Reliability Analysis of Maintenance Data for Medical Devices

Sharareh Taghipour, Dragan Banjevic, Andrew K.S. Jardine

University of Toronto, Toronto, ON M5S 3G8, CANADA

Abstract

This paper proposes a method to analyze statistically maintenance data for complex medical devices with censoring and missing information. It presents a classification of different types of failures and establishes policies for analyzing data at system and component levels taking into account the failure types. The results of this analysis can be used as basic assumptions in development of a maintenance/inspection optimization model. As a case study, we present the reliability analysis of a general infusion pump from a hospital.

Keywords: reliability; analysis; tend test; medical devices; maintenance data; inspection; work orders

1. Introduction

A common belief of a healthcare society is that medical devices fail independent of age, following exponential distribution. This belief originates from the general conviction that electronic equipment has a constant failure rate. Even the most widely known and used reliability prediction handbook, MIL-HDBK-217 [1], proposes reliability models whose construction is based on the constant failure rate assumption. This assumption has been criticized as inaccurate, however, and its use may result in erroneous decisions [2, 3, and 4]. Reliability and failure patterns of a device may be affected by external factors such as operating conditions, environmental stress, expertise level of operators, etc. Therefore, to realistically determine their reliability, devices should be studied in their operating context. To this end, learning about existing maintenance procedures of equipment and statistical analysis of field data are essential steps towards developing an optimal evidence-based maintenance/inspection plan.

Although a number of researchers have considered the reliability prediction of medical systems at the design and development stage over the last two decade [5, 6], less literature deals with the reliability prediction of medical devices while they are in use in hospitals. Ion et al. [7] analyze the field data for medical imaging systems during the warranty period. Roelfsema [8] presents the results of early reliability prediction for Philips medical systems based on field data.

In this study, we analyzed maintenance data (collected from 2000 to 2007) from a Canadian General hospital. We conducted statistical analysis to learn about the failure types and trends of a particular general infusion pump used to deliver liquids for therapeutic and/or diagnostic purposes. However, since all medical devices undergo periodical safety and performance inspections, the general policies derived from this study can be applied to other medical devices.

Although dealing with a specific pump, this paper describes how available maintenance data in a hospital can be statistically analyzed. In Section 1, we classify different types of failure and establish policies for analyzing the data at three different levels: system level, failure type, and component level; the results of the analysis can be used as a foundation for a maintenance/inspection optimization model. Section 2 explains general maintenance and inspection procedures employed by hospitals. Analysis of failure data, the proposed policy to treat the failure data, and the results of the analysis appear in Section 3. The final section presents our conclusions.

2. Maintenance/inspection procedures for medical devices

2.1. Major inspection and routine tests

Medical devices are complex repairable systems consisting of a large number of interacting components which perform a system's required functions. A repairable system, upon failure, can be restored to satisfactory performance by any method except replacement of the entire system [9]. Medical devices usually undergo several types of tests/inspections during their life cycles as described here [10]:

- *Acceptance Test*

A series of qualitative and quantitative tasks designed to verify the safety and performance of newly received equipment, as well as conformity to applicable codes, regulations and standards.

- *Operational Check*

Visual and operational check of the equipment's safety and functionality, typically performed at the beginning of the day or work period, or just before using equipment on a patient.

- *Safety and Performance Inspection (SPI)*

A set of qualitative and quantitative tasks designed to verify the safety and performance of each piece of equipment by detecting potential and hidden failures and taking appropriate actions.

After accomplishing the acceptance test for a newly received device, SPIs are scheduled to be performed periodically. If any problem is found at inspection, corrective actions are taken to restore the device or its defective parts to an acceptable level. In addition, a set of failure preventive actions may be taken to prevent future failures and/or restore device function; these include part replacement, calibration, lubrication, etc. to address age or usage related deterioration.

When a device fails while it is in use, the problem is reported by the operator, and again appropriate actions (corrective maintenance) are taken. When the repair of a device is no longer technically feasible or cost effective, replacement becomes the best or the only option [10]. Fig. 1 describes major tests and actions performed during a device's life cycle.

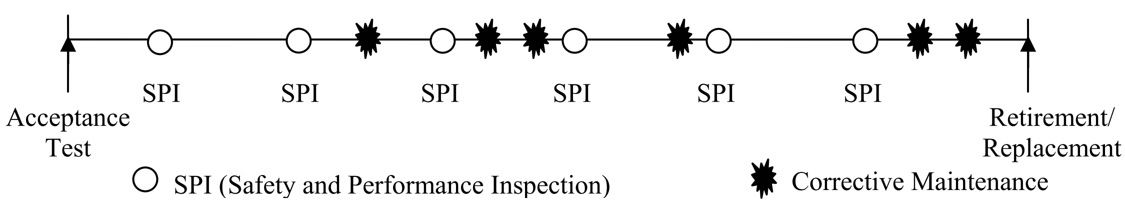


Figure 1. Major tests and actions performed during a device's life cycle

Currently, most hospitals merely follow manufacturers' recommended intervals for periodic SPI of devices. SPI intervals differ from 6 to 12 months depending on the device type and risk level

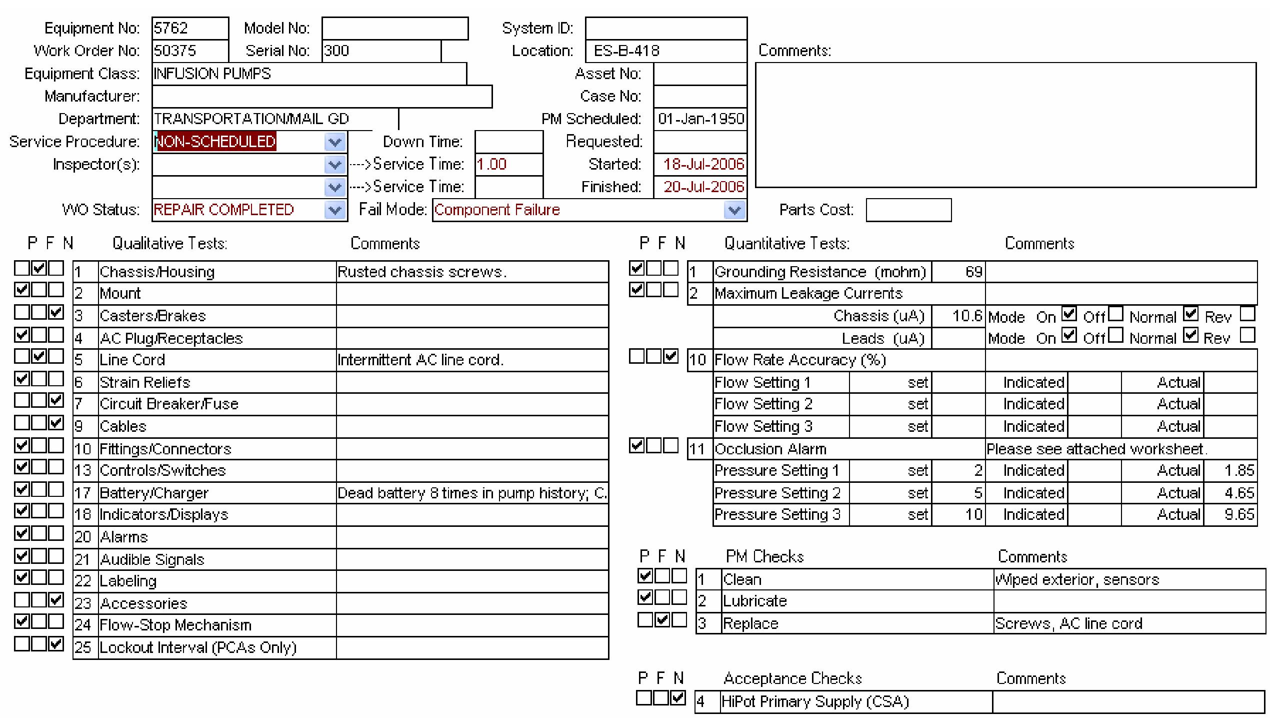
[11]. Class III (high risk) devices such as defibrillators should be inspected every 6 months, and class II (medium risk) devices like ECGs should be inspected annually. However, the optimality and even the necessity of these recommended intervals are questionable. It is essential to establish an evidence-based inspection or maintenance regimen derived from analysis of field data.

2.2. Scheduled and non-scheduled work orders

The maintenance and inspection data are usually available in the computerized maintenance management system (CMMS) of a hospital, stored in either scheduled or non-scheduled work orders. Scheduled work orders are used for routine tests (SPIs); however, when a device fails or has a defective part, a non-scheduled work order is requested to fix the problem. Both scheduled and non-scheduled work orders include the basic information of a device and a test checklist designed for a particular class of device. The checklist contains qualitative and quantitative tests; technicians or clinical engineers should use this list to ensure that all necessary tests and checks are accomplished. Fig. 2 shows a sample of a work order created for corrective maintenance (non-scheduled work order).

Qualitative tests mainly consist of visual inspection of the main parts/components of a device. For a general infusion pump, these include testing its chassis/housing, line cord, battery/charger, etc. Quantitative tests include measuring parameters of a device to check whether or not the parameters are in control. Grounding resistance and maximum leakage currents are among the quantitative tests for a general infusion pump.

A work order presents all PM checks and actions, such as cleaning, lubricating or replacement of a device or its parts. A work order is also created for an acceptance test of a newly received device.



Equipment No: 5762 | Model No: | System ID:
Work Order No: 50375 | Serial No: 300 | Location: ES-B-418
Equipment Class: INFUSION PUMPS | Asset No:
Manufacturer: | Case No:
Department: TRANSPORTATION/MAIL GD | PM Scheduled: 01-Jan-1950
Service Procedure: NON-SCHEDULED | Down Time: | Requested:
Inspector(s): | --->Service Time: 1.00 | Started: 18-Jul-2006
--->Service Time: | Finished: 20-Jul-2006
WO Status: REPAIR COMPLETED | Fail Mode: Component Failure
Comments:
Parts Cost:

P F N		Qualitative Tests:	Comments
F	1	Chassis/Housing	Rusted chassis screws.
P	2	Mount	
N	3	Casters/Brakes	
P	4	AC Plug/Receptacles	
F	5	Line Cord	Intermittent AC line cord.
P	6	Strain Reliefs	
N	7	Circuit Breaker/Fuse	
N	9	Cables	
P	10	Fittings/Connectors	
P	13	Controls/Switches	
P	17	Battery/Charger	Dead battery 8 times in pump history; C.
P	18	Indicators/Displays	
P	20	Alarms	
P	21	Audible Signals	
P	22	Labeling	
N	23	Accessories	
P	24	Flow-Stop Mechanism	
N	25	Lockout Interval (PCAs Only)	

P F N		Quantitative Tests:		Comments
P	1	Grounding Resistance (mohm)	69	
P	2	Maximum Leakage Currents		
		Chassis (uA)	10.6	Mode On ☑ Off ☐ Normal ☑ Rev ☐
		Leads (uA)		Mode On ☑ Off ☐ Normal ☑ Rev ☐
N	10	Flow Rate Accuracy (%)		
		Flow Setting 1	set	Indicated Actual
		Flow Setting 2	set	Indicated Actual
		Flow Setting 3	set	Indicated Actual
P	11	Occlusion Alarm		Please see attached worksheet.
		Pressure Setting 1	set 2	Indicated Actual 1.85
		Pressure Setting 2	set 5	Indicated Actual 4.65
		Pressure Setting 3	set 10	Indicated Actual 9.65

P F N		PM Checks	Comments
P	1	Clean	Wiped exterior, sensors
P	2	Lubricate	
F	3	Replace	Screws, AC line cord

P F N		Acceptance Checks	Comments
N	4	HiPot Primary Supply (CSA)	

Figure 2. A typical work order

"P" (pass), "F" (failed), or "N" (not applicable) are the possible results of each qualitative or quantitative test on the work order. When a test is performed on a component, and it is found to be non-defective, the result is "P"; however, in the case of failure of a part, the result is shown as "F". Since a general test checklist is designed for a class of equipment, some qualitative or quantitative tests may not be applicable to particular devices in that class; for these devices, the test result is given as "N". Quantitative and qualitative tests are specifically performed for components/features of a device, but PM checks show the actions performed at the device level.

3. Analysis of failure data

3.1. Preliminary analysis of failure data

For the purposes of this study, we selected a general infusion pump. General-purpose infusion pumps (Fig. 3) are used to accurately deliver liquids through intravenous or epidural routes for therapeutic and/or diagnostic purposes [12]. The reason for selecting this model is its significant number of non-scheduled work orders.

Figure 3. General Infusion Pump

The pilot hospital has 681 pieces of this model used in three departments. All scheduled and non-scheduled work orders were extracted from the CMMS, and we performed a visual examination of the data looking for possible errors.

Finding at the inspection	Non-scheduled work orders	Scheduled work orders
Device Contaminated	1	0
Component Failure	910	1182
User Input/Setup Error	3	1
Problem Not Found	275	316
Not Applicable	0	2
Device over used	1	**0**
NULL (the cause was not specified)	6	18
Total	**1196**	**1519**

Table 1: The total number of scheduled and non-scheduled work orders and findings at inspection

Table 1 shows the total number of scheduled and non-scheduled work orders and the associated findings at inspection. Minor defects are not usually reported by the operators and are rectified only at SPIs. Moreover, clinical engineers perform opportunistic maintenance of some components like batteries which may be replaced before complete depletion. Therefore, as Table 1 makes clear, more failed components are identified in scheduled work orders than in non-scheduled work orders. The average time between non-scheduled work orders is 452.8 days, with a standard deviation of 423.1 days. However, 661.6 days is the average time between scheduled work orders with relatively small variation (a standard deviation of 175.3 days).

We counted the percentage of times that a test resulted in "P", "F" and "N/A" for different components/features of the devices. Tables 2-4 summarize the results. Components/features with zero or few failures or with mostly "N/A" tests were excluded from the study.

The Pareto analysis of the qualitative test results of non-scheduled work orders identified the most problematic components (Fig. 4). "chassis/housing", "alarms", and "fittings/connectors" are the most problematic components/features with the total number of failures being 502, 321, and 240, respectively.

Non-scheduled work orders				Scheduled work orders			
Component	**P (%)**	**F (%)**	**N/A (%)**	**Component**	**P (%)**	**F (%)**	**N/A (%)**
AC Plug/Receptacles	98.92	1.08	0.00	AC Plug/Receptacles	90.75	0.54	8.71
Alarms	73.29	26.71	0.00	Alarms	76.20	15.08	8.71
Audible Signals	91.93	8.07	0.00	Audible Signals	75.12	16.17	8.71
Battery/Charger	97.92	2.08	0.00	Battery/Charger	79.39	11.90	8.71
Chassis/Housing	58.24	41.76	0.00	Chassis/Housing	37.26	54.03	8.71
Controls/Switches	91.26	8.74	0.00	Controls/Switches	86.12	5.17	8.71
Fittings/Connectors	80.03	19.97	0.00	Fittings/Connectors	64.48	26.80	8.71
Indicators/Displays	91.01	8.99	0.00	Indicators/Displays	83.11	8.17	8.71
Labeling	90.68	9.32	0.00	Labeling	63.88	27.34	8.77
Line Cord	88.77	11.23	0.00	Line Cord	67.13	24.16	8.71
Mount	83.94	15.89	0.17	Mount	79.27	12.02	8.71
Circuit Breaker/Fuse	3.99	1.08	94.93	Circuit Breaker/Fuse	1.98	0.18	97.84
Strain Reliefs	98.34	1.50	0.17	Strain Reliefs	88.52	2.46	9.01

Table 2. Qualitative tests results extracted from scheduled and non-scheduled work orders

Non-scheduled work orders				Scheduled work orders			
Component	**P (%)**	**F (%)**	**N/A (%)**	**Component**	**P (%)**	**F (%)**	**N/A (%)**
Flow Rate Accuracy (%)	12.90	0.00	87.10	Flow Rate Accuracy (%)	90.50	0.00	9.50
Grounding Resistance (mohm)	99.75	0.17	0.08	Grounding Resistance (mohm)	91.17	0.12	8.71
Maximum Leakage Currents	100.00	0.00	0.00	Maximum Leakage Currents	91.29	0.00	8.71
Occlusion Alarm	99.08	0.75	0.17	Occlusion Alarm	90.56	0.72	8.71

Table 3. Quantitative tests results extracted from scheduled and non-scheduled work orders

Non-scheduled work orders				Scheduled work orders			
PM Check	**P (%)**	**F (%)**	**N/A (%)**	**PM Check**	**P (%)**	**F (%)**	**N/A (%)**
Clean	99.75	0.25	0.00	Clean	91.29	0.00	8.71
Lubricate	99.67	0.33	0.00	Lubricate	91.11	0.18	8.71
Replace	29.37	70.63	0.00	Replace	13.04	78.19	8.77

Table 4. PM checks results extracted from scheduled and non-scheduled work orders

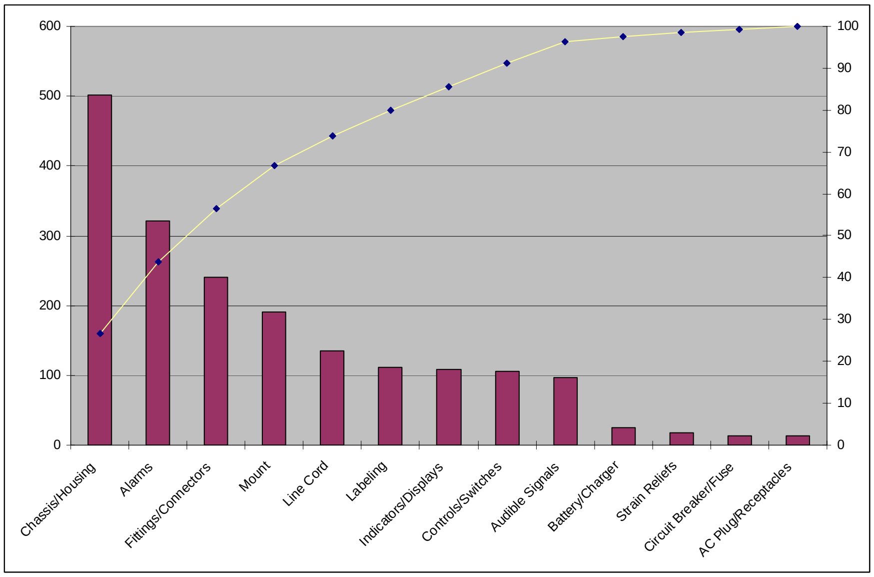


Figure 4. Pareto analysis for qualitative tests results

3.2. Hard and soft failures

Failures can be classified into two broad categories: soft and hard failures. Soft failures are the gradual loss of performance of a product, while hard failures cause it to stop working [13]. However, in this paper, “hard” failures signify those failures where the user is notified as soon as they occur. Hard failures either directly influence functioning of a device or are self-announcing. An example is a faulty alarm; as soon as the problem occurs, users are notified that something is wrong with the device, and it is sent for repair. We can assume that a hard failure occurs just before a SPI or corrective maintenance; therefore, the work order’s starting time can be considered as the failure time (complete data). Arguably, soft failures are identified at SPI when operators have postponed reporting them.

“Soft” failures are minor defects that have no or little influence on the device’s functioning. These defects such as “chassis/housing” defects may occur a long time before a scheduled or non-scheduled work order’s starting time, but the device still can function properly. The problem is not reported by the user and is detected only at the next SPI or during corrective maintenance triggered by a hard failure. Therefore, there is a time delay between the real occurrence of a soft failure and its detection. Ignoring this time is misleading; it should be considered as left or interval censoring time.

While we can interpret soft failures as the result of system deterioration over time, hard failures are mostly unpredictable and random in nature.

Interval censoring takes place when a minor failure or defect happens between a “no failure found” inspection and a “failure detected” inspection. If a defect occurs between the acceptance date and the first corrective maintenance or SPI, interval censoring is reduced to left censoring in which the starting time for the interval is zero.

Although there is not always a clear distinction between soft and hard failures, we used them to classify the components/features of a system. We categorized components/features in two groups depending on whether their associated failures were soft or hard failures and analyzed the failure data separately for each group. We classified components/features according to their failure percentage at scheduled and non-scheduled work orders. Classification of soft and hard failures can be found in Appendix A. This classification has been approved from a practical point of view by clinical engineers involved in this study.

In order to check the validity of our treatment with soft and hard failures, we calculated the number of hard and soft failures detected at each single scheduled and non-scheduled work order. Table 5 shows the results. The significant percentage of single hard failure detected in non-scheduled work orders emphasizes that it is correct to assume all hard failures' times as complete data. The close values obtained for one and two detected soft failures at both scheduled and non-scheduled work orders show that we can treat soft failures as left or interval censored data.

Failures frequency (%) in work orders			
	No	SPI	NS
Soft failures	1	33.92	57.55
	2	34.39	31.87
	3	22.13	8.24
	4	7.79	2.34
	5	1.77	
Hard failures	1	79.72	83.72
	2	18.88	14.88
	3	1.22	1.24
	4	0.17	0.16

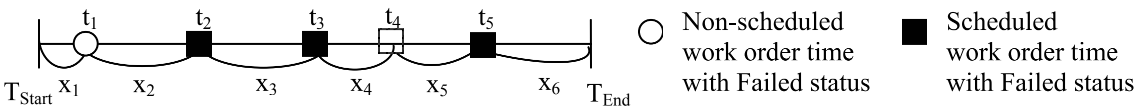
Table 5. Soft and hard failure frequencies at scheduled and non-scheduled work orders

3.3. Proposed policy for analyzing soft and hard failures

Because of the large number of censored data and combination of hard and soft failures, we propose separate policies for analyzing failure data at different levels, including system level, components corresponding to hard failures, and components corresponding to soft failures.

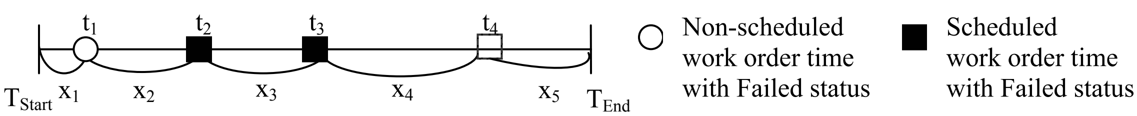
3.3.1. System level

All non-scheduled work orders' starting times should be considered as a system's exact failure times. The time between SPIs having at least one failure identified should be considered as either left or interval censored. The right censored data is specified according to the result of the last inspection (failure or pass), and its distance to the end of the test. Figures 5.a and 5.b illustrate the proposed policy, and the policy rules are explained in detail in Appendix B.



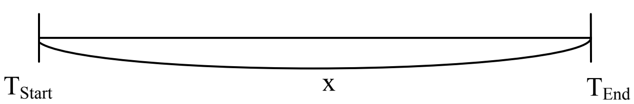
Event Times	WO Type	Data Type
x_1	NS	Complete
x_2	PM	Left Censored
x_3	PM	Left Censored
x_4		Pass
$(x_4, x_4+x_5]$	PM	Interval Censored
x_6		Right Censored (if x_6 <= 0. 5*MTBF)
0. 5*MTBF		Right Censored (if x_6 > 0. 5*MTBF)

MTBF= Mean time between failures



Event Times	WO Type	Data Type
x_1	NS	Complete
x_2	PM	Left Censored
x_3	PM	Left Censored
x_4		Pass
x_4+x_5		Right Censored (if x_5 <= 0. 5*MTBF)
x_4+0.5*MTBF		Right Censored (if x_5 > 0.5*MTBF)

Figure 5.a. The policy for analyzing failures at the system level



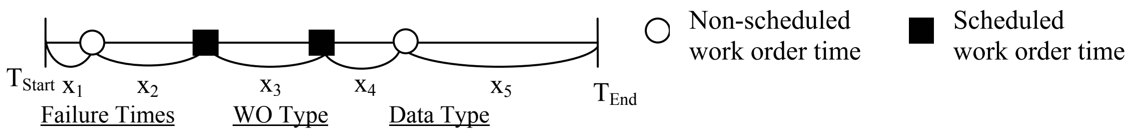
Event Times	WO Type	Data Type
x		Right Censored (if x <= 0.5*MTBF)
0.5*MTBF		Right Censored (if x > 0.5*MTBF)

Figure 5.b. The policy for dealing with "no failure" devices at the system level

When the time between the last inspection or acceptance test (when there is no failure) and the end of the test is long, particularly for soft failures, there are two possibilities: 1. a soft failure may have occurred somewhere between the last inspection and the end of the test; 2. no failure has occurred in the meantime. Since there is no information available regarding these two possible scenarios, and as it is expected that a failure occurs sometime between the average (or proportion of) times between failures and the end of the test, we consider a proportion of that time as the right censoring time. In order to be consistent, we consider (0.5*mean time between failures) as right censored data. We conducted different analysis with different values (0.5, 0.75 and 1) as the ratio to decide which proportion of mean time between failures should be considered as the right censoring rule. Even though we may have several failures in that interval, we only use an approximation to the first event.

3.3.2. Component level – hard and soft failures

For the components corresponding to hard failures, work orders' starting times are considered as a component's exact failure times if they report a failure of that component. The time interval between the last failure of the component and the end of the test is considered as right censored data.



Failure Times	WO Type	Data Type
x_1	NS	Complete
x_2	PM	Complete
x_3	PM	Complete
x_4	NS	Complete
x_5		Right Censored

Figure 6. The policy for dealing with hard failures – component level

For soft failures, all rules are similar to the system level's rules, except that we never have complete data. If there is no inspection with "no problem found" between two consecutive failures, the time interval between these two failures is considered as left censored data.

3.4. Trend analysis – the Laplace trend test

The "Laplace trend test", the "Reverse Arrangement test" or the "Military Handbook test" [14] can be used to quantitatively determine whether the failure times exhibit increasing or decreasing trend. In this paper we use the Laplace trend test to check whether the failures of the infusion pump system or its components exhibit any type of trend, that is, to check whether preventive or corrective maintenance has any influence on the system or its components. We will first analyze the failure data for trend at the system level, then trend in hard and soft failures, and finally trend in failures of individual components.

The Laplace trend test can be used for multiple repairable systems to check between no trend and non-homogenous Poisson process (NHPP) exponential models. Assume that there are m similar and independent systems. Let T_{ij} represent the time of the i^{th} failure of the j^{th} system, $i = 1,2,...,n_j$, $j = 1,2,...,m$, observed over the period $[T_{j_{start}}, T_{j_{end}}]$, where $T_{n_j} \le T_{j_{end}}$. Let:

$$\hat{n}_j = \begin{cases} n_j & \text{if the process is time truncated} \\ n_j - 1 & \text{if the process is failure truncated} \end{cases}$$

The trend test can then be conducted using the following formula:

$$LA = \frac{\sum_{j=1}^{m}\sum_{i=1}^{\hat{n}_j} T_{ij} - \sum_{j=1}^{m} \frac{1}{2}\hat{n}_j (T_{j_{start}} + T_{j_{end}})}{\sqrt{\frac{1}{12}\sum_{j=1}^{m} \hat{n}_j (T_{j_{end}} - T_{j_{start}})^2}} \quad . \tag{1}$$

The value of LA should be compared to high (for improvement) or low (for degradation) percentiles of the standard normal distribution. The Laplace trend test compares the average of failure arrival times to the midpoint of the observed time interval. There is a trend when the average of the failure arrival times deviates from the midpoint of the observation interval. In the Laplace trend test, when $n_j \geq 3$, the approximation with the Normal distribution generally provides good results.

For a trend test at a significant level of α, we have: If $LA > z_{\alpha/2}$ or $LA < z_{-\alpha/2}$, then there is a trend, with direction of the trend [15] indicated by the following:

If $LA < z_{-\alpha/2}$, the process is improving, so the times between failures are increasing.
If $LA > z_{\alpha/2}$, the process is deteriorating, so the times between failures are decreasing.

3.5. Trend analysis and distribution fitting of failure data

We analyze the failures data based on the proposed policy described in section 3.3. We use the Laplace trend test for trend analysis of failure data, taking the mid point of censoring intervals as an approximation of the actual failure times. Even though this is an approximation and therefore not precise, we opt to use it because no available trend test analysis in the literature incorporates interval censored data (except right censored).

The mean cumulative function (MCF) and confidence limits are plotted versus system age in days using the RELIABILITY procedure in SAS. The MCF plot describes average behavior of a system or multiple systems under study. It is constructed incrementally at each failure event by considering the number of systems at risk at that point in time. The number of systems at risk depends on how many systems are contributing information at that particular point in time. Information can be obscured by the presence of censoring and truncation [16]. The RELIABILITY procedure in SAS provides a nonparametric estimate and plot of the MCF for the number or cost of repairs for a population of repairable systems.

The nonparametric estimate of the MCF, the variance of the MCF estimate, and confidence limits for the MCF estimate are based on Nelson [17]. MCF, usually denoted by $M(t)$, is defined as the mean of the cumulative number of events up to time t. The method does not assume any underlying structure for the repair process [18].

3.5.1. Analysis of failures data at the system level

As shown in Fig. 7, the MCF plot at the system level is not a straight line. Its initial concave appearance is followed by a stable period, and a convex curve appears at the end. Therefore, the assumption of homogeneous Poisson process for failure data is not valid. The end-of-history ages are plotted in an area at the top.

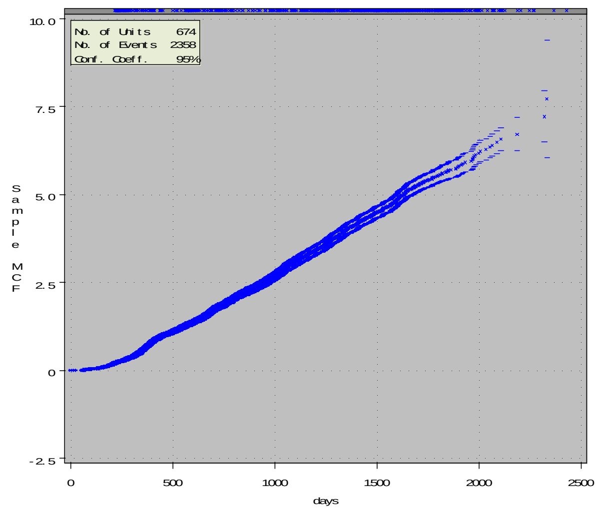


Figure 7. The MCF plot for the failures at the system level-0.5*MTBF

Using equation 1, the Laplace trend test value for the failure data is LA=12.59. At the significant level $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, the systems show a degrading trend.

Since the failure times are not independently and identically distributed, we first classify the times between two consecutive events (n-1th, nth) of each system. An event can be either a failure or a censoring event. The nth event of a system is censoring if at inspection no failure is found; otherwise, it is a failure event. We then pool the times between n-1th and nth events of all systems and conduct the Weibull analysis for each category separately. For example, for calculating the times to the first event, all devices which have at least one failure during their life are identified. For these devices, the time to the first failure is extracted with "failure" as an event. For devices with no failure at all, the first and only event is a censoring event with the censoring time from 0 to the last updating time of the CMMS database. Finally, the Weibull analysis is performed for the times between n-1th and nth events using LIFEEG procedure in SAS. Tables 6 and 7 summarize the Weibull parameters estimated for all times between events.

The results of the Weibull distribution fitted to the times between n-1th and nth event (failure/censoring)								
N	No of observations	No of non-censored	No of right censors	No of left censors	No of interval censors	β	η	$\mu = \eta\Gamma(\frac{1}{\beta}+1)$
1	674	237	28	348	61	1.5680	396.9852	356.6204
2	646	215	111	283	37	1.5175	318.9879	287.5696
3	536	177	100	222	37	1.3499	306.6525	281.2007
4	433	143	116	162	12	1.5622	311.9259	280.3152
5	318	101	115	96	6	1.3042	301.7416	278.5013
6	203	48	114	40	1	1.1024	381.6272	367.9733
7,8	129	34	58	35	2	0.9718	247.9506	251.0791
≥ 9	89	41	30	15	3	0.8472	156.9852	171.1464

Table 6. Weibull distribution parameters estimated for times between events – system level

n	β			η			$\mu = \eta\Gamma(\frac{1}{\beta}+1)$		
	estimate	**95% confidence limits**		**estimate**	**95% confidence limits**		**estimate**	**95% confidence limits**	
1	1.5680	1.4414	1.7058	396.9852	372.1665	423.4590	356.6204	333.3963	379.8445
2	1.5175	1.3867	1.6606	318.9879	297.9330	341.5308	287.5696	267.7068	307.4324
3	1.3499	1.2165	1.4868	306.6525	282.5753	332.7812	281.2007	257.6725	304.7289
4	1.5622	1.3977	1.7461	311.9259	288.1422	337.6727	280.3152	257.8563	302.7741
5	1.3042	1.1365	1.4967	301.7416	269.8376	337.4177	278.5013	246.4507	310.5519
6	1.1024	0.8865	1.3708	381.6272	307.6500	473.3927	367.9733	284.4419	451.5047
7,8	0.9718	0.7618	1.2395	247.9506	192.9140	318.6887	251.0791	182.2192	319.9390
≥ 9	0.8472	0.6710	1.0695	156.9852	115.1459	214.0272	171.1464	112.4674	229.8254

Table 7. 95% confidence limits estimated for the Weibull distribution parameters

As can be seen in the tables, earlier failures show a degrading trend; however, later failures appear more random, i.e. follow exponential distribution. The μ values are slightly decreasing for the first five events which have more non-censored data; thus, they constitute a more reliable result. The results are also consistent with the result of the Laplace trend test.

Since SAS only gives the 95% confidence limits estimation for the estimated parameters (β and η), we calculated the limits for μ using the Delta method [13]. For details, see Appendix C. The estimated lower and upper limits for μ using the Delta method appear in the last two columns of Table 7.

3.5.2. Analysis of all hard failures, soft failures, and the failures of individual components

Appendix D describes the results obtained from analysis of all hard and soft failures separately, and some of the system's components individually. An overview of the analysis results appears in Table 8.

Level	*LA	β	η	μ
System level	12.59 (↓)	↓	↓	↓
All hard failures	6.12 (↓)	↓	↓	↓
AC Plug/Receptacles	0.87 (--)	--	--	--
Alarms	4.13 (↓)	↓	↓	↓
Circuit Breaker/Fuse	2.39 (↓)	N/A	N/A	N/A
Controls/Switches	1.15 (--)	--	--	--
Indicators/Displays	-2.72 (↑)	↓	↑	↑
Mount	9.31 (↓)	↑	↓	↓
Occlusion Alarm	-1.11 (--)	--	--	--
All soft failures	11.08 (↓)	↕	↕	↕
Audible Signals	-7.94 (↑)	↓	↑	↑
Battery/Charger	5.69 (↓)	↓	↑	↑
Chassis/Housing	11.40 (↓)	↕	↓	↓
Fittings/Connectors	8.15 (↓)	↕	↓	↓
Labeling	4.13 (↓)	N/A	N/A	N/A
Strain Reliefs	3.03 (↓)	N/A	N/A	N/A
Line Cord	10.61 (↓)	↕	↕	↕
Grounding Resistance	-2.10 (↑)	N/A	N/A	N/A
Symbols: ↑ improving/increasing ↓ degrading/decreasing ↕ no obvious trend -- no trend *LA (the Laplace test value) N/A (not applicable)				

Table 8. An overview of the results obtained from statistical analysis at different levels

The total number of failures and their completeness or censoring status determine the accuracy of the results. With more complete failure data, the results are better and more consistent. For example, due to the large number of exact hard failure data, the results for this group exhibit a degrading trend consistent with the results of the Laplace trend test. However, in the case of only a few failures, nothing can be learned because of the large number of censored data. “N/A” and “↕” results in Table 8 stem from the scarceness of failure data. In most cases, there were a few failures as the first event, and after that, “censoring” was the only available event. Nor were the exact soft failure times available; thus, all data for this group were either left or interval censored, and consequently, the results showed a large variation. Overall, the Laplace trend test appears to yield more reliable results: analyzing the data marginally may show large variations due to the censoring and scarceness of failure data;, and when the results are aggregated, they do not show a clear trend.

4. Concluding remarks

The analysis of data from complex medical devices is not simple, largely because of the amount of censored and missing information. In this paper, we propose a solution to this problem. Even with scarce failure data and a large number of censoring events, we are able to conclude that the reliability of systems degrades over time. This result runs counter to the common belief that failures of electronic equipment, including medical devices, are exponentially distributed and the times between failures of the same device are independent. Practicing statisticians make this erroneous assumption when they fit a distribution to the failure data without checking whether or not the data exhibit a trend. When there is a trend, models such as non-homogenous Poisson process can be used to describe the failure process, and the time to each event can be analyzed separately.

The results of the statistical analysis can be used to develop an inspection/optimization model. An approach to incorporate both soft and hard failures in the model could treat soft failure processes as degradation and hard failures as shocks. Degradation-threshold-shock models [19] can then be employed to find the optimum inspection interval for the systems.

Different trends have been obtained at numerous levels: the system level, hard failures, soft failures, and individual components. To establish an optimal inspection policy at each level, a model should take into account the failure trend and life pattern at that level.

Obviously, precise and reliable results demand precise failure data. Since medical devices tend to be highly reliable, scarce failure data is always a problem in statistical analysis. A solution could be to aggregate the CMMS data from hospitals with similar equipment. This aggregation should be performed cautiously, however, since the same device may exhibit different failure patterns depending on operating and environmental conditions.

Acknowledgements

Our sincere thanks go to John Leung from Toronto General Hospital who provided the maintenance data and devoted his time and expertise to this research. We also would like to thank Neil Montgomery from C-MORE lab for his useful comments. We acknowledge the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Ontario Centre of Excellence (OCE), and the C-MORE Consortium members for their financial support.

References

1. United States Department of Defense, *US Mil-Hdbk-217. Reliability Prediction of Electronic Equipment*, version F, 1995.
2. Choi JG, Seong PH. Reliability of Electronic Components. *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems,* ch. 1, Seong PH (ed.). Springer: New York, 2008.
3. Puran L. Mil-Hdbk-217: What is Wrong with It. *IEEE transactions on Reliability* 1990; **39** (5): 518.
4. Mortin DE, Krolewski JG, and Cushing MJ. Consideration of Component Failure Mechanisms in the Reliability Assessment of Electronic Equipment- Addressing the Constant Failure Rate Assumption. *Annual Reliability and Maintainability Symposium* 1995; 54-59.
5. Fries RC. *Reliable Design of Medical Devices*. Boca Raton: CRC, 2005.
6. Dhillon BS. *Medical Device Reliability and Associated Areas*. Boca Raton: CRC, 2000.
7. Ion RA, Sonnemans PJM, Wensing T. Reliability Prediction for Complex Medical Systems. *Reliability and Maintainability Symposium* 2006; 368-373.
8. Roelfsema S. Early reliability prediction based on Field Data. http://alexandria.tue.nl/extra2/afstversl/tm/roelfsema2004.pdf [2004].
9. Ascher HE, Feingold H. *Repairable Systems Reliability: Modeling, Inference, Misconceptions and their Causes.* Marcel Dekker: New York, 1984.
10. Atles L. *A Practicum for Biomedical Technology and Management Issues.* Kendall/Hunt Publishing Company: Dubuque, 2008.
11. U.S. FDA/CDRH: Product code Classification Database. http://www.fda.gov/cdrh/prodcode.html [15 June 2005].
12. ECRI Institute- Infusion Pumps, General-Purpose. http://www.ECRI.org [2009].
13. Meeker W, Escobar L, *Statistical Methodology for Reliability Data*. Wiley: New York, 1998.
14. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ [18 July 2006].
15. Ansell JI, Phillips MJ. *Practical Methods for Reliability Data Analysis.* Oxford: Oxford University Press, 1994.
16. Trindade D, Nathan S. Field Data Analysis for Repairable Systems: Status and Industry Trends. *Handbook of Performability Engineering*, ch. 26, Misra KB (ed.). Springer: New York, 2008.
17. Nelson W. An application of graphical analysis of repair data. *Quality and Reliability Engineering International* 1998; **14** (1): 49-52.
18. Johnston G. Reliability Data Analysis in the SAS System. http://support.sas.com/rnd/app/papers/reliability.pdf.
19. Lehmann A. Degradation-Threshold-Shock Models. *Probability, Statistic and Modelling in Public Health*, Nikulin M, Commenges D, and Huber C (eds.). Springer: New York, 2006.

Appendix A – Soft and hard failures

Non-scheduled work orders							Scheduled work orders					
Component	**P**	**F**	**N/A**	**Total**	**P %**	**F %**	**P**	**F**	**N/A**	**Total**	**P %**	**F %**
Accessories	48	0	1154	1202	100	0	35	0	1629	1664	100	0
Cables	51	0	1151	1202	100	0	38	0	1626	1664	100	0
Casters/Brakes	49	0	1153	1202	100	0	33	0	1631	1664	100	0
Flow-Stop Mechanism	1202	0	0	1202	100	0	1519	0	145	1664	100	0
Lockout Interval (PCAs Only)	48	0	1154	1202	100	0	34	0	1630	1664	100	0

Table A-1. Components/features that can be ignored since they have no failure reported

Non-scheduled work orders							Scheduled work orders						
Component	**P**	**F**	**N/A**	**Total**	**P %**	**F %**	**P**	**F**	**N/A**	**Total**	**P %**	**F %**	***Ratio**
AC Plug/Receptacles	1189	13	0	1202	98.92	1.08	1510	9	145	1664	99.41	0.59	1.83
Alarms	881	321	0	1202	73.29	26.71	1268	251	145	1664	83.48	16.52	1.62
Circuit Breaker/Fuse	48	13	1141	1202	78.69	21.31	33	3	1628	1664	91.67	8.33	2.56
Controls/Switches	1097	105	0	1202	91.26	8.74	1433	86	145	1664	94.34	5.66	1.54
Indicators/Displays	1094	108	0	1202	91.01	8.99	1383	136	145	1664	91.05	8.95	1
Mount	1009	191	2	1202	84.08	15.92	1319	200	145	1664	86.83	13.17	1.21
Occlusion Alarm	1191	9	2	1202	99.25	0.75	1507	12	145	1664	99.21	0.79	0.95

Table A-2. Components/features that have more failure at non-scheduled work orders (hard failures)

Non-scheduled work orders							Scheduled work orders						
Component	**P**	**F**	**N/A**	**Total**	**P %**	**F %**	**P**	**F**	**N/A**	**Total**	**P %**	**F %**	***Ratio**
Audible Signals	1105	97	0	1202	91.93	8.07	1250	269	145	1664	82.29	17.71	0.46
Battery/Charger	1177	25	0	1202	97.92	2.08	1321	198	145	1664	86.97	13.03	0.16
Chassis/Housing	700	502	0	1202	58.24	41.76	620	899	145	1664	40.82	59.18	0.71
Fittings/Connectors	962	240	0	1202	80.03	19.97	1073	446	145	1664	70.64	29.36	0.68
Labeling	1090	112	0	1202	90.68	9.32	1063	455	146	1664	70.03	29.97	0.31
Strain Reliefs	1182	18	2	1202	98.5	1.5	1473	41	150	1664	97.29	2.71	0.55
Line Cord	1067	135	0	1202	88.77	11.23	1117	402	145	1664	73.54	26.46	0.42
Grounding Resistance	1199	2	1	1202	99.83	0.17	1517	2	145	1664	99.87	0.13	1.31

Table A-3. Components/features that have more failure at scheduled work orders (soft failures)

*Ratio = No of failures identified at NS/ No of failures identified at PM

We included "Occlusion Alarm" in the hard group because of its similarity to the "Alarms" component in the qualitative test. "Grounding Resistance" has only a few failures in both scheduled and non-scheduled work orders, so we included it in the soft group.

Appendix B - Rules for analyzing failure data at the system level

• We consider all non-scheduled work order starting times as system exact failure times (complete data) since we assume that a non-scheduled work order is requested upon a hard failure.

• The time interval between two consecutive failures ending in a SPI with at least one failure (scheduled work order) is treated as left censored data since we consider this detected failure as a soft failure.

• The time interval between an inspection time with "no problem found" and a SPI with at least one failure is treated as interval censored data.

• If the last available history ends in a failure and the time interval between this failure and the end of the test is short (<=0.5*mean time between failures), we consider the time interval as right censored data.

• If the last available history is a failure and the time interval between this failure and the end of the test is long (>0.5*mean time between failures), we consider 0.5*mean time between failures as right censored data.

• If the last available history is an inspection with "no failure found" and the time interval between this inspection and the end of the test is short (<= 0.5*mean time between failures), we consider the time interval between the last failure event and the end of the test as right censored data.

• If the last available history is an inspection with "no failure found" and the time interval between this inspection and the end of the test is long (> 0.5*mean time between failures), we consider the time interval between the last failure event + 0.5*mean time between failures as right censored data.

Appendix C- Confidence limits estimation for μ (the Delta method)

The "Delta method" can be applied to estimate $\mathrm{Var}(g(\hat{\theta}))$, if $g(\hat{\theta})$ is a smooth nonlinear function of $\hat{\theta}_i$ and can be approximated by a linear function of the $\hat{\theta}_i$ values in the region with no negligible likelihood. The Delta method is applied based on a first order Taylor series expansion of $g(\hat{\theta})$ about $\gamma = [E(\hat{\theta}_1),...,E(\hat{\theta}_r)]$.

If we apply the Delta method to the μ as a function of $\hat{\eta}$ and $\hat{\beta}$, and assuming low correlation between $\hat{\eta}$ and $\hat{\beta}$, we obtain:

$$Var[\mu(\hat{\eta},\hat{\beta})] \approx (\frac{\partial \mu}{\partial \eta})^2 Var(\hat{\eta}) + (\frac{\partial \mu}{\partial \beta})^2 Var(\hat{\beta}) = \left[\Gamma(1+1/\hat{\beta})\right]^2 Var(\hat{\eta}) + \left[\hat{\eta}\psi(1+\frac{1}{\hat{\beta}})\Gamma(1+\frac{1}{\hat{\beta}})/\hat{\beta}^2\right]^2,$$

where $\psi(x) = \text{digamma}(x) = \dfrac{\mathrm{d}(\log(\Gamma(x))}{\mathrm{d}x} = \dfrac{\mathrm{d}(\Gamma(x))/\mathrm{d}x}{\Gamma(x)}$.

The $\mathrm{Var}(\hat{\beta})$ and $\mathrm{Var}(\hat{\eta})$ can be obtained from standard error estimates of SAS output.

Appendix D- Analysis of failure data

D.1. Hard failures

The results of the Weibull distribution fitted to the times between n-1th and nth event (failure/censoring)								
n	No of observations	No of non-censored	No of right censors	No of left censors	No of interval censors	β	η	$\mu=\eta\Gamma(\frac{1}{\beta}+1)$
1	674	502	172	0	0	1.2982	1215.687	1123.1
2	502	317	185	0	0	1.1256	1077.554	1032.1
3	317	185	132	0	0	1.1007	924.7741	892.1
4	185	91	94	0	0	0.9864	768.3577	772.9
5	91	46	45	0	0	0.8626	701.4417	756.3
≥ 6	121	76	45	0	0	0.7785	310.3867	358.8

Table D-1. Weibull distribution parameters estimated for times between events – hard failures

LA=6.12 (systems are degrading)

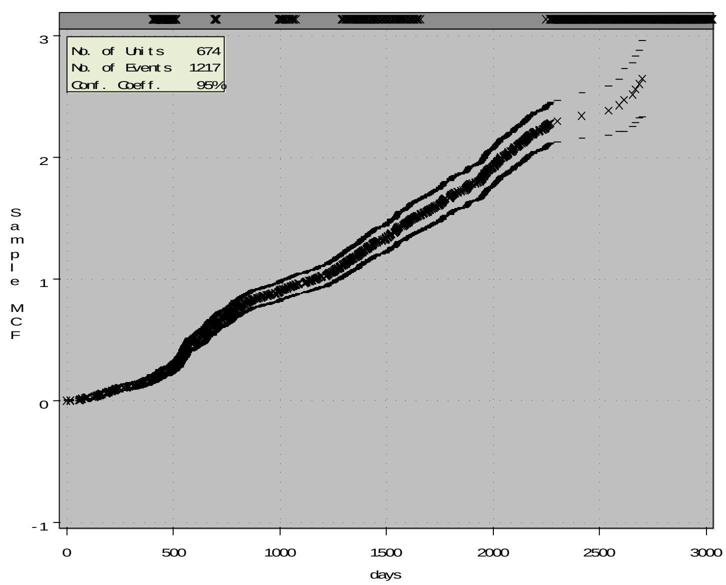


Figure D-1. The MCF plot of hard failures

D.2. Soft failures

The results of the Weibull distribution fitted to the times between n-1th and nth event (failure/censoring)								
n	No of observations	No of non-censored	No of right censors	No of left censors	No of interval censors	β	η	$\mu=\eta\Gamma(\frac{1}{\beta}+1)$
1	674	0	30	463	181	1.1162	350.7222	336.8
2	644	0	122	418	104	1.2525	254.8389	237.2
3	523	0	116	335	72	1.0424	247.6217	243.5
4	404	0	158	217	29	1.0465	303.4680	298
5	247	0	124	108	15	0.8027	409.1623	462.5
6	123	0	76	40	7	0.5665	701.3626	1142.3
≥ 7	82	0	46	26	10	0.7959	351.3244	399.5

Table D-2. Weibull distribution parameters estimated for times between events – soft failures

LA=11.08 (systems are degrading)

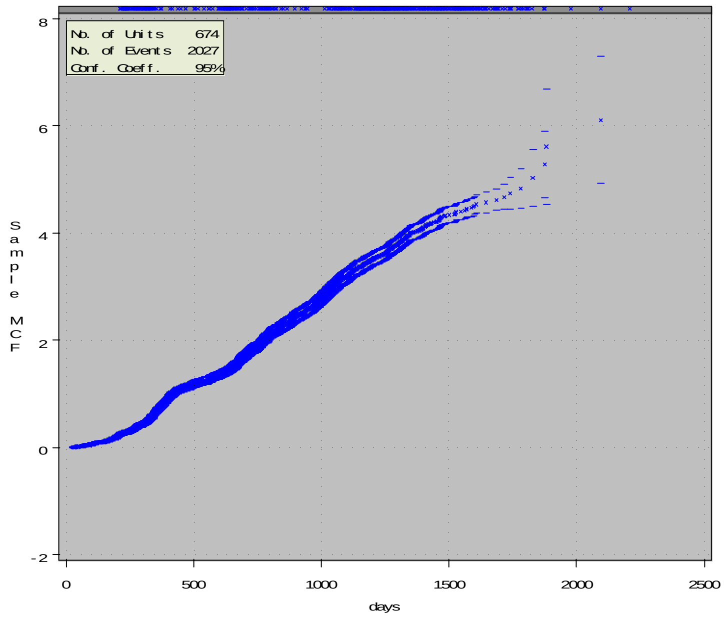


Figure D-2. The MCF plot of soft failures

D.3. Alarm failures

The results of the Weibull distribution fitted to the times between n-1th and nth event (failure/censoring)								
n	No of observations	No of non-censored	No of right censors	No of left censors	No of interval censors	β	η	$\mu=\eta\Gamma(\frac{1}{\beta}+1)$
1	674	306	368	0	0	1.1876	2564.701	2419.1
2	306	140	166	0	0	0.9368	2118.686	2183.1
3	140	64	76	0	0	0.8782	1353.484	1443.9
4	64	24	40	0	0	0.6894	1113.709	1430.8
≥ 5	61	38	23	0	0	0.6548	255.7360	346.6

Table D-3. Weibull distribution parameters estimated for times between events – Alarms failures (hard)

LA=4.13 (systems are degrading)

D.4. Chassis/Housing failures

The results of the Weibull distribution fitted to the times between n-1th and nth event (failure/censoring)								
n	No of observations	No of non-censored	No of right censors	No of left censors	No of interval censors	β	η	$\mu=\eta\Gamma(\frac{1}{\beta}+1)$
1	674	0	101	317	256	1.2662	645.6414	599.6
2	573	0	146	251	176	1.1804	497.9293	470.4
3	428	0	173	178	77	1.2267	477.8781	447.1
4	252	0	146	79	27	1.5471	473.8461	426.3
5	107	0	75	24	8	1.3434	597.5264	548.4
≥ 6	40	0	32	7	1	0.7905	1395.492	1594.9

Table D-4. Weibull distribution parameters estimated for times between events – Chassis/Housing failures (soft)

LA=11.40 (systems are degrading)