} & \Pr(S_{i1k} < C_{i1k}, \dots, S_{i, \ell-1, k} < C_{i, \ell-1, k}, S_{i\ell k} = U) \\ &= \sum_{u=L}^U \Pr(S_{i1k} < C_{i1k}, \dots, S_{i, \ell-2, k} < C_{i, \ell-2, k}, S_{i, \ell-1, k} = U - u) \\ & \quad \times \Pr(R_{i\ell k} = u) \end{aligned} \tag{10}$$

for $U = 0, 1, \dots$, where $L = \max(0, U + 1 - C_{i, \ell-1, k})$ and $R_{i\ell k} = S_{i\ell k} - S_{i, \ell-1, k} \sim \text{Poisson}(n_i f_{i\ell} \lambda_k^0)$.

3. Let r denote the smallest value of s such that

$$\sum_{y=0}^s \Pr(S_{i1k} < C_{i1k}, \dots, S_{i, \ell-1, k} < C_{i, \ell-1, k}, S_{i\ell k} = y) \geq 1 - \alpha_{\ell k}^{(i)}.$$

Then $C_{i\ell k} = r + 1$.

The recursive formula (10) can be derived along the same lines as that for the binomial distribution (Schultz et al. 1973). Such recursive formulas are computationally efficient and are widely used in group sequential tests (e.g., see p. 237 of Jennison and Turnbull 2000). To implement (10), first compute $\Pr(S_{i1k} = U)$ for $U = 0, \dots, C_{i1k} - 1$ (and save these results). Then if $\ell \geq 3$, compute $\Pr(S_{i1k} < C_{i1k}, \dots, S_{i, j-1, k} < C_{i, j-1, k}, S_{ij k} = U)$ using (10) for $U = 0, \dots, C_{ij k} - 1$ (and save these results), successively for $j = 2, \dots, \ell - 1$. Finally compute $\Pr(S_{i1k} < C_{i1k}, \dots, S_{i, \ell-1, k} < C_{i, \ell-1, k}, S_{i\ell k} = U)$ using (10). The appendix gives the sequence of steps used to compute the C_{ijk} values.

Table 1: *Example Data and Critical Values for the C0140 Monitoring Process in August 1997 for the April to July 1997 Production Months*

<i>Production Month</i>	<i>Service Month</i>	<i>Month R_{ijk} Becomes Available</i>	<i>Historical Number of Reports</i>	<i>Number of Reports</i>	<i>Cumulative Number of Reports</i>	<i>Critical Value</i>
i	k	$i + j + k - 1$	$n_{ij}\lambda_k^0$	R_{ijk}	S_{ijk}	C_{ijk}
APR 1997	1	MAY 1997	.612	0	0	6
APR 1997	1	JUN 1997	.785	3	3	8
APR 1997	1	JUL 1997	.435	3	6	9
APR 1997	1	AUG 1997	.366	2	8	10
APR 1997	2	JUN 1997	.371	2	2	5
APR 1997	2	JUL 1997	.475	3	5	7
APR 1997	2	AUG 1997	.263	1	6	7
APR 1997	3	JUL 1997	.462	2	2	6
APR 1997	3	AUG 1997	.592	2	4	7
APR 1997	4	AUG 1997	.394	2	2	5
MAY 1997	1	JUN 1997	.907	5	5	7
MAY 1997	1	JUL 1997	.791	4	9	9
MAY 1997	1	AUG 1997	.430	2	11	9
MAY 1997	2	JUL 1997	.549	2	2	6
MAY 1997	2	AUG 1997	.479	4	6	7
MAY 1997	3	AUG 1997	.684	4	4	6
JUN 1997	1	JUL 1997	.790	0	0	6
JUN 1997	1	AUG 1997	.531	0	0	8
JUN 1997	2	AUG 1997	.478	0	0	6
JUL 1997	1	AUG 1997	.022	0	0	3

5.2 Example

Table 1 gives the observed number of reports R_{ijk} and cumulative number of reports $S_{ijk} = R_{i1k} + R_{i2k} + \dots + R_{ijk}$ for labor code C0140 up to August 1997 for the April to July 1997 production months. The third column in Table 1 indicates when R_{ijk} and S_{ijk} values first become available and when the corresponding critical value C_{ijk} is computed. The computation of C_{ijk} is based on the recursive algorithm given in Section 5.1. Here we use $\alpha = .1\%$, $M = 4$, and $\rho = 1$. The historical sale patterns f_1^0, \dots, f_4^0 and report intensities $\lambda_1^0, \dots, \lambda_4^0$ were computed as described in Section 4.3. The values of $\alpha_1, \dots, \alpha_4$ are also given there.

For illustration, consider the first month in service for the May 1997 production month ($i = 29$). In June 1997, $f_{29,1} = n_{29,1}/n_{29}$ and $f_{29,2} = n_{29,2}/n_{29}$ are available and $\alpha_{11}^{(29)}$ is computed using (7). Then $C_{29,1,1} = 7$ is obtained using step 1 of the recursive algorithm. In July 1997, $f_{29,3}$ becomes available and $\alpha_{21}^{(29)}$ is computed using (7). Then $C_{29,2,1} = 9$ is obtained using steps 2 and 3 of the algorithm. Similarly, $C_{29,3,1} = 9$ is computed in August 1997. These $C_{29,j,1}$ and the corresponding

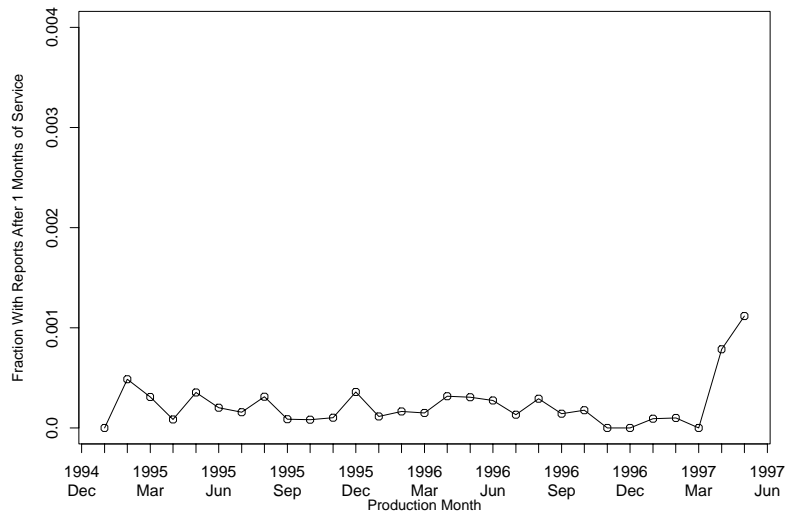


Figure 5: *Nonparametric Estimate of Fraction With Reports After One Month of Service as a Function of Production Month for Data Available in July 1997.*

$S_{29,j,1}$ values are plotted in the first plot of the second row of Figure 3. In general, for each production month and service month, there is one plot in the sequential test monitoring charts. Here Table 1 gives all of the S_{ijk} and C_{ijk} values needed for producing Figure 3. Note that $S_{29,2,1} = C_{29,2,1} = 9$ and $S_{29,3,1} = 11 > C_{29,3,1} = 9$, indicating that an alarm would be signaled, respectively, in both July and August 1997. The historical number of reports $n_{ij}\lambda_k^0$, which is $E(R_{ijk})$ under H_0 , is indeed much smaller than the observed R_{ijk} for the first month in service for the May 1997 production month, as can be seen from the table.

After an alarm is signaled, analysts should look more carefully at the reasons (both statistical and physical) for the signal. Physically, engineers would be interested in doing careful physical failure mode analysis of some of the units that failed (and it is often a good idea to look carefully at units of a similar vintage that have not failed to see if they are headed for the same fate). Statistically, several further graphical analyses can be used to supplement Figure 3. Figure 5 provides a followup chart giving a nonparametric (Kaplan-Meier) estimate of the fraction of automobiles with C0140 reports in the first month of service (because the alarm signal was for the first month of service) as a function of production month, based on data available in July 1997. This plot shows clearly the reason for the signal for production month May 1997. Figures 3 and 5 and Table 1 show that there was also some evidence of problems in April 1997, but not enough for the signal at the chosen $\alpha = .1\%$ false alarm probability and spending function.

After a problem is detected other important questions focus on the magnitude of the problem. Two general classes of problems exist. In some cases the detected problem is likely to affect all or nearly all units in service. In other cases the problem affects only a proportion of units (sometimes called a limited failure population or LFP model, as described in Meeker 1987). Figure 6 is a Weibull probability plot representing the 9 reports in the first service month and the 2 additional reports in

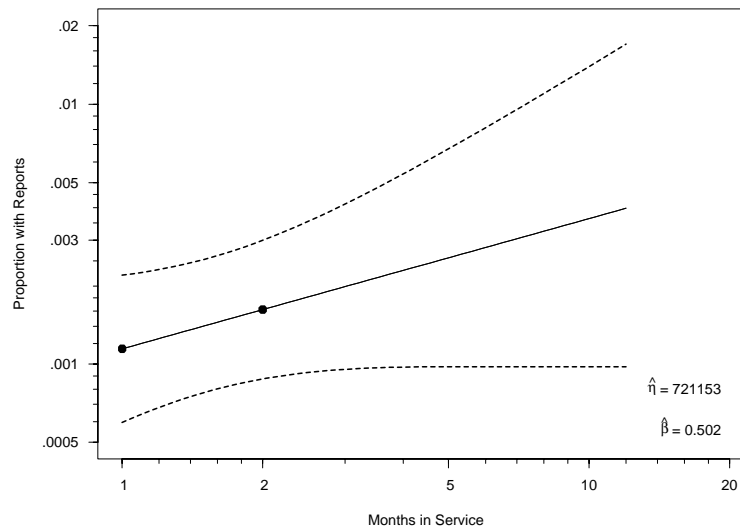


Figure 6: *Weibull Probability Plot and Weibull ML Estimate for Production Month May 1997 Based on Data Available in July 1997 (after two months in service).*

the second service month. The plotted points are nonparametric (adjusted Kaplan-Meier) estimates of fraction with reports as a function of month in service. The line is the corresponding Weibull ML estimate and approximate 95% pointwise confidence intervals, extrapolated to 12 months. Extrapolation has risks and can only be expected to provide adequate projections for a limited amount of time into the future, unless there is firm knowledge about the failure mode and its failure-time distribution.

Figure 7 is a Weibull probability plot representing the failure-time data available in November 1998. This plot provides a retrospective view of the reliability of the automobiles manufactured in May 1997. This plot shows that the Weibull distribution provided adequate, but somewhat pessimistic, projections out to 6 to 8 months in service. The nonparametric estimate begins to level off (and deviate from the Weibull distribution) after 6 to 8 months and the last reports occurred after 14 months in service, with a cumulative fraction with reports just over 1%, indicating that the problem should be described by an LFP model. Generally failure mode analysis of failed units and autopsy of some unfailed units can help provide more timely information on whether the population is LFP or not.

6 Average Run Length

Average run length (or ARL, the mean time until a signal as a function of the process state) is an important metric for a monitoring scheme. Under H_0 , when the process is in the in-control state, the ARL should be large. In the presence of an important reliability problem, the ARL should be small enough to quickly detect the problem.

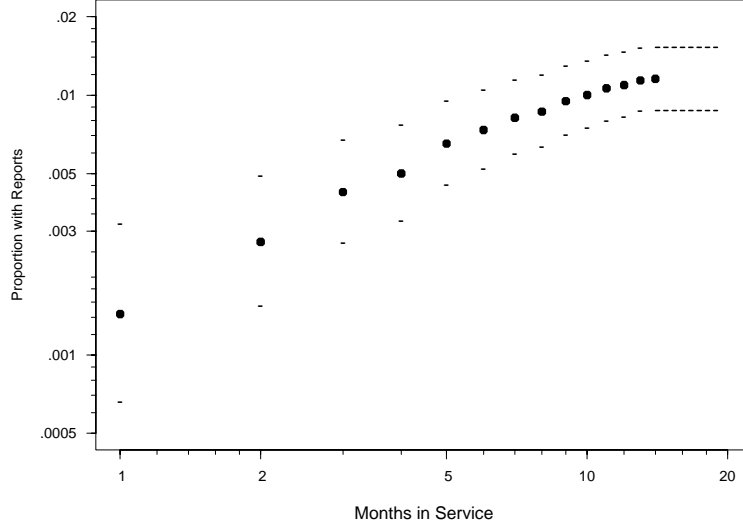


Figure 7: *Weibull Probability Plot for Production Month May 1997 Based on Data Available in November 1998.*

6.1 Average run length under H_0

Suppose the monitoring scheme is implemented for a particular labor code beginning with production period one. The first sequential test monitoring chart is plotted in period two. Suppose that the period in which the first alarm is triggered is denoted by $N + 1$. Then the ARL is $E(N)$, which can be expressed as

$$E(N) = \sum_{\ell=1}^{\infty} \Pr(N \geq \ell)$$

(e.g., Larsen and Marx 2001, p. 202). Note that “ $N \geq \ell$ ” implies that no alarm is triggered during the first $(\ell - 1)$ periods of monitoring. Let γ_{ij} ($j \leq M$) denote the probability that no alarm is triggered during the first j periods of monitoring of production period i . Let $\gamma_{ij} = \gamma_{iM}$ for $j > M$. Because each production period is monitored for M periods and the overall false alarm probability, as given in (3), is at most α , it follows that $\gamma_{ij} \geq 1 - \alpha$. During the first $(\ell - 1)$ periods of monitoring, $(\ell - 1)$ production periods are monitored independently. Thus

$$\Pr(N \geq \ell) = \prod_{i=1}^{\ell-1} \gamma_{i,\ell-i} \geq (1 - \alpha)^{\ell-1}.$$

Then

$$E(N) = \sum_{\ell=1}^{\infty} \Pr(N \geq \ell) \geq \sum_{\ell=1}^{\infty} (1 - \alpha)^{\ell-1} = \frac{1}{\alpha}.$$

That is, the ARL under H_0 is at least $1/\alpha$.

For the special case when production and sale patterns do not vary over time (i.e., $n_i = n$ is constant for all production periods, and $f_{ij} = f_j^0, j = 1, \dots, M$), $\gamma_{ij} = \gamma_j$ does not depend on i . In

this case,

$$\begin{aligned}
E(N) &= \sum_{\ell=1}^{\infty} \prod_{i=1}^{\ell-1} \gamma_{\ell-i} = \sum_{\ell=1}^{\infty} \prod_{j=1}^{\ell-1} \gamma_j \\
&= 1 + \gamma_1 + \gamma_1 \times \gamma_2 + \cdots + \gamma_1 \times \gamma_2 \times \cdots \times \gamma_{M-2} \\
&\quad + \gamma_1 \times \gamma_2 \times \cdots \times \gamma_{M-1} \times (1 + \gamma_M + \gamma_M^2 + \cdots) \\
&= 1 + \gamma_1 + \gamma_1 \times \gamma_2 + \cdots + \gamma_1 \times \cdots \times \gamma_{M-2} + \frac{\gamma_1 \times \gamma_2 \times \cdots \times \gamma_{M-1}}{1 - \gamma_M}.
\end{aligned}$$

Note that γ_j can be written as $\gamma_j = (1 - \alpha_{j1}^*) \times (1 - \alpha_{j-1,2}^*) \times \cdots \times (1 - \alpha_{1j}^*)$, where $1 - \alpha_{jk}^*$ represents the probability of no false alarm for monitoring the units produced in one period during their k th period in service and the first j periods after information about λ_k first becomes available. The computation of α_{jk}^* is as follows.

1. Let $\alpha_1, \dots, \alpha_M$ be the values computed in Section 4.1 and let

$$\begin{aligned}
\alpha_{jk} &= \alpha_k \times \left(\frac{f_1^0 + \cdots + f_j^0}{f_1^0 + \cdots + f_{M-k+1}^0} \right)^{\rho}, \text{ for } k = 1, \dots, M, \\
&\quad j = 1, \dots, M - k + 1.
\end{aligned}$$

2. Compute critical values C_{jk} using the recursive algorithm in Section 5.1, ignoring all the subscripts “ i ” involved there and noting that $S_{jk} \sim \text{Poisson}(n(f_1^0 + \cdots + f_j^0)\lambda_k^0)$.
3. Compute

$$\alpha_{jk}^* = 1 - \Pr(S_{1k} < C_{1k}, \dots, S_{jk} < C_{jk})$$

using the recursive algorithm in Section 5.1, again ignoring the subscripts “ i ” there.

6.2 Average run length under H_a

Suppose report intensities of manufactured products are in-control before period t_0 but that the report intensities change to $\lambda_k^a > \lambda_k^0$ for $k = 1, \dots, M$ in period t_0 and stay there in periods $t \geq t_0$. Denote the alternative hypothesis by H_a : $\lambda_k = \lambda_k^a, k = 1, \dots, M$. Suppose that the first alarm is signaled in period $N + t_0$ for monitoring the production periods $t \geq t_0$ under H_a . Similar to the derivations in Section 6.1, the ARL can be expressed as

$$E(N) = \sum_{\ell=1}^{\infty} \Pr(N \geq \ell),$$

where “ $N \geq \ell$ ” means that no alarm is triggered during the first $(\ell - 1)$ periods of monitoring of the production periods $t \geq t_0$. Note that $\Pr(\cdot)$ is with respect to H_a . Let δ_{ij} ($j \leq M$) denote the probability that no alarm is triggered during the first j periods of monitoring of production period $i \geq t_0$. Let $\delta_{ij} = \delta_{iM}$ for $j > M$. During the first $(\ell - 1)$ periods of monitoring production periods $i \geq t_0$, $(\ell - 1)$ such production periods are monitored independently. Since each production period is monitored for M periods,

$$\Pr(N \geq \ell) = \prod_{i=t_0}^{t_0+\ell-2} \delta_{i, (t_0+\ell-1-i)}.$$

For the special case when production and sale patterns do not vary over time, similar to the arguments in Section 6.1, $\delta_{ij} = \delta_j$ does not depend on i , and in this case, the ARL simplifies to

$$\begin{aligned} E(N) &= \sum_{\ell=1}^{\infty} \prod_{i=t_0}^{t_0+\ell-2} \delta_{t_0+\ell-1-i} = \sum_{\ell=1}^{\infty} \prod_{j=1}^{\ell-1} \delta_j \\ &= 1 + \delta_1 + \delta_1 \times \delta_2 + \cdots + \delta_1 \times \cdots \times \delta_{M-2} + \frac{\delta_1 \times \delta_2 \times \cdots \times \delta_{M-1}}{1 - \delta_M}. \end{aligned}$$

The computation of δ_j 's is similar to that of γ_j 's in Section 6.1 and is done by replacing λ_k^0 there by λ_k^a . That is, we now have $S_{jk} \sim \text{Poisson}(n(f_1^0 + \cdots + f_j^0)\lambda_k^a)$. Specifically, let $\beta_{jk}^* = 1 - \Pr(S_{1k} < C_{1k}, \dots, S_{jk} < C_{jk})$ for $k = 1, \dots, M, j = 1, \dots, M - k + 1$, where the C_{jk} values were computed as described in Section 6.1. Then β_{jk}^* can be computed the same way as α_{jk}^* in Section 6.1, and

$$\delta_j = (1 - \beta_{j1}^*) \times (1 - \beta_{j-1,2}^*) \times \cdots \times (1 - \beta_{1j}^*), \text{ for } j = 1, \dots, M.$$

6.3 Example

In Sections 6.1 and 6.2, formulas for computing ARL's under H_0 and H_a are presented for the special case when production and sale patterns do not vary over time. As the formulas indicate, these ARL's depend on many parameters, including $\alpha, M, f_1^0, \dots, f_M^0, \lambda_1^0, \dots, \lambda_M^0, \rho$, and n . Under H_a , they also depend on $\lambda_1^a, \dots, \lambda_M^a$. Table 2 gives some values of ARL's for $\alpha = .1\%, .5\%$, and 1% , $M = 4, 8$, and 12 , and $\rho = .5, 1$, and 2 under $H_0: \lambda_1 = \lambda_1^0, \dots, \lambda_M = \lambda_M^0$ and $H_a: \lambda_1 = \lambda_1^a, \dots, \lambda_M = \lambda_M^a$, where $\lambda_k^a = \lambda_k^0 + s\lambda_k^0$ for $k = 1, \dots, M$, with $s = 1, 2$ and 3 . In computing these ARL's, we choose $n = 13,000, f_1^0 = .15, f_2^0 = .25, f_3^0 = .15, f_4^0 = .12, f_5^0 = .09, f_6^0 = .07, f_7^0 = .05, f_8^0 = .04, f_9^0 = .03, f_{10}^0 = .02, f_{11}^0 = .01$, and $f_{12}^0 = .01$. This choice roughly represents an average production and a typical sale pattern based on the automobile database used in our examples. We also choose $\lambda_1^0 = .00025, \lambda_2^0 = .00015, \lambda_3^0 = .0002, \lambda_4^0 = .00015, \lambda_5^0 = .0001, \lambda_6^0 = .00015, \lambda_7^0 = .00005, \lambda_8^0 = .00005, \lambda_9^0 = .00005, \lambda_{10}^0 = .00007, \lambda_{11}^0 = .00008$, and $\lambda_{12}^0 = .00009$. This choice reflects the report intensities for the labor code C0140 used throughout this paper. Here for $s = 1, 2$, and 3 , $\lambda_k^a = \lambda_k^0 + s\lambda_k^0$ may indicate respectively a potentially mild, moderate, and serious reliability problem. (This is only for illustrative purposes. Whether a problem is serious or not depends on the specific application.)

As Table 2 shows, the ARL's under H_0 are greater than but close to $1/\alpha$. Due to the discreteness of the Poisson distribution involved, it is typically impossible to design a monitoring scheme such that the ARL's under H_0 are exactly equal to $1/\alpha$. As the last column of the table indicates, the monitoring schemes developed in this paper can often quickly detect a serious reliability problem. Here an alarm is typically triggered within three or four periods of monitoring after a serious problem occurs. The choice of α, M , and ρ can have a significant effect on the rate of false alarms and the ARL's, especially for detecting small to moderate reliability problems. This choice should be based on the specific application.

7 Summary of Analyses of Other Labor Codes

As described in Section 2.5, we retrospectively investigated the time series given by the number of warranty reports after four months of service and also after 12 months of service, as a function of

Table 2: *ARL's of Sequential Test Monitoring Charts With Parameters α , M , and ρ for Detecting Different Reliability Shifts*

α	M	ρ	Shift s ($\lambda_k = \lambda_k^0 + s\lambda_k^0, k = 1, \dots, M$)			
			0	1	2	3
.1%	4	.5	1510.20	32.41	6.57	3.61
.1%	4	1.0	1239.69	23.75	5.89	3.62
.1%	4	2.0	1376.23	23.86	5.92	3.67
.1%	8	.5	1262.29	18.37	5.76	3.76
.1%	8	1.0	1325.74	18.96	5.80	3.77
.1%	8	2.0	1477.81	18.58	6.16	4.19
.1%	12	.5	1139.33	18.14	5.97	3.81
.1%	12	1.0	1156.33	17.19	6.23	4.00
.1%	12	2.0	1294.60	17.43	6.38	4.22
.5%	4	.5	244.41	10.69	4.27	3.02
.5%	4	1.0	244.41	10.69	4.27	3.02
.5%	4	2.0	262.40	10.92	4.37	3.14
.5%	8	.5	232.59	9.74	4.40	3.07
.5%	8	1.0	260.96	9.86	4.60	3.20
.5%	8	2.0	276.32	10.13	5.01	3.61
.5%	12	.5	219.96	10.00	4.60	3.16
.5%	12	1.0	228.74	10.06	4.89	3.43
.5%	12	2.0	264.57	10.48	5.12	3.62
1%	4	.5	127.32	8.43	3.59	2.54
1%	4	1.0	132.98	8.29	3.64	2.68
1%	4	2.0	198.34	9.63	4.06	2.95
1%	8	.5	126.87	7.87	4.08	3.00
1%	8	1.0	133.00	8.04	4.24	3.05
1%	8	2.0	141.79	8.26	4.41	3.19
1%	12	.5	118.82	8.20	4.26	3.05
1%	12	1.0	120.05	8.31	4.36	3.10
1%	12	2.0	140.75	8.94	4.74	3.41

production month, for all of the 1,908 labor codes. Then we applied our detection procedure to 48 of the 1,908 labor codes for which there were at least 18 months with a stable report rate (so that we could establish a base line from the data) followed by a noticeable increase in the rate that persisted for at least two production months. The selection of these 48 labor codes was based on the retrospective plots for 12 months of service. For these evaluations we used an overall false alarm probability of $\alpha = .1\%$ and values of M ranging from 1 to 18. Without loss of useful information we report a summary of the results for $M = 4, 8,$ and 12 . There were some interesting differences among these three choices. Generally using M greater than 12 gave results that were similar to $M = 12$ and did not appear to importantly improve the ability to detect any problems that we discovered in our retrospective analysis of the warranty data.

In the labor code C0140 detailed example we used a monitoring period of $M = 4$ months and were able to detect the increase in report rate after only three months of monitoring (these three months include shipping time, time until sale, and time for the problem to be reported by the customer). Not all report rate increases can be detected so easily. For example, the increase in the report rate (starting from April 1997) for labor code C0176 is obvious in the retrospective plot (see Figure 1), but small relative to the background rate and could not be detected with $M = 4$ months of monitoring. This problem was, however, detected with both $M = 8$ and $M = 12$ months of monitoring.

Some labor codes emit signals very shortly after production (e.g., after two months) because at least some customers begin to have the problem shortly after the time of purchase. The fact that some problems were not detected until as many as 12 months after manufacturing is *not* a reflection of poor performance of the detection scheme. Unlike quality, field reliability can be quantified, in general, only after a sufficient amount of operating time has elapsed to excite the failure. This is because some physical failure modes require a certain amount of initiation time before they can occur. For example, with labor code T2020 (see Figure 1), deterioration of reliability started in October 1996 and gradually became very serious in July 1997. Our detection scheme first signaled the reliability problem in September 1997 (for $M = 8$ or 12) or October 1997 (for $M = 4$). To be conservative, we used October 1996 as the start of the reliability problem (which gradually became more serious in later months) and recorded the time to detection as 11 (for $M = 8$ or 12) or 12 (for $M = 4$) months. In this example, the earliest possible time of detection would be 11 months because there were only three reports before September 1997 (one in June and two in August 1997) for all the production months since October 1996 when a reliability problem started.

Table 3 summarizes the detection success for $M = 4, 8,$ and 12 , giving the distribution of time to a first detection among the 48 labor codes, as well as the distribution if one could have chosen the best M for each labor code. We would recommend doing such an analysis of different labor codes with historical data and using available engineering information in order to select an appropriate value of M for different labor codes. Generally, M should be smaller (larger) for labor codes with a reliability problem expected to appear earlier (later) in a product's life. Table 3 shows that with $M = 4$, some problems would not be detected and that for such labor codes a larger value of M should be used. The disadvantage of using $M = 8$ or $M = 12$ relative to $M = 4$ is that, for a fixed false alarm rate, available detection power has to be spread more broadly, generally lengthening the average time to detection; smaller values of M provide quicker detection for problems that can show up early. Table 4 provides a summary of the values of M that resulted in the quickest detection.

Table 3: *Frequency Distribution for Times to First Detection*

<i>Months to Detection</i>	<i>M = 4</i>	<i>M = 8</i>	<i>M = 12</i>	<i>Best of M = 4, 8, 12</i>
2	9	8	7	9
3	4	4	5	4
4	3	2	2	3
5	6	6	6	6
6	3	7	6	7
7	1	2	2	1
8	3	5	4	5
9	3	4	6	5
10	1	1	0	0
11	1	2	4	4
12	3	1	4	4
13	1	1	0	0
14	0	0	0	0
15	1	0	0	0
16	0	1	0	0
17	1	0	0	0
18	0	1	1	0
Undetected	8	3	1	0
Total number	48	48	48	48

Table 4: *Value of M Having the Quickest Detection*

<i>Quickest Detection</i>	<i>Number</i>	<i>Number Including Ties</i>
Same time for M=4, 8, and 12	23	
M=4 and 8	3	
M=8 and 12	8	
M=4	5	31
M=8	2	36
M=12	7	38
Total number	48	

8 Concluding Remarks and Directions for Future Research

This section describes some additional implementation issues and indicates some possible areas for future research.

1. For some labor codes in the warranty database there were long periods of time with 0 reports, followed by a single isolated report. If the base line λ_k^0 is 0, even a single such report will cause a signal. If it is desired not to have a signal for an isolated report, λ_k^0 should be adjusted to be slightly larger than 0.
2. In many applications it will be useful or necessary to adjust for special factors affecting the frequency of warranty reports. It is important, for example, to account for known systematic factors like seasonality or trend in warranty reports rates. This can be done by analyzing historical data and using these to fit a model that will provide appropriate values of λ_k^0 .
3. We define the beginning of product life as the date of sale because it is available in the database and because it coincides with the beginning of the warranty period. When date of sale is not in the production database, it is usually in the report database (as customers/repair shops have to present proof that the warranty is still in force). Thus information for R_{ijk} is available. The past patterns in the report rates used to determine the λ_k^0 values will reflect the distribution of time between manufacturing and sales. Thus an assumption of the detection procedure, as described here, is that the distribution of time between manufacturing and sales is approximately stable over time.
4. For detection purposes, we stratify on production period. Running the monitoring procedure separately on different plants, shifts, etc. could provide much more power to detect localized manufacturing problems. On the other hand, to detect new problems arising from a change in product design, higher power for detection will result from pooling data over different manufacturing lines.
5. Field product life distributions are almost always a mixture of different use rates and environments. This mixture can work to aid in early detection. In particular, reports for some failure modes will tend to occur earlier from segments of the population having high use rates. Our detection procedure accounts for this implicitly by using historical report rates to develop the critical values. Thus another assumption of our procedure, as described here, is that the product use-rate distribution is approximately stable over time. If it is not, then adjustment of the λ_k^0 reference values would be needed.
6. Beyond $M > 12$ months or so of observation with a two-dimensional warranty policy, the number of reports will begin to be affected by automobiles that experience “mileage-out” because they have accumulated (say) 36 thousand miles of service. The reduction in report rate resulting from mileage out will, however, be reflected in the historical reporting pattern and thus our monitoring procedure can be used without modification for two-dimensional warranties.
7. Most of the work cited in Section 1.2 is concerned with estimating the failure-time distribution of a manufactured product. The detection scheme presented here, however, is based on the

distribution of warranty reports over time. This simpler inference problem does not require adjustment for mileage-out, effects of the use-rate distribution and so on.

8. The most serious reliability problems tend to arise abruptly in time, generally caused by a change of product design, materials or components, or in a manufacturing process. The detection procedure described here was designed to detect such changes. To increase the power of detecting gradual or smaller changes (which we did observe in some labor codes), various kinds of runs rules can be added. Runs rules could be constructed in a manner that is similar to runs rules that have been associated with the Shewhart process monitoring schemes. The use of such rules would affect false alarm rates and the ARL distribution, but these characteristics would be easy to evaluate using either numerical or simulation techniques.
9. For given specific costs and cost functions, it would be possible to optimize the detection procedure to minimize some particular definition of total cost.

Acknowledgments

We would like to acknowledge some of the background work of our students in their M.S. creative component projects that helped lay the foundation for the results reported in this paper. In particular, Kim Wentzlaf developed a collection of SAS macros to extract and pre-process information from the part of the automobile warranty database that was used in our examples. Ed Staats further developed these macros to allow us to automate a survey of the reliability behavior of all of the labor codes in the database and to take account of date-of-sale information. Ed also wrote programs to explore some earlier forms of decision rules with which we experimented before developing the methods presented in this paper.

A Computing Monitoring Quantities S_{ijk} and C_{ijk}

This appendix outlines the sequence of steps used to compute the test statistics and corresponding critical values used in the warranty monitoring procedure. The inputs include $\alpha, M, \rho, \lambda_1^0, \dots, \lambda_M^0$, and f_1^0, \dots, f_M^0 .

1. Compute $\alpha_k, k = 1, \dots, M$ using the method in Section 4.1.
2. Suppose production starts in period 1. At the end of period 1, n_1 and n_{11} become available. Set $f_{11} = n_{11}/n_1$ and $f_{1k} = f_k^0$ for $k = 2, \dots, M$.
3. At the end of period i ($i \geq 2$), n_i and $n_{m,(i+1-m)}$, $m = i, i-1, \dots, \max(1, i+1-M)$ are available. Set $f_{m,(i+1-m)} = n_{m,(i+1-m)}/n_m$ for $m = i, \dots, \max(1, i+1-M)$ and $f_{ik} = f_k^0$ for $k = 2, \dots, M$. Compute the following for $m = i-1, \dots, \max(1, i-M)$ and $j = 1, \dots, i-m$.
 - (a) Obtain $R_{m,j,(i+1-m-j)}$ from the warranty report data set.
 - (b) Compute

$$S_{m,j,(i+1-m-j)} = \begin{cases} R_{m,1,i-m} & \text{if } j = 1 \\ S_{m,j-1,(i+1-m-j)} + R_{m,j,(i+1-m-j)} & \text{if } j > 1 \end{cases}$$

- (c) Compute $\alpha_{j,(i+1-m-j)}^{(m)} = \alpha_m \times \left(\frac{f_{m1} + \dots + f_{mj}}{f_{m1} + \dots + f_{m,M+m+j-i}} \right)^\rho$.
- (d) Compute $C_{m,j,(i+1-m-j)}$ using the recursive algorithm in Section 5.1.

References

- Blischke, W. R., and Murthy, D. N. P. (1994), *Warranty Cost Analysis*, New York: Marcel Dekker.
- Blischke, W. R., and Murthy, D. N. P. (eds.) (1996), *Product Warranty Handbook*, New York: Marcel Dekker.
- Hu, X. J., Lawless, J. F., and Suzuki, K. (1998), "Nonparametric Estimation of a Lifetime Distribution When Censoring Times Are Missing," *Technometrics*, 40, 3–13.
- Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods With Applications to Clinical Trials*, Boca Raton: Chapman & Hall/CRC.
- Kalbfleisch, J. D., and Lawless, J. F. (1988), "Estimation of Reliability in Field-Performance Studies" (with discussion), *Technometrics*, 30, 365–388.
- Kalbfleisch, J. D., Lawless, J. F., and Robinson, J. A. (1991), "Methods for the Analysis and Prediction of Warranty Claims," *Technometrics*, 33, 273–285.
- Karim, M. R., Yamamoto, W., and Suzuki, K. (2001a), "Statistical Analysis of Marginal Count Failure Data," *Lifetime Data Analysis*, 7, in print.
- Karim, M. R., Yamamoto, W., and Suzuki, K. (2001b), "Change-Point Detection From Marginal Count Failure Data," *Journal of the Japanese Society for Quality Control*, 31, in print.
- Kim, K., and DeMets, D. L. (1987), "Design and Analysis of Group Sequential Tests Based on the Type I Error Spending Rate Function," *Biometrika*, 74, 149–154.
- Lan, K. K. G., and DeMets, D. L. (1983), "Discrete Sequential Boundaries for Clinical Trials," *Biometrika*, 70, 659–663.
- Larsen, R. J., and Marx, M. L. (2001), *An Introduction to Mathematical Statistics and Its Applications*, 3rd ed., Upper Saddle River, NJ: Prentice-Hall.
- Lawless, J. F. (1998), "Statistical Analysis of Product Warranty Data," *International Statistical Review*, 66, 41–60.
- Lawless, J. F., Hu, J., and Cao, J. (1995), "Methods for the Estimation of Failure Distributions and Rates From Automobile Warranty Data," *Lifetime Data Analysis*, 1, 227–240.
- Lawless, J. F., and Kalbfleisch, J. D. (1992), "Some Issues in the Collection and Analysis of Field Reliability Data," in *Survival Analysis: State of the Art*, Editors J. P. Klein and P. K. Goel, Kluwer Academic Publishers, 141–152.

- Meeker, W. Q. (1987) Limited Failure Population Life Tests: Application to Integrated Circuit Reliability. *Technometrics* 29, 151-165.
- Pocock, S. J. (1977), "Group Sequential Methods in the Design and Analysis of Clinical Trials," *Biometrika*, 64, 191–199.
- Robinson, J. A. , and McDonald, G. C. (1991), "Issues Related to Field Reliability and Warranty Data," in *Data Quality Control: Theory and Pragmatics*, Editors Gunar E. Liepins and V. R. R. Uppuluri, Marcel Dekker, Inc, 69–89.
- Schultz, J. R., Nichol, F. R., Elfring, G. L., and Weed, S. D. (1973), "Multiple-Stage Procedures for Drug Screening," *Biometrics*, 29, 293–300.
- Shewhart, W. A. (1931), *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand. (Republished in 1981, with a dedication by W. Edwards Deming, by the American Society for Quality Control, Milwaukee, WI)
- Slud, E., and Wei, L. J. (1982), "Two-Sample Repeated Significance Tests Based on the Modified Wilcoxon Statistic," *Journal of the American Statistical Association*, 77, 862–868.
- Suzuki, K. (1985a), "Estimation of Lifetime Parameters From Incomplete Field Data," *Technometrics*, 27, 263–272.
- Suzuki, K. (1985b), "Nonparametric Estimation of Lifetime Distributions From a Record of Failures and Follow-ups," *Journal of the American Statistical Association*, 80, 68–72.
- Wang, L., and Suzuki, K. (2001a), "Nonparametric Estimation of Lifetime Distribution From Warranty Data Without Monthly Unit Sales Information," *The Journal of Reliability Engineering Association Japan*, 23, 145–154.
- Wang, L., and Suzuki, K. (2001b), "Lifetime Estimation on Warranty Data Without Date-of-Sale Information—Case Where Usage Time Distributions Are Unknown," *Journal of the Japanese Society for Quality Control*, 31, 148–167.