Eindhoven University of Technology

MASTER

Early reliability prediction based on field data

Roelfsema, S.

Award date:
2004

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Philips
Philips Medical Systems
Veenpluis 4-6
5684 PC Best

Business Group Digital Imaging Systems
Business Unit Cardio Vascular
Development department
Best, the Netherlands

Mentors from Philips Medical Systems:
M. v. Asten
ir. J.A.M. Versteijlen
Ing. J.P de Stigter

Technische Universiteit Eindhoven
Eindhoven, the Netherlands

1st supervisor
dr. R.A. Ion
Section: Quality and Reliability Engineering

2nd supervisor
dr. ir. J.J.M.Trienekens
Section: Information Systems

3rd supervisor
dr. ir. P.J.M. Sonnemans
Section: Quality and Reliability Engineering

Author
Stefan Roelfsema
Student number 451299
Technische Universiteit Eindhoven
Faculty: Technology Management
Study: Industrial Engineering and Management Science
Section: Quality and Reliability Engineering

Preface

This report is the result of a nine months graduation project performed at Philips Medical Systems in Best. Working on this project has been a great experience. It has been an opportunity for me to see what it is like to work in an international company with all its dynamics and to work on a large project. In the beginning it is hard to overlook the whole project and during the project I had to evaluate my progress several times in order to keep on track. Thankfully I had great support from both Philips and the Technical University Eindhoven (TU/e) to help me in this respect. In the end, looking back the time has gone by faster than I could ever expect, but I guess that is a good thing meaning that I enjoyed my time.

Philips has also given me the opportunity to enjoy other activities through their young employee's network called Global I's. This has enriched my internship even more, being able to go to the CEBIT in Hannover, participate in a swimming clinic with Pieter van den Hoogenband, and being a member of the team of the corporate social responsibility team contest 2004.

I am thankful to everyone making this project possible for me and I like to take this opportunity to thank a number of people in particular. My thanks go out to my supervisors at the TU/e, Roxana Ion, Peter Sonnemans and Jos Trienekens for their input, support, and feedback. I also like to thank Jan Rouvroye who helped me out with the Matlab programming.

At Philips Medical Systems my thanks go out to my mentor, Riny van Asten, who was always there for me, ready to help me in any way, giving me feedback, and introducing me to other people in the organization. Also to Jaap de Stigter and Martien van Overeem, for their feedback and support whenever I needed it. Next to that I like to thank Huub Gooren for the pleasant time in our office and the nice conversations. Also my thanks go out to Janwillem Schrier who did great work in providing data necessary for the project.

The other students doing their graduation project at Philips Medical Systems were great company and a great help by sharing of experiences and thoughts about all aspects of doing a graduation project. Especially the close cooperation with Crispijn Roos lead to new insights helpful to the project. And a special thanks goes out to Beerd van Schijndel who helped me with the cover of this report.

I would like to thank my family and my girlfriend Marisol for their support and for always believing in me, which has been a great support to me.

My thoughts go out to Joost Versteijlen, who was passionate about improving the reliability of the cardio vascular x-ray systems at Philips Medical Systems, and who passed away much too young. With my work I hope to have contributed to the ideas he had about the improvement of the reliability analysis capabilities within Philips Medical Systems.

Stefan Roelfsema

Eindhoven, July 12, 2004

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Abstract

This report presents the results of research performed at Philips Medical Systems in Best (the Netherlands) concerning the early reliability prediction based on field data. This implies an investigation on the (quality of) required data, choosing relevant models for the specific situation, and performing analyses on the data with those models. The analyses are split in three parts, one analysis focusing on data from one system, the second analysis focusing on data from multiple systems, and the third analysis focusing on the difference between two data sources.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Management summary

This report describes the graduation project performed at the Cardio Vascular development department of Philips Medical Systems. Philips Medical Systems is a large organization making complex high-tech products. The company is situated in a dynamic medical systems environment where there are trends towards a growing market, higher product utilization, increasing product functionalities and increasing pressure on cost of ownership. This means product reliability is an important issue.

The project falls within a research field of the Quality and Reliability Engineering (QRE) department at the Technical University in Eindhoven. The aim of that research field is to develop competence in quality and reliability by creating methods to predict the occurrence of product failures in the development process and early in the field introduction. PMS wants to be able to monitor, control and predict the product reliability in an earlier stage. In that way the feedback loop can be shortened, which leads to faster problem recognition.

The aim of this project is to investigate how the product failure pattern that is found at PMS can be modeled and to assess the prediction performance of those models.

Trends in the market are making reliability prediction using field data nowadays even more necessary then before. Time for extensive laboratory testing of products is not available anymore. The speed of evolution of new technologies and new product designs and the market-dictated short lead times between design start and product shipment, in parallel with longer Mean Time To Failure (MTTF) and Mean Time Between Failures (MTBFs), have made it difficult to find the test time to accumulate sufficient failure data to be useful for future prediction purposes. Added to this is the uncertainty about the customer use which, together with the mentioned trends, makes it hardly possible to develop and test the product on fitness for use for a reasonable time period (at least the warranty period). Related to these trends is the cost aspect. The sooner failures in products are recognized, the lower the cost of repairing those failures, since the number of products sold (that need to be repaired) is still relatively small.

The available data
There are two sources with field data available: service data and a data from software loggings of system at the customers, called FMT data. Both data sources are able to provide the necessary data for modeling system reliability. That is, they provide the failure moments relative to the moment of installation. There are however problems related to data quality for both sources of data. The service data has data quality problems related to all four data quality metrics, that is, completeness, consistency, timeliness, and accuracy. The FMT data has data quality problems related to completeness and accuracy. For the service data a manual filtering is performed to improve the data quality.

Failure moments
For the service data the failure moments are approximated by the difference between the warranty start date and the service call start date. This number gives the time to failure in days. The failure moments of the FMT data are obtained differently. The systems at the customer are able to provide the system online time enabling accurate determination of the failure moments. The failure moments are given in system online hours.

Models
The category of models that is found to be applicable has the following characteristics: continues parametric models for repairable systems with a mixture of hardware and software, without redundant parts. The model choice can be made using the general procedure for analyzing failure data of a repairable system. To determine whether there is trend in the failure data or not the Laplace trend test for multiple system is used. A straightforward generalization of the single system Laplace test if there are observations from m independent systems is:

$$L_c = \frac{\sum_{j=1}^{m}\sum_{i=1}^{\hat{n}_j} T_{ij} - \sum_{j=1}^{m}\frac{1}{2}\hat{n}_j(b_j + a_j)}{\sqrt{\frac{1}{12}\sum_{j=1}^{m}\hat{n}_j(b_j - a_j)^2}} \tag{1}$$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

The combined Laplace test indicates significant trend in five out of the six calculations, based on 1000 days of data. Using the general procedure for analyzing failure data of a repairable system this leads to the choice for Non Homogeneous Poisson Process models. There are two models most used in literature that are described as useful NHPP models for analyzing the failure pattern of repairable products: Power law and Exponential law. Both these models are investigated in this report.

Power law
Intensity:

$$\mu_1(T) = \lambda\beta T^{\beta-1} \text{ with } \lambda, \beta > 0,\ T \geq 0 \tag{2}$$

When the shape parameter β is equal to 1, the power law model reduces to the homogeneous Poisson process (HPP). When β>1 (β<1) the intensity function is monotonically increasing (decreasing) with the operating time T: this corresponds to the situation in which the times between successive failures become shorter (longer) with T. The parameters for this model are calculated using the following formulas:

$$\hat{\beta} = \frac{n}{\sum_{j=1}^{m} \sum_{i=1}^{n} \ln \frac{T_n}{T_{ij}}} \tag{3}$$

and

$$\hat{\lambda}^* = \frac{n}{T_n^{\beta}} \tag{4}$$

$$\lambda^* = m\lambda$$

The expected number of failures is: $E[N(T)] = \lambda T^{\beta}$ (5)

Exponential law
Intensity:

$$\mu_2(T) = e^{\alpha_0+\alpha_1 T},\ -\infty < \alpha_0, \alpha_1 < \infty,\ T \geq 0 \tag{6}$$

Since there were no articles or books found for calculating the parameters when including more than one system in the Exponential Law model, this formula has been derived by R. Ion from the Technische Universiteit Eindhoven.

Parameter estimation:

$$\alpha_0 = \ln\left(\frac{\alpha_1 \sum_{j=1}^{m} n_j}{\sum_{j=1}^{m} (e^{\alpha_1 t_{nj}} - 1)} \right) \tag{7}$$

and

$$\sum_{j=1}^{m} \sum_{i=1}^{n_j} T_{ij} + \frac{\sum_{j=1}^{m} n_j}{\alpha_1} - \frac{\sum_{j=1}^{m} n_j \sum_{j=1}^{m} T_{nj} \cdot e^{\alpha_1 T_{nj}}}{\sum_{j=1}^{m} (e^{\alpha_1 T_{nj}} - 1)} = 0 \tag{8}$$

Analyses
Both the service data and the FMT data are used in the failure modeling. The data sets that are used are equal for both model types. The analyses that are performed can be divided in the following manner:

- Analysis of service data of a single system
  This analysis investigates whether the power law model and the exponential law model are able to fit the failure pattern of the data that is used in this project. For this analysis unfiltered data is used, this has no influence on the question of whether the power law and exponential law are able to fit the failure pattern.
- Analysis of service data of multiple systems
  This analysis investigates the influence of using data from multiple systems for modeling the failure pattern. The analysis uses both unfiltered and filtered data to investigate the influence on the calculated values.
- FMT data versus service data
  This analysis shows the differences between using FMT data and using service data when trying to establish the failure pattern of a system.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

These analyses are used to determine parameter values of the model, to calculate the ROCOF, to determine the goodness-of-fit, and to construct the confidence intervals on the $MTBF_c$. To get more insight in the behavior of the power law and exponential law model, the models are created on the basis of data containing 365 days, 730 days, and 1000 days of operating time. For all three groups of data the model is calculated for 900 days of operating time. This means that the model based on 365 days and the one based on 730 days make a prediction of the number of failures until 900 days through extrapolation. The model based on 1000 days of data does not predict, it serves as a comparison of the line that the model would plot if it had all the necessary data.

Analysis one: Analysis of service data of a single system.
A first conclusion is that both modeling types, power law and exponential law, are able to fit the data of a single system based on 1000 days of data. But when prediction of the failure pattern for 900 days is concerned based on a limited amount of data (365 days), both model types give a wrong prediction of future failures. This has to do with the following: the models are flexible in a way that they can model a deteriorating system, an improving system and a system that shows no trend (a Homogeneous Poisson Process). The form that the model takes depends on the data that is put in the model. This means that the failure pattern early after field introduction of the system determines the pattern of the model. If a trend develops later on in the product life this will not be modeled when only using data from the period shortly after field introduction. The data needs to show at least a small amount of trend in order for the model to pick up the sign that the system is actually improving in time. When the model is based on 730 days of data it is able to predict the failure pattern of a 900 days period correctly.

Analysis two: Analysis of service data of multiple systems.
A first conclusion here is that both the power law and the exponential law models adapt well to the data from the number of systems included in the model building. Next to that, similar effects as when using a single system for building the model can be found, meaning an inaccurate representation of the failure pattern when the data of 365 days is used. This means that when the model is based on 365 days of data this model can only be used to determine the situation at that moment, not to predict the future failure pattern. The model based on 730 days of data is very close to the model based on 1000 days of data, like with the model based on a single system. This means that based on 730 days of data of good prediction of a 900 days period can be made.

The large difference in the number of failures that occurs per system has a big influence on the fit of the model to the data. The model is of course an average and therefore it will not fit individual systems well, especially if the systems show such a large difference in number of failures. This spread in failure pattern between different systems makes the predictive value less accurate and causes relatively wide confidence bounds. When the spread in the failure pattern is smaller this leads to predictions that are closer to the actual failure pattern, although based on 365 days of data it still deviates too much to be called a good prediction.

The power law and exponential law model do not always agree on the form the model should take. This even means that in some cases, based on the same data, one model type calculates an increasing reliability, while the other model type calculates a decreasing reliability. The cause of this can be found in the large difference in the number of failures that occur per system. One model type reacts differently than the other in case of such a diverse failure pattern between systems. When the failure patterns are not consistent the models neither will be.

The influence of pollution in the data was investigated by comparing filtered data to unfiltered data. The filtered data shows similar failure patterns as the unfiltered data, although clearly different lines appear when the models are plotted. This means that the filtered data leads to different values of the expected number of failures, the model intensity, the goodness-of-fit, and the MTBF. It proves that having a good data quality is essential for obtaining reliable answers from these models.

Analysis three: FMT data versus service data.
This analysis shows the differences between using FMT data and using service data. The difference between the two data sources is clearly visible when they are compared; the failure moments, as well as the number of failures do not correspond. This leads to differences in the model based on FMT data and the model based on service data. That leads to maybe one of the most important conclusions from this report, that is, there first

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

needs to be a clear understanding of the data that is put in the model before conclusions can be drawn about the final number that the model gives as an output. The quality of a model can be only as good as the quality of the data that is put in.

Overall conclusions

A basic question is which data source to use for making reliability analyses and predictions, service data or FMT data. The first relevant observation in that respect is that different data is obtained by these two data sources when failure data from the same systems over the same time period is collected. The FMT data has less human influence during the data collection and next to that the failure moments are recorded with much greater accuracy and therefore should be the data source used for the analyses. Service data should be used next to that for other analyses like call rate, material usage and material cost analysis.

Choice between the power law model, the exponential law model, or another type of model.
Both the power law model and the exponential law model have showed to be able to correctly fit to the failure pattern of the systems. And although there are some differences in the failure pattern that they show no definite conclusions about one model type being better than the other can be made. The problem with both these models is that early prediction based on data from a relatively small period of time (365 days) is not possible given this particular failure pattern shown by the systems analyzed in this project. This means that the search for a different model could be undertaken, looking for a model where earlier prediction of the reliability might be possible.

High data quality will not lead to better reliability prediction based on a small amount of data. The flexibility of both the Power law and Exponential law model – being able to model increasing, decreasing, and constant reliability – forms the problem in the reliability prediction. Based on data from a relatively small period of time (365 days) it is not possible to give a good prediction of the period until 900 days given the particular failure pattern shown by the systems analyzed in this project. This is due to the fact that the reliability improvement becomes visible later in the product lifetime. When 730 days of data are included a better prediction can be made.

Further research

A follow up should focus on the following aspects:

1. Improving data quality of FMT tool
   It is essential that the data quality of the data provided by the FMT tool is as good as possible. Since this tool is only being used since very recent for making reliability analyses there probably is room for improvement. Especially the question of accuracy of the failure data is important to research extensively.

2. Analysis of more FMT data sets
   This project provides the first analysis based on FMT field data. More analyses based on data of a larger time period need to be performed in order to conclude whether the behavior of the FMT data is similar to the service data.

3. Research on new model that makes a better prediction based on a small amount of data (early data).
   The analyses have indicated that the power law and exponential law model are not able to predict the failure pattern based on a small amount of data; that is, not based on the datasets that were used in this project. Therefore, if a different model can be found or developed which is able to make a better prediction based on a small amount of data this would mean a significant improvement.

4. Implementation of the model into the business procedures.
   When a model is found giving better early predictions, or when it is decided to use the power law or exponential law model given its limitations, a good implementation of the analyses that can be performed with these models needs to be assured. The model analyses need to fit in the procedures of the organization.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

List of abbreviations

CM	Corrective Maintenance
CSA	Customer Support Agreement
FCO	Field Change Order
FMT	Field Monitoring Team
FPR	Field Problem Report
FSE	Field Service Engineer
GDW	Global Data Warehouse
HPP	Homogeneous Poisson Process
MTBF	Mean Time Between Failures
MENHIR	Market Essential Harmonized Installed base, Reliability & Performance Reporting
NHPP	Non homogeneous Poisson Process
PM	Planned Maintenance
PMS	Philips Medical Systems
QMCB	Quality Maintenance Control Board
QRE	Quality and Reliability Engineering
ROCOF	Rate of occurrence of failure
SSD	Sales and Service District
SSR	Sales and Service Region
TU/e	Technical University of Eindhoven
TSO	Table Side Operation (module)

Symbols

α_0 = parameter for the exponential law model

α_1 = parameter for the exponential law model

β = shape parameter for the power law model

λ = scale parameter for the power law model

$\mu_1(T)$ = power law intensity function

$\mu_2(T)$ = exponential law intensity function

$\chi^2(n)$ = Chi-squared value

$C^2(n)$ = Cramér-von Mises value

L = Laplace value

L_c = Combined Laplace value

$MTBF_c = m_c$ = Cumulative Mean Time Between Failures

$MTBF_i = m_i$ = Instantaneous Mean Time Between Failures

N(T) = number of failures in (0,T)

E[N(T)] = expected number of failures in (0,T)

i = index for failure

j = index for system

m = total number of independent systems

n = total number of observed failures for a system

T = Time from start of system life

T_{ij} = Time to failure i of system j (i = 1 – n), (j = 1 – m)

T_n = Total observed time period

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Table of content

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Introduction

Reliability is a basic demand for customer satisfaction. During development of a new product the marketing department is already promising new and improved functionalities in the new product; customer expectations are created. However, although getting a customer may be achieved by promoting the functionalities and improvements of the product, when it comes to keeping the customer it is crucial to live up to the expectations; making sure the product provides the presumed functionalities and performs these functionalities to a satisfactory level during a specified period of time. And if the product does brake down the problem has to be resolved quickly and adequately.

Next to innovation, speed is a key word in the business environment of today. Strong competition makes it imperative to have the innovative products in the market before or at the same time competitors do. This faster environment asks for faster determination of the product reliability. Is the product reliability high enough? What kind of failure pattern does the product contain? How can reliability analysis help in these questions?

During development the necessary tests for reliability are performed, but in order to know how the product performs at the customer it is necessary to analyze the data that comes from the field. This report investigates how the product failure pattern that is found at Philips Medical Systems can be modeled and assesses the prediction performance those models. In order to be able to make argued statements about this question scientific articles and books were consulted to create a theoretical basis. Practical application of theory was performed through research at Philips Medical Systems.

Two important terms from this introduction are *field data* and *reliability*. The field data that is used in this report is broader than data accumulated through service on products only; data from a direct link to the product at hospitals will be used as well. This will be discussed in more detail in paragraph 2.1.3. The first definition that will be given is for reliability. Kales [1] gives in his book a detailed scientific definition of **reliability**:

> ***The reliability of an item is the probability that the item will perform a specified function under specified operational and environmental conditions, at and throughout a specified time.***
>
> ***This means that before we can deal with reliability, the producer and the user must reach formal agreements on what the product is to do, how the user is to use the product, the range of environments under which the product is expected to perform satisfactorily, and the instant or duration in time that the performance of the product or service is demanded.***

An overview of all the definitions given throughout this report is given in Appendix A. Throughout the report the words 'product' and 'system' are used exchangeably. When spoken of in general the word 'product' is used; specific entities are called 'system'.

Figure 1 describes the structure of this research report and the interaction of its different steps. This structure has four levels indicated as horizontal fields: The upper level indicates the wish to solution level; the second level indicates the current way of working at PMS, the third level contains the search for a useful model, and the fourth level indicates the investigated proposed model and changes for proposed data input. The report contains these levels and its steps in the following way:

Chapter I describes steps 1, 2 and 3: a background on the field data environment and the medical systems environment leads to a description of the research project. This is followed by the reliability analysis benefits and the specific project objectives.

Chapter II describes steps 4, 5 and 6: the current data collection, data quality, and data analysis situation at PMS is described. With this information the necessary background information is given to be able to look for a new reliability model.

Chapter III describes step 7, 8 and 9: a research on parametric reliability prediction models described in literature. It describes a framework for continues reliability models and gives a more detailed

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

description on models for repairable systems. Step 8 describes the process of making a model choice. Step 9 gives the conclusion of this chapter.

Chapter IV describes step 10, 11 and 12: issues concerning the data input and the proposed failure modeling are described followed by an evaluation of these steps.

Chapter V gives the conclusions in step 13 and therefore comes back to steps 3, 6, 9 and 12.

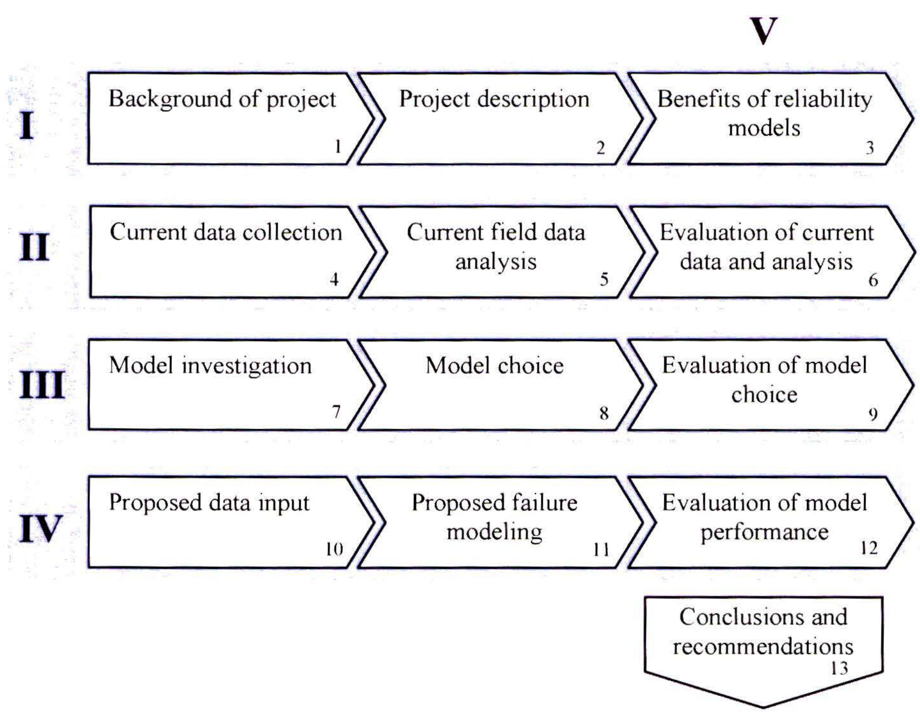


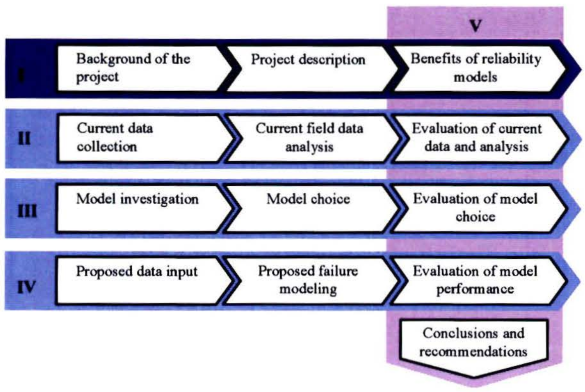
Figure 1 Report structure

Each chapter will start with a small version of figure 1 to indicate which steps of the project are explained in that chapter. Next to the figure is a quote that is related to that chapter.

In this version of the report actual data are replaced by fictive data due to the confidential character of the data.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Chapter I Project background and description



Let's make things better
(Philips slogan)

This first chapter starts with explaining the specific situation the project is performed in. It gives background information on the medical systems environment on one side and the importance of field data on the other side. This is step 1 of the report structure. After that the project description will be discussed (step 2) and subsequently the benefits of reliability models are described (step 3).

1.1 Background of the project

1.1.1 Medical systems environment

The medical systems environment discussed in this paragraph consists of facts and figures of PMS, especially zooming in on the department where the project is performed and relevant trends in healthcare related to the product under consideration in this report.

Philips Medical Systems

General Information

Philips Medical Systems is part of Royal Philips Electronics NV. It is one of the world leaders in diagnostic imaging, patient monitoring, clinical IT and related services. The division is active in healthcare for over 100 years and currently employs 30,000 people worldwide with over 6,000 service professionals. The CEO of PMS is Jouko Karvinen. There are 18 manufacturing sites over the world and over 450 products and services are sold in more than 100 countries. The R&D expenses amount more than 11% of system sales [2].

Ambition

The ambition of Philips Medical Systems is to become the premier healthcare technology company in the world through a relentless pursuit of innovation.

Strategy Objectives

- Achieve 14% EBITA (earnings before interest and tax) in 2004, and grow faster than the market
- Continuous innovation of products & clinical applications
- Improve patients' lives through technology
- Building the strongest customer relationships
- Clinical Excellence without compromise

Financial situation

The sales of PMS were around 6,850 million euro in 2002, which is 22% of total sales of Philips. This means a growth of 42%, of which 41% was caused by takeovers. The sales volume increased with 8%, while the average prices decreased with 3%. Regional growth was the largest in North America. In table 1 the sales growth over the period of 1996 – 2002 is given.

Table 1 Sales growth

Year	1996	1997	1998	1999	2000	2001	2002
Sales indexed with 1996 = 100	100	112	123	157	191	304	430

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

PMS Product groups
PMS has a wide range of medical products, see table 2, using different high-tech technologies. In the next overview the product groups are given, within each group there are again several types of systems. One of the product groups given in table 2 is Cardio Vascular X-ray / Cath labs. This is the product group that the focus of the research performed in this project will be on.

Table 2 product groups [3]

Cardio Vascular X-Ray/Cath labs		Magnetic Resonance Imaging (MRI)
Echo- / electrocardiography		Computed Tomography (CT)
Radiation Therapy Planning		Surgical X-Ray/C-arm
Defibrillation / resuscitation		Home monitoring
Multimodality Fusion		Digital X-ray
Medical IT/PACS		Diagnostic ECG
Nuclear Medicine		Positron Emission Tomography (PET)
Ultrasound		Customer Support, Healthcare Consulting and Financial/Leasing Service

Organization
In Figure 2 an overview of the organization is given. The practical research of this project is performed in the systems department. The figure shows that the Systems department falls under Development, which is part of the Business Unit Cardio Vascular (BU CV) within the Business Group Digital Imaging Systems (BG DIS) of the PMS organization. Another important part in this figure is the Sales and Service organization; this is where the field data is created. This part of the organization is divided into three parts, the Sales and Service Regions (SSR).

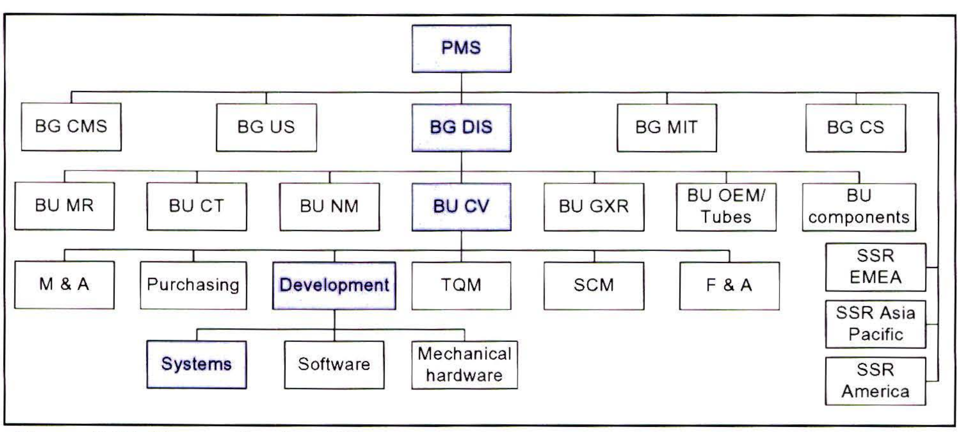


Figure 2 PMS organization [3]

Business Unit Cardio Vascular
In this Business Unit cardio vascular (CV) systems are made for cardiac and vascular imaging. These can be used in diagnostics and intervention. Techniques such as 3D image reconstruction and a flat detector, which transforms the x-ray beams directly into a digital signal, are important features of the system. Table 3 shows the current products of cardio / vascular x-ray and gives the application area of these products. The numbers in the product names reflect the size of the x-ray detector; biplane means there are two stands (two detectors).

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Table 3 cardio / vascular x-ray products

Product	Description
Integris Allura 9	Cardiac and mixed cardiovascular applications
Integris Allura 9 biplane	Pediatric and electro-physiology applications
Integris Allura 12	Mixed cardiovascular or general vascular applications
Integris Allura 12 biplane	Neurovascular applications
Integris Allura 15	Mixed cardiovascular or general vascular applications
Integris Allura 15 biplane	Neurovascular applications
Integris Allura CV	Versatile applications
Integris 3D RA	(neuro) vascular interventions
Allura Xper FD10	Cardiac and mixed cardiovascular applications
Allura Xper FD 20	Mixed cardiovascular or general vascular applications

Healthcare trends

There are several trends specifically in healthcare. First of all the aging of the human population, through the last century the percentage of old people has risen and it will keep rising in the future. This means that there are less people to finance the growing costs of healthcare. The hospitals are reacting to this by working hard on the efficiency. Next to that there is more emphasis on getting a better diagnosis for each patient. The *MD Buyline report* [4] gives a detailed report about the market development in the cardio/vascular labs. In the following lines some aspects of this report are mentioned.

The cardiac and vascular lab market is growing and will continue to do well as the baby boomer population ages and begins to reflect symptoms of cardiovascular disease. But although the market is growing, it continues to be largely a replacement market, particularly as digital detector flat panel technology is purchased.

Although the number of dedicated labs is shrinking, a sizeable increase in the number of combination cardiac/vascular labs is being purchased. Cardiac magnetic resonance imaging (MRI) and computed tomography (CT) are emerging as primary diagnostic tools to rule out cardiovascular disease in patients. The result should be a greater emphasis on therapeutic and interventional work with cardiac/vascular labs rather than basic diagnostic work: the patients going to the lab will have already been proven to have cardiovascular disease. This means increased procedure times for interventional versus diagnostic procedures as well as higher stress levels on the equipment. In turn, these factors should drive expansion of existing imaging services to address workload, utilization, and increased patient populations. These processes lead to higher reliability demands.

Healthcare organizations are also becoming more conscious of the cost of operating the labs. The pressure on costs will continue to increase as the procedure mix shifts to more costly interventional procedures. Also, the organizations strive to create a more consistent utilization pattern among all physicians to manage costs. Next to that facilities are seeking more comprehensive information management systems that may originate in the imaging lab but extend to all cardiac and/or vascular diagnostic and disease management areas in the hospital. This drive on cost reduction means that Philips will need to deliver the earlier mentioned increasing reliability for a decreasing cost of ownership of the labs.

Summarizing

Philips Medical Systems is a large organization making complex high-tech products. It is situated in a dynamic medical systems environment where there are trends towards a growing market, higher product utilization, increasing product functionalities and increasing pressure on costs. This means product reliability is an important issue. The next paragraph will discuss the importance of field data.

1.1.2 Field data environment

Life data analyses

According to Oh and Bai [5] life data analyses are commonly used to estimate the lifetime distribution of a product and to obtain information on the life characteristic such as reliability, failure rate, percentile and mean time to failure, etc. This information is then used in developing new products or improving the reliability of

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

existing products, in designing burn-in and warranty programs, and in planning the supply of replacement parts. In general there are two ways of performing life data analyses, one is on the basis of laboratory tests, the other one is based on field data.

Laboratory test analysis is not enough
Laboratory tests are performed on the product before it is put on the market. Oh and Bai [5] claim that most of the former works on life data analyses utilized laboratory life test data. However, many times field data is superior to laboratory data because it captures actual usage profiles and the combined environmental exposures that are difficult to simulate in the laboratory and it is more likely to observe longer time-to-failures.

Rai and Singh [6] reinforce this by stating that at development stage various activities including concept/design failure mode and effects analysis, design verification planning and reporting, and robust design experiments are performed. Effectiveness of these activities in developing reliable and robust products is very often judged through laboratory life testing. However, success from laboratory life testing alone does not give design engineers full confidence and feedback about field performance.

Field data analysis
Manufacturing companies collect field failure data of their products for various purposes. Apart from providing the data for any procurement incentive or penalty scheme and for vendor rating, a field failure reporting and analysis system has four main purposes according to Blanks [7]:
a) Obtaining data on which to base corrective action, i.e. action to eliminate, or at least reduce, future failures and to eliminate corrective and preventive maintenance deficiencies.
b) Obtaining data on which to base optimized equipment replacement and overhaul policies and schedules.
c) Obtaining data on which to base logistic support optimization, e.g. the supply of spares and provision of maintenance resources.
d) Obtaining input data for future reliability and maintainability analyses and predictions, e.g. as required in future system planning, in Life Cycle Cost tender validation and in product evaluation.

Run-in test analysis
Between laboratory test analysis and field data analysis, run-in tests can be performed. This is a test performed on a system after production, before it is delivered to the customer. Zaino [8] says the following about this: "typically, the run-in test surfaces early failures that result from assembly errors, calibration or adjustment errors, and/or part defects. This test is not to be confused with Reliability Growth tests, which are conducted earlier in the development cycle on a sample of machines. Reliability Growth tests usually surface design defects and process limitations that often lead to major or minor design modifications. In theory, the run-in test, and the subsequent adjustments and replacements of weak or nonconforming parts, improves the reliability of the individual machine only."

Market trends
Paragraph 1.1.1 showed that field data analysis is important, but it does not explain why this analysis should be performed as soon as possible after market introduction. Time and cost are important factors, as this paragraph will show.

Trends in the market are making reliability prediction using field data nowadays even more necessary then before. Time for extensive laboratory testing of products is not available anymore as is pointed out by Blanks [9]. He states that the speed of evolution of new technologies and new product designs and the market-dictated short lead times between design start and product shipment, in parallel with longer Mean Time To Failure (MTTF) and Mean Time Between Failures (MTBFs), have made it difficult to find the test time to accumulate sufficient failure data to be useful for future prediction purposes. These market trends are also recognized by Sander et al. [10] and Petkova et al. [11]. Ion et al. [12] add to this the uncertainty about the customer use which, together with the mentioned trends, makes it hardly possible to develop and test the product on fitness for use for a reasonable time period (at least the warranty period).

Related to these trends is the cost aspect. The sooner failures in products are recognized, the lower the cost of repairing those failures, since the number of products sold (that need to be repaired) is still relatively small.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Summarizing
These two subparagraphs have indicated the importance of field data analysis in reliability analyses, next to the importance of testing during development, and the need for this type of analysis to be performed soon after market introduction because of the trends in the market.

1.2 Project description

1.2.1 Problem / wish

Paragraphs 1.1.1 and 1.1.2 outlined the surroundings of the project. As shown field data plays an important factor in reliability analysis; 1) shorter development time, faster innovation and uncertainty about the customer use lead to the need for determination of the product reliability soon after market introduction of a product; 2) trends in the market of medical systems indicate a demand for higher product reliability with lower cost of ownership. How to cope with these issues while still maintaining a high reliability level forms the problem. This is the basis for the project aim, which is discussed in the next subparagraph.

Stakeholders and their focus
The two main stakeholders in this project are the Technical University of Eindhoven (TU/e) and Philips Medical Systems. At the Engineering department of 'Cardio Vascular x-ray' at Philips Medical Systems field data is used to get insight in the reliability of the products that are in the market. At the moment, only about two to three years after market introduction it is possible to draw conclusions from the collected data. PMS wants to be able to monitor, control and predict the product reliability in an earlier stage. In that way the feedback loop can be shortened, which leads to faster problem recognition.

At the TU/e this graduation project falls within a research field of the Quality and Reliability Engineering (QRE) department. The aim of the research field is to develop competence in quality and reliability by creating methods to predict the occurrence of product failures in the development process and early in the field introduction. As a first step an understanding of the characteristics of the prevailing reliability prediction models is necessary. This also implies knowledge on how to obtain data of the right quality for these kinds of models. Later in the research program improved models might be introduced to predict the occurrence of failures better.

In an internal report called "Reliability Growth Plan" an outline of the long-term reliability growth plan for the business unit CV is given. A reliability SWOT analysis on skills and knowledge in that report describes the knowledge within the QRE department at the TU/e as one of the opportunities for PMS to obtain knowledge in the field of reliability analyses. This has lead to the cooperation between PMS and the QRE department at the TU/e.

The aim project follows from the two focuses of the TU/e and PMS:

To investigate how the product failure pattern that is found at PMS can be modeled and to assess the prediction performance of those models.

1.2.2 Project research questions

As part of the competence development of the Quality and Reliability Engineering (QRE) group at the TU/e and the Business Unit CV at PMS this project will do three things: investigate the data that is available for making reliability analyses, identify models that can be used for reliability prediction, and perform analyses. These steps can be put as project questions leading to targets.

<u>Data question and targets</u>

1. Is the data that is necessary for making a reliability prediction available?
   - Investigate the databases with field data and determine whether the necessary data for making a reliability prediction is available.
   - Investigate the quality of the data to determine how accurate the results will be.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Model question and target

2. What model(s) are available for the situation at PMS?
   - Identify a reliability prediction model, based on field data, that is applicable for the situation of Philips Medical Systems.

Analyses questions and targets

3. Can these models give the expected early insight into future reliability?
   - Use field data to determine parameter values of the model, determine the goodness-of-fit, and construct the confidence intervals. Use this to determine whether early prediction is possible.

4. What are the improvement possibilities?
   - Make recommendations on the data collection and handling processes and on reliability prediction models to improve the accuracy of the prediction.

Relation with other chapters

The next chapters will try to give an answer to these project questions. Chapter II focuses on question one, chapter III focuses on question two, chapter IV focuses on question three and chapter V focuses on question four. Chapter V will also get back to questions 1-3, giving the overall conclusions on these questions.

1.3 Benefits of reliability models

1.3.1 Field data reliability analysis

To get a better understanding of the specific use of field data reliability models this subparagraph shows what the outcome of such a model is. Chapter III discusses these models in greater detail. But since a clear understanding of what is meant by a model is necessary, a definition of a model is presented first. There are many descriptions of a model given in literature. What they have in common is that they describe a model as a simplification of reality, and therefore necessarily incomplete, focusing on essential elements or characteristics. The definition used in this report, taken from the online dictionary of computing [13], is used since it also takes into account the limitations of a model.

> *A model is a description of observed behavior, simplified by ignoring certain details. Models allow complex systems to be understood and their behavior predicted within the scope of the model, but may give incorrect descriptions and predictions for situations outside the realm of their intended use.*

The outcome of a field data reliability model is both numeric and graphical. The numeric values that can be calculated are:

- Number of failures that the product is expected to have in a certain period of time. For example after one year, two years, and three years.
- The mean time between failures at different points in time.
- The model sensitivity to the data in the form of confidence intervals.

The graphical outcome provides:

- The expected failure pattern of a product.
- The expected moment that the reliability of the product comes into steady state.

The next paragraph shows how these outcomes can be used for specific means. It shows that by making these analyses at several moments in time comparison of the numeric values and graphs will lead to benefits for several departments.

1.3.2 Benefits of reliability models

This paragraph shows the potential benefits of making predictions of product reliability using field data. These reliability predictions based on field data are not a goal by itself. It provides an interpretation of data that can be used, together with other analyses, in order to gain more insight in the product reliability and to take actions based on this insight that is provided.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Determination of current situation / pattern
The first target is to determine the current situation. In order to know where to improve the product reliability there has to be a null measurement of the failure times to determine the current reliability situation. The null measurement gives specific numeric values and a graphical outcome calculated by the model, indicating the reliability performance at that moment. The outcome of this analysis is the reference point for the analyses performed after that.

Evaluation / control
For marketing

- Marketing can get a better understanding of the product reliability and failure behavior. This can help in the risk analysis for bringing a product on the market at a certain point in time. It forms a basis for marketing, together with development, to work on managing possible product reliability risks.

For development

- As mentioned earlier in this chapter tests are performed during development and after production. On the basis of these tests, predictions can be made about the reliability of the system. Field data can be used to check whether these predictions are close to reality or not.
- Reliability improvement efforts that are taken by development can be evaluated regarding their impact on system reliability.
- The failure pattern during the product life can be studied to determine which part of the product life requires the most attention for improvement efforts. Also, the difference in failure pattern for different products can be studied, leading to different strategies for improvement for different products.

For production

- The effects of the run-in tests performed after assembly can be evaluated; helping with the optimization of this test period.
- The same as with the improvement efforts from development also the reliability improvement efforts from production can be evaluated.

For sales / service

- Patterns in failure times can be studied, leading to a better anticipation (understanding) of the number of failures that the customer is going to be confronted with and around what time. The customer support agreements can be adjusted given the outcome of the analyses.
- Influence of planned maintenance influences on the failure pattern can be evaluated, leading to new strategies in the planned maintenance.

Warranty cost assessment
Before a product is put on the market an estimate of the warranty costs is made. Using the predicted number of failures after a certain period of time (1/2/3 years) an evaluation of that warranty cost estimate can be made. This warranty cost evaluation can lead to new insight in the total warranty costs that will be made and therefore can be used as an input for financial decision-making.

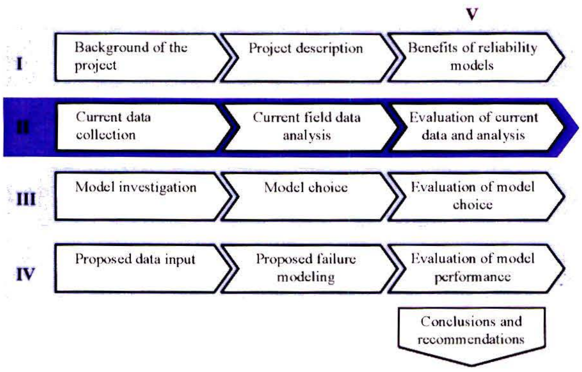
Resource assessment
The necessary effort in product improvement can be assessed at a strategic level. Some products might need large improvement efforts for the product on the market; in other cases effort for a follow up product should receive the most attention.

Learning organization
Comparing analyses of a system that are produced at different moments in time can give insight in the degree the organization is learning.

Basis for root cause analysis
When root causes of dominant failure mechanisms are to be discovered further analysis on a more detailed level needs to be performed. The analyses of failures at system level form a basis for this further research.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Chapter II Analysis of current reliability related processes



"The primary focus of data analysis is on detecting a trend, since one usually hopes that interarrival times tend to become larger, thus indicating reliability improvement"
(H.E. Ascher & C.K. Hansen)

This chapter describes the product and current processes for the collection of data (step 4). Next the current field data analysis is described (step 5), followed by an evaluation of these two steps (step 6). Since all discussed processes are aimed at improvement of the product, some details about the product will be given first.

2.1 Current data collection

2.1.1 The product

This paragraph is an extension of the remarks made about the product in paragraph 1.1.1. It is important to give some extra details on the product that is under consideration in this report to present a better understanding of the complexity of the product.

The product consists of modules, which can vary for different product types, but in general consist of: a stand, a table with controls, monitors in a ceiling suspension, a cabinet with computer hardware for image processing and storage, a cabinet for the x-ray generator, a cabinet with computer hardware for movement of the stand and table (geometry), and operator consoles (see figure 3). Many of these modules contain thousands of parts. Next to the hardware there is a large amount of software necessary for various reasons. There is geometry software for the movement of the stand and table, imaging software for the digital images made by the system, software for administrative purposes, etc.

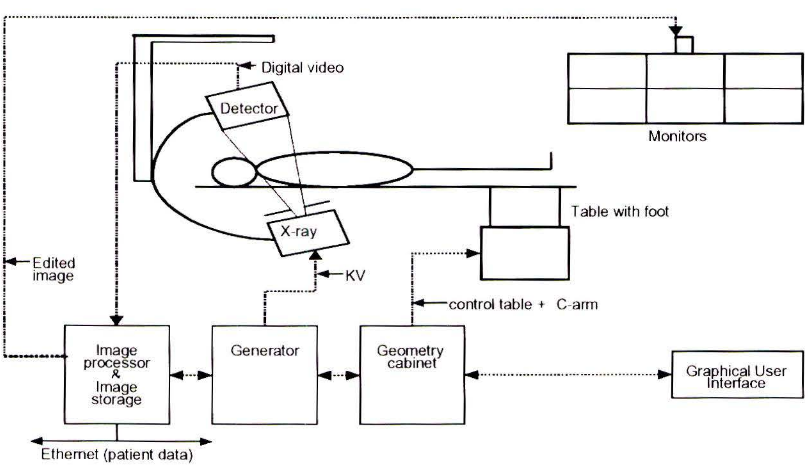


Figure 3 Schematic of the product [14]

A relatively small number of Cardio Vascular X-ray systems are sold around the world, especially when taken into account that there are different types of Cardio Vascular X-ray systems. Although the basic modules are

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

more or less the same for different product types, the customer has several options to customize the system. Next to the choice for monoplane or biplane and the x-ray detector size, a choice can be made between a floor stand and a ceiling stand, a flat detector or traditional image intensifier, the number of monitors, and lots of other options; but also in software different application options are possible.

2.1.2 Feedback loops for reliability improvement

By focusing on finding failures as early as possible steps can be taken in order to try to come up with timely solutions for the failures and thus keeping costs down. There are several loops in the Product Life Cycle that lead to improvement of the reliability; the phases on the Product Life Cycle and the improvement loops connected to these phases are visualized in figure 4.

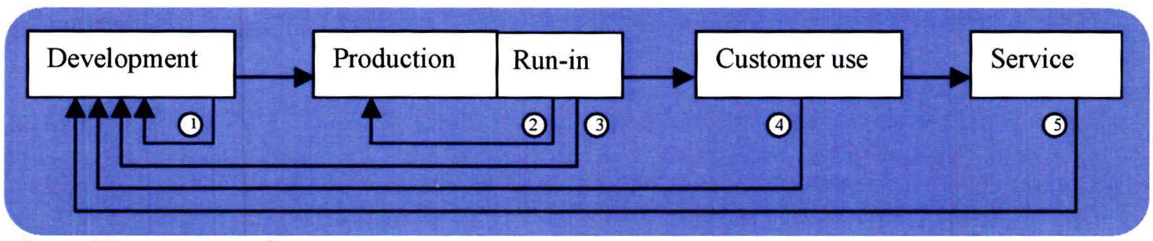


Figure 4 Improvement loops

1. Development loop
   First of all there are improvements through testing before the product is put on the market, see the graduation report of Roos [14]. Since the time to market is forced on the product some known problems are fixed after market introduction, this means that the development department keeps testing and fixing problems after market introduction. This loop also represents the development of new features that are added to the product after market introduction. Next to that there are also International Standards that change over time that lead to necessary changes in the product after market introduction.

2. Production loop
   When production of a new product type starts there is usually a learning curve, which means that the reliability should improve gradually. Every system that is manufactured is being tested during "run-in" tests when it is finished. The testing of these systems is a couple of days. In these tests log files are created which show the failures that occur, that is, the software detectable failures. Next to that the hardware failures are stored in a so called 'Forest Database'. When a failure on a system occurs that is due to manufacturing, the parts that cause the failure are replaced to make sure the failure does not occur again. There are no actions taken toward development for those kinds of failures.

3. Production to development loop
   When production finds mistakes in the product that lead back to design this is reported to development. These failures are found in the same database systems described at the Production loop.

4. Customer use to development loop
   Information from the customer normally comes through the service department (loop 5), however, as mentioned before recently the Field Monitoring project was started. On basis of this information development can take improvement actions. Data from the field monitoring database will be used in the analysis of this project (see chapter IV).

5. Service to development loop
   There are two ways in which service helps the development department with information about the reliability of the systems.
   - First of all the information from all the calls and jobs performed on the systems in the field are collected and used by development to perform analyses on. This data forms the initial data source used in the analysis of this project (see chapter IV).
   - The second way is through a Field Problem Report (FPR). When the service engineer encounters a structural problem an FPR can be made. With a structural problem, a problem is meant that has been

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

encountered several times on different systems. This FPR is sent to the Quality Maintenance Control Board (QMCB) at the engineering department. This board has to decide on the kind of action to be taken as a result of this FPR (see Appendix B).

In the near future there will also be more research performed on 'field returns' as an extra feedback from service to development. This means that engineers from development will investigate the parts that broke down at the customer. Conclusions from these investigations should lead to improvements.

The conclusion of this paragraph is that there are several feedback loops in place for reliability improvement. The loops that are of interest in this report are the "customer use loop" (4) and "service loop" (5). The data that is gathered in these loops and the analyses performed on this data is explained in the remainder of this chapter.

2.1.3 Data collection

Limnios and Nikulin [15] stress the importance of field data for making a reliability analysis. They state that reliability field data are the core of any reliability analysis. For systems in the field, these data are the basis for a realistic estimation of the achieved reliability level. For new designed systems they are used for a realistic evaluation of the reliability level. Field data are usually collected during a long time, for different climatic conditions, for different system configurations, etc. The problem is how to aggregate as much reliability information as possible, keeping control over quality of these data.

Field data that is used for reliability analysis at PMS is collected from two different data sources. One source provides data that is directly obtained from systems at the customer, arrow 1 in figure 5; the other source provides data obtained through two data collection point at the service organization in the form of call data (arrow 2) and job data (arrow 3). The next subparagraphs discuss each data collection point. Both these data sources are used for analysis in this project.

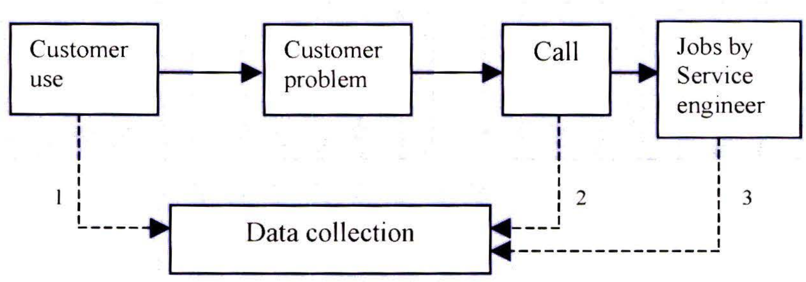


Figure 5 Points of data collection

Customer use

Recently a Field Monitoring project was started at PMS. A direct link to about 20 sites at this moment gives data about the activities performed on the product. Log files are generated which show the exact moment in time that a failure occurs and what kind of failure it is. Every day the log files of the day before are retrieved from the systems and put in a database. Problem with this data is that not only failures are logged but also much more details of things happening on the system. This makes filtering of the data necessary in order to be able to use the data for analysis.

From the log files a list is created with all failure types that have occurred at some point in time. At the moment the total number of failure types is quite large. System experts review this list to point out the relevant failures types as will be defined in paragraph 3.2, leaving a list of relevant failure types that is about 14% of total failures types at this moment. This list is used to identify the number of relevant failures per system.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Unfortunately only the last 24 hours online time are logged, meaning that if the system is online for longer than 24 hours there is loss of data. The total online time of the system is still recorded but information about the activities that take place before those last 24 hours is lost.

Next to that these software loggings are not able to identify all failures that occur on the system since the loggings are unable to detect failures that have no link to the software whatsoever, like a broken remote control for example.

For the analysis of failure moments the current procedure is as follows: the days on which a relevant failure takes place are identified; for those days the FMT database only registers the first failure moment that occurs on a certain day. Other possible failures that occur that same day are not recorded. The failure time is measured in system online hours since the installation of the system.

Customer problem
A customer has a certain goal with the product that has been bought. For example, it has to work fast and efficient. There are three ways of satisfying a customer, and therefore three areas in which a customer can become dissatisfied. These are:

- Functionality
  A customer has a certain idea of what the product must be able to perform, what functions there must be on the product.
- Reliability
  Next to having the desired functions the product should also be able to perform these functions without problems during a specified period of time; it should be reliable.
- Serviceability
  In the situation that the product does fail, the customer wants the problem fixed as soon as possible and in such a way that the problem does not come back any more.

When does a customer place a call?

- Perceived functionality problems
  When a customer is dissatisfied with the performance of the functions of the product a call can come in. This does not mean that something has stopped working, just that the customer is not satisfied with the way it is working. For example, this can happen because of wrong use of the product or because of wrong system installation / fine-tuning. Depending on how big the difference is between the expected functionality and the perceived functionality a call will be made.

- Failures
  When there is a product failure there are two possibilities. The first possibility is that there is an unacceptable failure. This kind of failure is reported right away since the product needs to be working properly. The other kind of failures are so called 'intermitting' failures, these are problems that occur once in a while and they can usually be solved by rebooting the system. When the user thinks the moment has come to report these failures the 'call' to the service desk will be made. In this situation it is not known when the first time was that the failure took place.

Warranty period / Customer Support Agreement
The warranty period and Customer Support Agreement (CSA) are important factors related to resolving a customer problem. The warranty period for the products is one year. During the warranty period both preventive maintenance and corrective maintenance is performed, as well as service on how to operate the system. After the warranty period there are Customer Support Agreements (CSA) in different gradations, starting with 'only preventive maintenance' to 'full maintenance' contracts. Everything that is not in the maintenance contract is paid repair. This has influence on the customer behavior for what to do when the customer has a problem. If the customer is under warranty, or has a full maintenance contract, he will not hesitate to make a call. When the customer only has a preventive maintenance or no service contract anymore, meaning all repairs are paid repairs, then this customer might wait with repairing problems or tries to find cheaper service outside Philips. The most reliable field data from service therefore comes from systems under warranty or CSA.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Call

When a corrective maintenance call comes in the employee puts a short description of the customer problem in the local service database. Next to that each call gets a priority code on the following scale based on seriousness of the problem:

1. Procedure in process; this means that the system stopped working while a procedure on a patient was performed.
2. System down; this means that the system has stopped working totally, without a patient being in danger.
3. System restricted; this means that the system has lost part of its functions but is still operational.
4. Intermittent problem; this is a problem that occurs occasionally. Since the cause of an intermittent problem is probably not changing, it could be counted as just one failure (instead of taking into consideration all calls made about this intermittent problem). Right now all calls for intermittent problems are registered and treated separately.
5. Schedule activity; this is a minor problem that does not need direct attention.

The date on which the call is made and the closing date of the call are also put into the database. These open and close dates of calls are not always representative of the moment work is performed. Calls of planned maintenance activities can be open long before the planned maintenance is performed; closing of calls is sometimes done days or even weeks after the job of that call is performed. Before a service engineer is sent to the customer a first diagnosis of the system is made by use of 'remote service'. This means that the service organization can connect to the system remotely from the SSD. The calls made by customers can be divided in five categories. These categories are explained next.

Installation activities

When a new system arrives at the hospital it is installed by Field Service Engineers. The total installation can take several weeks since it involves the complete setup of the operation room and the control room. Furthermore, the software of the system needs to be completely set to the customer's whishes.

Field Change Orders and Upgrade kit

There are two ways of product improvement when the product is in the market, an upgrade kit and a Field Change Order (FCO). When development comes up with an improved version of hardware or software, or with a new feature, these options can be offered to the customers in the form of an upgrade kit. This means that improvements are system dependent; this is no overall upgrade of all systems in the market at a prespecified moment.

It can also happen that PMS decides to install a new version of hardware or software on the installed base (the systems on the market). This is done in the form of an FCO. The cost of service calls (and the possible loss of customer satisfaction) versus cost of the FCO lead to the decision on whether or not to carry out the FCO. Hardware changes will generally only be made when the installed base is still small. For software, the size of the installed base is not such a big issue. Since the FCO's are initiated by PMS this upgrade is an overall upgrade, the moment on which the FCO is performed in the hospital depends on the available time and regulations at the hospital. PMS tries to execute the FCO on all systems that qualify for the upgrade within a given timeframe.

Planned Maintenance / Preventive Maintenance

This sort of maintenance is arranged in the CSA contract with the customer. Together with the customer a schedule is made for this type of maintenance. As an indication a preventive maintenance frequency of once every six months can be used. The relation with the Call Rate is absolutely present according to a service department employee. This means that systems that have had no planned maintenance in a long period of time have more failures than systems that did have planned maintenance.

Corrective maintenance

Corrective maintenance does not always mean that parts have to be replaced. The problem can be in the software, which does not lead to replacement of a part. Next to that a problem is sometimes solved through a calibration. When the corrective maintenance does concern part replacement there are several possibilities according to the Handbook of reliability engineering [16]. It states that in the case of a repairable product, the behavior of an item after a repair depends on the type of repair carried out. Various types of repair action can be defined:

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

- Good as new repair
  Here, the failure time distribution of repaired system is identical to that of a new system, and successive failures are modeled using an ordinary renewal process. In real life this type of repair would seldom occur, since only a part of the system is repaired, not the whole system.
- Minimal repair
  A failed system is returned to operation with the same effective age as it possessed immediately prior to failure. Failures then occur according to a non-homogeneous Poisson process with an intensity function having the same form as the hazard rate of the time to first failure distribution. This type of rectification model is appropriate when system failure is caused by one of many components failing and the failed component being replaced by a new one.
- Different from new repair (I)
  Sometimes when an system fails, not only the failed components are replaced but also other that have deteriorated sufficiently. The mean time to failure of a repaired system is assumed to be smaller than that of a new system. In this case, successive failures are modeled by a modified renewal process.
- Different from new repair (II)
  In some instances, the failure distribution of a repaired system depends on the number of times the system has been repaired. The mean time to failure decreases as the number of repairs on the system increases.

The kind of repair performed at systems under consideration is minimal repair, since it's not the whole system failing but just a part of a very complex system.

Customer support activities and visits
These are activities that do not involve maintenance. Examples of customer support activities are: helping out with conferences given in hospitals where the system is used, instructing personnel on using the system, etc.

Job by Field Service Engineer
The Field Service Engineer (FSE) performs the jobs as described in last paragraph. The jobs performed for a call are linked to the call through an ID number. When a job is completed the FSE fills out the jobsheet. The following details can be reported in the jobsheet:
- Operations; this is an overview of the hours worked, including open and close date of the job.
- Codes; these are dropdown menus of the cause, damage and activity of the job.
- Parts; details of part(s) replaced.
- Exceptions; space where exceptions can be put.
- Notes; space where the FSE can type notes on the corrective action performed.

The details in the jobsheets are not always complete; the cause of the problem and the damage are sometimes left open, but also examples of hours that are not filled out exist. A possible reason for not filling out the cause of the problem or the damage is that in these dropdown menu's the choice that be made does not always reflect what the FSE wants to put in. Next to the fact that jobsheets are not always filled out completely it is also not possible for the FSE to provide all the preferred data, like the total online time of the system until failure.

The corrective maintenance jobs performed by a service engineer where parts have failed lead to repair or replacement of the part. Parts that are replaced are replaced with the newest version of that part, this means it is not only a replacement but also an upgrade. Repairing a part does not mean 'on-site repairing'. The part that has failed is replaced with a 'new part' and taken back to the plant for repair. There the part is repaired and used as 'new part' in another system. On a database system there is an Electronic Spare parts Catalog that gives a full list of all parts that can be ordered by the service engineers. In this catalog it is stated whether the part has to be repaired or replaced/consumed.

A general rule for the choice between repairing and replacing is based on cost. This means that if the cost of repair is lower than the cost of replacement, the part will be repaired. An exception is made for Table Side Operation modules (TSO's), these modules have a so-called "biohazard" which means that they would need to be treated with special cleaning materials when being repaired. Therefore these parts are always replaced, not repaired.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Field data collection databases

Each call, with its accompanying jobs, is filed in a local database system at the SSR. At the moment it is not the same type of database system in every country, causing problems with transferring this data. Data from these local systems is put in the Global Data Warehouse (GDW), but only from the countries where the interfaces make data transfer to the GDW possible. Even for those, the interfaces between these database systems are not optimal. The notes made in the jobsheet are often not copied completely since there is a limit on the text transferred to the GDW. Also, the hours reported in the jobsheet are not always the same as the hours recorded in the GDW.

Next to problems with transferring data from one database system to another, there are also fields that are no longer included in the GDW. The priority code, the cause, damage, and activity of a call are not transferred to the GDW. On top of all this the language in the jobsheets is the native language of the place where the jobsheet is from. This means that when the data is collected in the GDW several different languages are put in.

In the USA there has been a change in local database system: since end of 1996 field data have been retrieved from the Fieldwatch database from PMSNA (Philips Medical Systems in North America). These data have been analyzed and published quarterly till end of 2000. Since September 2002 jobsheet data (from Jan 2001 onwards) from PMSNA are available via the Global Data Warehouse (GDW). One remark has to be made concerning the data from the USA. Systems are sold in the USA through PMSNA and through dealers who buy from Philips for a group of hospitals. The data is available from PMSNA; dealer data are not available. European data are only available from a number of countries. This has to do with differences in database systems, which makes data conversion not always possible.

2.1.4 Data quality

Studies described in literature found that often many end-users, including managers, are unaware of the lacking quality of data they use in a data warehouse [17]. According to Blischke and Murthy [18] some of the principal difficulties frequently encountered in service call data are as follows:

- Inaccurate, incomplete data – missing or incorrect entries; transpositions, and so on.
- Delays in reporting – periodic or haphazard reports of calls.
- Lags in making calls (particularly for minor failures or failures that do not seriously affect operation of the item)
- Invalid calls, calls after expiration of the warranty, failures due to misuse, or calls on items that did not fail.
- Valid calls that are not made – ignorance of warranty terms; compensation deemed not worth the effort of collecting, and so on.

In the previous paragraphs data quality problems were mentioned throughout the text. This paragraph puts these data quality problems into a framework using a categorization described in literature. This provides a way to structure the problems in order to find solutions for the data quality problems. There are many descriptions of data quality metrics in literature sometimes using up to seven [19] or even twenty [20] metrics. However, the four most important metrics found in literature are given by Ballou and Pazer [19]: data completeness, consistency, timeliness, and accuracy. These four metrics are described in this paragraph giving the problems related to these metrics. In paragraph 2.3 conclusions are drawn from these data quality problems.

Completeness

Definition of completeness: Presence of all defined content at both data element and data set levels [21].

Completeness problems in service data

Data set level:

- Systems sold through dealers are not in the GDW database.
- Transition to a new database system in the USA in 2001, causing that only data from after January 2001 is available in the GDW.
- Data in Europe is only available from a number of European countries.
- No data available from the Asian countries.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

- No complete data from systems that are not under warranty or CSA, and therefore not able to model the failure behavior of the total system life.

Data element level:
- FSE not filling out a jobsheet completely.
- Jobs registered in the SSR database, which are not found in the GDW database.
- Details from jobs in SSR database that are not transferred completely to the GDW database (e.g. notes made by the FSE).
- Fields in the local SSR database that are not transferred to the GDW (e.g. priority code).
- The time that the system has been online until a failure takes place cannot be recorded.
- There are jobs in the GDW database without calls and calls without jobs.

Completeness problems in FMT data
Data element level:
- Only some systems types are monitored by the field monitoring team. This means that there is no FMT data from all the older systems that are in the field (which is the large majority).
- At most one failure per day is registered.
- Potential loss of data when system is online for more than 24 hours.

Consistency
Definition of consistency: format and definition uniformity within and across all comparable data sets [21].

Consistency problems in service data
- The information in the jobsheets can differ from one SSR to the other since different systems are used for data storage in the different regions.
- Textual data arrive in the native language of each country.
- There are differences between the product types which have an influence if all this data is treated as one set.
- The fact that, as mentioned earlier, some fields in the SSR database are not transferred to the GDW database leads to a consistency problem since different call priority levels are not visible anymore.
- Different FSE's throughout the world operating differently; different customers throughout the world reacting differently on problems.
- CM calls that are in fact "site visits / customer visits".
- Installation activities, planned maintenance and FCO calls incorrectly booked as a CM call.

Consistency problems in FMT data
- There are no consistency problems found for the FMT data.

Timeliness
Definition of timeliness: extent to which the age of information is appropriate for the task and user; where age is the amount of time that has passed since the information was produced [22]

Timeliness problems in service data
- No 'live' data is available since the GDW is built offline and only uploaded with data once a month.

Timeliness problems in FMT data
- There are no timeliness problems found for the FMT data.

Accuracy
Definition of accuracy: extent to which data represents what it is supposed to represent [22]. In other words, the degree to which something is correctly documented in the data.

Accuracy problems with service data
- Calls that stay open long after job has been performed.
- Moment on which the failure is recorded can be impure, depending on when the call is made.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

- There are cases of an incorrect number of hours recorded in the GDW database for corrective maintenance activities; it is not the same as the number of hours actually booked.

Accuracy problems with FMT data

- The relevant failure types are identified by experts, which does not rule out the there are relevant failure types that are not identified or that a failure type identified as being relevant is not.

Possible causes of these data quality problems

The overview in these last subparagraphs shows that there are data quality problems related to all four metrics. These data quality problems are divers and have different causes. The causes can be divided into human errors, database and data transfer errors, and procedural errors.

- Human errors, like not filling out the job sheet completely, or filling it out the wrong way, cause their problems mostly in the consistency and accuracy of the data.
- Database and data transfer errors, like the data from several countries that cannot be taken into account, and the problems with incomplete or incorrect data transfer, cause problems in the completeness and consistency of the data.
- Procedural errors, like keeping a call open too long, only updating the database once a month cause problems in the consistency and timeliness of the data.

In order to resolve the problems with the databases and the data transfer between the databases action is being undertaken at the moment. Worldwide the service departments are converting to the same database system (SAP) which will have a positive effect on the completeness and consistency of the data. The Market Essential Harmonized Installed base, Reliability & Performance Reporting (MENHIR) project within PMS provides requirement specifications to have better (central) control of the reporting tools and methodologies in the new IT Landscape (SAP and the SAP Integrated Client). This is explained in more detail in the MENHIR report [23].

For the Human errors and procedural errors no specific structural plans for improvement are in place, at least not to my knowledge. These errors are likely to be more difficult to resolve, especially because they cannot always be seen as errors. Cultural influences, priority considerations, lack of diagnosis tools, etc. are possible reasons for the 'errors'. Making sure for example that every FSE processes jobs in the job sheet in exactly the same way all over the world is a big challenge.

2.2 Current field data analysis

Now that the recorded data has been explained, together with the problems in data quality, this paragraph shows the current analyses that are performed by PMS based on the data recorded in the GDW.

Data analysis

Sanity check

The data is first checked for calls without a job and jobs without a call, these are removed from the database. This is done under the assumption that a call has at least one job and a job belongs to exactly one call.

Filtering for Call rate analysis

The following filters are being used on the data in the database:

1. Only the Integris family and its successors are considered. This means that systems older than the Integris family that are still in the field are not considered. Neither are the systems from other suppliers that are serviced by PMS, for example a GE system. A list of system versions that is considered can be found in appendix C.
2. Calls per month are calculated for systems that are at maximum 3 years old and that are supported under Warranty or under a Customer Support Agreement (CSA) [24]. The reason for this is that the most reliable data is obtained from this group. For the systems that are under warranty or CSA calls will be registered when something is wrong, when there is no more warranty or CSA this cannot be assumed.
3. Systems that produce no calls in a single year are filtered out under the assumption that these systems are not in use.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Call rate analysis I MAT (Moving Annual Total) Call rate
Definition of Call rate according to the PMS Business Balance Score Card, where KPI stands for Key Performance Indicator:

> *The number of corrective maintenance calls per operational system year for systems with an age of 3 years or younger. This KPI is an internal measure for the quality of products/systems and is a leading indicator for Customer Satisfaction.*

In formula, keeping in mind the filter explained above:

$$\text{Call Rate I MAT} = \frac{\text{Number of calls per month over last 12 months}}{\text{Total number of systems under warranty over last 12 months}} \qquad (2\text{-}1)$$

This total number of systems has to be specified in more detail. For a system that is only one month old it is counted for as 1/12 system. Hence, only a system that is at least one year old counts for a full system. This MAT Call rate shows what has happened on average in the last 12 months. An example of this analysis is given in figure 6, taken from the Quarterly report [24].

On the horizontal axis of the figure is the time measured in quarters of a year. The vertical axis shows the call rate I (MAT). In the figure there are three lines representing data from two different geographic sections (North America and Europe) and a line for the world average.

What can be derived from this graph is that in the first and second quarter of 2002 the call rate in the United States went up and in the third quarter of 2003 the call rate started going down again, while in Europe the call rate kept stable during that period. It is not possible to see in this graph whether the system that caused this increase were in the beginning, middle, or towards the end of their technical lifetime.

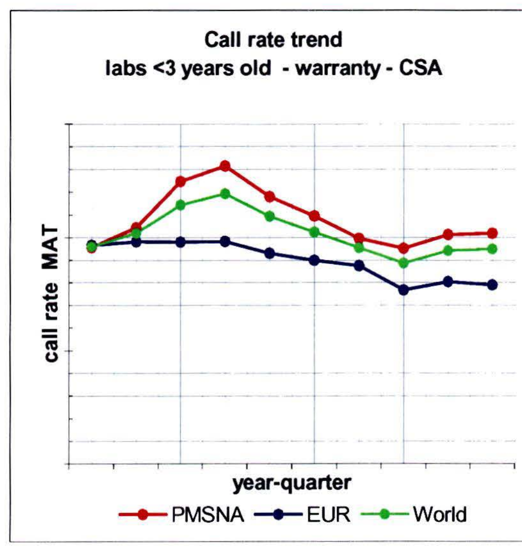


Figure 6 Call rate I MAT

Call rate II analysis "Call rate in the lifetime"
In this calculation systems with a maximum of 3 years of age are used, but to get more data from systems in the last age category, systems from a 42 months period are considered in the calculation. From that data systems are grouped by age. This means that all systems that are at least one month old are grouped, all systems that are at least two months old, etc. From these system groups the total number of calls per month is determined. This number is then divided by the total number of systems in that group. An example of this analysis is given in figure 7, taken from the Quarterly report [24].

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

On the horizontal axis the months in operation are given. The vertical axis shows the call rate II, in calls per month. This graph shows a decrease in the call rate as the systems are in operation for a longer period of time.

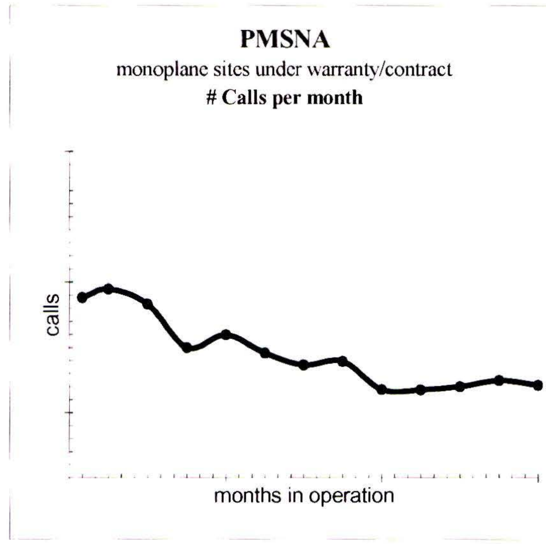


Figure 7 Call rate II

Other kinds of analyses

Material (pareto) analysis

To get insight in the material used during maintenance the 10 highest scores on material quantity and the 10 highest scores on material exchange costs are determined. This information is put in Pareto diagrams; the items in the diagrams are ranked on total exchange costs: Material and labor.

Top X list and Top Q list

Every Sales and Service Region keeps a list of the biggest quality related problems (Top X list). The Business Line Customer Service has a similar list. Together these lists form the Top Q list with priority problems, when developing structural solutions for problems in the field are concerned.

2.3 Evaluation of the current data and analysis

This paragraph sums up the most important statements made on data collection, data quality and current data analyses in this chapter. Next to that a short summarize is given of what has been discussed so far and how this fits in the report.

2.3.1 Important statements

Product related considerations

- The cardio vascular x-ray products are very complex, consisting of a large amount of modules. Furthermore the customer has many adjustment (fine tuning) options making it a customized system. These differences between systems make overall comparison less accurate.
- The product is a so-called "repairable product".
- The economic lifetime of the product is very long (sometimes more than 20 years). During its lifetime numerous improvements and updates are made on the system, so the systems do not remain the same.
- The products are being sold in relatively small numbers.

Reliability feedback loops

- There are five general feedback loops for reliability improvement identified from which the data of the 'customer use to development loop' and the 'service to development loop' are analyzed in this report.

Current data collection

- There are two sources for the collection of data: 1) service data from the Global Data Warehouse; 2) field monitoring data from the FMT database.
- FMT data is retrieved directly from the systems at the hospitals, it gives the failures noticed by software and the moment of failure is given with great accuracy in terms of online time.
- There are five major activities performed by Field Service Engineers: installation activities, FCO's, planned maintenance, corrective maintenance, and customer support activities. From these five activities

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

the interest of this report goes to the corrective maintenance activities since system failures are related to this type of activity.

Data quality

- The data quality problems are classified according to four data quality metrics: completeness, consistency, timeliness and accuracy.
- The causes of these data quality problems can be divided into human errors, database and data transfer errors, and procedural errors.
- The database and data transfer errors are worked on at the moment in the MENHIR project.
- The data quality problems with service data are much more extensive than the problems with the FMT data. Reasons for that are that the service data requires much more human input collected in several independent databases.

Current field data analysis

- The failure analyses that are done:
    - Call rate analysis I; moving annual total.
    - Call rate analysis II; call rate in the lifetime.
- The failure moment is estimated by the closing date of a call, not by the start date. There can be a large difference between these dates, sometimes several months. This influences the call rate calculated. This is analyzed in more detail in appendix D.
- These analyses are overall analyses, not per product type. This gives only a very general picture of the performance of all product types together.
- Call rate analysis I gives no insight in product reliability, only insight in overall number of calls of all systems in the field during a certain month. More about the disadvantages of the use of call rate for reliability purposes can be read in the article by Ion et al. [12].
- Data is aggregated into calls per month, which makes the Call Rate calculations less accurate. For calculations of mean time between failures this aggregation would make calculations very inaccurate.
- The analyses only look back in time, they are evaluative of nature; statistical predictions cannot be made for the future.

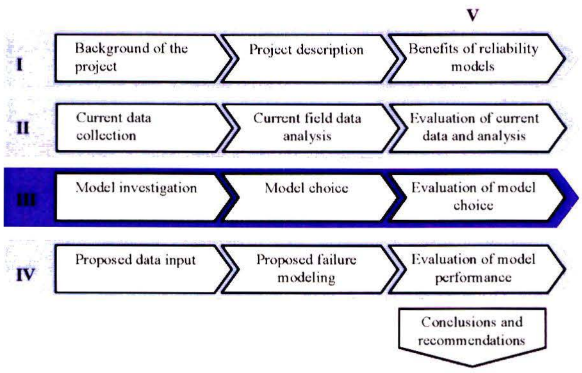
2.3.2 Summary

This chapter provided insight in the field data that is available for analysis and explained the current way of field data analysis. Two data sources are available for failure analysis, service data and FMT data, both have data quality problems which need to be kept in mind when drawing conclusions based on this data. The current field data analysis is not able to provide the desired reliability prediction capabilities in order to satisfy the project aim described in paragraph 1.2. Therefore, chapter III will describe models found in literature and uses a data analysis to determine what kind of models can be used in the situation of PMS. The study is not exhaustive and the many references in the chapter give the opportunity for further reading in this field. The focus is on models that are most used and therefore well described in literature. This choice is made with the reasoning that first a good understanding of more popular models should be obtained before the more exotic models are studied. Chapter IV analyses the prediction performance of the chosen models.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Chapter III Model investigation and choice



For a repairable system, one is rarely interested primarily in time to first failure. Rather, interest generally centers around the probability of system failure as a function of system age.

(L.H. Crow)

This chapter starts with a literature investigation for finding a model applicable to the situation at PMS (step 7). Subsequently the model choice is made based on actual field data (step 8). After that an evaluation of this model choice is presented (step 9).

3.1 Model investigation

Step 7 starts with a broad view on models for field data reliability analysis, followed by an overview of coming to the right model in case of repairable systems. After that two Nonhomogeneous Poisson Process (NHPP) models will be discussed and the tests that need to be performed to assess the accuracy of these models.

3.1.1 Model considerations

There is a large diversity of models described in literature all written from a certain perspective, depending on the system or part under consideration. First of all there is a difference between hardware and software reliability models. This has to do with the different reaction of hardware and software to repair of failures. Debugging in software normally leads to a permanent improvement, other than repair of hardware, which is subject to deterioration [25]. Since the product under consideration is a complex product with a lot of hardware and software, the models designed specifically for software will be excluded. With that the assumption is made that the model used is able to take both hardware and software failures into consideration, although the difference between these two kinds of failures should be kept in mind when drawing conclusions from the results given by the model.

Next is the difference between models for system versus models for parts (components). According to Ascher and Feingold [25] there is a widely held misconception that the same models and data analysis techniques can be used for repairable systems and parts. Parts usually only fail once and are discarded afterwards, and therefore only Time To First Failure is of interest. Models for repairable systems use Time Between Failures since the system can fail more than once. Related to this is the question of what state the system is in after repair, this was discussed in paragraph 2.1.3. This project deals with models for system reliability; models where reliability of parts is added to form system reliability are excluded.

Another decision that has to be made is between discreet and continues models. Discreet models are based on data in which the system has to operate only on discreet moments, like for instance a switch. This situation is not applicable in this report since the system has to operate for a continues time; more on discrete models can be read in [26].

Then there is the choice between parametric models and non-parametric models. The non-parametric models are not based on an expected underlying failure distribution. Non-parametric models use curve-fitting techniques, like regression analysis, see Blischke and Murthy [27], and Arkin [28]. In this project it is chosen to investigate the possibility to fit a model on an expected failure distribution, therefore non-parametric models are left out of the analysis.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

There are models for redundant and non-redundant systems; redundant systems have components in standby mode where there can also be failures of components that are in standby mode. The system at PMS is generally saying 'non-redundant', although there are possibilities build into the system to keep it running without the use of certain functions. This means that certain components can fall out (software/hardware) without the whole systems failing. More information about redundant systems by Seo et al. [29].

For lots of specific situations different kinds of special models have been constructed. For example models including pre-sales data (where there are repairs at the dealer). Unfortunately these models are only described for non-repairable systems in literature, see Majeske [30]. There are two dimensional models especially for the car industry where it is possible to look at both mileage related data as well as time related data. More on two-dimensional models by Chen [31] and Lu [32]. And models using neural networks for system reliability. These models can be found for example in an article by Xu et al. [33].

Summarizing, the models that this chapter will look at are continues parametric models for systems with a mixture of hardware and software, without redundant parts. The next paragraph will give a framework that can be used for the identification of a proper model in this situation.

3.1.2 Framework for continues parametric models

Model identification framework

In figure 8 an overview of continuous models used for reliability prediction is given. As can be seen in this figure several subdivisions in these models can be made. The most important subdivision is repairable versus non-repairable. This is in line with what Crow says in his article [34]: many systems can be categorized into two basic types; one-time or non-repairable systems, and reusable or repairable systems. This will be elaborated on in the next subparagraphs.

The reason to leave this choice explicitly in the model is because the project started with the objective to validate a reliability prediction model for continues parametric systems developed at the TU/e.

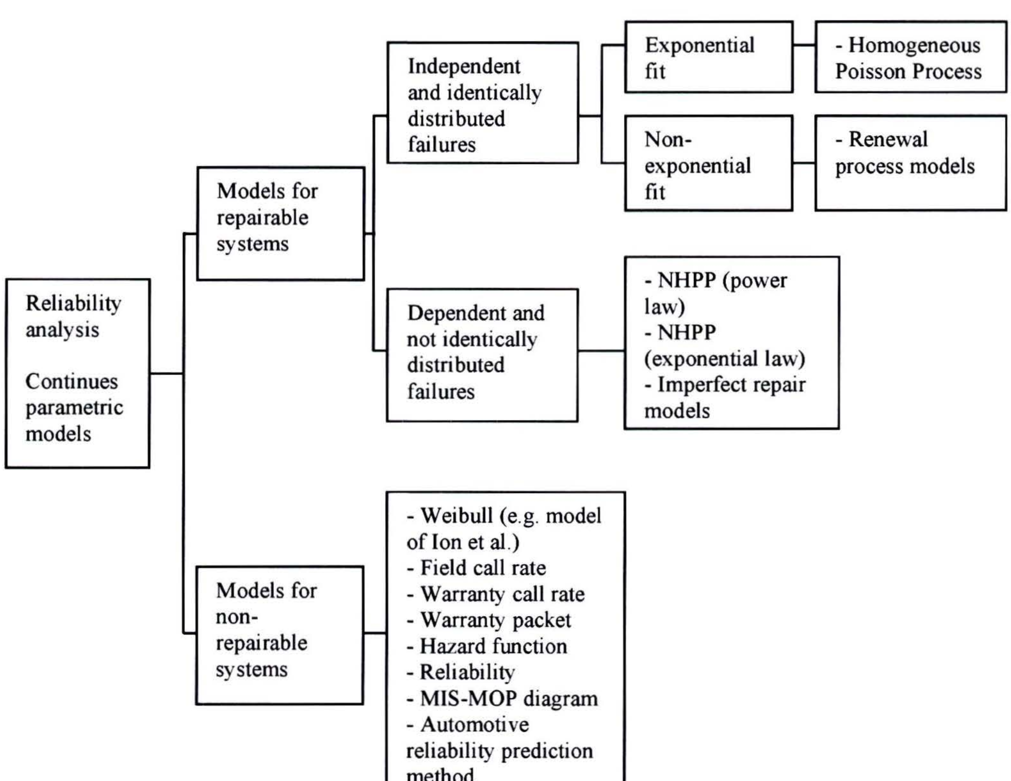


Figure 8 Model identification Framework

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Non-repairable systems
Models for non-repairable products include only Time To First Failure since the product is discarded right after this failure. Or the assumption is made that only the time to first failure is of interest because this one (usually) falls within the warranty period. Models that are used for non-repairable systems are, among others, the Weibull model [12], field call rate [10], warranty call rate [11,12], warranty packet method [10], hazard function [10,11], reliability function [11], and the MIS MOP method [10]. In the conclusions of the previous chapter it is said that the product under consideration is a so-called 'repairable system'. This means that the non-repairable systems models cannot be applied in this situation.

Repairable systems
When a system can be repaired other models are used. For example in "Modeling the reliability of repairable systems in the aviation industry" [35] this is the subject of study. This paper states that in complex machinery such as a jet engine, the systems are generally not replaced but are repaired when they fail. In this case, the usual non-repairable methodologies are simply not appropriate for repairable systems and the renewal process should not be used since the required refurbishment will typically not achieve a "same-as-new" status [34].

According to Crow [34] one is rarely interested primarily in Time To First Failure for a repairable system. Rather, interest generally centers on the probability of system failure as a function of system age. Exact reliability analyses for complex, repairable systems are often difficult because of the complicated failure process that may result from the replacement or repair policy. A common procedure in practice is to approximate the complicated failure process by a simpler failure process, which although not exact, still yields useful practical results. One such approach assumes that the failure times of the complex repairable system follow a Non Homogeneous Poisson Process (NHPP). Crow developed the Power Law NHPP as a model for the reliability of a complex, repairable system when data are generated from multiple systems. Another NHPP model often seen in literature [35, 35] is the Exponential Law model.

3.1.3 General procedure for analyzing failure data of a repairable system

When deciding among the three most basic model categories, Homogeneous Poisson Process (HPP), Renewal Process (RP), and Non Homogeneous Poisson Process NHPP, the procedure outlined in figure 9 can be used. This procedure is proposed by Ascher and Hansen [37] and will be explained in this paragraph.

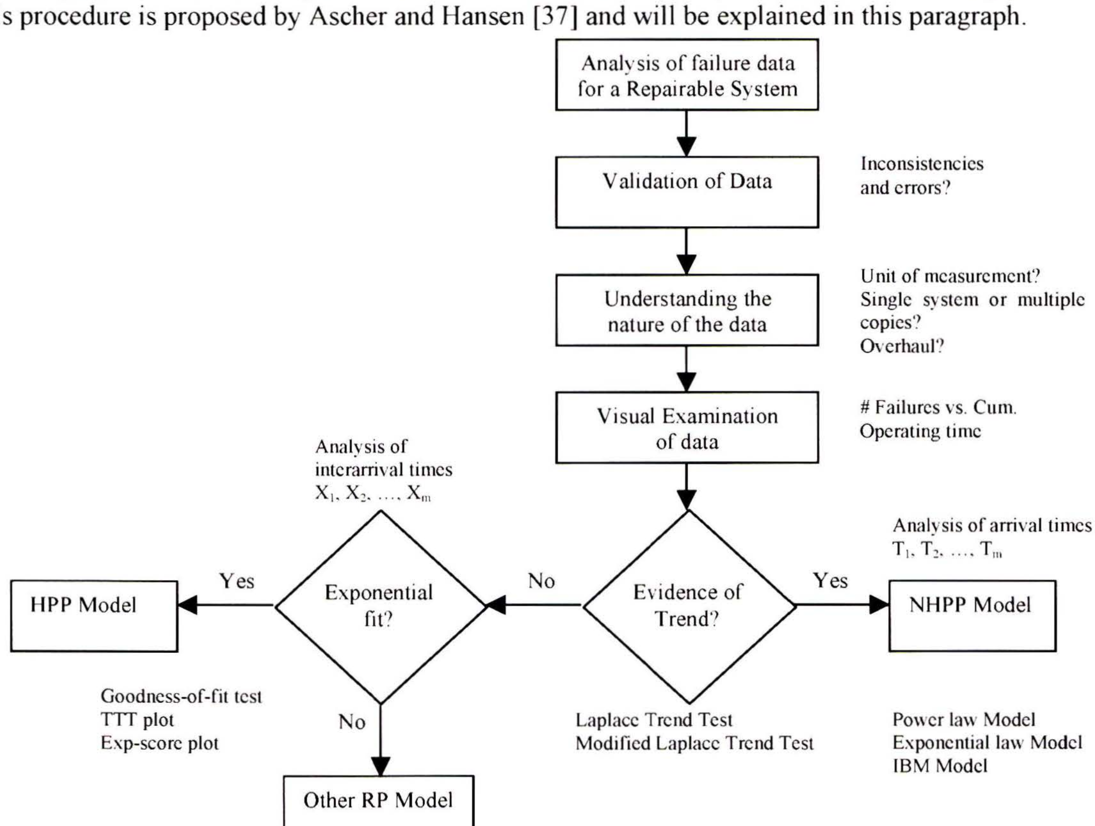


Figure 9 General Procedure for analyzing failure data of a repairable system

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Validation of data
This first check of the data is used to track down any inconsistencies and errors in the data. Through this check conclusions can be drawn on the quality of the data. This can lead to a number of assumptions that have to be made about the data. If too many inconsistencies and errors are found the conclusion can be drawn that the data is not useful for analysis.

Understanding the nature of the data
Single system or multiple copies?
This consideration speaks for itself. A choice has to be made between including a single system or multiple copies. For both choices different models are available.

Unit of measurement
In [36] Lindqvist says the following about this "one should remember the fact that the choice of time scale influences the pattern of failures. Using calendar time, operating time, mileage or cumulative repair cost as the time scale for a car will probably give quite different patterns of failures." This shows that the choice for unit of measurement should be made explicitly.

Overhaul?
It should be established whether or not situations of overhaul are included in the data analysis. Depending on whether or not overhaul is included different assumptions on repair, and therefore on models, should be made.

Visual examination of data
For this step there are several graphical techniques that can be used, Ascher and Feingold [25] mention (1) plotting cumulative failures versus cumulative time on linear paper, (2) estimating average "rate of occurrence of failure" (ROCOF) in successive time periods and (3) Duane plots. These graphs can be used to help determine whether a system is improving or deteriorating. Such a technique is particularly useful for seeking out the data's salient features and for checking the assumptions made in fitting formal models to data.

The interarrival times of an improving (deteriorating) system tend to become larger (smaller); hence a plot of cumulative number of failures on linear paper will tend to be concave down (up) [15].

Evidence of trend
In [37] it is stated that often the patterns of failures show evidence of some trend. They say that in fact, the primary focus of data analysis is on detecting a trend, since one usually hopes that interarrival times tend to become larger, thus indicating reliability improvement; in some cases, improvement is unrealistic so one hopes for renewal rather than deterioration. Kvaløy and Lindqvist [36] give the following definition of **trend**:

There is a trend in the pattern of failures if the inter-arrival times tend to alter in some systematic way, which means that the inter-arrival times are not identically distributed.

In Ascher and Feingold [25] tests are mentioned which have been proposed for distinguishing between an HPP and a monotonic trend. These tests are Laplace, Bates, Bartholomew, Boswell, Cox and Lewis, Boswell and Brunk, Lorden and Eisenberger, Saw and MIL-HDBK-189. In other literature [36], [37], [38] the Laplace test is proposed as a good test to use in this situation.

A generalized formula for trend in [36] makes it possible to look at both time-truncated and failure-truncated processes. Censoring strategies are important because data obtained by different censoring schemes are stochastically different. Hence, data must be treated differently depending on which censoring scheme is actually used. If n = total number of observed failures for a system, then the difference between time and failure truncation is:

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

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

L = the Laplace statistic
T_i = Time to failure i (i = 1, ..., $\hat{n}$)
(a,b] = time interval

In formula:

$$L = \frac{\sum_{i=1}^{\hat{n}} T_i - \frac{1}{2}\hat{n}(b+a)}{\sqrt{\frac{1}{12}\hat{n}(b-a)^2}} \tag{3-1}$$

These processes are observed in the time interval (a,b]. For time truncation the system is observed during a prespecified (operation) time. The observed number of failures is thus a random variable. For failure truncation the system is observed until a prespecified number of failures has occurred. The length of the observation interval is now random.

According to Kvaløy and Lindqvist [36] a straightforward generalization of the single system Laplace test if there are observations from m independent systems is:

$$L_c = \frac{\sum_{j=1}^{m}\sum_{i=1}^{\hat{n}_j} T_{ij} - \sum_{j=1}^{m}\frac{1}{2}\hat{n}_j(b_j + a_j)}{\sqrt{\frac{1}{12}\sum_{j=1}^{m}\hat{n}_j(b_j - a_j)^2}} \tag{3-2}$$

The *j*th system is observed in the time interval (a_j, b_j] with n_j failures occurring at times T_{ij}, i = 1, 2,..., n_j

There is evidence of trend at a significance level α if [37]:
$L > z_{\alpha/2}$ (reliability deterioration)
$L < -z_{\alpha/2}$ (reliability improvement)
The null hypothesis of the Laplace test is the HPP. When the trend test is performed and a significant likelihood of trend is determined, the model that should be chosen is the Non-homogeneous Poisson Process model.

If $-z_{\alpha/2} < L < z_{\alpha/2}$ then there is no evidence of positive or negative reliability growth at the α significance level and the growth analysis is terminated. In this case, the hypothesis of exponential times between successive failures (or a homogeneous Poisson process) is accepted at the α significance level [39].

Data Analysis for exponential fit
In case there is no trend the data should be analyzed for exponential fit. This can be done through analysis of the interarrival times of failures.

3.1.4 Repairable systems models
This paragraph discusses the repairable systems models. First an overview of the symbols used in this paragraph is given.

α_0 = parameter for the exponential law model
α_1 = parameter for the exponential law model
β = shape parameter for the power law model
λ = scale parameter for the power law model
$\mu_1(T)$ = power law intensity function
$\mu_2(T)$ = exponential law intensity function

E[N(T)] = expected number of failures in (0,T)
$MTBF_i$ = instantaneous MTBF

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

$MTBF_c$ = cumulative MTBF
i = index for failure
j = index for system
m = total number of independent systems
n = total number of observed failures for a system
T = Time from start of system life
T_{ij} = Time to failure i of system j (i = 1,..., n), (j = 1,..., m)
N(T) = number of failures in (0,T)

Homogeneous Poisson Process
This model comes about when the interarrival times between failures are independent and identically distributed according to the exponential distribution, with parameter λ. It implies that there is no improvement or wearout with age. This basic model is also known as a Homogeneous Poisson Process (HPP). According to [40] the following formulas apply:

The cumulative distribution function of the waiting time to the next failure (or “interval” time between failures):

$$F(T) = 1 - e^{-\lambda T} \tag{3-3}$$

N(T) = the cumulative number of failures from time 0 to time T

$$P\{N(T) = k\} = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \tag{3-4}$$

E[N(T)] = λT = the expected number of failures by time T (3-5)
λ = the Rate of Occurrence of Failures (ROCOF)
$\frac{1}{\lambda}$ = the Mean Time Between Failures (MTBF) (3-6)

In the HPP model, the probability of having exactly k failures by time T is given by the Poisson distribution with mean λT (see equation 3-4).

Despite the simplicity of this model, it is widely used for repairable equipment and systems throughout industry. Justification for this comes, in part, from the shape of the empirical Bathtub Curve. Most systems spend most of their "lifetimes" operating in the long flat constant ROCOF portion of the Bathtub Curve. The HPP is the only model that applies to that portion of the curve, so it is the most popular model for system reliability evaluation and reliability test planning [40].

Other Renewal Process (RP) models
In Ascher and Hansen [37] it is stated that the RP model implies that the system is restored to an “as new” condition by each repair. Models used in this situation are conventional analysis techniques. Coetzee [38] states that in this situation the data can be reordered in the conventional failure interval histogram. This can then be used to fit a standard statistical distribution, after which, optimization of the maintenance strategy for the component or system can take place.

NHPP Power Law model
The NHPP is a generalization of the HPP that allows for a change in the intensity as a function of system age [35]. Or put in another way, HPP is a process with no *trend*, while the NHPP permits the modeling of trend via the intensity function μ (T); another term used for μ (T) is peril rate or the ROCOF for a nonhomogeneous Poisson process. This way the other parts of the bathtub curve can be modeled.

A well-accepted format of the NHPP model is the ‘Power law process’ (PLP) [36, 38].

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

$$\mu_1(T) = \lambda\beta T^{\beta-1} \text{ with } \lambda, \beta > 0,\ T \geq 0 \quad (3\text{-}7)$$

When the shape parameter β is equal to 1, the power law model reduces to the homogeneous Poisson process (HPP) with constant failure intensity equal to 1/λ. When β>1 (β<1) the intensity function is monotonically increasing (decreasing) with the operating time T: this corresponds to the situation in which the times between successive failures become shorter (longer) with T [41].

Parameter estimation
According to [38] the maximum likelihood estimates for the parameters of the Power Law intensity are:

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ln \frac{T_n}{T_i}} \quad (3\text{-}8) \qquad \text{and} \qquad \hat{\lambda} = \frac{n}{T_n^{\beta}} \quad (3\text{-}9)$$

n = Number of observed failures for a system.
T_n = Total observed time period
T_i = Time to failure i

Parameter estimation in case of more than one system
For the superposition system, there is a "failure" each time any one of the j systems fails. Consequently, the intensity of failure for the superposition system is j times the intensity for each of the systems [34].

$$\hat{\beta} = \frac{n}{\sum_{j=1}^{m} \sum_{i=1}^{n} \ln \frac{T_n}{T_{ij}}} \quad (3\text{-}10) \qquad \text{and} \qquad \hat{\lambda}^* = \frac{n}{T_n^{\beta}} \quad (3\text{-}11)$$

$\lambda^* = m\lambda$
T_{ij} = Time to failure i of systems j
T_n = Total observed time period
m = Number of systems that are evaluated

Number of failures
The expected number of failures is: $E[N(T)] = \lambda T^{\beta}$ (3-12)
Where T = Time

Mean Time Between Failures
There are two types of mean time between failures (MTBF), the instantaneous MTBF and the cumulative MTBF. The $MTBF_i$ is equal to one divided by the model intensity as formula 3-13 shows. This shows the direct connection between the intensity and the $MTBF_i$. The $MTBF_c$ is calculated by dividing the time T by the expected number of failures at time T, as formulas 3-14 shows:

$$MTBF_i(T) = m_i(T) = \frac{1}{\mu_1(T)} = \frac{T^{1-\beta}}{\lambda\beta} \quad (3\text{-}13)$$

$$MTBF_c(T) = m_c(T) = \frac{T}{E[N(T)]} = \frac{T}{\lambda T^{\beta}} \quad (3\text{-}14)$$

NHPP Exponential Law Model
The formula for the Exponential law was introduced by Cox and Lewis [42]. Another name for this model is 'Log-linear model'. According to Coetzee [38] the formula for the intensity function (ROCOF) is:

$$\mu_2(T) = e^{\alpha_0 + \alpha_1 T},\ -\infty < \alpha_0, \alpha_1 < \infty,\ T \geq 0 \quad (3\text{-}15)$$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Where α_0 and α_1 are the parameter for the exponential law model. This format of the NHPP-model models repairable systems well with $\alpha_1 > 0$

Using maximum likelihood estimates, the parameters for $\mu_2(T)$ can be found from:

$$\sum T_i + n\alpha_1^{-1} - nT_n\{1 - e^{-\alpha_1 T_n}\}^{-1} = 0 \tag{3-16}$$

and

$$\hat{\alpha}_0 = \ln\left\{\frac{n\hat{\alpha}_1}{e^{\hat{\alpha}_1 T_n} - 1}\right\} \tag{3-17}$$

n = Number of observed failures for a system.
T_n = Total observed time period
T_i = Time to failure i

The parameter values are found by solving equation (3-16) for $\hat{\alpha}_1$ and then substituting this value in equation (3-17) to solve for $\hat{\alpha}_0$.

Number of failures
The expected number of failures in the interval [a,b] are:

$$E[N(T)_{ab}] = \frac{e^{\alpha_0}}{\alpha_1}(e^{\alpha_1 T_b} - e^{\alpha_1 T_a}) \tag{3-18}$$

Mean Time Between Failures
Here the same types of MTBF as with power law, the instantaneous MTBF and the cumulative MTBF. The formulas are:

$$MTBF_i(T) = m_i(T) = \frac{1}{\mu_2(T)} = \frac{1}{e^{\alpha_0 + \alpha_1 T}} \tag{3-19}$$

$$MTBF_c(T) = m_c(T) = \frac{T}{E[N(T)]} = \frac{T}{\frac{e^{\alpha_0}}{\alpha_1}(e^{\alpha_1 T_b} - e^{\alpha_1 T_a})} \tag{3-20}$$

Parameter estimation in case of more than one system
Since there were no articles or books found for calculating the parameters when including more than one system in the Exponential Law model, this formula has been derived by R. Ion from the TU/e.

Model:

$$\mu_2(T) = e^{\alpha_0 + \alpha_1 T}, \; -\infty < \alpha_0, \alpha_1 < \infty, \; T \geq 0$$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Parameter estimation:

$$\alpha_0 = \ln\left(\frac{\alpha_1 \sum_{j=1}^{m} n_j}{\sum_{j=1}^{m} (e^{\alpha_1 t_{nj}} - 1)}\right) \tag{3-21}$$

$$\sum_{j=1}^{m}\sum_{i=1}^{n_j} T_{ij} + \frac{\sum_{j=1}^{m} n_j}{\alpha_1} - \frac{\sum_{j=1}^{m} n_j \sum_{j=1}^{m} T_{nj} \cdot e^{\alpha_1 T_{nj}}}{\sum_{j=1}^{m} (e^{\alpha_1 T_{nj}} - 1)} = 0 \tag{3-22}$$

These parameters are found in a similar way as the parameters for one system.

3.1.5 Goodness-of-fit test

Two tests for Goodness-of-fit are discussed. First, the generally applicable Chi-squared test, which can be used for the Power Law model and the Exponential Law model. Next, the Cramér-von Mises test is given. This test is applicable only to the Power Law model.

Chi-squared test

The standard X^2 test can be applied to both NHPP models and has the benefit that the start time of the test need not be zero as is the case with Crow's application of the Cramér von Mises test [36, 38]. The X^2 test is applied in the customary way with the expected number of failures in any interval (T_a, T_b) given by:

For $\mu_1(T)$:

$$E[N(T)_{ab}] = \lambda(T_b^{\beta} - T_a^{\beta})$$

For $\mu_2(T)$:

$$E[N(T)_{ab}] = \frac{e^{\alpha_0}}{\alpha_1}(e^{\alpha_1 T_b} - e^{\alpha_1 T_a})$$

According to the book "Applied statistics and probability for engineers" [43] the Chi-Square Goodness-of-fit test is calculated in the following manner:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3-23}$$

O_i = observed frequency in the *i*-th class interval
E_i = expected frequency in the *i*-th class interval
k = number of classes
p = number of parameters
Degrees of freedom = k – p – 1

According to [43] the observed and expected frequencies in a class interval should preferable be at least 3-5. Next to that the class intervals don't have to be of equal size, if a class does not contain enough values the class can be enlarged. The approximation improves as the number of observations increases.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

A frequency distribution that uses either too few or too many bins will not be informative. Choosing the number of bins approximately equal to the square root of the number of observations often works well in practice [43].

The hypothesis that the distribution of the population is the hypothesized distribution is rejected if the calculated value of the test statistic $\chi_0^2 > \chi^2_{\alpha,k-p-1}$.

Yamada and Osaki [44] mention the following: "We should perform a goodness-of-fit test to check statistically whether the applied reliability growth model provides a good fit to the observed data. The well-known and commonly used test is χ^2 goodness-of-fit test. The goodness-of-fit test based on this statistic has the following advantage and disadvantage, respectively:

- The tabulated critical values of χ^2 are readily available for all sample sizes and significance levels.
- The accuracy of χ^2 goodness-of-fit test decreases, as the sample size gets smaller.

Cramér-von Mises test (for Power Law)

According to [39] the Cramér-von Mises statistic is given by the following expression:

$$C^2(n) = \frac{1}{12n} + \sum_{i=1}^{n}\left[\left(\frac{T_i}{T_n}\right)^{\hat{\beta}} - \frac{2i-1}{2n}\right]^2 \tag{3-24}$$

$C^2(n)$ = Cramér-von Mises statistic
n = Number of observed failures for a system.
T_i = Time to failure i
T_n = Total observed time period
β = shape parameter for the power law model

This statistic then has to be measured against the correct value from the table given in appendix E. This table gives the critical values of the Cramér-von Mises statistic for 10% significance level. If the statistic $C^2(n)$ exceeds the critical value corresponding to 'n' in the table, then the hypothesis that the Power law model adequately fits the data shall be rejected. Otherwise, the model shall be accepted [39].

3.1.6 Confidence bounds

With the formulas for the $MTBF_c$ and $MTBF_i$ given earlier in this chapter it is possible to plot the lines of the $MTBF_c$ and $MTBF_i$. It is however also important to know something about the variance of the TBF. With the formulas given in this paragraph it is possible to calculate the confidence bounds for the Power law model. Unfortunately no formulas were found applicable to the exponential law model.

The following symbols are used:
∂ = partial derivative
β = shape parameter for the power law model
λ = scale parameter for the power law model
m_c = Cumulative Mean Time Between Failures
m_i = Instantaneous Mean Time Between Failures
T = Time from start of life
n = Number of observed failures for a system.
Var = variance
Cov = covariance

Before the confidence bounds can be calculated the variance and covariance of the two parameters need to be calculated. This is done using Fisher's Matrix [45].

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Fisher's Matrix:

$$\begin{bmatrix} -\dfrac{\partial^2\Lambda}{\partial\lambda^2} & -\dfrac{\partial^2\Lambda}{\partial\lambda\partial\beta} \\ -\dfrac{\partial^2\Lambda}{\partial\lambda\partial\beta} & -\dfrac{\partial^2\Lambda}{\partial\beta^2} \end{bmatrix}^{-1}_{\beta=\hat{\beta},\lambda=\hat{\lambda}} = \begin{bmatrix} Var(\hat{\lambda}) & Cov(\hat{\beta},\hat{\lambda}) \\ Cov(\hat{\beta},\hat{\lambda}) & Var(\hat{\beta}) \end{bmatrix}$$

Where the following formulas need to be calculated:

$$\frac{\partial^2\Lambda}{\partial\lambda^2} = -\frac{n}{\lambda^2} \tag{3-25}$$

$$\frac{\partial^2\Lambda}{\partial\beta^2} = -\frac{n}{\beta^2} - \lambda T^{\beta}(\ln T)^2 \tag{3-26}$$

$$\frac{\partial^2\Lambda}{\partial\lambda\partial\beta} = -T^{\beta}\ln T \tag{3-27}$$

For the calculation of the confidence bounds (CB) of the $MTBF_c$ the following formulas need to be calculated:

$$CB_{mc} = \hat{m}_c(T) \pm z_{\alpha}\sqrt{Var(\hat{m}_c(T))} \tag{3-28}$$

Formula (3-29)

$$Var(\hat{m}_c(T)) = \left(\frac{\partial m_c(T)}{\partial\beta}\right)^2 Var(\hat{\beta}) + \left(\frac{\partial m_c(T)}{\partial\lambda}\right)^2 Var(\hat{\lambda}) + 2\left(\frac{\partial m_c(T)}{\partial\beta}\right)\left(\frac{\partial m_c(T)}{\partial\lambda}\right)Cov(\hat{\beta},\hat{\lambda})$$

$$\frac{\partial m_c(T)}{\partial\beta} = -\frac{1}{\hat{\lambda}}T^{1-\hat{\beta}}\ln T \tag{3-30}$$

$$\frac{\partial m_c(T)}{\partial\lambda} = -\frac{1}{\hat{\lambda}^2}T^{1-\hat{\beta}} \tag{3-31}$$

The calculation of the confidence bounds (CB) of the $MTBF_i$ is done using the following formulas:

$$CB_{mi} = \hat{m}_i(T) \pm z_{\alpha}\sqrt{Var(\hat{m}_i(T))} \tag{3-32}$$

Formula (3-33)

$$Var(\hat{m}_i(T)) = \left(\frac{\partial m_i(T)}{\partial\beta}\right)^2 Var(\hat{\beta}) + \left(\frac{\partial m_i(T)}{\partial\lambda}\right)^2 Var(\hat{\lambda}) + 2\left(\frac{\partial m_i(T)}{\partial\beta}\right)\left(\frac{\partial m_i(T)}{\partial\lambda}\right)Cov(\hat{\beta},\hat{\lambda})$$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

$$\frac{\partial m_i(T)}{\partial \beta} = -\frac{1}{\hat{\lambda}\,\hat{\beta}^2} T^{1-\hat{\beta}} - \frac{1}{\hat{\lambda}\,\hat{\beta}} T^{1-\hat{\beta}} \ln T \tag{3-34}$$

$$\frac{\partial m_c(T)}{\partial \lambda} = -\frac{1}{\hat{\lambda}^2 \hat{\beta}} T^{1-\hat{\beta}} \tag{3-35}$$

3.2 Model choice

The model choice is made using the general procedure for analyzing failure data of a repairable system as described in paragraph 3.1.3. This paragraph describes the analysis of data that is necessary in order to make a model choice. But before this analysis is described a clear definition of a relevant failure is necessary, this is given in the next subparagraph.

Definition of relevant failures for analysis

In order to be able to perform an analysis on failure data it is important to have a clear definition on what will be noted as a relevant failure. For this it is necessary to first define a failure [13]:

> *The inability of a system or system component to perform a required function within specified limits.*

A definition of a relevant failure used in the IEC 1014 standard [46] is:

> *A failure that should be included in interpreting test or operational results or in calculating the value of a reliability performance measure.*

A more specific way of looking at failure is given in [35]. There it is stated that the uncertainty of an engine failure or removal is dependent on a number of external and internal factors:
- Component-specific factors (design, manufacturing)
- Operational factors (pressure, temperature)
- Environmental factors (ambient conditions, temperature, humidity)
- Maintenance factors (servicing frequency, overhaul strategy)

For this project the above definitions are not giving the essence of a relevant failure. Therefore definitions for relevant failures specifically for this project will be described. There will be two definitions since there are two data sources used for data analysis. Both data sources require their own definition because of the difference characteristics of the sources.

Description of a failure for service data

As described in chapter II the Global Data Warehouse database contains call and job information on installation activities, corrective and planned maintenance, field change orders, and customer support activities. The basic filtering, as PMS performs it, is to use the corrective maintenance calls as a description of a failure.

To clean up these corrective maintenance calls partially some of the inconsistencies are filtered out:
- Filtering out the calls before the start use date
- Filtering out calls without jobs, and jobs without calls

In the remainder of this report this data will be called the unfiltered data since, next to corrective maintenance calls, it still contains incorrectly placed:
- Planned maintenance calls
- Installation activities
- Customer support activities

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

- Field Change Orders

Description of a relevant failure for service data
When the data in the Global Data Warehouse database is given a closer look calls that are no corrective maintenance calls can be deleted from the list manually, leaving only corrective maintenance calls. This will be called the filtered data and was done in the following manner:

- Filter out the Field Change Orders
- Manually filter out all planned maintenance calls, installation activities and site visits incorrectly booked as corrective maintenance.
- Filter out the calls before the start use date
- Filter out calls without jobs, and jobs without calls

Selection criterion used for the moment of failure
Since there is an open and a close date of a call a choice has to be made between them for establishing the failure moment. The influence of choosing one or the other is shown in appendix D. Since the open date is closer to the moment of failure this will be used as the failure moment.

Description of a relevant failure for FMT data
In consultation with persons involved in reliability analysis and the field monitoring team at PMS the customer is taken as the center point for defining a relevant failure. The definition of a **relevant failure** that will be used for data from the FMT database is:

Relevant failures are all failures that a customer <u>can</u> be confronted with.

In this definition reduction of performance of a certain function will not be taken into account while it cannot be retrieved from the available data. The criterion for determination of failures that a customer can be confronted with is based on 'the judgment of development engineers'.

Now that the definition of a relevant failure is established a dataset is chosen to determine the kind of model fits the failure pattern of the systems under consideration the best. As said, this is done using the general procedure for analyzing failure data of a repairable system.

Validation of data
Data sources
Both types of data sources discussed in chapter II are used, service data and FMT data. The FMT data is available late in the project and is only considering a limited time period, since the FMT-project started only recently. Therefore, in order to analyze data of a larger amount of systems over a longer period, service data is used as the primary data source.

System choice
Since there are different system families developed at the Cardio / Vascular X-ray department, the first step is to define which system family to analyze. In table 4 the four general product areas are presented. In the rows a difference is made between cardio and vascular systems, this has mostly to do with the diameter of the x-ray detector of the system; the cardio x-ray system has a smaller detector. The columns monoplane and biplane differentiate between a system with a single stand (one x-ray detector) and a system with a double stand (two x-ray detectors).

Table 4 systems

	Monoplane	Biplane
Cardio X-ray	●	
Vascular X-ray		

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

The systems that will be considered are the Cardio monoplane systems as shown by the dot in table 4. The reason for looking at the cardio monoplane systems is that the field monitoring project was started for collecting field data for the latest generation of this type of systems.

Understanding the nature of the data
The visual examination of the data of multiple systems gives an impression of the different failure patterns that these systems show. The Laplace trend test is performed both on individual systems as on a group of systems as a whole and gives a numeric value to the failure pattern indication significant trend or not. As said earlier, all analyzed systems are cardio monoplane, still under warranty or CSA, and younger than three years.

Time lines
Three kinds of timelines are used in this report: calendar time, operating time, and online time. It is important to give the descriptions of these timelines in order to be clear on what is meant by these terms.

Calendar time
Figure 10 shows the installation of systems in the field and the failure history subsequent to the installation moment. It shows that system number one, for example, has three failures in the observation period. The part of time that these systems have actually been online cannot be seen; it is just the calendar days that the system is in the field after installation. The call rate I (MAT) analysis discussed in chapter II uses this timeline.

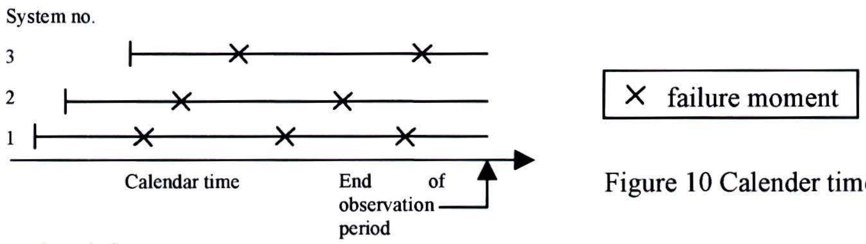


Figure 10 Calender time

Operational time
In figure 11 the same pattern of system history is illustrated, but now in terms of the operational time T. This way a comparison of the systems can be made all having the same age, e.g. all being in their first month after installation.

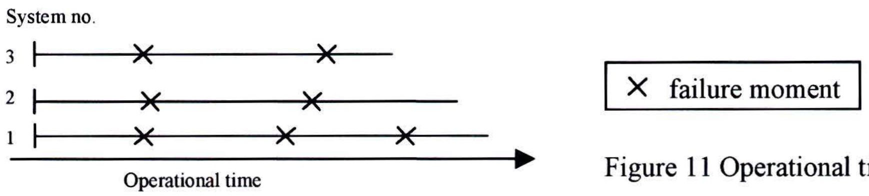


Figure 11 Operational time

Online time
The online time only takes into account the time the system has been switched on. This is of course more accurate than operational time. There is no assurance that the online time equals the exact productive time (time when medical procedures are performed), but it is the closest approximation that can be made right now.

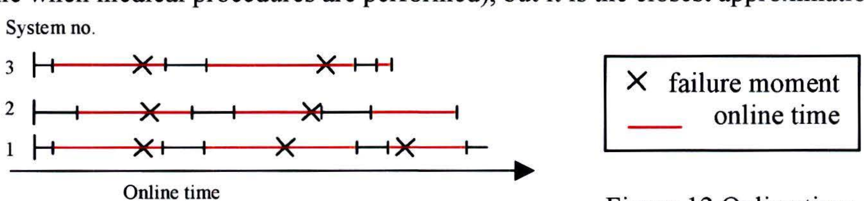


Figure 12 Online time

Service data
Location of the systems
The data from systems in the United States market is used for the analysis for a couple of reasons:
- It is the biggest market for PMS
- The data can be compared to each other since it comes out of the same database
- Philips uses this data in the analyses that it makes right now

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Number of systems
For the visual examination and the trend test twenty systems are used. This way the effects of having more than one system for an analysis can be taken into account without the data becoming unmanageable. These twenty systems have all been in use for at least 1000 days. Given that criterion the twenty systems are chosen randomly.

Next, these twenty systems are divided into two groups of ten systems: systems 1-10 and systems 11-20. This division is based on the installation date, meaning that systems 1-10 were installed earlier than systems 11-20. In this paragraph the following notation will be used for the datasets in this analysis:
- Data from system 1-20: dataset 2
- Data from system 1-10: dataset 2A
- Data from system 11-20: dataset 2B

Furthermore the analysis is also split into unfiltered data and filtered data. This is done to examine how big the influence is of the data pollution of the unfiltered data. Appendix H shows the differences in the recorded number of failures for filtered and unfiltered data of the 20 systems. From that overview it becomes clear that there are quite large differences between filtered and unfiltered failure data for some systems; on average the number of failures from filtered data is about 30% lower than the number of failures from unfiltered data in this dataset .

Unit of measurement
The unit of measurement is in days. This is due to the fact that the data stored on failures in the field is not recorded in greater detail.

Operational time
The problem with the data at hand is that it is not possible to determine the time a system has been online until a failure occurs or the online time between failures. The service engineer cannot record this online time. Therefore operating time is used in calculations with service data.

Field Monitoring Team data
Location of the systems
The data comes from systems that are located all over the world. All systems are of the same version.

Number of systems
At the moment there is a reasonable number of systems being monitored by the Field Monitoring Team. A selection of eight systems is made based on the following considerations.
- Some of the systems were used as prototypes. These were excluded since they can be seen as an extension of the test phase, not as normal customer use (see Roos [14]).
- The time the systems were in the market differs. Chosen are systems with at least 1000 hours of online time.
- Service data needed to be available of these systems in order to be able to analyze the differences between the results from the FMT data and the service data from these systems.

Online time
The online time is recorded in the database, which makes detailed determination of the online time possible.

Visual examination of data
For the visual examination of the data the plotting of cumulative failures versus cumulative time on linear paper is used. For the service data the graphs of all the used systems' failure patterns are given in appendix I. Two graphs of dataset 2A are presented in this paragraph to illustrate the failure patterns, figure 13 and 14; where figure 13 shows the unfiltered data and figure 14 the filtered data.

As can be seen in these plots the line tends to be concave down in almost all instances; only system 7 clearly shows a greater amount of failures than the other systems. This means that the interarrival times (time between failures) of the products tend to become larger. The conclusion from this visual examination of data is that it is plausible to assume that there is a trend.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Another observation that can be made is that there is a quite large spread in the failure pattern that the different systems show. This spread makes it more difficult to have the data be represented by one line of a model.

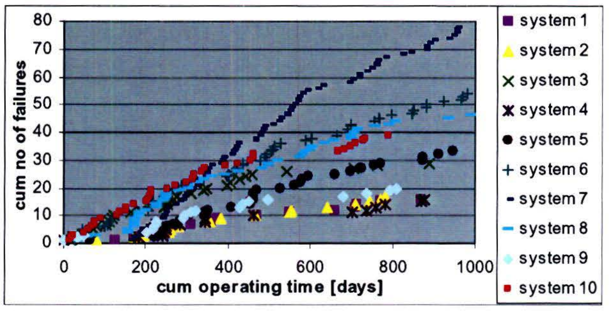


Figure 13 Failure pattern of dataset 2A, unfil. data

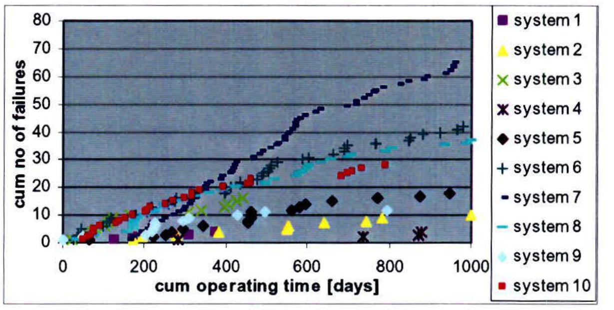


Figure 14 Failure pattern of dataset 2A, fil. data

Evidence of trend with Laplace test

Both individual Laplace tests and combined Laplace tests are performed. The Laplace tests of individual systems are performed in order to determine whether the individual systems show a significant trend. The combined Laplace test of 10 and 20 systems are in order to establish whether there is a significant trend when the data of all these systems is combined. This trend analysis forms an important step in the "General procedure for analyzing failure data of a repairable system", since the result of this step determines whether to use NHPP models or whether further research is necessary when the use of NHPP models is rejected.

Significance level

At a two-sided significance level of 90%, the null hypothesis of HPP should be rejected if the test statistic z is:

$$z \leq -1.64 \text{ or } z \geq 1.64$$

Service data results

The individual Laplace tests:

The results of the individual Laplace tests vary to large extend. In table 5 are the results of the three out of the twenty systems that an individual Laplace test was performed on. The total list can be found in appendix F.

Table 5 Results individual Laplace test, service data

System	Unfiltered	z	Significant trend		Filtered	z	Significant trend	
1	0.390	-1.64	No		-1.762	-1.64		Yes
8	-0.029	-1.64	No		-1.683	-1.64		Yes
15	-3.051	-1.64		Yes	-2.917	-1.64		Yes

The combined Laplace test:

The results of the combined Laplace test show a very significant trend for the twenty systems that are included in the test. However, if only ten systems are included in the test very different results appear, see table 6 for the unfiltered data and table 7 for the filtered data. The fact that dataset 2A shows no significant trend is probably caused by system 7 which shows a much higher number of failures and therefore a different failure pattern.

Table 6 Combined Laplace unfiltered data

Unfiltered data	L_c	z	Significant trend	
Combined Laplace 20 systems	-5.865	-1.64		Yes
Combined Laplace 10 systems (A)	-0.299	-1.64	No	
Combined Laplace 10 systems (B)	-7.760	-1.64		Yes

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Table 7 Combined Laplace filtered data

Filtered data	L_c	z	Significant trend	
Combined Laplace 20 systems	-8.497	-1.64		Yes
Combined Laplace 10 systems (A)	-4.142	-1.64		Yes
Combined Laplace 10 systems (B)	-7.758	-1.64		Yes

FMT data results compared to service data

In this comparison between FMT data and service data there is a problem of comparing two different time scales, the operating time and the online time. Using the dates on which relevant failures took place according to the FMT data a time period is set for operating time of the service data. Appendix G shows a graph in which the operating time is plotted against the online time. It shows that there is a linear relation between these two timelines. For 1000 hours of online time there is an average of 150 days of operating time.

Although there is a relation between operational time and the online time, there are differences found between the failure moments and the number of failures registered by the two data sources. For instance, the FMT data indicating no failures in a certain month for a system, while the service data indicates four failures in that same month for that same system. It is not the case that FMT data structurally registers more (or less) failures than the service data; this is different from system to system.

The individual Laplace tests:

The results of the individual Laplace tests for FMT data show no significant trend in all but one case. The time period that is observed is much shorter than the time period observed for the 20 systems in the calculations made earlier in this paragraph (1000 hours, equal to 150 days, versus 1000 days of data). This might explain why no trend is found. The service data of these systems leads to the same conclusion, no trend within these first 150 days.

Table 8 Results individual Laplace test, FMT data and Service data

System	FMT data	z	Significant trend		Service data	z	Significant trend	
1	-0.638	-1.64	No		-1.185	-1.64	No	
2	0.102	-1.64	No		-0.619	-1.64	No	
3	-2.389	-1.64		Yes	-1.431	-1.64	No	
4	-0.985	-1.64	No		-1.214	-1.64	No	
5	-0.144	-1.64	No		0.968	-1.64	No	
6	-0.951	-1.64	No		1.170	-1.64	No	
7	1.630	-1.64	No		-1.012	-1.64	No	
8	-0.200	-1.64	No		-0.585	-1.64	No	

The results of the combined Laplace test for these eight systems are given in table 9. This table shows the same conclusion as for the individual Laplace test: no trend in the first 1000 hours (150 days).

Table 9 Results combined Laplace test, FMT data and Service data

FMT data	L_c	z	Significant trend	
Combined Laplace 8 systems	-1.100	-1.64	No	
Service data	L_c	z	Significant trend	
Combined Laplace 8 systems	-1.320	-1.64	No	

3.3 Evaluation of model choice

The Laplace test shows that individual systems all have their own degree of trend, some being significant trend and some not. However, the combined Laplace for 10 and 20 systems give significant trend in all but one case. Therefore, given the visual examination of the data and the Laplace tests performed on the data, the models that should be used to analyze the failure data are the Nonhomogeneous Poisson Process models. The two most used and best described models in literature are the Power law and Exponential law model. The choice is made to investigate the prediction performance of these two models since first a good understanding of more popular models should be obtained before the more exotic models are studied, as mentioned in chapter II.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

The FMT data for 1000 hours and the service data over the same period of time did not show any trend. It is still possible to use NHPP models in this situation since the formula changes into an HPP formula when $\beta = 1$, as explained by Ascher and Feingold [25]. The graduation project by Roos [14] shows this as well.

As a starting point in this analysis of NHPP-models individual systems will be used for fitting a Power law and an Exponential law model. This is done to get an idea of how the model reacts to the data. After that both Power law and exponential law models for more than one system will be fitted to the data.

Causes of trend

Since trend is indicated, the question raises what causes this trend. There are several processes that influence the system reliability, which can cause the reliability improvement. First of all the replacement of a bad part, or a part that is outside specification. The term 'outside specification' is used to include parts that have not broken down, but that do not function correctly. The cause of these kinds of parts to be replaced can be multiple. Wrong specification, wrong design, fault during production, damage during shipping, etc. An extra effect of the replacement of those kinds of parts to the reliability improvement is that the parts are normally replaced with the newest version of the part, which is not necessarily the same version as was in the original system.

The second cause of the reliability improvement is the implementation of an FCO. This is similar to the replacement of parts that are bad, or out specification, with the difference that an FCO is coordinated by PMS, meaning that the change might be to parts that have not failed at that moment. The second difference is that with an FCO the change applies to all systems, not to a single system.

User

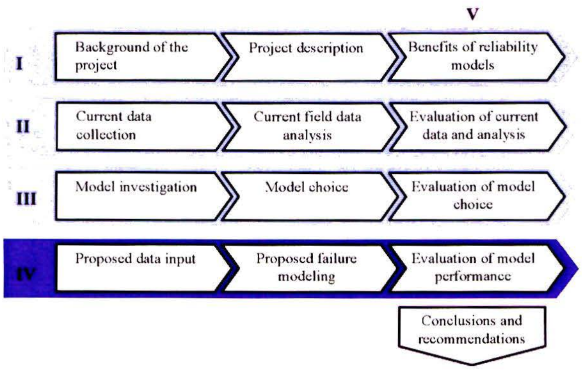
A different perspective on reliability improvement which can play a role in the perception of reliability improvement is habituation by the customer. This only plays a role in case service data is used as an input for failure data. With habituation is meant the effect of 'knowing the system better as time goes by' on placing a call by the user. This can be compared to having a new car. In the beginning each little strange sound the car makes can raise an alarm bell, leading to a call at the service department. Once the owner of the car has the car for a longer period of time, he will be more familiar with it and won't be as anxious to call service. This effect on the calls that come in is not a proven fact; the information comes from someone within the PMS service organization. However, this effect does sound plausible and needs further research to really determine the effect on calls that come in.

The initial reliability level can also change through time. The first system sold of a totally new version will likely have a lower reliability level than a system that comes on the market long after the first introduction of the new version to the market. Both in development and in production initial problems will exist with a new version. A second form of influence on the initial reliability level are the tests performed after assembly at production. The period that these tests are performed will influence the initial reliability level for the customer.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Chapter IV Analysis of field data



Not everything that can be counted counts, and not everything that counts can be counted.

(Albert Einstein)

Now that the choice for the models applicable in the situation of the cardio monoplane systems at PMS is clear the actual analyses can be performed. First the data input is discussed (step 10), followed by an analysis of the data in the proposed models (step 11). After that the model performance is evaluated (step 12).

4.1 Proposed data input

The kind of call data used for the calculations in this chapter is in line with the description of a failure (unfiltered data) and the description of a relevant failure (filtered data) given in chapter III. The FMT data is in line with the description of a relevant failure in case of FMT data. The moment the warranty starts is used as the starting point for the collection of failure data for the field data analyses. Furthermore, there is no grouping of failures; all failures are treated individually, which gives a more accurate reflection of reality.

Calculation of failure moments

For the service data the failure moments are approximated by the difference between the warranty start date and the call start date. This number gives the time to failure in days. Table 10 gives an example of the data necessary for the calculations. The failure moments of the FMT data is obtained differently. The systems at the hospital are able to provide the system online time enabling accurate determination of the failure moments. The failure moments are given in system online hours.

Table 10 Example of failure data

Warranty start date	Call open date	Time to failure	cum no of failures
2001-03-30	2001-04-02	3	1
2001-03-30	2001-04-17	18	2

Visualization of the period that can be analyzed with the available data

As mentioned in chapter II the available data only contains complete information about the failures in the first three years a system is in use. This means only part of the bathtub curve can be visualized with the available data. In order to get a better understanding of the phases in the bathtub curve some background information is now presented.

There are two types of bathtub curves [25], one for parts (non repairable systems) and one for repairable systems. As stated before, this project will only discuss system reliability. This leads to a bathtub curve where the rate of occurrence of failure (ROCOF) is plotted against cumulative operating time. For further discussion on the difference between the two bathtub curves I refer to [25]. The bathtub curve (figure 15) provides insight into the nature of the three classes of failure mechanisms: early failures (A), random failures (B), and aging (C). The models discussed in this report are related to the first two classes, aging is not included since there is no reliable data available to examine the behavior in aging class.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

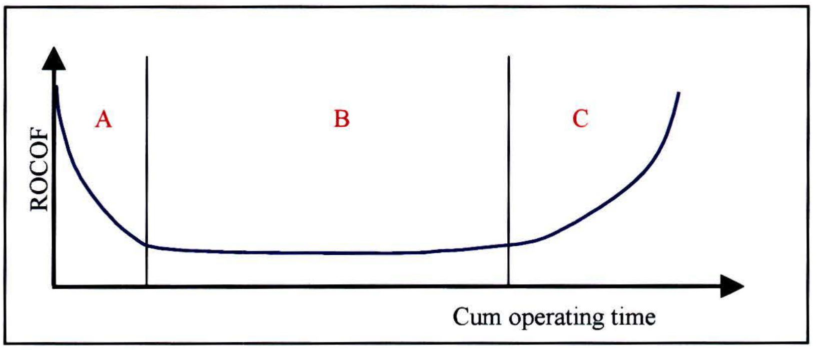


Figure 15 Bathtub curve

Prediction window

To get more insight in the behavior of the power law and exponential law model, the models are created on the basis of data containing 365 days, 730 days, and 1000 days of operating time. For all three groups of data the model is calculated for 900 days of operating time. This means that the model based on 365 days and the one based on 730 days make a prediction of the number of failures until 900 days through extrapolation. The model based on 1000 days of data does not predict, it serves as a comparison of the line that the model would draw if it had all the necessary information.

The goodness-of-fit test that is used for the models based on 365 and 730 days of data fits these models to the individual failures of the 900 days period. This means that it is used as a sort of goodness-of-prediction-fit, since the model uses data from 365 days (and 730 days) to fit a 900 days period. The models based on 1000 days of data are also fit to the individual failures of the 900 days period, which is a normal fitting of data.

Assumptions for the data

The fact that the data is used for modeling the real life situation means that a simplification of reality is used. In the next overview the most important assumptions are presented.

General assumptions

- Systems under consideration are representative for the cardio monoplane system population (see [34]).
- The repair on the products can be seen as minimal repair (see chapter II).
- The features to the system stay the same, none are removed, and none are added.
- Quality of data: the necessary data for analyses is complete, consistent, timely and accurate.
  Problems with the data quality were discussed in chapter II.
- The failure is solved at the moment the corrective maintenance job is executed (no need to come back for it a second time, 'No Trouble Found repair').
- Repair time is not taken into account. Ascher and Feingold formulate this in the following manner in [25]: "We will also devote little attention to repair times, assuming in most cases that repair is either completed instantaneously or is measured on a different time scale than system operating time."
- Products in population are assumed to be identical, this means that the same sort of failures can occur on all products under consideration.
  This is achieved partly by keeping a dataset restricted to only cardio monoplane systems and looking at the same system version within the cardio monoplane systems. There will however still be small differences given the customization possibilities the customer has described in chapter II. More about the tradeoffs between completeness and consistency of data can be found in Ballou and Pazer [21].

Specific assumptions for the data from the service data

- Usage of the customer is constant in time and the usage is constant among the users (see [8]).
  It is known that this assumption is not true. The usage depends on the hospital, the kind of system (cardio, vascular, or cardio/vascular), and the procedure that is performed by the doctor. Service data however is not able to record the online time as explained in chapter III.
- The moment a call takes place is equal to the moment the failure has taken place (there is no reporting delay).
  In chapter II it was explained that there can be such a reporting delay. However, those kinds of delays are usually restricted to a couple of days maximum.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

- At moments that there are no calls, there are no failures.
  This is also debatable since not certain that every time something happens that falls under the definition of a relevant failure a call is placed.

Specific assumptions for the data from the FMT database
- When there is no log made, there is no failure.
- If there is more than one failure in a day the other failures that day are the result of that first failure and therefore do not need to be taken into account.

4.2 Proposed failure modeling

4.2.1 Data analyses using the proposed models

As concluded from last chapter NHPP models will be used; specifically the power law and the exponential law model. Both the service data and the FMT data are used in the failure modeling. The data sets that are used are equal for both model types. The analyses that are performed can be divided in the following manner:

1. Analysis of service data of a single system
   This analysis investigates whether the power law model and the exponential law model are able to fit the failure pattern of the data that is used in this project. For this analysis unfiltered data is used, this has no influence on the question of whether the power law and exponential law are able to fit the failure pattern since chapter III showed that for both the unfiltered and filtered data NHPP models should be used.
2. Analysis of service data of multiple systems
   This analysis investigates the influence of using data of multiple systems for modeling the failure pattern. The analysis uses both unfiltered and filtered data to investigate the influence of filtering on the calculated model values.
3. FMT data versus service data
   This analysis shows the differences between using FMT data and using service data when trying to establish the failure pattern of a system.

4.2.2 Analysis one: Service data of a single system

This analysis uses the data from one system; appendix J shows the failure data that is used for the calculation of the model parameters. The system is one of the 20 systems that are used in analysis two, that is, system 15. This is a system that shows significant trend, but there are no special characteristics why this system is chosen for the analysis of the single system, the other systems would qualify just as well. First is the analysis made with the power law model, after that the analysis with the exponential law model.

Power law

As explained in chapter III, the formulas to estimate the model parameters based on maximum likelihood estimates are:

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ln \frac{T_n}{T_i}} \quad \text{and} \quad \hat{\lambda} = \frac{n}{T_n^{\hat{\beta}}},$$

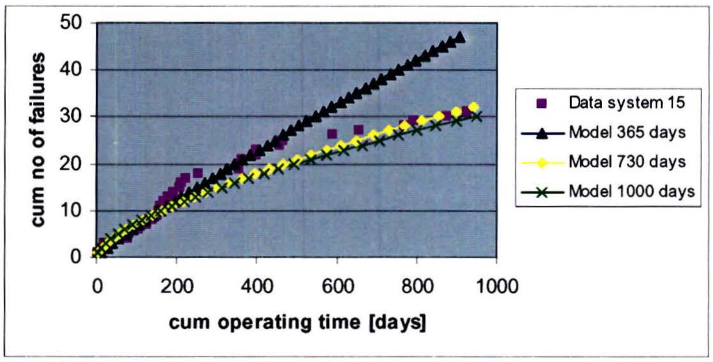
These parameter estimates are calculated based on 365, 730, and 1000 days of failure data leading to the parameters values given in table 11.

Table 11 parameters $\hat{\lambda}$ and $\hat{\beta}$, formulas for E[N(T)] and $\hat{\mu}_1(T)$

	$\hat{\lambda}$	$\hat{\beta}$	E[N(t)]	$\hat{\mu}_1(T)$
365 days	0.112	0.887	$E[N(T)] = 0.1118T^{0.887}$	$\hat{\mu}_1(T) = 0.0992T^{-0.113}$
730 days	0.339	0.664	$E[N(T)] = 0.3393T^{0.664}$	$\hat{\mu}_1(T) = 0.2252T^{-0.336}$
1000 days	0.423	0.622	$E[N(T)] = 0.4233T^{0.622}$	$\hat{\mu}_1(T) = 0.2631T^{-0.378}$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

With these parameters the expected number of failures E[N(T)] are calculated, see table 10. There is a clear 'pattern' visible in the values of these equations. The constants in the formula are increasing and the values of the power are decreasing. The decreasing value of $\hat{\beta}$ indicates an increasing trend since when $\hat{\beta}$ =1 the model becomes a Homogeneous Poisson Process. When the expected number of failures E[N(T)] are plotted, as shown in figure 16, it is clear that based on the smallest amount of data (365 days) the model reflects the actual data the least accurate. This indicates that for this dataset, prediction of the failure pattern for 900 days based on 365 days of data is not accurate. For the models based on 730 and 1000 days of data the middle part of the graph shows a slight deviation from the actual data.



The actual number of failures after 900 days is 30; E[N(900)] based on 365 days is 47; E[N(900)] based on 730 days is 31; E[N(900)] based on 1000 days is 29. This shows how close to the actual number of failures the model is based on both 730 and 1000 days of data.

Figure 16 Expected number of failures power law model

Model goodness-of-fit: Chi-squared test and Cramér-von Mises
As explained in chapter III the Chi-squared test works by dividing the observed and expected number of failures into classes. At a 95% level of significance, with five classes, and two parameters in the model, the test statistic is 5.99. The number of classes is determined using the rules for given in paragraph 3.1.5. The test statistic for the Cramér-von Mises test is taken from table 25 in appendix E. Both tests test the models against a 900 days period, as explained in 4.1 (prediction window). Table 12 gives the results.

Table 12 results goodness-of-fit tests

System	Days	χ^2	Test statistic	Result χ^2	$C^2(n)$	Test statistic	Result $C^2(n)$
15	365	9.55	5.99	Reject	0.806	0.171	Reject
15	730	2.56	5.99	Accept	0.264	0.171	Reject
15	1000	2.89	5.99	Accept	0.209	0.171	Reject

Conclusion from Chi-squared goodness-of-fit
- The model based on 365 days of data is rejected. This means that this model cannot give a good representation of the failure behavior and is not able to predict the near future.
- The model based on 730 days of data is accepted. This means that this model gives a good representation of the failure behavior and is able to predict the near future.
- The model based on 1000 days of data is accepted. This means that this model gives a good representation of the failure behavior.

Conclusion from Cramér-von Mises
- This test does not work with classes, like the Chi-squared test. This takes away the problem of having to choose classes of time periods.
- The difference with the Chi-squared test is that the Cramér-von Mises test rejects all models! The values of the models based on 730 days and 1000 days of data are however very close to being accepted. But the difference in result between the two tests is striking. The reason for not accepting the model might be explained by the deviation of the models from the data points in the middle part of figure 16.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Bathtub curves

The intensity function $\hat{\mu}_1(T)$ is also calculated using the estimated parameter values $\hat{\lambda}$ and $\hat{\beta}$, the formulas are shown in table 11. Since $\hat{\beta}$ is smaller than 1 in all three cases the rate of occurrence of failure is decreasing in each case. Figure 17 shows the plots of the ROCOF against the cumulative operating time. With these plots it is possible to visualize when the system comes into the 'steady state' phase, indicated as phase B in figure 15 where the bathtub curve was explained. The three lines are all still decreasing at the end of the graph which means they have not reached their steady state, although the lines are leveling out indicating that they are close to reaching the steady state.

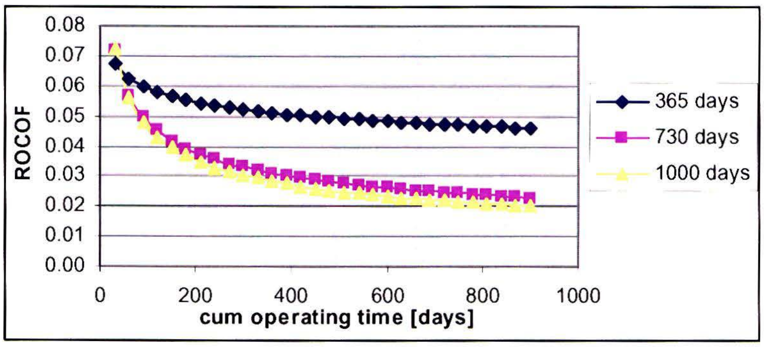


Figure 17 Bathtub curves power law model

The graph clearly shows the difference between the line based on 365 days compared to the lines based on 730 and 1000 days of data.

MTFBc

Based on the Fisher Matrix the variance and covariance of λ and β are calculated. These are used to calculate the upper and lower 90% confidence bounds of the MTBFc. The figures below show the measured MTBFc and the MTBFc calculated by the model based on an increasing amount of failure data with its confidence bounds. These three graphs clearly visualize the adaptation of the calculated MTBFc to the extra data it is given. The confidence bounds are quite wide indicating that the necessary precaution needs to be taken when drawing conclusions based on the actual MTBFc values in the figures.

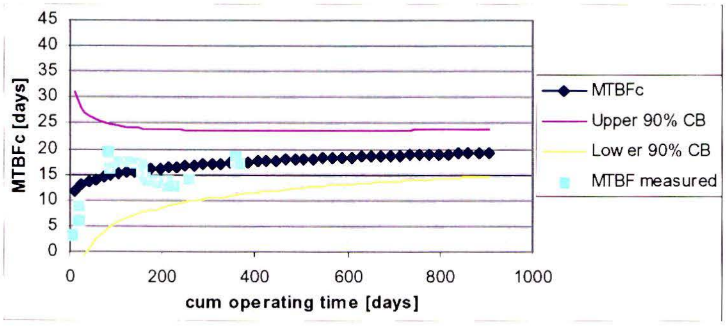


Figure 18 $MTBF_c$ based on 365 days of data

In figure 18 the MTBFc starts at a value around 10 days, with a confidence interval of 0 < MTBFc < 31, and ends up at a value a little less than 20 days, with a confidence interval of 14 < MTBFc < 23.

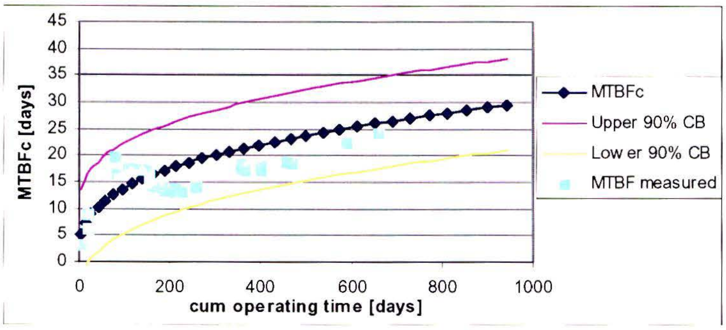
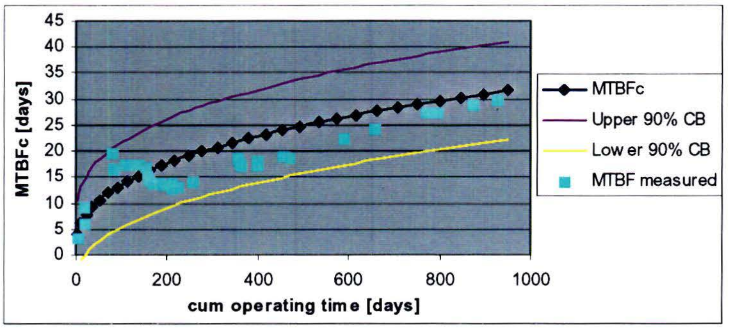


Figure 19 $MTBF_c$ based on 730 days of data

In figure 19 the MTBFc starts at a value around 5 days, with a confidence interval of 0 < MTBFc < 13, and ends up at a value a little less than 30 days, with a confidence interval of 21 < MTBFc < 38.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved



In figure 20 the MTBFc starts at a value around 4 days, with a confidence interval of 0 < MTBFc < 10, and ends up at a value around 31 days, with a confidence interval of 22 < MTBFc < 41.

Figure 20 $MTBF_c$ based on 1000 days of data

Concluding from this power law analysis of one system it can be said that, looking at the graph where the cumulative operating time is set against the cumulative number of failures, the models based on 730 days and 1000 days of data are able to represent the real data quite good. The two goodness-of-fit tests show a very different outcome, which could be explained by the deviation in the middle part of the graph in figure 16. The graphs of the cumulative mean time between failure show that the value of this statistic should be dealt with care since the confidence bounds are quite wide. So far the analysis based on the power law formulas. The next paragraph will use the same dataset for calculations based on the exponential law formulas.

Exponential law

The formulas to estimate the model parameters based on maximum likelihood estimates are:

$$\sum T_i + n\alpha_1^{-1} - nT_n\{1 - e^{-\alpha_1 T_n}\}^{-1} = 0 \qquad \text{and} \qquad \hat{\alpha}_0 = \ln\left\{\frac{n\hat{\alpha}_1}{e^{\hat{\alpha}_1 T_n} - 1}\right\}$$

Like with the power law calculations, these parameter estimates are calculated based on 365, 730, and 1000 days of failure data leading to the parameters values given in table 13. The parameters of this model are calculated using a mathematical program called Matlab. The interesting conclusion that can be drawn when observing the parameter values is that the values of 1000 days of data are in between the values of 365 days of data and 730 days of data. This is different from the power law model.

Table 13 Parameters α_0 and α_1, formulas for E[N(T)] and $\hat{\mu}_2(T)$

	α_0	α_1	E[N(T)]	$\hat{\mu}_2(t)$
365 days	-2.6506	-0.0011	$E_2[N(T)] = \frac{e^{-2.6506}}{-0.0011}(e^{-0.0011T_b} - e^{-0.0011T_a})$	$\hat{\mu}_2(T) = e^{-2.6506-0.0011T}$
730 days	-2.4720	-0.0025	$E_2[N(T)] = \frac{e^{-2.4720}}{-0.0025}(e^{-0.0025T_b} - e^{-0.0025T_a})$	$\hat{\mu}_2(T) = e^{-2.4720-0.0025T}$
1000 days	-2.5617	-0.0021	$E_2[N(T)] = \frac{e^{-2.5617}}{-0.0021}(e^{-0.0021T_b} - e^{-0.0021T_a})$	$\hat{\mu}_2(T) = e^{-2.5617-0.0021T}$

Figure 21 shows the expected number of failures E[N(T)] which are calculated using the Exponential law formula for a single system. Here it is visible what is stated above: the line based on 1000 days of data is in between the line of 365 days of data and 730 days of data, although these two lines are almost the same line.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

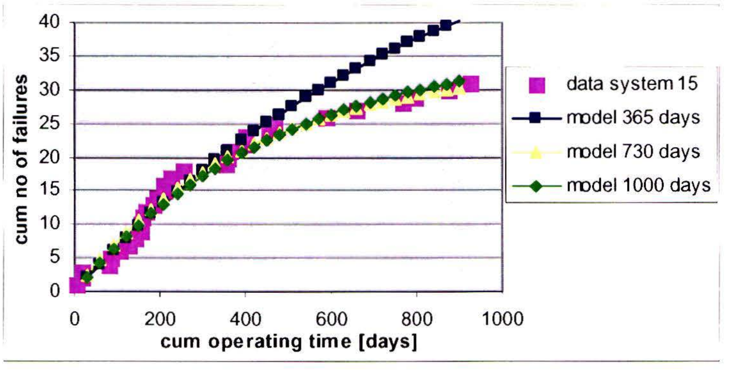


Figure 21 number of failures exponential law model

The actual number of failures after 900 days is 30; E[N(900)] based on 365 days is 40; E[N(900)] based on 730 days is 30; E[N(900)] based on 1000 days is 31. This is similar to the power law model, only the expected value based on 365 days is closer to the actual value compared to the power law model.

Model goodness-of-fit: Chi-squared test

The chi-squared test tests the models against a 900 days period, as explained in 4.1 (prediction window). Unfortunately it is only possible to perform a Chi-squared test on the exponential law models since the Cramér-von Mises test is only built for the power law models. At a 95% level of significance, with five classes, and two parameters in the model, the test statistic is 5.99. The results are presented in table 14.

Table 14 results goodness-of-fit test

System	Days	χ^2	Test statistic	Result
15	365	4.70	5.99	Accept
15	730	0.83	5.99	Accept
15	1000	1.81	5.99	Accept

Conclusion from Chi-squared goodness-of-fit

- The models based on 365, 730, and 1000 days of data are accepted. This means that the exponential law model gives a good representation of the failure behavior and is able to predict the near future.
- Interesting detail is that the value of χ^2 is the lowest for the model that is based on 730 days of data, instead of the model based on 1000 days of data. This means the best fit to the data points by the model based on 730 days of data.

Bathtub curves

The model intensity $\hat{\mu}_2(T)$ is also calculated using the estimated parameter values α_0 and α_1, the formula values are shown in table 13. Since α_1 is smaller than 0 in all three cases the rate of occurrence of failure is decreasing in each case. Figure 22 shows the plots of the ROCOF against the cumulative operating time. With these plots it is possible to visualize when the system comes into the 'steady state' phase, indicated as phase B in figure 15 where the bathtub curve was explained. The three lines are all still decreasing at the end of the graph which means they have not reached their steady state, although the lines are leveling out indicating that they are close to reaching the steady state.

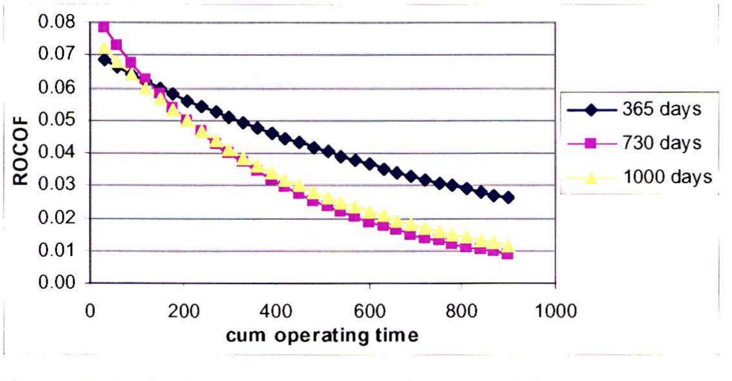


Figure 22 Bathtub curves exponential law model

The difference between the curves based on the exponential law model (figure 22) and the power law model (figure 17) is that the curves of the exponential law model are closer together and have lower final values.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

$MTBF_c$

Unfortunately there are no formulas found in literature to calculate the confidence bounds for the exponential law model, therefore only the $MTBF_c$ calculated by the model based on 365, 730, and 1000 days of data is given set out against the measured $MTBF_c$ values. Interesting in these graphs compared to the graphs of the $MTBF_c$ based on the power law model is the beginning of the modeled line. Where the $MTBF_c$ based on the power law model showed a curve in the beginning of the line, the $MTBF_c$ based on exponential law is an almost straight line. The power law model gives a better fit to the measured MTBF in the beginning of the graphs.

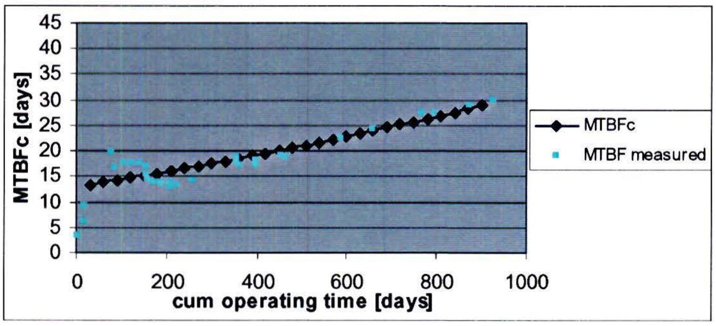


Figure 23 MTBFc based on 365 days of data

In figure 23 the $MTBF_c$ starts at a value around 14 days and ends up at a value around 23 days.

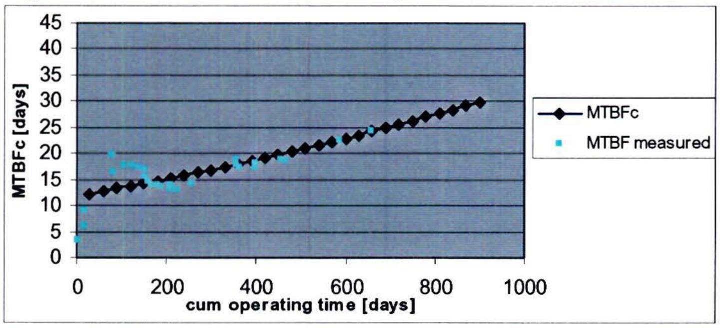


Figure 24 MTBFc based on 730 days of data

In figure 24 the $MTBF_c$ starts at a value around 12 days and ends up at a value a little less than 30 days.

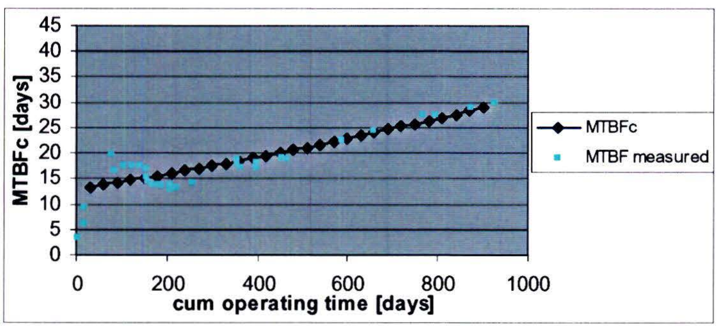


Figure 25 MTBFc based on 1000 days of data

In figure 25 the $MTBF_c$ starts at a value around 14 days and ends up at a value around 29 days.

Concluding from this Exponential law analysis for one system is that the fit of the model to the data points seems to be good, as well as the predictive value based on 730 days of data, and even the predictive value based on 365 days of data according to the Chi-squared goodness-of-fit test. Visual examination shows however that this model starts to deviate from the data points after about 450 days. For the $MTBF_c$ the model fits the measured $MTBF_c$ very well except for the first part.

Overall conclusions of analysis one

The fit to this dataset is accepted using the Chi-squared goodness-of-fit test in all but one case, 365 days of data in the power law model. The Cramér-von Mises goodness-of-fit rejects the power law model in all cases, for which an explanation could be found in the visual examination of the data. Visual examination shows that

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

especially the middle part of the fitted power law model deviates from the actual data leading to the rejection by the Cramér-von Mises test.

For the exponential law model the fit to this dataset is accepted for the model based on 365, 730 and 1000 days of data using the Chi-squared goodness-of-fit test. Visual examination shows however that the model based on 365 days does deviate from the data points after about 450 days. The main difference between Power law and Exponential law when the $MTBF_c$ is concerned is the beginning of this $MTBF_c$ for Exponential law. This first part does not follow the data points like the Power law model does. On the other side, the $MTBF_c$ for Power law does not fit the data points as well as Exponential law in the middle part. It seems that both models focus on a different part. The next step is to see whether the formula also works when more than one system is included.

4.2.3 Analysis two: Service data from multiple systems

The same twenty systems are used for this analysis as the ones used in paragraph 3.2, where the service data for trend analysis is discussed. These twenty systems have all been in use for at least 1000 days. The same notation for the datasets in this analysis will be used:

- Data from system 1-20: dataset 2
- Data from system 1-10: dataset 2A
- Data from system 11-20: dataset 2B

Furthermore the analysis is also split into unfiltered data and filtered data to examine how big the influence is of the data pollution of the unfiltered data.

Power law

The combined systems formula is used (given in paragraph 3.1.4) for the parameter estimation of the power law model. The formulas are calculated using 365 days, 730 days and 1000 days of data. Tables 15 and 16 give the values of the estimated model parameters $\hat{\lambda}$ and $\hat{\beta}$ calculated for each of the cases.

Table 15 Parameter values

Data	Dataset 2A				Dataset 2B			
	Unfiltered		Filtered		Unfiltered		Filtered	
	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\beta}$
365 days	0.0088	1.2840	0.0089	1.2071	0.0783	0.9632	0.0437	0.9961
730 days	0.0387	1.0048	0.0268	1.0019	0.2189	0.7617	0.1318	0.7817
1000 days	0.0763	0.8864	0.0537	0.8798	0.3167	0.6931	0.1801	0.7242

A conclusion from these model parameters is that the $\hat{\lambda}$ values are increasing when the number of days included in estimating the parameters increases. For the $\hat{\beta}$ values this is the other way around, their values decrease when the number of days included in estimating the parameters increases. If $\hat{\beta} < 1$ this means that the times between failures become longer (increasing reliability); if this $\hat{\beta}$ value gets closer to zero there is more trend in the model, which is the case when the number of days included goes from 365 to 1000 days. Another conclusion can be drawn when the $\hat{\beta}$ values from dataset 2A are compared to the $\hat{\beta}$ values from dataset 2B is that for dataset 2A, based on 365 and 730 days of data, $\hat{\beta} \geq 1$ while for dataset 2B $\hat{\beta} \leq 1$ in all cases. This means that these $\hat{\beta}$ values for dataset 2A indicate an increasing rate of occurrence of failure. In other words, the reliability is not increasing but decreasing.

Table 16 Parameter values

Data	Dataset 2			
	Unfiltered		Filtered	
	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\beta}$
365 days	0.0347	1.0780	0.0236	1.0737
730 days	0.1083	0.8589	0.0685	0.8706
1000 days	0.1737	0.7737	0.1076	0.7893

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

In comparison with the parameter values based on dataset 2A and 2B the overall parameter values (of dataset 2) are in between the values of datasets 2A and 2B, as would be expected. The $\hat{\lambda}$ value based on dataset 2 is larger than the $\hat{\lambda}$ value based on dataset 2A and smaller than the $\hat{\lambda}$ value based on dataset 2B. A similar conclusion can be drawn for $\hat{\beta}$.

A comparison of the unfiltered values and the filtered values of $\hat{\lambda}$ and $\hat{\beta}$ shows that the influence of the data pollution on the pattern of the models is not large. This means that the pollution in the unfiltered data does not lead to a model going from decreasing reliability over time based on unfiltered data to increasing reliability based on filtered data, or the other way around. The differences in parameter values will however lead to different values of the intensity function, expected number of failures and the MTBF.

Using the parameter values from table 15 and 16 the intensity function, the expected number of failures at a certain point of time, and the MTBFc can be obtained. In order to draw conclusions on the models it is not necessary to show all graphs of all models, therefore the graphs of dataset 2A are presented to explain the effects that take place. The choice for presenting the graphs of dataset 2A is based on two considerations. First, graphs from dataset 2 become more or less unreadable because of the many data points. Second, the failure patterns of the systems in dataset 2A is more divers then that of the systems in dataset 2B, which makes the complications of having more than one system in the data analysis more clear. Appendix K gives the intensity functions and the function for calculation of the expected number of failures for dataset 2, 2A, and 2B.

Expected number of failures
In figures 26 and 27 the cumulative numbers of failures against the cumulative operating time for unfiltered and filtered data of dataset 2A are given. The figures show the actual data points and the values calculated by the models based on 365, 730, and 1000 days of data from these ten systems. Clearly visible is the fact that the lines of the models based on filtered data are lower than the lines of the models based on unfiltered data. Also clearly visible is the large spread of the actual data points around the modeled lines. The other graphs of the cumulative number of failures against the cumulative operating time are given in Appendix L. In these other graphs the most distinct difference compared to the graphs given below is that the line based on 365 days of data is more in line with the lines based on 730 and 1000 days of data. This is due to the fact that the data of dataset 2B shows less spread.

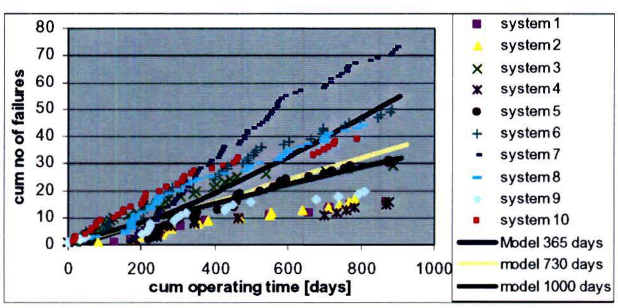


Figure 26 Power law unfiltered data

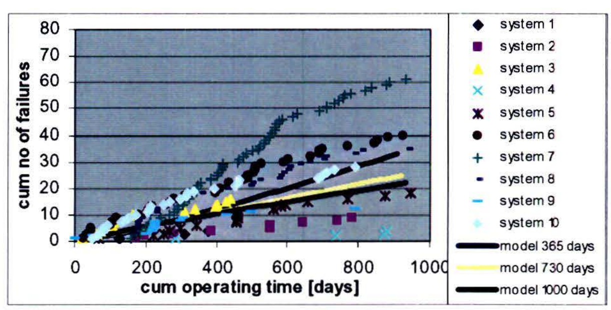


Figure 27 Power law filtered data

Conclusions on the expected number of failures of dataset 2A:

- Using 365 days of data gives a graph with a slight curve upwards, implying that the reliability is deteriorating with time.
- Using 730 days of data gives a graph with a slight curve downwards, implying that the reliability is slightly improving with time.
- Using 1000 days of data gives a graph with more curve downwards, implying that the reliability is improving with time.

Model Goodness-of-fit (Chi-squared test)
In appendix M the results of the chi-squared test are given. These are the goodness-of-fit results on individual systems, for the models based on combined systems formula. The model is fitted to 900 days of data, the

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

failures are divided into classes based on the rules for making classes explained in chapter III. The number of parameters p = 2 and the level of significance $\alpha = 0.05$. The test statistic varies per system to which the goodness-of-fit test is performed since the number of failures, and therefore the number of classes, is different per system.

A critical note to this analysis is that the models are based on 10 (and 20) systems, while the goodness-of-fit is performed on a single system. Since the model is based on an average of all systems, it is logical that the goodness-of-fit to an individual system is not always perfect.

To get some insight in the results of the chi-squared test two histograms are made, figure 28 and 29, indicating the values at 365, 730, and 1000 days of data per system. The shorter the bar in the histogram the better the fit of the model to the data. These figures show that in most cases the values "365 days" are much higher than the values "730 days" and "1000 days", indicating a worse fit. This is logical since the model based on 365 days of data predicts based on the smallest amount of data. When the model is fitted to system 7 the situation turns around, the best fit is obtained with the model based on 365 days of data, although that fit is still very far from being a significant fit.

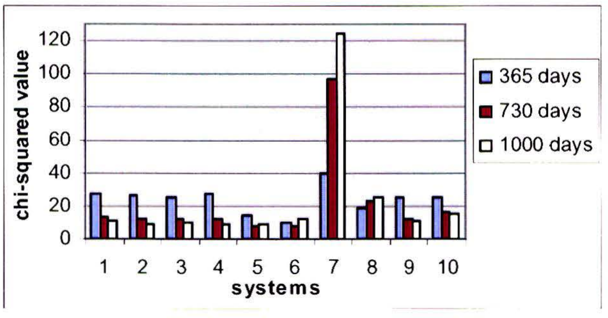


Figure 28 Dataset 2A Chi-squared unfiltered data

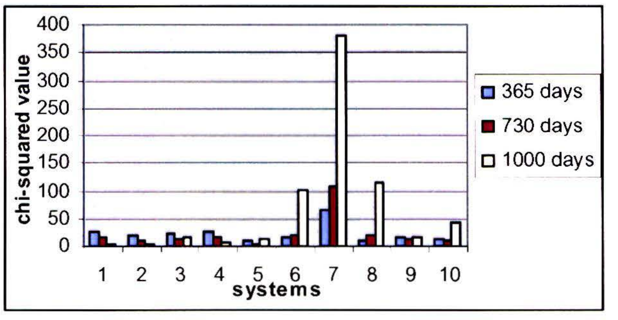


Figure 29 Dataset 2A Chi-squared filtered data

Conclusions of Chi-squared goodness-of-fit

- For datasets 2, 2A, and 2B most goodness-of-fit results show an improvement in fit when more data is included.
- Dataset 2A has one system giving a totally different reaction than the other systems.
- The results of the filtered data are similar to the results of the unfiltered data, although the chi-squared values are different.

Model Goodness-of-fit (Cramér-von Mises)

As said earlier, this test does not work with classes, like the Chi-squared test. Figures 30 and 31 show the Cramér-von Mises results for dataset 2A, the results of dataset 2 and 2B are given in appendix N.

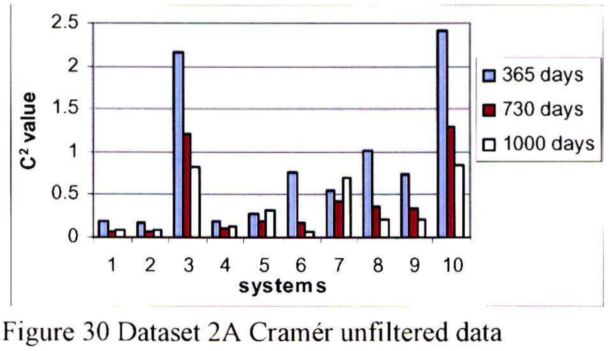


Figure 30 Dataset 2A Cramér unfiltered data

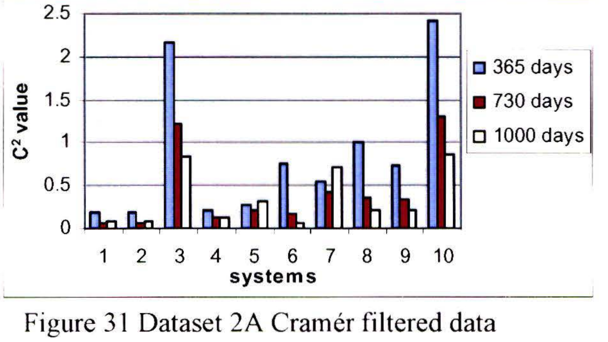


Figure 31 Dataset 2A Cramér filtered data

Conclusions from Cramér-von Mises

- It can be concluded that more models are accepted by this test than with the Chi-squared test, see appendix N for details.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

- The same critical note to this analysis can be made as with the Chi-squared test, meaning that the models are based on 10 (and 20) systems, while the goodness-of-fit is performed on only 1 system. Since the model is based on an average of all systems, it is logical that the goodness-of-fit to an individual system is not always perfect.

Bathtub curves
Figures 32 and 33 show the bathtub curves of the unfiltered and filtered data for the models based on dataset 2A. In appendix L the other intensity functions based on the calculated parameter values are plotted representing the parts A and B of the bathtub curve discussed in the beginning of this chapter.

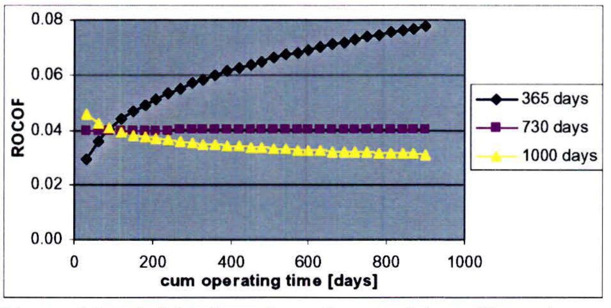


Figure 32 ROCOF based on dataset 2A unfiltered

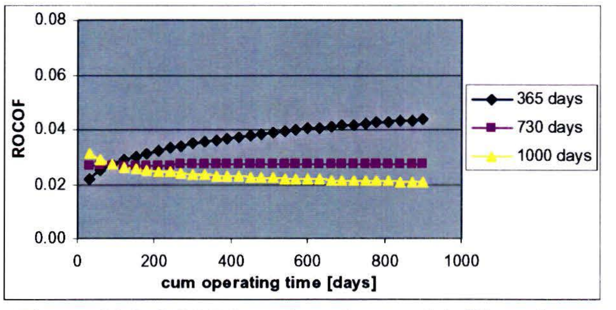


Figure 33 ROCOF based on dataset 2A filtered

Conclusions on the bathtub curves, plotted using the calculated ROCOF:

- Using 365 days of data the prediction is that the ROCOF of the systems is deteriorating in time and no steady state can be predicted. Next to that there is a clear difference between the values of the curve based on filtered data compared to the curve based on unfiltered data. The final value of the unfiltered data (around 0.08) is almost twice as high as the value based on filtered data (around 0.04).
- Using 730 days of data the prediction is that the ROCOF of the systems is constant in time. This would imply an HPP where the moment steady state is achieved from the beginning.
- Using 1000 days of data the prediction is that the ROCOF of the systems is improving in time. Here it is possible to make an estimate of the moment steady state is achieved. For the filtered data this is around 500 days.

MTBFc
The $MTBF_c$ is calculated the same way as for the analysis based on a single system. The figures 34 and 35 show the $MTBF_c$ with its confidence bounds for the model based on 1000 days of operating time of dataset 2A. Also the points of the actual $MTBF_c$ values measured for the ten systems are given in the graphs. The graphs show that the modeled lines represent the data reasonably well, although the confidence bounds indicate that the actual values of the lines should be treated with a certain amount of reservation. The other $MTBF_c$ graphs are in appendix O.

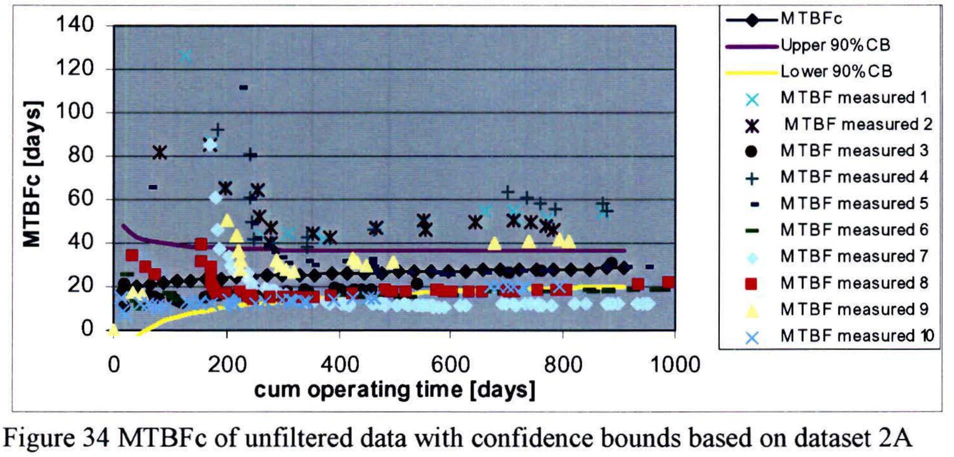


Figure 34 MTBFc of unfiltered data with confidence bounds based on dataset 2A

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

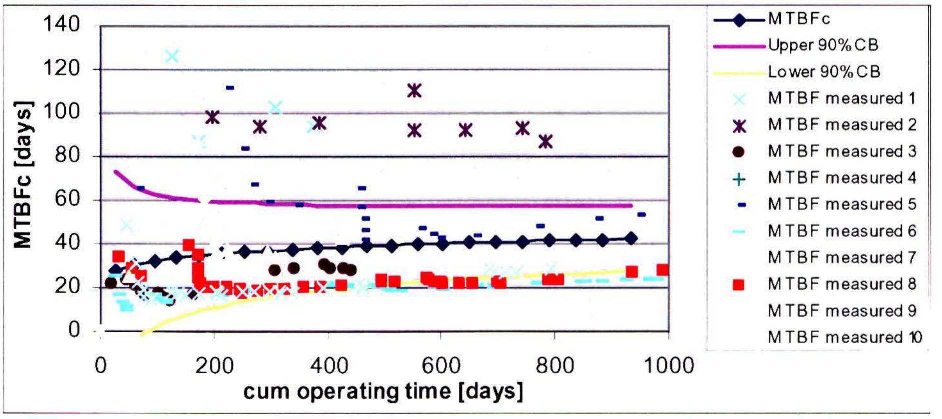


Figure 35 MTBFc of filtered data with confidence bounds based on dataset 2A

Exponential Law

The values of the exponential law model parameters are calculated using the formulas for multiple systems. Again, the parameter values are calculated using 365, 730 and 1000 days of data. Matlab is used to calculate the parameters α_0 and α_1. Table 17 gives the parameter values.

Table 17 Parameter values of dataset 2A and 2B, exponential law

Data	Dataset 2A				Dataset 2B			
	Unfiltered		Filtered		Unfiltered		Filtered	
	α_0	α_1	α_0	α_1	α_0	α_1	α_0	α_1
365 days	-3.7475	0.00380	-4.1077	0.0041	-2.6067	-0.00071	-3.0933	0.00040
730 days	-3.1315	-0.00016	-3.6898	0.0010	-2.4813	-0.00180	-2.9698	-0.00090
1000 days	-3.1070	-0.00030	-3.5130	-0.000022	-2.5394	-0.00160	-2.9734	-0.00130

The first thing that strikes here is the sign change of α_1 in dataset 2A and 2B; where $\alpha_1 < 0$ means an increasing reliability and $\alpha_1 > 0$ means a decreasing reliability. This is similar to the change of $\hat{\beta}$ in the power law models being smaller or larger than 1. Only in case of unfiltered data in dataset 2B there is no value of $\alpha_1 > 0$. In that case there is an increasing reliability for the models based on 365, 730, and 1000 days of data.

When the parameter values of power law and exponential law are compared on the basis of whether increasing or decreasing reliability is modeled, it can be concluded that there are some cases where power law and exponential law disagree on whether there is an increasing or a decreasing reliability. This shows the caution that needs to be taken when interpreting these models, since the same data can lead to a model with increasing or decreasing reliability depending on the model chosen. In the conclusions of this chapter an attempt will be made to decide which model type (power law or exponential law) has the best prediction performance, based on which model provides the most significant fits to the 900 days period.

Table 18 Parameter values of dataset 2, exponential law

Data	Dataset 2			
	Unfiltered		Filtered	
	α_0	α_1	α_0	α_1
365 days	-3.0715	0.00120	-3.5051	0.0018
730 days	-2.7815	-0.00099	-3.2991	0.00001
1000 days	-2.8049	-0.00095	-3.2241	-0.00071

Conclusion from the parameter values of dataset 2 is that these values are in between the values of dataset 2A and 2B. Appendix P gives the intensity functions and the function for calculation of the expected number of failures for dataset 2, 2A, and 2B.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Expected number of failures

In figure 36 and 37 the cumulative number of failures is plotted against the cumulative operating time for dataset 2A. The lines of the models based on 365, 730, and 1000 days of data are plotted together with the actual number of failures that occur in this time period. The other graphs of the expected number of failures are given in Appendix Q.

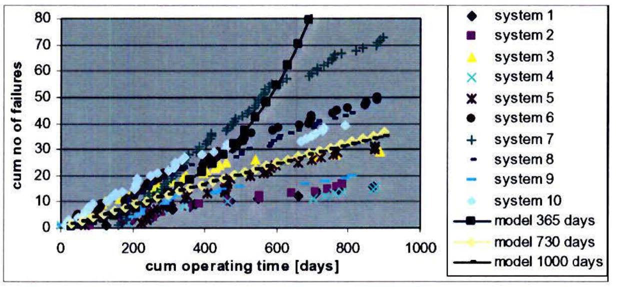


Figure 36 dataset 2A unfiltered data

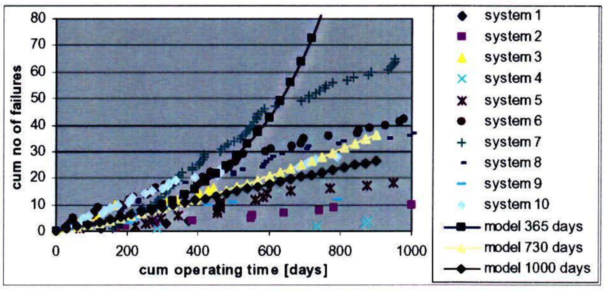


Figure 37 dataset 2A filtered data

Conclusions on expected number of failures:

- For the model based on dataset 2A
  - Using 365 days of data gives a graph with exponentially faster occurrence of failure, this is a very bad fit to the data.
  - Using 730 days of data gives a graph with a slight curve, implying an increasing rate at which the failures are occurring in time; in other words a decreasing reliability.
  - Using 1000 days of data gives a graph with a more or less straight line, implying no trend.

- For the model based on dataset 2B and dataset 2 (appendix Q)
  - Using 365 days of data gives a graph with an almost straight line, implying no trend.
  - Using 730 days of data gives a graph with a slight curve, implying a slightly decreasing rate at which the failures are occurring in time; in other words a slightly increasing reliability.
  - Using 1000 days of data gives a graph with more curve, implying a decreasing rate at which the failures are occurring in time; in other words a increasing reliability.

Model Intensity $\hat{\mu}_2(t)$

In figures 38 and 39 the ROCOF is plotted showing a clearly different pattern in case of 365 days of data, both in case of filtered and unfiltered data. The pattern based on 365 days of data predicts a strongly increasing ROCOF, meaning that the reliability is decreasing strongly. This is not a good prediction since the graphs based on 730 and 1000 days of data show a totally different ROCOF. When compared to the Power law graphs (figure 32 and 33) the most striking is the difference between the patterns based on 365 days of data. Although in both cases the predictive value is not good, the ROCOF stabilizes in the Power law model, while it only gets worse in the Exponential law. The other graphs of the ROCOF (intensity function) are given in Appendix Q. There the plots of dataset 2B show that when there is less spread in the dataset, the plot based on 365 days of data shows a less radical deviation from the plots based on 730 and 1000 days of data.

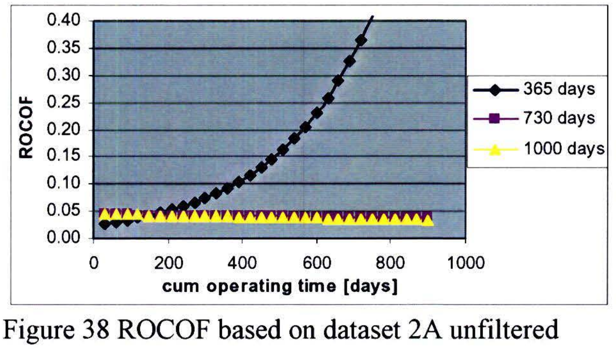


Figure 38 ROCOF based on dataset 2A unfiltered

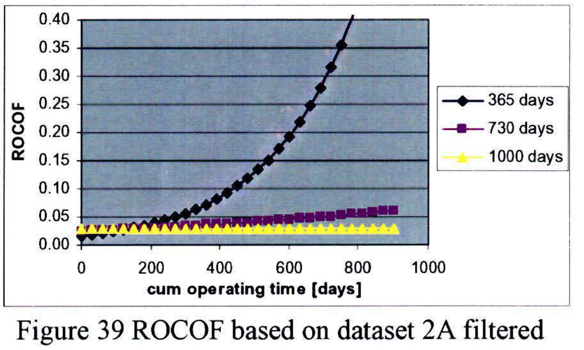


Figure 39 ROCOF based on dataset 2A filtered

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Conclusions on the bathtub curves (see also appendix Q), plotted using the calculated ROCOF:
- Based on 365 days of data the plots give a diverse image. It can be concluded that the predictive value of these models is low.
- Using 730 days of data the prediction is that the systems are improving in time. It is not possible to make an estimate of the moment steady state is achieved.
- Using 1000 days of data the prediction is also that the systems are improving in time. The form of the trend has only changed slightly. It is not possible to make an estimate of the moment steady state is achieved.

Model Goodness-of-fit
In figure 40 and 41 the histograms with the results of the chi-squared goodness-of-fit test of dataset 2A are presented. These are the chi-squared goodness-of-fit results on individual systems, for the models based on combined systems formula. The model is fitted to 900 days of data, the failures are divided into classes based on the rules for making classes explained in chapter III. The number of parameters p = 2 and the level of significance $\alpha = 0.05$. The results for dataset 2 and 2B are presented in appendix R.

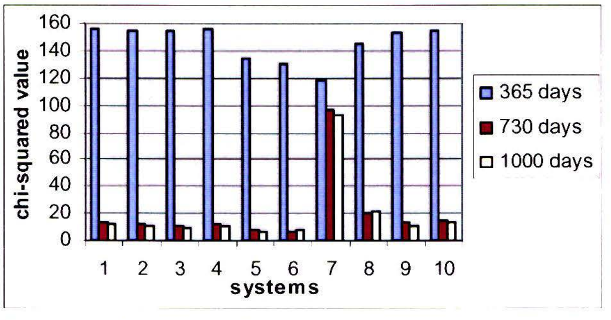


Figure 40 Dataset 2A Chi-squared unfiltered data

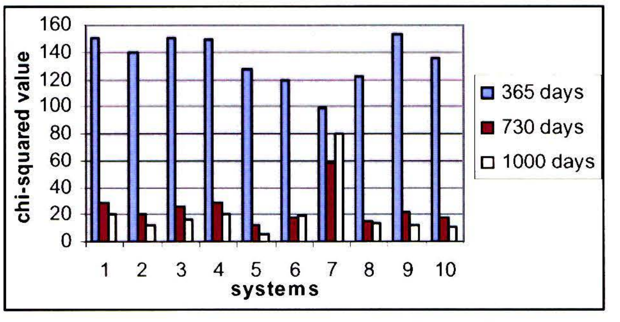


Figure 41 Dataset 2A Chi-squared filtered data

Conclusions of Chi-squared goodness-of-fit
- For the model based on dataset 2A
    - The chi-squared values of the model that is based on 365 days of data are extremely high indicating a very bad fit. This is in accordance with the remarks made earlier about the intensity and expected number of failures based on 365 days of data.
    - System 7 shows to be totally different from the other systems where the failure pattern is concerned, which was also seen for the power law model. In most other cases it shows that including more data leads to better fits.
- For the model based on dataset 2B and dataset 2 (appendix R).
    - Dataset 2B shows to have better fits to the models based on that data. Especially the fit of the model based on 365 days is much better compared to the model for dataset 2A.
    - When dataset 2 is observed as a whole the earlier made conclusion about system 7 is confirmed.
    - The models based on filtered data generally show a worse fit to the systems than the models based on unfiltered data.

In appendix S graphs are presented of the actual and expected number of failures for the individual systems of dataset 2 based on 365, 730, and 1000 days of data. The continuing line represents the model based on all systems of dataset 2. This graph gives an impression of the differences in individual failure patterns and how those lines compare to the continuing line based on all systems. The conclusion from these graphs is that the differences between the individual failure patterns cause a large spread around the continuing line.

$MTBF_c$
As explained in the single system analysis where the exponential law model of one system was analyzed there are no formulas found in literature for confidence bounds of the exponential law model. To get an idea of the spread around the $MTBF_c$ of dataset 2 (the thick black line in figure 42), based on 1000 days of data, the individual $MTBF_c$ plots of the 20 systems in dataset 2 are plotted as well. These give an idea the lines that the

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

thick black line is representing. Interesting to see is that the MTBF lines in figure 42 are almost straight lines. This in contrast to the lines modeled by Power law.

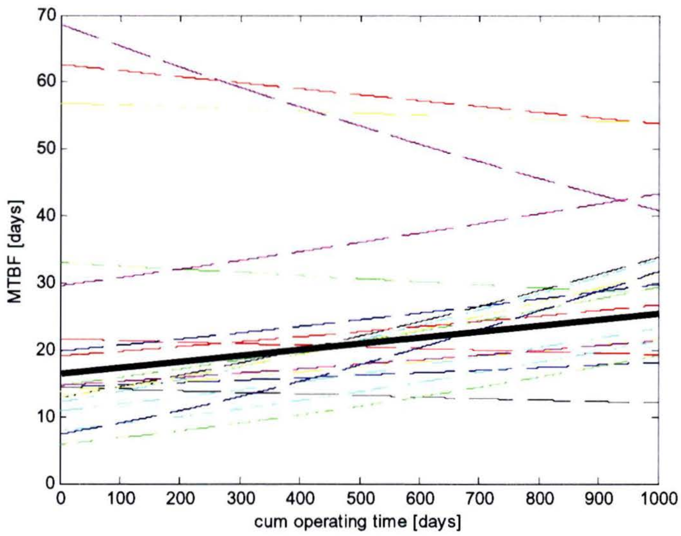


Figure 42 $MTBF_c$ of systems from dataset 2

The $MTBF_c$ line of dataset 2 starts around 17 days and ends around 23 days. The individual $MTBF_c$ lines start between 8 and 69 days and end between 12 and 54 days.

Overall conclusions of analysis two
The Power law model seems to have a reasonably good fit to the data when there is trend. The predictive value depends on the amount of data that the model is based on, based on 365 days of data the predictions made are not accurate. When 720 days of data is used, the predictive value is a lot better. More accurate data should be used to validate these conclusions.

For the exponential law model it can be concluded that the model based on 365 days of data shows a radically different failure pattern, especially for dataset 2A. The model based on 730 days of data is close to the model based on 1000 days of data indicating a much better fit to the data points based on these 730 days of data. The influence of the spread in data seems to be larger for the Exponential law model than for the Power law model, since dataset 2B (in appendix Q) shows a better fit for the model based on 365 days of data. The fact that no confidence intervals can be calculated for the Exponential law is a set back.

4.2.4 Analysis three: FMT data versus service data

Number of systems taken into account
The same eight systems are used for this analysis as the ones used in paragraph 3.2, where the FMT data for trend analysis is discussed.

Points of time chosen to fill into model
In this comparison between FMT data and service data there is a problem of comparing two different time scales, the operating time and the online time. Using the dates on which relevant failures took place according to the FMT data a time period is set for operating time of the service data, as explained in chapter III. Appendix G shows a graph in which the operating time is plotted against the online time. It shows that there is a linear relation between these two timelines. For 1000 hours of online time there is an average of 150 days of operating time.

Power law
First the parameter values of both the FMT data and the Service data from the chosen eight systems are calculated. These parameters are given in table 19.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Table 19 Parameter values

Data	System 1-8			
	FMT data		Service data	
	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\beta}$
1000 hours / 150 days	0.0307	0.8627	0.0689	0.9487

Both for the FMT data and the Service data the value of $\hat{\beta}$ is close to one, indicating there is almost no trend in these data. This is in line with the observations from the data in dataset two, where the first 365 days show almost no trend (or even a negative trend). Another observation is that the value $\hat{\lambda}$ for the Service data is more than double compared to the value of the FMT data.

Using the formula for the expected number of failures the following graphs are made (figure 43 and 44). These graphs visualize the difference between using FMT data to model the failure pattern and using Service data. It is clear that according to the model based on the Service data of these systems the number of failures after 150 days (1000 hours) is lower than the number of failures according to the FMT data. This shows the importance of the data source that is used to perform the calculations with; a different data source gives a different outcome.

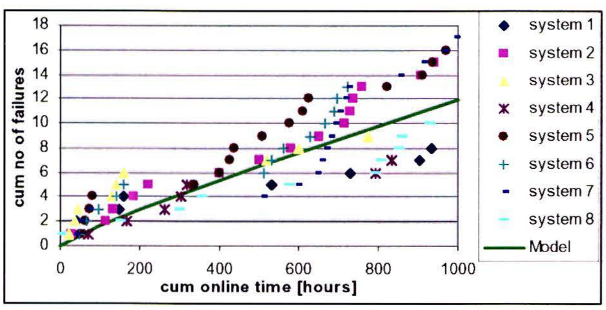


Figure 43 Expected number of failures FMT data

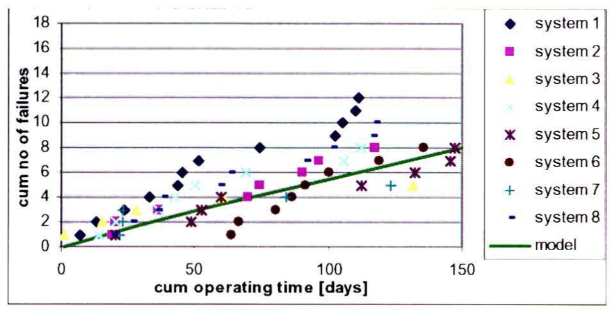


Figure 44 Expected number of failures Service data

Model Intensity $\hat{\mu}_1(t)$

The intensity functions of the FMT data and the Service data are given in table 20. When graphs would be made from the equations given in table 20 they would show that there are no bathtub curve effects in this data. This would indicate that these systems show no reliability improvement at all during their lifetime. However, based on the knowledge from the second data set it is plausible that later in the lifetime of these systems reliability growth will occur. Therefore, using these models for prediction purposes is not wise. Data from a longer period of time should be accumulated in order to be able to model the trend that presumably will be visible later on in the lifetime.

Table 20 Intensity, formula based on system 1-8

Data	FMT data	Service data
1000 hours / 150 days	$\hat{\mu}_1(T) = 0.0264T^{-0.137}$	$\hat{\mu}_1(T) = 0.0654T^{-0.051}$

Goodness-of-fit

In appendix T the results of the Cramér-von Mises and the Chi-squared goodness-of-fit test are given. Both tests show a similar view on the goodness-of fit. In the Cramér-von Mises test a significant fit is found for five of the eight systems, for both the FMT data and the Service data. This indicates that the spread of the data around the calculated model line is not very large. The Chi-squared test also accepts five out of the eight systems, although the accepted systems are not totally in accordance with the Cramér-von Mises test results.

MTBFc

In figure 45 and 46 the $MTBF_c$ is plotted together with the 90% confidence bounds. The increase in $MTBF_c$ value is much larger in the FMT data than the increase in the Service data. Another observation that can be made from these graphs is that the spread in the FMT data is much larger than the spread in the Service data.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Statements about the value of the MTBF$_c$ based on FMT data should therefore be made with great reservations.

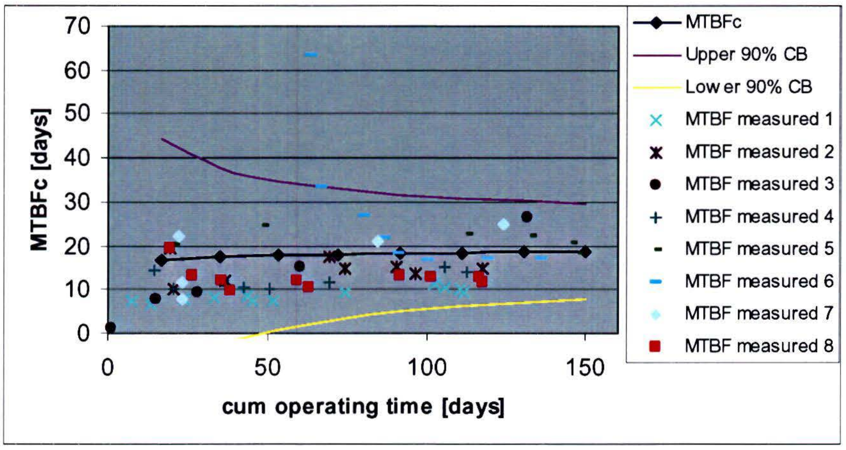


Figure 45 MTBF$_c$ based on Service data from eight systems

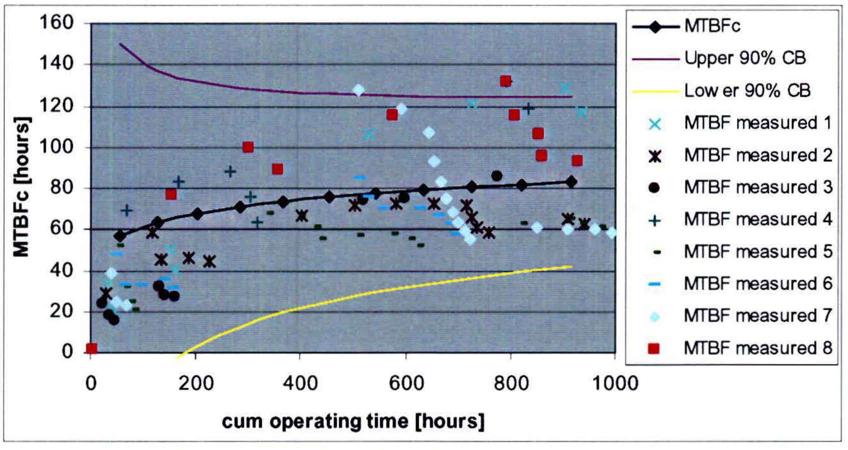


Figure 46 MTBF$_c$ based on FMT data from eight systems

Exponential law

In table 21 the parameter values of both the FMT data and the Service data from the chosen eight systems are given calculated using exponential law.

Table 21 Parameter values

Data	System 1-8			
	FMT data		Service data	
	α_0	α_1	α_0	α_1
1000 hours / 150 days	-4.5304	0.00047	-2.9766	0.00360

The differences found between the parameter values from the FMT data and the Service data are similar to the differences found using power law. The parameter values calculated using exponential law show almost no trend, like the values found using power law. However, interesting detail is that where the little bit of trend in the power law model is indicating increasing reliability, the little bit of trend in the exponential law model is indicating decreasing reliability! Since there is almost no trend in both cases the influence of this difference will not be large, it is however a totally different sort of trend.

Figures 47 and 48 show the expected number of failures based on the exponential law model. The modeled lines of both the FMT data and the Service data are increasing faster than the lines calculated using power law. This is due to the explained difference between decreasing and increasing reliability. The expected

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

number of failures after 1000 hours of online time calculated from the FMT data with exponential law is 14, while calculated by the power law system the expected number of failures is 12. A similar difference is visible for the Service data.

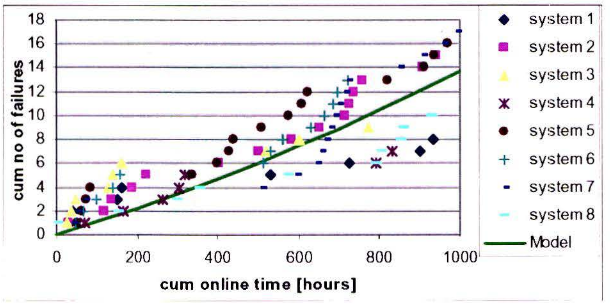


Figure 47 Expected number of failures FMT data

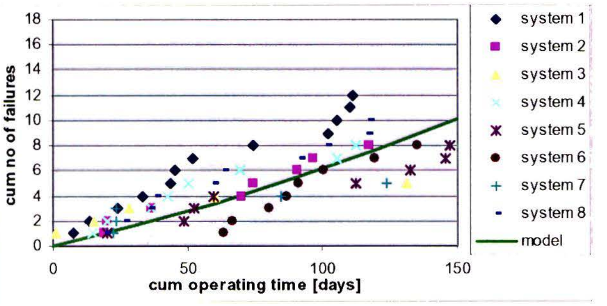


Figure 48 Expected number of failures Service data

Model Intensity $\hat{\mu}_2(t)$

In table 22 the equations for calculating the model intensity are given. As said before there is almost no trend in these models, plotting the ROCOF against the cumulative operating time is therefore not useful.

Table 22 Model intensity based on system 1-8

Data	FMT data	Service data
1000 hours / 150 days	$\hat{\mu}_2(T) = e^{-4.5304+0.00047T}$	$\hat{\mu}_2(T) = e^{-2.9766+0.003607T}$

Chi-squared goodness-of-fit test

When the values of the chi-squared test based on exponential law are compared with the values based on power law (both given in appendix T) the observation can be made that now for only four out of the eight systems there is a significant fit for the FMT data, and only three out of eight for the Service data. This indicates that for this dataset the power law gives a better fit to the data. Although this statement is based on the chi-squared test, which is not the most reliable goodness-of-fit test, the test is performed exactly the same way for the power law and the exponential law model and therefore seems credible.

4.3 Evaluation of the model performance

This chapter describes three analyses:

1. Analysis of service data of a single system
   This analysis investigated whether the power law model and the exponential law model are able to fit the failure pattern of the data that is used in this project.

   The analysis showed that both the power law and the exponential law model are able to fit the failure pattern of the data when enough data is provided to the model (1000 days of data). Based on a limited amount of data (365 days of data) both models give an incorrect prediction of the situation over a 900 days time period. The models based on 730 days of data are very close to the models based on 1000 days of data and therefore able to give a good prediction of the situation over a 900 days period of time.

   The $MTBF_c$ plots of Power law and Exponential law shows a clear difference between these lines; where the Power law model is able to fit the data points better in the beginning, the Exponential law model provides a better fit in the middle part of the plots.

2. Analysis of service data of multiple systems
   This analysis investigates the influence of using data from multiple systems for modeling the failure pattern. The analysis uses both unfiltered and filtered data to investigate the influence of filtering on the calculated model values.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

When multiple systems are used for the building of a model the diversity of the failure times of systems that are included has a big influence on the models. The more spread there is in the failure patterns of the systems, the less accurate the model can represent the data.

Similar to analysis of a single system the models based on multiple systems are not able to give a good prediction of the failure pattern based on a limited amount of data (365 days). When the model is based on 730 days of data the model is very close to the model based on 1000 days of data indicating a good prediction of the 900 days period, provided that the model based on 1000 days of data gives a good representation of the 900 days period. The fact that the goodness-of-fit test fits the model to the data points of the individual systems, which show a large spread in the failure pattern, is the reason that the fit is not significant in most cases. This means that it is hard to determine what line represents these systems the best.

The influence of pollution in the data is clearly visible in this analysis when the unfiltered and filtered plots are compared. The expected number of failures of the filtered data is lower than that of the unfiltered data. This difference between models based on filtered data and models based on unfiltered data proves that having a good data quality is essential for obtaining reliable answers from these models.

The question of which model type (Power law or Exponential law) is able to provide a better goodness-of-fit to the actual failure data can be found through analyzing the chi-squared goodness-of-fit of both the power law and the exponential law model. The Cramér-von Mises test cannot be used since this test can only be used for the power law model. An overall conclusion for these models based on the dataset used in this analysis of multiple systems is that the exponential law gives a slightly better fit than the power law model. But this difference is so small that the conclusion can be drawn that they are both equal when their performance is considered. Using the goodness-of-fit test when more than one system is included gives problems however as discussed on the top of this page.

The fact that they are equally able to model the data does not mean that they are the same. The analysis has shown that there are distinct differences in the models based on power law and exponential law in several cases, meaning that the power law model suggests an increasing reliability while the exponential law model suggests a decreasing reliability.

3. FMT data versus service data

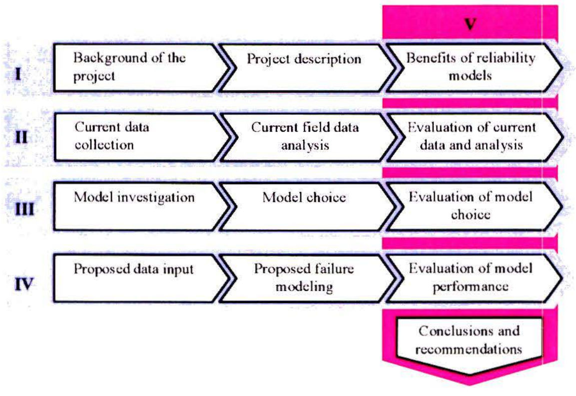
This analysis shows the differences between using FMT data and using service data when trying to establish the failure pattern of a system.

This analysis shows that the model based on FMT data shows a similar pattern as the model based on Service data. However, due to the fact that the actual data differs (the FMT failure moments do not coincide with the service failure moments), the models provide different model values.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Chapter V Conclusions and recommendations



Life can only be understood backwards, but it must be lived forwards.

(Soren Kierkegaard)

This chapter starts with the conclusions, which are directly related to the project research questions that were stated in chapter I. Subsequently the recommendations are presented including a proposal for further research.

5.1 Conclusions

The following aims of PMS and the QRE department at the TU/e were given in chapter I: For the QRE department the aim of the research field is to develop competence in quality and reliability by creating methods to predict the occurrence of product failures in the development process and early in the field introduction. PMS wants to be able to monitor, control and predict the product reliability in an earlier stage. In that way the feedback loop can be shortened, which leads to faster problem recognition.

This has lead to the following project formulation:

To investigate how the product failure pattern that is found at PMS can be modeled and to assess the prediction performance of those models.

The project research questions and targets stated in chapter I will be used as a framework for forming the conclusions of the project in this chapter.

Data questions and targets

Question 1: Is the data that is necessary for making a reliability prediction available?

Target a: Investigate the databases with field data and determine whether the necessary data for making a reliability prediction is available.

There are two sources with field data available: a Global Data Warehouse database where service data is collected and a Field Monitoring Team database where system software loggings are collected. Both data sources are able to provide the necessary data for modeling system reliability. That is, they provide the failure moments relative to the moment of installation.

Target b: Investigate the quality of the data to determine how accurate the results will be.

Both sources of data have problems related to data quality, although there are more problems with the service data. The service data has data quality problems related to all four data quality metrics, that is, completeness, consistency, timeliness, and accuracy.

<u>Positive points of service data</u>

- Data from all system types and versions is available.
- A broad spectrum of failures is covered, including failures that would not lead to a software logging (like in the FMT data).

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Negative points of service data

- Completeness: far from complete on data set level; only data from a number of countries is available. On a data element level the failure moments are generally complete only for the period that the systems are under warranty or CSA, and there are problems with the completeness of the data necessary for the determination of severity of the failures.
- Consistency: the wrong transfer of data between databases, differences between procedures throughout the world, and personnel that has a certain amount of freedom when filling out the jobsheets lead to consistency problems.
- Timeliness: timely availability of data has to improve, especially if considered that the FMT data is updated daily.
- Accuracy: failure moment registration has to improve since there is too much diversity in the registration of moment of failure.

The FMT data has data quality problems related to completeness and accuracy.

Positive points of FMT data

- Consistency: the failure data is consistent since all loggings are evaluated using the same description of a relevant failure.
- Timeliness: the database is update with new failure data on a daily basis.

Negative points of FMT data

- Completeness: far from complete on data set level since only data from the newest system version is available.
- Accuracy: no certainty on whether the list of registered relevant failures is an accurate representation of all relevant failures.

The data quality problems with the service data, which were the largest, lead to defining a specific group of systems for the analysis of this service data. That way a part of the data quality problems could be coped with. Only cardio monoplane systems were considered, other system types were left out of the analysis; the data was taken from systems in the USA only, to improve the consistency of the data; only systems under warranty or CSA were analyzed; and a manual filtering was used to filter out calls wrongly booked as corrective maintenance.

Model questions and targets

Question 2: What model(s) are available for the situation at PMS?

Target: Identify a reliability prediction model, based on field data, which is applicable for the situation of Philips Medical Systems.

Answer: The category of models that is found to be applicable for situation of PMS has the following characteristics: continues parametric models for repairable systems with a mixture of hardware and software, without redundant parts. The model choice can be made using the general procedure for analyzing failure data of a repairable system. The fact that the models that are considered by this general procedure are designed for hardware failures indicates that the assumption is made that both hardware and software failures can be considered by these models. There are however differences between these types of failures.

To determine whether there is trend in the failure data or not the Laplace trend test is performed. Conclusion from the individual Laplace tests is that, based on 1000 days of data, the systems show a heterogeneous picture with some of the systems showing significant trend, while other systems do not show this significant trend. The combined Laplace test indicates significant trend in five out of the six calculations, based on 1000 days of data. This leads to the choice for Non Homogeneous Poisson Process models. There are two models in literature that are most used and described as useful NHPP models for analyzing the failure pattern of repairable products: Power law and Exponential law. Both these models are investigated in this report.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Analyses questions and targets

Question 3: Can these models give the expected early insight into future reliability?

Target: Use field data to determine parameter values of the model, determine the goodness-of-fit, and construct the confidence intervals. Use this to determine whether early prediction is possible.

Answer: In chapter IV three analyses of datasets are presented, the answer to this third question will therefore be separated in three parts, A, B and C.

Part A: Analysis of service data of a single system.
This analysis was performed to investigate whether the power law model and the exponential law model are able to fit the failure pattern of the data that is used in this project.

A first conclusion is that both modeling types, power law and exponential law, are able to fit the data of a single system based on 1000 days of data. But when prediction of the failure pattern for 900 days is concerned based on a limited amount of data (365 days), both model types give a wrong prediction of future failures. This has to do with the following: the models are flexible in a way that they can model a deteriorating system, an improving system and a system that shows no trend (a Homogeneous Poisson Process). The form that the model takes depends on the data that is put in the model. This means that the failure pattern early after field introduction of the system determines the pattern of the model. If a trend develops later on in the product life this will not be modeled when only using data from the period shortly after field introduction. The data needs to show at least a small amount of trend in order for the model to pick up the sign that the system is actually improving in time. When the model is based on 730 days of data it is able to predict the failure pattern of a 900 days period correctly.

Part B: Analysis of service data of multiple systems.
This analysis was performed to investigate the influence of using data from multiple systems for modeling the failure pattern.

A first conclusion here is that both the power law and the exponential law models adapt well to the data from the number of systems included in the model building. Next to that, similar effects as when using a single system for building the model can be found, meaning an inaccurate representation of the failure pattern when the data of 365 days is used. This means that when the model is based on 365 days of data this model can only be used to determine the situation at that moment, not to predict the future failure pattern. The model based on 730 days of data is very close to the model based on 1000 days of data, like with the model based on a single system. This means that based on 730 days of data of good prediction of a 900 days period can be made.

The large difference in the number of failures that occurs per system has a big influence on the fit of the model to the data. The model is of course an average and therefore it will not fit individual systems well, especially if the systems show such a large difference in number of failures. This spread in failure pattern between different systems makes the predictive value less accurate and causes relatively wide confidence bounds. When the spread in the failure pattern is smaller this leads to predictions that are closer to the actual failure pattern, although based on 365 days of data it still deviates too much to be called a good prediction.

The power law and exponential law model do not always agree on the form the model should take. This even means that in some cases, based on the same data, one model type calculates an increasing reliability, while the other model type calculates a decreasing reliability. The cause of this can be found in the large difference in the number of failures that occur per system. One model type reacts differently than the other in case of such a diverse failure pattern between systems. When the failure patterns are not consistent the models neither will be.

The influence of pollution in the data was investigated by comparing filtered data to unfiltered data. The filtered data shows similar failure patterns as the unfiltered data, although clearly different lines appear when the models are plotted. This means that the filtered data leads to

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

different values of the expected number of failures, the model intensity, the goodness-of-fit, and the MTBF. It proves that having a good data quality is essential for obtaining reliable answers from these models.

Part C: FMT data versus service data.
This analysis shows the differences between using FMT data and using service data.

The difference between the two data sources is clearly visible when they are compared; the failure moments, as well as the number of failures do not correspond. This leads to differences in the model based on FMT data and the model based on service data. That leads to maybe one of the most important conclusions from this report, that is, there first needs to be a clear understanding of the data that is put in the model before conclusions can be drawn about the final number that the model gives as an output. The quality of a model can be only as good as the quality of the data that is put in.

Aim of the project
The aim of this project was to investigate how the product failure pattern that is found at PMS can be modeled and to assess the prediction performance of those models. The direct answer to this aim is that the investigation for a model lead to the choice of the NHPP power law and exponential law models, which are able to predict the failure pattern when based on enough data to pick up signs of trend. For the datasets used in this project this is not after 365 days, but after 730 days.

Use of the model
In chapter I the potential benefits of prediction of product reliability using field data were discussed. It described the use for: determination of the current situation; evaluation and control; warranty cost assessment; resource assessment; learning organization. Next to that it could serve as a basis for root cause analysis. Question then is whether the models researched in this project can be put into practice now in order to benefit from this potential. Conclusion is that it is too early for using the models for these purposes. In the recommendations an answer will be given on what needs to be done in order to arrive at the desired situation.

5.2 Recommendations

The recommendations are split into two parts, first the last question that was formulated in paragraph 1.2.2 will be answered. After that the recommendations on further research will be explained.

The fourth question with a project target described in paragraph 1.2.2 was as follows:
Question 4: What are the improvement opportunities?
Target: Make recommendations on the data collection and handling processes and on reliability prediction models to improve the accuracy of the prediction.

Answer: First the recommendations on the data collection will be given, then on the handling processes and last on the models.

Data collection: The most basic question in this respect is the question of which data source to use for making reliability analyses and predictions, service data or FMT data. Both sources of data have positive sides and negative sides as explained in the conclusions. But a big difference exists in what needs to be done to influence the quality of data that is produced by these two data sources.

Part A: When service data is continued to be used the following points should be worked on:
<u>Database and data transfer</u>
Actions have to be taken in improving the completeness and consistency of the databases by creating one database system. Promising are the actions taken by the MENHIR project to improve the problems related to the databases. Since this is already being worked on at the moment no extra actions need to be taken in this respect. However, at least until the MENHIR project is in place problems with data completeness and consistency of the data will remain.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Personnel
The value of complete, consistent, accurate, timely and valid data for the purpose of analysis should be a part of the training of field service engineers and other personnel involved in the creation of failure data. This should lead to the creation of awareness of providing data of good quality. The problem is that there are differences of interest between the SSRs and the Business Units. The SSRs have no primary interest in providing reliability data; their main interest when filling out the jobsheets is to make sure they are able to account for the hours booked and to make sure the billing is done correctly.

Procedures
Actions have to be taken in improving the consistency and timeliness of the data by adaptation of procedures. Registration of the online time at the moment a job is performed on a system should be done when possible.

For some problems the origin of the problem can be unclear, being for example whether it is a human error or a procedural error. The discussion in this report about data quality is not exhaustive. To find out more about the root causes of data quality problems a study should be performed at the service department (SSR and/or SSD).

Part B: When FMT data will be used the following points should be worked on:
An investigation is necessary to determine to what extend the failure data produced by the FMT tool contains the relevant failures of the systems. This comes down to answering the following two questions:
- To what extend is the tool able to log the relevant types of failure? The fact that this will never be 100% is already certain since the software loggings are unable to detect failures that have no link to the software whatsoever, like a broken remote control for example.
- Are the registered relevant failures giving a complete and valid representation of the failures defined as relevant failures? This is questionable since it is the opinion of a system engineer, which is not always unambiguous.

In general: If not all problems of the service data and FMT data can be worked on at the same time a division of the metrics has to be made indicating the importance of each metric. For instance first concentrating on improving the consistency data, then on the accuracy, then on completeness and last on the timeliness. The priority depends on the severity of problems related to the metrics.

Independent of the data source that is used attention should be given to the gravity of the failures that occur. The use of a scale to categorize the failures leads to more insight in the sort of failures that occur. This leads to a further refinement of the relevant failure definition. The service data is already capable of providing such a categorization with its priority code, if this code would be available in the Global Data Warehouse.

In addition, the following categorization was found in literature [45] and could be used in making a distinction in the data now booked as "failure".
Type I - *Failure* – Severe operational incidents that would definitely result in a service call, such as part failures, unrecoverable equipment hangs, DOAs, consumables that fail/deplete before their specified life, onset of noise, and other critical problems. These constitute "hard-core" failure modes that would require the services of a trained repair technician to recover.
Type II - *Intervention* - Any unplanned occurrence or failure of product mission that requires the user to manually adjust or otherwise intervene with the product or its output. These tend to be "nuisance failures" that can be recovered by the customer, or with the aid of phone support. Depending on the nature of the failure mode, groups of the Type II failures could be upgraded to Type I if they exceed a predefined frequency of occurrence.
Type III - *Event* - Events will include all other occurrences that do not fall into either of the categories above. This might include events that cannot be directly classified as failures, but

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

whose frequency is of engineering interest and would be appropriate for statistical analysis. Examples include failures caused by test equipment malfunction or operator error.

Data handling: My recommendation for the future situation would be to use FMT data for reliability analysis, and to use the service data next to that for other analyses like call rate, material usage and material cost analysis. When data from the FMT becomes available over a longer period of time this should be used to model the total system life. Furthermore, the failure moments stored in the FMT database provided for the analyses in online hours is detailed enough for the calculations made with this data, it is not necessary to provide this data in minutes or seconds.

Models: Choice between the power law model, the exponential law model, or another type of model. Both the power law model and the exponential law model have showed to be able to correctly fit to the failure pattern of the systems. And although there are some differences in the failure pattern that they show no definite conclusions about one model type being better than the other can be made. The problem with both these models is that early prediction based on data from a relatively small period of time (365 days) is not possible given this particular failure pattern shown by the systems analyzed in this project. This means that the search for a different model could be undertaken, looking for a model where earlier prediction of the reliability might be possible.

Further research

A follow up should focus on the following aspects:

1. Improving data quality of FMT tool
   It is essential that the data quality of the data provided by the FMT tool is as good as possible. Since this tool is only being used since very recent for making reliability analyses there probably is room for improvement. Especially the question of accuracy of the failure data is important to research extensively.

2. Analysis of more FMT data sets
   This project provides the first analysis based on FMT field data. More analyses based on data of a larger time period need to be performed in order to conclude whether the behavior of the FMT data is similar to the service data.

3. Research on new model that makes a better prediction based on a small amount of data (early data).
   The analyses have indicated that the power law and exponential law model are not able to predict the failure pattern based on a small amount of data; that is, not based on the datasets that were used in this project. Therefore, if a different model can be found or developed which is able to make a better prediction based on a small amount of data this would mean a significant improvement.

4. Implementation of the model into the business procedures.
   When a model is found giving better early predictions, or when it is decided to use the power law or exponential law model given its limitations, a good implementation of the analyses that can be performed with these models needs to be assured. The model analyses need to fit in the procedures of the organization.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Literature

[1] Kales, P. (1998), "Reliability, for technology, engineering, and management", Upper Saddle River, Prentice-Hall Inc.

[2] Internet site of Medical.Philips.com in subsection 'about us', date visited: 14-01-2004

[3] Intranet site of Philips Medical systems, visited 20-03-2004

[4] MD Buyline intelligence report: Full Report, Angiography, Cardiac Cath and Special Procedures, 14-10-2003, internal PMS report

[5] Oh, Y.S.; Bai, D.S. (2001), "Field data analyses with additional after warranty failure data", Reliability Engineering and System Safety **72,** p 1-8

[6] Rai, B; Singh, N (2003), "Hazard rate estimation form incomplete and unclean warranty data", Reliability Engineering and System Safety **81,** p 79-92

[7] Blanks, H.S. (1992), "Reliability in procurement and use, from specification to replacement", Chichester, John Wiley & Sons Ltd

[8] Zaino, N.A.; Berke, T.M. (1992), "Determining the effectiveness of Run-In: a case study in the analysis of repairable system data", Proceedings of the Annual Reliability and Maintainability Symposium 1992, p 58-70

[9] Blanks, H.S. (1998), "The Challenge of Quantitative Reliability", Quality and Reliability Engineering International **14,** p 167-176

[10] Sander, P.C.; Toscano, L.; Luitjens S.; Petkova, V.; Huijben, A.; Brombacher, A.C. (2003), "Warranty Data Analysis for Assessing Product Reliability", Blischke & Murthy (eds): Case studies in Reliability and Maintenance, Chichester, Wiley-Interscience

[11] Petkova, V.T.; Sander, P.C.; Brombacher, A.C. (2000), "The use of quality metrics in service centers", International journal of Production Economics **67,** p 27-36

[12] Ion, R.A.; Petkova, V.T.; Peeters, B.H.J.; Sander, P.C. (2003), "Early reliability prediction in consumer electronics", paper submitted and under revision at the journal Quality and Reliability Engineering.

[13] Online dictionary of Computing, http://foldoc.doc.ic.ac.uk/foldoc/dictionary.gz, 1993-2004 Denis Howe, date visited: 06-07-2004

[14] Roos, C.S. (2004), "Reliability Groei bij PMS Cardiovascular", Technische Universiteit Eindhoven

[15] Limnios, N.; Nikulin, M. (2000), "Recent advances in reliability theory: methodology, practice, and inference", Birkhäuser, Boston.

[16] Pham, H. et al. (2003), "Handbook of reliability engineering", Springer-Verlag London.

[17] Rudra, A.; Yeo, E. (1999), Key Issues in Achieving Data Quality and Consistency in Data Warehousing among Large Organisations in Australia", Proceedings of the 32nd International Conference on System Sciences

[18] Blischke , W.R.; Murthy, P.D.N. (2003), "Case studies in Reliability and Maintenance", Chichester, Wiley-Interscience

[19] Ballou, D.P.; Pazer, H.L. (1985), "Modeling data and process quality in multi-input, multi-output information systems", Management Science **31** No 2, p 150-162

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

[20] Wang, R.Y.; Strong, D. (1996), "Beyond accuracy: what data quality means to data consumers", Journal for Management Information Systems **12** No 4 p 5-34

[21] Ballou, D.P.; Pazer, H.L. (2003), "Modeling Completeness versus Consistency Tradeoffs in Information Decision Contexts", IEEE Transactions on knowledge and data engineering, **15** No 1, p 240-243

[22] Alter, S. (2002), "Information Systems, the foundation of E-business", Pearson Education, Inc. New Yersey

[23] Thoma, A.; Tholen, R. (2004), "MENHIR", internal PMS report

[24] Asten, M. v. (2004), Field data analysis – Quarterly report, internal PMS report

[25] Ascher, H.E.; Feingold, H. (1984) "Repairable Systems Reliability, modeling, inference, misconceptions and their causes", New York, Marcel Dekker, Inc.

[26] Dhillon, B.S. (1980), "Reliability Growth: a Survey", Microelectronics Reliability **20,** p 743-751

[27] Blischke, W.R.; Murthy, P.D.N. (2000), "Reliability: Modeling, Prediction and Optimization", New York: Wiley-Interscience

[28] Arkin, B.L.; Leemis, M. (2000), "Nonparametric estimation of the cumulative intensity function for a Nonhomogeneous Poisson Process from overlapping realizations", Management Science **46**, 989-998

[29] Seo, J.H.; Jang, J.S.; Bai, D.S. (2003) "Lifetime and reliability estimation of repairable redundant system subject to periodic alternation" Reliability Engineering and System Safety **80,** p 197-204

[30] Majeske, K.D. (2003), "A mixture model for automobile warranty data", Reliability Engineering and System Safety **81,** p 71-77

[31] Chen, T; Popova, E (2002), "Maintenance policies with two-dimensional warranty", Reliability Engineering and System Safety **77,** p 61-69

[32] Lu, M-W. (1998), "Automotive reliability prediction based on early field failure warranty data", Quality and Reliability Engineering International **14,** p 103-108

[33] Xu, K.; Xie, M.; Tang, L.C.; Ho, S.L. (2003), "Application of neural networks in forecasting engine systems reliability", Applied Soft Computing **2,** p 255-268

[34] Crow, L.H. (1993), "Confidence Intervals on the Reliability of Repairable Systems", Proceedings of the Annual Reliability and Maintainability Symposium 1993, p 126-133

[35] Weckmar, G.R.; Shell, R.L.; Marvel, J.H. (2001), "Modeling the reliability of repairable systems in the aviation industry", Computers & Industrial Engineering **40,** p 51-63

[36] Kvaløy, J.T.; Lindqvist, B.H. (1998), "TTT-based tests for trend in repairable systems data", Reliability Engineering and System Safety **60,** p 13-28

[37] Ascher, H.E.; Hansen, C.K. (1998), "Spurious Exponentiality observed when incorrectly fitting a distribution to nonstationary data", IEEE transactions on reliability **47** p 451-459

[38] Coetzee, J.L. (1997), "The role of NHPP models in the practical analysis of maintenance failure data", Reliability Engineering and System Safety **56,** p 161-168

[39] IEC 1164 (1995), "Reliability growth – Statistical test and estimation methods", Geneva: IEC 1970 – 2000

[40] Natrella M, et al. (2003), NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

[41] Pulcini, G. (2001), "Modeling the failure data of a repairable equipment with bathtub type failure intensity", Reliability Engineering and System Safety **71** p 209-218

[42] Cox, D.R.; Lewis, P.A.W. (1966), "The statistical Analysis of Series of Events", London: Methuen & Co LTD, New York: John Wiley & Sons Inc

[43] Montgomery, D.C.; Runger, G.C. (1994), "Applied statistics and probability for engineers", New York, John Wiley & Sons, Inc.

[44] Yamada, S.; Osaki, S. (1983), "Reliability growth models for hardware and software systems based on Nonhomogeneous Poisson Processes: a survey", Microelectronics reliability **23** No 1, p 91-112

[45] Internet site of Reliasoft, www.reliasoft.com and rg.reliasoft.com, visited 04-06-2004

[46] IEC 1014 (1989), "Programmes for reliability growth", Geneva: IEC 1970 – 2000

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendices

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix A Definitions

Failure
Given by the online dictionary of computing [13]:

> *The inability of a system or system component to perform a required function within specified limits.*

Model
Definition of a model given by the online dictionary of computing [13]:

> *A model is a description of observed behavior, simplified by ignoring certain details. Models allow complex systems to be understood and their behavior predicted within the scope of the model, but may give incorrect descriptions and predictions for situations outside the realm of their intended use.*

Relevant failure
A definition of a relevant failure used in the IEC 1014 standard [46] is:

> *A failure that should be included in interpreting test or operational results or in calculating the value of a reliability performance measure.*

The definition of a relevant failure that will be used in this report is:

> *Relevant failures are all failures that a customer can be confronted with.*

In this definition reduction of performance of a certain function will not be taken into account while it cannot be retrieved from the available data. The criterion for determination of failures that a customer <u>can</u> be confronted with is 'the judgment of development engineers'.

Reliability
Detailed definition of reliability given by Kales[1]:

> *The reliability of an item is the probability that the item will perform a specified function under specified operational and environmental conditions, at and throughout a specified time.*
>
> *This means that before we can deal with reliability, the producer and the user must reach formal agreements on what the product is to do, how the user is to use the product, the range of environments under which the product is expected to perform satisfactorily, and the instant or duration in time that the performance of the product or service is demanded.*

Trend
Definition of trend [36]:

> *There is a trend in the pattern of failures if the inter-arrival times tend to alter in some systematic way, which means that the inter-arrival times are not identically distributed.*

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix B Procedures for calls and jobs

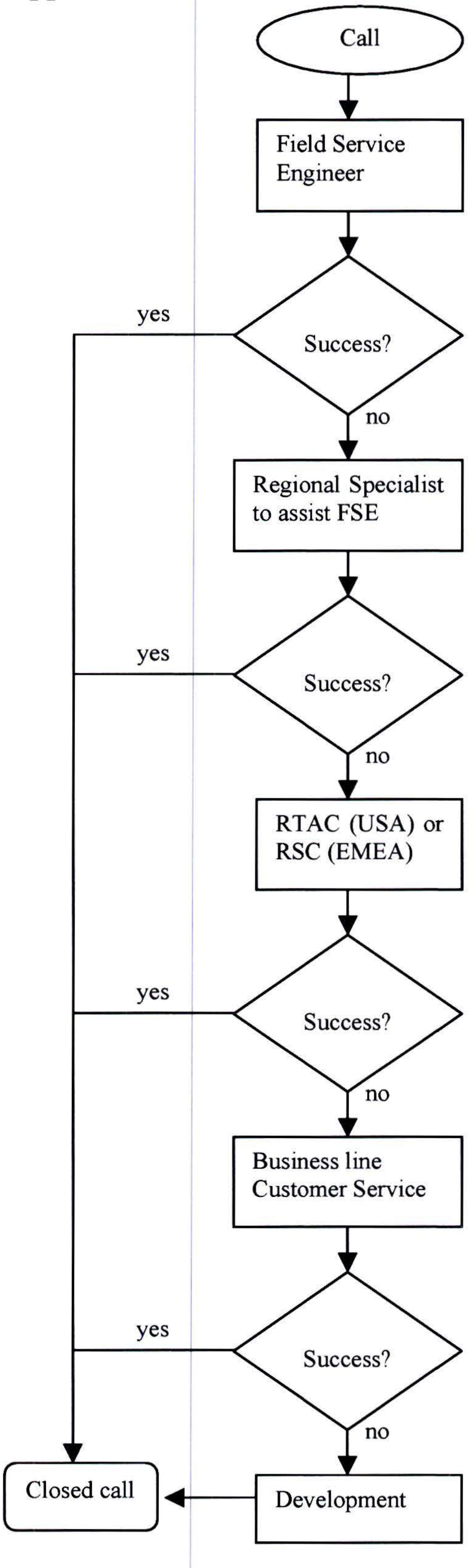


Figure 49 Call flow diagram

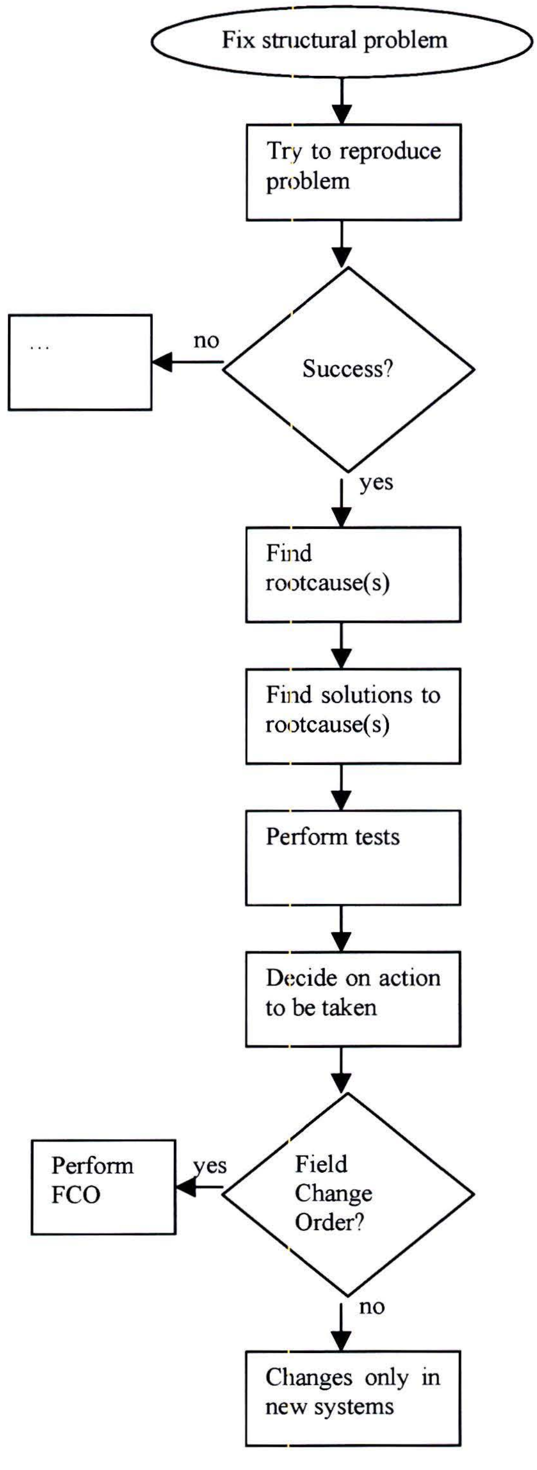


Figure 50 Problem fixing flow diagram

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

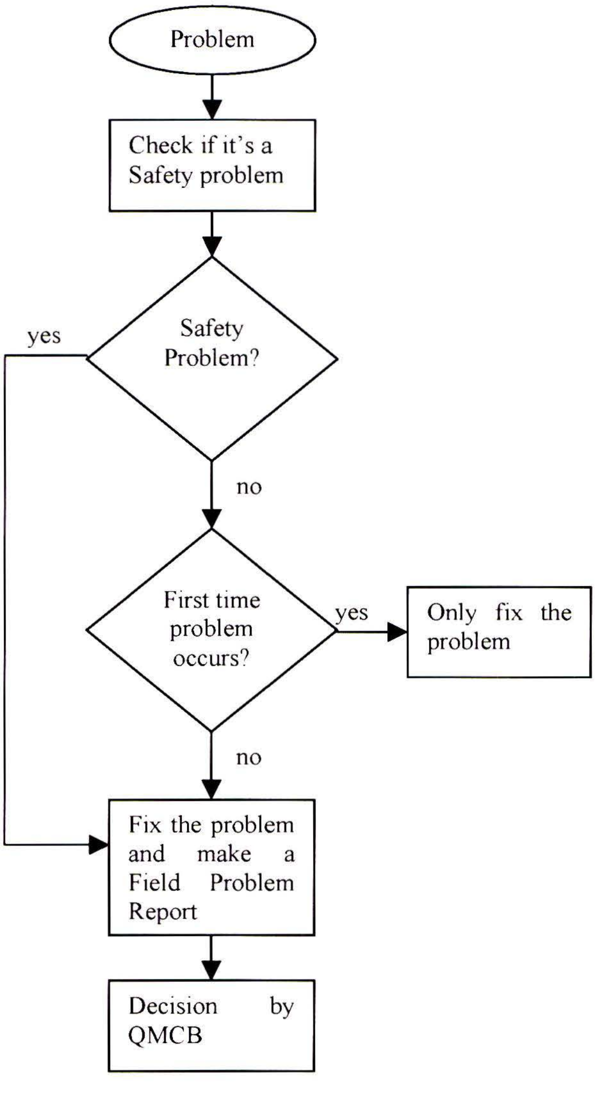


Figure 51 Problem flow diagram

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix C System versions

Table 23 System versions

Description	Family	MoBi	Applic
Allura 9C Xper FD	Allura Xper	mono	card
Allura 9F Xper FD	Allura Xper	mono	card
Allura 9C mo	Allura	mono	card
Allura 9F mo	Allura	mono	card
Allura 9 bi	Allura	bipl	card
Ints CV	Integris 5000	mono	mix
Allura 12/15 mo	Allura	mono	vasc
Allura 12/15 bi	Allura	bipl	vasc
V3000	Integris 3000	mono	vasc
V3000	Integris 3000	mono	vasc
V3000	Integris 3000	mono	vasc
V4000	Integris 3000	mono	vasc
V4000	Integris 3000	mono	vasc
V4000	Integris 3000	mono	vasc
H3000	Integris 3000	mono	card
HM3000	Integris 3000	mono	mix
BN/BV3000 Bi	Integris 3000	bipl	vasc
BN/BV3000 Mo	Integris 3000	mono	vasc
BH3000	Integris 3000	bipl	card
V3000P	Integris 3000	mono	vasc
V3000P	Integris 3000	mono	vasc
V3000P	Integris 3000	mono	vasc
H5000	Integris 5000	mono	card
BH5000	Integris 5000	bipl	card
V5000	Integris 5000	mono	vasc
BV5000	Integris 5000	bipl	vasc
Allura 9F mo FD	Allura FD	mono	card
Allura 9C mo FD	Allura FD	mono	card

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix D Call open and close date investigation

Four randomly chosen systems to investigate the influence of choosing the start date or the close date of a call on the number of calls per month. M stands for month, so M0 is 0 months after the system was taken into use.

Table 24 Call information according to overview

System a

	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Start	1	4	2	0	3	6	3	1	2	3	2	2	1
End	0	2	0	1	2	10	2	2	2	3	1	4	0
Diff	1	2	2	1	1	4	1	1	0	0	1	2	1

System b

Start	0	2	11	7	3	1	0	1	3	1	3	0	0
End	0	2	7	11	2	2	0	1	3	1	3	0	0
Diff	0	0	4	4	1	1	0	0	0	0	0	0	0

System c

Start	1	5	3	1	2	4	9	4	7	2	3	2	0
End	1	0	3	3	3	5	2	5	7	6	3	3	0
Diff	0	5	0	2	1	1	7	1	0	4	0	1	0

System d

Start	0	5	7	7	2	5	6	3	10	3	2	2	2
End	0	5	0	13	3	4	5	3	6	6	5	1	2
Diff	0	0	7	6	1	1	1	0	4	3	3	1	0

Conclusions drawn from the data of the four systems:

- The difference between the starting date and the closing date of a call can cause large differences in the number of calls in a month.
- The difference between opening and closing date of a call varies from 'a couple of hours' to 'more than four months'.
- The start date of a job (within a call) is sometimes earlier than the start date of the call itself. Sometimes this can be more than a month difference. This is probably due to a call on the same problem that has been worked on before. The job start date is then equal to the job start date of the job that handled the same problem earlier. This can be caused by the time of ordering parts.
- The other way around also happens; the start date of the job is then a lot later than the start date of the call. Question is what happens in the time between the start of the call and the start of the job.
- There are cases of calls with a job open for a month while only a couple of hours are booked within this period.
- Problems are not always solved right away; there is an example where one problem that leads to six separate calls.

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix E Critical values for Cramér-von Mises

Table 25 Critical values for Cramér-von Mises goodness-of-fit test at 10% level of significance

n	Critical value of statistic
3	0.154
4	0.155
5	0.160
6	0.162
7	0.165
8	0.165
9	0.167
10	0.167
11	0.169
12	0.169
13	0.169
14	0.169
15	0.169
16	0.171
17	0.171
18	0.171
19	0.171
20	0.172
30	0.172
60	0.173

This table is taken from 1164 IEC (1995) [29]

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix F Individual Laplace test

Table 26 Results individual Laplace test, service data

System	Unfiltered	Significant trend		Filtered	Significant trend	
1	0.390	**No**		-1.762	**No**	
2	0.096	No		-0.228	No	
3	-2.813		Yes	-4.136		Yes
4	-1.022	**No**		1.325	**No**	
5	1.231	No		-0.221	No	
6	-0.715	No		-1.440	No	
7	1.473	No		0.233	No	
8	-0.029	No		-1.683		Yes
9	-1.535	No		-2.542		Yes
10	-2.423		Yes	-3.394		Yes
11	-0.633	**No**		0.971	**No**	
12	-2.073		Yes	-2.075		Yes
13	-0.605	**No**		-1.255	**No**	
14	-3.690		Yes	-3.162		Yes
15	-3.051		Yes	-2.917		Yes
16	-1.089	**No**		-1.744		Yes
17	-3.523		Yes	-4.487		Yes
18	-0.499	**No**		-0.781	**No**	
19	-4.275		Yes	-5.150		Yes
20	-4.968		Yes	-4.639		Yes

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix G Relation between operating time and online time

Figure 52 shows the relation between the cumulative operating time in days and the cumulative online time in hours. It can be concluded that for all systems in the graph there is a linear relation between the operating time and the online time.

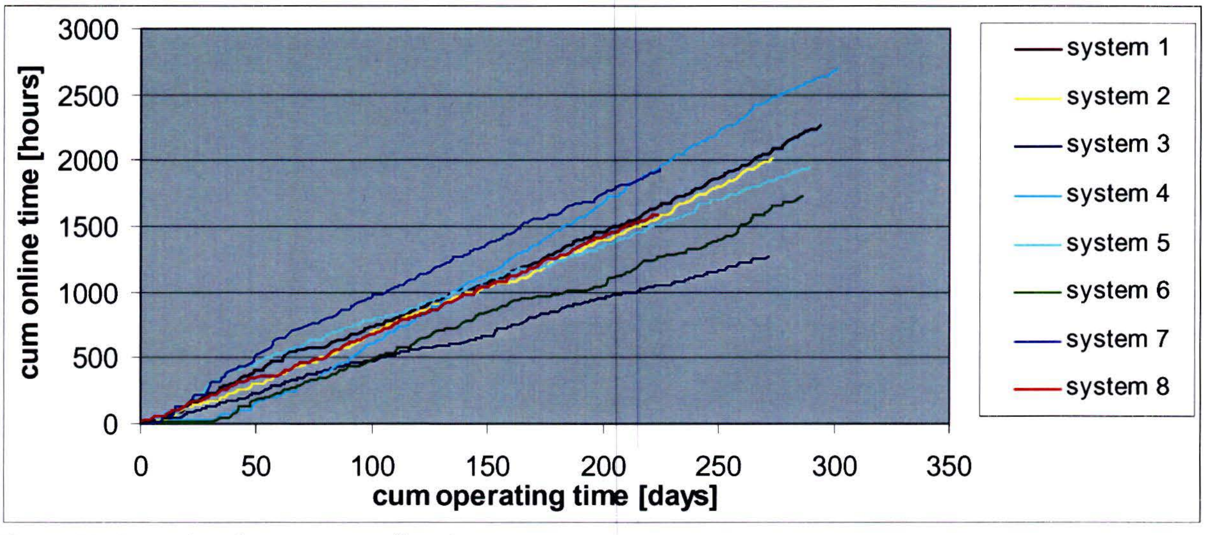


Figure 52 Operating time versus online time

In table 27 the number of days connected to 1000 hours of online time is given. From this the average and standard deviation are calculated showing an average of 150 days of operating time connected to 1000 hours of online time.

An interpretation of this relation between the operating time and the online time is that the system is online for 6.67 hours per day (1000/150 = 6.67 hours per day) if the system is used every day of those 150 days. This would mean an online time of 9.33 hours per day if the system is used 5 days per week.

Table 27

System	Hours	Days
1	1000	137
2	1000	145
3	1000	214
4	1000	135
5	1000	137
6	1000	185
7	1000	106
8	1000	144
Average		150
Standard deviation		33.55

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix H Difference between filtered and unfiltered data

Table 28 Failures of filtered and unfiltered data for the 20 systems

System	365 days			730 days			1000 days		
	Failures		Diff	Failures		Diff	Failures		Diff
	Unfiltered	Filtered		Unfiltered	Filtered		Unfiltered	Filtered	
1	8	3	63%	14	4	71%	16	4	75%
2	8	3	63%	14	7	50%	17	9	47%
3	20	12	40%	27	16	41%	29	16	45%
4	9	1	89%	11	1	91%	16	4	75%
5	12	6	50%	27	15	44%	33	18	45%
6	23	17	26%	43	35	19%	54	42	22%
7	28	22	21%	61	51	16%	78	65	17%
8	24	18	25%	41	32	22%	46	36	22%
9	12	9	25%	17	11	35%	20	12	40%
10	27	19	30%	37	26	30%	39	28	28%
11	18	12	33%	35	25	29%	51	39	24%
12	17	12	29%	29	22	24%	33	25	24%
13	17	11	35%	28	18	36%	31	21	32%
14	22	16	27%	25	16	36%	29	19	34%
15	21	14	33%	27	18	33%	31	21	32%
16	23	18	22%	37	30	19%	44	37	16%
17	34	27	21%	42	34	19%	45	36	20%
18	13	5	62%	31	16	48%	34	19	44%
19	39	27	31%	49	35	29%	51	36	29%
20	26	14	46%	29	14	52%	31	15	52%
Av.	20	13	34%	31	21	32%	36	25	31%

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix I Graphs of the failure versus operating time of systems 1-20

These are the systems that are used for the trend analyses and the fitting of the models.

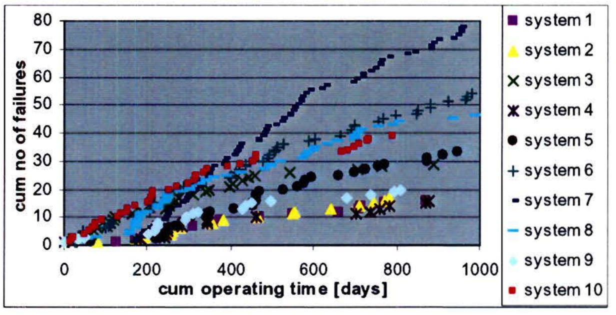


Figure 53 Failure pattern of system 1-10, unfil. data

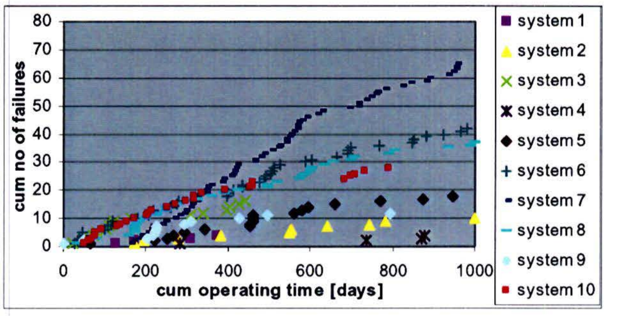


Figure 54 Failure pattern of system 1-10, fil. data

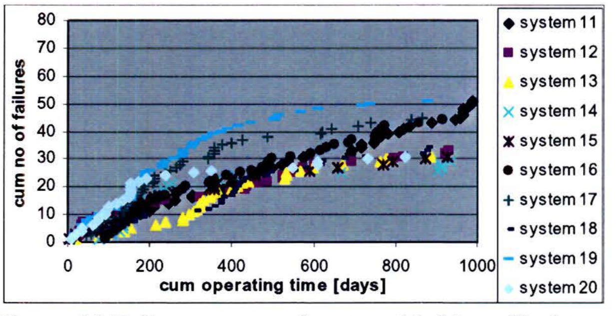


Figure 55 Failure pattern of system 11-20, unfil. data

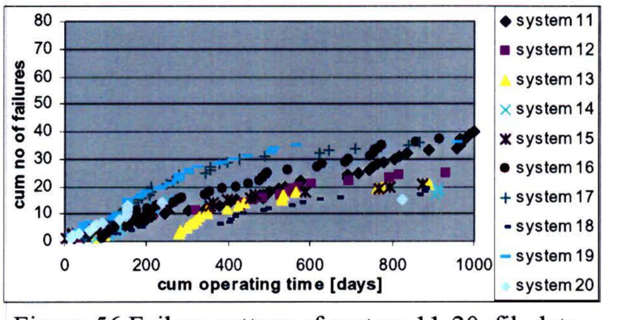


Figure 56 Failure pattern of system 11-20, fil. data

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix J Data set one system

Table 29 Data used for analysis 1 in chapter IV

Warranty start date	Call open date	TTF	TBF	cum no of failures
2001-03-30	2001-04-02	3	3	1
2001-03-30	2001-04-17	18	15	2
2001-03-30	2001-04-17	19	0	3
2001-03-30	2001-06-16	79	60	4
2001-03-30	2001-06-20	83	4	5
2001-03-30	2001-07-13	105	23	6
2001-03-30	2001-07-31	124	18	7
2001-03-30	2001-08-16	140	16	8
2001-03-30	2001-08-30	154	14	9
2001-03-30	2001-08-31	154	1	10
2001-03-30	2001-09-05	160	5	11
2001-03-30	2001-09-13	168	8	12
2001-03-30	2001-09-27	181	14	13
2001-03-30	2001-10-06	190	9	14
2001-03-30	2001-10-24	209	19	15
2001-03-30	2001-10-25	209	1	16
2001-03-30	2001-11-08	223	14	17
2001-03-30	2001-12-10	255	32	18
2001-03-30	2002-03-20	356	100	19
2001-03-30	2002-03-21	357	1	20
2001-03-30	2002-03-27	363	6	21
2001-03-30	2002-05-02	398	36	22
2001-03-30	2002-05-02	399	0	23
2001-03-30	2002-06-28	456	57	24
2001-03-30	2002-07-09	466	11	25
2001-03-30	2002-11-08	588	122	26
2001-03-30	2003-01-16	657	69	27
2001-03-30	2003-05-07	768	111	28
2001-03-30	2003-06-02	794	26	29
2001-03-30	2003-08-20	874	79	30
2001-03-30	2003-10-11	925	52	31

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix K Power law: model intensity $\hat{\mu}_1(T)$ and expected number of failures E[N(T)]

Power law unfiltered

Table 30 Formula based on dataset 2A

365 days	730 days	1000 days
$\hat{\mu}_1(T)=0.0113T^{0.284}$	$\hat{\mu}_1(T)=0.0389T^{0.005}$	$\hat{\mu}_1(T)=0.0673T^{-0.114}$
$E[N(t)]=0.0088t^{1.284}$	$E[N(T)]=0.0387T^{1.005}$	$E[N(T)]=0.0763T^{0.886}$

Table 31 Formula based on dataset 2B

365 days	730 days	1000 days
$\hat{\mu}_1(T)=0.0754T^{-0.037}$	$\hat{\mu}_1(T)=0.1668T^{-0.238}$	$\hat{\mu}_1(T)=0.2195T^{-0.307}$
$E[N(T)]=0.0783T^{0.963}$	$E[N(T)]=0.2189T^{0.762}$	$E[N(T)]=0.3167T^{0.693}$

Table 32 Formula based on dataset 2

365 days	730 days	1000 days
$\hat{\mu}_1(T)=0.0374T^{0.078}$	$\hat{\mu}_1(T)=0.0930T^{-0.141}$	$\hat{\mu}_1(T)=0.1344T^{-0.226}$
$E[N(T)]=0.0347T^{1.078}$	$E[N(T)]=0.1083T^{0.859}$	$E[N(T)]=0.1737T^{0.774}$

Power law filtered

Table 33 Formulas based on dataset 2A

365 days	730 days	1000 days
$\hat{\mu}_1(t)=0.0107t^{0.207}$	$\hat{\mu}_1(t)=0.0268t^{0.002}$	$\hat{\mu}_1(t)=0.0472t^{-0.120}$
$E[N(t)]=0.0089t^{1.207}$	$E[N(t)]=0.0268t^{1.002}$	$E[N(t)]=0.0537t^{0.880}$

Table 34 Formulas based on dataset 2B

365 days	730 days	1000 days
$\hat{\mu}_1(t)=0.0436t^{-0.004}$	$\hat{\mu}_1(t)=0.1030t^{-0.218}$	$\hat{\mu}_1(t)=0.1304t^{-0.276}$
$E[N(t)]=0.0437t^{0.996}$	$E[N(t)]=0.1318t^{0.782}$	$E[N(t)]=0.1801t^{0.724}$

Table 35 Formulas based on dataset 2

365 days	730 days	1000 days
$\hat{\mu}_1(t)=0.0253t^{0.074}$	$\hat{\mu}_1(t)=0.0596t^{-0.129}$	$\hat{\mu}_1(t)=0.0849t^{-0.211}$
$E[N(t)]=0.0236t^{1.074}$	$E[N(t)]=0.0685t^{0.871}$	$E[N(t)]=0.1076t^{0.789}$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix L Graphs of Power law model intensity $\hat{\mu}_1(T)$ and expected number of failures E[N(T)]

Expected number of failures E[N(T)]:

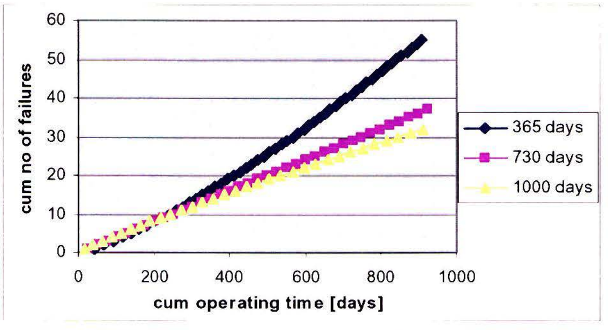


Figure 57 E[N(T)] of dataset 2A unfiltered power law

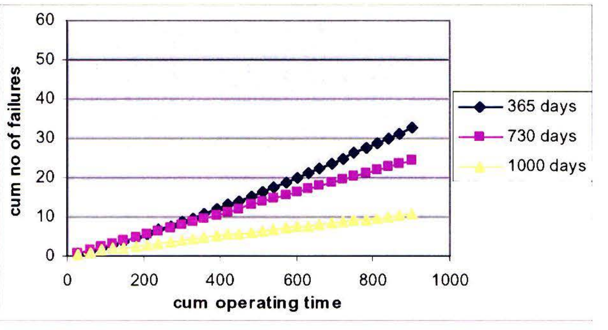


Figure 58 E[N(T)] of dataset 2A filtered power law

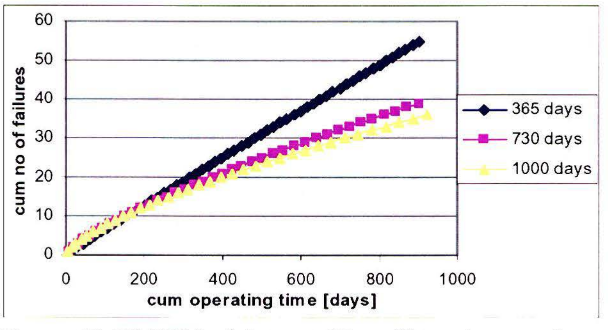


Figure 59 E[N(T)] of dataset 2B unfiltered power law

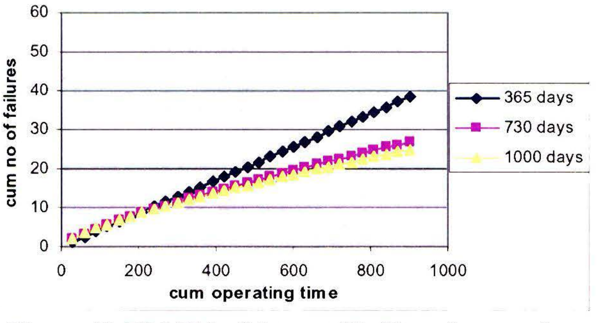


Figure 60 E[N(T)] of dataset 2B filtered power law

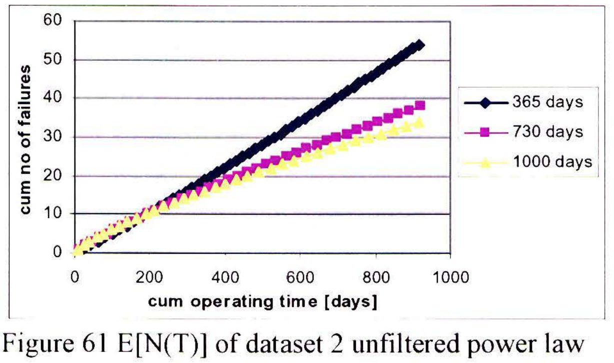


Figure 61 E[N(T)] of dataset 2 unfiltered power law

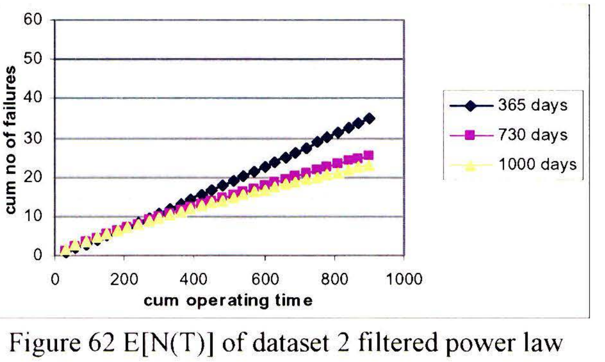


Figure 62 E[N(T)] of dataset 2 filtered power law

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Intensity functions:

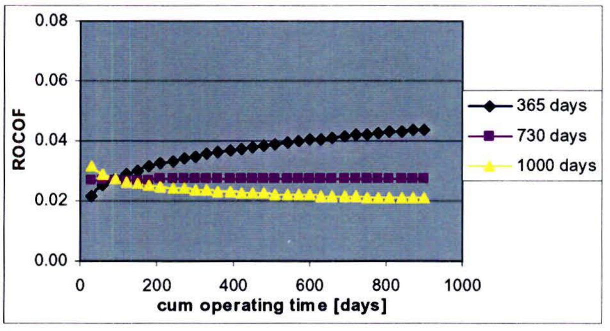


Figure 63 ROCOF of dataset 2A unfiltered power law

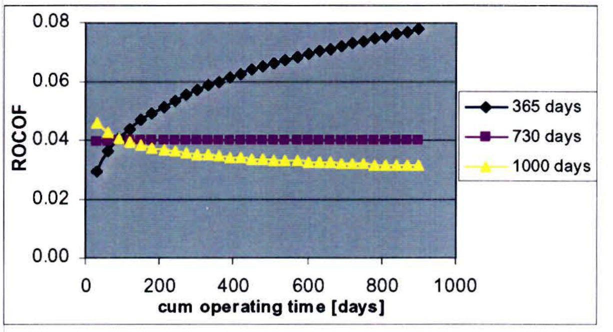


Figure 64 ROCOF of dataset 2A filtered power law

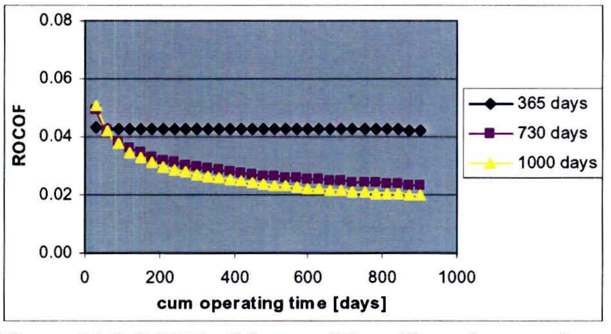


Figure 65 ROCOF of dataset 2B unfiltered power law

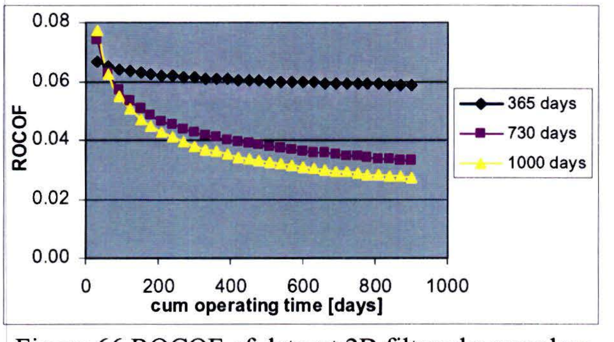


Figure 66 ROCOF of dataset 2B filtered power law

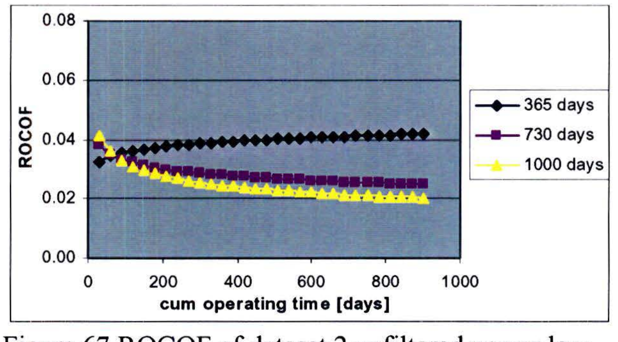


Figure 67 ROCOF of dataset 2 unfiltered power law

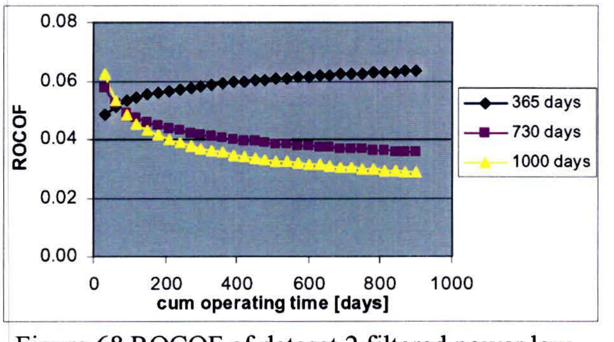


Figure 68 ROCOF of dataset 2 filtered power law

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix M Chi-squared goodness-of-fit

Table 36 unfiltered data Power law

Chi-squared goodness-of-fit based on formula for dataset 2A

Sys	χ^2 365 days	**Test statistic**	Result	χ^2 730 days	**Test statistic**	Result	χ^2 1000 days	**Test statistic**	Result
1	28.07	5.99	Reject	13.07	5.99	Reject	10.50	5.99	Reject
2	26.46	5.99	Reject	11.79	5.99	Reject	9.33	5.99	Reject
3	25.00	3.84	Reject	12.00	3.84	Reject	9.38	3.84	Reject
4	27.72	3.84	Reject	11.89	3.84	Reject	8.82	3.84	Reject
5	14.38	5.99	Reject	7.75	5.99	Reject	9.15	5.99	Reject
6	10.31	9.49	Reject	8.05	9.49	Accept	11.92	9.49	Reject
7	39.63	12.59	Reject	96.49	12.59	Reject	124.17	12.59	Reject
8	18.53	9.49	Reject	23.54	9.49	Reject	25.60	9.49	Reject
9	25.46	3.84	Reject	12.00	3.84	Reject	10.63	3.84	Reject
10	25.10	5.99	Reject	16.84	5.99	Reject	15.00	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2B

11	3.57	5.99	Accept	3.19	5.99	Accept	6.64	5.99	Reject
12	15.86	5.99	Reject	5.60	5.99	Accept	3.34	5.99	Accept
13	15.71	5.99	Reject	9.20	5.99	Reject	10.62	5.99	Reject
14	29.48	3.84	Reject	17.80	3.84	Reject	18.73	3.84	Reject
15	16.31	5.99	Reject	5.73	5.99	Accept	2.91	5.99	Accept
16	6.18	7.81	Accept	4.17	7.81	Accept	7.85	7.81	Reject
17	20.04	7.81	Reject	17.38	7.81	Reject	20.22	7.81	Reject
18	15.71	7.81	Reject	12.72	7.81	Reject	10.88	7.81	Reject
19	28.70	9.49	Reject	28.86	9.49	Reject	33.11	9.49	Reject
20	36.05	5.99	Reject	24.95	5.99	Reject	18.21	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2

1	26.90	5.99	Reject	13.47	5.99	Reject	12.08	5.99	Reject
2	25.35	5.99	Reject	12.30	5.99	Reject	10.91	5.99	Reject
3	19.49	3.84	Reject	8.40	3.84	Reject	6.49	3.84	Reject
4	26.60	3.84	Reject	12.92	3.84	Reject	10.24	3.84	Reject
5	13.96	5.99	Reject	11.97	5.99	Reject	10.10	5.99	Reject
6	3.45	9.49	Accept	5.59	9.49	Accept	9.84	9.49	Reject
7	37.74	12.59	Reject	94.11	12.59	Reject	126.40	12.59	Reject
8	15.30	9.49	Reject	23.39	9.49	Reject	27.12	9.49	Reject
9	23.55	3.84	Reject	11.64	3.84	Reject	10.59	3.84	Reject
10	16.69	5.99	Reject	9.26	5.99	Reject	8.16	5.99	Reject
11	4.02	5.99	Accept	3.21	5.99	Accept	5.18	5.99	Accept
12	18.32	5.99	Reject	7.02	5.99	Reject	5.02	5.99	Accept
13	14.32	5.99	Reject	7.71	5.99	Reject	9.04	5.99	Reject
14	31.66	3.84	Reject	18.35	3.84	Reject	18.24	3.84	Reject
15	18.27	5.99	Reject	6.64	5.99	Reject	4.01	5.99	Accept
16	7.38	7.81	Accept	3.77	7.81	Accept	7.20	7.81	Accept
17	25.79	7.81	Reject	19.48	7.81	Reject	23.92	7.81	Reject
18	14.66	7.81	Reject	11.62	7.81	Reject	10.33	7.81	Reject
19	35.46	9.49	Reject	39.01	9.49	Reject	37.34	9.49	Reject
20	47.41	5.99	Reject	32.73	5.99	Reject	25.15	5.99	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Table 37 filtered data power law

Chi-squared goodness-of-fit based on formula for dataset 2A

Sys	χ^2 365 days	Test statistic	Result	χ^2 730 days	Test statistic	Result	χ^2 1000 days	Test statistic	Result
1	25.94	3.84	Reject	17.01	3.84	Reject	4.83	3.84	Reject
2	18.04	3.84	Reject	10.52	3.84	Reject	3.83	3.84	Accept
3	22.39	3.84	Reject	13.94	3.84	Reject	17.00	3.84	Reject
4	25.55	3.84	Reject	16.83	3.84	Reject	4.97	3.84	Reject
5	9.83	3.84	Reject	4.38	3.84	Reject	12.50	3.84	Reject
6	15.80	9.49	Reject	20.92	9.49	Reject	103.00	9.49	Reject
7	65.55	12.59	Reject	109.80	12.59	Reject	380.33	12.59	Reject
8	13.55	9.49	Reject	19.92	9.49	Reject	114.00	9.49	Reject
9	16.83	3.84	Reject	11.51	3.84	Reject	15.08	3.84	Reject
10	14.57	5.99	Reject	10.20	5.99	Reject	43.33	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2B

11	1.96	5.99	Accept	5.93	5.99	Accept	8.38	5.99	Reject
12	8.23	5.99	Reject	1.93	5.99	Accept	2.43	5.99	Accept
13	13.98	5.99	Reject	11.35	5.99	Reject	11.20	5.99	Reject
14	23.35	3.84	Reject	8.06	3.84	Reject	16.07	3.84	Reject
15	10.68	5.99	Reject	3.98	5.99	Accept	2.83	5.99	Accept
16	3.52	7.81	Accept	14.90	7.81	Accept	15.99	7.81	Reject
17	18.88	7.81	Reject	23.82	7.81	Reject	23.15	7.81	Reject
18	10.69	7.81	Reject	4.12	7.81	Accept	3.87	7.81	Accept
19	22.29	5.99	Reject	23.87	5.99	Reject	22.19	5.99	Reject
20	29.26	3.84	Reject	18.34	3.84	Reject	16.36	3.84	Reject

Chi-squared goodness-of-fit based on formula for dataset 2

1	27.80	3.84	Reject	18.87	3.84	Reject	15.90	3.84	Reject
2	19.74	3.84	Reject	11.93	3.84	Reject	9.64	3.84	Reject
3	20.97	3.84	Reject	12.48	3.84	Reject	9.80	3.84	Reject
4	27.56	3.84	Reject	18.82	3.84	Reject	15.98	3.84	Reject
5	10.77	3.84	Reject	5.75	3.84	Reject	4.57	3.84	Reject
6	8.74	9.49	Accept	16.12	9.49	Reject	16.70	9.49	Reject
7	58.50	12.59	Reject	96.67	12.59	Reject	137.00	12.59	Reject
8	12.87	9.49	Reject	16.30	9.49	Reject	26.97	9.49	Reject
9	19.30	3.84	Reject	11.93	3.84	Reject	10.59	3.84	Reject
10	11.22	5.99	Reject	6.77	5.99	Reject	7.12	5.99	Reject
11	2.15	5.99	Accept	6.13	5.99	Reject	8.42	5.99	Reject
12	9.05	5.99	Reject	5.05	5.99	Accept	3.80	5.99	Accept
13	11.79	5.99	Reject	9.17	5.99	Accept	9.37	5.99	Reject
14	23.51	3.84	Reject	16.66	3.84	Reject	15.24	3.84	Reject
15	10.01	5.99	Reject	4.17	5.99	Accept	2.87	5.99	Accept
16	4.83	7.81	Accept	8.20	7.81	Reject	16.67	7.81	Reject
17	24.13	7.81	Reject	24.10	7.81	Reject	27.58	7.81	Reject
18	8.47	7.81	Reject	4.10	7.81	Accept	4.17	7.81	Accept
19	29.44	5.99	Reject	25.54	5.99	Reject	32.98	5.99	Reject
20	33.31	3.84	Reject	24.37	3.84	Reject	21.40	3.84	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Histograms of the chi-squared test results

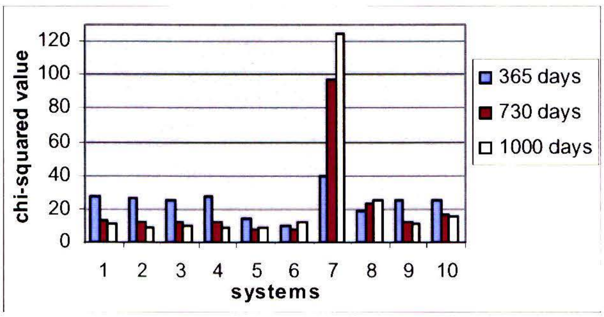


Figure 69 Dataset 2A Chi-squared unfiltered data

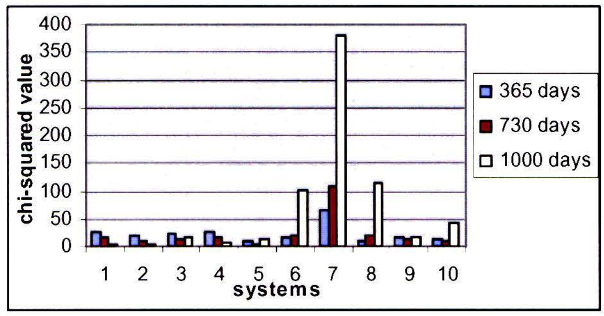


Figure 70 Dataset 2A Chi-squared filtered data

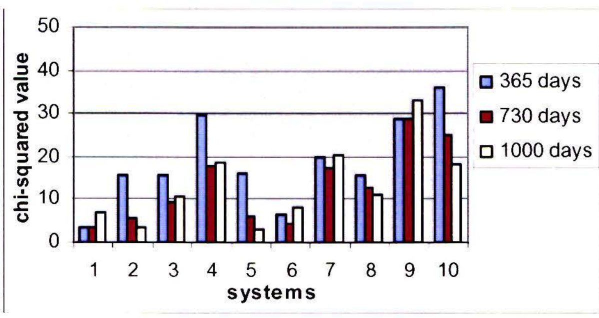


Figure 71 Dataset 2B Chi-squared unfiltered data

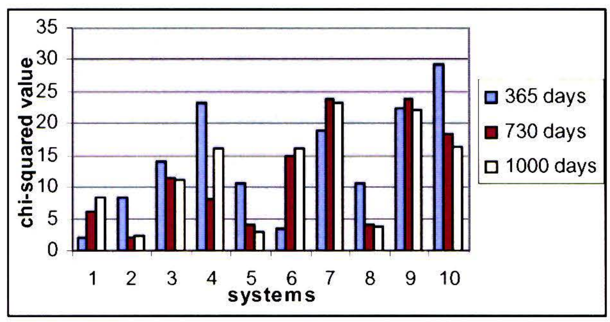


Figure 72 Dataset 2B Chi-squared filtered data

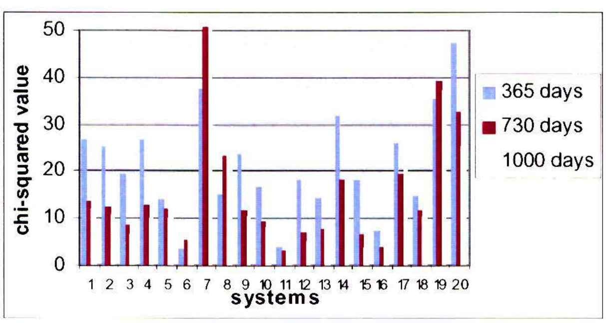


Figure 73 Dataset 2 Chi-squared unfiltered data

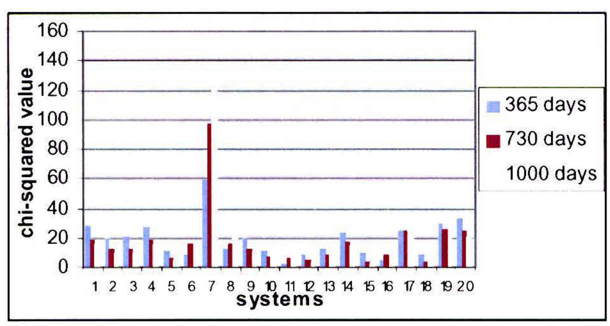


Figure 74 Dataset 2 Chi-squared filtered data

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix N Cramér-von Mises goodness-of-fit Power law

Table 38 unfiltered data
Formula based on dataset 2A

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
1	0.190	0.171	Reject	0.073	0.171	Accept	0.082	0.171	Accept
2	0.179	0.171	Reject	0.064	0.171	Accept	0.079	0.171	Accept
3	2.156	0.172	Reject	1.217	0.172	Reject	0.834	0.172	Accept
4	0.201	0.171	Reject	0.116	0.171	Accept	0.127	0.171	Accept
5	0.267	0.172	Reject	0.201	0.172	Reject	0.312	0.172	Reject
6	0.759	0.172	Reject	0.170	0.172	Accept	0.056	0.172	Accept
7	0.547	0.173	Reject	0.427	0.173	Reject	0.709	0.173	Reject
8	1.014	0.172	Reject	0.357	0.172	Reject	0.210	0.172	Reject
9	0.742	0.172	Reject	0.332	0.172	Reject	0.203	0.172	Reject
10	2.417	0.172	Reject	1.297	0.172	Reject	0.851	0.172	Reject

Formula based on dataset 2B

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
11	0.034	0.172	Accept	0.122	0.171	Accept	0.256	0.171	Reject
12	0.701	0.172	Reject	0.268	0.171	Reject	0.163	0.171	Accept
13	0.374	0.172	Reject	0.289	0.172	Reject	0.354	0.172	Reject
14	2.551	0.172	Reject	1.657	0.171	Reject	1.385	0.171	Reject
15	1.042	0.172	Reject	0.461	0.172	Reject	0.314	0.172	Reject
16	0.276	0.172	Reject	0.151	0.172	Accept	0.227	0.172	Reject
17	1.859	0.172	Reject	0.900	0.173	Reject	0.656	0.173	Reject
18	0.248	0.172	Reject	0.156	0.172	Accept	0.211	0.172	Reject
19	2.702	0.172	Reject	1.373	0.172	Reject	0.992	0.172	Reject
20	2.766	0.172	Reject	1.743	0.172	Reject	1.389	0.172	Reject

Formula based on dataset 2

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
1	0.089	0.171	Accept	0.092	0.171	Accept	0.146	0.171	Accept
2	0.078	0.171	Accept	0.090	0.171	Accept	0.150	0.171	Accept
3	1.464	0.172	Reject	0.750	0.172	Reject	0.511	0.172	Reject
4	0.126	0.171	Accept	0.136	0.171	Accept	0.184	0.171	Reject
5	0.181	0.172	Accept	0.354	0.172	Reject	0.534	0.172	Reject
6	0.291	0.172	Reject	0.047	0.172	Accept	0.079	0.172	Accept
7	0.370	0.173	Reject	0.814	0.173	Reject	1.256	0.173	Reject
8	0.497	0.172	Reject	0.195	0.172	Reject	0.204	0.172	Reject
9	0.430	0.172	Reject	0.179	0.172	Reject	0.128	0.172	Accept
10	1.589	0.172	Reject	0.755	0.172	Reject	0.486	0.172	Reject
11	0.118	0.172	Accept	0.033	0.172	Accept	0.105	0.172	Accept
12	0.988	0.172	Reject	0.461	0.172	Reject	0.290	0.172	Reject
13	0.546	0.172	Reject	0.287	0.172	Reject	0.283	0.172	Reject
14	3.061	0.172	Reject	2.080	0.172	Reject	1.708	0.172	Reject
15	1.414	0.172	Reject	0.722	0.172	Reject	0.491	0.172	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

16	0.490	0.172	Reject	0.160	0.172	Accept	0.146	0.172	Accept
17	2.472	0.172	Reject	1.331	0.172	Reject	0.949	0.172	Reject
18	0.412	0.172	Accept	0.162	0.172	Accept	0.152	0.172	Accept
19	3.504	0.172	Reject	1.989	0.172	Reject	1.445	0.172	Reject
20	3.296	0.172	Reject	2.247	0.172	Reject	1.806	0.172	Reject

Table 39 filtered data
Formula based on dataset 2A

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
1	0.480	0.171	Reject	0.346	0.171	Reject	0.269	0.171	Reject
2	0.038	0.171	Accept	0.050	0.171	Accept	0.091	0.171	Accept
3	2.084	0.172	Reject	1.548	0.172	Reject	1.207	0.172	Reject
4	0.323	0.171	Reject	0.378	0.171	Reject	0.420	0.171	Reject
5	0.170	0.172	Reject	0.143	0.172	Reject	0.195	0.172	Reject
6	0.562	0.172	Reject	0.200	0.172	Reject	0.074	0.172	Accept
7	0.393	0.173	Reject	0.416	0.173	Reject	0.683	0.173	Reject
8	0.663	0.172	Reject	0.291	0.172	Reject	0.162	0.172	Accept
9	0.898	0.172	Reject	0.599	0.172	Reject	0.432	0.172	Reject
10	1.438	0.172	Reject	0.865	0.172	Reject	0.554	0.172	Reject

Formula based on dataset 2B

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
11	0.039	0.172	Accept	0.259	0.171	Reject	0.383	0.171	Reject
12	0.320	0.172	Reject	0.095	0.171	Accept	0.069	0.171	Accept
13	0.271	0.172	Reject	0.311	0.172	Reject	0.375	0.172	Reject
14	2.288	0.172	Reject	1.567	0.171	Reject	1.378	0.171	Reject
15	0.623	0.172	Reject	0.245	0.172	Reject	0.167	0.172	Accept
16	0.191	0.172	Reject	0.110	0.172	Accept	0.164	0.172	Accept
17	1.754	0.172	Reject	0.880	0.173	Reject	0.693	0.173	Reject
18	0.069	0.172	Accept	0.147	0.172	Accept	0.204	0.172	Reject
19	2.580	0.172	Reject	1.419	0.172	Reject	1.140	0.172	Reject
20	2.276	0.172	Reject	1.618	0.172	Reject	1.421	0.172	Reject

Formula based on dataset 2

Sys	$C^2(n)$ 365 days	Test statistic	Result	$C^2(n)$ 730 days	Test statistic	Result	$C^2(n)$ 1000 days	Test statistic	Result
1	0.393	0.171	Reject	0.263	0.171	Reject	0.217	0.171	Reject
2	0.039	0.171	Accept	0.095	0.171	Accept	0.144	0.171	Accept
3	1.742	0.172	Reject	1.181	0.172	Reject	0.952	0.172	Reject
4	0.357	0.171	Reject	0.423	0.171	Reject	0.457	0.171	Reject
5	0.139	0.172	Reject	0.201	0.172	Reject	0.297	0.172	Reject
6	0.310	0.172	Reject	0.069	0.172	Accept	0.650	0.172	Reject
7	0.357	0.173	Reject	0.713	0.173	Reject	0.713	0.173	Reject
8	0.404	0.172	Reject	0.156	0.172	Accept	0.139	0.172	Accept
9	0.702	0.172	Reject	0.421	0.172	Reject	0.324	0.172	Reject
10	1.062	0.172	Reject	0.532	0.172	Reject	0.360	0.172	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

11	0.025	0.172	Accept	0.128	0.172	Accept	0.245	0.172	Reject
12	0.432	0.172	Reject	0.169	0.172	Accept	0.100	0.172	Accept
13	0.313	0.172	Reject	0.261	0.172	Reject	0.304	0.172	Reject
14	2.536	0.172	Reject	1.868	0.172	Reject	1.592	0.172	Reject
15	0.776	0.172	Reject	0.389	0.172	Reject	0.256	0.172	Reject
16	0.292	0.172	Reject	0.098	0.172	Accept	0.106	0.172	Accept
17	2.098	0.172	Reject	1.219	0.172	Reject	0.907	0.172	Reject
18	0.078	0.172	Accept	0.092	0.172	Accept	0.141	0.172	Accept
19	3.012	0.172	Reject	1.887	0.172	Reject	1.458	0.172	Reject
20	2,478	0.172	Reject	1.908	0.172	Reject	1.644	0.172	Reject

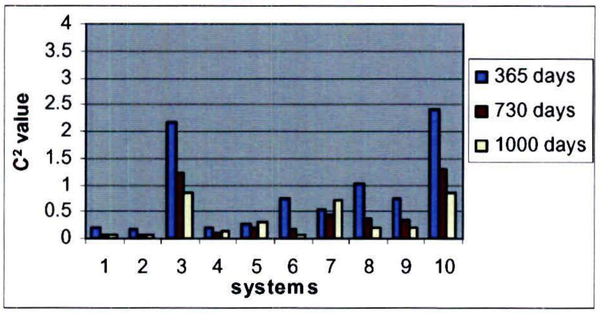


Figure 75 Dataset 2A Cramér-von Mises unfiltered data

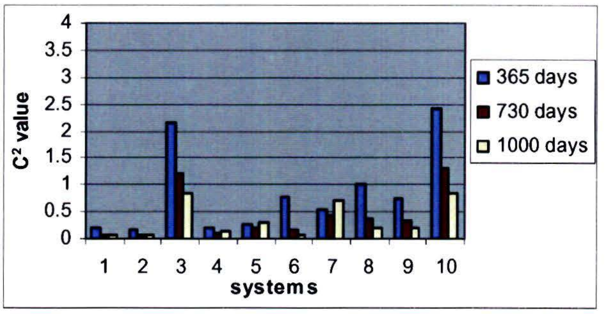


Figure 76 Dataset 2A Cramér-von Mises filtered data

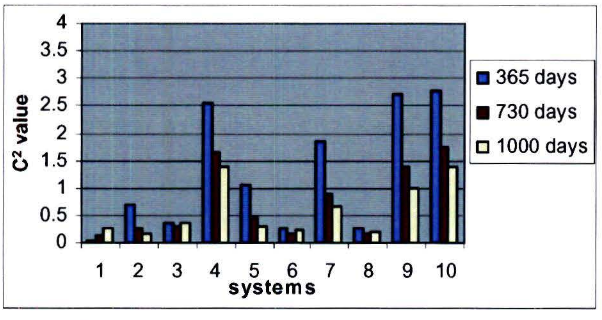


Figure 77 Dataset 2B Cramér-von Mises unfiltered data

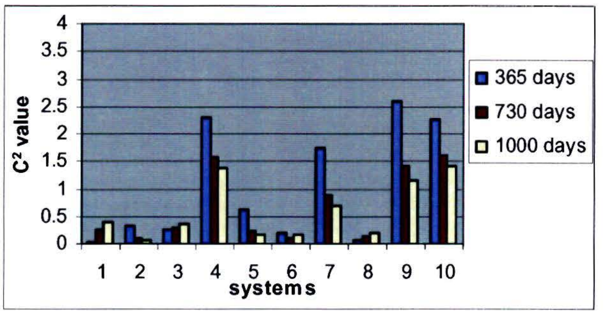


Figure 78 Dataset 2B Cramér-von Mises filtered data

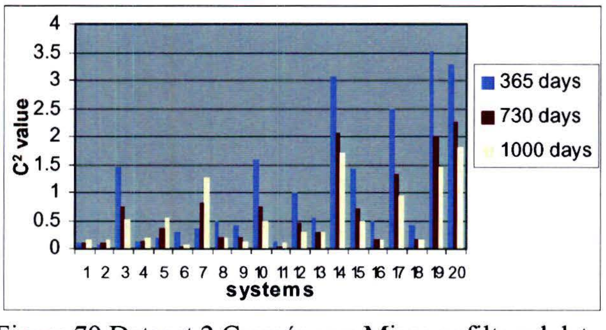


Figure 79 Dataset 2 Cramér-von Mises unfiltered data

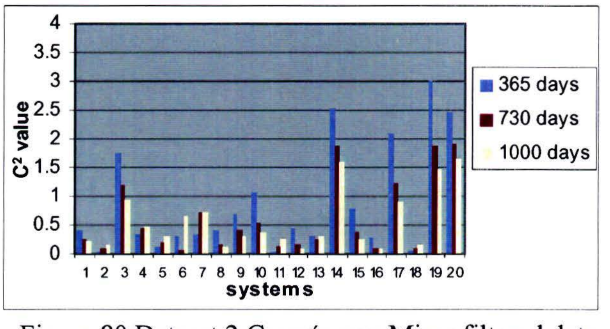


Figure 80 Dataset 2 Cramér-von Mises filtered data

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix O MTBF$_c$ power law

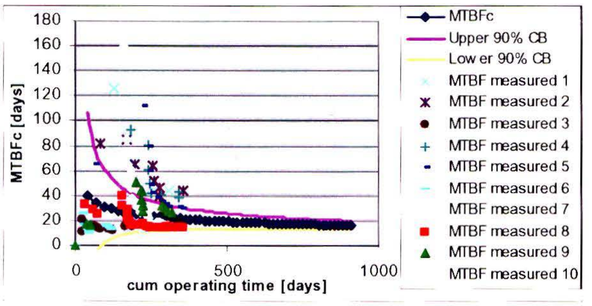


Figure 81 Power law dataset 2A unfiltered 365 days

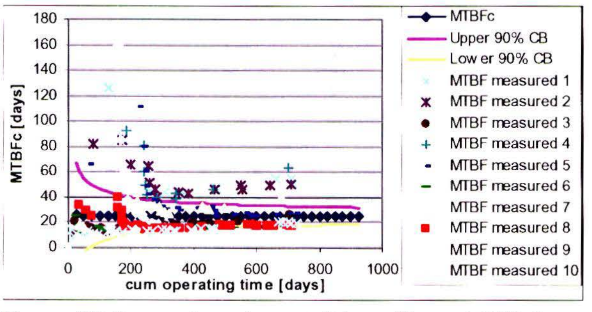


Figure 82 Power law dataset 2A unfiltered 730 days

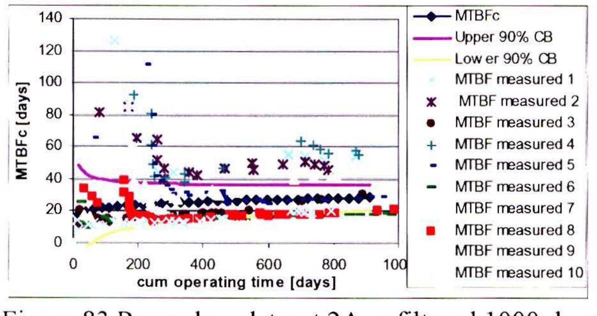


Figure 83 Power law dataset 2A unfiltered 1000 days

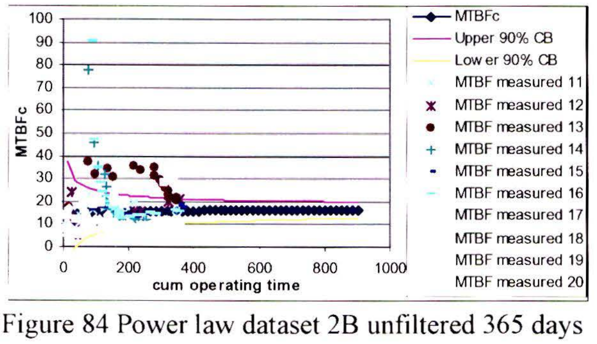


Figure 84 Power law dataset 2B unfiltered 365 days

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

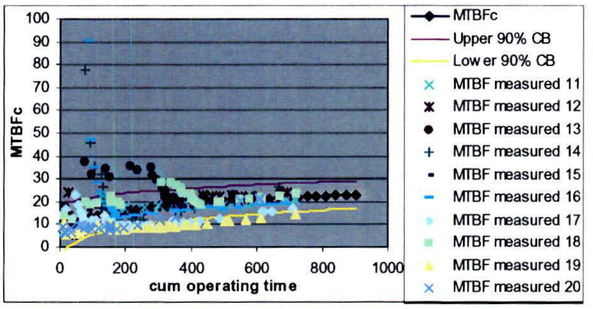


Figure 85 Power law dataset 2B unfiltered 730 days

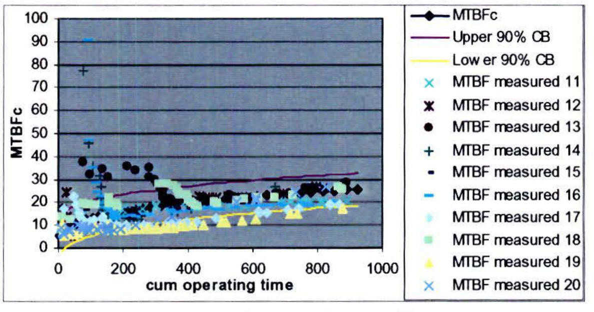


Figure 86 Power law dataset 2B unfiltered 1000 days

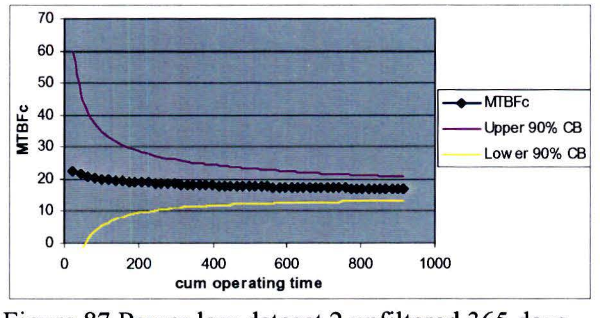


Figure 87 Power law dataset 2 unfiltered 365 days

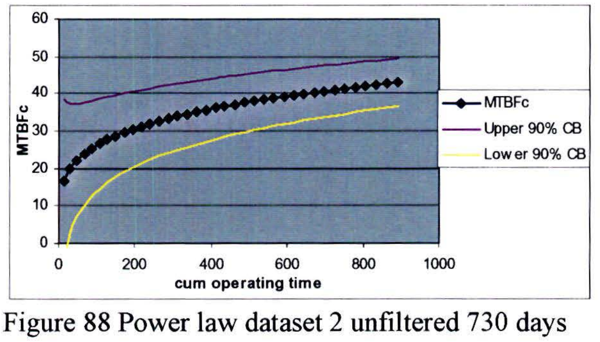


Figure 88 Power law dataset 2 unfiltered 730 days

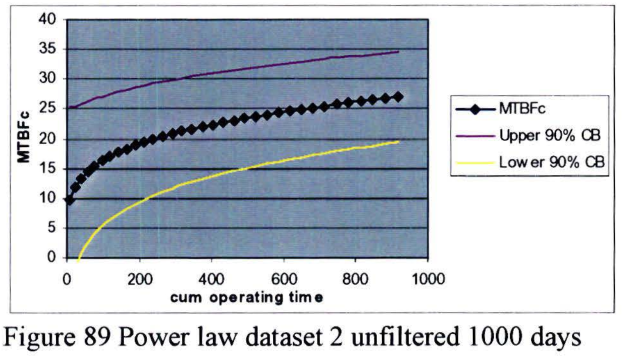


Figure 89 Power law dataset 2 unfiltered 1000 days

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix P Exponential law: model intensity $\hat{\mu}_2(T)$ and expected number of failures E[N(T)] filtered

Exponential law unfiltered

Table 40 Dataset 2A

365 days	$\hat{\mu}_2(T) = e^{-3.7441+0.0038T}$
	$E_2[N(T)] = \frac{e^{-3.7441}}{0.0038}(e^{0.0038T_b} - e^{0.0038T_a})$
730 days	$\hat{\mu}_2(T) = e^{-3.1320-0.00016T}$
	$E_2[N(T)] = \frac{e^{-3.1320}}{-0.00016}(e^{-0.00016T_b} - e^{-0.00016T_a})$
1000 days	$\hat{\mu}_2(T) = e^{-3.1074-0.00030T}$
	$E_2[N(T)] = \frac{e^{-3.1074}}{-0.00030}(e^{-0.00030T_b} - e^{-0.00030T_a})$

Table 41 Dataset 2B

365 days	$\hat{\mu}_2(T) = e^{-2.60734-0.00071T}$
	$E_2[N(T)] = \frac{e^{-2.60734}}{-0.00071}(e^{-0.00071T_b} - e^{-0.00071T_a})$
730 days	$\hat{\mu}_2(T) = e^{-2.47218-0.0018T}$
	$E_2[N(T)] = \frac{e^{-2.47218}}{-0.0018}(e^{-0.0018T_b} - e^{-0.0018T_a})$
1000 days	$\hat{\mu}_2(t) = e^{-2.53035-0.0016T}$
	$E_2[N(T)] = \frac{e^{-2.53035}}{-0.0016}(e^{-0.0016T_b} - e^{-0.0016T_a})$

Table 42 Dataset 2

365 days	$\hat{\mu}_2(T) = e^{-2.8595+7.654E-16T}$
	$E_2[N(T)] = \frac{e^{-2.8535}}{7.654E-16}(e^{7.654E-16T_b} - e^{7.654E-16T_a})$
730 days	$\hat{\mu}_2(t) = e^{-2.782-0.00098T}$
	$E_2[N(T)] = \frac{e^{-2.782}}{-0.00098}(e^{-0.00098T_b} - e^{-0.00098T_a})$
1000 days	$\hat{\mu}_2(t) = e^{-2.8054-0.00095T}$
	$E_2[N(T)] = \frac{e^{-2.8054}}{-0.00095}(e^{-0.00095T_b} - e^{-0.00095T_a})$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Exponential Law filtered

Table 43 Formulas based on dataset 2A

365 days	$\hat{\mu}_2(T) = e^{-4.1154+0.0041T}$
	$E_2[N(T)] = \frac{e^{-4.1154}}{0.0041}(e^{0.0041T_b} - e^{0.0041T_a})$
730 days	$\hat{\mu}_2(T) = e^{-3.6886+0.0001T}$
	$E_2[N(T)] = \frac{e^{-3.6886}}{0.0001}(e^{0.0001T_b} - e^{0.0001T_a})$
1000 days	$\hat{\mu}_2(T) = e^{-3.5136-0.000021T}$
	$E_2[N(T)] = \frac{e^{-3.5136}}{0.000021}(e^{0.000021T_b} - e^{0.000021T_a})$

Table 44 Formulas based on dataset 2B

365 days	$\hat{\mu}_2(T) = e^{-3.0933+0.0004T}$
	$E_2[N(T)] = \frac{e^{-3.0933}}{0.0004}(e^{0.0004T_b} - e^{0.0004T_a})$
730 days	$\hat{\mu}_2(T) = e^{-2.9704-0.0009T}$
	$E_2[N(T)] = \frac{e^{-2.9704}}{-0.0009}(e^{-0.0009T_b} - e^{-0.0009T_a})$
1000 days	$\hat{\mu}_2(T) = e^{-2.9865-0.0013T}$
	$E_2[N(T)] = \frac{e^{-2.9865}}{-0.0013}(e^{-0.0013T_b} - e^{-0.0013T_a})$

Table 45 Formulas based on dataset 2

365 days	$\hat{\mu}_2(T) = e^{-4.2088+0.0054T}$
	$E_2[N(T)] = \frac{e^{-4.2088}}{0.0054}(e^{0.0054T_b} - e^{0.0054T_a})$
730 days	$\hat{\mu}_2(T) = e^{-3.2998+0.000012T}$
	$E_2[N(T)] = \frac{e^{-3.2998}}{-0.000012}(e^{-0.000012T_b} - e^{-0.000012T_a})$
1000 days	$\hat{\mu}_2(T) = e^{-3.2247-0.00071T}$
	$E_2[N(T)] = \frac{e^{-3.2247}}{-0.00071}(e^{-0.00071T_b} - e^{-0.00071T_a})$

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix Q Graphs of Exponential law model intensity $\hat{\mu}_2(T)$ and expected number of failures E[N(T)]

Expected number of failures E[N(T)]:

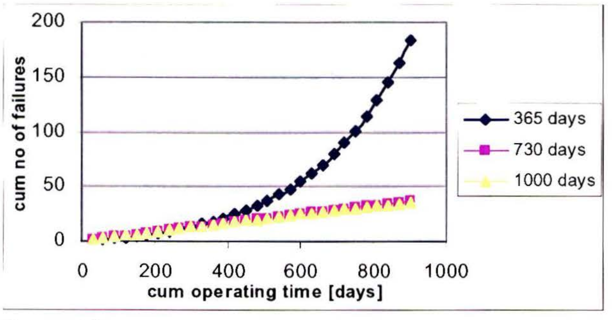


Figure 90
E[N(T)] of dataset 2A unfiltered exponential law

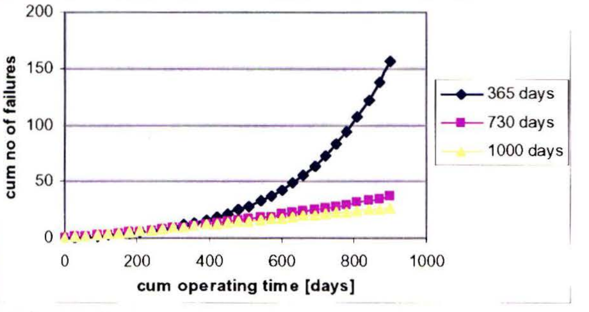


Figure 91
E[N(T)] of dataset 2A filtered exponential law

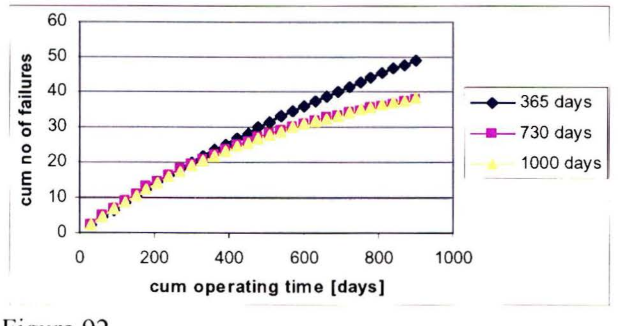


Figure 92
E[N(T)] of dataset 2B unfiltered exponential law

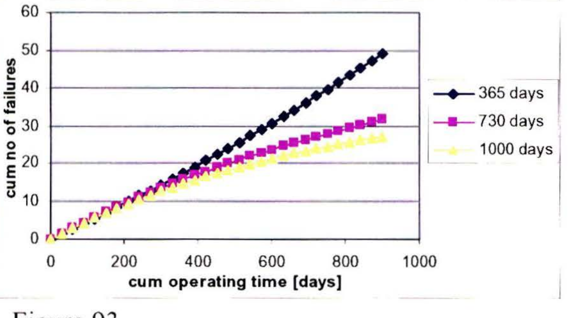


Figure 93
E[N(T)] of dataset 2B filtered exponential law

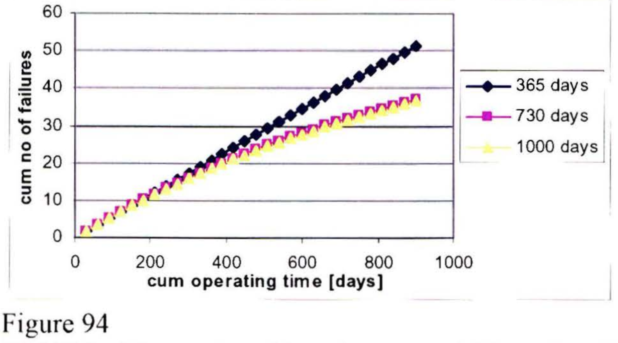


Figure 94
E[N(T)] of dataset 2 unfiltered exponential law

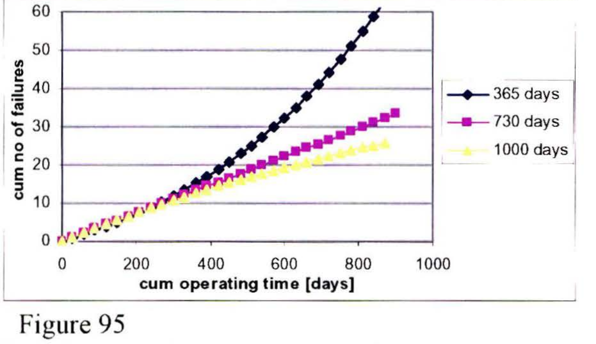


Figure 95
E[N(T)] of dataset 2 filtered exponential law

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Intensity functions:

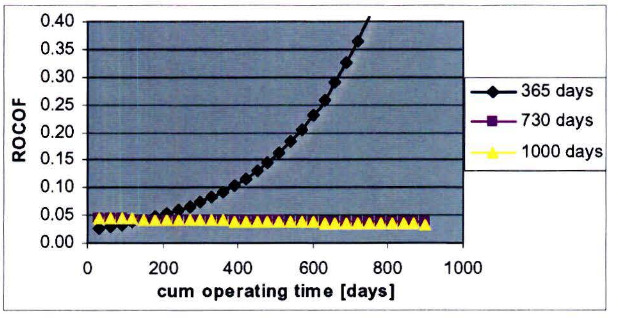


Figure 96
ROCOF of dataset 2A unfiltered exponential law

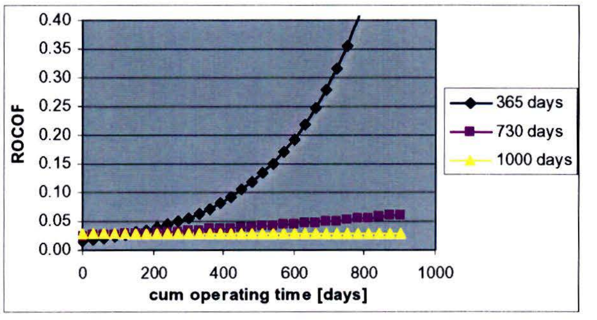


Figure 97
ROCOF of dataset 2A filtered exponential law

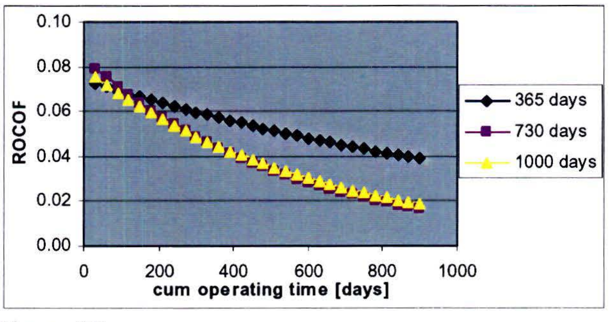


Figure 98
ROCOF of dataset 2B unfiltered exponential law

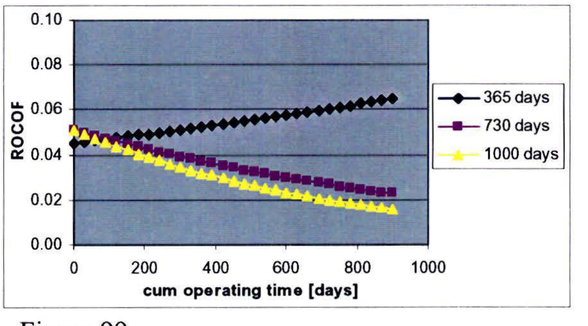


Figure 99
ROCOF of dataset 2B filtered exponential law

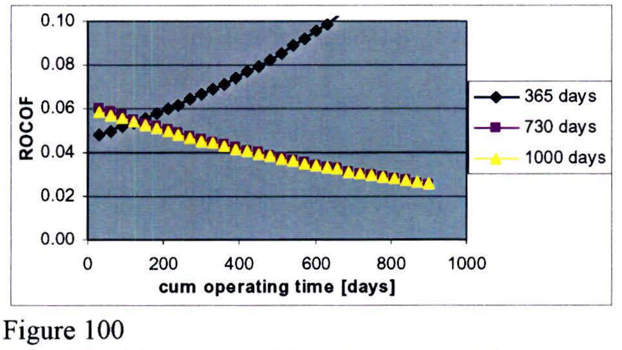


Figure 100
ROCOF of dataset 2 unfiltered exponential law

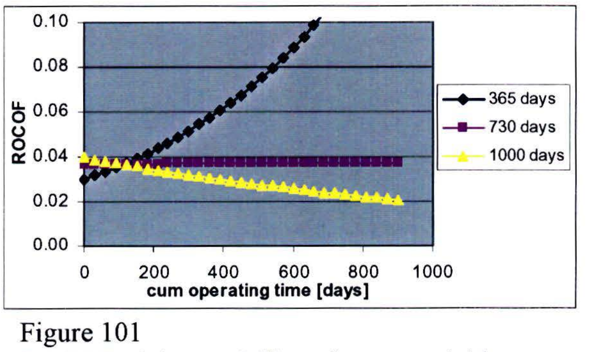


Figure 101
ROCOF of dataset 2 filtered exponential law

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix R Chi-squared goodness-of-fit Exponential law

Table 46 unfiltered data

Chi-squared goodness-of-fit based on formula for dataset 2A

Sys	χ^2 365 days	**Test statistic**	Result	χ^2 730 days	**Test statistic**	Result	χ^2 1000 days	**Test statistic**	Result
1	156.21	5.99	Reject	14.00	5.99	Reject	12.17	5.99	Reject
2	154.26	5.99	Reject	12.71	5.99	Reject	11.00	5.99	Reject
3	154.48	3.84	Reject	10.84	3.84	Reject	9.53	3.84	Reject
4	156.05	3.84	Reject	12.62	3.84	Reject	11.30	3.84	Reject
5	134.17	5.99	Reject	8.23	5.99	Reject	7.06	5.99	Reject
6	129.90	9.49	Reject	6.60	9.49	Accept	7.57	9.49	Accept
7	117.66	12.59	Reject	97.42	12.59	Reject	92.42	12.59	Reject
8	144.55	9.49	Reject	20.49	9.49	Reject	21.15	9.49	Reject
9	153.03	3.84	Reject	12.84	3.84	Reject	11.21	3.84	Reject
10	154.22	5.99	Reject	14.34	5.99	Reject	13.00	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2B

11	2.39	5.99	Accept	7.29	5.99	Reject	7.22	5.99	Reject
12	10.79	5.99	Reject	3.78	5.99	Accept	4.68	5.99	Accept
13	10.86	5.99	Reject	7.12	5.99	Reject	5.68	5.99	Accept
14	22.27	3.84	Reject	12.12	3.84	Reject	13.58	3.84	Reject
15	11.05	5.99	Reject	2.81	5.99	Accept	3.02	5.99	Accept
16	2.11	7.81	Accept	4.35	7.81	Accept	5.49	7.81	Accept
17	13.46	7.81	Reject	9.27	7.81	Reject	10.87	7.81	Reject
18	13.11	7.81	Reject	13.42	7.81	Reject	9.62	7.81	Reject
19	20.60	9.49	Reject	17.55	9.49	Reject	17.55	9.49	Reject
20	30.36	5.99	Reject	17.81	5.99	Reject	20.73	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2

1	26.04	5.99	Reject	13.44	5.99	Reject	13.87	5.99	Reject
2	24.49	5.99	Reject	12.44	5.99	Reject	12.87	5.99	Reject
3	17.49	3.84	Reject	6.06	3.84	Reject	6.54	3.84	Reject
4	25.51	3.84	Reject	13.43	3.84	Reject	12.77	3.84	Reject
5	14.45	5.99	Reject	12.35	5.99	Reject	8.35	5.99	Reject
6	2.34	9.49	Accept	5.75	9.49	Accept	4.94	9.49	Accept
7	41.27	12.59	Reject	94.20	12.59	Reject	91.32	12.59	Reject
8	15.06	9.49	Reject	18.63	9.49	Reject	18.14	9.49	Reject
9	22.63	3.84	Reject	10.64	3.84	Reject	11.48	3.84	Reject
10	13.88	5.99	Reject	5.42	5.99	Accept	7.32	5.99	Reject
11	3.34	5.99	Accept	4.20	5.99	Accept	2.93	5.99	Accept
12	16.19	5.99	Reject	5.64	5.99	Accept	4.99	5.99	Accept
13	13.71	5.99	Reject	5.94	5.99	Accept	6.80	5.99	Reject
14	29.27	3.84	Reject	16.75	3.84	Reject	16.19	3.84	Reject
15	15.94	5.99	Reject	4.55	5.99	Accept	4.99	5.99	Accept
16	6.38	7.81	Accept	4.48	7.81	Accept	3.00	7.81	Accept
17	22.36	7.81	Reject	16.26	7.81	Reject	17.51	7.81	Reject
18	13.56	7.81	Reject	8.08	7.81	Reject	12.13	7.81	Reject
19	30.89	9.49	Reject	27.48	9.49	Reject	30.01	9.49	Reject
20	41.12	5.99	Reject	27.48	5.99	Reject	27.70	5.99	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Table 47 filtered data

Chi-squared goodness-of-fit based on formula for dataset 2A

Sys	χ^2 365 days	Test statistic	Result	χ^2 730 days	Test statistic	Result	χ^2 1000 days	Test statistic	Result
1	150.03	3.84	Reject	28.93	3.84	Reject	19.98	3.84	Reject
2	139.99	3.84	Reject	20.63	3.84	Reject	12.87	3.84	Reject
3	150.51	3.84	Reject	25.31	3.84	Reject	16.67	3.84	Reject
4	149.11	3.84	Reject	28.48	3.84	Reject	19.72	3.84	Reject
5	128.03	3.84	Reject	12.13	3.84	Reject	5.78	3.84	Reject
6	119.30	9.49	Reject	17.40	9.49	Reject	19.25	9.49	Reject
7	99.43	12.59	Reject	58.17	12.59	Reject	80.13	12.59	Reject
8	121.36	9.49	Reject	15.45	9.49	Reject	13.83	9.49	Reject
9	153.12	3.84	Reject	21.59	3.84	Reject	12.86	3.84	Reject
10	135.21	5.99	Reject	17.31	5.99	Reject	10.17	5.99	Reject

Chi-squared goodness-of-fit based on formula for dataset 2B

11	6.83	5.99	Reject	3.48	5.99	Accept	11.08	5.99	Reject
12	17.44	5.99	Reject	4.25	5.99	Accept	2.99	5.99	Accept
13	20.51	5.99	Reject	10.02	5.99	Reject	7.52	5.99	Reject
14	33.17	3.84	Reject	17.71	3.84	Reject	12.74	3.84	Reject
15	19.41	5.99	Reject	5.25	5.99	Accept	2.72	5.99	Accept
16	9.32	7.81	Reject	5.40	7.81	Accept	7.90	7.81	Reject
17	24.20	7.81	Reject	14.02	7.81	Reject	13.15	7.81	Reject
18	19.20	7.81	Reject	6.57	7.81	Accept	5.54	7.81	Accept
19	27.01	5.99	Reject	14.83	5.99	Reject	14.06	5.99	Reject
20	40.22	3.84	Reject	20.95	3.84	Reject	18.34	3.84	Reject

Chi-squared goodness-of-fit based on formula for dataset 2

1	60.73	3.84	Reject	25.73	3.84	Reject	18.78	3.84	Reject
2	51.45	3.84	Reject	18.00	3.84	Reject	12.43	3.84	Reject
3	53.41	3.84	Reject	16.75	3.84	Reject	10.18	3.84	Reject
4	60.24	3.84	Reject	25.62	3.84	Reject	18.88	3.84	Reject
5	40.38	3.84	Reject	9.45	3.84	Reject	5.68	3.84	Reject
6	28.81	9.49	Reject	9.57	9.49	Reject	12.87	9.49	Reject
7	32.14	12.59	Reject	71.40	12.59	Reject	102.40	12.59	Reject
8	32.64	9.49	Reject	11.17	9.49	Reject	16.05	9.49	Reject
9	50.47	3.84	Reject	18.39	3.84	Reject	11.66	3.84	Reject
10	39.70	5.99	Reject	9.02	5.99	Reject	4.24	5.99	Accept
11	22.42	5.99	Reject	1.55	5.99	Accept	6.22	5.99	Reject
12	38.22	5.99	Reject	5.88	5.99	Accept	3.32	5.99	Accept
13	39.10	5.99	Reject	11.86	5.99	Reject	7.77	5.99	Reject
14	54.96	3.84	Reject	22.01	3.84	Reject	15.03	3.84	Reject
15	39.93	5.99	Reject	7.93	5.99	Reject	2.70	5.99	Accept
16	26.15	7.81	Reject	7.53	7.81	Accept	9.45	7.81	Reject
17	50.22	7.81	Reject	21.60	7.81	Reject	19.63	7.81	Reject
18	37.30	7.81	Reject	6.27	7.81	Accept	4.55	7.81	Accept
19	52.97	5.99	Reject	23.97	5.99	Reject	21.63	5.99	Reject
20	66.23	3.84	Reject	25.31	3.84	Reject	20.35	3.84	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

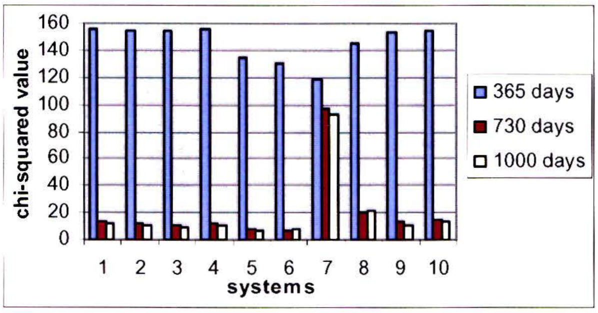


Figure 102 Dataset 2A Chi-squared unfiltered data

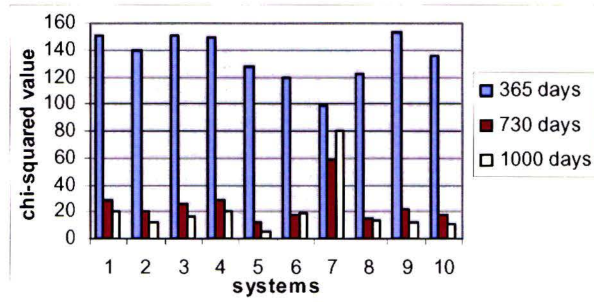


Figure 103 Dataset 2A Chi-squared filtered data

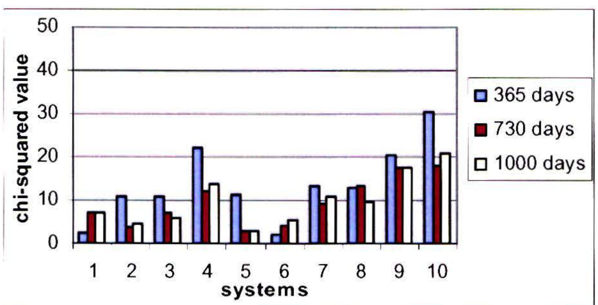


Figure 104 Dataset 2B Chi-squared unfiltered data

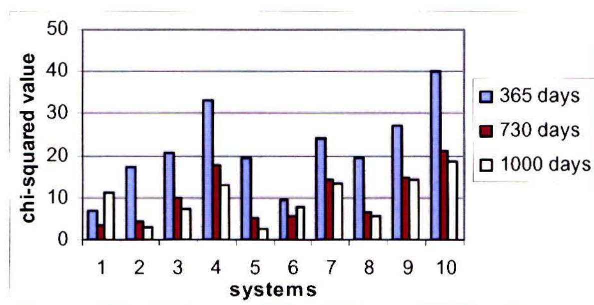


Figure 105 Dataset 2B Chi-squared filtered data

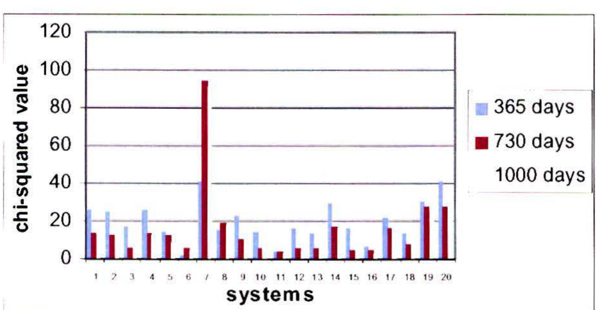


Figure 106 Dataset 2 Chi-squared unfiltered data

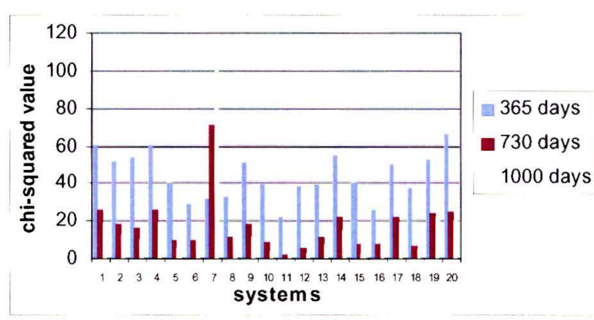


Figure 107 Dataset 2 Chi-squared filtered data

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

Appendix S Actual and expected number of failures, exponential law

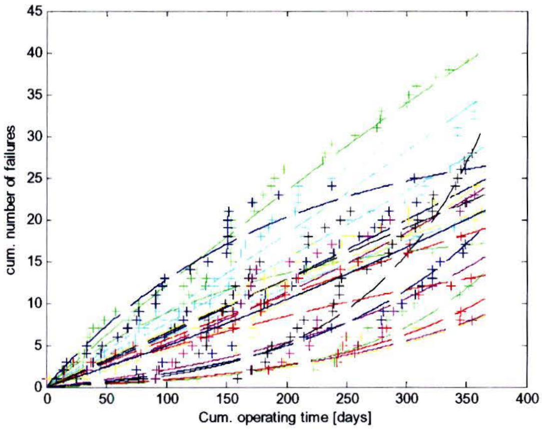


Figure 108 dataset 2, 365 days, unfiltered data

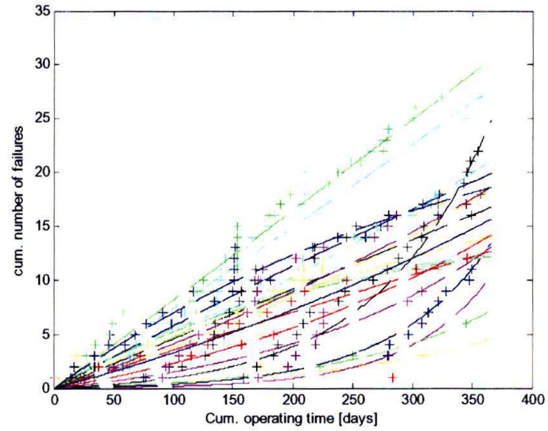


Figure 109 dataset 2, 365 days, filtered data

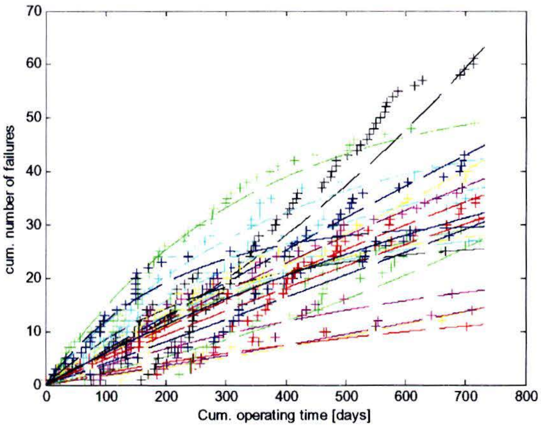


Figure 110 dataset 2, 730 days, unfiltered data

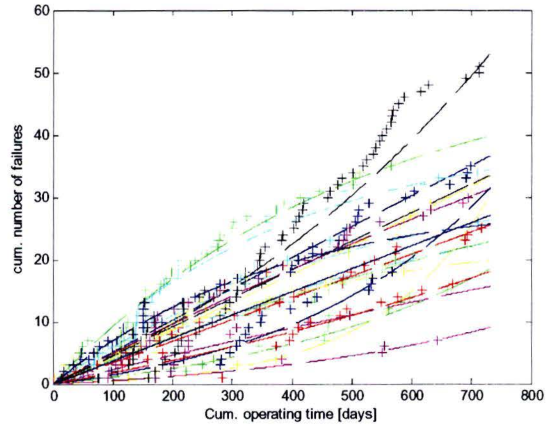


Figure 111 dataset 2, 730 days, filtered data

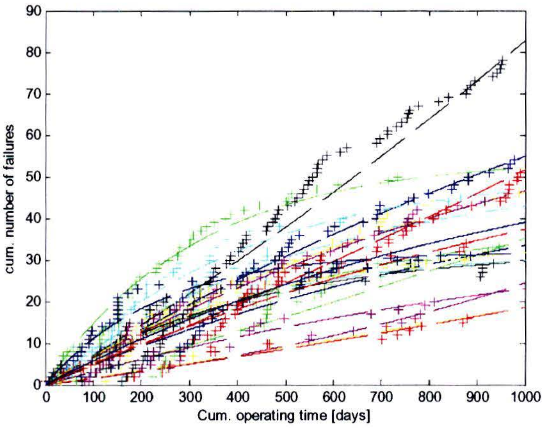


Figure 112 dataset 2, 1000 days, unfiltered data

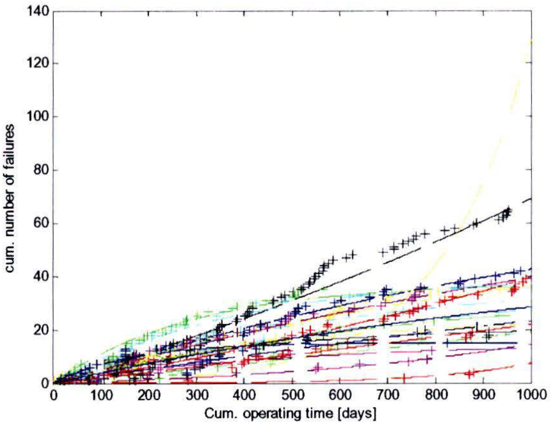


Figure 113 dataset 2, 1000 days, filtered data

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved

PHILIPS

Appendix T Dataset 3 FMT versus Service data

Results of the goodness-of-fit tests.

Power law

Table 48 Cramér-von Mises

System	FMT hours	$C^2(n)$	Test statistic	Result	Service days	$C^2(n)$	Test statistic	Result
1	1000	0.109	0.165	Accept	150	0.143	0.169	Accept
2	1000	0.083	0.169	Accept	150	0.075	0.165	Accept
3	1000	0.455	0.167	Reject	150	0.667	0.160	Reject
4	1000	0.114	0.162	Accept	150	0.161	0.165	Accept
5	1000	0.067	0.165	Accept	150	0.153	0.165	Accept
6	1000	0.140	0.171	Accept	150	0.317	0.165	Reject
7	1000	0.675	0.169	Reject	150	0.165	0.160	Reject
8	1000	0.176	0.171	Reject	150	0.071	0.167	Accept

Table 49 Chi-squared test

System	FMT hours	χ^2	Test statistic	Result	Service days	χ^2	Test statistic	Result
1	1000	2.00	3.84	Accept	150	10.33	5.99	Reject
2	1000	5.00	5.99	Accept	150	3.17	3.84	Accept
3	1000	8.50	3.84	Reject	150	4.70	3.84	Reject
4	1000	4.40	3.84	Reject	150	4.00	3.84	Reject
5	1000	5.00	5.99	Accept	150	0.67	3.84	Accept
6	1000	2.55	5.99	Accept	150	2.17	3.84	Accept
7	1000	21.25	5.99	Reject	150	1.83	3.84	Accept
8	1000	3.75	3.84	Accept	150	4.80	3.84	Accept

Exponential law

Table 50 Chi-squared test

System	FMT hours	χ^2	Test statistic	Result	Service days	χ^2	Test statistic	Result
1	1000	3.00	3.84	Accept	150	10.20	5.99	Reject
2	1000	3.92	5.99	Accept	150	5.99	3.84	Reject
3	1000	14.75	3.84	Reject	150	0.70	3.84	Accept
4	1000	5.44	3.84	Reject	150	4.57	3.84	Reject
5	1000	1.83	5.99	Accept	150	2.13	3.84	Accept
6	1000	6.37	5.99	Reject	150	4.03	3.84	Reject
7	1000	9.33	5.99	Reject	150	3.70	3.84	Accept
8	1000	2.29	3.84	Accept	150	6.29	3.84	Reject

Copyright © 2004 Philips Medical Systems Nederland B.V.
All rights reserved