Data-driven, context-aware improvement of the periodic maintenance performance of IXR-devices

van de Ven, T.

Award date:
2020

EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

Master

Data-driven, context-aware improvement of the periodic maintenance performance of IXR-devices

by

Ties van de Ven

Graduation thesis

Data-driven, context-aware improvement of the periodic maintenance performance of IXR-devices

In partial fulfilment of the requirements for the degree of

Master of Science

In Operations Management and Logistics

Author:	ing. Ties van de Ven
Student ID:	1276514
University:	Eindhoven University of Technology
Program:	Operations Management & Logistics
Capacity group:	Information Systems
Supervisor:	Dr. Ir. H. Rik Eshuis
Second supervisor:	Dr. Yingqian Zhang
Company:	Philips Healthcare
Company supervisors:	Jan Korst, Verus Pronk and Maikel Boumans
Date:	March 10, 2020

Abstract

At Philips planned maintenance (PM) of the medical devices is executed on a periodic interval, it is aimed to minimize planned downtime while ensuring safety and reliability of the medical devices. The current static PM schedule applies to all Interventional X-Ray (IXR) devices and is based on worst-case (usage) behavior, even though not all medical devices behave similar. This research focusses on the IXR-devices. Here it is proposed to enable a dynamic PM schedule for PM tasks considering the actual usage and operational context of a medical device. The proposed model to improve the current PM policy exists of three maintenance policies: Calendar-Time-Based Maintenance, Usage-Based Maintenance and Usage-Severity-Based Maintenance. The model was tested on two different PM tasks. The availability of data that could be derived from the Failure Mode and Effect Analysis report determined to what extent the model was tested. Next to the model, this research proposes a redesign of the business process in order to fully apply the model and to improve the PM policy periodically.

Summary

Problem context

This report describes research on improving the current planned maintenance (PM) policy of Philips with regard to medical Interventional X-Ray (IXR) devices. This research is highly relevant given the current large number of labor hours associated with PM of IXR-devices, that cause a significant part of the total maintenance costs. Maintenance of medical devices is important and in some cases mandatory because of strict safety regulations. A safe and reliable usage of a device must be ensured with the help of maintenance actions.

Currently, the installed base exists of roughly 10.000 IXR-devices that all require PM. Because of that, the labor costs associated with PM of IXR-devices are a significant part of the total maintenance costs. PM consists of multiple specific PM tasks that are determined during a Failure Mode and Effect Analysis (FMEA) that is performed by Research and Development (R&D). Generally, potential failures are prevented by executing PM tasks at a periodic interval. The FMEA report considers safety factors and therefore the interval of each PM task is rather conservative. The determined intervals for the execution of PM tasks are similar for all IXR-devices, even though research has determined that there is a large variation in usage among devices. This makes the current PM schedule static. Next to the static PM schedule, a lack of insight in available data and a lack of evidence regarding the effectivity of PM are regarded as other causes of the high labor hours spent on PM.

Research goals and questions

Philips desires to ensure a safe, reliable IXR-device with minimal downtime for the customer. In terms of productivity, Philips wants to achieve those goals with minimal costs. The research goal is to provide insights on how the large number of labor hours associated with PM could be reduced, such that the costs for the supplier, Philips, as well as downtime for the customer can be minimized. Research on PM concluded that in order to address this goal, improvement of the current PM policy should be based on contextual factors: actual usage and operational context, since this is most relevant from both scientific and business point of view. To the best of our knowledge, the possibilities of considering operational context (context-awareness) to the existing maintenance policies in the field of medical devices (or related fields) is still, to a large extent, unexplored. Therefore, this research focusses on the optimization of a PM policy for medical devices by considering the contextual factor: operational context. This is done such that variation in usage and degradation of a sub-system is removed and PM tasks could be executed more accurately. By doing this, the aim of the transformation from a static PM checklist to a dynamic PM checklist can be realized.

Philips extracts large amounts of monitored data of devices and therefore the application of data for determining the PM interval gets more attractive. This finally leads to the main research question and the following deliverables:

How to improve the current planned maintenance policy of Interventional X-Ray devices with the use of data, considering contextual factors?

1. Report of qualitative and quantitative analysis that shows the current PM performance from both perspectives. Overview of contextual factors that influence degradation of a device.
2. A single parameter-model and multi-parameter model that consider the contextual factors in order to improve the current PM policy.
3. Redesigned business processes in order to improve the PM policy periodically.

Methods

In order to effectively and thoroughly answer the research question, the regulative cycle was used as the main methodology. Moreover, the cross industry standard process for data mining (CRISP-DM) methodology was integrated within this main methodology.

The first phase of the regulative cycle is the *problem formulation,* this phase was integrated with *business understanding (CRISP-DM)* since both phases contain overlap. This phase focused on establishing a deeper understanding of the current PM policy and the clarification of the research questions.

The subsequent *diagnose* phase was integrated with the *data understanding* and *data preparation (CRISP-DM)* in order to identify the quality and availability of monitored data on PM. By using literature and semi-structured interviews with subject matter experts (SMEs) of R&D, the process of how a PM task is developed and what contextual factors play a role in degradation of a device in general were determined. Within this phase, data was prepared to test the hypotheses regarding the contextual factors and to act as input variables for useful datasets dedicated to prognostics. By having semi-structured interviews with SMEs from service innovation (SI), which are responsible for the PM manual, it was determined what tasks are currently executed during PM, what the main bottlenecks are that cause the large number of labor hours associated with PM and of which PM tasks the interval can possibly be stretched.

The subsequent *design* phase showed two artifacts: firstly, a model to improve the current PM policy, that considers the contextual factors, and secondly, a redesign of the business process aimed at improving the current PM policy periodically. The proposed model contains three maintenance policies that consider contextual factors ‘actual usage’ and ‘operational context’. The applied maintenance policies were: calendar-time-based maintenance (CTBM), usage-based maintenance (UBM) and usage-severity-based maintenance (USBM). The model consists partly of prognostics of the defined failure parameter(s). Therefore, the *modeling* and *evaluation* phase were a necessity. Since actual usage of a device does not change substantially over time, single linear regression was used to do prognostics for the failure parameter(s). A single-parameter model (usage) and a multi-parameter model (including operational factors) were applied in order to determine the interval of execution of a PM task.

Model to improve current planned maintenance policy

The model was experimentally tested on two of the scoped PM tasks (PM task J and K), because these two tasks had sufficient usage data available (e.g. number of clinical procedures or distance travelled by axis). For both PM task J and K, the model was applied to maintenance policies CTBM and UBM considering mainly actual usage. For CTBM the operational context was considered as well. Figure 1 and 2 show the results of applying the proposed model to PM task J and K. Over a time period of 4.5 years, application of CTBM lead to a 25% reduction of the number of executions of PM tasks J. CTBM does not directly reduce the number of executions of PM task K. Usage of the multi-parameter model including operational context CTBM (OC) lead to a reduction of 44% and 10% respectively. Looking at UBM, execution of PM task J and K is even reduced with 80% and 94% respectively. The total reduction of PM labor hours of a maintenance cycle was 2%-6% when adapting the model to PM task J and K. On the first eye, this is not a large difference, but it has to be considered that PM task J and K are relatively small compared to the remaining PM tasks upon which is scoped. The model was validated by SMEs of R&D. Especially CTBM can be considered as a quick win, because of the impact and easiness of implementation.

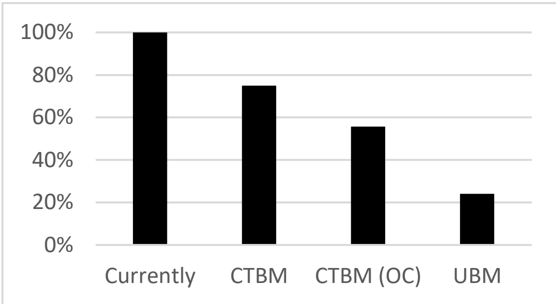


Figure 1: Fraction of number of executions PM task J.

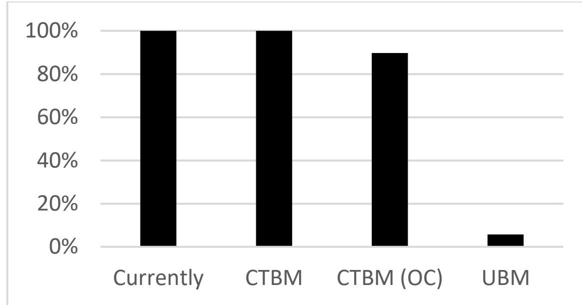


Figure 2: Fraction of number of executions PM task K.

Redesign of business process

In order to periodically improve the maintenance policy, the business process of the development of PM tasks needed redesigning. As a requirement of this business redesign, the failure data regarding the PM should be gathered by adding a form to the current PM checklist where field service engineers (FSEs) sign to indicate the necessity of performing PM tasks. By assessing these checklists, failure parameter(s) that indicate the actual need for a PM task can be determined/updated. Redesigning the business process required an extra action to be executed after performing the FMEA by R&D. In fact, R&D checks the application of dynamic maintenance instead of applying the fixed conservative interval for all devices.

The actions associated with this extra action are shown in Figure 3. A PM task prevents a sub-system from failures. PM tasks of a new sub-system or sub-systems with unknown failure parameter(s) start with the current conservative intervals. Within the redesign, the failures of PM are monitored constantly and failure parameter(s) are defined/updated every six months. With the defined failure parameter(s) and their corresponding threshold values, the proposed model could be applied to eventually establish a dynamic execution of a specific PM task. Every single sub-system, that is related to such a PM task, should be investigated separately by looking at what failure parameter(s) influence degradation. The failure

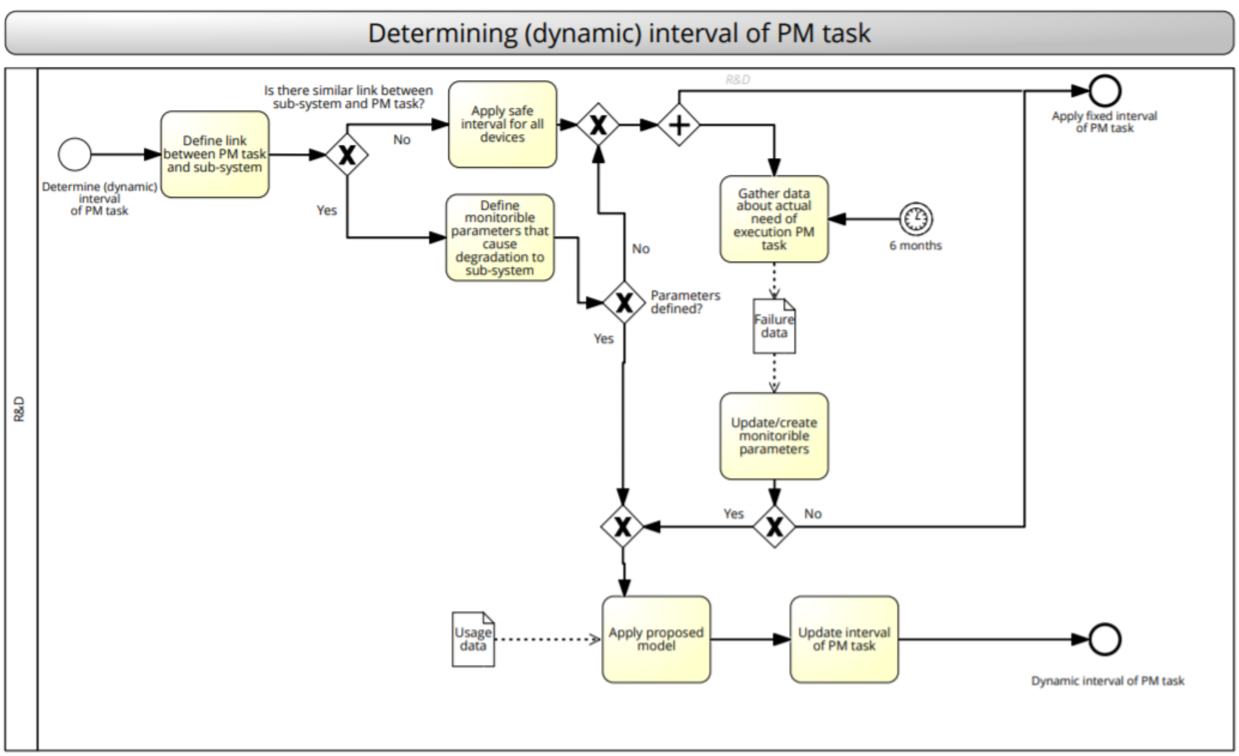


Figure 3: Redesign of business process (determining (dynamic) interval of PM task).

parameters include all types of indicators that influence degradation of a device such as number of clinical procedures, type of clinical procedures, number of times the sub-system went on/off, meters of movement, wear, temperature, time etc. According to SMEs of R&D, the proposed redesign of the business process would be promising for the long term since customized maintenance can be realized.

Conclusions and recommendations

This research concludes that the current static PM checklist can be transformed to a dynamic PM checklist where PM tasks are only executed if completely necessary. The proposed model shows that updating the fixed interval of the whole fleet according to actual usage and operational context could, depending on sub-system, reduce significantly the number of executions (CTBM and CTBM (OC)). Following that, looking at the actual usage of a device individually (UBM), reduces the number of executions even more. These reductions in execution of maintenance reduced the total number of PM labor hours with 2%-6%. This research also found that when redesigning the business process carefully the model to improve PM policy can potentially be applied to the remaining PM task at Philips as well as other companies that deal with high safety factors.

This research concludes that a periodic maintenance policy can be improved with the proposed model that uses data (data-driven) and considers the operational context (context-awareness). However, some dimensions of the model are still unexplored.

Firstly, the proposed model was applied to two out of the seven PM tasks. Future research regarding the remaining PM tasks is required to indicate the impact of applying this model to those PM tasks. In case of poorly known failure parameters for a particular sub-system, failure data regarding the related PM task should be gathered first.

Secondly, for the prognostics, a single linear regression model was used with fixed parameter values. The performance of the proposed model could be further improved if the prognostics method would be more accurate. Perhaps other applicable will perform better relative to the current method.

Thirdly, the role of other factors that might influence degradation of a sub-system, next to the failure parameters, should be explored. Now, failure parameters are based on the FMEA report. But, for example, the role of age is not considered, even though this might play a role in degradation of the sub-system.

Fourthly, within this research the role of degrading sub-systems that cause degradation of other sub-systems in not taken into account. It would be relevant to investigate if particular sub-systems do influence the rate of degradation to other sub-systems.

Contents

Table of figures

List of tables

List of equations

Table of Abbreviations

Abbreviation	**Meaning**
BIU	Business Innovation Unit
CBM	Condition-based Maintenance
CM	Corrective Maintenance
CRISP-DM	Cross Industry Standard Process for Data Mining
CTBM	Calendar Time-Based Maintenance
CTQ	Critical To Quality
CTS	Critical To Safety
DHF	Design History Form
e.g.	“For example”
FDA	Food and Drug Administration
FDA	Food and Drug Administration
FMEA	Failure Mode and Effect Analysis
FSE	Field Service Engineer
FTE	Full-Time Equivalent
IXR	Interventional X-ray
OEM	Original Equipment Manufacturer
PM	Planned Maintenance
R&D	Research and Development
RMM	Risk Management Matrix
SI	Service Innovation
SME	Subject Matter Expert
SWO	Service Work Order
UBM	Usage-Based Maintenance
USBM	Usage Severity-Based Maintenance

1. Introduction

The aim of this research project is to investigate the possibilities to improve a periodic maintenance policy by using data-driven and context-aware methods to predict the moment of execution of planned maintenance (PM) at a medical device. This is useful since it could potentially reduce costs from supplier perspective and reduces planned downtime for the customer.

As this research project is conducted within a company, first the company context will be introduced in this chapter. After that, the scientific relevance, research goal and deliverables will be explained.

1.1 Company info

Gerard Philips and his father Frederik Philips founded Royal Philips of the Netherlands (Philips) in 1891. Frederik financed the purchase of an empty factory building in Eindhoven where production of carbon-filament lamps and other electro-technical products started in 1892. In 1912 Philips became a company listed on the Amsterdam Stock Exchange. To stimulate product innovation, Philips established a research laboratory, NatLab, in 1914. From that moment onwards, Philips invented many products in different fields as X-rays, radio reception, radios, televisions and electric shavers (N.V., 2020).

Nowadays, Philips is mostly known for its strong presence in the healthcare and lighting industry. Because of strategic reasons, Philips mainly focuses on healthcare. Therefore, Philips split off their lighting department in 2016. This department continued with a new name called Signify N.V.. Philips still owns a small part of Signify N.V. but the annual report of 2016 states that they will reduce their stake to zero eventually (N.V., 2020).

Philips consists of four divisions namely: ‘Personal Health’, ‘Diagnosis and Treatment’, ‘Connected Care’ and ‘Other’. In 2018, Philips had around 77.000 FTE and a revenue of EUR 18.1 billion. The largest part of Philips' research division is located at the High Tech Campus Eindhoven and the headquarter resides in Amsterdam since 2007.

Philips aims at improving healthcare by the use of advanced technology. Downtime of the medical devices provided by Philips may have large consequences from both economic and social perspective. Therefore, the downtime of the medical devices should be minimized. In order to achieve this, maintenance is performed in a proactive way by executing maintenance on a periodic interval.

1.2 Scientific relevance

Many companies are depending on the availability of their capital goods to keep the primary processes running (e.g. aircraft, trains and MRI scanners). Downtime of those capital goods could lead to a significant loss of revenue and unsatisfied customers. The unavailability of medical devices could even lead to loss of human life under some circumstances. It is thus important to keep capital goods running even though it could be very expensive to realize (Arts, 2016).

Within maintenance, different policies exist to maintain a particular capital good. Those policies were developed during the past decades. The earliest maintenance technique is corrective or unplanned maintenance (CM). This takes place when a (unexpected) breakdown occurs. Because of its simplicity, this maintenance approach is widely used in many industries.

Later on, the availability of capital goods gained more importance. To avoid unexpected downtime, companies started to use a preventive maintenance approach, called time-based maintenance or PM. This maintenance approach determines a time/planned interval to perform maintenance regardless of the health status of the capital good. Consequently, many failures were avoided, but the downside of this approach is that it caused a rise in maintenance costs. In order to save costs but still being reliable, multiple maintenance policies emerged.

First, this section shows four well known maintenance policies: calendar-time-based maintenance (CTBM), usage-based maintenance (UBM), usage-severity-based Maintenance (USBM) and condition-based maintenance (CBM). Secondly, the application of maintenance policies that consider contextual factors in the field of healthcare will be discussed.

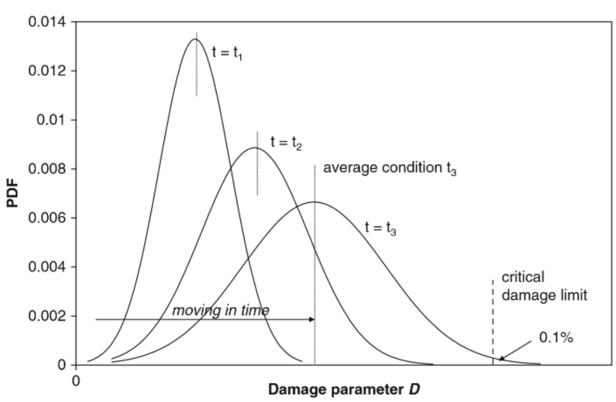
Factors that cause a conservative PM interval

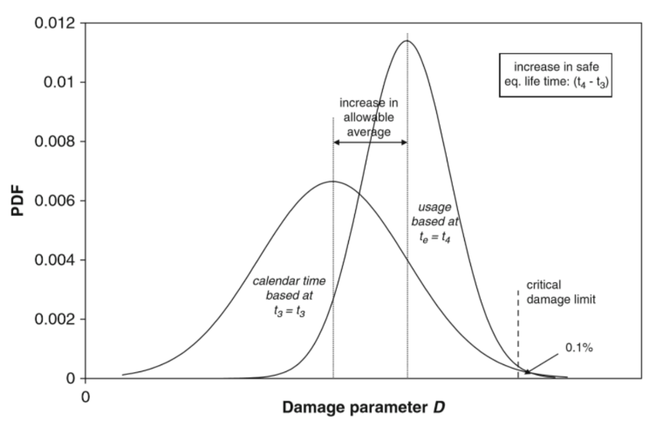
Tinga (2010) shows how physical models deal with uncertainty and how it causes very conservative time intervals of PM tasks. During the design of a device, the original equipment manufacturer (OEM) must determine the maintenance intervals. At that stage, the actual number of usages is unknown in most cases. The same holds for the severity per type of usage. Because of missing this crucial information, OEM assumes a conservative usage parameter and conservative severity of the type of usage. Lifetime of the device is computed that leads immediately to prescribed maintenance intervals. Generally, these assumptions made by OEM are covered by safety factors.

Figure 4a shows the procedure for the computation of lifetime. Possible values of the damage parameter D (e.g. amount of wear) at three moments in time (t_1, t_2, t_3) represent three normal distributions (Tinga, 2010). The uncertainty within this damage parameter is caused by three sources: the actual usage is unknown, uncertainty in the effect of usage (severity per unit type) and variations in the life consumption for a given internal load.

Another factor that causes uncertainty, that is not mentioned by Tinga (2010), is the operational context of a device. Moubray (1999) describes that reliability-centered maintenance should be performed in only one specific operational context (e.g. country, type of (surgery) room, type of device or climate) in order to remove the variation caused by context.

Since degradation proceeds in time, the damage gradually increases. This is represented by the moment of the distribution of D to the right (Figure 4a). At the same time, the width of distribution increases since the relative uncertainties are assumed to be constant. The sub-system must be replaced (or repaired) when the tail of the distribution reaches a threshold value (probability of being damaged). Critical components will be replaced (or repaired) when only, for example, 0.1% of the distribution or less has reached the threshold value (damage limit). The figure shows that, due to large uncertainties in condition prediction, the sub-system is already replaced (or repaired) at t_3, even though the average sub-system is far below this critical point. Since actual usage and operational context is unknown, intervals must be computed based on most severe usage which clearly results in conservative maintenance intervals. This is the statistical background of safety factors that is mentioned before and applied by OEM.





Left: Figure 4a, width of distribution increases because of uncertainties. Right: Figure 4b, less uncertainties causes less variation, this narrows the distribution and expected lifetime increases (CTBM vs. UBM) (Tinga, 2010).

Calendar-time-based maintenance (CTBM)

The easiest way of defining intervals is by using fixed periods. Those intervals can be obtained based on experience (historical data) or model-prognostics. For a fleet of devices with minimal variation in usage among each other or with a degradation process that is directly related to time, CTBM is very suitable (Tinga, 2010). However, if a fleet of devices has significant variation in usage or degradation that does not depend directly on time, it is almost impossible to accurately determine an interval by CTBM because a lot of assumptions have to be made on the usage of the devices in a certain period of calendar time. This yields, in practice, that a very wide distribution and a very conservative failure time will be selected in order to obtain an acceptable probability of failure.

Usage-based maintenance (UBM)

By applying the concept of UBM, the conservatism in the maintenance intervals can be reduced by taking into account the actual usage of a device. Very common parameters to use are usage parameters such as operating hours or distance of a conveyer belt. The advantage of UBM compared to CTBM is that intervals of devices can be extended easily when the device is not or only partly used in some period of time. The uncertainty in usage will not be present anymore since it is assumed actual usage is monitored and the relation between usage and degradation is known. Removing the uncertainty in usage reduces the width of the distribution, which leads to an increase of service life. This can be seen in Figure 4b. Tinga (2010) shows that the tail of the original distribution reaches the damage limit at t_3. By considering actual usage, the damage limit is reached significantly later. The service life of a sub-system thus increased from t_3 to t_4 (Figure 4b). Nevertheless, applying UBM still assumes a conservative severity per parameter unit.

In literature more papers are present that apply this concept of UBM. Compared to Tinga (2010) their approaches differ because they use experience-based models instead of physical-based models (Fraser, 1994; Molent, 1998; Hunt & Hebden, 2001; Lieven, 2006).

Usage-severity-based maintenance (USBM)

To take into account the usage severity during operation of a system, Tinga (2010) proposes the USBM approach. This method is only applicable if variation in type of usage is monitored. This includes for example, a measurement tool that measures speed/load per operating hour. Next to that, the relation between usage severity and degradation of the sub-system must be known. Therefore, physical-based models have to be created. By applying USBM, the uncertainty of the effect of usage (severity per unit type) will be removed and prediction of the lifetime (t_5) will be more accurate. Therefore, lifetime is expected to increase in general ($t_5 > t_4$).

In literature similar approaches of USBM have been applied. Heine (2007) describes terrain identification for a military vehicle with the application of accelerometers. By correlating the data on damage of different terrain types, an estimate of severity per different terrain is made. Pecht and Vichare (2006) consider the actual monitored usage of electronic devices (e.g. memory usage, number of on/off switches) to compute the expected service life. Within this analysis, variations in temperature are considered to determine the severity. Both approaches include severity of actual usage by applying an experience-based model. The approach proposed by Tinga (2010), as previously mentioned, applies a physical-based model.

Replacements or reparations associated with UBM or USBM can be done in a passive or pro-active way. By doing this passively, the actual usage is monitored until it reaches a certain threshold value and replacements/reparations take place immediately. The pro-active way uses prognostics and therefore the replacement/reparation can be planned in the future, while the latter is only suitable when usage can be predicted.

Condition-based maintenance (CBM)
Condition-based maintenance (CBM) is the most direct way of predicting the expected service life of a component. CBM requires a direct health assessment of a component on a regular basis. Because CBM assesses the health condition of a sub-system, the remaining uncertainty is removed. Therefore, the width of the distribution (Figure 5) will be narrowed even further. A CBM program consists of three key steps: data acquisition, data processing and maintenance decision making (Jardine, Lin, & Banjevic, 2006). Data acquisition includes collecting of relevant information. Data processing, information handling, includes handling and analyzing the data generated from the data acquisition. The last step: maintenance decision making, can be divided into two main categories: diagnostics and prognostics. Diagnostics focuses on detection, isolation and identification of faults. Prognostics in turn attempts to predict faults or failures before they occur. This approach is the most accurate one compared to the mentioned policies. Nevertheless, investing in a CBM monitoring system is more expensive.

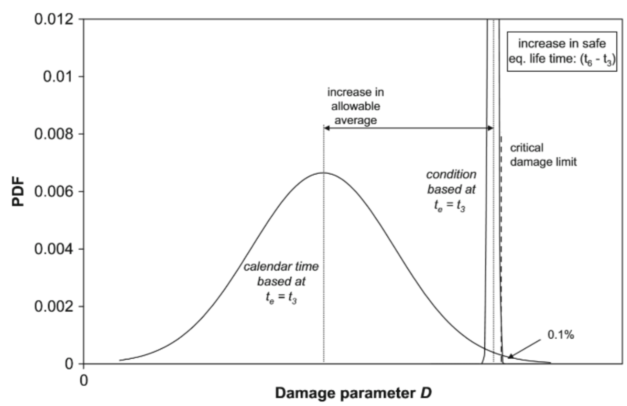


Figure 5: CBM increases expected lifetime significantly (Tinga, 2010).

Maintenance policies that consider contextual factors in the field of healthcare
Looking at the maintenance policy of healthcare organizations, most of them merely follow manufacturer's recommendations and do not profit from maintenance excellence to the same extent as other industries. Maintenance optimization models for medical devices are still, to a large extent, an underexplored area. Based on a mapping review, Rincon (2012) states that applying models to optimize maintenance strategies in the healthcare domain is scarce and still in its infancy stage.

The prognostics of the remaining useful life (RUL) hardly depends on contextual factors (usage and operational context) (Jardine, Lin, & Banjevic, 2006; Altay & Green, 2006; Kim & Kuo, 2009). The RUL is critically important for a used device because of their impact on the planning for maintenance (Cui, Loh, & Xie, 2004; Wang & Zhang, 2008; Lee, Lapira, Bagheri, & Kao, 2013). Especially flexibility in operational context has emerged as an important requirement in the design of business processes (Rosemann, Recker, & Flender, 2008). Nevertheless, the concepts of "context" and "context aware system" have not been well utilized by researchers from field of predictive maintenance (Kumar, Ahmadi, Verma, & Varde, 2016). Analysis of ten review and survey papers in the area from time period 2005–2014 mentioned in literature reveals that term the "context", or words similar to "context", are never directly mentioned in the context of predictive maintenance (Schmidt & Wang, 2015). At this moment of time, a few papers related to contextualized predictive maintenance have been published, those papers describe methods and models for predictive maintenance from a generic academic perspective without any practical application (Si, Wang, Zhou, & Chang-Hua, 2010; Carl-Anders Johansson, 2014; Galar, Thadari, Catelani, & Ciani, 2015).

To the best of our knowledge, considering context-awareness to the proposed maintenance policies has not been tested before in the field of medical devices (or related fields). Contextual factors can be classified in two categories: usage and operational context. The maintenance policies mentioned apply actual usage but do not take the operational context (context-awareness) of a device into account.

Therefore, this research focusses on the optimization of the maintenance policies for medical devices by considering contextual factor ‘operational context’, such that variation in usage and variation in degradation is removed. So that, PM tasks could be executed more accurately.

1.3 Research goal

This research aims to investigate the current PM approach of IXR-devices and explore opportunities to improve the PM performance without effecting the corrective maintenance (CM) performance.

IXR-devices are medical devices that are used in the field of image-guided therapy. These devices are used to obtain live images of soft tissues during interventions, using X-rays. Image guided therapy is the use of medical imaging to plan, perform and evaluate medical interventions. This research focusses on the releases: Allura R8.1, Allura R8.2 and Azurion, because those devices are sold most frequently globally (around 10.000 devices).

Currently, the labor costs associated with PM of IXR-devices are a significant part of the total maintenance costs. The causes of high labor costs and corresponding downtime for the customer are shown in Figure 6 using a cause-effect diagram (Rasul, 2005). It was decided that the research focuses on causes within the orange circle, because they are considered the main reason for the high labor costs and subsequent downtime for the customer.

From *procedures* perspective, the PM checklist and schedule are similar for all IXR-devices. This makes PM static, because there might be contextual differences in usage or operational context. The main reason of the current static PM policy is a lack of *measurements* regarding the effectivity of PM tasks and involves a lack of insight in the current available data. The remaining causes are not considered for this research, because it is expected that solving the mentioned main causes will (partly) solve the remainders. Obtaining *measurements* helps to challenge the current PM intervals (from *people perspective*) and the current performance can be reviewed (*performance perspective*). The safety regulations associated with medical devices (*procedures perspective*) cause fixed intervals for particular PM tasks and are considered a requirement.

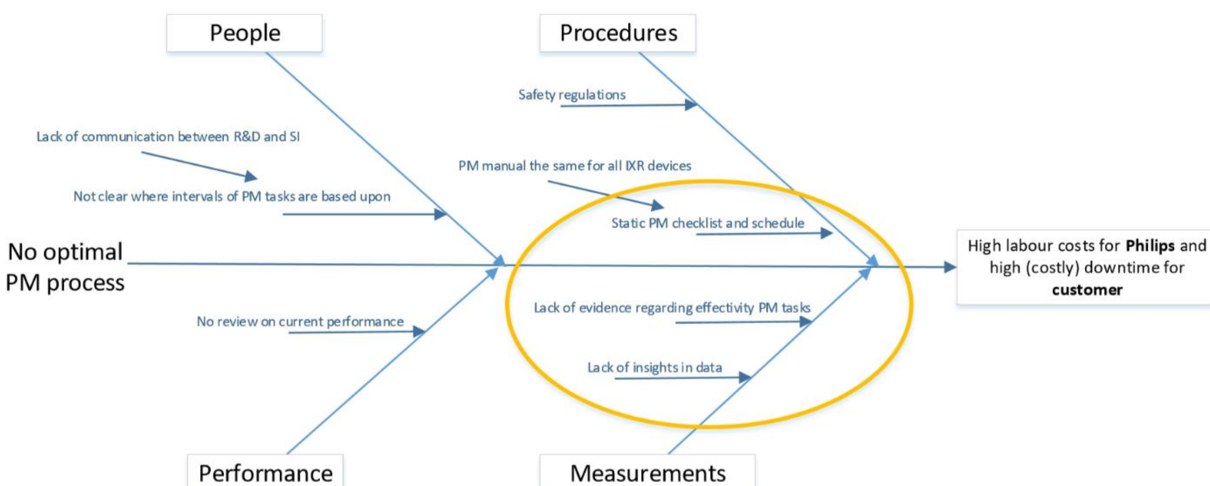


Figure 6: Causes of high labor costs for Philips and high (costly) downtime for the customer.

Since many years Phillips extracts large amounts of monitored data of the IXR-devices and therefore the application of usage data for determining a PM interval gets more attractive. The availability of usage data is essential to get insight in the current performance of the maintenance policy and therefore possible improvements of the frequency and effectivity of maintenance tasks can be explored.

Considering the aim from both business and scientific perspective, leads to the following main research question:

How to improve the current planned maintenance policy of IXR-devices with the use of data, considering contextual factors?

The following research questions (RQs) are formulated in order to answer the main research question:

1. What is Philips' strategy regarding the maintenance policy?
2. What is the role of the different stakeholders within the current PM process?
3. What data is available to analyze the current PM tasks?
   3.1. What are the opportunities and limitations of PM data?
   3.2. How to evaluate the effectivity and efficiency of a PM task with data?
4. How do the contextual factors influence the number of failures in general?
   4.1. What are the contextual factors from literature, expert and data perspective?
   4.2. What is the impact of the contextual factors?
5. Which tasks are currently executed during PM of an IXR-device?
   5.1. Which tasks does a PM visit consist of?
   5.2. How much time does each PM task require?
6. Combining the qualitative and quantitative analyses, which intervals of PM tasks are feasible to stretch?
7. How to design a model to improve the current PM policy considering contextual factors?
   7.1. Which maintenance policies can be used to improve PM?
   7.2. How to design a single-parameter model?
   7.3. How to design a multi-parameter model including operational context?
   7.4. What is impact of applying the model?
8. How to implement the new PM policy?

1.4 Deliverables

In order to improve the current PM policy three deliverables are developed:

1. Report of qualitative and quantitative analysis that shows the current PM performance from both perspectives. The analysis shows the proportion of time of a PM task within a maintenance cycle in order to focus on the main bottlenecks (quantitative). It becomes clear whether the interval of a PM task can possibly be stretched (qualitative). Next to that, the role of the contextual factors that influence the number of failures is explained.
2. A model that considers the contextual factors in order to improve the current PM policy. The model uses prognostics to determine whether execution of a PM task can be postponed to the next visit. First, a single-parameter model is shown which considers actual usage of the device. After that, a multiple-parameter model shows the impact of considering the operational context parameters. The model is able to estimate an improved interval of a certain PM task considering the contextual factors.
3. Redesigned business processes in order to improve the PM policy periodically.

2. Research methodology

This chapter describes the methodology used in order to analyze the problem and to answer the research questions in a structured way. This research applies a combination of two methodologies, namely the regulative cycle and the cross industry standard process for data mining (CRISP-DM). Both methodologies complement each other since the analytic model (CRISP-DM) describes the steps executed within the regulative cycle detailed.

2.1 Regulative cycle

The regulative cycle is a widely used research methodology with the focus on practical application (Aken, Berends, & Bij, 2007), it consists of five phases: problem formulation, diagnose, design, intervention and evaluation (Figure 7). To get a better understanding of the phases, a short description of every phase is given.

- *Problem formulation:* define the research objective which is confirmed by the stakeholders.
- *Diagnose:* map out the current state and analyze it in order to find bottleneck(s) and the causes of the bottleneck(s).
- *Design:* design plan to remove bottlenecks and to implement solutions.
- *Intervention:* execute plan. The plan can be adjusted during execution, as indicated by the cyclic character of the model.
- *Evaluation:* check if the initial goal is reached. If further improvements are necessary, formulate new problems and the regulative cycle starts again.

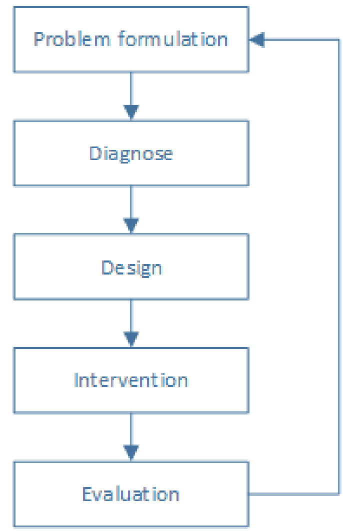


Figure 7: Regulative cycle steps.

2.2 CRISP-DM methodology

The CRISP-DM is an analytic model (Chapman, Clinton, & Kerber, 2000), it contains the life cycle of data mining projects and consists of six phases, shown in Figure 8. The sequence of the phases is not fixed, moving back and forth between different phases is required (Chapman, Clinton, & Kerber, 2000). The outcome of each phase determines which phase, or particular task of a phase, has to be executed next. The arrows indicate the most important and frequent dependencies between phases. The outer circle symbolizes the cyclic nature of data mining that does not end once a solution is deployed. The lessons learned can trigger new business questions. The six phases of the CRISP-DM model are shorty described:

- *Business understanding*: understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition and preliminary plan to achieve the objectives.
- *Data understanding:* initial data collection and proceeds with activities that enable you to become familiar with the data, identify data quality problems and discover first insights into data.
- *Data preparation:* covers all activities needed to construct the final dataset from the initial raw data. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools.
- *Modeling:* select and apply various modeling techniques. There can be a loop back to data preparation.

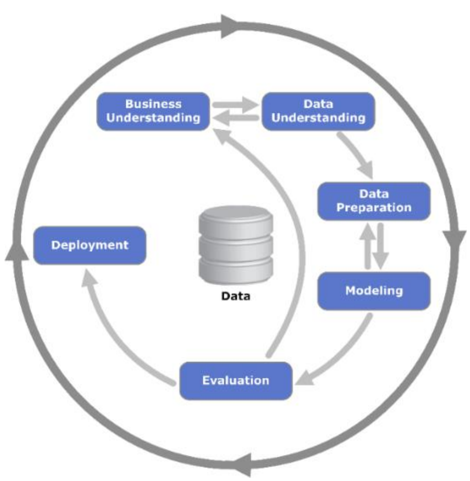


Figure 8: CRISP-DM steps.

- *Evaluation:* evaluate and review the model, to be certain that the model achieves the business objectives.
- *Deployment:* consider evaluation results and determine a strategy for their deployment.

2.3 Project approach

This section describes the application of both methodologies during the research in order to answer the research questions. The regulative cycle was used as the main methodology and the CRISP-DM methodology was integrated within this model. Because of limited time of this research, the focus was on the problem formulation, diagnose and design phase. Little attention was given to the intervention and evaluation phases. An overview of the research approach is given in Figure 9.

2.3.1. Problem formulation and business understanding

The first phase of the regulative cycle is the *problem formulation*. Since the *business understanding* phase of the CRISP-DM methodology partly overlaps with the *problem formulation* phase of the regulative cycle, they were merged in this section. Within this phase an initial understanding of the current situation was created by executing semi-structured interviews with supervisors of Philips and with SMEs of service innovation (SI) and research and development (R&D). Based on that, research questions and deliverables were determined.

2.3.2 Diagnose, data understanding and data preparation

The *diagnose* phase answers RQ 1 - RQ 6. Since the characteristics of the *data understanding* phase of the CRISP-DM contains overlap with the *diagnose* phase of the regulative cycle, both phases were interwoven in this section.

1. *What is Philips' strategy regarding the maintenance policy?*

To determine Philips' maintenance strategy, a semi-structured interview with the manager of the business innovation unit (BIU) was conducted. The BIU is the department that is responsible for creating the PM manual. The reason why this interview was held with the manager of BIU is his role in the organization, his responsibility of the PM manual, and his more than five years of experience in that role. The interview was structured such that the researcher could give direction to the interview in order to make sure that the Philips maintenance strategy was clear at the end.

2. *What is the role of the different stakeholders within the current PM process?*

The goal of RQ 2 was to determine the influential parties that are involved within the current PM process. During this phase, it was important to get a clear view of the current maintenance policy. Therefore, it was important to acquire information from the stakeholders such as R&D, SI and field service engineers (FSE). R&D develops the devices and the corresponding requirements regarding safety. SI, in turn, develops the PM manual to make sure all safety requirements are met. FSEs finally execute the PM tasks derived from the PM manual on site.

Information of these stakeholders was acquired using unstructured interviews. Since the researcher required a clear view of all stakeholders' perspectives, it was important to perform unstructured interviews and to ask stakeholders' opinion about the current PM policy as well as their exact role in this. The interviewed stakeholders were selected based on function (FSE, R&D and SI) and experience (at least four years) to make sure that a clear and reliable view of the current PM policy was obtained.

3. What data is available to analyze the current PM tasks?

In order to investigate the effectivity per PM task a quantitative analysis was conducted with the use of *data understanding* and *data preparation*.

3.1. What are the opportunities and limitations of the PM data?
3.2. How to evaluate the effectivity and efficiency of a PM task with data?

Within this phase the *data understanding* and *data preparation* phases were required in order to gather clean data for the contextual factors data analysis and for the prognostic purposes.

Data understanding
First of all, relevant data was extracted from the available sources. This phase started with exploring the database Vertica, which is primarily constructed for research purposes. Vertica contains daily log files from a variety of devices, including the data of the IXR-devices. Exploring this database generated knowledge about the type and relevance of data. Since Vertica includes many attributes, knowing the most relevant attributes was deemed necessary. Initial data collection took place during this phase in order to become familiar with the data. A crucial part of data understanding is the validation of the collected data, which was done with the assistance of data scientists. While collecting the data it turned out that the data could be categorized into three groups: device data, maintenance data and customer data.

Data preparation
The initial data obtained during the *diagnose* phase was converted to a structured dataset. In the process of realizing this conversion, different data preparation steps were executed. Firstly, the relevant data required for analysis was extracted from the initial dataset. Secondly, to increase the data quality to the level required, the data was cleaned from unwanted data. Thirdly, intermediate datasets were constructed by creating derived attributes or new records. The reasons for including or excluding data were documented, because that would prevent wrong interpretations and conclusions during the evaluation phase. Fourthly, data extracted from Vertica was merged with Python. Lastly, the used data was reformatted to be able to do the contextual factors data analysis and to do prognostics of the usage.

By executing semi-structured interviews with data scientists that are familiar with Vertica, multiple insights were obtained. The opportunities, limitations and in which way the effectivity and efficiency of a PM task is evaluated, were found. Next to that, to validate the data, a semi-structured interview took place with stakeholders from SI because of their experience and know-how with the data.

4. How do the contextual factors influence the number of failures?

The goal of RQ 4 was to investigate how the contextual factors influence degradation of a sub-system that is related to a PM task. Therefore, it was required to specify the contextual factors. In order to formulate contextual factors three perspectives were considered: interviews with SMEs of R&D and FSE, data of contextual factors and literature.

4.1. What are the contextual factors from literature, expert and data perspective?
4.2. What is the impact of the contextual factors?

To gain an understanding of the contextual factors, semi-structured interviews with SMEs were held. The SMEs in this case are the FSEs and employees of R&D, with at least four years of experience. FSEs execute the PM tasks daily and R&D designed the PM tasks. This means both stakeholders knew which contextual factors could play a role in degradation of a sub-system. Next to that, it was checked if data regarding the contextual factor was available. Finally, literature research concluded to what extent the contextual factors were considered as factors that influence degradation of a sub-system from literature perspective. The acquired contextual factors were categorized in usage and operational context. The hypotheses dedicated to the contextual factors were tested in order to determine whether they influence the degradation of a device on system-level. The degradation of a device was specified in the number of failures (CM cases) because of the missing preferable failure data of PM tasks.

5. *Which tasks are currently executed during PM of an IXR-device?*

The goal of RQ 5 was to investigate the tasks that are currently executed during PM by performing a qualitative analysis.

5.1. Which tasks does a PM visit consist of?
5.2. How much time does each PM task require?

During the qualitative analysis, gathering information happened by interviewing stakeholders (FSE, R&D and SI). The interviews happened in a structural way in order to get to know what tasks a PM visit consist of and the planned time per PM visit. Structuring the interviews made sure that all PM tasks would be evaluated and could be categorized. To gain a full picture of the current PM policy, we took observation of multiple experienced (at least four years) FSEs during the execution of a PM visit. A Pareto analysis (Stojcetovic, Živče, Lazarević, & Marjanović, 2015) was conducted, such that the project could be scoped upon the PM tasks that are causing the largest number of PM labor hours.

6. *Combining the qualitative and quantitative analyses, which intervals of PM tasks are feasible to stretch?*

By combining the qualitative and quantitative analysis, and showing this during a workshop with relevant, experienced (at least four years) stakeholders (SI and R&D), it was found of which PM tasks the interval could be stretched. Within this phase the safety regulations, which are established by the Food and Drug Administration (FDA) (Amant, 2017), were considered. The safety regulations gave insight in what PM tasks are mandatory to execute. For practical reasons the stakeholders could not be booked for the full day, and therefore the workshop took a couple of hours.

The deliverable of the *diagnose* phase was a report of the quantitative and qualitative analysis that showed the current PM performance from both perspectives. Within this phase, it turned out what the bottlenecks of the current PM policy were. Also, it was found whether the interval of a PM task can be possibly stretched and what the role of contextual factors, that influence the number of failures, is. Next to that, it turned out what the possibilities and limitations of the available PM data were.

2.3.3 Design, data modeling and evaluation

The *design* phase answers RQ 7 and RQ 8.

7. *How to design a model to improve the current PM policy considering contextual factors?*

The goal of this phase was to investigate how the interval of a PM task could be stretched by testing multiple maintenance policies (CTBM, UBM and USBM), considering contextual factors and prognostics. There is a trade-off because extending the interval should not lead to negative downsides such as more unplanned downtime of a device. For this part of the research the phases *data modeling* and *evaluation* of the CRISP-DM methodology were used as a guideline.

7.1 Which maintenance policies can be used to improve PM?
7.2 How to design a single-parameter model?
7.3 How to design a multi-parameter model including operational context?
7.4 What is impact of applying the proposed model?

A literature study was conducted in order to investigate the existing maintenance policies that could improve the PM policy at Philips. Knowing the available data at Philips, the contextual factors on system-level and knowledge sourcing from the SMEs of R&D, it was indicated what maintenance policies were applicable for IXR-devices.

Modeling
At the modeling phase, the *data preparation* phase is finished. With the use of the clean datasets and expert knowledge from data scientists at Philips, a suitable prognostics model was developed. Since the usage data does not change substantially over time, single linear regression was applied in order to do prognostics:

- *Linear regression based model.* Linear regression based methods are commonly used in industry and is possibly the simplest model to use for trending the degradation path (Si, Wang, Zhou, & Chang-Hua, 2010).

Evaluation
In this phase, the acquired results were evaluated. To make sure no activities would be missed during the development of the models, a process review took place. It was decided whether the research would proceed to deployment or that some adaptations would have to be done. This phase also evaluated whether the model could be generalized such that it potentially could be useful for improving the interval of other PM tasks.

8. How to implement the new PM policy?

During the *design* phase the model was tested on two PM tasks. This part described what would need to happen in order to apply the model for the remaining PM tasks (short term) and what business processes would need to be changed in order apply the model fully and to improve the PM policy periodically (long term). In order to control the implementation of the model successfully, the RACI matrix (Rasul, 2005) is designed to show which stakeholders are responsible for which tasks. The RACI matrix was developed in collaboration with the manager of BIU to make sure that possibly adaptations in employees' responsibilities could be executed.

The deliverables within the *design* phase were both a model that considers the contextual factors in order to improve the current PM policy and a redesigned business process in order to improve the PM policy periodically. A single-parameter model was applied, which considers the actual usage of a device. Next to that, a multi-parameter model was applied, that considers the impact of operational contexts parameters. The model is able to estimate an improved interval of a certain PM task considering the contextual factors.

2.3.4 Intervention and deployment

Since the *intervention* phase of the regulative cycle has overlap with the *deployment* phase of the CRISP-DM methodology, both were merged in this section. Within this phase plans to improve the current PM policy have to be implemented. Because of limited time of this research, this phase was not reached.

2.3.5 Evaluation

After implementation of the model and the redesign of the business process, a check is required in order to verify whether the initial goal is reached. In case there would be still room for improvement, a new problem formulation would be defined and the regulative cycle starts again. Again, because of limited time of this research, this phase was not reached yet.

To conclude, this chapter described the steps within the research method applied on detail in order to answer the research questions that lead to the deliverables. All research questions will be answered and elaborated on in the next chapters. Since the main phases were *diagnose* and *design* of the regulative cycle, the following chapters are named according to these phases. Both the *diagnose* phase and *design* phase are divided into two chapters.

Chapter 3 shows the current maintenance policy at Philips and availability of PM data (*diagnose).*

Chapter 4 describes the role of contextual factors, the bottlenecks within the current PM policy and concludes with an overview of which PM tasks the interval can possibly stretched of (report of quantitative and qualitative analysis) (*diagnose)*.

Chapter 5 describes the feasible maintenance policies within the model with regard to IXR-devices and shows how the model should be applied (single-parameter and multi-parameter model) (*design).*

Chapter 6 shows the results of applying the model on two PM tasks. Next to that, chapter 6 explains how the model could be applied to the remaining PM tasks and a redesign of the business process is proposed. Chapter 6 concludes with a validation of both artifacts the model and the redesign of the business process (*design)*.

This research concludes with chapter 7, a conclusion of the research from both business and scientific perspective. Next to that, this chapter shows the limitations of the research and suggestions for future research.

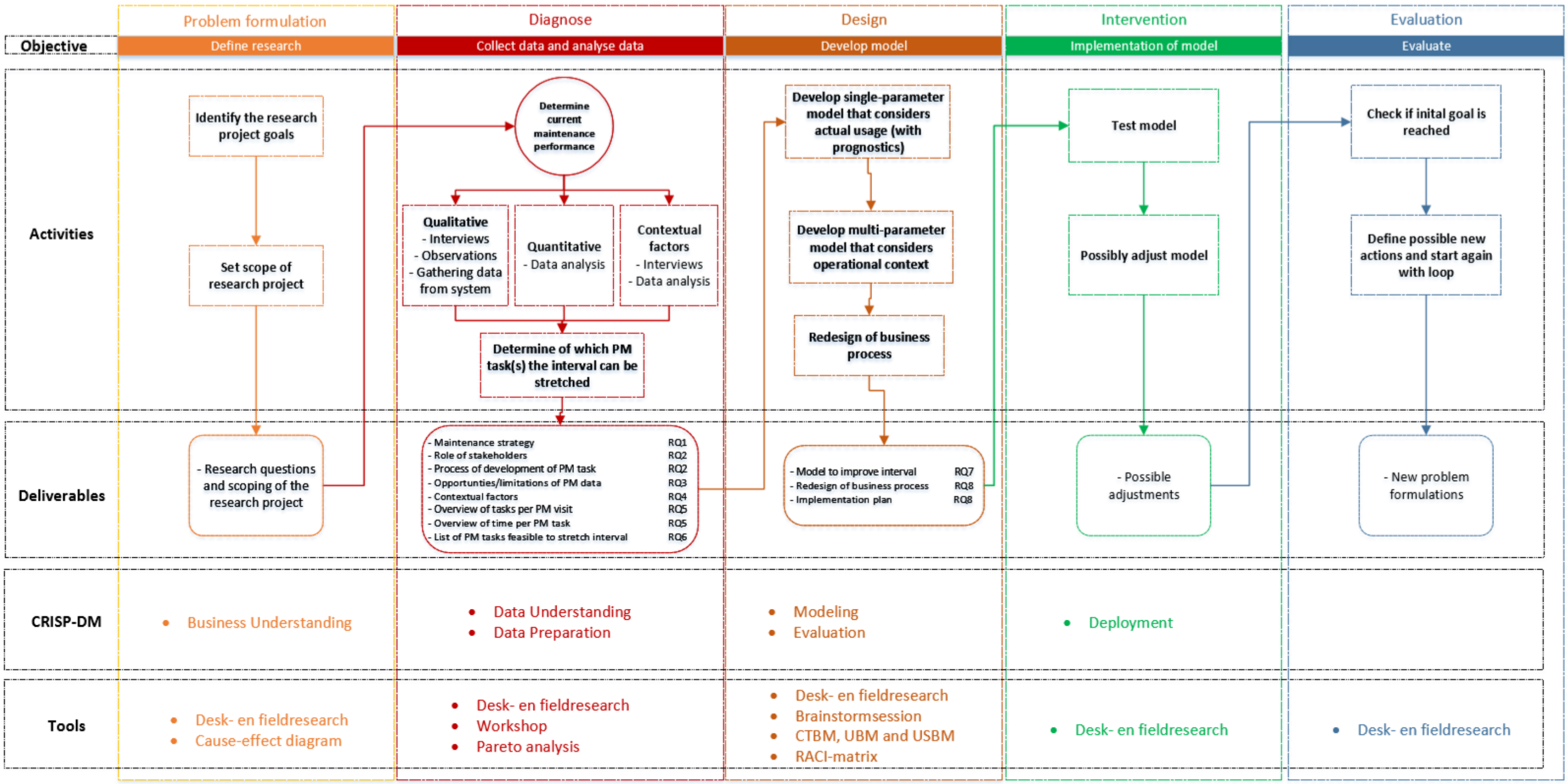


Figure 9: Research approach: regulative cycle and CRISP-DM.

3. Diagnosis: current maintenance policy at Philips

This chapter describes the current state of the PM policy of IXR-devices at Philips. Firstly, this section describes the maintenance strategy (RQ 1) and the relevant stakeholders within the process of developing PM tasks (RQ 2). Secondly, the possibilities and limitations of the available data on PM are shown (RQ 3).

3.1 Planned maintenance and corrective maintenance at Philips

This section describes the maintenance strategy (RQ 1) and the relevant stakeholders within the process of developing PM tasks (RQ 2). Next to that, this section describes the relation between PM and CM.

3.1.1 Stakeholders within the process of developing PM tasks

Philips strategy regarding maintenance is to minimize downtime of the medical device by performing PM. Philips offers service agreements to customers, depending on the service contract it includes PM and/or CM. According to Philips, PM means the following:

Planned Maintenance is used for the service request to deliver the scheduled maintenance activities which are contractually agreed upon in advance by a warranty or service contract, or purchased by customer (Lu, 2013).

A PM visit consists of multiple tasks. Figure 10 shows the process of how PM tasks are developed, but also the stakeholders involved within the process. The process starts with a change request. This request can be seen as a new sub-system or an update of a sub-system which is part of a device that has to be developed. With the introduction of a new device, the main part of the sub-systems remains the same. Thus, for such sub-systems there is no need to develop new PM tasks. R&D is responsible for the development aspect, where they use historical data of the devices with the goal to learn from experiences and to minimize CM and PM tasks associated with the change. After development of the updated/new sub-system, R&D checks the sub-system by creating a failure mode & effects analysis (FMEA). FMEA is a systematic procedure for identifying the potential failure modes and their causes and effects on product performance. The purpose of conducting a FMEA during the develop process is to:

- Ensure that the product design meets its requirements
- Address foreseeable issues regarding service and initial quality
- Avoid factory defects

The FMEA classifies all possible failures based on severity, occurrence and detection (see Appendix A). Multiplying the three factors gives a value. If this value is above a certain threshold then further action is required. Depending on the failure either no further action is required or the failure is classified as critical to quality (CTQ) or critical to safety (CTS). R&D makes sure that all those service specifications are defined per possible failure. A CTQ or CTS classified failure involves a defined action in order to prevent the failure. This action can be related to either production or maintenance.

Next to the FMEA, a risk management is executed by R&D. FMEA focusses on what sub-systems do have the potential to fail and their corresponding impact. The risk management focusses on the user risks that are associated with the usage of the device. Depending on the risk either no further action is required, or the failure is classified as CTQ or CTS into the service specifications of an updated/new sub-system.

Eventually, R&D has a list with all service specifications including the CTS and CTQ per updated/new sub-system. The next step is on the account of SI. SI translates the service specifications into PM tasks. For example, the service specifications note that for an updated/new sub-system an action should be performed on a specific interval. SI translates the action to be performed into tasks and creates a PM schedule. During this process, the earlier PM manual issues experienced by the FSEs are considered. After that, a review by R&D takes place dedicated to the defined PM tasks. In case of incomplete PM tasks, they have to be evaluated again by SI. If the PM tasks cover the service specifications, SI will be allowed to update the PM manual. When the PM manual is updated, it is launched to all service organizations of Philips. The service organizations check whether the updated PM manual is realistic. If not, SI will have to evaluate the new PM tasks again. It depends on the feedback of the service organizations whether SI has to adapt the PM tasks in collaboration with R&D or not. If there is no feedback from the service organizations, the updated PM manual will be introduced to the FSEs. The FSE executes PM according to the updated PM manual. If FSEs experience issues because of the PM manual it can be reported, such that SI can take those issues into account when creating PM tasks.

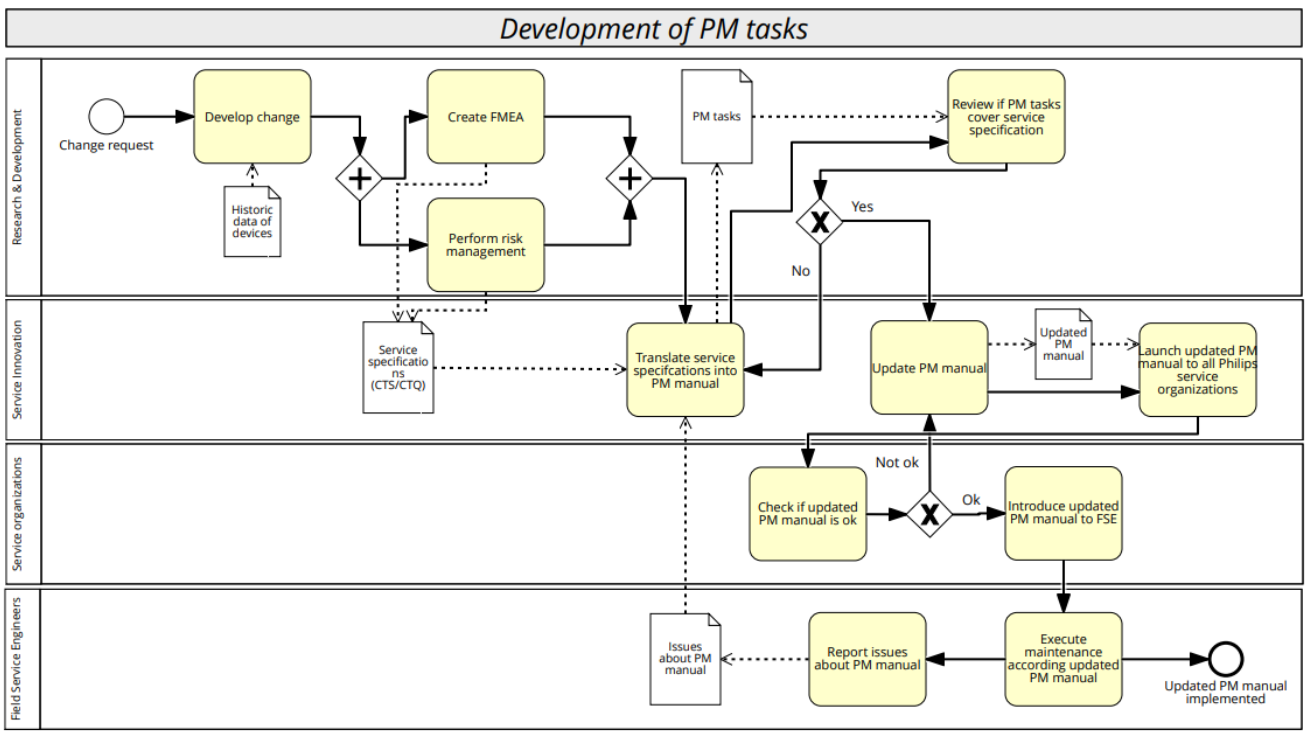


Figure 10: Current business process of developing PM tasks.

Corrective maintenance is determined as follows within Philips:

Corrective Maintenance is used for the service request to fix a problem of broken equipment (Customer, Demo); the fix may be delivered remotely with or without parts, onsite with or without parts, or delivered in a bench repair center (Lu, 2013).

A CM case is a result of device failure. A CM request can be initiated either by the customer, by a call, or via Philips. The latter creates CM cases remotely based on patterns in the log files of the device. CM cases can be solved either remotely by Philips, on site by an FSE, self-service by the customer or via a subcontractor. Priority codes are used in combination with the service level agreement to determine the urgency of a CM. The priority is based on a range from 1-5. It has to be noted that most CM cases are solved without part replacement. Such CM cases do not contain a material number but a piece of text that explains the solved case in the log file.

3.1.2 Interval of execution of PM task based on worst-case behavior of device

Figure 11 shows that the service life of a device differs depending on the contextual factors (usage and operational context). Both aspects influence the degradation of a sub-system and therefore the service life per sub-system. All PM tasks cover one or multiple sub-systems, which they prevent from failures. When R&D performs the FMEA, a physical-based model, including safety factors, is applied to determine at what threshold value of each failure parameter a failure is likely to occur. Based on the conservative threshold value of the failure parameter and the used (worst-case) behavior/load(s) of a device, the interval of a PM task is determined. For the used worst-case behavior/load(s) of a device, actual worst-case behavior/load(s) are not considered.

After execution of a PM task, the corresponding failure parameters of the sub-system start from their initial value again. For example, if failure parameter 1 is a certain number of clinical procedures (threshold value), the cumulative actual number of clinical procedures (load) will start from zero after execution of PM. This means, it is assumed that the state of sub-systems is 100% after execution of PM. We set this assumption for model purposes, knowing that this is not very realistic since, for example age, might play a role in degradation of a sub-system. Next to that, when performing the FMEA, it is assumed that other sub-systems do not influence the degradation of the sub-system that is considered. This might also not be realistic since degradation of other sub-systems might cause faster degradation to other sub-systems.

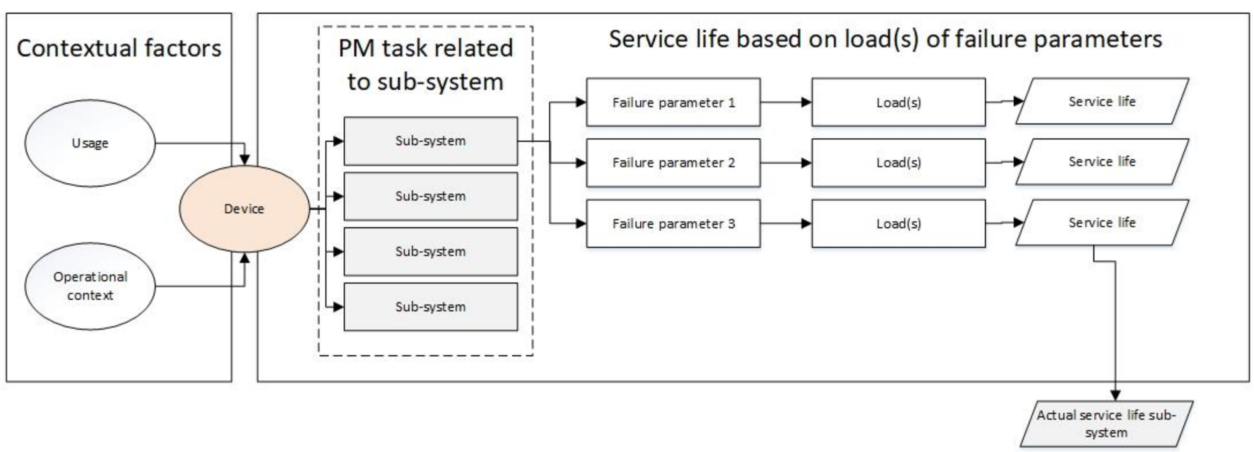


Figure 11: Relation between PM tasks and sub-systems.

A small analysis was executed to indicate the impact of not performing PM on time (Appendix B). Customers can decide whether they want a service contract that includes PM or not. If customers cooperate with Philips and choose to include the PM in the contract, a FSE from Philips will execute PM at least once per six months. It occurs that PM is not executed within six months. The small analysis indicated that a delay in performing PM, does not cause more failures. According to R&D, this is according to expectations, since safety factors are used to determine the threshold value of a failure parameter and intervals of PM are based on the worst-case behavior of a device.

This section showed the maintenance strategy aimed at maximizing the uptime of a device (RQ 1). It turned out that R&D and SI were the main stakeholders within the current process of developing PM tasks (RQ 2). Next to that, it was found that R&D uses physical-based models to determine the threshold values of failure parameters and to determine the interval of a PM. By doing this, safety factors are used when determining the threshold values. Next to that, worst-case behavior/load(s), without considering actual behavior/load(s), of a device is assumed.

3.2 Possibilities and limitations of PM data

Within the current PM policy much usage data is logged. Firstly, this section describes the available PM data. Secondly, the steps within data preparation are shown to prepare the data for the contextual factor analysis and for prognostic purposes. Next to that, this section describes the opportunities and limitations of the PM data (RQ 3).

3.2.1 Data understanding

In order to gain insight in PM data, it was necessary to understand the available data. It was found that the data could be divided in three parts: device data, maintenance data and customer data.

Device data

All IXR-devices worldwide contain a unique code called equipment number. Since Philips sells many different types of IXR-devices all equipment numbers are related to a system code. The system code contains information about the release and the configuration of the device. The installation date of a device at a customer is logged too, such that the age per device can be determined. Next to that, it can be seen how many clinical procedures and the type of clinical procedure were executed per day. Because of that, the number of workdays per year of a device can also be determined. A summary of the available device data is shown in Table 1.

Table 1: Device data (IXR).

Feature	Description
Equipment number	Unique number of a device.
System code	Code that determines release and configuration of device. Release: R8.1, R8.2 and Azurion. Configuration: monoplane and biplane.
Installation date	Date that device is installed at customer.
Number of clinical procedures per clinical procedure type per day	The number of clinical procedures executed on a device per day. The type of clinical procedures are: cardio, neuro, vascular, other and unknown.
Number of workdays	The total number of workdays for a period of time for a device.
Distance travelled by axes	Distance travelled by geometric axes of device in meters per day.

Maintenance data

Of all the devices that have a service contract at Philips, data regarding the maintenance is logged in Vertica. Per device the total number of PM cases and CM cases is logged. Also, the labor time spent to execute the PM/CM case is monitored (Table 2). The labor time per visit type is logged but it is not feasible to gain insight in the actual duration per PM task. During a PM visit, FSEs have to monitor whether the PM tasks is executed successfully (passed/failed). Unfortunately, based on practical experience it is known that the FSE repeats a PM task, possibly after making some intermediate adjustments, until the PM task is executed successfully. Therefore, the stored data only contains PM tasks that are passed. For example, if it appears that the verification values are out of range after calibration, the FSE will execute the calibration again until the verification measurement values are within the required range. Table 2 shows a summary of the relevant maintenance data.

Table 2: Maintenance data of IXR-devices.

Feature	Description
Number of PM cases	The number of PM cases per equipment number for a period of time.
Number of CM cases	The number of CM cases per equipment number for a period of time.
Labor time	The number of hours used per PM/CM visit. Labor time is divided into travel time, waiting time and working time.
Passed/failed per PM task per visit	The result of an executed PM task, result is either passed or failed.

Customer data

The last category of available data is customer related. Philips has many customers around the world. Additionally, the country and location of all devices are monitored. Also, the type of service agreement with the customer is contained in the log files. Table 3 shows a summary of the available customer data.

Table 3: Customer data of IXR-devices.

Feature	Description
Demographics	The country and name of hospital where device is located.
Type of service contract	The type of service contract, including/excluding PM/CM.

3.2.2 Data preparation

The last section has taught what PM data is available. This section contains the data preparation phase and is divided into multiple, but iterative, steps:

- Selecting data: deciding what data to include and exclude
- Integrating data: merge datasets
- Cleaning data: refer to actions that have been executed in order to clean the included data
- Constructing data: derive features and attributes
- Formatting data: convert formats

All relevant data was extracted from the Vertica database. Moreover, the selected data was merged and cleaned using Excel. After that, the remaining data was cleaned with the use of Python. With a clean dataset new features were added, eventually the dataset is formatted to Excel again in order to use it for model purposes (Figure 12).

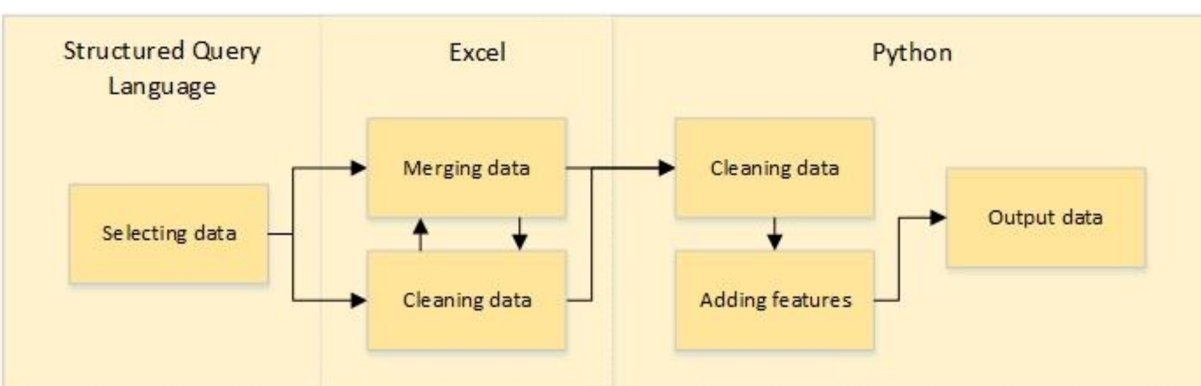


Figure 12: Steps of data preparation.

Selecting data
All relevant data was selected from the Vertica environment with the use of structured query language (SQL). Vertica contains of many data tables that are updated daily. Within this phase the relevant tables where joined based on equipment number of the device. Not all tables use the same feature name for the equipment number. This had to be taken into account when joining multiple tables. The relevant tables to include were those that contained contextual factors and monitored usage data for prognostic purposes. In total, five relevant tables were selected from Vertica:

Tables to indicate contextual factors:

1. Age per device per year because this might be a usage factor.
2. Number of PM/CM per device per year, including country and system code of device. Because CM is required to indicate the contextual factors that influence this value. From the system code the release and configuration of a device can be derived. These, and the country might be operational context factors.
3. Workdays per year per device, because this might be a usage factor.
4. Number of clinical procedures per type per device per year. Because this might be a usage factor.

Tables for model and prognostics purposes:

5. Number of clinical procedures per type per device per year/half a year.
6. Distance travelled by longitudinal and lateral axes in meters of a device per year/half a year.

Integrating data
Within this phase three datasets were conducted:

- Dataset 1: to indicate contextual factors that influence the number of failures (tables 1-4)
- Dataset 2: to model and predict the behavior of devices with regard to number of clinical procedures (table 5).
    - Dataset 2.1: data structured per year.
    - Dataset 2.2: data structured per half a year.
- Dataset 3: to model and predict the behavior of devices with regard to distances travelled by longitudinal and lateral axes in meters (table 6).
    - Dataset 2.1: data structured per year.
    - Dataset 2.2: data structured per half a year.

During this phase the six tables selected from Vertica were converted to Excel and Python. Within Excel and Python cleaning and merging the data were iterative processes. For Dataset 1, the equipment number and year were concatenated in all tables in order to create a new unique code. Since all the selected tables 1-4 show data per year of each equipment number, it was easy to merge all the data now based on the unique code (Figure 13). Table 2 was used as main table since that table contained the number failures (CM) per year, the remaining tables were merged to table 2. It was not required to merge table 5 or table 6. Within Figure 13 it can be seen how many rows each table consisted of. After integrating, some data cleaning steps took place, the exact steps of data cleaning are described in next section.

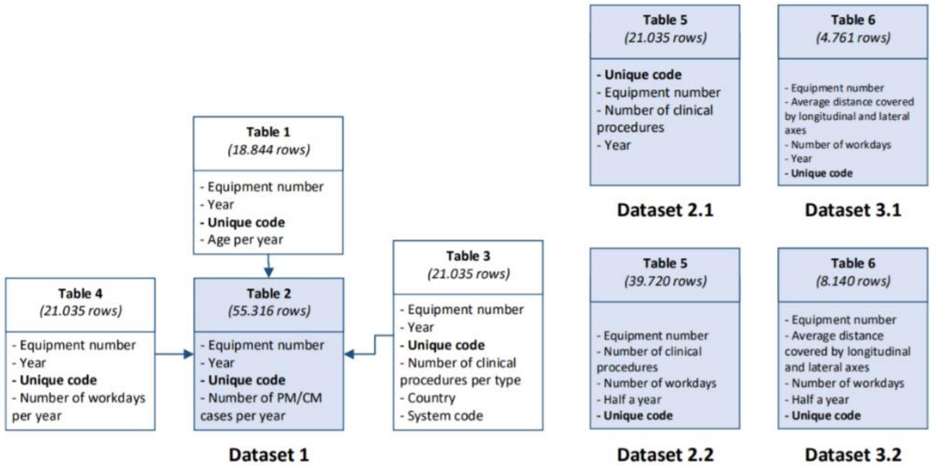


Figure 13: Process of integrating data.

Data cleaning

Data cleaning took place in multiple phases of the data preparation phase. Table 4 shows the steps that were taken in order to clean the data. It has to be considered that the data derived from Vertica has already been cleaned before it got stored in the database.

Table 4: Process of data cleaning.

Data cleaning step	Explanation	Cleaning took place in	Applied to table/dataset
1. Remove all rows that contain data before 2012.	Data before 2012 is seen as outdated data that does not reflect the current situation.	Vertica by SQL	All tables
2. Only include the relevant system codes.	The relevant system codes are the ones that are part of the releases R8.1, R8.2 or Azurion, they are scoped upon for the research.	Vertica by SQL	All tables
3. Remove all rows that contain zero CM in a year.	When zero CM cases took place in a year, it means that CM maintenance is executed by a different party.	Excel	Dataset 1
4. Remove all devices that never had any PM.	In order to only include the devices that have PM in contract, all devices that never had any PM visit were removed.	Excel	Dataset 1
5. Remove row if row contains a ‘nan’ value.	In order to get a clean dataset with no ‘nan’ values. Except for age since a feature is added to indicate if a device is at least one year old (next section).	Python	Dataset 1
6. Include devices (rows) that have at least twelve months of historical data.	For simulation purposes it is important to only add new devices when they have at least twelve months of historical data. So predictions can be made and the device enters directly with a PM task execution for CTBM.	Python	Dataset 2.2 and 3.2
7. Remove devices (rows) that do not have data from their start point to the first half of 2019.	Devices can be added within time of the simulation. But once in, the device cannot leave the dataset since it might reflect a skewed situation where no maintenance is needed because devices leave the dataset after a period of time.	Python	Dataset 2.2 and 3.2

8. At least ten working days monitored.	Of some rows the cumulative distance per year/half a year is based on less than ten workdays. According SMEs of research department the distance travelled by the axes is not yet measured accurately every day. Therefore, an average per workday needs to be define to estimate the usage per year/half a year. Because of this, it was decided that a row is included if it has at least monitored data of ten workdays.	Python	Dataset 3.1 and 3.2
9. Only include rows with age of devices > 0	Since not all install dates were known the datasets contained data points without age. One goal of the age is to filter the year of install (age = 0) in order to get an accurate view of the remaining features. If one device has no install date but contains data points of at least two years, then it is known that the age of that device is not 0 in the second year.	Python	Dataset 1, 2 and 3
10. Remove all rows that is not in top 5 country: US, JP, DE, FR or GB	In order to apply the operational context 'country', it is chosen to consider the top 5 counties that consists of the most data points.	Python	Dataset 2.1 and 3.1

Constructing data

After the iterative processes of merging and cleaning the data three datasets remained. Table 5 shows the added features to the datasets.

Table 5: Added features to datasets.

Added feature	**Explanation**	**Applied to dataset**
Release	Based on the system code per device the release of a device was derived: R8.1, R8.2 or Azurion. Because release might be an operational context factor.	Dataset 1
Configuration	Based on the system code per device the configuration of a device was derived: monoplane or biplane. Because configuration might be an operational context factor.	Dataset 1
Distance travelled by axes per year/half a year	With regard to geometry of the longitudinal and lateral axes, the average distance per workday is multiplied by 272 (per year) or 136 days (per half a year). It was decided to do this, since this data is not logged consequently daily. By doing this, a model to decide whether to execute PM was developed.	Dataset 3.1 and 3.2
Average number usage per workday	In order to prognose the usage for next period, the average number of usage per workday was computed.	Dataset 2.2 and 3.2

3.2.3 Data overview

After data preparation, the final datasets were determined. Table 6, Table 7 and Table 8 show snapshots of how each dataset looks like. Appendix C shows the descriptive statistics of the features.

Table 6: Dataset 1.

Dataset 1													
Feature (rows)	Equipment number (19.682)	Configuration (19.682)	Year (19.682)	Number of workdays (19.682)	Number of clinical procedures (19.682)	Country (19.682)	Release (19.682)	Number of PM cases (19.682)	Number of CM cases (19.682)	Age (15.768)	Neuro procedures (19.682)	Vascular procedures (19.682)	Cardio procedures (19.682)
Unique code *(equipment number + year)*	XXXXXX	Monoplane/Biplane	2012-2019	1-365	1-6088	All countries	R8.1/R8.2/Azurion	0-27	1-51	1-22	0-2679	0-5885	0-5115

Table 7: Dataset 2.

Dataset 2													
Dataset 2.1	*Feature (rows)*	2014 (774)		2015 (1.379)		2016 (1.851)		2017 (2.318)		2018 (2.607)		2019 (2.709)	
Dataset 2.2	*Feature (rows)*	1st half of 2014 (1.234)	2nd half of 2014 (1.234)	1st half of 2015 (1.234)	2nd half of 2015 (1.449)	1st half of 2016 (1.740)	2nd half of 2016 (2.068)	1st half of 2017 (2.567)	2nd half of 2017 (3.162)	1st half of 2018 (3.517)	2nd half of 2018 (3.665)	1st half of 2019 (3.747)	2nd half of 2019 (3.795)
Dataset 2.2	*Equipment number*	Number of workdays per half a year											
Dataset 2.2	*Equipment number*	Average number of clinical procedures per workday in half a year											
Dataset 2.1 and 2.2	*Equipment number*	Number of clinical procedures per year / half a year											

Table 8: Dataset 3.

Dataset 3													
Dataset 3.1	*Feature (rows)*	2014 (171)		2015 (275)		2016 (454)		2017 (599)		2018 (796)		2019 (870)	
Dataset 3.2	*Feature (rows)*	1st half of 2014 (157)	2nd half of 2014 (157)	1st half of 2015 (157)	2nd half of 2015 (208)	1st half of 2016 (279)	2nd half of 2016 (366)	1st half of 2017 (466)	2nd half of 2017 (645)	1st half of 2018 (779)	2nd half of 2018 (886)	1st half of 2019 (1.040)	2nd half of 2019 (1.078)
Dataset 3.2	*Equipment number*	Average number of travelled distance in meters of per workday in half a year											
Dataset 3.1 and 3.2	*Equipment number*	Travelled distance in meters per year / half a year											

This section showed that both effectivity and efficiency of PM cannot be evaluated at this moment. Nevertheless, data regarding usage and operational factors is available to test what contextual factors play a role in degradation of a device since the number of failures per year is monitored (Dataset 1). Next to that, Dataset 2 and Dataset 3 are suitable to simulate maintenance policies including prognostics when the number of clinical procedures or the travelled distance of the longitudinal/lateral axis in meters are used as failure parameters of a sub-system (RQ 3). Later on, the mentioned datasets will be used and referred to as Dataset 1, Dataset 2.1/2.2 or Dataset 3.1/3.2.

To conclude, this chapter showed the maintenance strategy aimed at maximizing the uptime of a device (RQ 1). It turned out that R&D and SI were the main stakeholders within the current process of developing PM tasks (RQ 2). Next to that, it was found that R&D uses physical-based models to determine the threshold values of failure parameters and to determine the interval of a PM. By doing this, safety factors are used when determining the threshold values. Next to that, worst-case behavior of a device is assumed without considering the actual usage of a device.

Finally, Section 3.2 showed the available PM data. It turned out that there is no useful data to challenge the effectivity and efficiency of PM. Nevertheless, three datasets were acquired that will be used in next chapters to indicate the role of contextual factors on degradation of a device (Dataset 1), and to simulate maintenance policies within the model including prognostics (Dataset 2 and Dataset 3).

4. Diagnosis: contextual factors and bottlenecks within PM

The last chapter described the current PM process and the available PM data. Firstly, this chapter describes the role of contextual factors to degradation of a device (RQ 4). After that, this chapter describes the bottlenecks within this PM policy (RQ 5) and indicates of which intervals of PM tasks could be stretched (RQ 6). Since this chapter identifies the PM tasks that cause the main PM labor time within a maintenance cycle (bottlenecks) and the role of contextual factors, this chapter includes the report of the quantitative and qualitative analysis (first deliverable).

4.1 Contextual factors that influence the number of failures of a device

This section describes the contextual factors (RQ 4). Firstly, this section describes the initial analysis in order to identify relevant contextual factors. Secondly, hypotheses are formulated to test whether the found contextual factors influence degradation of a device. Finally, the formulated hypotheses are tested.

4.1.1 Initial analysis of contextual factors

In order to determine the relevant contextual factors that influence the degradation of an IXR-device, we investigated the contextual factors from three perspective: initial (quick) data analysis, literature and interviews (Table 9). An extended analysis from all perspectives is included in Appendix D. From Section 3.2 it turned out that there is no data related to both contextual factors environment and educational level of clinical users. Literature on contextual factors shows that all contextual factors in Table 9 are relevant since the named papers conclude that the contextual factor might influence degradation. Interviews with SMEs of R&D show that all contextual factors in Table 9 are relevant except for the educational level of clinical users. Table 9 shows that the following contextual factors are supported from all three perspectives: the number of clinical procedures, number of workdays, age, type of clinical procedure, culture/country, configuration and release.

Table 9: Relevant contextual factors.

Contextual factor	Category	Data available	Literature	Interviews
Number of clinical procedures	Usage	Yes	(Müller, Staudacher, Friedl, Köhler, & Weißschuh, 2010; Carl-Anders Johansson, 2014; Galar, Thadari, Catelani, & Ciani, 2015)	Yes
Number of workdays	Usage	Yes	(Müller, Staudacher, Friedl, Köhler, & Weißschuh, 2010; Carl-Anders Johansson, 2014; Galar, Thadari, Catelani, & Ciani, 2015)	Yes
Environmental	Operational context	No	(Müller, Staudacher, Friedl, Köhler, & Weißschuh, 2010)	Yes
Age	Usage	Yes	(Peng & Dong, 2010).	Yes
Type of clinical procedure	Usage	Yes	(Tinga, Wubben, Tiddens, Wortmann, & Gaalman, 2019)	Yes
Culture/country	Operational context	Yes	(Rosemann, Recker, & Flender, 2008)	Yes
Education level of clinical user	Operational context	No	(Rosemann, Recker, & Flender, 2008)	No
Configuration	Operational context	Yes	(Moubray, 1999)	Yes
Release	Operational context	Yes	(Moubray, 1999)	Yes

4.1.2 Formulate hypotheses

In this section hypotheses were formulated to check whether the contextual factors are significantly influencing the number of failures (CM cases) per year. Based on the initial analysis conducted in the previous section, it is expected that most of the contextual factors are related to the degradation of a device. The contextual factors: environment and educational level of the clinical user, do not have data monitored. Therefore, it is not possible to test the hypotheses. The remaining contextual factors can be divided into different types: discrete (usage) and categorical data (operational context).

Since the goal of this part of the research is to find out if these contextual factors actually influence the number of failures per year, multiple hypotheses were determined. The dependent variable is the number of failures per year, which is discrete and not normally distributed (Appendix F). Therefore, non-parametric methods were used. Depending on the independent variable, a statistical method was used to determine if there is a significant relation between both variables. Table 10 shows the most frequently used statistical methods to determine the degree of significance between two or more variables.

Table 10: Common statistical methods (Zulfigar & Bhaskar, 2018).

Scenario	**Dependent variable**	**Independent variable**	**Parametric**	**Non-parametric**
Mean of two independent groups	Discrete	Categorical/nominal	Independent t-test	Mann-Whitney U test
Mean of 2 paired samples	Discrete	Time variable (time before & after, t0 & t1)	Paired t-test	Kruskal-Wallis (Bonferroni correction)
Mean of 3+ independent groups	Discrete	Categorical/nominal	One-way ANOVA	Kruskal-Wallis (Bonferroni correction)
3+ measurement on the same subject	Discrete	Time variable	Repeated measures ANOVA	Friedman Test
Relationship between 2 continuous variables	Discrete	Continuous/scale	Pearson's correlation coefficient	Spearman's correlation coefficient
Predicting the value of one variable from other	Discrete	Any	Linear regression	Kernel test
Survival distribution	Time to failure	Any, often time	Weibull life	Kaplan-Meier
Relationship between two categorical variables	Categorical/nominal	Categorical/nominal	-	Chi-squared test

In order to test the relation between contextual factors and the number of failures, the following null-hypotheses have been formulated:

For discrete variables:

Null-hypothesis: contextual factor X is unrelated to the number of failures (CM cases) of a device with p-value <0.01.

For categorical variables:

Null-hypothesis: there is no difference in the number of failures (CM cases) between categories of contextual factor X with p-value <0.01.

In case of a rejected null-hypothesis, the sample would give reasonable evidence to support the alternative hypothesis. This does not automatically imply a relevant or important difference in general (Zulfigar & Bhaskar, 2018). In our case it is considered highly relevant because the contextual factors are based on the initial analysis conducted in Section 4.1.1.

4.1.3 Test hypotheses

In this section the hypotheses described in Section 4.1.2 were tested (Table 11). In order to test the hypotheses, Dataset 1 (Section 3.2.3) was used. The contextual factors 'age' and 'clinical procedure: Unknown' are contextual factors that were unrelated to the number of failures. Table 11 shows the correlation value of the usage factors. 'Number of clinical procedures', 'number of workdays' and the 'clinical procedure: Neuro', have a small (positive) correlation with the mean number of failures (CM cases) per year.

Table 11: Hypotheses of contextual factors tested.

Null-hypothesis: Discrete variable (usage) (Spearman's correlation)				**Null-hypothesis: Categorical variable (operational context)** (Kruskal-Wallis, with Bonferroni correction)		
X	P<0.01	H0 rejected?	Correlation	X	P<0.01	H0 rejected?
Number of clinical procedures	✓	Yes	0.226	Country/Culture	✓	Yes
Age	✓	No	0.023	Configuration	✓	Yes
Number of workdays	✓	Yes	0.226	Release	✓	Yes
Clinical procedure: Cardio	✓	Yes	0.101			
Clinical procedure: Neuro	✓	Yes	0.207			
Clinical procedure: Vascular	✓	Yes	0.098			
Clinical procedure: Other	✓	Yes	0.067			
Clinical procedure: Unknown	✓	No	0.039			

The remaining usage factors contain a weaker correlation, below 0.2. Because of the weak correlation of the discrete variables, there seems no clear linear relation between the variables.

All null-hypotheses for the categorical variable were rejected, this means that there are significant differences between the mean number of failures per category within each operational context. Appendix E contains the Kurskal-Wallis and Bonferroni test of the categorical variables. With this test it can be seen whether there is a significant difference between the categories of an operational context factor.

It is important to notice that this section describes the contextual factors that influence the mean number of failures on system-level in general. This means, this analysis provides merely a general overview of contextual factors that do play a role in degradation of a device (all sub-systems).

This section showed the initial found contextual factors from expert, literature and data perspective. After that, it was tested what contextual factors definitely influence the mean number of failures of a device. It was concluded that all usage and operational context features do influence the mean number of failures on a device, except for usage features 'Age' and 'Clinical procedure: Unknown'. Therefore, this section has answered RQ 4. The operational context factors defined within this section, will be used in Chapter 6 in order to apply the multi-parameter model.

4.2 Bottlenecks within PM

This section describes the tasks to execute within PM and the planned time per PM task (RQ 5). Because of that, it turns out what the bottlenecks of PM are. Also, this section combines the qualitative and quantitative knowledge gained in order to define of which PM tasks the interval can possibly be stretched (RQ 6).

The current PM process consist of a policy that incorporates two visits per year. Since one cycle consists of two years, four (slightly) different visit types per cycle exist. The maintenance cycle starts again after four visits. The PM tasks that have to be executed by the FSEs depend on the type of visit (Figure 14).

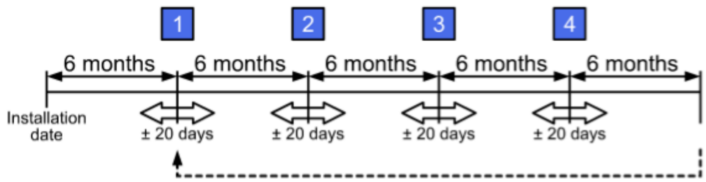


Figure 14: Maintenance cycle.

The total planned execution time for PM tasks during each visit is shown in the PM manual. One PM manual is written for every release. For the releases R8.1, R8.2 and Azurion, the PM tasks in the manual are similar. A release is divided into two configurations namely: monoplane or biplane, there are specific PM tasks that only apply to monoplanes or biplanes.

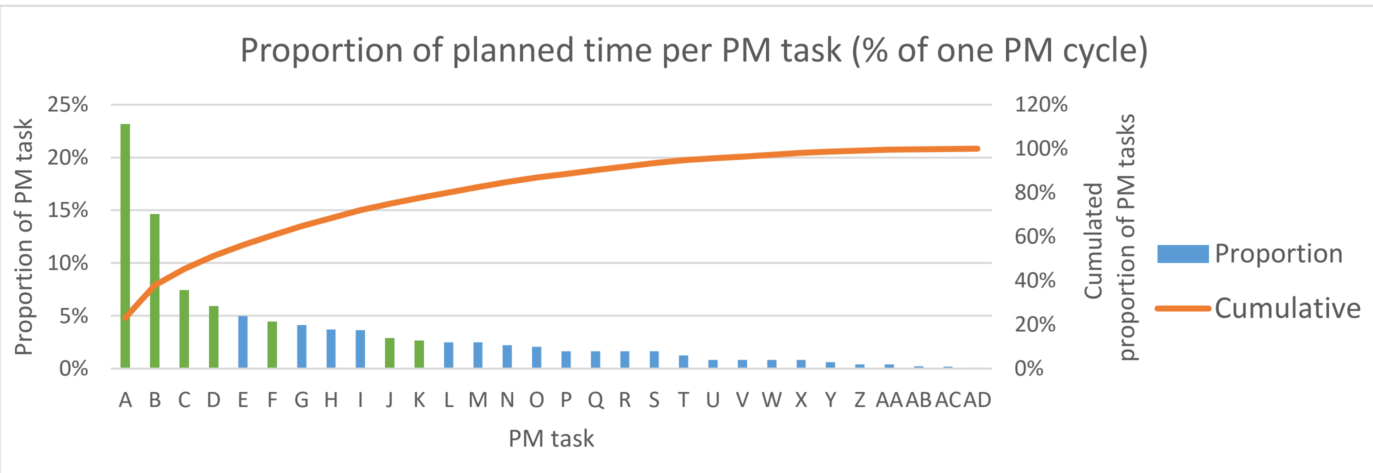


Figure 15: Pareto analysis of planned time per PM task within a maintenance cycle.

Based on the installed devices, the proportion of time per PM task within one maintenance cycle was computed (Figure 15). This plot excludes travel time and off-site preparations. The full list of PM tasks to perform consists of 30 activities. In order to scope the research, a Pareto analysis (Stojcetovic, Živče, Lazarević, & Marjanović, 2015) was conducted. Based on the Pareto it can be seen that twelve PM tasks (A-L) are responsible for 80% of the total PM time of one maintenance cycle. Some of those PM tasks are mandatory because of safety reasons set up by the FDA (Amant, 2017) and therefore not included in the scope of this research. The PM tasks within this scope can be categorized as follows: calibration/verification, handing over, safety related, cleaning/checking and preparation tasks. By consulting the product owner(s) of each PM task related sub-system(s), the potential of stretching the interval of a PM task with the use of usage data was indicated. Because of mandatory PM tasks this research focusses on the following PM tasks (marked green in Figure 15 and Table 12):

Table 12: PM task within scope of research.

Planned maintenance task	Type	Product owner(s)	Potential to stretch interval with usage data
A	Calibration/verification	R&D	Yes
B	Calibration/verification	R&D	Yes
C	Creating a backup	Customer Service (CS)	Ye
D	Calibration/verification	R&D	Yes
E	Handing over	CS	No, mandatory task
F	Calibration/verification	R&D	Yes
G	Safety related	R&D	No, mandatory task
H	Safety related	R&D	No, mandatory task
I	Safety related	R&D	No, mandatory task
J	Cleaning/checking	R&D	Yes
K	Cleaning/checking	R&D	Yes
L	Preparations	CS	No, mandatory task

This section showed the planned time of the categorized PM tasks that take place within a maintenance cycle. Besides that, it turned out what the main bottlenecks of PM are (RQ 5). This section also indicated whether the interval of execution of PM tasks can possibly be stretched with the use of usage data by consulting the product owner(s) (RQ 6).

To conclude, this chapter contained the first deliverable: report of qualitative and quantitative analysis that shows the current PM performance from both perspectives. The analysis showed the proportion of time of a PM task within a maintenance cycle in order to focus on the main bottlenecks PM task A-L (RQ 5) (quantitative). Next to that, it turned out that the interval of execution of PM tasks: A, B, C, D, F, J and K, can possibly be stretched (RQ 6) (qualitative). Besides that, the role of the contextual factors 'usage' and 'operational context' that influence the number of failures was explained (RQ 4).

5. Design: feasible maintenance policies for IXR-devices

The next step in the regulative cycle is the design phase, which combines previously gained knowledge on PM sourcing from experience, interviews and data analysis. Firstly, this section describes what maintenance policies are feasible for IXR-devices. Secondly, this section describes the application of the proposed model while considering contextual factors.

5.1 Maintenance policies feasible for IXR-devices

The feasibility of different maintenance policies at Philips, derived from literature, were considered within the context of IXR-devices. This was based on the available usage data at Philips, the contextual factors on system-level and knowledge sourcing from the SMEs of R&D.

As mentioned in Section 3.1, the interval of a PM task is based on an FMEA. Based on interviews with SMEs of R&D, it was found that intervals of a PM task are rather conservative. Within the service specifications of the FMEA it can be seen what failure parameter(s) the interval per PM task is based on. The maintenance policies mentioned in Section 1.2 were indicated as possible policies that could improve the current periodic PM policy. Considering these policies, it was concluded that, depending on the details of the specifications, application of CTBM, UBM or USBM is feasible for IXR-devices. Potentially CBM is the best maintenance policy to improve the current periodic PM policy at Philips. Nevertheless, this maintenance policy is excluded for this research since CBM requires direct monitoring of health indicators by sensors not installed yet. Therefore, this research was focused on the application of CTBM, UBM and USBM.

Some differences between literature and practical application at Philips had to be taken into account when applying the proposed maintenance policies. The acquired papers about different maintenance policies do mainly consider manufacturing areas. There, the need for safety factors is lower since a failure of a device does not endanger humans directly. These safety factors are reflected in the FMEA service specifications that are used to determine the interval of execution of PM tasks. Next to that, the acquired papers use either experience-based models or physical-based model. When R&D performs the FMEA, a physical-based model is applied to determine at what threshold value of each failure parameter a failure is likely to occur at a sub-system. The physical-based model applied during FMEA is considered accurate in predicting failures according to SMEs of R&D. This means, for this research the FMEA will be used as a starting point. So, for this research it is not necessary to determine the threshold value of a failure parameter again by applying an experience-based model or physical-based model.

In order to reduce variation caused by usage, all three maintenance policies take into account contextual factor ‘usage’. In fact, CTBM considers actual worst-case usage behavior of the whole fleet and takes the devices that behave worst as a reference. The interval of execution of a PM task for the whole fleet is similar and based on the actual worst-case behavior. At this moment, the interval of execution of PM tasks is determined by assumed worst-case behavior without considering actual behavior of the devices. The actual worst-case behavior can be compared with the threshold value of a failure parameter that is fixed for a time period. Depending on the threshold value of the failure parameter, the interval of execution of a PM task can possibly be stretched.

Looking at UBM, contextual factor usage is considered on individual level of a device, such that each device is maintained on time by monitoring the failure parameter of each device. This policy needs to be applied in a proactive way, by using prognostics, in order to decide whether execution of a PM task has to be performed.

USBM is similar to the UBM policy. The only difference is that USBM takes the severity per type of usage into account as explained in Section 1.2.. Prognostics have to be done for each type of usage.

All three maintenance policies consider contextual factor ‘usage’ in order to improve accuracy in determining when PM has to be executed. Nevertheless, none of the mentioned maintenance policies consider the operational context (country/culture, room of device, type of device, specifications of device), even though they might influence the usage of a device or degradation of a sub-system.

Therefore, in order reduce variation in usage or degradation of a sub-system caused by operational context factors, this research considers the operational context within the proposed maintenance policies. Applying operational context factors to the mentioned maintenance policies is relevant from both scientific and business perspective. Because, to the best of our knowledge, considering context-awareness to the proposed maintenance policies has not been tested before in the field of medical devices (or related fields). Next to that, Philips is mainly interested in improving the PM policy by using data-driven maintenance policies, considering context-awareness is seen as a contribution to this since it might even further improve the PM policy.

This section described the maintenance policies that are feasible for application at IXR-devices based on the available data and knowledge of SMEs of R&D. CTBM, UBM and USBM consider the actual usage as a contextual factor in order to tune the interval of execution of maintenance. Within these policies the operational context, that might cause variation in usage or in degradation of a sub-system, is not considered. Therefore, this research attempts to include the operational context when tuning the interval of execution of maintenance. Next section describes how the mentioned maintenance policies could be applied to IXR-devices in more detail.

5.2 Application of model to IXR-devices

This section describes the theoretical application of the proposed model where CTBM, UBM and USBM for IXR-devices were applied in order to improve the PM policy (RQ 7). To what extent a maintenance policy is applicable depends on the available data within the FMEA report. The first policy, CTBM, is relatively straightforward and considers one failure parameter (contextual factor ‘usage’) that determines the fixed interval of the whole fleet. The second policy, UBM, considers one failure parameter (contextual factor ‘usage’), but treats devices individually by using the actual usage of a monitored device. The third policy assumes that severity among different types of usage is known. Finally, this section explains how including contextual factor ‘operational context’ could increase the expected lifetime of a sub-system and therefore the interval of execution of a PM task.

5.2.1 Calendar-Time-Based Maintenance

As mentioned before, R&D defines an interval of execution of a PM task within the FMEA based on conservative behavior/load(s) of the device. The failure parameter(s) and their threshold values that determine the interval of execution of a PM task should be derived from the FMEA if possible. Since the failure parameter(s) applied at this level are sub-system specific, this CTBM policy should be executed for every PM task that is related to a sub-system. Application of CTBM is feasible when actual values of the failure parameter(s) are roughly known. It is not required for those failure parameter(s) to be monitored on a daily basis since CTBM attempts to define an updated (fixed) interval of execution of a PM task based on the devices that actually show worst-case behavior. For each failure parameter the threshold value needs to be compared to the new threshold value determined by the actual worst-case behavior. Based on the actual worst-case behavior and the acceptable level of exceeding the threshold value, a new threshold value can be computed in order to determine a new CTBM interval for the whole fleet. An acceptable exceeding rate might be larger if the sub-system is not considered critical. The corresponding equation is shown in Equation 1.

Equation 1: Equation to determine interval CTBM.

Equation to determine interval of execution of a PM task with CTBM

$$F_{sub-system} = \{f_0, \ldots, f_n\}, \qquad f = failure\ parameter$$

$$CTBM\ interval_{PM\ task} = \min_{f \in F_{sub-system}} \left(\frac{\text{Threshold value}_f}{\text{New threshold value}_f} * current\ interval_{PM\ task}\right) \tag{1.1}$$

$$New\ threshold\ value_f = \mu_f + i * \sigma_f,\ \text{i} \in [0,5] \tag{1.2}$$

$$\mu_f = mean\ value\ of\ failure\ parameter$$

$$\sigma_f = standard\ deviation\ of\ failure\ parameter$$

$$i = trade - off\ value\ to\ increase\ or\ decrease\ exceeding\ rate\ and\ CTBM\ interval$$

$$i = \frac{New\ threshold\ value_f - \mu_f}{\sigma_f} \quad \textit{(Derived from 1.2)} \tag{1.3}$$

$$Exceeding\ rate_f = \frac{Number\ of\ cases\ that\ exceed\ new\ threshold\ value_f}{Total\ number\ of\ cases_f} \tag{1.4}$$

5.2.2 Usage-Based Maintenance

UBM considers devices on an individual level and without differentiations between severity of the failure parameter(s). UBM is applicable if the actual behavior of each device is monitored on a daily basis.

To decide whether a certain PM task requires to be performed during a PM visit, this part of the design phase requires prognostics of the monitored failure parameter(s) for the period until the next PM visit. At this moment, one maintenance cycle of two years consists of four visits, such that every six months a visit is scheduled. However, the proposed model is also applicable for less visits per maintenance cycle, as long as the next PM visit date is roughly known. Obviously, prognostics will be less accurate if time between two visits increases. The value of a failure parameter from the last time a certain PM task was performed, adds the predicted value for the next period. If the total value does not exceed the threshold value of the failure parameter that is considered, a PM task will have the possibility to be postponed to the next visit.

As an example, Figure 16 shows that at t_c a prognosis was used in order to determine if execution of a PM task could be postponed to the next PM visit. The last time this PM task was executed is t_l. In order to determine whether a PM task can be postponed to the next PM visit, these prognostics were executed before every PM visit. When applying prognostics on it, a PM task will be more dynamic and execution of this task is only performed when necessary. Equation 2 shows the corresponding equation.

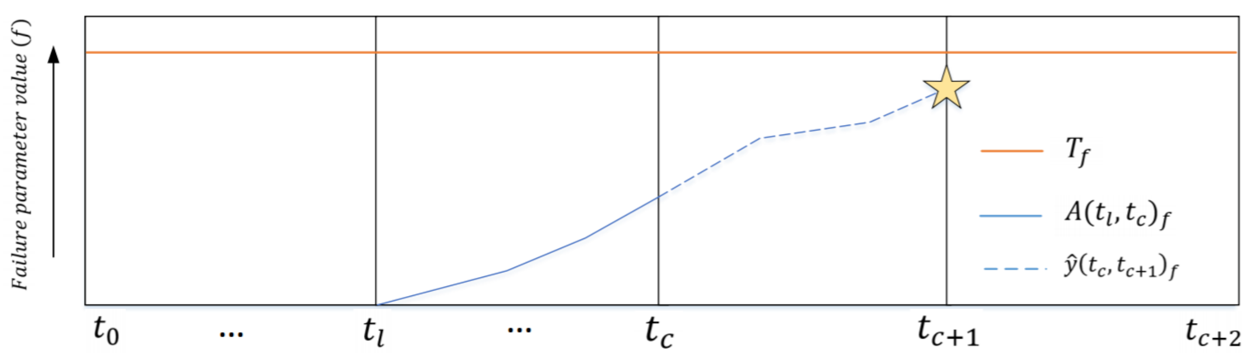


Figure 16: Prognostics of a failure parameter.

Equation 2: Equation to decide whether PM task has to be executed.

Equation for deciding whether PM task has to be executed

$f = failure\ parameter$

For all $f \in F_{sub-system}$ with $F_{sub-system} = \{f_0, \ldots, f_n\}$

$T_f = threshold\ value\ of\ failure\ parameter\ (constant\ value)$

$t_0, t_1, t_2, t_3 \ldots denote\ the\ successive\ PM\ visits$

$t_l = last\ PM\ visit\ for\ which\ task\ is\ executed$

$t_c = Current\ PM\ visit, so\ decide\ whether\ PM\ task\ needs\ to\ be\ executed$

$A(t_l, t_c)_f = value\ of\ failure\ parameter\ since\ last\ time\ of\ execution\ PM\ task$

$\hat{y}(t_c, t_{c+1})_f = prediction\ value\ of\ failure\ parameter$

At t_c:

Postpone execution of PM task for each $F_{sub-system} = \{f_0, \ldots, f_n\}$ *we have*

$$A(t_l, t_c)_f + \hat{y}(t_c, t_{c+1})_f < T_f \tag{2.1}$$

If not: execute PM task

It is important to notice that some sub-systems might degrade based on multiple failure parameters. In this case UBM uses all failure parameters and a PM task will be executed when one of them exceeds the threshold value.

Equation 2 shows how it is decided whether a PM task has to be executed. Within this equation, prognostics are required in order to determine $\hat{y}(t_c, t_{c+1})_f$. According to SMEs of R&D, the usage behavior of a device does not change substantially over time. Therefore, it was decided that single linear regression will be applied to determine $\hat{y}(t_c, t_{c+1})_f$.

In order to determine the optimal parameter values for prognostics, the mean absolute percentage error (MAPE) was used. This is done because this regression error is most intuitive, applies absolute errors and is robust to outliers (Mir & Ahmad, 2018).

It was considered unnecessary to compute new parameter values at each moment of t_c. Therefore, the parameter values (β_0 and β_1) that correspond to the minimal MAPE value when predicting the usage of the 1st half of 2019 with the data of the 2nd half of 2018, were used at each time of performing prognostics (t_c). These periods of time were chosen since they contain the most data points within Dataset 2.2 and 3.2. Equation 3 shows how prognostics are performed during simulation of the UBM approach.

Equation 3: Equation to perform prognostics.

Equation to perform prognostics for all $\boldsymbol{f} \in \boldsymbol{F_{sub-system}}$ with $\boldsymbol{F_{sub-system}} = \{\boldsymbol{f_0}, \dots, \boldsymbol{f_n}\}$

D = number of days until next PM visit = maximal number of workdays in period of six months based on history of device (if not reliable D = 136)

$X_i(k) = average\ value\ of\ failure\ parameter\ per\ workday\ in\ period\ [t_{i-1}, t_i)\ for\ device\ k$

$\hat{y}_{workday_{i+1}}(k) = prediction\ of\ failure\ parameter\ value\ per\ workday\ in\ period\ [t_i, t_{i+1})\ for\ device\ k$

$\hat{y}_{i+1}(k) = prediction\ of\ failure\ parameter\ value\ in\ period\ [t_i, t_{i+1})\ for\ device\ k$

$N(c) = the\ set\ of\ all\ devices\ that\ are\ active\ both\ at\ t_{c-1}\ and\ at\ t_{c+1}$

$$MAPE(c) = \frac{1}{|N(c)|} * \sum_{k \in N(c)} \left| \frac{y_{workday_{i+1}}(k) - \hat{y}_{workday_{i+1}}(k)}{y_{workday_{i+1}}(k)} \right| \tag{3.1}$$

$\sigma_{MAPE(c)} = standarddeviation\ of\ MAPE(c)\ when\ doing\ 10 - fold\ cross\ validation$

$j \in [-5,5]$ *(Varies to show trade-off of increasing/decreasing factor of $\sigma_{MAPE(c)}$)*

$$\hat{y}_{workday_{i+1}}(k) = \beta_0 + X_i(k) * \beta_1 \tag{3.2}$$

$$\hat{y}_{i+1}(k) = \left(\hat{y}_{workday_{i+1}}(k) * \left(1 + MAPE(c) + \sigma_{MAPE(c)} * j\right)\right) * D \tag{3.3}$$

When simulating UBM over 2015-2019, Equation 3.2 and Equation 3.3 apply the fixed parameter values of $\boldsymbol{\beta_0}$ *and* $\boldsymbol{\beta_1}$ at $\boldsymbol{t_c}$ derived from minimizing the *MAPE* value (Equation 3.1) over the following period:

$[t_{i-1}, t_i) = 2nd\ half of\ 2018$

$[t_i, t_{i+1}) = 1st\ half of\ 2019$

Next to that, the derived minimal $\boldsymbol{MAPE(c)}$ value and corresponding $\boldsymbol{\sigma_{MAPE(c)}}$ are fixed factors in Equation 3.3, such that conservative predictions are used.

5.2.3 Usage-Severity-Based Maintenance

Some sub-systems might vary in degree of degradation because of variation in the considered severity of actual usage failure parameter(s). For example, in the current phase not only the number of clinical procedures is taken into account, but also the type of clinical procedure. USBM is applicable on the PM tasks if sub-system(s) differs in degradation per unit of the failure parameter(s).

USBM is only applicable if distinction between the type of usage is considered during the FMEA. If this is not the case, expert knowledge of the SMEs of R&D could be applied instead. To illustrate, imagine that the failure parameter considered at UBM is the number of clinical procedures and applying USBM leads to a differentiation in severity of degradation per type of clinical procedure. Additionally, imagine it is known that the number of movements of the sub-system depends per type of clinical procedure. If there is no data available of the severity of degradation of each type of clinical procedure, a SME might know the number of movements per type of clinical procedure. This example would be suitable for applying SME knowledge because the SMEs recognize the failure parameter(s) and can link it directly to each type of clinical procedure. Nevertheless, if degradation of the sub-system is not affected by the number of movements, PM failure data is required in order to determine the severity in degradation per type of clinical procedure with regard to that particular sub-system.

It could be that severity per type is derivable from the FMEA. For example, a certain interval of execution of PM is based on either two 'Cardio' clinical procedures or one 'Vascular' clinical procedure per day. This indicates that a 'Vascular' clinical procedure causes twice as much degradation compared to a

‘Cardio’ clinical procedures. If this distinction between type of clinical procedures would be considered during the execution of an FMEA, then USBM would be applicable as well. A prognostics method similar to the one proposed in the previous section should be developed for all types of clinical procedures.

Currently, the FMEA report does not take into account the severity per type of usage. Therefore, testing this part of the model is not possible within this research.

5.2.4 Adding operational context to maintenance policies

This section describes how the operational context could contribute to the mentioned maintenance policies since it removes variation in usage (CTBM) and degradation (UBM and USBM). By doing this, the single-parameter model (usage) becomes a multi-parameter model (including operational context).

In perspective of CTBM, adding the operational context factors is feasible. Similar computations as stated in Equation 1.1 and Equation 1.2 should be performed but now within each operational context. The operational context factors can be derived from the statistical analysis in Section 4.1. The operational context factors found by Section 4.1 were: country/culture, release and configuration. Applying CTBM within each operational context leads to a differentiation in the interval of execution of a PM task. Therefore, flexible intervals of PM visits are considered when adding the operational context factors.

From UBM and USBM perspective, devices are considered from an individual perspective. Therefore, it is not possible to remove variation in usage anymore. Currently, R&D does not include the operational context when executing an FMEA, even though a device or sub-system might be more likely to degrade in a particular operational context. Because of that, including operational context factors to UBM and USBM is not possible within this research.

However, if R&D includes the operational context factors when executing the FMEA, it will result in different threshold values for each failure parameter within each operational context. The operational context factors can be defined by applying the statistical analysis used in Section 4.1. In this case, the statistical analysis should be executed for the categorical variables and including the failures of the sub-system (instead of the whole device). The statistical analysis will define the significant differences between operational context factors. In other words, if the statistical analysis shows that failure rates differ because of operational context, an identical FMEA should be performed in all different operational contexts.

Moubray (1999) confirms this approach, the paper mentions that reliability centered maintenance should be performed in all specific contexts. Considering relevant operational context factors applied by R&D during the life tests removes uncertainty and therefore adapts the interval of execution a PM task when applying UBM or UBSM. Since currently the worst case operational context is assumed during FMEA, it is expected that the threshold values in other operational contexts will increase. Therefore, the intervals of execution of PM tasks will increase when considering relevant operational contexts separately.

This section described the theoretical part of the proposed model where CTBM, UBM and USBM for IXR-devices is applied in order to improve the PM policy and therefore answers RQ 7. Application of a single-parameter model (usage) and a multi-parameter model (including operational context) were explained.

To conclude, this chapter described that CTBM, UBM and USBM are maintenance policies that are considered feasible for IXR-devices (RQ 7). Therefore, these maintenance policies are part of the proposed model. Within these policies the operational context that might cause variation in usage or degradation of a sub-system, is not considered yet. Therefore, the proposed model exists of single-parameter (usage) models and a multi-parameter model (including operational context) in order to determine the interval of execution of a PM task.

6. Design: applying model

The previous chapter described the feasible maintenance policies within the model and how to apply them with regard to IXR-devices from a theoretical perspective. This chapter, firstly, describes the application of the proposed model on two PM tasks. Also, the benefits of the model compared to the current PM policy are discussed. This section includes the second deliverable: a single-parameter (usage) model and multiple-parameter (including operational context) model to improve the PM policy of IXR-devices. Secondly, this section focuses on how the model could be applied to the remaining PM tasks. Thirdly, this section describes a redesign of the business process in order to periodically improve the PM policy. Therefore, this section answers RQ 8 and includes the third deliverable of this research. Finally, the proposed model and redesign of the business process are validated.

6.1 Results

The model discussed in the Section 5.2 was applied on two PM tasks namely PM task J and K, to demonstrate the improvements of the model compared to the current PM policy. The remaining PM tasks are not included because of missing data regarding failure parameters and a lack of time. Since sub-systems of both PM tasks do have different failure parameter(s), each PM task is discussed separately. Also, to what extent the model was tested, hardly depends on the availability of data derived from the FMEA. Because of that, the model is partly tested for both PM tasks.

6.1.1 PM task J

Based on the FMEA where PM task J is derived of, CTBM and UBM were applied. Table 13 shows the failure parameter that causes degradation to the sub-system of PM task J.

Table 13: Failure parameter of sub-system related to PM task J.

Failure parameter	Value	Unit (f)	Current interval
Clinical procedures	Number of clinical procedures performed at device	U	1 year

Failure parameter of sub-system

The model applied for PM task J with regard to its sub-system is based on the number of clinical procedures and is derived from the FMEA. Execution of PM task J is required before exceeding the threshold value of the failure parameter:

$$U(t_l) \leq 2920$$

$$t_l = time\ since\ last\ execution\ of\ PM\ task\ J$$

$$U = number\ of\ clinical\ procedures\ performed\ at\ device$$

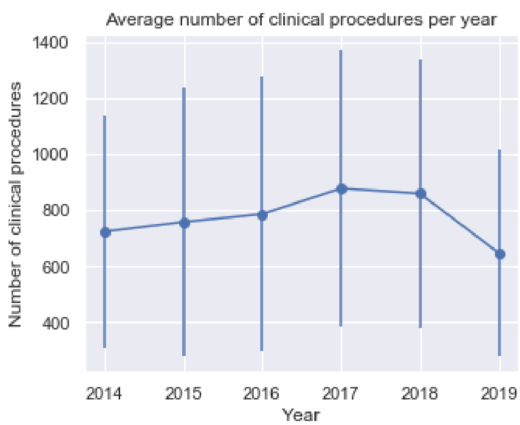
6.1.1.1 PM task J (CTBM)

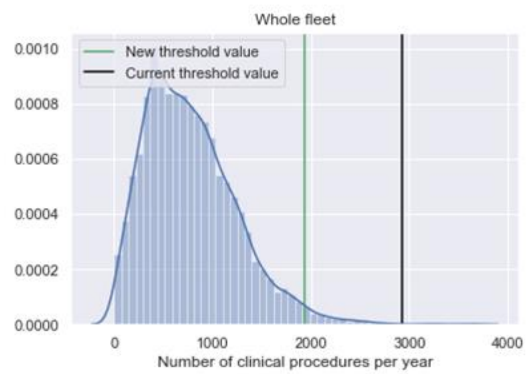
It is important to analyze the behavior of the failure parameter that covers the actual number of clinical procedures, per device. It can be clearly seen that the average number of clinical procedures/year is slightly increasing until 2018 and that the corresponding standard deviation is relatively large (Figure 17a and Figure 17b). The relatively large standard deviation means that there is a lot of variation in number of clinical procedures/year among the devices. The mean (μ) is 780 clinical procedures/year with a standard deviation (σ) of 466 clinical procedures/year .

Considering this mean of 780 clinical procedures/year, it seems evident that the interval of execution of PM task J can, on average, be extended with roughly four years. This is not necessarily the case, because we have to take into account the relatively large standard deviation. With a perfect normal distribution of the data, adding three times the standard deviation to the mean would lead to 0.1% probability of exceeding this threshold value. Based on our data that is not perfectly normally distributed, and the fact that we have an acceptable probability of 2% that exceeds the threshold value, a new threshold value for all devices was computed.

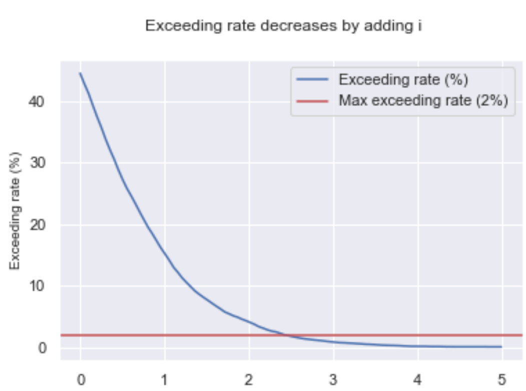
Figure 18a shows that the exceeding rate is decreasing when *i* increases. Where, *i* is the number of times that the standard deviation is added to the mean value in order to determine the new threshold value (Equation 1.2). Following the interviews with SMEs of R&D, it was concluded that 2% is an acceptable rate of PM tasks that are performed delayed (the failure parameter exceeds the threshold value), because the determined threshold value derived from FMEA includes safety factors. Next to that, this sub-system is not seen as critical. A 2% exceeding rate corresponds to an *i* value of 2.27 (Figure 18a). This leads to a new threshold value of 1934 clinical procedures/year (Figure 17b). The current threshold value was 2920 clinical procedures/year. Therefore, the proposed interval by applying CTBM, and looking at the new threshold value of the whole fleet, is 1.51 years (Figure 18b). Which is an improvement according to the current interval of 1 year.

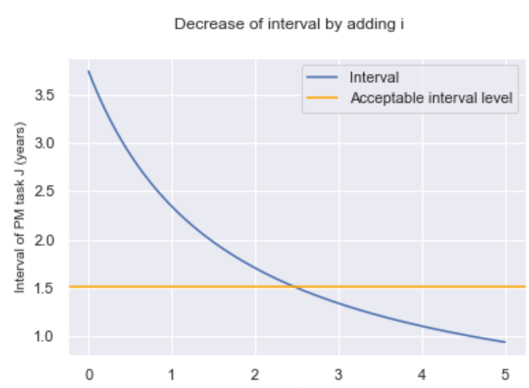
$$New\ threshold\ value = 780 + \mathbf{2.27} * 466 = 1934\ clinical\ procedures \qquad \text{(Equation 1.2)}$$
$$CTBM\ interval_{PM\ task\ J} = \frac{2920}{1934} * 1 = 1.51\ year \qquad \text{(Equation 1.1)}$$





Left: Figure 17a, average number of clinical procedures/year. Right: Figure 17b, distribution of number of clinical procedures/year with current threshold value and new threshold value [Dataset 2.1].





Left: Figure 18a, exceeding rate of failure parameter (Equation 1.3). At the exceeding rate of 2%, i = 2.27 that corresponds to the new threshold value of 1934 (Equation 1.2: $New\ threshold\ value_f = \mu_f + i * \sigma_f$). Right: Figure 18b, the interval of PM task J becomes 1.503 year (Equation 1.1). Obviously, a smaller new threshold value (smaller i) causes a larger exceeding rate and a larger interval of execution of PM task J.

Application of contextual factors (CTBM (OC))
In order to remove the variation in usage, this section includes the operational context (OC). The same probability of exceeding the threshold value of the last section is assumed valid (2%). However, in this case distributions were taken separately depending on the operational context. Next to that, dynamic intervals of PM visits are considered. By doing this, the interval of execution of PM task J can potentially be extended in many of the cases (Table 14). In practice, this means that the interval of execution of PM task J and therefore a PM visit, is adapted depending on each operational context. A 'nan' value means that there are yet insufficient data points in this operational context (less than 50). Appendix G shows the fraction of devices within each operational context. The actual usage of the devices in some operational contexts is relatively large, for example, in FR. Therefore, it may happen that the interval of execution of PM task J is decreasing. Appendix H contains the distribution of usage within each operation context, similar to Figure 17b. Appendix I contains a reliability test of 100 iterations, the coefficient of variation shows that the dispersion of intervals per operation context is fairly low (1%-33%). The lower the coefficient of variation, the more precise the estimate of the interval. Again, for computing the intervals per operational context Equation 1.1 and Equation 1.2 were applied. In this case, each operational context has a unique μ_f, i and σ_f.

Table 14: Interval per operational context with dynamic PM visits. The heatmap indicates number of data points in each operational context [Dataset 2.1].

Interval of execution of PM task J per operational context (years)					
Operational context	**US**	**JP**	**DE**	**FR**	**GB**
R8.1 – Monoplane	1.61	1.86	1.51	0.92	1.85
R8.1 – Biplane	1.70	1.33	1.52	1.50	2.29
R8.2 – Monoplane	1.74	1.54	1,24	1.04	1.47
R8.2 – Biplane	2.01	1.91	1.53	1.58	2.08
Azurion - Monoplane	2.21	1.74	1.12	1.15	nan
Azurion - Biplane	3.05	1.39	nan	nan	nan

The results in Table 14 were validated by SMEs of R&D. According to R&D, the intervals are according to expectations, since in some countries they maximize capacity of IXR-devices, where other countries do not.

6.1.1.2 PM task J (UBM)
For UBM, the interval of execution of PM task J per device was determined by looking at the actual usage of a device on individual level. Since UBM has been applied in a pro-active way, prognostics regarding the number of clinical procedures until next PM visit were executed. In order to decide what model technique to apply Dataset 2.2 was investigated. It turned out that the usage data does not change substantially over time (Figure 19).

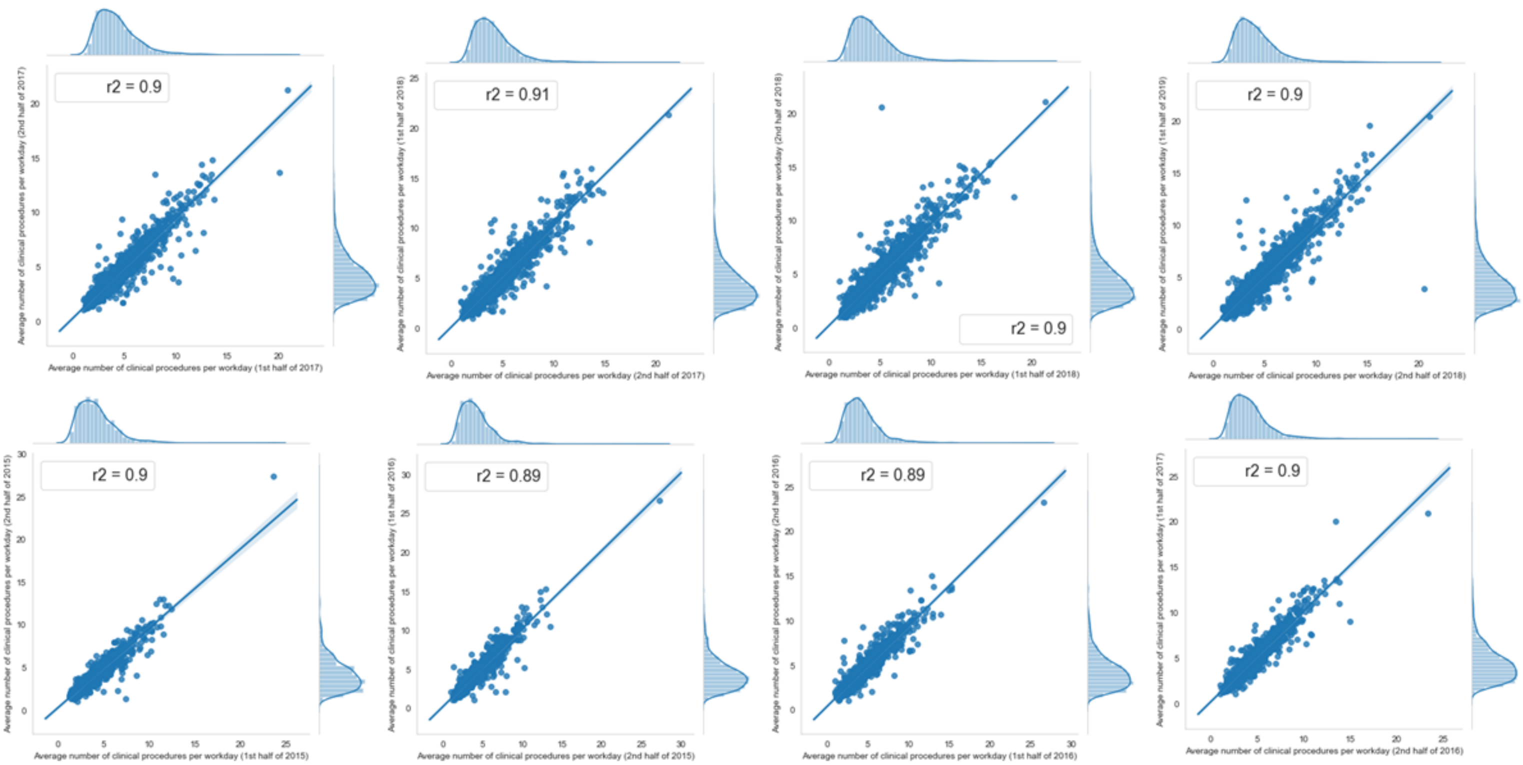


Figure 19: Average number of clinical procedures per workday (2015-2019) [Dataset 2.2].

Figure 19 shows per device the average number of clinical procedures per workday from 2015 to the first half of 2019. The large R^2 values confirm the statement made by SMEs of R&D saying that the number of clinical procedures does not change substantially over time. In this case, the large R^2 values indicate that single linear regression might be a suitable model for prognostics.

Since the behavior in usage is relatively constant, it was decided to do prognostics within the simulation based on a fixed β_0 and β_1. As stated in Equation 3, data of the second half of 2018 and the first half of 2019 was used in order to define the optimal parameter values β_0 and β_1 of a single linear regression. A 10-fold cross validation was applied in order to determine the optimal parameter values that correspond with the minimum MAPE. The single linear regression method performed with:

$MAPE(c) = 0.10$ and $\sigma_{MAPE(c)} = 0.11$ where $\beta_0 = 0.14, \beta_1 = 1.00.$ (Equation 3.1)

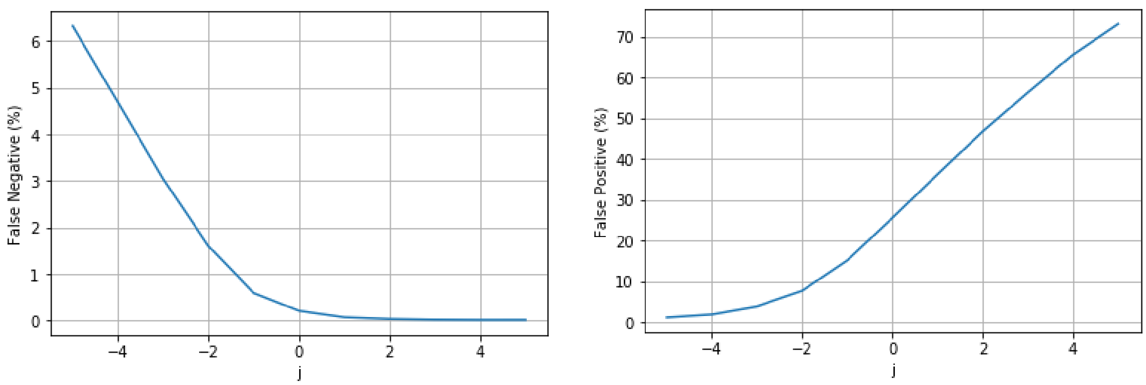
Again, both values of β_0 and β_1 indicate that behavior does not change substantially over time. Note, β_0 and β_1 were determined based on average usage per workday.

To indicate the performance of UBM regarding the current PM performance, the UBM approach was simulated over 4.5 years (01-01-2015 until 01-06-2019). In order to determine whether PM task J could be postponed Equation 2.1 was applied. For predicting usage of each device, both Equation 3.2 and Equation 3.3 were applied with $\beta_0 = 0.14, \beta_1 = 1.00$. The monitored number of clinical procedures are considered reliable and therefore the number of workdays as well. Because of that, the maximum number of workdays (per half a year) in the history of a device was used as input for parameter *D.* A device was included in the dataset in case it contained at least one year of monitored data and had data until the first half of 2019. This is done because it is unpreferable to add devices to the simulation that leave the data pool when the threshold value is almost reached. Therefore, the number of devices in the data pool increases over time (Dataset 2.2).

Trade-off depends on 'j' value

The *j* value varied from -5 to 5 which means a trade-off between the rate of exceeding the failure parameter and the number of times executing PM task J were made. A higher value of *j* causes a higher (conservative) prediction $\hat{y}_{i+1}(k)$ (Equation 3.3). Obviously, more executions of PM task J would be done when *j* increases, even though in a later stage it was not required (false positive) (Figure 20b).

Vice versa, decreasing *j* causes lower predictions of $\hat{y}_{i+1}(k)$ (Equation 3.3). Therefore, postponing of the execution of PM task J occurs more frequently and the failure parameter gets exceeded more often (false negative) (Figure 20a).



Left: Figure 20a, impact of j value on false negative ratio. Right: Figure 20b, impact of j value on false positive ratio.

The results of applying UBM on PM task J compared to the current periodic maintenance policy are shown in Figure 21. Looking into one failure parameter: the number of clinical procedures, shows us that a significant reduction in executions of PM task J is theoretically possible. The number of executions was, depending on the *j,* reduced to 22%-34% compared to the current situation. Considering the acceptable exceeding ratio of 2%, the fraction of execution of PM task J would be 24% (*j=-2).* Since the number of devices in Dataset 2.2 increases over time, it is expected that the fraction of executions of PM tasks J, compared to the current situation, will be slightly higher.

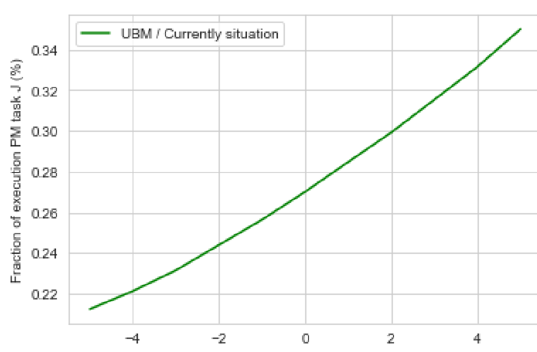


Figure 21: Impact of j value to fraction of executions PM task J (UBM compared to the current situation).

6.1.1.3 Comparing results of applying model to PM task J

Application of the proposed model results in a significant reduction of the number of executing PM task J. The policies applied in the model were simulated over 4.5 years with actual usage data. Moreover, a fixed visit each six months was assumed for CTBM and UBM. This means, if the interval associated with CTBM is 2.4 years, PM task J is executed every 2 years. Flexible intervals of PM visits were considered regarding CTBM (OC). Figure 22 shows the fraction of the number of executions of PM task J compared to the current periodic maintenance policy that assumes worst-case behavior without considering actual usage data or operational context. It can clearly be seen that the application of CTBM reduces the number of executions with 25%. When taking the operational context into consideration, this reduction amounts to 44%. Considering each device individually, UBM can be applied which would even lead to a reduction of 76% compared to the current situation ($j = -2$).

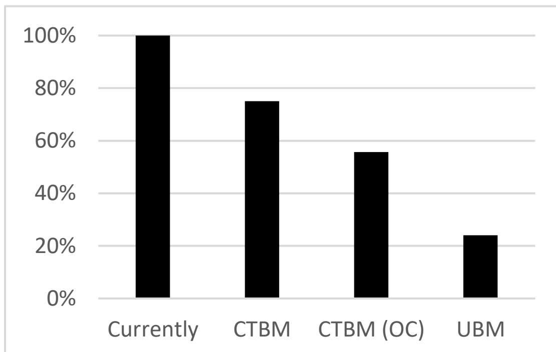


Figure 22: Fraction of number of executions of PM task J per maintenance policy.

6.1.2 PM task K

Based on data derived from the FMEA of the sub-system that is related to PM task K, CTBM and UBM were applied. The sub-system related to PM task K exists of two axes (Table 15). It turned out that this sub-system degrades based on the failure parameter: distance in meter, which is based on the FMEA and expert knowledge of the SMEs of R&D.

Table 15: Failure parameters of sub-system related to PM task K.

Parameter	Value	Unit (f)	Current interval
Meter	Distance of longitudinal axis	M_{Long}	1 year
Meter	Distance of lateral axis	M_{Lat}	1 year

Failure parameters of sub-system

The failure parameters for PM task K related to its sub-system is based on the distances travelled by both axes and is derived from the FMEA:

$$M_{long}(t_l) \leq 4032 \; OR \; M_{lat}(t_l) \leq 443$$

$$t_l = time\ since\ last\ execution\ of\ PM\ task\ K$$

$$M_{Long} = total\ distance\ travelled\ in\ longitudinal\ direction, in\ meters$$

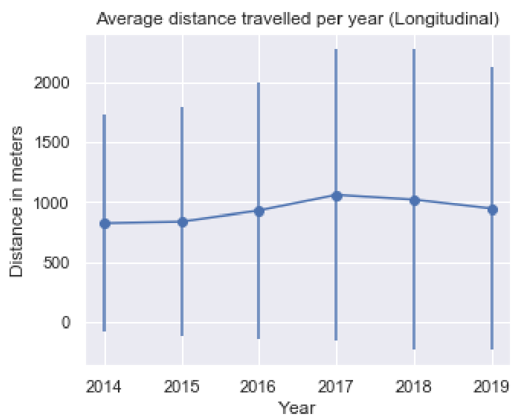
$$M_{Lat} = total\ distance\ travelled\ in\ lateral\ direction, in\ meters$$

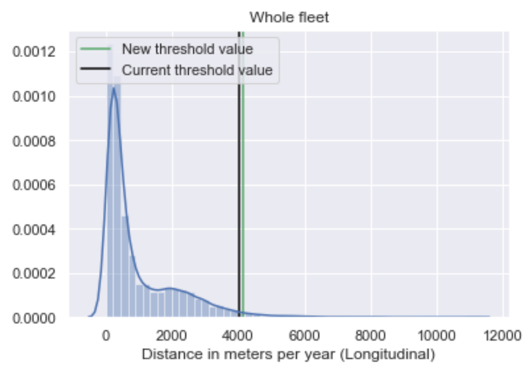
6.1.2.1 PM task K (CTBM)

The mean (μ) distance for the longitudinal axis is 951 meter/year, accompanied by a standard deviation (σ) of 1124 meter/year (Figure 23a and Figure 23b). Following the interviews with SMEs, it was concluded that a 2% probability of a delayed execution of a PM task is acceptable. This corresponds to an *i* value of 2.82 (Figure 24a) and a new threshold value of 4132 meters/year. The current threshold value of this failure parameter is 4032 meters/year. Therefore, the proposed interval by applying CTBM is 1 year for the longitudinal axis (Figure 24b). So, no further stretching of the interval of execution of PM task K is considered realistic since the current threshold value is already exceeded with roughly 2%.

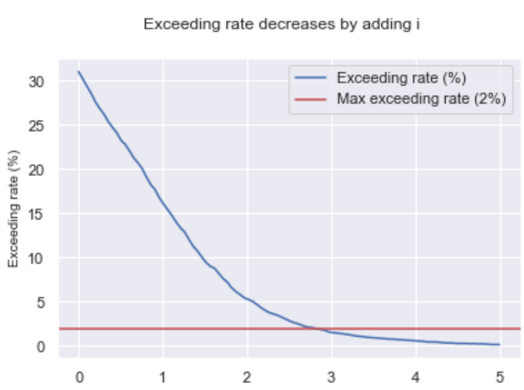
$$New\ threshold\ value_{long} = 951 + \mathbf{2.82} * 1124 = 4132\ meters \qquad \text{(Equation 1.2)}$$

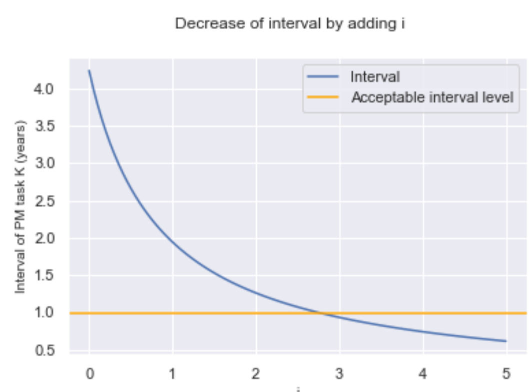
$$CTBM\ interval_{PM\ taks\ K} = \frac{4032}{4132} * 1 = 0.97 \approx\ 1\ year \qquad \text{(Equation 1.1)}$$





Left: Figure 23a, average number distance in meters/year. Right: Figure 23b, distribution of distance in meters/per year with current threshold value and new threshold value [Dataset 3.1].



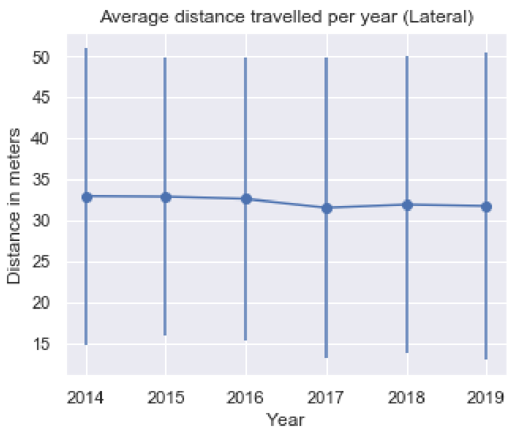


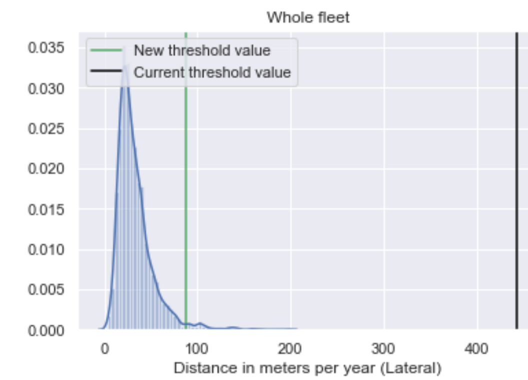
Left: Figure 24a, exceeding rate of failure parameter (Equation 1.3). At the exceeding rate of 2%, i = 2.82 and the corresponding new threshold value is 4132 (Equation 1.2: $New\ threshold\ value_f\ =\ \mu_f\ +\ i * \sigma_f$). Right: Figure 24b, the interval of PM task J becomes 1 year (Equation 1.1). Obviously, a smaller new threshold value (smaller i) causes a higher exceeding rate and a higher interval.

The mean (μ) distance for the lateral axis equals 32.79 meter/year with a standard deviation (σ) of 18.91 meter/year (Figure 25a and Figure 25b). The *i* value is equal to 2.92 (with exceeding rate of 2%). Therefore, the new threshold value at the lateral axis, before PM has to be executed, is 88.17 meters/year. Since the current threshold value of the failure parameter is 443 meters/year, the interval was stretched to 5 years for the lateral axis.

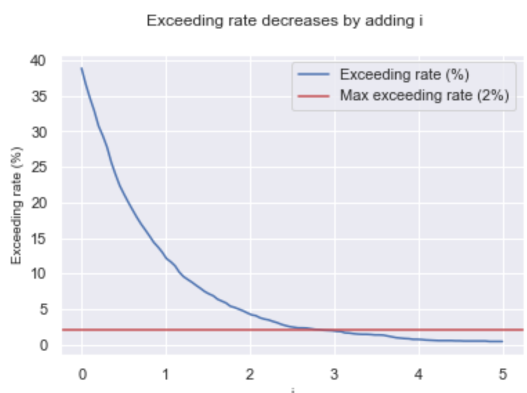
$$New\ threshold\ value_{lat} = 32.79 + \mathbf{2.92} * 18.91 = 88.17\ meters \quad \text{(Equation 1.2)}$$

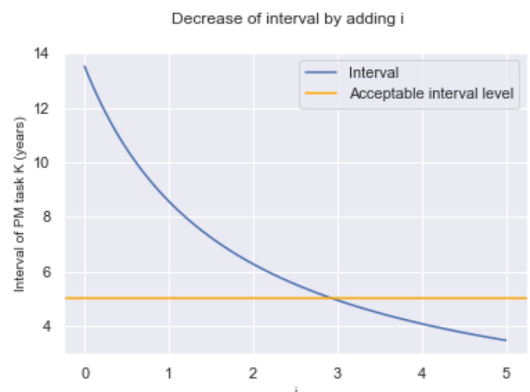
$$CTBM\ interval_{PM\ task\ K} = \frac{443}{88.17} * 1 \approx\ 5\ years \quad \text{(Equation 1.1)}$$





Left: Figure 25a, average number distance in meters/per year. Right: Figure 25b, distribution of distance in meters/per year with current threshold value and new threshold value [Dataset 3.1].





Left: Figure 26a, exceeding rate of failure parameter (Equation 1.3). At the exceeding rate of 2%, i = 2.92 and the corresponding new threshold value is 88.17 (Equation 1.2: $New\ threshold\ value_f\ =\ \mu_f\ +\ i * \sigma_f$). Right: Figure 26b, the interval of PM task J becomes 5 years (Equation 1.1). Obviously, a smaller new threshold value (smaller i) causes a higher exceeding rate and a higher interval.

Since the need for PM task K depends on two failure parameters, execution of PM task K is required when one of those failure parameters exceeds the threshold value. Therefore, the new interval for PM task K by applying CTBM would be similar to the current situation, 1 year:

$$CTBM\ interval_{PM\ task\ K} = \min_{f \in F_{sub-system}} \left(\frac{\text{Thresholdvalue}_f}{\text{Boundary value}_f} * current\ interval_{PM\ task}\right) = \min_{f \in F_{sub-system}} (1, 5) = 1\ \text{year} \quad \text{(Equation 1.1)}$$

Nevertheless, it would be worthwhile to investigate the opportunity to split PM task K into two tasks that each have one failure parameter. By doing this, the interval of execution of the PM task on the lateral axis, can be stretched to 5 years.

Application of contextual factors

In order to remove the variation in usage, this section includes the operational context. The same probability of exceeding the threshold value of the last section is assumed valid (2%). However, in this case distributions were taken separately depending of the operational context and therefore dynamic intervals of PM visits were considered. By doing this, the interval of execution of PM task J can potentially be extended in many of the cases (Table 16). In practice, this means that the interval of execution of PM task K and therefore a PM visit, is adapted depending on each operational context. A 'nan' value means that there is yet insufficient data available within this operational context yet (less than 50). Appendix G shows the fraction of devices per operational context. The actual usage of the devices in some operational contexts is larger, for example, in FR. Therefore, it occurs that the interval of execution of PM task K decreases. Appendix H contains the distribution of usage within each operation context, similar to Figure 23b and Figure 25b. Appendix I contains a reliability test of 100 iterations, the coefficient of variation shows that the dispersion is fairly low for the operational contexts that include relatively more data points (4%-10%). The lower the coefficient of variation, the more precise the estimate of the interval. Again, for computing the intervals per operational context Equation 1.1 and Equation 1.2 were applied. In this case, each operational context has a unique μ_f, i and σ_f.

Table 16: Interval per operational context. The heatmap indicates the number of data points in each operational context [Dataset 3.1].

Interval of execution of PM task K per operational context (years)					
Operational context	***US***	***JP***	***DE***	***FR***	***GB***
R8.1 – Monoplane	0.97	1.10	3.62	0.63	8.03
R8.1 – Biplane	1.40	1.29	1.14	nan	7.99
R8.2 – Monoplane	0.85	nan	1.39	0.97	5.56
R8.2 - Biplane	0.99	1.25	0.95	nan	4.41
Azurion - Monoplane	1,18	nan	3.11	nan	6.77
Azurion - Biplane	2.36	nan	nan	nan	nan

The results in Table 16 were validated by SMEs of R&D. According to the SMEs of R&D, the intervals are according to expectations, since in some countries they tend to make more use of the axes than other countries. Since correctly positioning the device for a treatment can also be done by moving other sub-systems.

6.1.2.2 PM task K (UBM)

For UBM the interval of execution of PM task K per device was determined by looking at the actual usage of a device on individual level. Since UBM has been applied in a pro-active way, prognostics regarding the travelled distance in meters until the next visit, were executed. In order to decide what model technique to apply, the data of Dataset 3.2 was investigated. It turned out that the usage data of Dataset 3.2 does not change substantially over time (Figure 27 and Figure 28).

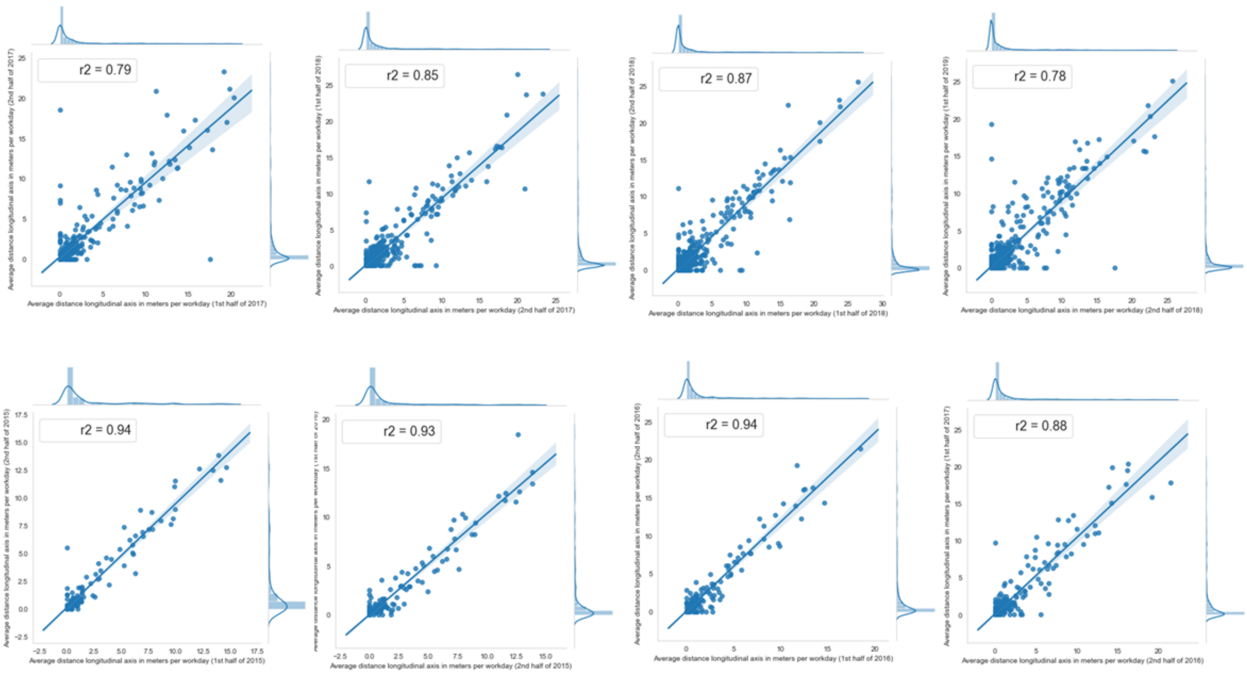


Figure 27: Average distance travelled by longitudinal axis per workday (2015-2019) [Dataset 3.2].

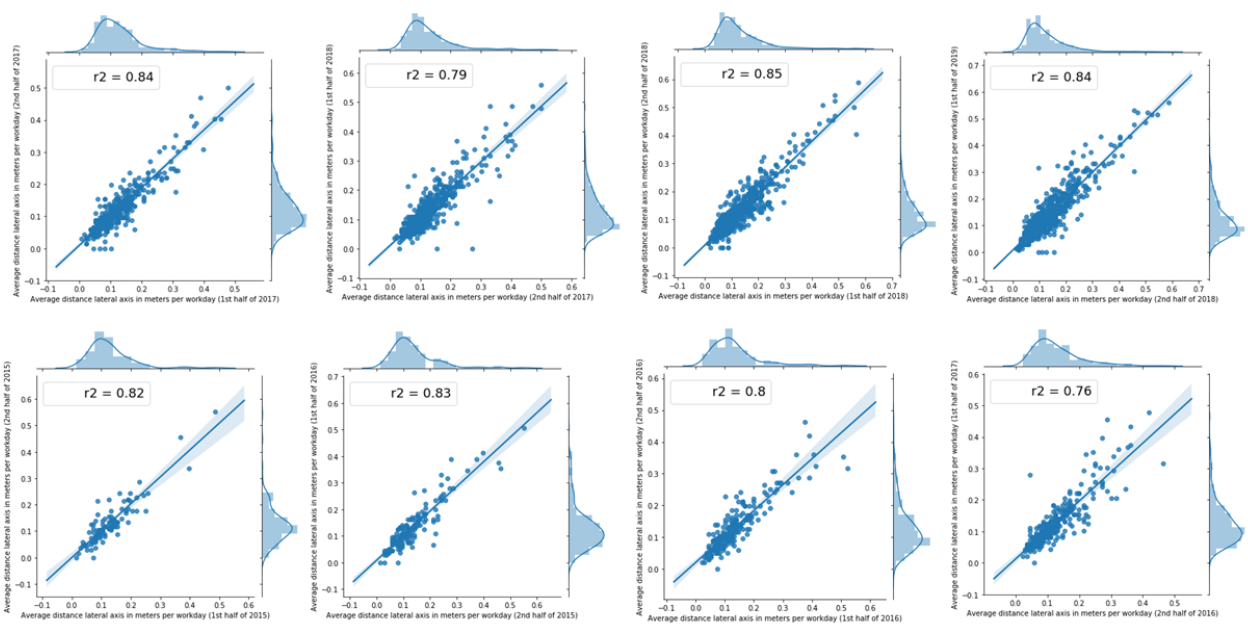


Figure 28: Average distance travelled by lateral axis per workday (2015-2019) [Dataset 3.2].

Figure 27 and Figure 28 show per device the distance travelled in meters by the longitudinal and lateral axis from 2015 to the first half of 2019. The large R^2 values confirm the statement made by SMEs of R&D, saying that the number of travelled distance in meters does not change substantially over time. In this case, the large R^2 values indicate that single linear regression is potentially a suitable model for prognostics.

Since the behavior in usage does not change substantially over time, it was decided to do prognostics within the simulation based on a fixed β_0 and β_1. As stated in Equation 3, data of the second half of 2018 and the first half of 2019 was used in order to define the optimal parameter values β_0 and β_1 of a single linear regression.

During the research project it turned out that MAPE was not the most suitable method to measure accuracy of the forecasts of the failure parameter(s) of the sub-system related to PM task K, because the longitudinal axis of a relatively large fraction of the devices contains a travelled distance equal to zero ($y_{workday_k}$= 0). This means a relatively large fraction of the devices does not use the longitudinal axis (45% in the second half of 2018). Because of this, the MAPE values would grow to an infinite error value. The devices that do not use the longitudinal axis, cannot possibly exceed the threshold value. Therefore, it was decided to exclude the devices that do not use the longitudinal axis in order to determine β_0 and β_1.

A 10-fold cross validation was applied in order to determine the optimal parameter values that correspond with the minimum MAPE. The single linear regression method performed with:

$Longitudinal: MAPE(c) = 0.62$ and $\sigma_{MAPE(c)} = 0.06$ where $\beta_0 = 0.32, \beta_1 = 0.93$ (Equation 3.1)
$Lateral: MAPE(c) = 0.21$ and $\sigma_{MAPE(c)} = 0.01$ where: $\beta_0 = 0.01, \beta_1 = 0.98$ (Equation 3.1)

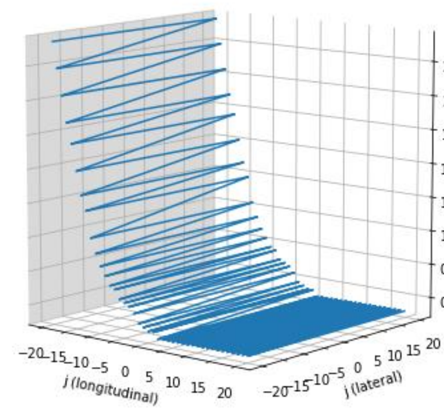
Again, both β_0 are close to zero and both β_1 are close to 1, this indicates that usage is not changing substantially over time.

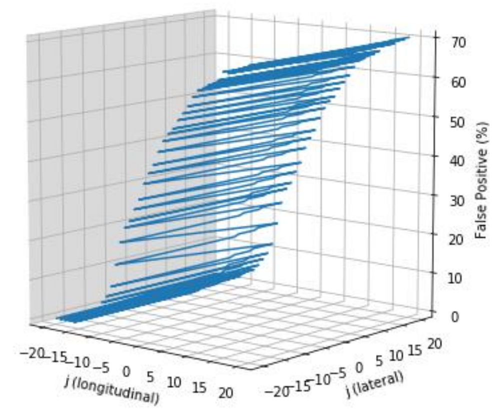
To indicate the performance of UBM compared to the current PM performance, the UBM approach was simulated over 4.5 years (01-01 2015 until 01-06-2019). In order to determine whether PM task K could be postponed, Equation 2.1 was applied. For predicting usage of each device both Equation 3.2 and Equation 3.3 were applied. The monitored distance of both axes were not considered reliable and therefore the number of workdays neither (Section 3.2). Because of that, the maximum number of workdays in half a year was used as input for parameter *D (D=136).* A device was included in the dataset in case it contained at least one year of monitored data and included data until the first half of 2019 (Dataset 3.2). This is done because it is unpreferable to add devices to the simulation that leave the data pool when the threshold value is almost reached. Therefore, the number of devices in the data pool increases over time.

Trade-off depends on 'j' value
The *j* value varies from -20 to 20, which means a trade-off between the rate of exceeding the failure parameter and the number of executions of PM task J can be made. A higher value of *j* causes a higher (conservative) prediction $\hat{y}_{i+1}(k)$ (Equation 3.3). Obviously, more executions of PM task J would be done when *j* increases, even though in a later stage it was not required (false positive) (Figure 29b).

Vice versa, decreasing *j* causes lower predictions of $\hat{y}_{i+1}(k)$ (Equation 3.3). Therefore, postponing of the execution of PM task J occurs more frequently and a failure parameter gets exceeded more often (false negative) (Figure 29a).

It can be seen that mostly *j (longitudinal)* influences both false negative and false positive ratios. This is according to expectations since the distribution of the longitudinal axis is wider and the threshold value relatively closer to the actual distance travelled looking at usage per year, compared to the lateral axis (Figure 23b and Figure 25b).





Left: Figure 29a, impact of j value to false negative ratio. Right: Figure 29b, impact of j value on the false positive ratio.

Applying UBM to PM task K leads, again, to a significant reduction in number of executions of PM task K. The number of executions is, depending on *j,* reduced to 5%-11% compared to the current policy (Figure 30). When *j (longitudinal) = -17* and *j (lateral) = -20* the rate of exceeding the threshold value is below the acceptable exceeding rate of 2%. The corresponding fraction of execution of PM tasks K is 5.7% compared to the current policy.

The large reduction is according to expectations because of the wide distribution in travelled distance of the longitudinal axis. Such that, only a small percentage of the devices has an actual high usage. Since the number of devices in Dataset 3.2 increases over time, it is expected that the fraction of executions of PM tasks K, compared to the current situation, will be slightly higher.

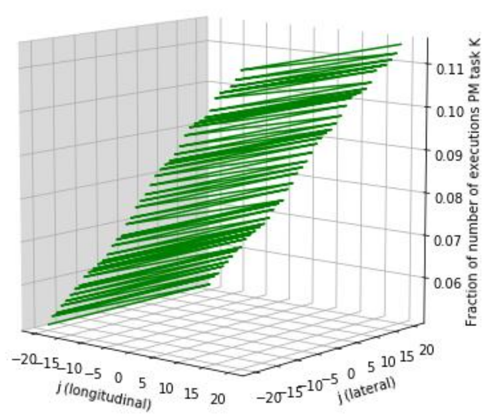


Figure 30: Impact of j value on the fraction of execution of PM task K (UBM compared to the current situation).

Application of contextual factors

Since failure data of PM tasks is not available yet, the failure data belonging to failures (CM) of the related sub-system of PM task K was used, even though there is not a causal relation between the failures (CM) and PM. Applying the statistical analysis proposed in Section 4.1.3, resulted in a significant difference in the number of failures of operational factor: release (Table 17). There was no significant difference in the number of failures of operational contexts: country/culture and configuration. Therefore, only operational context: release, was considered as an operational context factor. By knowing this, R&D should perform the FMEA separately for all releases in order to determine (different) threshold values of both failure parameters.

Table 17: Pairwise comparisons of release.

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
R8.1-R8.2	-56,974	10,893	-5,230	,000	,000
R8.1-Azurion	172,991	31,304	5,526	,000	,000
R8.2-Azurion	116,017	32,104	3,614	,000	,001

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

6.1.2.3 PM task K (USBM)

The sub-system of PM task K includes differences in severity that degrade both axes. During interviews with SMEs it turned out that, next to the distance travelled by the axes the weight of the patient and the cradle/tilt degree of this sub-system, are the most important factors that cause degradation to both axes. Therefore, there is a difference in the degree of degradation of the axes depending on the weight of the patient and the degree of cradle/tilt. At this moment, when doing an FMEA, a patient's weight of 100 kilo is assumed for every meter an axis moves even though not all patients have this weight. Unfortunately, the weight of patients is not monitored because of privacy reasons and therefore weight cannot be considered. The degree of cradle/tilt is monitored daily and therefore applicable as a parameter to determine severity per meter travelled by both axes. For this sub-system, the degree of cradle/tilt can be included in order to apply USBM. Currently, it is not possible to determine the exact severity level per degree of cradle/tilt since this is not considered when performing the FMEA.

6.1.2.4 Comparing results of applying model to PM task K

Application of the proposed model results in a significant reduction of executing PM task K. Figure 31 shows the fraction of number of executions of PM task K relative to the current periodic maintenance policy that assumes worst-case behavior without considering actual usage or operational context. All policies were simulated over 4.5 years and a fixed visit each six months was assumed for CTBM and UBM. This means, if the interval associated with CTBM is 2.4 years, PM task K will be executed every 2 years. The application of CTBM does not reduce the number of executions, since the computed interval is equal to the current interval. The wide distribution of travelled distance in meters causes that only a small percentage of the devices has an actual high usage, these devices are in the tail of the distribution and therefore the new threshold value is relatively large and no stretching of the interval of execution of PM task K is possible with CTBM. Taking the operational context with flexible intervals into consideration, the reduction in number of executions of PM task K would be 10%. Considering each device individually, UBM can be applied which would lead to a reduction of 94% compared to the current situation. This large reduction is according to expectations since the distribution in travelled distance of the longitudinal axis is wide (large σ). With UBM, the sub-systems in the tail of the distribution will have, chronologically, the first maintenance. Due to the large σ, the expected lifetime of the remaining sub-systems is larger.

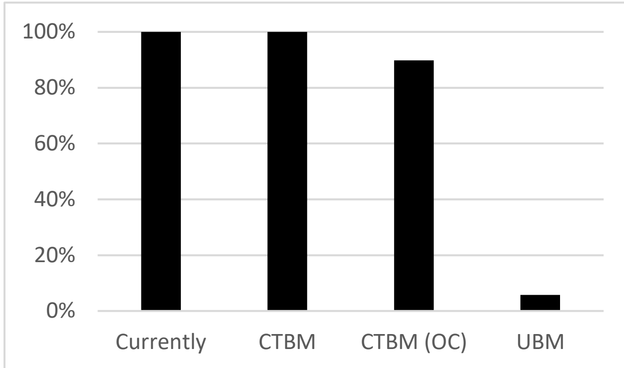


Figure 31: Fraction of number of executions of PM task J per maintenance policy.

This section included the second deliverable: a single-parameter (usage) model and multiple-parameter (including operational context) model to improve the PM policy of IXR-devices. It showed the application of the proposed model on PM task J and PM task K. To what extent the model was applied hardy depended on the available data within the FMEA report. Results showed that for PM task J, a single-parameter model CTBM resulted in significant improvements (25%). Because of the wide distribution of the distance travelled by the longitudinal axis, only a small percentage of the devices has an actual high usage (tail of the distribution). This caused that CTBM did not directly resulted in a reduction of number of executions for PM task K. Applying the single parameter model UBM resulted in a significant improvement of 76% and 94% for both PM tasks J and K. The multi-parameter model (CTBM (OC)), in turn, resulted in a significant improvement compared to the single-parameter model of CTBM for PM task J and K, respectively 19% and 10%.

This concludes that variation in actual usage was reduced when considering the operational context of a device. The number of total labor hours reduced 2%-6% when using the proposed model for PM tasks J and K (RQ 7). On the first eye, this is not a large difference, but it has to be considered that PM task J and K are relatively small compared to the remaining PM tasks upon which is scoped. Next to that, since the total labor hours are significantly high, a small reduction of labor hours already causes large savings. In order to be able to consider the operational context within policies UBM and USBM, R&D should derive a threshold value of the failure parameter(s) within each relevant operational context during the FMEA. Section 6.2 describes how the proposed model can be applied to the remaining PM tasks within scope of this research.

6.2 Apply model to remaining PM tasks

Currently, the model proposed was applied to two PM tasks which are focused on: PM task J and K. In order to further improve the current PM policy, the proposed model should be applied to the remaining PM tasks. Again, depending on the threshold value(s) of the failure parameters used in the FMEA and the measurability of the failure parameter(s), the model could potentially be applied to improve the PM policy of the remaining tasks. On the short term, product owners of the sub-systems related to PM tasks A, B, C, D and F should check which failure parameter(s) mostly influence degradation of the sub-system and what the corresponding threshold value(s) are. By doing this, the opportunity for improvement considering actual usage and operational context can be investigated using the proposed model. It is recommended to focus first on PM tasks A and B since those PM tasks have the largest impact on the total PM time per maintenance cycle. For PM task J and K, PM failure data should be collected in order to fully apply the proposed model. An overview of actions to perform on short term are shown in Table 18.

Table 18: Action per PM task upon which is scoped.

Planned maintenance task	Type	CTBM	UBM	USBM	Action
A	Calibration/verification	T.b.d.	T.b.d.	T.b.d.	Apply model by product owner
B	Calibration/verification	T.b.d.	T.b.d.	T.b.d.	Apply model by product owner
C	Creating a backup	-	-	-	Nothing. Mandatory task
D	Calibration/verification	T.b.d.	T.b.d.	T.b.d.	Apply model by product owner
E	Handing over				Nothing. Mandatory task
F	Calibration/verification	T.b.d.	T.b.d.	T.b.d.	Apply model by product owner
G	Safety related	-	-	-	Nothing. Mandatory task
H	Safety related	-	-	-	Nothing. Mandatory task
I	Safety related	-	-	-	Nothing. Mandatory task
J	Cleaning/checking	Yes	Yes	No	Gather PM failure data (see 6.3)
K	Cleaning/checking	Yes	Yes	Yes	Gather PM failure data (see 6.3)
L	Preparations	-	-	-	Nothing. Mandatory task

This section described the actions that need to be performed in order to (1) fully apply the proposed model and (2) to apply the model to the remaining PM tasks. Section 6.3 describes a redesign of the business process in order to improve the PM policy periodically.

6.3 Redesign of the business process

This section includes the third deliverable: a redesign of the business process that is required in order to periodically update the intervals of each PM task. Philips should change their business process in order to go from the current PM policy, with a fixed checklist of maintenance tasks for each PM visit, to a policy where execution maintenance tasks depends on the actual need (dynamic checklist). Next to that, this section describes how the new maintenance policy should be implemented (RQ 8).

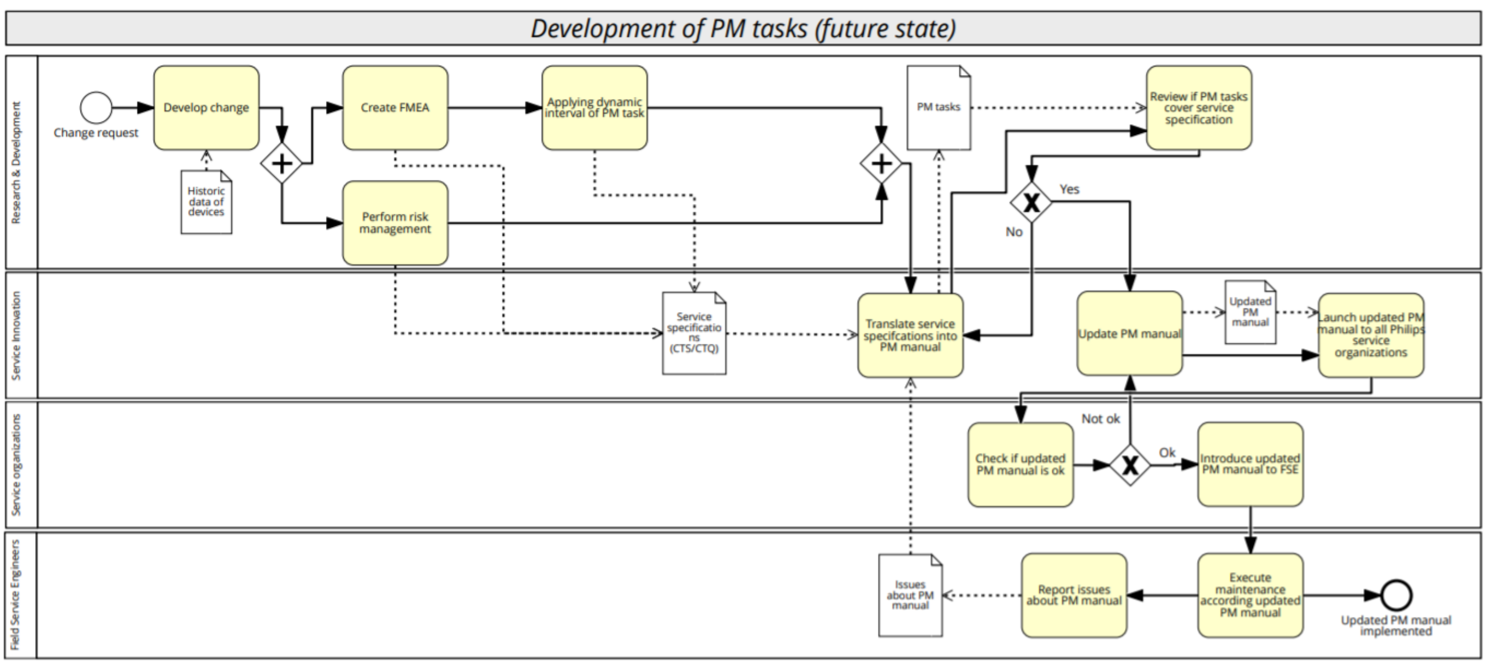


Figure 32: Redesign of business process of the development of PM tasks.

In Section 3.1 it was found that during the FMEA actual usage behavior and operational contexts are not considered. For the PM task checklist to become more dynamic, Philips should add an extra process after the process ‘Create FMEA’, is performed. This extra process is called ‘Applying dynamic interval of PM task’ (Figure 32).

Please note that the business process change is most beneficial if failure data regarding execution of each PM task is gathered. In order to gather continuously PM failure data, Philips should adapt the checklist of FSEs. As mentioned before, FSEs sign all maintenance tasks they perform at every single customer. FSEs do this regardless of whether execution of this task is definitely necessary. The checklist of FSEs requires adaptation in a way such that it is noticeable what PM tasks were definitely necessary to execute. By doing this, failure data regarding each PM task would be gathered continuously. R&D can use this failure data to see what failure parameter(s) do influence degradation of a sub-system. When this is measured continuously, R&D can update the FMEA on a periodic bases with actual data, considering actual usage and operational context.

After R&D develops an FMEA and it appears that a PM task should be executed in order to prevent failures of a sub-system, the application of dynamic intervals is investigated within the new business process. The process of applying a dynamic interval to a PM task is shown in Figure 33. First, R&D defines a link between the PM task and the sub-system(s). If this sub-system is similar to other sub-systems (but older versions), then failure parameter(s) that indicate degradation of this old sub-system will need to be defined. It is thereby required that those failure parameter(s) are monitored. The defined failure parameter(s) can be applied at the new sub-system. If R&D is unfamiliar with this sub-system, a fixed (conservative) time interval will be defined. When the new sub-system is used in the field, data is gathered about the failure of the PM task. Also, if R&D is familiar with the sub-system, but they do not know what failure parameter(s) influence degradation, a fixed (conservative) time interval is defined. In the meantime, data is gathered about the failure of the PM task.

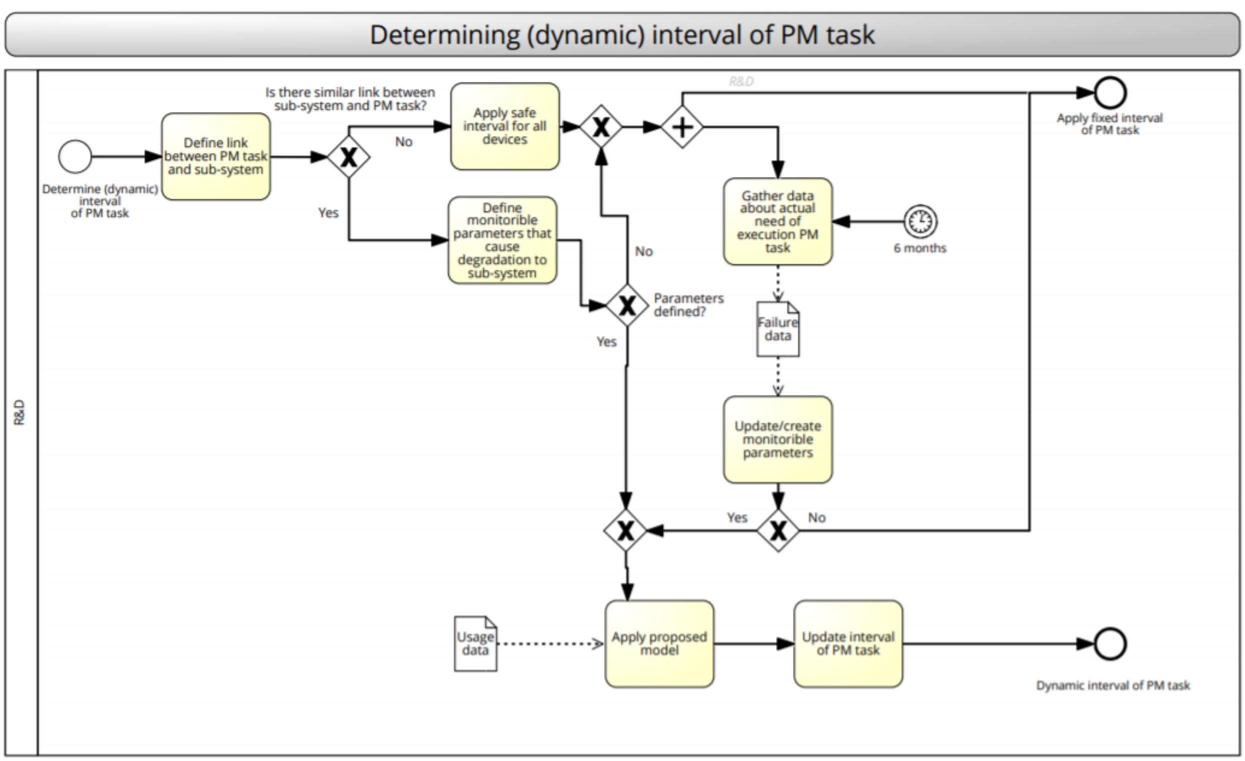


Figure 33: New business process to determine (dynamic) interval of PM tasks on a periodic basis.

If failure parameter(s) are definable (whether directly or after six months), the proposed model will be applied considering actual usage and the operational context of the devices. Additionally, the interval of execution of PM tasks is updated. From now on the interval of this particular PM task is dynamic and depends on the actual usage and operational context of the device. Since the actual failure data is monitored continuously, the applied failure parameter(s) are updated every six months. Every single sub-system related to a PM task, should be investigated separately by looking at which failure parameter(s) influence degradation. Failure parameters in this situation are all types of indicators that influence degradation of a device. This includes for example: number of clinical procedures, type of clinical procedures, number of times the sub-system went on/off, meters of movement, wear, temperature, time etc.

If it is not clear what failure parameters do influence the degradation of a sub-system or the failure parameters are not monitored, the proposed model is unfeasible and the interval of the PM task remains fixed.

In order to fully apply the proposed model and to improve the PM policy periodically, the mentioned business redesign requires practical implementation. Table 19 shows for every function the tasks that need to be executed in order change the current static PM policy to a more dynamic maintenance policy on the long term (RQ 8). A RACI-matrix (Smulders, 2018) was used to show the roles of each function related to the tasks.

Table 19: RACI-matrix that shows implementation plan of both artifacts.

Task/Function	*Product owner of sub-system (PM task A)*	*Product owner of sub-system (PM task B)*	*Product owner of sub-system (PM task D)*	*Product owner of sub-system (PM task F)*	*Product owner of sub-system (PM task A)*	*Product owner of sub-system (PM task J)*	*Product owner of sub-system (PM task K)*	*Service Innotvation*	*Data Scientist*	*Project leader*
Apply model	R	R	R	R	R					A
Define information required to check if PM task was actually needed	R	R	R	R	R	R	R	I		A
Change work orders of FSEs, so required input data is gathered								R		A
Analyse PM failure data and see what parameter(s) do influence degradation of sub-system	R	R	R	R	R	R	R		C	I
Determine threshold values per parameter	R	R	R	R	R	R	R		C	
Apply model	R	R	R	R	R	R	R			A
Adapt intervals of PM tasks								R		A
Apply changed business process	R	R	R	R	R	R	R	R		A
Investigate opportunities to implement UB/UBSM								R		C/I

0-1 year	R: Responsible
1-2 years	A: Accountable
2-2.5 years	C: Consulted
> 2.5 years	I: Informed

This section included the third deliverable: a redesign of the business process that is required in order to periodically update the intervals of each PM task. In order to fully apply the proposed model and the redesign of the business process, it is required to gather failure data of actual need for execution of PM tasks. Next to that, this section described the tasks that need to be performed by multiple functions in order to implement the proposed model and the redesign of the business process successfully (RQ 8). Section 6.4 describes the validation of both artifacts: the proposed model and the redesign of the business process.

6.4 Validation

Since the use of the proposed model and redesign of the business process cannot be tested in practice, because of a lack of time during this research, both artifacts were validated by presenting them to the SMEs of R&D.

The model was presented to multiple owners (R&D) of sub-systems related to PM tasks. The majority of SMEs suggested that the proposed model could be useful. Since there is no system or software yet that can automatically decide whether a PM task has to be performed for an individual device, they preferred using the CTBM and CTBM (OC) policy on the short term. They mentioned that it could give direct insights in how fleets of devices are used, based on operational context. The conservative intervals can potentially be transformed to intervals based on to actual (worst-case) usage within each operational context. SMEs of R&D consider this as a quick win in order to reduce the large number of PM labor hours. Nevertheless, in case of CTBM (OC), it is required that all tasks within PM visit can be adapted to this flexible interval. Therefore, CTBM (OC) needs to be applied at the remaining PM tasks first.

Applying the UBM or USBM could be useful in the future, in combination with the proposed business process redesign. Next to the fact that the current information system is not ready for tuning the PM checklist of each individual device at this moment, no PM failure data is collected to determine the failure parameter(s) that influence the actual need for execution of a PM task. The SMEs would greatly prefer to gain more information about the actual need of PM and therefore they supported the idea of adapting the checklists of FSEs to gather failure data of PM. The SMEs state that reviewing the FMEA on periodic basis is useful in order to update the interval of execution of PM tasks. Nevertheless, they state it is only beneficial to do this for the PM tasks that cause the largest PM time.

To conclude, this chapter described the application of the proposed model on two PM tasks. The model is only partly applied since it was not possible to derive all required data from the FMEA reports. Therefore, single-parameter models (CTBM and UBM) and a multi-parameter model (CTBM (OC)) were applied to PM task J and K (second deliverable). Applying the model resulted in a significant reduction in number of executions of PM tasks J and K.

Results showed that for PM task J, a single-parameter model CTBM resulted in significant improvement of 25%. Because of the wide distribution of the distance travelled by the longitudinal axis, only a small percentage of the devices has an actual high usage (tail of the distribution). This caused that CTBM did not directly resulted in a reduction of number of executions for PM task K. Applying the single parameter model UBM resulted in a significant improvement of 76% and 94% for both PM tasks J and K. The multi-parameter model (CTBM (OC)), in turn, resulted in a significant improvement compared to the single-parameter model of CTBM for PM task J and K, respectively 19% and 10%. Therefore, CTBM (OC) outperforms CTBM and is it proven that variation in actual usage was reduced by considering the operational context of a device.

Removing variation in degradation with UBM and USBM is feasible when R&D considers the relevant operational context when performing a FMEA. Next to that, this chapter described how the proposed model can potentially be used for the remaining PM tasks which is scoped upon and how a redesign of the business process can contribute in improving the PM policy periodically (third deliverable). In order to fully use the proposed model and the redesign of the business process, failure data of PM should be gathered continuously by adapting work order of the FSEs. SMEs of R&D validated the proposed model and the redesign of the business process. CTBM was considered as a quick win because of its ease of implementation. For CTBM (OC), flexible intervals of other sub-systems have to be determined first in order to adapt the interval of a PM visit. The remaining part of the model (UBM and USBM), in combination with the redesign of the business process, were considered as something useful for the future. This holds because the information system is not ready to tune the PM checklist for each individual device at this moment. Next to that, PM failure data is not gathered yet.

7. Conclusion

The aim of this research was to improve the current PM policy of IXR-devices with the use of data (data-driven) and considering the operational context (context-awareness). This research is highly relevant given the current large number of labor hours associated with PM of IXR-devices, that cause a significant part of the total maintenance costs. Next to that, the large number of labor hours associated with PM of IXR-devices invokes (costly) downtime for the customer. In the proposed design, three deliverables were determined: (1) a report of the qualitative and quantitative analysis that shows the bottlenecks and contextual factors within the current maintenance policy, (2) a single-parameter and multiple-parameter model to improve the current periodic maintenance policy and (3) a redesign of the business process in order to improve the current PM policy periodically. To achieve these deliverables, eight research questions required answering, where the research methodology to accomplish this consisted of the regulative cycle and CRISP-DM. This chapter describes the conclusion of the research, the limitations and recommendations for future research.

7.1 Conclusion of research

This section describes the conclusion of the research from both business and scientific perspective.

Define and business understanding

Firstly, an explorative research was conducted as part of the first phase of both the regulative cycle (*define research*) and CRISP-DM (*business understanding*). This was executed in order to get an understanding of current maintenance policy for IXR-devices, to formulate research questions and to scope the research project.

Diagnose, data understanding and data preparation

Secondly, during the *diagnose* phase of the regulative cycle, questions concerning the strategy of the maintenance policy (RQ 1) and which stakeholders were involved (RQ 2) were established. During this phase it turned out that R&D is the most important stakeholder, since they perform FMEAs for each sub-system. Based on the results of the FMEA, it is decided whether PM tasks have to be executed to prevent failures. Consequently, R&D has to determine an interval of execution of each PM task that is based on a certain conservative threshold value that includes safety factors. Next to that, R&D assumes worst-case usage behavior of the device without considering actual usage or operational context. The second most import stakeholder is SI. SI translates the service specifications (results of FMEA) into the PM manual, after a review of the PM manual by R&D the (updated) PM manual is released to the field where the FSEs perform the PM tasks. The PM schedule is similar for all IXR-devices, which means no differentiations based on contextual factors (usage and operational context) are made.

During this phase, a quantitative analysis was conducted as well. There, with assistance of data-scientists understanding of the available data was gained and data was prepared. It turned out that available data on the effectivity and efficiency of the PM tasks is limited. The effectivity of a PM task is not monitored, since FSEs perform the tasks on a checklist without checking the actual need. The actual time per maintenance visit is monitored, but this time consists of all activities that have been performed during that visit. Depending on the specific country's regulations it can occur that extra PM tasks (that are not in the PM manual) need to be performed. This makes it hard to measure the efficiency of each PM task. Nevertheless, three useful datasets were acquired in order to indicate the contextual factors and to simulate the multiple maintenance policies using prognostics (RQ 3).

By combining the qualitative research, quantitative research and literature, statistical hypotheses were formulated. This has been done to check how contextual factors influence the number of failures (RQ 4). The contextual factors tested, exist of discrete variables (usage) and categorical variables (parameters that determine the operational context). With the application of statistics (Spearman's correlation and Kruskal-Wallis test) the contextual factors that influence the mean number of failures per year were pinpointed. The usage parameters 'Number of clinical procedures', 'Number of workdays', 'Number of clinical procedures per clinical type' turned out to influence the number of failures per year (except for clinical procedure type: 'Unknown'). The 'Number of clinical procedures' is most strongly correlated with the mean number of failures. The operational context factors country/culture, configuration and release, were found to influence the mean number of failures per year.

Next to that, it was determined that a maintenance cycle consists of two years, so that in two years all PM activities are performed. Since some PM tasks are mandatory to execute because of safety regulations, an FSE has to visit a device every six months at least once. During this phase, the type, characteristics and the planned time of each PM task were established (RQ 5). From that moment the research was scoped upon seven PM tasks that are causing most of the PM time within a maintenance cycle. Noteworthy, none of the PM tasks includes replacements of sub-systems. In fact, the PM tasks exist mostly of cleaning/checking functionality, calibrating/verification and safety related tasks.

Within the *diagnose* phase a qualitative research was conducted in order to determine which PM tasks allow for stretching of the interval of execution. The qualitative aspect involved semi-structured interviews with SMEs of R&D, and included a workshop where they had to indicate the opportunities of increasing the interval of a PM task with the use of usage data. Gathering all the knowledge gained by interviews and data analysis, it turned out that the interval of execution of PM tasks: A, B, C, D, F, J and K, can possibly be stretched (RQ 6).

During the *diagnose* phase, the first deliverable was conducted: a report of the qualitative and quantitative analysis that shows the bottlenecks and contextual factors within the current maintenance policy. Next to that, it turned out of which PM tasks the interval of execution can be stretched.

Design, data modeling and evaluation

Thirdly, during the *design* phase of the regulative cycle it was investigated what model was suitable to improve the current maintenance policy, considering the contextual factors (RQ 7). By doing a literature review it was found what maintenance policies were available and which one were applicable to IXR-devices. It turned out that the single-parameter (usage) models: CTBM and UBM, achieved a significant reduction in number of executions of both PM task J and K. Results showed that for PM task J a single-parameter model CTBM resulted in a reduction of number of executions of 25%. Because of the wide distribution of the distance travelled by the longitudinal axis, only a small percentage of the devices has an actual high usage (tail of the distribution). This caused that CTBM did not directly result in a reduction of number of executions for PM task K. For the single-parameter model UBM, both PM tasks J and K resulted in significant reductions in number of executions, respectively 80% and 94%.

Since none of the available literature included the role of operational context within the applied maintenance policies, this research newly applies the role of operational context by adding the parameter(s) defined during diagnose phase (multi-parameter model). Adding operational context parameters to the model CTBM (OC), and considering flexible intervals, proves that an extra reduction of 10%-25% is accomplished compared to CTBM. In order to apply the parameters of the operational context to UBM, and in order to apply the proposed model USBM, failure data regarding PM tasks is required. By gathering PM failure data, relevant operational contexts can be identified by using the Kruskal-Wallis test, subsequent R&D should perform the FMEA in each identified operational context. Such that, an operational context has a unique threshold value for all failure parameters. Within this phase the second deliverable was conducted: a single-parameter and multiple-parameter model to improve the current periodic maintenance policy.

Next to the proposed model, redesigning the business process is recommended in order to fully apply the proposed model and to further improve the current maintenance policy periodically. The redesign of the business process shows how the new business process should look like and how the new maintenance policy should be implemented (RQ 8). By continuously monitoring the failure data of PM, failure parameter(s) and their threshold values that indicate failure of the related sub-system can be defined/updated on a periodic basis. By doing this, intervals of PM tasks can be adapted periodically on the actual data in each relevant operational context. By defining/updating the failure parameter(s) and determining threshold value(s) for each operational context, the proposed model could potentially be applied in order to determine the (dynamic) interval for each PM task in each relevant operational context. Within this phase the third deliverable was conducted: a redesign of the business process in order to improve the current PM policy periodically.

Conclusion from business perspective

This research showed how the current periodic maintenance policy of IXR-devices can be improved with data (data-driven) and considering the operational context (context-awareness). The main causes of the large number of labor hours were lack of insights in data, lack of evidence regarding effectivity of PM and a static PM manual. By gaining insights in the data, a solution to the large number of labor hours was given. Implementing the proposed model and the redesign of the business process lead to a PM checklist that is dynamic rather than static. Adopting the intervals proposed by the model for PM task J and K, could already reduce the total number of labor hours with 2%-6% within one maintenance cycle. The lack of evidence regarding effectivity of PM can be solved by adding an extra checklist to the work orders of FSE, as proposed in Section 6.3. By applying the proposed model to the remaining PM tasks and implementing the redesign of the business process it is expected that the number of labor hours will even reduce further, since PM task J and K are causing a relatively small number of labor hours. Therefore, the aim of reducing the large number of labor hours associated with PM is achieved.

Conclusion from scientific perspective

From a scientific perspective, much is written about maintenance policies to improve preventive maintenance. To the best of our knowledge, none of the policies CTBM, UBM or USBM is yet applied in the field of healthcare (or a related field), where safety factors cause small, conservative, intervals of preventive maintenance. Safety factors are reflected in the conservative threshold value of a failure parameter. Also, worst-case behavior of a device is assumed without considering the data on actual usage. Next to that, the application of operational context on the mentioned maintenance policies is still, to a large extent, unexplored, even though it is generally highlighted that each reliable performance test has to be executed in all possible contexts according to Moubray (1999). Therefore, this research newly applied the operational context parameters to the current existing maintenance policies (context-awareness). This research showed that applying contextual factor 'operational context' to CTBM, and considering flexible PM visits, causes a significant reduction in the number of executions of PM task J and K, since it removes variation in actual usage.

Next to the application of operational context to CTBM, other relevant knowledge of the mentioned policies was obtained. None of the policies were tested in the field of healthcare. This research, importantly, strongly suggests that especially within the field of healthcare, these policies can achieve major improvements. Because of the safety factors there is a tendency to use conservative time intervals of maintenance, even though in some cases safety can be ensured while stretching intervals.

CTBM

Tinga (2010) mentioned that CTBM is not suitable for devices with high variation in usage among each other because a lot of assumptions have to be made regarding the usage of a device. This would result in the selection of a conservative interval such that an acceptable failure rate would be obtained. This research acquired new insights on the usage of CTBM saying that it is also suitable for devices that do not have a similar usage rate to each other, as long as actual worst-case usage of the devices is used in order to determine the CTBM interval for the whole fleet. Our case is highly suitable for applying this, because of high (conservative) safety factors that are considered when determining the threshold values of the failure parameters and worst-case behavior of a device is assumed (without looking at the actual worst-case usage).

High safety factors are distinctive within healthcare, which makes CTBM suitable in the field of healthcare or related fields. For that, it is required that threshold values of the failure parameters are defined (with experience-based models but preferably physical-based models) and that the failure parameters are monitored (not necessary daily monitored). Of course, an exceeding rate for the threshold value of the failure parameter(s) should be defined (can be 0%) to ensure safety of the device. A trade-off between costs of failures and preventive maintenance costs should define the exceeding rate (for example, an exceeding rate of 5% when component is not critical). Again, an exceeding rate of 5% does not necessarily mean that 5% of the sub-systems fails. It only suggests that 5% of sub-systems is more likely to fail. Obviously, a higher exceeding rate causes a larger periodic interval of the whole fleet.

UBM/USBM

This research confirms that UBM removes the uncertainty in usage and therefore execution of maintenance can be determined more accurate per device. Within this research, UBM causes 80%-94% reduction in number of executions of PM tasks J and K. This large reduction is caused by (1) safety factors that cause conservative threshold values of failure parameters and (2) assumed worst-case usage without considering actual usage data. Therefore, it is likely that UBM can achieve (major) improvements in the field of healthcare (or related fields). Consequently, other companies are able to use this part of the model if the following requirements are met: the threshold values of the failure parameters are defined (with experience-based models but preferably physical-based models) and the failure parameters are monitored daily. An add on to UBM is that, theoretically, variation in degradation rates could be removed by determining the threshold value of failure parameter(s) in each relevant operation context. The Kruskal-Wallis test is suitable for determining the relevant operational context factors.

Unfortunately, this research was not able to test the performance of USBM. Since USBM takes the severity of each usage type into account, it is expected that determining the time of executing maintenance will be even more accurate. Nevertheless, in the hypothetical case USBM is used, a similar add on would apply as was the case for UBM: determining the threshold value of failure parameter(s) in each relevant operation context.

For applying the proposed model at other companies, some important requirements have been set. If one of these is missing, and in order to fully apply the proposed model, it is recommended to implement the redesign of the business process according to this research. By acquiring failure data of preventive maintenance, failure parameters and threshold values of failure parameters can be determined. Subsequently, the proposed model can be applied to an extent preferable.

7.2 Limitations

Even though this research has promising results, multiple cases can potentially be improved.

Firstly, there is a potential bias towards certain outcomes because this research was conducted by a single case study, researchers individual perception could be dominant (Almeida, Faria, & Queirós, 2017). This bias is minimized by using objective methods (regulative cycle and CRISP-DM), verifying analysis and results by other researchers and showing the reproducibility of the research.

Secondly, within this research the maintenance policies CTBM, UBM and UBSM were considered, but the application of condition-based maintenance (CBM) might result in even more savings. Especially, when failure parameters of sub-system cannot be defined based on PM failure data, an investment of sensors might be useful to predict failures applying CBM. Especially when considering that scientific literature shows that CBM potentially leads to the most optimized PM schedule (Tinga, 2010).

Thirdly, within this research it is assumed that the state of a sub-system is 'new' after execution of the related PM task, such that that the cumulated failure parameter value starts from zero again until the threshold value is (almost) reached. However, this is not very realistic, since other factors, such as age, might influence the degradation of a sub-system.

Fourthly, within this research it is assumed that degradation of a sub-system is not influenced by degradation of other sub-systems, even though this might be the case in practice.

Fifthly, another factor that has affected the feasibility of this research is the availability of failure data regarding PM. Therefore, the proposed model could not be fully tested. However, this research describes how PM failure data should be applied within the model when it is gathered.

7.3 Future research

This research concluded that a periodic maintenance policy can be improved with the proposed model that uses data (data-driven) and considers the operational context (context-awareness). However, some dimensions of the model are still unexplored.

Firstly, the proposed model was only applied to two out of the seven PM tasks upon which is scoped. Future research regarding the remaining PM tasks is required to indicate the real impact of applying this model to those PM tasks. In case a particular sub-system has failure parameters that are yet unknown, failure data regarding the related PM task should be gathered first.

Secondly, by implementing the proposed redesign of the business process, lots of failure data on each PM task will be gathered. The failure data firstly needs to be investigated in order to check whether failure parameter(s) do influence degradation of a sub-system and what their threshold values are. It might occur that it is not possible to predict the need for PM based on the monitored parameter(s). In such cases, the current conservative interval should be applied and possibilities for other maintenance policies should be explored, such as CBM.

Thirdly, since this research was conducted by a single case study, potential bias was created. In order to prevent this, objective validation methods can check or improve the current work. An example would be to reproduce the research partly again. Thus, performing the same analysis and apply the same proposed model to PM task J and check whether the results are similar.

Fourthly, the performance of the proposed model for prognostics could be further improved if the prognostics method would be more accurate, since a single linear regression model was used with fixed parameter values. The performance of the proposed model could be further improved if the prognostics method would be more accurate. Perhaps other applicable will perform better relative to the current method.

Fifthly, the role of other factors that might influence degradation to a sub-system but are not part of the failure parameters should be explored. Now, failure parameters are based on FMEA. But, for example, the role of age is not considered even though this might play a role in degradation of the sub-system.

Finally, within this research the role of degrading sub-systems that cause degradation to other sub-systems is not taken into account. It would be relevant to investigate if particular sub-systems do influence the rate of degradation to other sub-systems.

References

Aken, J. v., Berends, H., & Bij, H. v. (2007). *Problem solving in organizations.* Groningen, The Netherlands: Cambridge University Press.

Almeida, F., Faria, D., & Queirós, A. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, (pp. 369-387).

Altay, N., & Green, W. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research 175*, 475-493.

Amant, K. (2017). The cultural context of care in international communication design: a heuristic for addressing usability in international health and medical communication. *Communication Design Quarterly*, 62-70.

Arts, J. (2016). *Maintenance modeling and optimization.* Eindhoven: Eindhoven University of Technology.

Carl-Anders Johansson, V. S. (2014). Context Driven Remaining Useful Life Estimation. Luleå, Sweden: Luleå University of Technology.

Chapman, P., Clinton, J., & Kerber, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide.*

Chen Xiongzi, Y. J. (2011). Remaining Useful Life Prognostic Estimation for Aircraft. *(School of Automation Science and Electrical Engineering, Beihang University*, 5.

Cui, L., Loh, H., & Xie, M. (2004). Sequential inspection strategy for multiple systems under availability requirement. *European Journal of Operational Research*, 170-177.

Fraser, K. (1994). An overview of health and usage monitoring systems (HUMS) for military helicopters. *Melbourne: Defence Science and Technology Organisation*.

Galar, D., Thadari, A., Catelani, M., & Ciani, L. (2015). Context awareness for maintenance decision making: A diagnosis and prognosis approach. *Measurement*, 137-150.

Hunt, S., & Hebden, G. (2001). Validation of the Eurofigther typhoon structural health and usage monitoring system. *Smart Materials and Structures*, 10.

Jardine, A., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing 20*, 1483-1510.

Kim, K., & Kuo, W. (2009). Optimal burn-in for maximizing reliability of repairable non-series systems. *European Journal of Operational Research 193*, 140-151.

Kumar, U., Ahmadi, A., Verma, A., & Varde, P. (2016). *Current trend in reliability, availability, maintainability and safety: an industry perspective.* Springer.

Lee, J., Lapira, E., Bagheri, B., & Kao, H.-A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 38-41.

Lieven, F. (2006). Damage prognosis: the future of structural health monitoring. *Mathematical, physical and engineering sciences*, 365.

Lu, X. (2013). *Service Coding (internal document).* Eindhoven: Philips.

Mir, R., & Ahmad, S. (2018). Advances in Statistical Forecasting Methods: An Overview. *New Delhi Publishers*, 40-52.

Molent, L. (1998). A unified approach to fatigue usage monitoring of fighter aircraft based on F/A-18 experience. *Proceedings of ICAS98*, 25.

Moubray, J. (1999). *Reliability-Centered Maintenance.* The Aladon Network.

Müller, M., Staudacher, S., Friedl, W., Köhler, H., & Weißschuh, M. (2010). Probabilistic engine maintenance modeling for varying environmental and operating conditions. *Proceedings of the ASME Turbo Expo*, 6.

N.V., K. P. (2020). *Philips*. Opgehaald van https://www.philips.com/a-w/about/company/our-heritage.html

Peng, Y., & Dong, M. (2010). A prognosis method using age-dependent hidden semi-markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 25.

Rasul, C. L. (2005). *Project management handbook.* Project Management Institute.

Rosemann, M., Recker, J., & Flender, C. (2008). Contextualization of business processes. *International Journal of Business Process Integration and Management*, 3.

Schmidt, B., & Wang, L. (2015). Predictive maintenance: literature review and future trends. *Proceedings of the 25th International Conference on Flexible Automation and Intelligent Manufacturing*, 232-239.

Si, X.-S., Wang, W., Zhou, D.-H., & Chang-Hua, H. (2010). Remaining useful life estimation - A review on the statistical data driven approaches. *European Journal of Operational Research*, 1-14.

Smulders, M. (2018). *Business & Management modellen.* Noordhoff Uitgevers.

Stojcetovic, B., Živče, Š., Lazarević, D., & Marjanović, D. (2015). Application of the pareto analysis in project management. *International Quality Conference*, 3.

Tappert, S.-H., & Charles, D. (2009). A Genetic Algorithm for Constructing Compact Binary Decision Trees. *Computer Science Department, Pace University*, 13.

Tinga, T., Wubben, J., Tiddens, W., Wortmann, J., & Gaalman, G. (2019). Dynamic maintenance based on functional usage profiles. 18.

Wang, W., & Zhang, W. (2008). An asset residual life prediction model based on expert judgements. *European Journal of Operationa Research*, 496-505.

Zulfigar, A., & Bhaskar, C. (2018). Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*, 60.