# A review on coarse warranty data and analysis

Shaomin Wu[1]

*Kent Business School, University of Kent, Canterbury CT2 7PE, UK*

**Abstract**

Warranty data contain useful information about product quality and reliability, but they are usually coarse data because they may be aggregated, delayed, censored, missing or vague. They might, however, be the only forms of warranty data a manufacturer has, analysing such data are therefore needed and can also 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 warranty reserves plans.

In last two decades, considerable research has been conducted in analysing coarse warranty data (CWD) from several different perspectives. This paper categorises different types of CWD and reviews techniques to analyse such data. It concludes with research needs in CWD.

**Keywords**: Aggregated data, reporting delay, sales delay, warranty data, incomplete censored data, coarse data.

## 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 has become increasingly more important in consumer and commercial transactions and is widely used to serve many different purposes [2-7]. The US Congress has enacted several acts (UCC, Magnusson Moss Act, Tread Act, etc.) over the last 100 years. The European Union (EU) passed legislation requiring a two-year warranty for all products sold in Europe [8].

Warranty has been studied from many different perspectives by researchers from diverse disciplines and the literature on warranty is vast. As early as 1996, Djamaludin et al. [9] listed over 1500 papers on warranties. In recent years, research in warranty has attracted wide attention of researchers, as can be seen from the review papers [8,10-15] and the books [16-21].

---

[1] Email: s.m.wu@kent.ac.uk; Tel: 0044 1227 827940

Warranty data are comprised of claims data and supplementary data. Warranty claims data are the data collected during the servicing of items 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 reliability. Starting with Suzuki [22,23], considerable research on warranty data analysis has been conducted, see [12,24-30], for example. Karim et al. [12] is an excellent review paper that summarises the different statistical models and methods used to analyse warranty claims data. Wu [15] presents a new review on warranty data analysis.

Unlike data collected from laboratories where high quality data can be guaranteed, warranty data collected from the field usually have low quality for a variety of reasons that we will discuss in this paper. They can be aggregated, delayed, censored, missing, vague, etc. We refer to such data as "coarse warranty data". The term *coarse* is borrowed from Heitjan and Rubin [31], who define it as "heaped, censored and missing". As can be seen from our discussion in Section 2 of this paper, the coarse data problem arises in warranty data collection and analysis from time to time. However, no review paper specifically on this topic has been found, and it sometimes may confuse readers on the same term with different meanings (for example, different types of reporting delays). The aim of this paper is therefore to review different types of coarse warranty data and approaches to analysing them.

The remainder of this paper is structured as follows. Section 2 describes the relevant data that should ideally be collected for warranty data analysis. Section 3 categorises coarse warranty data and reviews approaches to analysing them. Section 4 summarises papers relating to the subject topic, and Section 5 concludes the paper with a discussion of the topics for research in the future.

## 2. Warranty data collection

Warranty claims data are lifetime data collected during the servicing of items over of warranty period. Analysing such lifetime data requires good quality data. The data collected during pre-launch are done in controlled conditions – failure and censoring times properly recorded. If all items are tested to failure then the data are said to be complete (as all items have failed). If not, the data contain some censored data in addition to failure data. In contrast the data collect during pre-launch, data from the field are often coarse for a variety of reasons that we discuss in the remainder of the section.

For proper analysis, one need to collect data such as amounts and dates: a typical sequence of events leading to warranty data are characterised by six time points (h) – (m), as indicated in Figure 1.
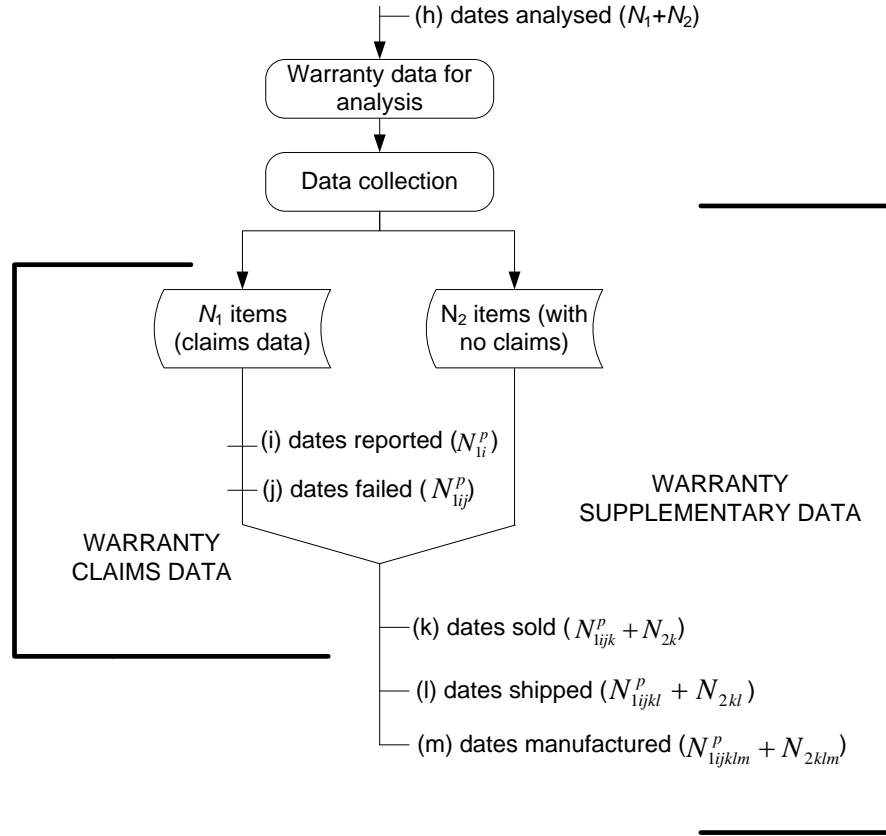


Figure 1: Warranty data: time history of events

Ideally, one can collect all of the relevant data from the points (h) – (m) indicated in Figure 1 as well as data relating to the time instant when the warranty for items expires. The data collection starts from the time instant that the product is launched in the market, not all relevant data, however, might be collected. Table 1 shows the relevant data that should ideally be collected.

Table 1: Description of Figure 1: relevant data that should ideally be collected.

| | |
|---|---|
| $N_1$: | Number of items with one or more warranty claims over the data collection period. |
| $N_1^p$: | Number of items that have had $p$ warranty claims over the data collection period $(N_1 = \sum_p N_1^p)$ |
| $N_2$: | Number of items with no warranty claims over the data collection period |
| $N_{1i}^p$: | Number of items with the $p^{th}$ $(p \geq 1)$ warranty claims on day $i$ $(N_1^p = \sum_i N_{1i}^p)$ |
| $N_{1ij}^p$: | Number of items that had the $p^{th}$ $(p \geq 1)$ warranty claim on day $i$ after failing on day $j$ |

$(N_{1i}^p = \sum_j N_{1ij}^p)$

$N_{1ijk}^p$ : Number of items that had the $p^{th}$ $(p \geq 1)$ warranty claim on day $i$, after failing on day $j$ and sold on day $k$ $(N_{1ij}^p = \sum_k N_{1ijk}^p)$

$N_{1ijkl}^p$ : Number of items that had the $p^{th}$ $(p \geq 1)$ warranty claim on day $i$, after failing on day $j$, sold on day $k$, and shipped to retailers on day $l$ $(N_{1ijk}^p = \sum_l N_{1ijkl}^p)$

$N_{1ijklm}^p$ : Number of items that had the $p^{th}$ $(p \geq 1)$ warranty claim on day $i$, after failing on day $j$, sold on day $k$, shipped to retailer on day $l$ and manufactured on day $m$ $(N_{1ijkl}^p = \sum_m N_{1ijklm}^p)$

$N_2$: Number of items with no warranty claims over the data collection period

$N_{2k}$: Number of items with no claims and sold on day $k$ $(N_2 = \sum_k N_{2k})$

$N_{2kl}$: Number of items with no claims, sold on day $k$ and shipped out to retailers on day $l$ $(N_{2k} = \sum_l N_{2kl})$

$N_{2klm}$: Number of items with no claims, sold on day $k$, shipped out to retailers on day $l$, and manufactured on day $m$ $(N_{2kl} = \sum_m N_{2klm})$

$t_{ij}^*$: Day when the $j^{th}$ warranty claim ( $j \geq 1$) for item $i$ was analysed $(\tilde{t}_{ij} \leq t_{ij} \leq t_{ij}^*)$

$t_{ij}$: Day when the $j^{th}$ warranty claim ( $j \geq 1$) for item $i$ was conducted $(1 \leq i \leq N_1)$

$\tilde{t}_{ij}$: Day when the $j^{th}$ failure ( $j \geq 1$) occurred for item $i$ $(\tilde{t}_{ij} \leq t_{ij})$

$t_{i0}$: Day when item $i$ was sold

$\tilde{t}_{i0}$: Day when item $i$ was produced $(\tilde{t}_{i0} \leq t_{i0})$

## 3. Coarse warranty data and analysis

Figure 2 illustrates several different scenarios when complete data are not collected or available to the warranty analyst.
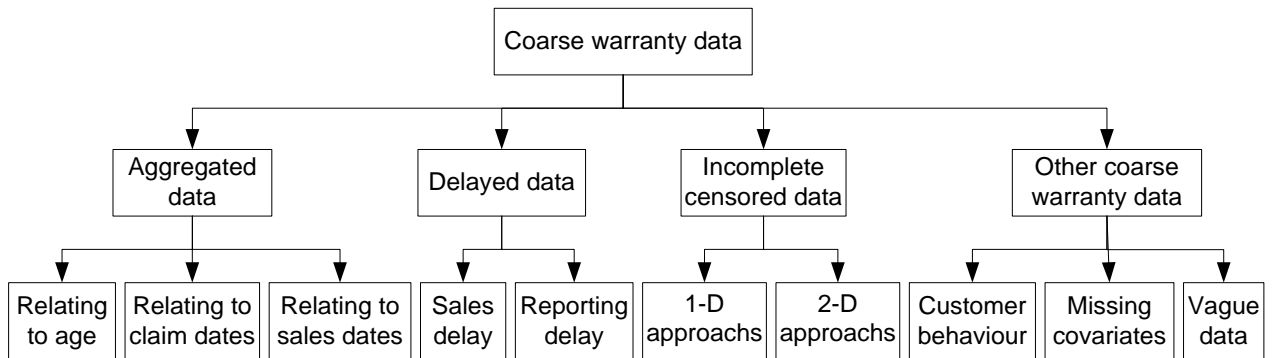


Figure 2: Different scenarios resulting in coarse data collection.

## 3.1 Aggregated data and analysis

Warranty data might only be available in the form of aggregated claims. That is, they might be aggregated into groups. Different parties (sales people, service agents, information processing team) might provide different types of aggregated data.

We refer to the time since an item is sold as the 'age' of the item. It is consequently informative to analyse claims as a function of age, bearing in mind that other factors may also need to be examined.

From aggregated data, one might not be able to obtain the exact age of failed items, but only knows that the age lies within an interval. As such, estimating warranty claims for aggregated warranty claims has been a focus in the literature. When one analyses aggregated claims, a commonly used assumption is that the number of warranty claims follows a Poisson distribution.

Warranty data might be aggregated at different time points in Figure 1. In the literature, the following three types of aggregated claims have been studied.

- *Relating to age* – or *type I aggregated claims.* The claims in this type are aggregated so that only the numbers of repairs or total claims for items with their ages lying in various intervals are observed [24]. This can occur at point (i) in Figure 1. For example, a data analyst might be only given the total number of claims for items in age 0-30 days, 31-60 days, etc. The length of time intervals can be constant or variable.

  Kalbfleisch et al. [24] derive a nonparametric approach to estimating the expected number of claims, where the age intervals in the aggregated claims are of the same lengths, with a NHPP (non-homogeneous Poisson process) model. Kalbfleisch and Lawless [32] further give an estimate of the expected number of claims when the age intervals are variable.

- *Relating to claim dates* – or *type II aggregated claims*. In this type, the total number of claims within a specific time period is known, but the exact date when an individual claim made is unknown. For example, at point (i) in Figure 1, the number of claims, $N_{1i}^p$, on a specific date $i$ might not be available, but the sum of claimed items, $N_1^p$, within a fixed time period can be obtained. This can be due to the fact that the maintenance database only contains the dates of the maintenance activities but not the age of items [27].

  Suzuki et al. [27,33] and Karim et al. [34] present NHPP models for repairable items, and a multinomial model and its Poisson approximation for non-repairable items, where

the expectation-maximization (EM) algorithm is also employed. The EM algorithm is a method for finding maximum likelihood estimates (MLE) of parameters for incomplete data problems [35]. It is an iterative algorithm that alternates between performing an expectation (E) step to compute the expectation of the likelihood evaluated, given observed data, and a maximization (M) step to compute parameters maximizing the expected log-likelihood found on the E step.

- *Relating to sales dates* – or *type III aggregated claims*. In this type, the number of items sold within a time period is known, but the number of items sold on a specific date is unknown. For example, at point (k) in Figure 1, the number $N^p_{1ijk}$ of items sold on day $k$ might not be available, but the sum of items sold, $\sum_{k \in T_k} N^p_{1ijk}$, within a fixed time period $T_k$, can be obtained. This type can also cause a problem of sales delay, which will be discussed in Section 3.2.2.

  On type III aggregated claims, Lawless and Kalbfleisch [36] derive an estimator of the expected number of claims per item based on a NHPP model. Wang et al. [37] discuss the properties of this estimator for the repairable case and derive a non-parametric MLE of the claim frequency in a multinomial model for the non-repairable case.  Similar discussion is made in Suzuki et al. [27,33].

## 3.2 Delayed data and analysis

There are two main types of delays: reporting delay and sales delay.

### 3.2.1  Reporting delay

Reporting delay is the period between the time an event occurs and the time when it is reported [38]. Presence of reporting delay can cause a problem in monitoring and analysing occurrences of the events, since at any time point, many recent events may still be unreported.

There are two types of reporting delay that have been reported and studied in the literature.

- *Type I reporting delay*— which assumes that the failure of an item under warranty will be reported immediately; but it might take some time before the reported claim is entered into the warranty database and is accessible for analysis. Type I reporting delay is the delay between the time when a failure is reported and the time when the report is entered into a database for analysis.

Type I reporting delay is caused by the manufacturers who might need time to verify the claims before the claims are entered into the database. It is $t_{ij} - \check{t}_{ij}$, the time difference between points (i) and (h) in Figure 1, assuming that time point (i) is located at the same point (j). The length of type I reporting delay is commonly shorter than three months [39].

Parametric and nonparametric approaches have been developed to deal with reporting delay problem. On type I reporting delay, two approaches have been developed: (1) One approach is to remove recently reported cases; this is undesirable if information is wanted promptly [36]. For example, at mid-month we might report warranty claims made up to the end of the second last month so that only claims with reporting delays exceeding 1.5 months would be missed [39]. (2) Another approach is to introduce reporting delay probabilities. For example, the NHPP model is used in estimating the reporting delay probability. When assuming the probabilities of reporting delay are given, Lawless and Kalbfleisch [36] and Kalbfleisch et al. [24] estimate the expected number of warranty claims. A NHPP model based method on estimating concurrently the reporting lag distribution and the expected number of claims has also been developed [24,33,39]. When the probabilities of reporting delay are typically estimated from historical data, [26] gives the variance of the estimates.

- *Type II reporting delay*— which assumes that the failure of an item might not be reported immediately; but a reported claim will be immediately entered into the warranty database and will be accessible for analysts. Type II reporting delay is the delay between the time when an item fails and the time when the warranty of the item is reported. It is usually caused by the product users who might not claim warranty on failures until the warranty coverage is about to expire. For example, a vehicle with some failure modes (ie., for vehicles, minor oil leaks, engine slow to start, unusual engine noise, etc.) can still be operated, but its user might delay reporting the warranty claim until the warranty coverage is about to expire [40,41]. It is $t_{ij}^{\bullet} - t_{ij}$, the time difference between points (j) and (i) in Figure 1, assuming that time point (i) overlaps time point (h). On type II reporting delay, Rai and Singh proposed a nonparametric maximum likelihood approach to estimate hazard rate functions [41].

### 3.2.2  Sales delay

Sales delay is caused by the fact that one might not be able to obtain the exact dates of sales or the exact dates when products were put into service or the time $\tilde{t}_{i0}$ are unknown. The length of sales delay for item $i$, $t_{i0} - \tilde{t}_{i0}$, is the time interval between the time when the item is produced and the time when the item is sold (or time when the sold item is put into service).

Presence of sales delay can cause a problem that the exact age of an item is unknown. It cannot be ignored in analysing warranty claims data. The larger the sales delay, the more likely a product will observe warranty claims [42].

The following two types of sales delay have been studied.

- *Type I sales delay*—For some types of products, the manufacturer might not be able to collect the dates of sales. As such, type I sales delay is the elapsed time between the time when an item is sold and the time when the warranty of the item is reported---for un-failed items might not be obtainable. For items that fail under warranty, the failure time and the potential censoring time might be obtainable as the date of sale is verified as part of the warranty claims process [27,43,44].

    Hu et al. [45] consider nonparametric estimation of the lifetime distribution for the population with type I sales delay. Karim and Suzuki [46] propose to a NHPP model to estimate the distribution of type I sales delay, where the sales delay is caused by type II aggregated claims. They also consider incorporating follow-up information on sales dates in estimating the distribution of sales delay and the number of claims.

    Parametric approaches are also applied in estimating sales delay distributions. Ion et al. [44] and Karim [47] use the Weibull distribution and the lognormal distribution to fit type I sales delay, respectively. Zhao and Steffey [48] treat claims data as interval-censored observations in time-to-failure analysis.

- *Type II sales delay*—Sales dates may be unavailable for both failed and un-failed items. As such, type II sales delay is the elapsed time between the time when an item is sold or manufactured and the time when the item starts to operate. This can occur in type II aggregated claims. For example, a manufacturer may have warranty data, from which only manufacturer dates can be found, but dates of sales are not available [49].

    Approaches to estimating the probability of sales delay can be parametric or nonparametric. Baxter [50] introduce an approach to constructing a nonparametric estimator of the discrete lifetime distribution from quasi-life tables for the scenarios

where type II sales delay exists, assuming that the same numbers of items are put into service in different time periods. Tortorella [51,52] considers a more general case where the different numbers of items are assumed to put into service in different time periods. Crowder and Stephens [52] offer an analysis of moment-based estimators is made and a derivation of their asymptotic distribution. Akbarov and Wu [53] estimate the expected number of warranty claims under both renewing and non-renewing warranty policies taking into account the sales delay.

For aggregated claims data with unknown date entering service, Lim [54] present a nonparametric approach to estimating sales delay probabilities with the multinomial model, assuming that sales amount for each period is provided. Karim and Suzuki [46] estimate the distribution of sales delay based on the Poisson model, assuming the sales dates are unknown. Both the approaches are non-parametric approaches based and use the EM algorithm to search optimal parameters.

However, the above research only considers either reporting delay or sales delay. Wilson et al. [55] propose an approach to estimating the lifetime distribution (or the distribution of time to failure ) of items when both type II reporting delay and type II sales delay are concurrently considered. They allow both delays to take arbitrary parametric distributions and use a Bayesian inference procedure to fit a model for the time in operation.

### 3.2.3 Comments

The word *delay* in *type I reporting delay* and in *type II reporting delay* has different meanings. In type I reporting delay, the exact times $t_{ij}^*$, $t_{ij}$ and $\tilde{t}_{ij}$ are eventually obtainable.

It is only a matter of time when those values will be available for analysis. However, in type II reporting delay, one does not know the exact time $\tilde{t}_{ij}$ and it might never be obtained.

On reporting delay, a more complicating question is a combination of type I and type II reporting delays. This is the situation that the time when a failure occurs, the time when the failure is reported, and the time when the report is entered into the database for analysis are at different time points. Nevertheless, it has not yet attracted any attention in the literature.

It should be noted that *sales delay* discussed in this paper is different from the term "sales delay" used in marketing, where sales delay can be caused by various uncertainties such as price increase. In marketing, delaying the launch of new products, or sales delay, can adversely affect a company's operating results and financial condition.

The lifetime of an item can be regarded as $w^- + w^0 + w^+$, where $w^-$ is sales delay, $w^0$ is time to failure, and $w^+$ is the period from the time when it fails to the time when it is analysed, or a type of reporting delay. It should be noted that from a perspective of the arithmetic operation one can equally treat $w^-$ and $w^+$. This does not imply that those approaches developed for reporting delays can be applied indiscriminately to the case of sales delays, as one can find the following differences between reporting delays and sales delays.

- Their causes are different. Type I reporting delay is caused by the warranty processing procedures of a manufacturer, which is a problem of the internal management of the manufacturer. Type II reporting delay is caused by customers and might be affected by failure modes (for example, soft failure or hard failure) and other factors. Type I or II sales delays are largely due to marketing behaviour, for example, advertisements, the quality of the products, etc. These differences may cause that the probability distributions are difference.

- The length of sales delay can impact on product lifetimes, whereas that of reporting delay cannot.

We can see that the above-reviewed approaches to dealing with aggregated data and delayed data can be either nonparametric- or parametric-based. When a parametric approach is used, the expectation-maximization (EM) algorithm is borrowed to obtain the maximum likelihood estimates (MLE) [27,34,49].

### 3.3 Incomplete censored data and analysis

Warranty policies can be categorised into one- and two-dimensional policies. 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 some scenarios, however, collecting data about the un-failed items for data analysis might not be easy, which can occur in both 1-D and 2-D scenarios. For example, in Figure 1, the age $T$ and/or usage rate $U$ of the $N_2$ un-failed items may not be obtained.

For items under 1-D warranty policy, as mentioned in Section 3.2.2, one might not be able to obtain their sales dates, and therefore the censoring times for the un-failed items might not be obtained.

For items under 2-D warranty policy, Figure 3 illustrates possible behaviour of four items sold with a 2-D warranty policy. Item 1 failed within the warranty region and the failure time and usage at failure are reported to the manufacturer. The other items do not fail within the warranty region and as such no information about them is reported. The warranty of item 2 ceases due to usage exceeding the limit $U_w$, although its age is still below the limit $T_w$. For item 3, its warranty ceases due to age exceeding the limit $T_w$, although its usage is still below the usage limit $\underline{U}_w$. Item 4 has both the age and usage at failure above the age and the usage limit $U_w$ and $T_w$, respectively.
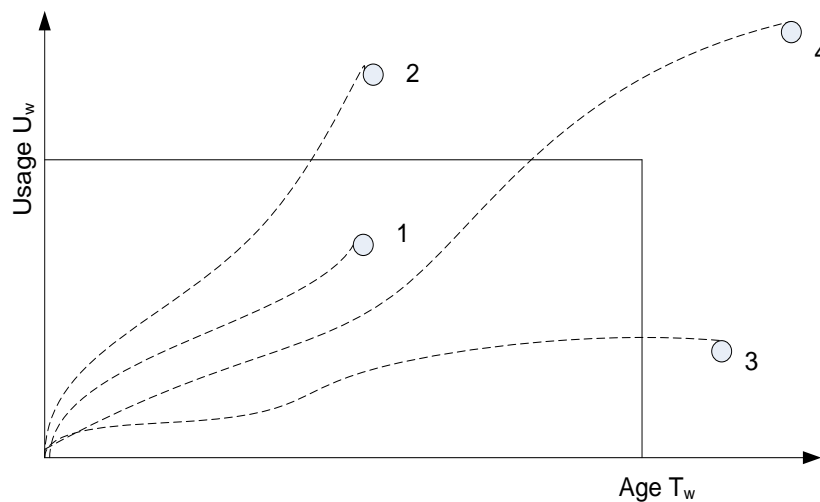


Figure 3. Two-dimensional warranty.

In analysing 2-D warranty data, for example, if one wants to estimate the distribution $F(T,U)$, we need to consider both age and usage. A complicating issue is that the age and usage might not be known to the manufacturer for some items. For example, in Figure 3, the age and usage of items 2, 3, and 4 might not be available. This presents a challenge that we can only collect data for those failed and reported items, but we do not know the age and the usage for those items whose warranty has expired.

Incomplete censored data can cause problems in analysing warranty claims data. For proper reliability assessment one needs both failure and censored data. If one estimates the reliability based solely on failure data then she draws biased inference.

The following two approaches have been developed to attack the problem of incomplete censored data.

### 3.3.1 Using both warranty claims data and supplementary data

As we discussed in the preceding contents, a manufacturer may not have data on un-failed items. One of the often used approaches to dealing with incomplete censored data is the supplementary data approach. This approach randomly selects a follow-up sample of items from the un-failed items under warranty and obtains their censoring times, usage history and/or any covariate values. Such a follow-up research can be follow-up studies, customer surveys, postal reply cards, and periodic inspections. The supplementary data approach is commonly used in lifetime distribution estimation.

In the case that follow-up data are available, a pseudo likelihood approach to estimating the parameters of survivor distributions has been developed. The pseudo likelihood approaches can be parametric or non-parametric based on Suzuki's work [22,23], and it can also be extended to analyse claims data with covariate information [56,57].

Lawless et al. [36] also offer a discussion on approaches to collecting supplementary data.

### 3.3.2 Using warranty claims data only

One might only use warranty claims data in data analysis when analysing 2-D warranty data. This leads a number of approaches developed. In the literature, three approaches have been developed to attack the problem of incomplete censored data for the case when supplementary data are not available. Those are univariate approach, bivariate approach, and composite scale approach. The univariate approach indirectly fits a joint distribution, the bivariate approach directly estimates a joint probability distribution, and the composite scale approach reduces the two-dimensional warranty problem to a one-dimensional formulation.

- *Univariate approach*: It indirectly estimates $F(T,U)$ through the following approaches: $F(T,U)= F(T|U) F(U)$ (where $F(T|U)$ is the conditional probability, T represents time scale, U represents usage scale, and F(.) represents probability distribution) or $F(T,U)= F(U|T) F(T)$. This is a univariate approach that concentrates on failure times by treating usage as a random function of age or time as a random function of usage: one might assume $U_i = g(T)$, where $g(.)$ is a random function, for example, $U_i = \alpha_i T$, where $\alpha_i$ is the usage rate that is a positive random variable. The research interest mainly focuses on estimation of the parameters in $F(T,U)$ [25,28,29,58-63].

- *Bivariate approach*. This approach directly estimates $F(T,U)$ from warranty data. Moskowitz and Chun [64] assume that the number of failures under the two-

dimensional warranty policies is distributed as a Poisson distribution with parameters that can be expressed by a regression function of the age and usage amounts of an item, and suggest a Poisson regression model to model failure times. Pal and Murthy [65] use Gumbel's bivariate exponential distribution to fit warranty claims. Jung and Bai [66] consider a bivariate approach and assume that age and usage are statistically correlated in a bivariate distribution; they regard each failed item within the warranty region $(0,T_w] \times (0,U_w]$ contributes a term f($t_i,u_i$) to the likelihood, and each unreported item contributes a term $1 - F(T_w, U_w)$. In case observations outside the warranty region

$(0,T_w] \times (0,U_w]$ are available (for example, if one casually obtains the usage rate of item 2 in Figure 2), however, the term $1 - F(T_w, U_w)$ cannot be used.

- *Time scale approach.* In addition to the above two approaches, Gertsbakh and Kordonsky [67], and Duchesne and Lawless [68] proposed methods of creating an alternative composite scale from age and usage, which integrates the two scales (age and usage) to create a single composite and scale and failures are modelled as a counting process using this composite scale. For example, in Gertsbakh and Kordonsky [67], 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. [69] present the power law process with the new time scale as a model for the reliability of a repairable system. The time scale approach is also used by Iskandar and Blischke [70] to model the warranty claims from a motorcycle manufacturer.

In addition to estimation of field reliability discussed in the above subsections, estimating warranty claims with incomplete censored data has also been studied. Chukova and Robinson [71] 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.

### 3.3.3 Comments

We conclude subsection 3.3 with two additional comments.

- It should be noted that the warranty policies for the same type of products can be different from region to region, because the different legislations exist in different countries. For example, cars with the same make may have a 1-D warranty policy, say, 5 year warranty in one country, but they may have a 2-D warranty policy, say, 3 year or

36,000km in other countries. An interesting question raised is to sufficiently use the warranty claims data collected from both regions in warranty data analysis.

- As far as we know, 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, she will find that many other failed items of the same product may be claimed due to the other failure modes. She will then have these censored and partial uncensored observations to estimate survivor distributions for the different failure modes. From this regard, there is no need to conduct any follow-up survey.

## 3.4 Other coarse warranty data and analysis

The above categorisation does not cover all the possible scenarios of incompleteness of warranty data. The following two areas are also studied in the literature.

### 3.4.1 Consumer behaviour

The human behaviours of product users and manufacturers can inevitably impact warranty execution. For example, they might have the following impacts.

- For un-failed products, the users may claim warranty, which can be due to a variety of reasons. For example, a user by accident spilled water into a laptop computer, which causes the computer failed. The user then claims warranty [72].

- Product users might not be bothered to execute warranty for failed items although the items are still under warranty [72-74]. Approaches to predicting occurred but not reported events are given by Lawless [38], in which cases about reportable diseases and insurance are considered.

### 3.4.2 Missing covariates

Many factors, or covariates, might contribute to product failures that result in warranty claims. In addition to factors such as age and usage rates, seasonal factors and production lines, for example, might be important. However, those factors are usually difficult to collect. Analysing warranty data with missing covariate information is another interesting topic [56,75-77].

### 3.4.3 Unclean claims data

Factors such as inexact time/mileage data and vague reported failures in warranty claims can also cause problems in data analysis and modelling [78]. Rai and Singh [78] discuss two parameter estimation methods, considering the incompleteness and uncleanness caused by the fact that warranty data are restricted only to the reported failures within warranty coverage and factors such as inexact time/mileage data and vague reported failures in a warranty claim make warranty data unclean.

## 4. Summary

In this section, we summarise the types of coarse warranty data discussed above with a quasi-life table, the three most frequently used techniques, and then the papers that have been reviewed in the above.

### 4.1 Quasi-life table and the coarse warranty data

We can express warranty claims data with a quasi-life table as shown in Table 2, where $r_{x,y}$ represents the number of claims received on the $(x + y - 1)$-th day and the claimed items are sold on day $D_x$, $M = \sum_{s_2}^{m_0} M_{s_2}$, and $r_{s_1} = \sum_{s_2=1}^{\min\{m_0, s_1\}} r_{s_2, s_1 - s_2 + 1}$, where $s_1 = 1, 2, \ldots, n_0$. Then the coarse cases discussed above and their causes can be summarised in Table 3.

Table 2: Part of warranty data

| Date of sales | Sold volume | Claims received | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | … | $m_0$ | $m_0 + 1$ | … | $n_0 - 1$ | $n_0$ |
| $D_1$ | $M_1$ | $r_{11}$ | $r_{12}$ | … | $r_{1,m_0}$ | $r_{1,m_0+1}$ | … | $r_{1,n_0-1}$ | $r_{1,n_0}$ |
| $D_2$ | $M_2$ | | $r_{21}$ | … | $r_{2,m_0-1}$ | $r_{2,m_0}$ | … | $r_{2,n_0-2}$ | $r_{2,n_0-1}$ |
| … | … | | | … | … | … | … | … | … |
| $D_{m_0-1}$ | $M_{m_0-1}$ | | | … | $r_{m_0-1,2}$ | $r_{m_0-1,3}$ | … | $r_{m_0-1,n_0-m_0+1}$ | $r_{m_0,n_0-m_0+2}$ |
| $D_{m_0}$ | $M_{m_0}$ | | | | $r_{m_0,1}$ | $r_{m_0,2}$ | … | $r_{m_0,n_0-m_0}$ | $r_{m_0,n_0-m_0+1}$ |
| Total | $M$ | $r_1$ | $r_2$ | … | $r_{m_0}$ | $r_{m_0+1}$ | … | $r_{n_0-1}$ | $r_{n_0}$ |

Table 3: Coarse warranty data and their causes.

| Coarse warranty data | Causes |
|---|---|

| | | |
|---|---|---|
| Aggregated data | Relating to age | individual $r_{uv}$ is not available, but $\sum_{v \in T_v} r_{uv}$ is obtainable for a fixed time interval $T_v$ |
| | Relating to claim dates | Individual $r_{uv}$ is not available, but $\sum_{u+v \in T_D} r_{uv}$ is obtainable for a fixed time interval $T_D$ |
| | Relating to sales dates | $M_u$ and $r_{uv}$ are not available, but $\sum_{u \in T_u} M_u$ and $\sum_{u \in T_u} r_{uv}$ are obtainable for a fixed time interval $T_u$ |
| Delayed data | Type I reporting delay | $r_{uv}$ are not temporarily available for $u + v > x$, ($x$ is a date), but they are available for $u + v \leq x$ |
| | Type II reporting delay | $r_{uv}$ may not include all of failed items at age $v$ and sold at date $u$; |
| | Type I sales delay | $D_i$ are only available for claimed items; |
| | Type II sales delay | $D_i$ are unavailable for failed or un-failed items; |
| Incomplete censored data | 1-dimensional | $D_i$ might not be available for un-failed items |
| | 2-dimensional | $r_{uv}$ might not include all of failures occurring under warranty limit |
| Other coarse data | Customer behaviour | un-failed items might be included in $r_{uv}$, but failed items might not be reported and not be included from $r_{uv}$ |
| | Unclean data | $D_i$ can be inaccurate, or some numbers counted in $r_{uv}$ might need to be counted in $r_{u'v'}$ |

## 4.2 Techniques used in analysing coarse warranty data

From the above review, one can find that three techniques have been most frequently used. These techniques are: non-homogeneous Poisson process, estimation for censored lifetime data, and estimation for truncated lifetime data.

- *Non-homogeneous Poisson process*. A widely used assumption is that the number of claims at age *t* and certain additional assumptions are independently distributed as Poisson distributions, see [24,27,37,79], for example. Here, the certain additional assumptions can be with a reporting lag *l* for cars putting into service on day *x*, for example.

- *Estimation for censored lifetime data*. Data on un-failed items might not be obtained and therefore are regarded as censored data. The main problem is that censoring time might not be available, as discussed above.

- *Estimation for truncated lifetime data*. Warranty data can also be dealt with based on the concept of truncated data. Truncated data arise when a variable is observable only if

it lies in some specified portion of its range. For observations that fall outside the certain range, the existence is not known (and this is what distinguishes truncation from censoring). It is important to distinguish between truncated data and censored data. Censoring occurs when there is a bound on an observation (lower bound for observations censored on the right, upper bound for observations censored on the left, and both upper and lower bounds for observations that are interval censored). Truncation, however, arises when even the existence of a potential observation would be unknown if its value lies in a certain range [80].

In Figure 1, the exact number $N_1$ of items might be unknown. That is, the exact number of items under warranty is not obtainable. This can be due to various reasons, for example, in the following scenarios, one can treat the warranty data as truncated data.

- o Sales delay, including type I and type II [45,75].
- o for items protected by two-dimensional warranty (e.g. time and usage limits), their usage limit might have expired but their calendar time limit is still under warranty. In this case, usage (or mileage for automobiles, for example) can only be observed when warranty claims are made. Warranty claims under this scenario has also be studied by [45,71,75].
- o Claims with type I reporting delay [39], for example.

## 4.3 Summary of the publications

Table 5 in Appendix summaries papers on coarse warranty data. In the table, the first column shows the authors; the second column includes the type of coarse data; the third column briefly summarises the purpose of the research in the paper.

Table 4 in Appendix is a legend table including the acronyms used in Table 5. Furthermore, Table 6 in Appendix lists the top four journals that publish CWD papers most frequently.

## 5. Conclusions and future research

This paper reviewed the existing work in coarse warranty data and analysis. It can be found that the major challenges facing warranty data analysts are poor data quality. As such, it has been mainly pursued to develop better improved techniques to solve the problems resulted from the poor data quality.

Below some interesting topics for further research are listed.

- *Human factors*. Human factors have seldom been considered in warranty claim analysis. However, product users might not be bothered to claim warranty, or some warranty claims might not valid (see [72], for example), analysing warranty claims data considering human factors can be interesting and important. This is especially true for those claims in which fault cannot be found, or known as NFF (no-fault-found). NFF contributes on average to 45% of reported service faults in electronic products [57]. The NFF problem is not new, but many believe it is getting worse, in part because today's highly complex products are equipped with more and more electronic sensors, computers, control functions and wires [81].

- *Warranty data collection*. The scenarios of incompleteness discussed in Section 2 might just be a few. In practice, there can be more cases of incompleteness deserving attention and research.

- *Long term warranty*. The existing research on warranty claims analysis 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 products, as discussed in a review paper [14]. 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.

## 6. Remarks

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. Our apologies are extended to both the researchers and readers if any relevant papers have been omitted.

### For further readings

Other issues on warranty data collection may be the problem caused by short claim history [7], warranty policy optimisation based on maintenance optimisation (see [2,6,82-84], for example), which are worth reading.

## Acknowledgement

## References

[1] Blischke WR, Murthy DNP. Product warranty management - I: A taxonomy for warranty policies. European Journal of Operational Research 1992;62(2):127-148.

[2] Wu S, Li H. Warranty cost analysis for products with a dormant state. European Journal of Operational Research 2007;182(3):1285-1293.

[3] Jung KM, Park M, Park DH. System maintenance cost dependent on life cycle under renewing warranty policy. Reliability Engineering and System Safety 2010;95(7):816-821.

[4] Shafiee M, Finkelstein M, Chukova S. On optimal upgrade level for used products under given cost structures. Reliability Engineering and System Safety 2011;96(2):286-291.

[5] Wu S, Longhurst P. Optimising age-replacement and extended non-renewing warranty policies in lifecycle costing. International Journal of Production Economics 2011;130(2):262-267.

[6] Park M, Mun Jung K, Park DH. Optimal post-warranty maintenance policy with repair time threshold for minimal repair. Reliability Engineering and System Safety 2013;111:147-153.

[7] Wu S, Akbarov A. Forecasting warranty claims for recently launched products. Reliability Engineering and System Safety 2012;106:160-164.

[8] Murthy DNP, Djamaludin I. New product warranty: A literature review. International Journal of Production Economics 2002;79(3):231-260.

[9] Djamaludin D, Murthy D, Blischke W. Bibliography on warranties. In: Blischke W, Murthy D, editors. Product Warranty Handbook, Marcel Dekker Inc; 1995, p. 839-916.

[10] Thomas MU, Rao SS. Warranty economic decision models: A summary and some suggested directions for future research. Operations Research 1999;47(6):807-820.

[11] Murthy DNP, Solem O, Roren T. Product warranty logistics: Issues and challenges. European Journal of Operational Research 2004;156(1):110-126.

[12] Karim MR, Suzuki K. Analysis of warranty claim data: A literature review. International Journal of Quality and Reliability Management 2005;22(7):667-686.

[13] Murthy DNP. Product warranty and reliability. Annals of Operations Research 2006;143(1):133-146.

[14] Rahman A, Chattopadhyay G. Review of long-term warranty policies. Asia-Pacific Journal of Operational Research 2006;23(4):453-472.

[15] Wu S. Warranty data analysis: A review. Quality and Reliability Engineering International 2012;28(8):795-805.

[16] Blischke WR, Murthy DNP. Warranty Cost Analysis. New York: CRC Press; 1994.

[17] Brennan JR. Warranties, Planning, Analysis and Implementation. New York: McGraw-Hill Professional; 1994.

[18] Murthy DNP, Blischke WR. Warranty Management and Product Manufacture. London: Springer; 2006.

[19] Sahin I, Polatoglu H. Quality, Warranty and Preventive Maintenance. Boston: Kluwer Academic Pub; 1998.

[20] Rai BK, Singh N. Reliability Analysis and Prediction with Warranty Data: Issues, Strategies, and Methods. London: Taylor & Francis Group; 2009.

[21] Thomas MU. Reliability and Warranties: Methods for Product Development and Quality Improvement. London: Taylor & Francis Group; 2006.

[22] Suzuki K. Nonparametric estimation of lifetime distribution form a record of failures and follow-ups. Journal of American Statistical Association 1985;80:68-72.

[23] Suzuki K. Estimation of lifetime parameters from incomplete field data. Technometrics 1985;27(3):263-271.

[24] Kalbfleisch JD, Lawless JF, Robinson JA. Methods for the analysis and prediction of warranty claims. Technometrics 1991;33(3):273-285.

[25] Lawless JF, Crowder MJ. Models and estimation for systems with recurrent events and usage processes. Lifetime Data Analysis 2010;16(4):547-570.

[26] Lawless JF. Statistical analysis of product warranty data. International Statistical Review 1998;66(1):41-60.

[27] Suzuki K, Karim MR, Wang L. Statistical analysis of reliability warranty data. In: Rao CR, Balakrishnana N, editors. Handbook of Statistics: Advances in Reliability, Amsterdam: Elsevier; 2001, p. 585-609.

[28] Alam M,M., Suzuki K. Lifetime estimation using only failure information from warranty database. IEEE Transactions on Reliability 2009;58(4):573-582.

[29] Lawless JF, Crowder MJ, Lee K-. Analysis of reliability and warranty claims in products with age and usage scales. Technometrics 2009;51(1):14-24.

[30] Rai BK. Warranty spend forecasting for subsystem failures influenced by calendar month seasonality. IEEE Transactions on Reliability 2009;58(4):649-657.

[31] Heitjan DF, Rubin DB. Ignorability and coarse data. Ann.Statist. 1991;19(4):2244-2253.

[32] Kalbfleisch JD, Lawless JF. Statistical analysis of warranty claims data. In: Blischke WR, Murthy DNP, editors. Product Warranty Handbook, New York: Marcel Dekker; 1996, p. 231-259.

[33] Suzuki K, Yamamoto W, Karim MR, Wang L. Data analysis based on warranty database. In: Limnios N, Nikulin M, editors. Recent Advances in Reliability Theory, Boston: Birkhauser; 2000, p. 213-227.

[34] Karim R, Yamamoto W, Suzuki K. Statistical Analysis of Marginal Count Failure Data. Lifetime Data Analysis 2001;7(2):173-186.

[35] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 1977;39:1-38.

[36] Lawless JF, Kalbfleisch JD. Some issues in the collection and analysis of field reliability data. In: Klein J, Goel P, editors. Survival Analysis: State of the Art, Dordrecht: Kluwer Academic; 1992, p. 141-152.

[37] 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(3):323-337.

[38] Lawless JF. Adjustments for reporting delays and the prediction of occurred but not reported events. Canadian Journal of Statistics 1994;22(1):15-31.

[39] Kalbfleisch, J, Lawless, J. Truncated data arising in warranty and field performance studies, and some useful statistical methods. 1991; RR-91-02. www.bisrg.uwaterloo.ca/archive/RR-91-02.pdf, accessed on 29th December 2012.

[40] Rai B, Singh N. Modeling and analysis of automobile warranty data in presence of bias due to customer-rush near warranty expiration limit. Reliability Engineering and System Safety 2004;86(1):83-94.

[41] Rai B, Singh N. Customer-rush near warranty expiration limit, and nonparametric hazard rate estimation from known mileage accumulation rates. IEEE Transactions on Reliability 2006;55(3):480-489.

[42] Robinson J, McDonald G. Issues related to field reliability and warranty data. In: Liepins G, Uppuluri V, editors. Data Quality Control: Theory and Pragmatics, New York: Marcel Dekker; 1991, p. 69-90.

[43] Hu XJ, Lawless JF, Suzuki K. Nonparametric estimation of a lifetime distribution when censoring times are missing. Technometrics 1998;40(1):3-13.

[44] Ion RA, Petkova VT, Peeters BHJ, Sander PC. Field reliability prediction in consumer electronics using warranty data. Quality and Reliability Engineering International 2007;23(4):401-414.

[45] Hu XJ, Lawless JF. Estimation of rate and mean functions from truncated recurrent event data. Journal of the American Statistical Association 1996;91(433):300-310.

[46] Karim MR, Suzuki K. Analysis of field failure warranty data with sales lag. Pakistan Journal of Statistics 2004;20(1):93-102.

[47] Karim MR. Modelling sales lag and reliability of an automobile component from warranty database. International Journal of Reliability and Safety 2008;2(3):234-247.

[48] Zhao K, Steffey D. Analysis of field performance using interval-censored incident data. Proceedings of Annual Reliability and Maintainability Symposium. 2009:43-46.

[49] 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.

[50] Baxter LA. Estimation from quasi life tables. Biometrika 1994;81(3):567-577.

[51] Tortorella M. Life estimation from pooled discrete renewal counts. In: Jewell NP, Kimber AC, Lee MLT, Whitmore GA, editors. Lifetime Data: Models in Reliability and Survival Analysis, Norwell: Springer; 1996, p. 331-338.

[52] Crowder M, Stephens D. On the analysis of quasi-life tables. Lifetime Data Analysis 2003;9(4):345-355.

[53] Akbarov A, Wu S. Warranty Claims Data Analysis Considering Sales Delay. Quality and Reliability Engineering International 2012; DOI: 10.1002/qre.1302, in press.

[54] Lim TJ. Nonparametric estimation of the product reliability from grouped warranty data with unknown start-up time. International Journal of Industrial Engineering: Theory Applications and Practice 2003;10(4):474-481.

[55] Wilson S, Joyce T, Lisay E. Reliability estimation from field return data. Lifetime Data Analysis 2009;15(3):397-410.

[56] Kalbfleisch JD, Lawless JF. Estimation of reliability in field-performance studies. Technometrics 1988;30(4):365-388.

[57] Jones J, Hayes J. Investigation of the occurrence of: No-faults-found in electronic equipment. IEEE Transactions on Reliability 2001;50(3):289-292.

[58] Singpurwalla ND, Wilson S. Warranty problem: its statistical and game theoretic aspects. SIAM Review 1993;35(1):17-42.

[59] Singpurwalla ND, Wilson SP. Failure models indexed by two scales. Advances in Applied Probability 1998;30(4):1058-1072.

[60] Phillips MJ, Sweeting TJ. Estimation from Censored Data with Incomplete Information. Lifetime Data Analysis 2001;7(3):279-288.

[61] Phillips MJ, Sweeting TJ. Estimation for censored exponential data when the censoring times are subject to error. Journal of the Royal Statistical Society. Series B (Methodological) 1996;58(4):775-783.

[62] Phillips MJ. Bayesian Estimation from Censored Data with Incomplete Information. Quality and Reliability Engineering International 2004;20(3):237-245.

[63] Yang SC, Nachlas JA. Bivariate reliability and availability modeling. IEEE Transactions on Reliability 2001;50(1):26-35.

[64] Moskowitz H, Chun YH. Poisson regression model for two-attribute warranty policies. Naval Research Logistics 1994;41(3):355-376.

[65] 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(4):488-502.

[66] Jung M, Bai DS. Analysis of field data under two-dimensional warranty. Reliability Engineering and System Safety 2007;92(2):135-143.

[67] Gertsbakh IB, Kordonsky KB. Parallel time scales and two-dimensional manufacturer and individual customer warranties. IIE Transactions (Institute of Industrial Engineers) 1998;30(12):1181-1189.

[68] Duchesne T, Lawless J. Alternative Time Scales and Failure Time Models. Lifetime Data Analysis 2000;6(2):157-179.

[69] 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(2):127-132.

[70] Iskandar BP, Blischke WR. Reliability and warranty analysis of a motorcycle based on claims data. In: Blischke W, Murthy D, editors. Case Studies in Reliability and Maintenance, Hoboken, NJ: John Wiley & Sons, Inc.; 2003, p. 623-656.

[71] Chukova S, Robinson J. Estimating mean cumulative functions from truncated automotive warranty data. In: Wilson A, Limnios N, Keller-McNulty S, Armijo Y, editors. Series on Quality Reliability and Engineering Statistics -10: Modern Statistical and Mathematical Methods in Reliability, Singapore: World Scientific; 2005, p. 121-136.

[72] Wu S. Warranty claim analysis considering human factors. Reliability Engineering and System Safety 2011;96(1):131-138.

[73] Patankar JG, Mitra A. Effects of warranty execution on warranty reserve costs. Management Science 1995;41(3):395-400.

[74] Patankar JG. Warranty and consumer behavior: Warranty execution. In: Blischke WR, Murthy DNP, editors. Product Warranty Handbook, New York: Marcel Dekker Inc; 1996, p. 421-438.

[75] Hu JX, Lawless JF. Estimation from truncated lifetime data with supplementary information on covariates and censoring times. Biometrika 1996;83(4):747-761.

[76] Hu XJ, Lawless JF. Pseudolikelihood estimation in a class of problems with response-related missing covariates. Canadian Journal of Statistics 1997;25(2):125-142.

[77] Karim MR, Suzuki K. Analysis of warranty data with covariates. Journal of Risk and Reliability 2007;221:249-255.

[78] Rai B, Singh N. Hazard rate estimation from incomplete and unclean warranty data. Reliability Engineering and System Safety 2003;81(1):79-92.

[79] 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.

[80] Meeker WQ, Escobar LA. Statistical Methods for Reliability Data. New York: John Wiley & Sons; 1998.

[81] Ramsey, J. Special Report: Avoiding NFF. *Avionics Today* 2005; .

[82] 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.

[83] Yang G. Warranty analysis. In: Yang G, editor. Life Cycle Reliability Engineering, Hoboken, New Jersey: John Wiley & Sons; 2007, p. 442-485.

[84] Wu S, Xie M. Warranty cost analysis for nonrepairable services products. International Journal of Systems Science 2008;39(3):279-288.

[85] Baik J, Prabhakar Murthy DN. 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.

[86] Davis T. A simple method for estimating the joint failure time and failure mileage distribution from automobile warranty data. Ford Technical Journal 1999;2(6):1-11.

[87] Eliashberg J, Singpurwalla ND, Wilson SP. Calculating the reserve for a time and usage indexed warranty. Management Science 1997;43(7):966-975.

[88] Kalbfleisch JD, Lawless JF. Some useful statistical methods for truncated data. Journal of Quality Technology 1992;24(3):145-152.

[89] Kalbfleisch JD, Lawless JF. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. Statistica Sinica 1991;1:19-32.

[90] Lawless J, Hu J, Cao J. Methods for the estimation of failure distributions and rates from automobile warranty data. Lifetime Data Analysis 1995;1(3):227-240.

[91] Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. Technometrics 1995;37:158-168.

[92] Majeske KD, Lynch-Caris T, Herrin G. Evaluating product and process design changes with warranty data. International Journal of Production Economics 1997;50(2-3):79-89.

[93] Majeske KD. A non-homogeneous Poisson process predictive model for automobile warranty claims. Reliability Engineering and System Safety 2007;92(2):243-251.

[94] Oh YS, Bai DS. Field data analyses with additional after-warranty failure data. Reliability Engineering and System Safety 2001;72(1):1-8.

[95] Yang G, Zaghati Z. Two-dimensional reliability modeling from warranty data. Proceedings of the Annual Reliability and Maintainability Symposium. 2002:272-278.

[96] Yun KW, Kalivoda FE. Model for an estimation of the product warranty return rate. Proceedings of the Annual Reliability and Maintainability Symposium. 1977:31-37.

# Appendix

Table 4: Legend

| | |
|---|---|
| A-RA | aggregated data---relating to age |
| A-RC | aggregated data---relating to claim dates |
| A-RS | aggregated data---relating to sales dates |
| D-RD | delayed data---reporting delay |
| D-SD | delayed data---sales delay |
| I-WD | incomplete censored data--- using warranty claims data only |
| I-SD | incomplete censored data--- using both warranty claims data and supplementary data |
| O-CB | other coarse data scenario---customer behaviour |
| O-MC | other coarse data scenario ---missing covariance |
| O-VD | other coarse data scenario ---vague data |

Table 5. Existing publications (listed in alphabetical order of first authors' surnames)

| (A) | (B) | (C) |
|---|---|---|
| Alam & Suzuki [28] | I-WD | Dealing with the problem that censored data are unavailable |
| Baik & Murthy [85] | I-WD | Estimating reliability using an accelerated life test model |
| Chukova & Robinson [71] | I-WD | Estimating the mean cumulative number of claims with a nonparametric approach |
| Davis [86] | I-WS | estimating the joint distribution with a 1-D conditional approach |
| Duchesne & Lawless [68] | I-WS | Dealing with 2-D warranty data with the time scale approach |
| Eliashberg et al. [87] | I-WS | dealing with warranty reserve problems with a 1-D conditional approach |
| Gertsbakh & Kordonsky [67] | I-WS | Dealing with 2-D warranty data with the time scale approach |
| Hu & Lawless [45] | I-SD, A-RS, A-RC | estimating rate and mean functions from truncated recurrent event data |
| Hu & Lawless [75] | I-SD | Proposing approaches to obtaining supplementary data and developing estimation approaches to lifetime distributions |
| Hu & Lawless [76] | I-SD O-MC | proposing pseudo-likelihood estimates in a class of problems with response-related missing covariates |
| Hu, et al. [43] | D-SD | Proposing a nonparametric approach to estimating the lifetime distribution |
| Ion et al. [44] | D-SD | Assuming to use the distribution of the sales- delay for a previous generation of products |
| Iskandar & Blischke [70] | I-SD | using two approaches to estimating claims: 1-D conditional approach and time scale approach |
| Jung & Bai [66] | I-WD | Directly estimating two-dimensional lifetime distribution |
| Kalbfleisch & Lawless [56] | I-SD | Estimating product reliability with particular attention on the estimation of regression coefficients in parametric models. |
| Kalbfleisch, et al. [24] | D-RD, A-RA | Estimating and predicting the number of warranty claims considering reporting delay and aggregated claims |

| | | |
|---|---|---|
| Kalbfleisch & Lawless [39] | D-RD | Treating claims with reporting delay as truncated data |
| Kalbfleisch & Lawless [88,89] | D-RD | Treating claims with reporting delay as truncated data |
| Kalbfleisch & Lawless [32] | A-RA, A-RS, D-RD | developing methods dealing with reporting delay, type I and type III aggregated claims |
| Karim [34] | A-RC | Developing a Poisson model and using the EM algorithm to estimate the probability of failure for repairable and nonrepairable items, respectively |
| Karim & Suzuki [46] | D-SD | estimating the number of claims with the EM algorithm when sales delay exists |
| Karim & Suzuki [77] | I-WD O-MC | using covariates in a Weibull regression model for estimating the lifetime distribution |
| Karim [47] | D-SD | estimate the lifetime distribution when sales delay is considered |
| Lawless [36] | D-RD, A-RS | Nonparametric approaches to estimating the expected warranty claims when reporting delay exists and claims may be aggregated |
| Lawless [38] | D-RD | Development of an approach that uses recent reporting data and incorporates random effect |
| Lawless et al. [90] | I-SD | estimating the two-dimensional joint distribution based on a 1-D conditional approach |
| Lawless & Nadeau [91] | D-RD | introducing an approach to estimating the cumulative mean functions |
| Lawless [26] | D-RD, A-RA, A-RS | A review paper |
| Lawless, et al. [29] | I-WD, I-SD | Using a 1-D conditional approach and assessing the dependence on age or usage in heterogeneous populations of products |
| Lawless et al. [25] | I-SD | estimating bivariate joint distributions |
| Lim [54] | D-SD | warranty estimation considering sales delay |
| Majeske, et al. [92] | D-SD | Considering sales delay in modelling |
| Majeske [93] | D-SD | develop a NHPP model to predict automobile warranty claims |
| Mohan [49] | A-RC | Using the EM algorithm to estimate a lifetime distribution |
| Moskowitz & Chun [64] | I-WD | using a bivariate Poisson model to predict claims for a two dimensional warranty |
| Oh & Bai [94] | I-SD | estimating the lifetime distribution when additional field data are available |
| Phillips & Sweeting [61] | I-SD | estimating the exponential distribution with incomplete censored data |
| Phillips & Sweeting [60] | I-SD | estimating the exponential distribution with incomplete censored data |
| Phillips [62] | I-SD | estimating the exponential distribution with incomplete censored data |
| Rai & Singh [78] | I-SD, O-VD | estimating the hazard rate with incomplete, and biased warranty claims |
| Rai & Singh [40] | O-CB | estimating hazard rate considering customer behaviour |
| Rai & Singh [41] | O-CB | estimate hazard rate consdering customer behaviour |
| Singpurwalla & Wilson [58] | I-WD | Using a 1-D conditional approach to estimating the lifetime distribution |
| Singpurwalla & Wilson [59] | I-WD | Estimating the lifetime distribution using, and using the Gamma process and Poisson processes to describe usage |
| Suzuki [22] | I-SD | Estimating the lifetime distribution using a nonparametric method with follow-up information |
| Suzuki [23] | I-SD | Estimating the lifetime distribution using a parametric method with follow-up information |

| Suzuki, et al. [33] | A-RS, A-RC | Using NHPP to estimate warranty claims based on type II and type III aggregated claims, respectively; and estimating lifetime distributions based on type III aggregated claims |
|---|---|---|
| Suzuki, et al. [27] | A-RS, A-RC, I-SD | Using NHPP to estimate warranty claims based on type II and type III aggregated claims, respectively; and reviewing nonparametric and parametric approaches to estimating lifetime distributions |
| Wang et al. [37] | A-RS | estimating the number of claims for both repairable and non-repairable items based on type III aggregated claims |
| Wilson, et al. [55] | D-RD, D-SD | estimating the survivor distributions considering sales and reporting delays with Bayesian inference |
| Yang & Nachlas [63] | I-WD | constructing a general framework for specifying bivariate longevity models, 1-D conditional approach based. |
| Yang & Zaghati [95] | I-WD | presenting a sequential regression method to model mileage accumulation |
| Yun & Kalivoda [96] | O-CB | estimating warranty return rate considering human factors |
| Zhao & Steffey [48] | D-SD | using interval-censored incident data to deal with sales delay |

(A)= Author(s); (B) = type of coarse warranty data; (C) = one sentence summary

Table 6: The top four journals in which papers on CWD were most frequently published

| Journals | Numbers of papers published in the journal |
|---|---|
| Lifetime Data Analysis | 7 |
| Reliability Engineering and System Safety | 6 |
| Technometrics | 6 |
| IEEE Transactions on Reliability | 4 |