Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1282

Published online 10 January 2012 in Wiley Online Library

Warranty Data Analysis: A Review

Shaomin Wu*

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 article attempts to summarise and review the research and developments in WDA with emphasis on models, methods and applications. It concludes with a brief discussion on current practices and possible future trends in WDA. Copyright © 2012 John Wiley & Sons, Ltd.

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.[1] Product warranty is becoming increasingly more important in consumer and commercial transactions and is widely used to serve many different purposes.[2–4]

A vast literature on warranty can be found; for example, in 1996, Djamaludin *et al*.[5] listed more than 1500 articles in this area. Recently, research in warranty has attracted even more attention, as can be seen from the review articles[6–11] and the books.[12–17]

Warranty data are composed 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 field reliability. Beginning with Suzuki,[18,19] extensive research on warranty data analysis (WDA) has been dome.[9,20–27] The study of Karim *et al*.[9] is an excellent review article on warranty claim data analysis (for other review articles on WDA before 2005, see Kalbfleisch *et al*.,[20] Lawless,[22] Suzuki *et al*.,[23] Robinson and McDonald[28] and Kalbfleisch and Lawless[29,30]). Over the last 5 years, more articles have been published, and a comprehensive review article 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 (e.g. 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, and so forth;
- 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 earlier, the main objectives of this article are (i) to review the existing research in WDA and (ii) to suggest new directions for future research on the basis of the trends and issues identified.

The rest of this article 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 methods that have been proposed and studied in the literature. Section 4 explains two tables summarising the articles reviewed. Section 5 concludes the article with a brief discussion on current practices and possible future trends in WDA.

School of Applied Sciences, Cranfield University, Bedfordshire MK43 0AL, UK
**Correspondence to: Shaomin Wu, School of Applied Sciences, Cranfield University, Bedfordshire MK43 0AL, UK.*
†E-mail: Shaomin.wu@cranfield.ac.uk

2. Warranty claims

A typical life cycle 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.

We can express warranty claim data as shown in Table I, 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.

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

- 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, and so forth.
- 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.[31]

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.

Figure 1. A typical life cycle of failed products

Table I. 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}

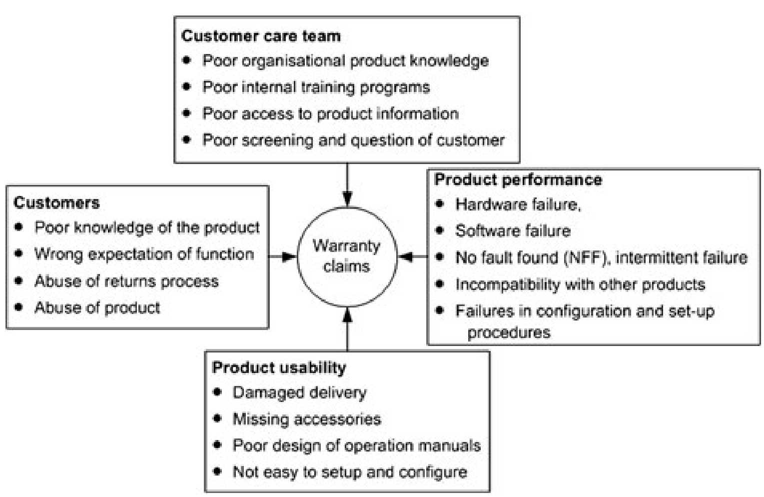

Figure 2. Examples of causes of warranty claims

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 (e.g. see Wu[32]).

3. Warranty data analysis

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

3.1. *Early detection of reliability problems*

The intention of the 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 in 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.
- Nonstationary 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 nonstationary.
- 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, several techniques have been developed to detect product abnormalities from warranty data. Karim *et al.*[33] proposed 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 postdesign change point. Wu and Meeker[34] used statistical detection rules to provide an early indication of reliability changes with the Poisson distribution estimation. Grabert *et al.*[35] developed an early detection system using neural networks and probability distribution estimation. Honari and Donovan[36] used control charts to monitor any changes and to validate their approach on the basis of both artificially generated data and warranty claims data.

Vittal and Neuman[37] surveyed the emerging science of early detection for the warranty and reliability issues. They summarised that three measures should be required to assess the efficacy of any 'early warning' detection system/algorithm: (i) the probability of detection of a change, (ii) the probability of false alarm and (iii) the alarm time ('time to detect' a change). However, in the literature, not all of the authors used 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. The early detection of reliability problems using warranty data can be more difficult compared with machinery health monitoring. 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.

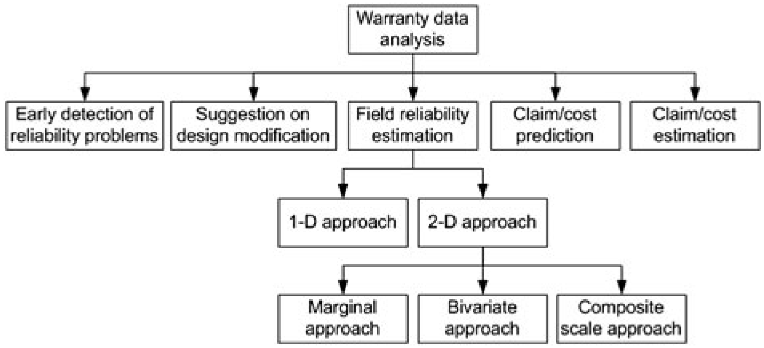

Figure 3. Analysis of warranty data

3.2. *Suggestion on design modification*

Warranty data can also be used 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, online monitoring techniques such as control charts might be pursued. Design modification, however, emphases 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.*[38] and Majeske and Herrin[39] used graphical tools to compare warranty claims with their benchmark data and suggested that manufacturers should analyse design changes using *post hoc* tests on warranty data. Yang and Cekecek[40] developed a method by using warranty data to prioritise design improvement efforts on the basis of design vulnerability and the feedback from warranty claims data. They used cost as a measure of design vulnerability to indicate the weaknesses of the design and provided feedback for design improvements and then used 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 (e.g. failure analysis). Buddhakulsomsiri *et al.*[41,42] used the elementary set concept and database manipulation techniques to search patterns and relationships among occurrences of warranty claims and to 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[43] developed 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 on the basis of warranty data can provide manufacturers with more important information.

When estimating product reliability from warranty claims data, however, we need to notice[22] the following:

- 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.[32]

Warranty policies can be categorised into one dimension and two dimensions. A one-dimensional (1D) policy is characterised by an interval (age only or usage only) as warranty limit. A two-dimensional (2D) policy is represented by a region in the 2D 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.) and 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 analyse warranty data collected from the products with 1D or 2D policies can be different. In the following sections, we reviewed publications in two aspects: 1D and 2D analysis for field reliability estimation.

3.3.1. One-dimensional approach. It should be noted that 1D 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 because techniques on estimating field reliability are well established, given complete age information.

Approaches in estimating the lifetime distribution include estimating mixed distributions,[44] fitting the Weibull distribution on the basis of a small number of warranty claims,[45] estimating the lifetime distribution considering sales delay[46] and estimating the intensity of a nonhomogeneous Poisson process (NHPP) for repairable items.[47]

3.3.1.2. Usage-based analysis. The 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; hence, modelling 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 nonfailure automobiles) because the usage time distributions of nonfailed products are different from those of failed products.

Starting from Suzuki,[18,19] there are a series of research on estimating product reliability when incomplete censored data, that is, incomplete usage data, are presented. Approaches in dealing with the case of the incomplete usage data have been studied by Lawless *et al.*[48–51]

One of often used approaches in dealing with incomplete censored data is the supplementary data approach. This approach randomly selects a follow-up sample of products from the nonfailed products under warranty, obtains their censoring times, usage history and/or any covariate values and then uses a pseudo-likelihood approach in 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 nonparametric based.[18,19] The pseudo-likelihood approach can also be extended to analyse claims data with covariate information.[48,52]

When additional field data are available, Oh and Bai[53] proposed methods in estimating the lifetime distribution. Attardi *et al.*[54] used 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[55] proposed to use NHPPs. Suzuki *et al.*[56] presented two methods, parametric and semiparametric, to estimate product field reliability, where nonfailure data are not included.

The usage intensity of products by specific groups of consumers are surveyed and analysed by Vintr and Vintr.[57]

However, it has been noted that little attention in the previous 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, he or she will find that many other failed items may be claimed due to the other failure modes. He or she 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. Covariate 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[48]:

- Information on the types and frequencies of problems (e.g. failures, replacements, etc.) and on the 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.[22] Covariates may be introduced at an individual unit level or at an aggregate claims level. In either case, log linear models are convenient.[22]

Kalbfleisch and Lawless[48] suggested procedures for the collection of field data and then used a regression model to estimate lifetime distributions from field failure data with supplementary information about covariates. Hu and Lawless[52] developed an estimation procedure with supplementary information about covariates and censoring times, suggested a technique for modelling warranty claims as truncated data and assumed warranty claims follow a Poisson process. Hu and Lawless[50] considered situations involving a response variable and covariates, both of which are incomplete. They discussed two types of pseudo-likelihood estimation and provided methods for the nonparametric estimation of lifetime distributions. Attardi *et al.*[54] used a mixed-Weibull regression model for the analysis of automotive warranty claims data, assuming that the products are a mixture of weak and strong subpopulations with respect to their reliabilities. The engine type and car model are used as covariates in their regression model. Karim and Suzuki[58] considered covariates associated with some reliability-related factors and presented a Weibull regression model for the lifetime of the component as a function of such covariates. The expectation maximisation 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 Vinta[59] and Hrycej and Grabert.[59,60]

3.3.2. Two-dimensional approach. Much of the literature on warranty analysis considers failure models that 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 on the basis of both age and usage amounts.

For items under the 2D warranty policy, Figure 4 shows the possible ages and usage rates of four items covered with a 2D 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 the usage at failure above the age limit and the usage limit.

In analysing 2D 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 the data of those failed and reported items are available, but the usage for those items whose warranty has expired cannot be obtained.

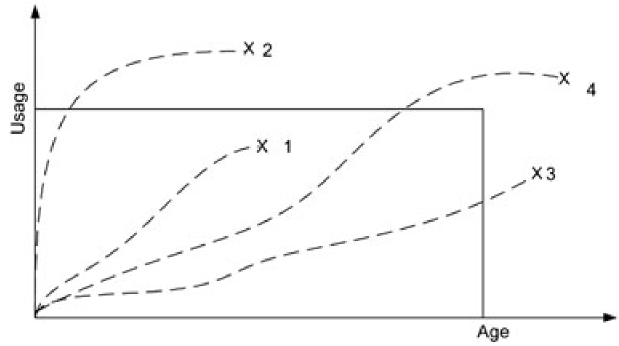

Figure 4. Two-dimensional warranty

To analyse 2D 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 2D warranty problem to a 1D 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 (e.g. low, medium and high users) or as a continuous variable with a density function.[61] A joint distribution can then be calculated from the marginal distributions. With this approach, Lawless *et al.*[21] considered 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[62] discussed a similar approach. Singpurwalla and Wilson[63] proposed 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[64] took age and mileage as the usage measure, respectively, and evaluated the mean cumulative number of claims or cost of claims and its standard error as functions of the usage measure, with both parametric and nonparametric approaches. Kleyner and Sanborn[65] presented a model where the usage time is a primary variable and the mileage accumulation is estimated from field return data. Their modelling 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[66] developed 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[67] assumed that the number of failures under the 2D 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[68] developed bivariate renewal models and also explained their inclusion in availability models. Pal and Murthy[69] used Gumbel's bivariate exponential distribution to fit warranty claims. Jung and Bai[70] considered a bivariate approach and assumed that age and usage are statistically correlated in a bivariate distribution. Lawless and Crowder[25] presented models to assess the dependence on age or usage in heterogeneous populations of products and showed how to estimate model parameters on the basis of different types of field data. Lawless and Crowder[27] provided joint models for the recurrent events and usage processes, which facilitated the analysis of their relationship as well as the prediction of failures.

3.3.2.3. Composite scale approach. In addition to the previously mentioned two approaches, Gertsbakh and Kordonsky[71] and Duchesne and Lawless[72] 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 the study of Gertsbakh and Kordonsky,[71] a new variable $V = \varepsilon T + (1\text{-}\varepsilon)U$ 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.*[73] presented the power law process with the new time scale as a model for the reliability of a repairable system. This approach was used by Iskandar and Blischke[74] 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 1D warranty, 5-year warranty, in one country, but have 2D warranty, 4-year warranty or 48,000 km, in other countries. An interesting question is to sufficiently use the warranty data collected from both the countries to estimate survivor distributions.

Data (e.g. 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.

3.4.1 Lifetime distributions. This approach is to estimate a time-to-claim distribution. In the literature, Kleyner and Sandborn[75] presented a warranty claim forecasting model on the basis of a piecewise application of Weibull and exponential distributions, which tries to capture the dynamic characteristic features of failure rates in both the early failure period and the intrinsic failure period of the bathtub curve. Rai[26] presented 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.

3.4.2 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.*[20] used a log-linear Poisson model to analyse and forecast warranty claims. In their work, they modelled warranty claims on the basis of 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 were also used by Singpurwalla and Wilson.[66] Kaminskiy and Krivtsov[76] developed warranty claim forecasting models with the G-renewal process–generalised renewal processes introduced by Kijima and Sumita,[77] the ordinary renewal process and the NHPP. They found that GRP provides a higher accuracy compared with the ordinary renewal process or the NHPP. Majeske *et al.*[78] presented an NHPP-based technique that forecasts the total number of claims and the timing of claims during the vehicle lifetime. Fredette and Lawless[79] presented forecasting methods for warranty claims, using mixed NHPP, and possible heterogeneity among the individuals is modelled using random effects. Kleyner and Sanborn[80] presented a Monte Carlo simulation procedure to generate the missing time from build to sale data and to conduct a statistical analysis of build to sale data for various vehicle name plates and platforms to enhance the forecasting procedure and improve its accuracy.

The stochastic process approaches might require assumptions such as the claim rates following a specific law (e.g. NHPP). Such an assumption might be violated because the quality of maintenance might not be difficult to assess.[81]

3.4.3 Artificial neural networks. Nonparametric approaches such as neural networks have also been applied to predict warranty claims. Starting from Wasserman and Sudjianto,[82,83] multilayer perceptron (MLP)[82,83] and radial basis function[84] have been used.

3.4.4. 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 I, the claim rates of each month can then been seen as time series. The techniques used in time series prediction can then borrowed. Singpurwalla and Wilson[66] considered using the Kalman filter to build forecasting models. Wasserman and Sudjianto[85] developed linear regression models, first-order uto-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[83] further compared the linear regression models, the time series models, the Kalman filter, the orthogonal series and the MLP from artificial neural networks in modelling and forecasting warranty claims and concluded 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.*[86] proposed 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 because 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 previously mentioned: (i) 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 (ii) they are developed on the basis of 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[31] introduced two different approaches in 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 more than 6 months ahead forecasting, the results show that the proposed models exhibit superior performance compared with that of MLP and radial basis function neural networks and ordinary SVR 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.*[20] used a log-linear Poisson model to estimate warranty claims, where the Poisson parameter is a function of time in service. Lawless and Kalbfleisch[49] introduced a moment estimator of the expected number of warranty claims for a single product.

The NHPP is one of most widely used techniques in estimating the number of warranty claims. Lawless and Nadeau[87] developed point estimates on the basis of Poisson models for the cumulative mean function of warranty claims. Hu and Lawless[88] presented a nonparametric approach in estimating the rate and mean function from truncated recurrent event data, assuming that the observation period for products under warranty is unknown until it experiences at least one claim. For aggregated warranty claims data, Lawless[22] estimated the expected number of claims and also gave the variance of the estimate.

When warranty claims are aggregated, Suzuki *et al.*[23] and Karim *et al.*[89] presented an NHPP model for repairable products, a multinomial model and its Poisson approximation for nonrepairable products and then used the expectation maximisation algorithm to estimate the expected number of claims. Wang *et al.*[90] and Suzuki *et al.*[23] used NHPPs to model warranty claims for repairable products.

The methods developed for estimating warranty claims have been surveyed by Kalbfleisch *et al.*[20] Robinson and McDonald,[28] and Kalbfleisch and Lawless.[30]

It might be confusing with the difference between warranty claim estimation and prediction. The difference of these two issues lies in the study of Kalbfleisch and Lawless[30]: 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 article reviewed the existing work in WDA. Historically, WDA has been mainly focussed 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, although 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, that is, 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 miniaturisation of radio frequency devices and the microelectromechanical systems as well as the advances in wireless technologies make 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*. The early detection of reliability problems requires efficient algorithms, which needs more research. The most important challenges might be to detect (i) critical failure modes that can result in human injuries and significant property losses and (ii) 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 because of 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 article.[11] 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, the data of long-term warranty have received little attention. For example, the research on warranty prediction reviewed earlier 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 Murthy and Djamaludin[7,22] and Lawless[22] for information in this area.

This review has attempted to be reasonably complete. However, those articles 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 the readers if any relevant articles have been omitted.

For further readings

The following studies are also of interest to the reader: Krivtsov and Frankstein,[91] Marcorin and Abackerli,[92] Suzuki,[93] Phillips and Sweeting,[94,95] Phillips,[96] Mohan *et al.*,[97] Meeker *et al.*,[98] Zuo and Meeker,[99] Escobar and Meeker[100] and Elkins and Wortman.[101]

Acknowledgements

This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) of the UK (EPSRC grant reference no. EP/G039674/1).

The authors 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. Djamaludin D, Murthy D, Blischke W. Bibliography on warranties. *Product Warranty Handbook*. Marcel Dekker: New York,1996; 839–916.
6. Thomas MU, Rao SS. Warranty economic decision models: A summary and some suggested directions for future research. *Operations Research* 1999; **47**:807–820.
7. Murthy DNP, Djamaludin I. New product warranty: A literature review. *International Journal of Production Economics* 2002; **79**:231–260.
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 Operational Research* 2006; **23**:453–472.
12. Blischke WR, Murthy DNP. *Warranty cost analysis*. CRC Press: New York, 1994.
13. Brennan JR. *Warranties, Planning, Analysis and Implementation*. McGraw-Hill Professional: New York, 1994.
14. Sahin I, Polatoglu H. *Quality, Warranty and Preventive Maintenance*. Kluwer Academic Pub: Boston, 1998.
15. Murthy DNP, Blischke WR. *Warranty Management and Product Manufacture*. Springer: London, 2005.
16. Thomas MU. *Reliability and warranties: methods for product development and quality improvement*. Taylor & Francis Group: New York, 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-A. 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; **48**:649–657.
27. Lawless JF, Crowder MJ. Models and estimation for systems with recurrent events and usage processes. *Lifetime Data Analysis* 2010; **16**:547–570.
28. Robinson JA, McDonald GC. Issues related to field reliability and warranty data. In *Data Quality Control: Theory and Pragmatics*, Liepins GE, 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, 2011. 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)* 2004; **3275**: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. Proceedings of the Annual Reliability and Maintainability Symposium, Anaheim, CA, USA, 19-22 Jan 1998, 94–99.

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 SVarma 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. Pseudo-likelihood estimation in a class of problems with response-related missing covariates. *The Canadian Journal of Statistics* 1997; **25**:125–142.
51. Hu XJ, Lawless JF, Suzuki 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 non-homogeneous 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 behaviour. *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 and an accelerated failure time model. *International Journal of Reliability and Safety* 2008; **2**(3):190–208.
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**:1–11.
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. *Series on Quality Reliability and Engineering Statistics −10: Modern Statistical and Mathematical Methods in Reliability* 2005; 121–136.
65. 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.
66. Singpurwalla ND, Wilson S. Warranty problem: its statistical and game theoretic aspects. *SIAM Review* 1993; **35**:17–42.
67. Moskowitz H, Chun YH. Poisson regression model for two-attribute warranty policies. *Naval Research Logistics* 1994; **41**:355–376.
68. Yang S-C, Nachlas JA. Bivariate reliability and availability modelling. *IEEE Transactions on Reliability* 2001; **50**:26–35.
69. 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.
70. Jung M, Bai DS. Analysis of field data under two-dimensional warranty. *Reliability Engineering and System Safety* 2007; **92**:135–143.
71. 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.
72. Duchesne T, Lawless J. Alternative time scales and failure time models. *Lifetime Data Analysis* 2000; **6**:157–179.
73. 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.
74. 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.
75. Kleyner A, Sandborn P. A warranty forecasting model based on piecewise statistical distributions and stochastic simulation. *Reliability Engineering and System Safety* 2005; **88**:207–214.
76. Kaminskiy MP, Krivtsov VV. G-renewal process as a model for statistical warranty claim prediction. *Proceedings of the Annual Reliability and Maintainability Symposium* 2000; 276–280.
77. 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.
78. Majeske KD. A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering and System Safety* 2007; **92**:243–251.
79. Fredette M, Lawless JF. Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims. *Technometrics* 2007; **49**:66–80.
80. Kleyner A, Sanborn K. Modelling automotive warranty claims with build-to-sale data uncertainty. *International Journal of Reliability and Safety* 2008; **2**:179–189.
81. Wu S, Zuo MJ. Linear and nonlinear preventive maintenance models. *IEEE Transactions on Reliability* 2010; **59**:242–249.
82. Wasserman GS, Sudjianto A. Neural networks for forecasting warranty claims. *Intelligent Engineering Systems Through Artificial Neural Networks* 2001; 901–906.
83. Wasserman GS, Sudjianto A. A comparison of three strategies for forecasting warranty claims. *IIE Transactions (Institute of Industrial Engineers)* 1996; **28**:967–977.
84. Rai B, Singh N. Forecasting warranty performance in the presence of the 'maturing data' phenomenon. *International Journal of Systems Science* 2005; **36**:381–394.

85. Wasserman GS. An application of dynamic linear models for predicting warranty claims. *Computers and Industrial Engineering* 1992; **22**:37–47.
86. Chen J, Lynn NJ, Singpurwalla ND. Forecasting warranty claims, chapter 31. In *Product Warranty Handbook*, Blischke WR, Murthy DNP (eds). Marcel Dekker: New York, 1996.
87. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**:158–168.
88. 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.
89. Karim MR, Yamamoto W, Suzuki K. Statistical analysis of marginal count failure data. *Lifetime Data Analysis* 2001; **7**:173–186.
90. 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.
91. 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.
92. Marcorin AJ, Abackerli AJ. Field failure data: An alternative proposal for reliability estimation. *Quality and Reliability Engineering International* 2006; **22**:851–862.
93. Suzuki K. Estimation of usage lifetime distribution from calendar lifetime data. *Reports of Statistical Application Research* 1993; **40**:10–22.
94. 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.
95. Phillips MJ, Sweeting TJ. Estimation from censored data with incomplete Information. *Lifetime Data Analysis* 2001; **7**:279–288.
96. Phillips MJ. Bayesian Estimation from censored data with incomplete information. *Quality and Reliability Engineering International* 2004; **20**:237–245.
97. Mohan K, Cline B, Akers J. A practical method for failure analysis using incomplete warranty data. *Proceedings of Annual Reliability and Maintainability Symposium* 2008:193–199.
98. Meeker WQ, Escobar LA, Hong Y. Using accelerated life tests results to predict product field reliability. *Technometrics* 2009; **51**:146–161.
99. Zuo J, Meeker WQ, Wu H. Analysis of window-observation recurrence data. *Technometrics* 2008; **50**:128–143.
100. Escobar LA, Meeker WQ. Statistical prediction based on censored life data. *Technometrics* 1999; **41**:113–124.
101. Elkins DA, Wortman MA. An examination of correlation effects among warranty claims. *IEEE Transactions on Reliability* 2004; **53**:200–204.

Author's biography

Dr Shaomin Wu is a lecturer in risk and decision analysis at Cranfield University. His research interests include warranty data analysis, risk analysis, reliability analysis, maintenance policy optimisation and warranty policy optimisation. He has served as a co-chairman of an international conference and a PC Member of a number of international conferences in reliability engineering and computer sciences. He is also an Editorial Board Member on three international journals. He has published over 35 papers in international journals.