

# Warranty data analysis: a review<sup>1</sup>

Shaomin Wu

*School of Applied Sciences, Cranfield University, Bedfordshire MK43 0AL, UK  
shaomin.wu@cranfield.ac.uk*

---

**Abstract:** Warranty claims and supplementary data contain useful information about product quality and reliability. Analysing such data can therefore be of benefit to manufacturers in identifying early warnings of abnormalities in their products, providing useful information about failure modes to aid design modification, estimating product reliability for deciding on warranty policy, and forecasting future warranty claims needed for preparing fiscal plans.

In the last two decades, considerable research has been conducted in warranty data analysis (WDA) from several different perspectives. This paper attempts to summarise and review the research and developments in WDA with emphasis on models, methodologies and applications. It concludes with a brief discussion on current practices and possible future trends in WDA.

**Keywords:** early warning, field reliability estimation, warranty data analysis (WDA), two-dimensional warranty, warranty claim prediction, design modification.

---

## 1. Introduction

A warranty is a contractual obligation incurred by a manufacturer (vendor or seller) in connection with the sale of a product. In broad terms, the purpose of warranty is to establish liability in the event of a premature failure of an item or the inability of the item to perform its intended function<sup>1</sup>. Product warranty is becoming increasingly more important in consumer and commercial transactions, and is widely used to serve many different purposes<sup>2-5</sup>.

A vast literature on warranty can be found, for example, in 1996, Djameludin et al<sup>6</sup> listed over 1500 papers in this area. Recently, research in warranty has attracted even more attention, as can be seen from the review papers<sup>5, 7-11</sup> and the books<sup>12-17</sup>.

Warranty data is comprised of claims data and supplementary data. Claims data are the data collected during the servicing of claims under warranty and supplementary data are additional data (such production and marketing related, items with no claims, etc.) that are needed for effective warranty management. Warranty data provide valuable information to indicate product quality and

- 
- <sup>1</sup> Suggested citation: Wu, S. (2012). Warranty data analysis: a review. *Quality and Reliability Engineering International*, 28(8), 795-805.
  - A review paper on the quality of warranty data: Wu, S. (2013). A review on coarse warranty data and analysis. *Reliability Engineering & System Safety*, 114, pp 1-11, is also recommended.
  - Author's current email: s.m.wu@kent.ac.uk

field reliability. Beginning with Suzuki<sup>18, 19</sup>, extensive research on warranty data analysis (WDA) has been done, see<sup>9, 20-27</sup>, for example. Karim et al<sup>9</sup>, published in 2005, is an excellent review paper on warranty claim data analysis. Other review articles on WDA prior to 2005 are references<sup>20, 22, 23, 28-30</sup>. Over the last five years, more papers have been published and a comprehensive review paper is needed to summarise the state-of-the-art developments in WDA.

The aim of WDA is to extract useful information and help in decision making by analysing warranty data with either statistical or computer algorithms (for example, neural network models). Warranty data can be used in many other ways by a manufacturer and include the following:

- To detect early warning of faulty designs, flawed production lines, defective parts, etc,
- To provide useful information for product modification and improvement,
- To estimate and explain the costs of warranty claims,
- To predict future claims and warranty cost, and
- To estimate product reliability for deciding on warranty policy and appropriate maintenance policy.

Relating to the five areas discussed above, the main objectives of this paper are (i) to review the existing research in WDA and (ii) to suggest new directions for future research based on the trends and issues identified.

The rest of this paper is structured as follows. Section 2 discusses the causes of warranty claims and the characteristics of warranty data. Section 3 looks at WDA, the different kinds of models and methodologies that have been proposed and studied in the literature. Section 4 explains two tables summarising the papers reviewed. Section 5 concludes the paper with a brief discussion on current practices and possible future trends in WDA.

## 2. Warranty claims

A typical lifecycle of failed products is shown in Figure 1, where the process starts from product manufacture time and ends at the time when they are returned to the manufacturer.

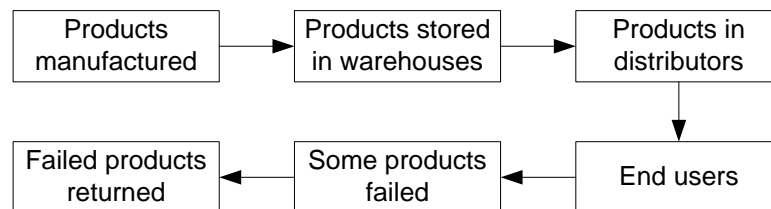


Figure 1. A typical lifecycle of failed products

We can express warranty claim data as shown in Table 1, where  $r_{x,y}$  represents the number of claims received on the  $(x + y - 1)$ -th day and the claimed items were sold on day  $d_x$  and manufactured on day  $D_x$ .

Table 1: Part of warranty data

Manufacture date	Manufacture volume	Sales date	Sales volume	Claims received (month in service)						
				1	2	...	$m_0$	...	$n_0 - 1$	$n_0$
$D_1$	$N_1$	$d_1$	$M_1$	$r_{11}$	$r_{12}$	...	$r_{1,m_0}$	...	$r_{1,n_0-1}$	$r_{1,n_0}$
$D_2$	$N_2$	$d_2$	$M_2$		$r_{21}$	...	$r_{2,m_0-1}$	...	$r_{2,n_0-2}$	$r_{2,n_0-1}$
...	...	...	...			...	...	...	...	...
$D_{m_0-1}$	$N_{m_0-1}$	$d_{m_0-1}$	$M_{m_0-1}$			...	$r_{m_0-1,2}$	...	$r_{m_0-1,n_0-m_0+1}$	$r_{m_0,n_0-m_0+2}$
$D_{m_0}$	$N_{m_0}$	$d_{m_0}$	$M_{m_0}$				$r_{m_0,1}$	...	$r_{m_0,n_0-m_0}$	$r_{m_0,n_0-m_0+1}$
Total	$M$		$M$	$r_1$	$r_2$	...	$r_{m_0}$	...	$r_{n_0-1}$	$r_{n_0}$

It should be noted that the quality of warranty data is usually not perfect, as they might be

- aggregated data: warranty data might only be available in the form of aggregated claims. That is, warranty data might be aggregated into groups. For example, a warranty claim analyst might be only given the total number of claims for items in age 0-30 days, 31-60 days, etc.
- delayed data: warranty data can include sales delay and reporting delay. For example, reporting delay might be caused by manufacturers who might need time to verify the claims before the claims are entered into the database.
- incomplete censored data: warranty data are commonly right censored data, which is caused by the fact that warranty can expire.

For more detailed discussion on the quality of warranty data and its analysis, the reader is referred to Wu<sup>31</sup>.

Warranty claims might be caused by various forms of failures. For example, Figure 2 shows possible causes of warranty claims, which can be roughly categorised into four types of failures: hardware failures, software failures, human errors and organisational errors.

It might be noted that most of publications in the reliability literature simply assume that warranty claims are due to hardware and/or software failures. Such an assumption might not necessarily hold. For example, an end-user might claim warranty although the item has not failed, or an end-user might not claim warranty although the item has already failed, see Wu<sup>32</sup>, for example.

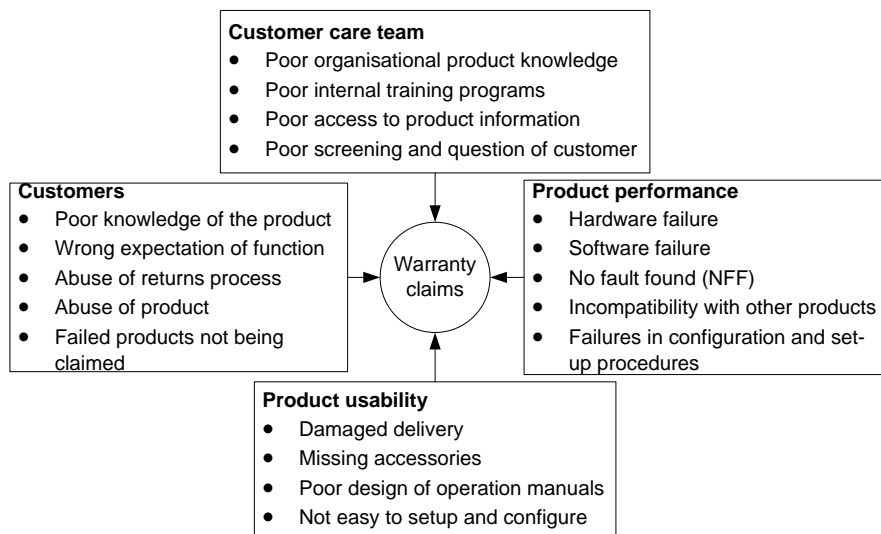


Figure 2. Examples of causes of warranty claims

### 3. Warranty data analysis

Warranty data analysis can broadly be categorised into five areas, as shown in Figure 3. In this section, we review existing research in these five areas.

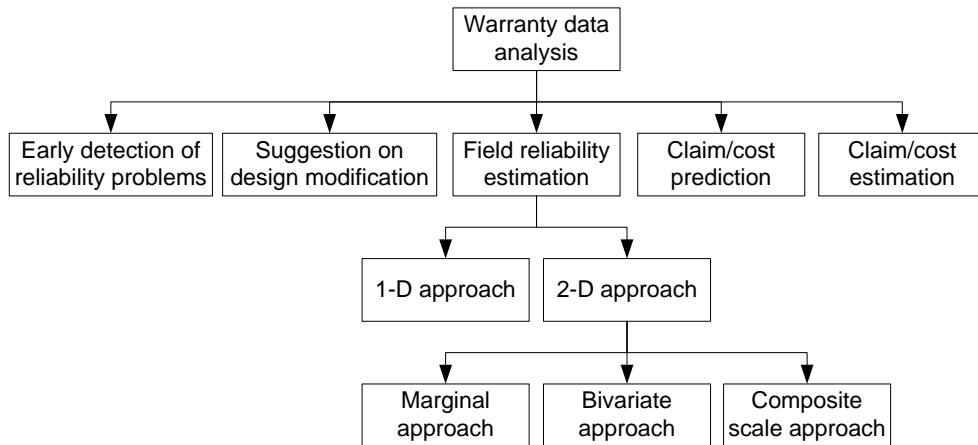


Figure 3: Analysis of warranty data.

#### 3.1 Early detection of reliability problems

The intention of early detection of reliability problems is to provide manufacturers with the opportunity to discover early indications of unexpected quality and reliability problems through WDA. In essence, this intention can be achieved through detecting abnormal change points in warranty data by using a variety of statistical techniques such as control charts or comparing probability distributions to the benchmarking distribution, or artificial intelligent techniques.

When developing an early warning algorithm, we should bear in mind that there are following four points deserving attention.

- New products. New products might have limited warranty claim data. Detecting abnormalities on the basis of such small sample size can be troublesome.
- Non-stationary process. Products that have already been launched might be modified from time to time. It is often necessary to assume that the warranty claim process is non-stationary.
- Common causes. The failure of a type of component that is installed in many different types of products can cause more serious problems than those only installed in one type of product. For example, a type of AC/DC adaptor might be used in many different types of laptop computers. If a large number of claims are due to the failure of the AC/DC adaptor, signalling a warning on this component can be important.
- Text mining. Warranty claims are usually reported by end-users and the failure modes are expressed in text documents. Developing algorithms to analyse failure modes from the documents is essential. Text mining from computer science can be a useful technique.

In the literature, a number of techniques have been developed to detect product abnormalities from warranty data. Karim et al<sup>33</sup> propose a method to detect change points from marginal count

warranty claims data through modelling and comparing the mean number of failures at the stages of the pre- and post- design change point. Wu and Meeker<sup>34</sup> use statistical detection rules to provide an early indication of reliability changes with the Poisson distribution estimation. Grabert et al<sup>35</sup> develop an early detection system using neural networks and probability distribution estimation. Honari and Donovan<sup>36</sup> use control charts to monitor any changes and validate their approach based on both artificially generated data and warranty claims data.

Vittal and Neuman<sup>37</sup> surveys the emerging science of early detection for the warranty and reliability issues. They summarise that three measures should be required to assess the efficacy of any “early warning” detection system/algorithm, (1) probability of detection of a change, (2) probability of false alarm, and (3) the alarm time (“time to detect” a change). However, in the literature, not all of authors use these three measures to assess the efficacy of the techniques/algorithms.

Techniques used in early detection analysis have a long application history in many other areas such as machinery health monitoring. Comparing to machinery health monitoring, early detection of reliability problems using warranty data can be more difficult. In machinery health monitoring, an item being monitored is usually not modified. In WDA, however, products might be continually modified; consequently, warranty claims can be due to a series of changing failure modes of the products.

### **3.2 Suggestion on design modification**

Warranty data can also be utilised to assist engineers in improving their product design. This can be done if failure causes can be analysed and further detected with warranty data.

It should be noted that both early detection analysis and design modification aim to detect abnormalities from warranty data. However, in early detection analysis, timeliness is an important factor: it focuses on techniques that can detect abnormalities at the earliest opportunity. As such, on-line monitoring techniques such as control charts might be pursued. Design modification, however, emphasises on using warranty databases to aid engineers to change their system design and aims to improve the reliability and quality of their products. Hence, timeliness might be less important.

Majeske et al<sup>38</sup> and Majeske and Herrin<sup>39</sup> use graphical tools to compare warranty claims to their benchmark data, and suggest that manufacturers should analyze design changes using post hoc tests on warranty data. Yang and Cekecek<sup>40</sup> develop a method of utilizing warranty data to prioritize design improvement efforts based on design vulnerability and the feedback from warranty claims data. They use cost as a measure of design vulnerability to indicate the weaknesses of the design and provides feedback for design improvements, and then use linear programming to prioritise those areas that are crucial for improvement, subject to limited budget.

Modern computing techniques such as data mining and text mining have also been introduced to extract meaningful knowledge from databases containing warranty claims. Extracting useful rules such as IF-THEN rules can assist engineers in their design analysis (for example, failure analysis).

Buddhakulsomsiri et al<sup>41,42</sup> uses the elementary set concept and database manipulation techniques to search patterns and relationships among occurrences of warranty claims and create IF-THEN rules, where the IF portion includes a set of attributes representing product features (e.g. production date, repair date, mileage-at-repair, transmission, engine type, etc.) and the THEN portion includes a set of attributes representing decision outcome (e.g. problem related labour code). These rules are used to identify root causes of a particular warranty problem or to develop meaningful conclusions.

Some manufacturers might have a warranty database and a customer survey database. Commonly, the warranty database is maintained by engineering departments, whereas the customer survey database is maintained by customer relationship departments. Linking the two databases and then analysing them can increase the understanding of both reliability/quality problems and customer expectation, which might result in a modification that can satisfy customer needs and improve product reliability. Sureka and Varma<sup>43</sup> develop a rule-based system for extracting named entities from customer complaint, technician comments and action taken field of the warranty claim forms.

### **3.3 Field reliability estimation**

Estimating the reliability of products from warranty data, or field reliability estimation, is important for manufacturers as it can help in various aspects such as selecting warranty policy, planning maintenance regimes and preparing spare parts. As warranty data reflect real operating environment and usage rate, they are more informative than testing data collected from laboratories. As such, estimating product reliability based on warranty data can provide manufacturers with more important information.

When estimating product reliability from warranty claims data, however, we need to notice<sup>22</sup> that:

- Warranty claims data are usually incomplete. Such incompleteness might result in biased inference.
- Warranty claims data are only collected from the early life of products and might provide little direct information about longer term reliability or durability.

We also need to notice that warranty claims can also contain claims due to human factors<sup>32</sup>.

Warranty policies can be categorised into one- and two-dimensional. A one-dimensional (1-D) policy is characterised by an interval (age only or usage only) as warranty limit. A two-dimensional (2-D) policy is represented by a region in the two-dimensional plane: generally one dimension representing age and the other representing usage. For different types of products, usage can be different, for example, output-based (miles for cars, copies made for photocopier, etc.), time-based (fraction of the time used – air-conditioners, heaters, etc.), stress level (used continuously but different stress levels – air conditioners on hot or very hot days).

In the literature, approaches developed to analysing warranty data collected from the products with 1-D or 2-D policies can be different. In the following, we review publications in two aspects: 1-D and 2-D analysis for field reliability estimation.

### **3.3.1 One-dimensional (1-D) approach**

It should be noted that 1-D warranty might also be usage based, for example, the warranty limit for a copy machine can be the number of copies that it has made.

#### **3.3.1.1 Age-based analysis**

In the literature, age-based field reliability estimation has not been the main focuses. This might be due to the fact that techniques on estimating field reliability are well established, given complete age information.

Approaches to estimating the lifetime distribution include estimating mixed distributions<sup>44</sup>, fitting the Weibull distribution based on a small number of warranty claims<sup>45</sup>, estimating the lifetime distribution considering sales delay<sup>46</sup>, and estimating the intensity of a NHPP (Non-Homogenous Poisson Process) for repairable items<sup>47</sup>.

#### **3.3.1.2 Usage-based analysis**

Effective estimation of usage-based lifetime distributions for warranty claims requires complete information of the usage intensity of all items, including censored and claimed. However, the usage intensity distributions of the items failing within the warranty limit might be different from those of products surviving the warranty. This causes a problem of obtaining censoring times (or usage intensity) for those products that have not reported to the product manufacturer.

For engineering purposes, usage time is more relevant; and hence, modeling usage accumulation is an integral part of reliability analysis. However, for a usage-based analysis, such as of mileage or copy number, it is difficult to estimate the lifetime distribution without having censored data (e.g., mileage of non-failure automobiles) because the usage time distributions of non-failed products are different from those of failed products.

Starting from Suzuki et al<sup>18,19</sup>, there are a series of research on estimating product reliability when incomplete censored data, ie., incomplete usage data, are presented. Approaches to dealing with the case of the incomplete usage data have been studied by references<sup>48-51</sup>.

One of often used approaches to dealing with incomplete censored data is the supplementary data approach. This approach randomly selects a follow-up sample of products from the unfailed products under warranty, obtains their censoring times, usage history and/or any covariate values, and then uses a pseudo-likelihood approach to estimating the parameters of survivor distributions. Such a follow-up research can be follow-up studies, customer surveys, postal reply cards, and periodic inspections. The pseudo-likelihood approaches can be parametric or non-parametric based<sup>18,19</sup>. The

pseudo-likelihood approach can also be extended to analyse claims data with covariate information<sup>48, 52</sup>.

When additional field data are available, Oh and Bai<sup>53</sup> proposes methods of estimating the lifetime distribution. Attardi, et al<sup>54</sup> use a mixed-Weibull regression model to estimate the failure time of components of the gear-box mounted on some FIAT automobiles.

When the usage time of censored items cannot be obtained, Suzuki<sup>55</sup> proposed to use non-homogeneous Poisson processes. Suzuki et al<sup>56</sup> present two methods, parametric and semi-parametric, to estimate product field reliability, where non-failure data is not included.

The usage intensity of products by specific groups of consumers are surveyed and analysed in reference<sup>57</sup>.

However, it has been noted that little attention in the above research has been paid to the causes of the warranty claims. Warranty claims of a product can be due to many different failure modes. If one is concerned with one of the failure modes, s/he will find that many other failed items may be claimed due to the other failure modes. S/He will then have these censored and uncensored observations to estimate survivor distributions for the different failure modes. This phenomenon has been observed from the warranty database of an automobile manufacturer in the UK.

### **3.3.1.3 Co-variate analysis**

Claim frequencies may be relating to many other factors. The following three aspects of data can be important and collectable in field-performance studies<sup>48</sup>:

- Information on types and frequencies of problems (e.g., failures, replacements, etc.) and on time patterns of problems (e.g., times to failure, performance degradation over time, life of the product, etc.).
- Manufacturing characteristics of items in use (e.g., model, place or time of manufacture, etc.).
- Usage intensity and operating conditions (e.g., personal characteristics of users, climatic conditions, etc.).

If only a few separate conditions are of interest, the simplest approach is to estimate the expected claims for each condition separately. More generally, covariates can be used to represent various factors and build regression models<sup>22</sup>. Covariates may be introduced at an individual unit level or at an aggregate claims level. In either case log linear models are convenient<sup>22</sup>.

Kalbfleisch and Lawless<sup>48</sup> suggest procedures for the collection of field data and then use a regression model to estimate lifetime distributions from field failure data with supplementary information about covariates. Hu and Lawless<sup>52</sup> develop an estimation procedure with supplementary information about covariates and censoring times, suggest a technique for modelling warranty claims as truncated data, and assume warranty claims follow a Poisson process. Hu and Lawless<sup>50</sup> consider situations involving a response variable and covariates, both of which are incomplete. They discuss two types of pseudo-likelihood estimation and provide methods for non-parametric estimation of



lifetime distributions. Attardi et al<sup>54</sup> use a mixed-Weibull regression model for the analysis of automotive warranty claims data, assuming that the products are a mixture of weak and strong sub-populations with respect to their reliabilities. The engine type and car model are used as covariates in their regression model. Karim and Suzuki<sup>58</sup> consider covariates associated with some reliability-related factors and present a Weibull regression model for the lifetime of the component as a function of such covariates. The expectation maximization (EM) algorithm is applied to obtain the maximum likelihood estimates of the parameters of the model.

Regions, in which the products are operated, are considered as impact factors by references<sup>59, 60</sup>.

### 3.3.2 Two-dimensional (2-D) Approach

Much of the literature on warranty analysis considers failure models which are indexed by a single variable, such as either age or mileage. There are situations where several characteristics are used together as criteria for judging the warranty eligibility of a failed product. For example, for automobiles, warranty coverage has sometimes both age and mileage limits, and it is often important to develop methods based on both age and usage amounts.

For items under 2-D warranty policy, Figure 4 shows the possible ages and usage rates of four items covered a 2-D warranty policy. Item 1 failed within both the age limit and the usage limit and it might be reported to the manufacturer. Item 2 failed within the age limit but beyond the usage limit and its warranty expired. Item 3 failed within the usage limit but beyond the age limit, and its warranty expired. Item 4 has both the age and usage at failure above the age limit and the usage limit.

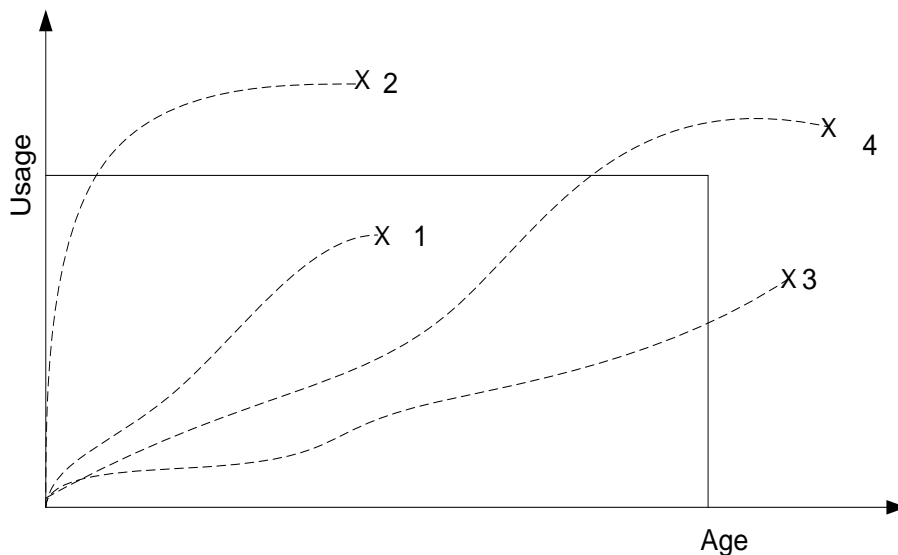


Figure 4. Two-dimensional warranty.

In analysing 2-D warranty data, for example, if we want to estimate the distribution, we need to collect data on both the age and usage of products. A complicating issue is that the age and usage might be unknown to the manufacturer for some items, for example, the age and usage of items 2, 3,

and 4 in Figure 4 might not be available. This poses a challenge that only data of those failed and reported items are available, but the usage for those items whose warranty has expired cannot be obtained.

To analyze 2-D warranty data with unknown censoring times, three approaches have been proposed in the literature: marginal approach, bivariate approach, and composite scale approach. The marginal approach indirectly fits a joint distribution, the bivariate approach directly estimates a bivariate distribution, and the composite scale approach reduces the two-dimensional warranty problem to a one-dimensional formulation.

### **3.3.2.1 Marginal approach**

The marginal approach assumes that the usage intensity for a customer is constant over the warranty period, but varies across the customer population. As a result, the usage rate is a random variable. It can be modelled either as a discrete variable (for example, low, medium and high users) or as a continuous variable with a density function<sup>61</sup>. A joint distribution can then be calculated from the marginal distributions. With this approach, Lawless et al<sup>21</sup> considers the occurrence of warranty claims for automobiles when both age and mileage accumulation affect failure. They discussed models to assess the dependence of failures on age and mileage, and to estimate survival distributions and rates from warranty claims data using supplemental information about mileage accumulation. Davis<sup>62</sup> discusses a similar approach. Singpurwalla and Wilson<sup>63</sup> propose an approach for developing probabilistic models in a reliability setting indexed by two variables, time and a time-dependent quantity such as the amount of use. They used these variables in an additive hazard model. Chukova and Robinson<sup>64</sup> take age and mileage as the usage measure, respectively, and evaluate the mean cumulative number of claims or cost of claims and its standard error as functions of the usage measure, with both parametric and non-parametric approaches. It presents a model where the usage time is a primary variable and the mileage accumulation is estimated from field return data. At each time interval the probability of exceeding the upper mileage limit on the warranty is calculated based on the mileage distribution and an analytical cumulative distribution function of exceeding the upper mileage limit at any point time is constructed. This modeling procedure accounts for an observed reduction in the number of warranty claims in the second half of the warranty period.

### **3.3.2.2 Bivariate approach**

The bivariate approach is to directly estimate a joint bivariate distribution from warranty data. Singpurwalla and Wilson<sup>65</sup> develop a bivariate failure model for automobile warranty claims data indexed by time and mileage with a method to estimate the density function of failure using a log-log model. Moskowitz and Chun<sup>66</sup> assume that the number of failures under the two-dimensional warranty policies is distributed as a Poisson process with parameters that can be expressed by a regression function of the age and usage amounts of a product. Yang and Nachlas<sup>67</sup> develop bivariate

renewal models and also explain their inclusion in availability models. Pal and Murthy<sup>68</sup> use Gumbel's bivariate exponential distribution to fit warranty claims. Jung and Bai<sup>69</sup> consider a bivariate approach and assume that age and usage are statistically correlated in a bivariate distribution. Lawless and Crowder<sup>25</sup> present models to assess the dependence on age or usage in heterogeneous populations of products, and show how to estimate model parameters based on different types of field data. Lawless and Crowder<sup>27</sup> provide joint models for the recurrent events and usage processes, which facilitate analysis of their relationship as well as prediction of failures.

### 3.3.2.3 Composite scale approach

In addition to the above two approaches, Gertsbakh and Kordonsky<sup>70</sup>, and Duchesne and Lawless<sup>71</sup> proposed methods of making an alternative composite scale from age and usage, which integrates the two scales (age and usage) to create a single composite scale and failures are modelled as a counting process using this composite scale. For example, in Gertsbakh and Kordonsky<sup>70</sup>, a new variable  $V = \varepsilon T + (1 - \varepsilon)U$  is introduced, where  $\varepsilon \in (0, 1)$ , the time scale of the variable  $V$  is a linear combination of the age scale  $T$  and the usage scale  $U$ , and it does not have a physical meaning. Ahn et al<sup>72</sup> present the power law process with the new time scale as a model for the reliability of a repairable system. This approach is used by Iskandar and Blischke<sup>73</sup> to model the warranty claims from a motorcycle manufacturer.

### 3.3.3 Some comments

It should be noted that the warranty policies for the same type of products can be different from region to region because the legislations can be different. For example, automobiles with the same make may have 1-D warranty, 5 year warranty, say, in one country, but have 2-D warranty, 4 year or 48,000km, say, in other countries. An interesting question is to sufficiently use the warranty data collected from both the countries to estimate survivor distributions.

Data (for example, covariate data) collected from both the warranty database and the customer survey database can also be used to develop field reliability models.

## 3.4 Warranty claim prediction

Warranty claim prediction in general terms is to predict the expected number of claims and/or the respective warranty cost at the warranty coverage. Predicting warranty claims can critically important for the finance departments of a company in preparing their fiscal plans. It can be found that the following techniques have been developed to predict warranty claims.

*Lifetime distributions.* This approach is to estimate a time-to-claim distribution. In the literature, Kleyner and Sandborn<sup>74</sup> present a warranty claim forecasting model based on a piecewise application of Weibull and exponential distributions, which tries to capture the dynamic characteristic features of failure rates in both early failure period and the intrinsic failure period of the bathtub curve. Rai<sup>26</sup>

presents a forecasting model incorporating calendar month seasonality, business days per month for authorised service centres, and sales ramp-up in addition to the earlier mentioned variables.

When fitting a lifetime probability distribution, both the exact number of claims and the exact number of unclaimed products should be known. Those two numbers, however, might not be available in the reality, due to the incompleteness.

*Stochastic processes.* The Poisson process has been the workhorse used in predicting warranty claims. The mean number of warranty claims is assumed to be the parameter of the process. Kalbfleisch et al<sup>20</sup> used a log-linear Poisson model to analyse and forecast warranty claims. In their work, they modelled warranty claims based on the date of warranty claim rather than the failure date, and therefore the reporting lag between occurrence of a claim and its entry to a database was taken into consideration. Dynamic linear models with leading indicators are also used in<sup>65</sup>. Kaminskiy and Krivtsov<sup>75</sup> develop warranty claim forecasting models with the G-renewal process—generalized renewal processes introduced by Kijima and Sumita<sup>76</sup>, the ordinary renewal process (ORP) and the non-homogeneous Poisson process (NHPP). They found that GRP provides a higher accuracy compared to the ORP or the NHPP. Majeske<sup>77</sup> present a NHPP-based technique that forecasts the total number of claims and the timing of claims during the vehicle lifetime. Fredette and Lawless<sup>78</sup> present forecasting methods for warranty claims, using mixed NHPP, and possible heterogeneity among the individuals is modelled using random effects.

The stochastic process approaches might require assumptions such as the claim rates following a specific law (for example, NHPP). Such an assumption might be violated, as the quality of maintenance might not be difficult to assess<sup>79</sup>.

*Artificial neural networks.* Non-parametric approaches such as neural networks have also been applied to predict warranty claims. Starting from Wasserman<sup>80</sup> and Wasserman and Sudjianto<sup>81</sup>, MLP (multi-layer perceptron)<sup>80,81</sup> and RBF (radial basis function)<sup>82</sup> have been used.

*The Kalman filter and time series models.* If we simply use the claim rates  $r_{x+y-1}$  to represent warranty claims  $r_{x,y}$  shown in Table 1, the claim rates of each month can then be seen as time series. Techniques used in time series prediction can then be borrowed. Singpurwalla and Wilson<sup>65</sup> consider using the Kalman filter to build forecasting models. Wasserman<sup>83</sup> develop linear regression models, first-order auto-regression time series models, and also the Kalman filter models to forecast warranty claims. In the linear regression models, the number of months in service is used to forecast the number of repairs per 1000 items. Wasserman and Sudjianto<sup>81</sup> further compare linear regression models, time series models, the Kalman filter, the orthogonal series, and the MLP from artificial neural networks in modelling and forecasting warranty claims, and conclude that the Kalman filter model offers a significant improvement over simple linear regression approach, but both the orthogonal series and the neural network models outperform the Kalman filter. In the same year, Chen et al<sup>84</sup> propose to model and forecast the number of warranty claims with the Kalman filter.

It should be noted that those approaches developed on the basis of repair rates (or claims rates) may cause information loss, as they are obtained as a ratio of warranty claims to the number of products in service (i.e., they integrate two observations into one).

Nevertheless, one might find two weaknesses existing in the approaches mentioned above: (1) they do not consider the fact that warranty claims reported in the recent months might be more important in forecasting future warranty claims than those reported in the earlier months, and (2) they are developed based on repair rates (i.e, the total number of claims divided by the total number of products in service), which can cause information loss through such an arithmetic-mean operation. To overcome these two weaknesses, Wu and Akbarov<sup>31</sup> introduce two different approaches to forecasting warranty claims: the first is a weighted support vector regression (SVR) model and the second is a weighted SVR-based time series model. These two approaches can be applied to two scenarios: when only claim rate data are available and when original claim data are available. Two case studies are conducted to validate the two modelling approaches. On the basis of model evaluation over six months ahead forecasting, the results show that the proposed models exhibit superior performance compared to that of MLP and RBF neural networks and ordinary support vector regression models.

### **3.5 Warranty claim estimation**

Estimating the number of warranty claims is another interesting topic. In this area, the Poisson process has been the workhorse.

Kalbfleisch, et al<sup>20</sup> used a log-linear Poisson model to estimate warranty claims, where the Poisson parameter is a function of time in service. Lawless and Kalbfleisch<sup>49</sup> introduce a moment estimator of the expected number of warranty claims for a single product.

The non-homogeneous Poisson process is one of most widely used techniques in estimating the number of warranty claims. Lawless and Nadeau<sup>85</sup> develop point estimates based on Poisson models for the cumulative mean function of warranty claims. Hu and Lawless<sup>86</sup> present a non-parametric approach to estimating the rate and mean function from truncated recurrent event data, assuming the observation period for products under warranty is unknown until it experiences at least one claim. For aggregated warranty claims data, Lawless<sup>22</sup> estimates the expected number of claims and also gives the variance of the estimate.

When warranty claims are aggregated, Suzuki et al<sup>23</sup> and Karim et al<sup>87</sup> present a non-homogeneous Poisson process (NHPP) model for repairable products, a multinomial model, its Poisson approximation for nonrepairable products, and then use the EM (Expectation-maximization) algorithm to estimate the expected number of claims. Wang et al<sup>88</sup> and Suzuki et al<sup>23</sup> use non-homogeneous Poisson processes to model warranty claims for repairable products.

Methods developed for estimating warranty claims have been surveyed by Kalbfleisch et al<sup>20</sup>, Robinson and McDonald<sup>28</sup>, and Kalbfleisch and Lawless<sup>30</sup>.

It might be confusing with the difference between warranty claim estimation and prediction. The difference of these two issues lies in<sup>30</sup>: warranty claim estimation is for a hypothetical infinite population of items, of which those sold are considered a random sample, whereas in warranty claim prediction, the population of items that is eventually sold is finite.

## 4. Conclusions and further research

This paper reviewed the existing work in warranty data analysis. Historically, warranty data analysis has been mainly focused on looking for new methods to estimate product field reliability and to estimate warranty claims from warranty data with poor quality. However, little attention has been paid to develop early warning algorithms and to propose suggestions on design modification, while these two areas are extremely important for manufacturers. Unlike warranty prediction or warranty estimation that only relate to the finance aspects of a manufacturer, an effective early warranty system and effective design modification can mitigate risk on human injuries and significant property losses. The well-known crisis of Toyota recall in 2010 might remind us of the importance of the development of early warning algorithms.

The techniques developed in one of the three areas: warranty estimation, warranty prediction and field reliability estimation, might be used in the other two areas. For example, techniques used in field reliability estimation can also be applied in warranty prediction or warranty estimation.

There is a need for future research to address the following research questions.

- The increasing miniaturization of radio frequency (RF) devices and microelectro-mechanical systems (MEMS), as well as the advances in wireless technologies, makes it increasingly easier to collect data of product performance. Researchers will need to process more data, most of which are streamline data. This requires developing more advanced analytical techniques, especially covariate analysis, for WDA.
- *Early detection of reliability problems and design modification.* Early detection of reliability problems requires efficient algorithms, which needs more research. The most important challenges might be to detect: (1) critical failure modes that can result in human injuries and significant property losses; and (2) common failure modes that can cause failures of many different products.
- *Field reliability estimation.* Estimating the reliability of newly launched products can be beneficial to manufacturers and/or warranty suppliers in their fiscal planning. For those products, few warranty data can be collected. Estimating the reliability of such products can be extremely difficult.
- *Claim prediction.* Existing research on WDA has been concentrated on short term warranty. Among various warranties, long term warranty is becoming increasingly more important, due to its application to longer-life assets and enhanced customer demand on service from a product

instead of procurement of new products, as discussed in a review paper<sup>11</sup>. As such, in recent years, some manufacturers such as electronics manufacturers have started contracting long term warranties. Apparently, offering long term warranty results in additional complexities. Thus, new problems arise for long term warranty. However, analysing claims data of long-term warranty has received little attention. For example, the research on warranty prediction reviewed above has been only concentrated on short-term, such as 6-month ahead prediction. However, medium-term and long-term prediction of warranty claims can be more important for manufacturers in fiscal planning and should be studied in the future.

## Remarks

This review does not cover the topic of estimation of warranty cost. The reader is referred to<sup>7,22</sup> for information in this area.

This review has tried to be reasonably complete. However, those papers that are not included were either considered not to bear directly on the topic of the review or inadvertently overlooked. My apologies are extended to both the researchers and readers if any relevant papers have been omitted.

## For further readings

The following works are also of interest to the reader: <sup>89,90,91,92,93,94,95,96,97,98,99</sup>.

## Acknowledgement

This research is supported by the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (EPSRC Grant reference: EP/G039674/1).

We are grateful to the anonymous referees for their useful comments. Valuable comments from Professor D.N.P. Murthy are gratefully acknowledged.

## References

1. Blischke WR, Murthy DNP. Product warranty management - I: A taxonomy for warranty policies. *European Journal of Operational Research* 1992; **62**: 127-148.
2. Wu S, Li H. Warranty cost analysis for products with a dormant state. *European Journal of Operational Research* 2007; **182**: 1285-1293.
3. Wu S, Xie M. Warranty cost analysis for nonrepairable services products. *International Journal of Systems Science* 2008; **39**: 279-288.
4. Wu S, Longhurst P. Optimising age-replacement and extended non-renewing warranty policies in lifecycle costing. *International Journal of Production Economics* 2011; **130**: 262-267.

5. Murthy DNP, Djameludin I. New product warranty: A literature review. *International Journal of Production Economics* 2002; **79**: 231-260.
6. Djameludin D, Murthy D, Blischke W. Bibliography on warranties. In *Product Warranty Handbook*. Marcel Dekker: New York, 1996; 839-916.
7. Thomas MU, Rao SS. Warranty economic decision models: A summary and some suggested directions for future research. *Operations Research* 1999; **47**: 807-820.
8. Murthy DNP, Solem O, Roren T. Product warranty logistics: Issues and challenges. *European Journal of Operational Research* 2004; **156**: 110-126.
9. Karim MR, Suzuki K. Analysis of warranty claim data: A literature review. *International Journal of Quality and Reliability Management* 2005; **22**: 667-686.
10. Murthy DNP. Product warranty and reliability. *Annals of Operations Research* 2006; **143**: 133-146.
11. Rahman A, Chattopadhyay G. Review of long-term warranty policies. *Asia-Pacific Journal of Operations Research* 2006; **23**: 453-472.
12. Blischke WR, Murthy DNP. Warranty cost analysis. CRC Press, 1994.
13. Brennan JR. Warranties, Planning, Analysis and Implementation. McGraw-Hill Professional, 1994.
14. Sahin I, Polatoglu H. Quality, Warranty and Preventive Maintenance. Kluwer Academic Pub: 1998.
15. Murthy DNP, Blischke WR. Warranty Management and Product Manufacture. Springer, 2005.
16. Thomas MU. Reliability and warranties: methods for product development and quality improvement. Taylor & Francis Group, 2006.
17. Rai B, Singh N. Reliability analysis and prediction with warranty data: Issues, Strategies, and Methods. Taylor & Francis Group: London, 2009.
18. Suzuki K. Estimation of lifetime parameters from incomplete field data. *Technometrics* 1985; **27**: 263-271.
19. Suzuki K. Nonparametric Estimation of Lifetime Distributions from a Record of Failures and Follow-ups. *Journal of the American Statistical Association* 1985; **80**: 68-72.
20. Kalbfleisch JD, Lawless JF, Robinson JA. Methods for the analysis and prediction of warranty claims. *Technometrics* 1991; **33**: 273-285.
21. Lawless JF, Hu XJ, Cao J. Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis* 1995; **1**: 227-240.
22. Lawless JF. Statistical analysis of product warranty data. *International Statistical Review* 1998; **66**: 41-60.
23. Suzuki K, Karim MR, Wang L. Ch. 21. Statistical analysis of reliability warranty data. *Handbook of Statistics* 2001; **20**: 585-609.



24. Alam MM, Suzuki K. Lifetime Estimation Using Only Failure Information From Warranty Database. *IEEE Transactions on Reliability* 2009; **58**: 573-582
25. Lawless JF, Crowder MJ, Lee K-. Analysis of Reliability and Warranty Claims in Products With Age and Usage Scales. *Technometrics* 2009; **51**: 14-24.
26. Rai B. Warranty Spend forecasting for subsystem failures influenced by calendar month seasonality. *IEEE Transactions on Reliability* 2009; **58**: 649-657.
27. Lawless JF, Crowder MJ. Models and estimation for systems with recurrent events and usage processes. *Lifetime Data Analysis* 2010; **16**: 1-24.
28. Robinson JA, McDonald GC. Issues related to field reliability and warranty data. In *Data Quality Control: Theory and Pragmatics*, Liepins GE and Uppuluri VRR(eds). Marcel Dekker: New York, 1991.
29. Kalbfleisch JD, Lawless JF. Some useful statistical methods for truncated data. *Journal of Quality Technology* 1992; **24**: 145-152.
30. Kalbfleisch JD, Lawless JF. Statistical analysis of warranty claims data. In *Product Warranty Handbook*, Blischke WR Murthy DNP(eds). Marcel Dekker: New York, 1996.
31. Wu S. A review on coarse warranty data and analysis. submitted;
32. Wu S. Warranty claim analysis considering human factors. *Reliability Engineering and System Safety* 2011; **96**: 131-138.
33. Karim MR, Yamamoto W, Suzuki K. Change-point detection from marginal count failure data. *Journal of the Japanese Society for Quality Control* 2001; **31**: 318-338.
34. Wu H, Meeker WQ. Early detection of reliability problems using information from warranty databases. *Technometrics* 2002; **44**: 120-133.
35. Grabert M, Prechtel M, Hrycej T, Günther W. An early warning system for vehicle related quality data. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science 3275)* 2004; 88-95
36. Honari B, Donovan J. Early detection of reliability changes for a non-Poisson life model using field failure data. *Proceedings of Annual Reliability and Maintainability Symposium, RAMS* 2007; 346-349
37. Vittal S, Neuman H. Early detection of warranty issues: A multi-disciplinary literature survey. *Proceedings of Annual Reliability and Maintainability Symposium* 2008;28-31
38. Majeske KD, Lynch-Caris T, Herrin G. Evaluating product and process design changes with warranty data. *International Journal of Production Economics* 1997; **50**: 79-89.
39. Majeske KD, Herrin GD. Determining warranty benefits for automobile design changes.
40. Yang K, Cekecek E. Design Vulnerability Analysis and Design Improvement by Using Warranty Data. *Quality and Reliability Engineering International* 2004; **20**: 121-133.

41. Buddhakulsomsiri J, Siradeghyan Y, Zakarian A, Li X. Association rule-generation algorithm for mining automotive warranty data. *International Journal of Production Research* 2006; **44**: 2749-2770.
42. Buddhakulsomsiri J, Zakarian A. Sequential pattern mining algorithm for automotive warranty data. *Computers and Industrial Engineering* 2009; **57**: 137-147.
43. Sureka A, De S Varma K. Mining automotive warranty claims data for effective root cause analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2008; **4947 LNCS**: 621-626.
44. Majeske KD. A mixture model for automobile warranty data. *Reliability Engineering and System Safety* 2003; **81**: 71-77.
45. Ion RA, Petkova VT, Peeters BHJ, Sander PC. Field reliability prediction in consumer electronics using warranty data. *Quality and Reliability Engineering International* 2007; **23**: 401-414.
46. Wilson S, Joyce T, Lisay E. Reliability estimation from field return data. *Lifetime Data Analysis* 2009; **15**: 397-410.
47. Marshall SE, Chukova S. On analysing warranty data from repairable items. *Quality and Reliability Engineering International* 2010; **26**: 43-52.
48. Kalbfleisch JD, Lawless JF. Estimation of reliability in field-performance studies. *Technometrics* 1988; **30**: 365-388.
49. Lawless JF, Kalbfleisch JD. Some issues in the collection and analysis of field reliability data. *In Survival Analysis: State of the Art*, Klein JP Goel PK(eds). Kluwer Academic: Dordrecht, 1992.
50. Hu XJ, Lawless JF. Pseudolikelihood estimation in a class of problems with response-related missing covariates. *Canadian Journal of Statistics* 1997; **25**: 125-142.
51. Hu XJ, Lawless JFSuzuki K. Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics* 1998; **40**: 3-13.
52. Hu XJ, Lawless JF. Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika* 1996; **83**: 747-761.
53. Oh YS, Bai DS. Field data analyses with additional after-warranty failure data. *Reliability Engineering and System Safety* 2001; **72**: 1-8.
54. Attardi L, Guida M, Pulcini G. A mixed-Weibull regression model for the analysis of automotive warranty data. *Reliability Engineering and System Safety* 2005; **87**: 265-273.
55. Suzuki K. Analysis of field failure data from a nonhomogeneous Poisson process. *Reports on Statistical Applications in Research* 1987; **34**: 6-15.
56. Suzuki K, Alam MM, Yoshikawa T, Yamamoto W. Two methods for estimating product lifetimes from only warranty claims data. *Proceedings of the 2nd IEEE International Conference on Secure System Integration and Reliability Improvement* 2008; 111-119.

57. VINTR Z, VINTR M. Estimate of warranty costs based on research of the customer's behavior. *Proceedings of Annual Reliability and Maintainability Symposium, RAMS 2007*, 323-328
58. KARIM MR, SUZUKI K. Analysis of warranty data with covariates. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2007*; **221**: 249-255.
59. VINTA S. Analysis of data to predict warranty cost for various regions. *Proceedings of Annual Reliability and Maintainability Symposium 2009*. 78-82
60. HRYCEJ T, GRABERT M. Warranty cost forecast based on car failure data. *IEEE International Conference on Neural Networks - Conference Proceedings 2007*, 108-113
61. BAIK J, MURTHY DNP. Reliability assessment based on two-dimensional warranty data analysis. *Proceedings of the European Safety and Reliability Conference 2006*; 2307-2311.
62. DAVIS T. A simple method for estimating the joint failure time and failure mileage distribution from automobile warranty data. *Ford Technical Journal 1999*; **2**:
63. SINGPURWALLA ND, WILSON SP. Failure models indexed by two scales. *Advances in Applied Probability 1998*; **30**: 1058-1072.
64. CHUKOVA S, ROBINSON J. Estimating mean cumulative functions from truncated automotive warranty data. *Modern Statistical and Mathematical Methods in Reliability 2005*; 121-136.
65. SINGPURWALLA ND, WILSON S. Warranty problem: its statistical and game theoretic aspects. *SIAM Rev 1993*; **35**: 17-42.
66. MOSKOWITZ H, CHUN YH. Poisson regression model for two-attribute warranty policies. *Naval Research Logistics 1994*; **41**: 355-376.
67. YANG S, NACHLAS JA. Bivariate reliability and availability modeling. *IEEE Transactions on Reliability 2001*; **50**: 26-35.
68. PAL S, MURTHY GSR. An application of Gumbel's bivariate exponential distribution in estimation of warranty cost of motor cycles. *International Journal of Quality & Reliability Management 2003*; **20**: 488-502.
69. JUNG M, BAI DS. Analysis of field data under two-dimensional warranty. *Reliability Engineering and System Safety 2007*; **92**: 135-143.
70. GERTSBAKH IB, KORDONSKY KB. Parallel time scales and two-dimensional manufacturer and individual customer warranties. *IIE Transactions (Institute of Industrial Engineers) 1998*; **30**: 1181-1189.
71. DUCHESNE T, LAWLESS J. Alternative Time Scales and Failure Time Models. *Lifetime Data Analysis 2000*; **6**: 157-179.
72. AHN CW, CHAE KC, CLARK GM. Estimating parameters of the power law process with two measures of failure time. *Journal of Quality Technology 1998*; **30**: 127-132.

73. Iskandar BP, Blischke WR. Reliability and warranty analysis of a motorcycle based on claims data. *In Case Studies in Reliability and Maintenance*, Blischke WR Murthy DNP(eds). John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003.
74. Kleyner A, Sandborn P. Forecasting the cost of unreliability for products with two-dimensional warranties. *Proceedings of the European Safety and Reliability Conference 2006*, 1903-1908
75. Kaminskiy MP, Krivtsov VV. G-renewal process as a model for statistical warranty claim prediction. *Proceedings of Annual Reliability and Maintainability Symposium 2000*; 276-280.
76. Kijima M, Sumita N. A Useful Generalization of Renewal Theory: Counting Process Governed by Non-negative Markovian Increments. *Journal of Applied Probability* 1986; **23**: 71-88.
77. Majeske KD. A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering and System Safety* 2007; **92**: 243-251.
78. Fredette M, Lawless JF. Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims. *Technometrics* 2007; **49**: 66-80.
79. Wu S, Zuo MJ. Linear and nonlinear preventive maintenance models. *IEEE Transactions on Reliability* 2010; **59**: 242-249.
80. Wasserman GS, Sudjianto A. Neural networks for forecasting warranty claims. *Intelligent Engineering Systems Through Artificial Neural Networks* 2001; 901-906.
81. Wasserman GS, Sudjianto A. A comparison of three strategies for forecasting warranty claims. *IIE Transactions (Institute of Industrial Engineers)* 1996; **28**: 967-977.
82. Rai B, Singh N. Forecasting warranty performance in the presence of the 'maturing data' phenomenon. *International Journal of Systems Science* 2005; **36**: 381-394.
83. Wasserman GS. An application of dynamic linear models for predicting warranty claims. *Computers and Industrial Engineering* 1992; **22**: 37-47.
84. Chen J, Lynn NJ, Singpurwalla ND. Forecasting warranty claims, chapter 31. *In Product Warranty Handbook*, Blischke WR Murthy DNP(eds). Marcel Dekker: 1996.
85. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**: 158-168.
86. Hu XJ, Lawless JF. Estimation of rate and mean functions from truncated recurrent event data. *Journal of the American Statistical Association* 1996; **91**: 300-310.
87. Karim MR, Yamamoto W, Suzuki K. Statistical Analysis of Marginal Count Failure Data. *Lifetime Data Analysis* 2001; **7**: 173-186.
88. Wang L, Suzuki K, Yamamoto W. Age-based warranty data analysis without date-specific sales information. *Applied Stochastic Models in Business and Industry* 2002; **18**: 323-337.

89. Krivtsov V, Frankstein M. Nonparametric estimation of marginal failure distributions from dually censored automotive data. *Proceedings of the Annual Reliability and Maintainability Symposium* 2004; 86-89.
90. Marcorin AJ, Abackerli AJ. Field failure data: An alternative proposal for reliability estimation. *Quality and Reliability Engineering International* 2006; **22**: 851-862.
91. Suzuki K. Estimation of usage lifetime distribution from calendar lifetime data. *Reports of Statistical Application Research* 1993; **40**: 10-22.
92. Phillips MJ, Sweeting TJ. Estimation for censored exponential data when the censoring times are subject to error. *Journal of Royal Statistics Society-B* 1996; **58**: 775-783.
93. Phillips MJ, Sweeting TJ. Estimation from Censored Data with Incomplete Information. *Lifetime Data Analysis* 2001; **7**: 279-288.
94. Phillips MJ. Bayesian Estimation from Censored Data with Incomplete Information. *Quality and Reliability Engineering International* 2004; **20**: 237-245.
95. Mohan K, Cline B, Akers J. A practical method for failure analysis using incomplete warranty data. *Proceedings of Annual Reliability and Maintainability Symposium*, 2008
96. Meeker WQ, Escobar LA, Hong Y. Using accelerated life tests results to predict product field reliability. *Technometrics* 2009; **51**: 146-161.
97. Zuo J, Meeker WQ, Wu H. Analysis of window-observation recurrence data. *Technometrics* 2008; **50**: 128-143.
98. Escobar LA, Meeker WQ. Statistical prediction based on censored life data. *Technometrics* 1999; **41**: 113-124.
99. Elkins DA, Wortman MA. An examination of correlation effects among warranty claims. *IEEE Transactions on Reliability* 2004; **53**: 200-204.