MASTER

Towards more efficient NMSC healthcare

Decision space

Jagt, J.D.

Award date:
2023

Towards more efficient NMSC healthcare: Decision space

J.D. Jagt (2022)

Mentors

Name	Role
M. Razavian	First assessor
M. Veta	Second assessor
L. Genga	Third assessor
G. Krekels	MohsA mentor

In partial fulfillment of the requirements for the degree of
Master of Science in Innovation Management

Eindhoven, April 4, 2023

Preface

Skin cancer is a global public health problem, with non-melanoma skin cancer (NMSC) being the most common type of skin cancer. The rising incidence of NMSC places a significant burden on healthcare systems worldwide, calling for the development of innovative solutions to increase the efficiency of NMSC healthcare. Artificial intelligence (AI) has shown great potential in improving healthcare efficiency and patient outcomes. In this thesis, a decision model is developed to assist decision makers in optimizing the efficiency of NMSC healthcare through AI.

I would like to express my sincere gratitude to my mentors, Maryam Razavian from the Eindhoven University of Technology and Gertruud Krekels from MohsA, for their invaluable guidance and support throughout this thesis. Their expertise and feedback have been instrumental in shaping this work.

I would also like to thank my parents for providing me with the brains and means for studying, as well as my friends for their unwavering support and encouragement throughout this journey.

This thesis is dedicated to all those who have been affected by skin cancer, and to the healthcare professionals who tirelessly work towards improving patient outcomes.

Abstract

With skin cancer being one the most occurring types of cancer worldwide and a rapid increase in yearly incidences, there is an urgent need for more efficient skin cancer healthcare. Patients that have fallen victim to a form of (non-melanoma) skin cancer in the past, will often develop new lesions in the future resulting in the diagnosis of 'chronic' skin cancer. Redesigning the healthcare process to create more efficiency, impacts a wide variety of stakeholders. Due to the large differences in perspective of these stakeholders, and the high impact healthcare decisions have on the well-being of patients, the need for a decision tool to guide decision makers towards finding an appropriate solution for this redesign is more relevant than ever. This thesis sets out to develop such a decision tool by following the design science methodology in context of a real-world skin clinic. The combination of a literature review uncovering the state-of-the-art/state-of-practice AI solutions with a stakeholder analysis and business process modeling approach, allow for the design of an artefact in the form of a question, option, criteria (QOC) model. This QOC model should guide decision makers on how to redesign the current healthcare process through providing a simple but complete overview of the decision making activities and perspectives that need to be taken into consideration. Validation of the artefact through a focus group study showed that, despite a few shortcoming, the model assisted in the decision making activities of the participants. A simple overview of the broad range of perspectives and criteria that need to be considered allows for substantiated decision making. A business process simulation of the current process and redesigned process was performed to test whether additional insights could be given for the options of the decision model. The results of these simulations show that simple simulations using estimates based on real-world data can provide additional insights for deciding on certain elements of the decision model by providing estimates of the workload and throughput time differences. Further recommendations are provided to further examine the support of decision making tools in the healthcare context, and to continue the road towards a clinician-AI synergy.

Executive Summary

INTRODUCTION - This thesis focuses on how decision making regarding the implementation of an AI-enabled process for non-melanoma skin cancer (NMSC) healthcare can be supported by usage of a decision model. Due to an increasing burden on NMSC healthcare as a result of ageing and tanning behaviour, there is the need for a redesign of the current process in order to increase the healthcare efficiency and remain feasible over time. The research follows the design science methodology and aims to provide the decision makers tasked with redesigning the process with a tool that can aid them in their decision making activities. For the development of this tool, questions are answered regarding the AI applications that are currently available, the stakeholders must be considered and what their requirements are, how the current process looks like, and possible directions on how this process can be enhanced.

CONTEXT - The project was performed in the context of a skin clinic based in Eindhoven named Mohs Academy (MohsA). At this clinic, patients are treated for a variety of skin problems, predominantly consisting out of skin cancer. Employees at MohsA have performed research into making healthcare more efficient through a one-stop-shop system that allows patients to receive a consult, diagnosis, and treatment on the same day. This is enabled through their offering of Mohs Micrographic Surgery (MMS). Their goal is to further improve the healthcare process, not only for their own clinic but as an example for others as well.

LITERATURE REVIEW - The exploration of the different AI applications that are currently available was done through a systematic literature review. The research method of snowballing was applied to retrieve the list of papers to be reviewed in the literature review. During snowballing, a start set of papers is selected through performing a search query, followed by multiple iterations of forward and backward snowballing during which citing papers and references are considered for inclusion in the literature review respectively. This method resulted in a total of 27 papers that were included for the literature review. From the review became evident that there is a lot of potential in AI applications in terms of performance (measured in terms of sensitivity and specificity). Multiple papers showed that the algorithms used for the classification of skin cancer lesions were able to outperform dermatologists. Both the usage of patient clinical data and images as a data source for training these algorithms has proven effective in the past. Research performed at MohsA confirmed these claims by showing its potential in their clinical setting.

Even though these AI applications have a high potential, there are pitfalls that need to be addressed before these applications can be implemented in a real-world scenario. Through the nature in which many AI algorithms are designed, it is difficult to substantiate how an AI arrives to a conclusion. This poses a problem as the decisions that are made in healthcare may have a big impact on a patient's health, and are thus important to be justifiable. Another aspect is the threat of a biased AI as a result of incomplete or unbalanced training data. In case not all lesion types are covered by the training data, an AI might be unable to correctly classify all cases. In case structural errors are made in e.g. dark-skinned versus fair-skinned patients, this may lead to discrimination. Additionally, due to the vast amount of data required for developing a high performance model, patient data needs to be shared amongst developers which raises privacy concerns.

PROBLEM INVESTIGATION - Development of a decision model requires extensive knowledge on the goals and requirements of the relevant stakeholders, and the context in which the decision model must be applied. This knowledge was acquired through performing a stakeholder analysis during which interviews were held with the problem owner and stakeholder representatives. Additionally, a document analysis was performed of the transcripts of earlier interviews with stakeholders performed at MohsA. Based on these analyses, the six stakeholder types: MohsA, dermatologist, insurance company, data supplier, patient, and data analyst were defined. Through further exploration, the stakeholder interests were defined that shaped the requirements for the development of the decision model. The stakeholder interest were described using the goal, question, metric (GQM) method to provide a simple overview and further guide towards the decision space requirements. The analysis of the current process was done through business process modeling. Process identification and process discovery identified the tasks, resources, and hand-offs of the process. Based on this, the current business process was modeled using the business process modeling notation of OMG [65].

ARTEFACT DESIGN - The design of the artefact was initiated by defining a set of draft questions with options and criteria to act as a basis for the further definition of the decision model based on the literature and GQM method. Using this draft to develop an interview guide, semi-structured interviews were held with representatives of the stakeholder groups to acquire their perspective on these questions

and build relations between questions, options, and criteria (QOC). Through qualitative data analysis (QDA) and conceptual validation, the most important elements were retrieved to build the final QOC model. This model consists out of six questions that concern 1) what process to enhance, 2) how to enhance this process, 3) what type of data to collect, 4) who analyzes the data, 5) how the data will be collected, and 6) how patients will be informed on the outcome of the analysis. For each of these elements relevant options and criteria have been described including a rationale of the relation between them. The final QOC model was represented as a tree-like structure displaying the depth of the decisions that are being made.

VALIDATION - The decision model was validated through performing a focus-group interview with four representatives of MohsA. These included two dermatologists, the problem owner/board member, and a skin therapist. During the focus-group interview, the QOC model was discussed and observations were made on how the model was perceived and what influence it had on the decision making activities of the participants. It became evident that the QOC model assisted with the decision making in multiple ways. First off, the model provided context for the decisions that were to be made. The participants of the focus group showed understanding of the problem and could place it in their daily work activities. Additionally, the model clearly contributed to the decision making by providing the decision makers with the variety of decisions that need to be made and the differences in perspective. Decision makers were able to provide a line of argument for their decisions using information from the decision model.

At the same time, the focus group unveiled downsides of the model. First off, a few missing elements were identified. These elements were considered important for the completeness of the decision model. Furthermore, some participants showed difficulty with interpretation of certain elements. Therefore, the model should be clarified further to adjust for this ambiguity. Based on these aspects, a redesign of the QOC model was proposed that should improve the validity of the model.

TESTING - To provide additional grounds for the selection of a specific option in the decision model, the contribution of a business process simulation was tested. The goal was to examine whether a simple simulation tool can complement the decision model by giving insights in the values of a set of performance indicators (resource consumption and cycle time). The simulation tested the scenario where an AI will be implemented into the healthcare process to enhance the efficiency through establishing a probability diagnosis that can either be used for the prioritization of appointments or the scheduling of direct treatments. By using data provided by MohsA, 6 different simulations were performed. These consisted out of a lower-bound, upper-bound and average scenario of the current and redesigned process. Based on the resulting performance indicators, the conclusion could be drawn that the redesigned process does indeed achieve the goal of reducing the workload of dermatologists and reducing the throughput time of patients (simultaneously indicating a decrease in number of patient visits). However, due to the limited nature of these simulation, more extensive tests must be done to validate these results.

Contents

List of Figures

List of Tables

1 | Introduction

This thesis is considered part of the research into how to overcome the growing need of healthcare in nonmelanoma skin cancer treatment. The goal is to make use of new technologies in the area of artificial intelligence to increase the efficiency of the current healthcare. The next subsection will describe the reasons for the increase in burden on the healthcare, afterwards the company for which the research will be performed is described and more information on the background of the research is presented. Finally, the problem and its relevance for literature is described, and the research design is outlined.

1.1 | Epidemiology

Cancer is a well-known disease that has had an impact on many lives. In 2019, cancer was the disease that posed the highest clinical, social and economic burden when measured in cause-specific Disability-Adjusted Life Years (DALYs) compared to all other human diseases. [53] Skin cancer is one of the most occurring types of cancers. According to the World Health Organisation one in three diagnosed cancers is skin cancer. Within skin cancer, nonmelanoma skin cancer (NMSC) comprises by far the largest part. To illustrate, in 2017 the World Health Organisation documented around 1.2 million cases of nonmelanoma skin cancer anually versus 300.000 cases of melanoma [22]. Nonmelanoma skin cancer (NMSC) can be divided in two major types, namely basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). From these two types of NMSC, BCC is most prominent with a BCC:SCC ratio between 1:1 and 10:1 depending on multiple individual characteristics like ethnic group, sex, and population. The reason for this wide range in ratio is that most European countries do not distinguish between different types of NMSC in their national cancer registry data [9].

The incidences of NMSC are increasing rapidly. Records have shown an annual increase of around 3-7& over the past decades [63, 25]. In the Netherlands, the number of cases has quadrupled since 1990 and this growth threatens to increase only further (between a factor of 2 and 5) [80]. Two main reasons can be identified for this exponential increase: an ageing society and a change in behaviour leading to increased exposure to UV radiation [2].

The reason that ageing has a large effect on skin cancer incidences is that NMSC mostly affects patients older then 65 due to an ageing skin and a change in their immune system [70, 84]. Especially actinic keratosis (AK), a form of pre-cancer that can develop into SCC (see section 1.3.2), has been reported to increase in prevalence with age [31]. Due to an ageing world population, the incidences of skin cancer increase with it. Between 2015 and 2050 the proportion of the world's population above 60 years old is likely to double from 12% to 24% [66]. On top of this, NMSC can be regarded as a chronic disease as an increasing number of patients develop multiple tumors during a lifetime. Thus when elderly start developing skin cancer or AK chances are that they will be subject to regular checks for the rest of their lives [26].

The second main reason for the exponential increase in skin cancer incidences is an increase in exposure to UV radiation. UV exposure has been classified an upmost carcinogen factor by the International Agency for Research on Cancer (IARC) [21]. The increase of exposure follows from multiple factors. One of these factors is climate change. As the ozone layer depletes the UV-radiation intensifies on terrestrial area while at the same time, there is less overall cloud coverage and more high-temperature areas emerge where people tend to go into the sun more often [16]. In addition to the climate change, a change in tanning behavior has taken place over the past decades. As a tanned skin has become a trait of attractiveness, people tend to ignore health information and neglect to use UV protection [37]. Only recently, more research has been published on the harmful effects of UV radiation. Since then, campaigns have started to increase awareness. However, as it takes decades before the effects of UV radiation turn into skin cancer, a decline in incidences due to successful campaigns will not show for a while.

1.2 | Company Description

This thesis will be based on a case from the company MohsA, a skin clinic with three locations in the Netherlands. This thesis will only be related to the MohsA location based in Eindhoven. The name MohsA is short for Mohs Academy. Mohs relates to Mohs Micrographic Surgery (MMS), a clinical procedure that is exercised at this clinic. Since MohsA has an ASDS certificate they are a licensed academy to train Mohs surgeons. In addition, they attribute to the quality of dermatolo-oncology and dermato-surgery by e.g.

initiating research in dermato-oncology and being in the board of the European Society for Micrographic Surgery [55].

At the time of writing the team at MohsA consists out of 25 employees that include dermatologists, skin therapists, doctor assistants, nurses, and administrative employees. The team of employees treat all different skin conditions of which skin cancer is a large part [55]. From 2016 till October 2022 MohsA has seen a little over 18.000 individual patients according to their healthcare records. One of MohsA's unique offerings is their one-stop-shop treatment for NMSC. This treatment focuses on patients that have had NMSC before (the so-called, chronic patients) and sets out to perform diagnosis, treatment advise, and, if possible, treatment, on the same day that there is found suspicion for skin cancer [56].

1.3 | Background Information

As explained before the two major types of nonmelanoma skin cancer are BCC and SCC. There are some other NMSCs like Merkel cell carcinoma, sebaceous carcinoma, and apocrine adenocarcinom. However, as about 99% of NMSCs can be contributed to BCC or SCC these other types will not be considered in this thesis [9]. Actinic keratosis is actually not a form of NMSC but a precancer, however, as this is a common type of lesion that has the risk of developing into skin cancer, it is included in this subsection.

1.3.1 | Basal Cell Carcinoma

BCC is the most common type of skin cancer. Basal cells are in the top layer of the skin and shed as new ones form. When the DNA in these cells is damaged by UV radiation, this can result in uncontrolled growth. This growth is a slow process, however, if left untreated BCC can grow wide and deep into the skin. This may have disfiguring effects and result in damage to tissue and bone. The dangers of BCC are moderate as BCC barely spreads further then the initial site the tumor started and, when acted at an early stage, can be treated well. If you leave a BCC untreated for a while, chances that the lesion will recur increase [23].

To detect a BCC it is possible to check whether the lesion checks the boxes below that are outlined by the Skin Cancer Foundation [23]. Often 2 or more of these symptoms can be recognized in a BCC. It is however noted that BCCs have many different appearances and may thus look different than described in this list.

- An open sore that does not heal, and may bleed, ooze or crust. The sore might persist for weeks, or appear to heal and then come back.
- A reddish patch or irritated area, on the face, chest, shoulder, arm or leg that may crust, itch, hurt or cause no discomfort.
- A shiny bump or nodule that is pearly or clear, pink, red or white. The bump can also be tan, black or brown, especially in people of color, and can be mistaken for a normal mole.
- A small pink growth with a slightly raised, rolled edge and a crusted indentation in the center that may develop tiny surface blood vessels over time.
- A scar-like area that is flat white, yellow or waxy in color. The skin appears shiny and taut, often with poorly defined borders. This warning sign may indicate an invasive BCC.

1.3.2 | Actinic Keratosis

The most common form of precancer is AK. AK develops on skin that has been damaged by long-term exposure to UV radiation. AK in itself is harmless, however, between 5-10% of these lesions develop into SCC, which can be invasive. Therefore, it is necessary that AK is treated timely. Another aspect of AK is that when a person has developed a case of AK, it is likely that they will develop more AKs in the future. Due to this constant monitoring and visits to a dermatologist are necessary. To identify AK it is easier to feel the surface opposed to looking for odd spots since the surface of an AK is often rough and may feel dry. In addition, they can be painful and sensitive, or be itchy with a burning sensation.

The early treatment of AK is relatively easy. Minor surgical procedures like cryosurgery, where the physician applies liquid nitrogen to the lesion for it to freeze and fall of, can be efficient and quick. For more widespread AK it is possible to apply photodynamic treatment. This treatment uses a light sensitive topical agent that is then exposed to red or blue light to kill the cancer cells. The advantage of this treatment is that no healthy tissue will be damaged [23].

1.3.3 | Squamous Cell Carcinoma

SCC is, behind BCC, the second most common form of skin cancer. Just like basal cells, squamous cells occur in the top layer of the skin and shed as new ones form. Just like AK, SCC can form when long-term exposure to UV radiation damages the DNA structure in the cells. Besides that, AK can develop into SCC as explained in section 1.3.2. When treated early, most SCCs are curable. However, when left to grow, lesions can become dangerous, grow into deeper layers, and spread into other parts of the body. In some cases SCCs can even be deadly.

Due to the fact that SCC develops as a result of exposure to UV radiation, lesions are often found on sun epxosed areas like the neck, face, scalp, lips, ears, shoulders, back of the hands, and forearms. In addition, SCC may develop in areas of skin injury like scars and sores. Characteristics of SCC may include thick scaly patches that easily bleed. In addition, it can be spots that are similar to warts or open sores that do not fully heal [23].

1.3.4 | Healthcare Costs and Projections

The total cost for skin cancer (benign, premalignant and malignant) in 2017 was €465 million in the Netherlands, which comes down to around €804 per patient. Compared to €723 per patient in 2007 this shows a large increase [63]. From these total costs, around €188 million is related to dermatological skin tumor management of which in turn 40% is comprised of benign and premalignant tumors [63]. Noels et al. [63] projected that the total costs for benign, premalignant and malignant skin cancer would increase up to €1.35 billion in 2030. A similar trend was found in the US by Guy Jr et al. [32] where there was a 126.2% increase in annual costs for skin cancer treatment between the period of 2002-2006 and 2007-2011. In addition, they showed that office-based visits made up nearly three quarters of the total $4.8 billion dollars in NMSC costs during 2007-2011.

1.3.5 | Artificial Intelligence in NMSC Healthcare

To enhance the efficiency of healthcare, research has been done in the application of artificial intelligence (AI). AI in healthcare can be divided into two branches: a virtual, and a physical branch. The virtual branch relates to machine learning applications, mathematical algorithms that improve through learning from experience. The physical branch includes physical objects like surgical robots or elderly care helpers [34]. This thesis will solemnly focus on the application of machine learning, thus the virtual branch. In specific, two applications of AI will be the topic of interest during this thesis. These are a logistic regression model for the analysis of a patient questionnaire and a deep learning algorithm for image classification. These two applications stem from earlier research performed at MohsA in these areas [26, 39].

Previous research has shown the potential of deep learning in diagnosing different types of (non)melanoma skin cancer [17]. High accuracies were achieved using pictures and a deep learning architecture [1, 83, 39, 79, 17, 58]. The research of Hoepel [39] was performed at MohsA. The goal of this research was to create a model that successfully classifies NMSC. This goal derived from the fact that most research in this area is done for melanoma instead of NMSC. Hoepel [39] built a deep learning architecture that has been tested with data from MohsA. The results looked promising in the sense that a high classification accuracy was reached with regards to the type of skin condition. Some remarks were made on the fact that a higher accuracy could be gained by using extra techniques for inclusion of (more) heterogenous data and the addition of extra features. This has however not yet been tested nor implemented.

The other application of AI that was developed and tested at MohsA by Geer-Rutten [26] is the usage of a logistic regression model trained on a patient questionnaire. Two models were created, one to test for AK and one to test for BCC. These models used the answers of a patient questionnaire as an input and selected multiple characteristics that defined AK and BCC for the regression variables. Both models displayed high accuracy. In addition, they compared the result of the model to the diagnosis a nurse made that filled out the same questionnaire and diagnosed the lesion. The accuracy of the nurse and the model were quite similar. Finally, when the diagnosis of the nurse was implemented in the model, it was shown that the accuracy increased even further concluding that the model can serve as an additive for the nurse's diagnosis.

1.4 | Problem Definition

Applications that have the potential of releasing the burden on the NMSC healthcare have proven to be effective in showing high accuracies in NMSC detection. However, as of yet, these novelties have not been implemented in everyday NMSC healthcare. The components are there but in the current situation, the integration of the components into the real-life setting has not yet been achieved. To our knowledge, no literature has of yet described the usage of an AI-enabled process in the healthcare of NMSC patients. In addition, as explained in section 1.1, NMSC can be regarded as a chronic disease with the largest risk group being elderly. Due to the ageing of society, in combination with the lifelong checks this group will undergo if they develop NMSC, there is a need for a redesign of the current healthcare process that incorporates AI to increase efficiency. The increased efficiency should ensure that, with the increase in burden, the healthcare remains economically feasible. Due to the exponential increase in cases, a small percentage of increased efficiency could already have a major impact in the long run.

The healthcare is subject to multiple stakeholders. In healthcare these stakeholders are generally defined as "a person or group with a vested interest in a particular clinical decision and the evidence that supports that decision" [8, 12]. In order to create an effective redesign, the stakeholders should be known and their requirements incorporated. In context of the problem three stakeholders can be identified: the care-receiver (patients), the care-giver (dermatologists, nurses, skin-therapists), and the payers (insurance companies). The project will set out to define stakeholder requirements to be considered during the (re)design process.

Within MohsA, past research displays their willingness to increase the efficiency of their healthcare [26, 39]. However, these research topics have not yet included the socio-economic and business process side of the issue. Therefore, they are lacking a process that could help them achieve their goals. A change strategy is necessary to provide direction for the implementation of the process and to know what stakeholders are affected [48]. Required is a sound process implementation that will endure and result in the desirable outcomes. To finalize, the following problem statement can be formulated:

MohsA is missing an AI-enabled process in the NMSC healthcare, which is necessary for achieving an increase in efficiency of chronic NMSC patient treatment.

1.5 | Research Gap

There has been extensive research in the area of applying deep learning algorithms in skin cancer diagnostics for both melanoma [1, 17, 58, 79] as nonmelanoma skin cancer [17, 39, 58, 79, 83]. Similarly, there has been research on detection models using risk factors and skin cancer characteristics that were collected through patient questionnaires [26, 28, 72]. Recent research also identified the patient perspective towards AI in skin cancer diagnostics [46, 61] showing that patients are amenable to using AI as a support tool for dermatologists in decision making.

However, even though there is enough evidence that the tools might be beneficial in real-life, there is only a limited number of healthcare providers that implement it. Research on this topic is absent, the effects of implementing AI on the efficiency of healthcare have not been documented. Xiang et al. [96] outlined key issues in the implementation of AI in healthcare, under which the integration into an existing clinical workflow. Further research on how to do so has to our knowledge not been published. This research sets out to fill that gap by focusing on mapping the current healthcare process and defining the decision space through which a new AI-enabled healthcare process can be designed.

1.6 | Research Design

The goal of this research is to identify what questions, options, and criteria are present when redesigning an NMSC healthcare process to increase efficiency through incorporating AI diagnostics. To perform this research, the design science methodology will be followed. As the problem aims to make use of user data uploaded through the internet to be processed and used by the software systems, we can define this system as an information system. On that note, the design science methodology as described by Wieringa [93] will be followed. This methodology follows the design cycle that iterates through 4 phases: problem investigation, artefact design, artefact validation and artefact implementation [93]. In figure 1.1 the phases and research topics per phase can be found. The time span of this thesis is limited, and therefore, the implementation of the artefact will not be considered. This is indicated in the figure by the arrow in the middle of the circle.

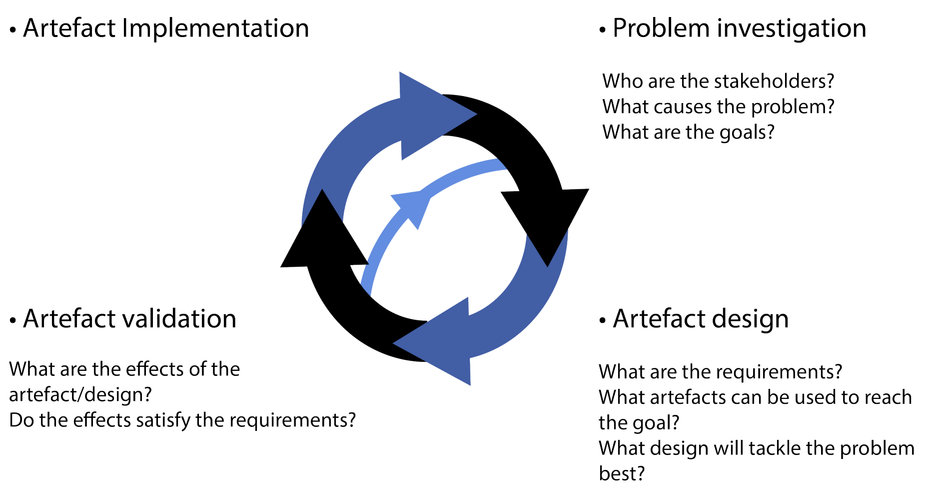


Figure 1.1: Design Cycle (based on Wieringa [93])

As described in section 1.4 the goal is to create an AI-enabled process in the healthcare of (chronic) patients. The process must be usable by the stakeholders at MohsA as well as by the patients on the 'customer' side of the process. As the process may include multiple new technologies in the redesign, a description of the interaction between these technologies is sought to describe the best possible solutions for the stakeholders. Additionally, the goal is to research which decisions may lead to the largest increase in efficiency of healthcare. This information can be summarized in terms of an artefact, goals, and requirements as displayed in table 1.1.

Table 1.1: Design problem

Design Problem	
Problem statement	MohsA is missing a process for the implementation of AI in their healthcare, which is necessary for achieving an increase in efficiency of chronic NMSC patient treatment.
Artefact	A decision space that identifies the questions, options and criteria surrounding the implementation of an AI-enabled process
Requirements	Provide context and insight in the decision making activities, provide best practices for an AI-enabled process
Goals	Decrease number of patient visits, reduce costs, maintain patient satisfaction

The research topics defined in this research cycle can be specified for this current research into research questions. These questions are divided per phase of the design cycle. During the problem investigation phase the following research questions will be answered:

Q_1) **What state-of-practice/state-of-the-art AI solutions are available to increase the healthcare efficiency and how do they interact?**

Q_2) **Who are the stakeholders and what are their requirements?**

Q_3) **What is the current process?**

The answers to these questions will describe the problem in more detail and create a solution space of stakeholders, goals, and solution directions. To answer these questions, first off the process must be outlined and the question must be asked, who are the stakeholders in this process. From literature, stakeholders in similar healthcare processes can be identified. In order to learn their requirements, interviews must be held with the relevant stakeholders. As the design is in a early stage of development, the requirements of the most prominent stakeholders will be prioritized. Furthermore, documentation of the healthcare clinic will be used to describe the current status, the related costs, and to provide evidence for the problem. As the problem involves the implementation of AI in the healthcare process, this phase will already go into the available possibilities.

The second phase, artefact design, will focus on what artefact to use, what requirements this artefact must satisfy, and how this artefact can lead to a treatment of the problem. For this the following research question is defined:

Q_4) **How to (re)design the process so that the efficiency increases?**

The context for this question is formed by the answers found in the problem definition phase. The answer to this solution follows from the design requirements and goals. From literature, processes from similar cases can be studied to present possible solution directions. Also, the users of the process must be included in the design of the process in order to generate usefulness, willingness, and adaptation. For this the interviews from the problem definition phase can offer guidance to present an implementation guide of the artefact.

Finally, the phase of artefact validation focuses on whether the artefact satisfies the requirements and what the effects are of the artefact. In other words, is the artefact implementable as a treatment of the problem? No separate research questions are formulated for this phase. The validation of the artefact is done by using input from the users on the design and by doing alpha tests with the users to test whether unforeseen interaction occurs, and what the effects are of the implementation.

2 | Literature Review

This section delves deeper into the literature around AI implementations in healthcare. First off, the methodology of the systematic literature review is explained. Secondly, the papers that are included in the systematic literature review are listed and finally the relevant literature is extracted.

2.1 | Methodology

The implementation of artificial intelligence in medicine is not new. The earliest research on computer systems aiding physicians in diagnosing patients ranges back to 1960 [77]. However, it is not until more recently that research into AI has seen a rapid increase. From 2017 to 2019 the number of papers indexed on PubMed with 'AI' in the title has increased tenfold. At the same time, this increased interest is also seen outside of the academic world, with a doubling of the number of AI endowed devices that are approved by the FDA [6]. This section delves into the more recent research of AI applications in medicine by means of a literature review. To perform this literature review, the method of snowballing is used to select articles for the review. Snowballing is a systematic literature method that makes use of the reference list and citations of an paper as a research method for finding relevant papers [94]. This approach starts by defining a set of papers with which the snowballing process is started (the start set). This start set is build up out of papers that are relevant to the research question and are found through searching in academic databases. Consequently, iterating between forward and backward snowballing is performed during which you look at the reference list of papers and at papers citing in order to identify new papers. The papers that are identified are measured against inclusion and exclusion criteria in order to decide whether they will be included in the final set of papers.

The start set is defined through performing two search queries in Google Scholar. Google Scholar is chosen in order to prevent any bias towards certain publishers. Since the subject of AI is very broad, it is chosen to perform two separate search queries. The first search query explores AI in healthcare in general to find papers related to all disciplines. Applications in other disciplines may lead to insights that can assist in exploring opportunities in the MohsA context. The second search query delves deeper into AI diagnostics that are specifically designed for skin cancer. In this way current research into the topic can be evaluated. The used search queries were `"artificial intelligence" AND applications AND (medicine OR healthcare)` and `("artificial intelligence" OR "deep learning" OR "machine learning") AND "skin cancer" AND (diagnostics OR diagnosis OR detection)`. From the results, papers have been selected based on relevance. To select the papers, first the title was read followed by the abstract of the paper. If this was not sufficient the introduction and conclusion were read. If there was still doubt about the inclusion, the entire paper was reviewed. In order to generate a diverse start set, different publishers and years have been chosen for examination. Papers that reference to another paper in the start set are excluded as these will be examined during the snowballing iterations. In addition to the papers that were found during the search queries in Google Scholar, 2 papers (**S5** & **S6**) were added to the start set that are relevant to the literature as they concern research into the topic that has been performed at MohsA. It is important that these papers are added as this ensures a complete and varied basis of literature. The chosen start set can be found below.

S1. Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10):719–731, 2018

S2. Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017

S3. Andre GC Pacheco and Renato A Krohling. The impact of patient clinical information on automated skin cancer detection. *Computers in biology and medicine*, 116:103545, 2020

S4. Claire M Felmingham, Nikki R Adler, Zongyuan Ge, Rachael L Morton, Monika Janda, and Victoria J Mar. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. *American Journal of Clinical Dermatology*, 22(2):233–242, 2021

S5. Mike CM van Zon, José D van der Waa, Mitko Veta, and Gertruud AM Krekels. Whole-slide margin control through deep learning in mohs micrographic surgery for basal cell carcinoma. *Experimental Dermatology*, 30(5):733–738, 2021

S6. Simone van der Geer-Rutten, P Ad M Kleingeld, Chris CP Snijders, Frank JCH Rinkens, Geert AE Jansen, HA Martino Neumann, and Gertruud AM Krekels. Development of a non-melanoma skin cancer detection model. *Dermatology*, 230(2):161–169, 2015

In order to decide on what papers to include and exclude during the snowballing iterations, a list of inclusion and exclusion criteria was established. These inclusion and exclusion criteria have been summarized below. The main criterion is that articles should present relevant findings in the field of AI in healthcare. Similarly, these findings should explore new peculiarities or strengthen the argument of already present peculiarities. With this, articles that are too far of topic or replicate other articles, are excluded.

- Only articles that present relevant findings regarding artificial intelligence implementations in healthcare are included
- Only articles that provide new insights or elaborate on the rationale of present insights are included
- Only articles that are written in English are included
- Only articles that are electronically available are included

Starting with the papers in the start set, forward and backward snowballing was performed. Papers that cite papers in the start set were examined through Google scholar as well as papers in the reference list of the papers in the start set. From titles that deemed interesting the abstract was read to check whether a paper would be included. If this was not enough, the context of the citation or entire paper was read. This resulted in the following papers:

P1. David Ben-Israel, W. Bradley Jacobs, Steve Casha, Stefan Lang, Won Hyung A. Ryu, Madeleine de Lotbiniere-Bassett, and David W. Cadotte. The impact of machine learning on patient care: A systematic review. *Artificial Intelligence in Medicine*, 103:101785, 2020. ISSN 0933-3657. doi: https://doi.org/10.1016/j.artmed.2019.101785

P2. Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1):30–36, 2019

P3. Rahul Kapoor, Stephen P Walters, and Lama A Al-Aswad. The current state of artificial intelligence in ophthalmology. *Survey of ophthalmology*, 64(2):233–240, 2019

P4. Camille Nebeker, John Torous, and Rebecca J Bartlett Ellis. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC medicine*, 17(1):1–7, 2019

P5. Shamim Nemati, Andre Holder, Fereshteh Razmi, Matthew D Stanley, Gari D Clifford, and Timothy G Buchman. An interpretable machine learning model for accurate prediction of sepsis in the icu. *Critical care medicine*, 46(4):547, 2018

P6. IS Stafford, M Kellermann, E Mossotto, Robert Mark Beattie, Ben D MacArthur, and Sarah Ennis. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ digital medicine*, 3(1):1–11, 2020

P7. Manu Goyal, Thomas Knackstedt, Shaofeng Yan, and Saeed Hassanpour. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine*, 127:104065, 2020

P8. Julia Höhn, Eva Krieghoff-Henning, Tanja B Jutzi, Christof von Kalle, Jochen S Utikal, Friedegund Meier, Frank F Gellrich, Sarah Hobelsberger, Axel Hauschild, Justin G Schlager, et al. Combining cnn-based histologic whole slide image analysis and patient data to improve skin cancer classification. *European Journal of Cancer*, 149:94–101, 2021

P9. Breno Krohling, Pedro BC Castro, Andre GC Pacheco, and Renato A Krohling. A smartphone based application for skin cancer classification using deep learning with clinical images and lesion information. *arXiv preprint arXiv:2104.14353*, 2021

P10. Andre G.C. Pacheco, Gustavo R. Lima, Amanda S. Salomão, Breno Krohling, Igor P. Biral, Gabriel G. de Angelo, Fábio C.R. Alves Jr, José G.M. Esgario, Alana C. Simora, Pedro B.C. Castro, Felipe B. Rodrigues, Patricia H.L. Frasson, Renato A. Krohling, Helder Knidel, Maria C.S. Santos, Rachel B. do Espírito Santo, Telma L.S.G. Macedo, Tania R.P. Canuto, and Luíz F.S. de Barros. Pad-ufes-20:

A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, 32:106221, 2020. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2020.106221

P11. Andre GC Pacheco and Renato A Krohling. An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, 25(9):3554–3563, 2021

P12. Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017

P13. Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2):94, 2019

P14. Titus J Brinker, Achim Hekler, Alexander H Enk, Joachim Klode, Axel Hauschild, Carola Berking, Bastian Schilling, Sebastian Haferkamp, Dirk Schadendorf, Tim Holland-Letz, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113:47–54, 2019

P15. David Roffman, Gregory Hart, Michael Girardi, Christine J Ko, and Jun Deng. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Scientific reports*, 8(1): 1–7, 2018

P16. Y Fujisawa, Y Otomo, Y Ogata, Y Nakamura, R Fujita, Y Ishitsuka, R Watanabe, N Okiyama, K Ohara, and M Fujimoto. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *British Journal of Dermatology*, 180(2):373–381, 2019

P17. S Van der Geer, M Frunt, HL Romero, NP Dellaert, MH Jansen-Vullers, TBJ Demeyere, HAM Neumann, and GAM Krekels. One-stop-shop treatment for basal cell carcinoma, part of a new disease management strategy. *Journal of the European Academy of Dermatology and Venereology*, 26(9):1154–1157, 2012

From the papers that were included, another iteration of snowballing was performed where papers that were referenced in the new papers that deemed relevant were considered for inclusion. This iteration was performed in order to make sure that information in the included papers that was not originally from these papers was cited correctly. This resulted in the following papers:

P18. Holger A Haenssle, Christine Fink, Roland Schneiderbauer, Ferdinand Toberer, Timo Buhl, Andreas Blum, A Kalloo, A Ben Hadj Hassen, Luc Thomas, A Enk, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology*, 29(8):1836–1842, 2018

P19. Noel CF Codella, Q-B Nguyen, Sharath Pankanti, David A Gutman, Brian Helba, Allan C Halpern, and John R Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4/5):5–1, 2017

P20. Roman C Maron, Michael Weichenthal, Jochen S Utikal, Achim Hekler, Carola Berking, Axel Hauschild, Alexander H Enk, Sebastian Haferkamp, Joachim Klode, Dirk Schadendorf, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *European Journal of Cancer*, 119:57–65, 2019

P21. Christian Murray, D Sivajohanathan, TP Hanna, S Bradshaw, N Solish, B Moran, R Hekkenberg, AC Wei, and T Petrella. Patient indications for mohs micrographic surgery: A clinical practice guideline. *Current Oncology*, 26(1):94–99, 2019

P22. David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. *The Lancet Digital Health*, 2021

P23. Tamar Sharon. *Data-driven decision making, AI and the Googlization of health research*. Pavia: Pavia University Press, 2020

P24. Caroline A Nelson, Lourdes Maria Pérez-Chada, Andrew Creadore, Sara Jiayang Li, Kelly Lo, Priya Manjaly, Ashley Bahareh Pournamdari, Elizabeth Tkachenko, John S Barbieri, Justin M Ko, et al.

Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA dermatology*, 156(5):501–512, 2020

P25. Tanja B Jutzi, Eva I Krieghoff-Henning, Tim Holland-Letz, Jochen Sven Utikal, Axel Hauschild, Dirk Schadendorf, Wiebke Sondermann, Stefan Fröhling, Achim Hekler, Max Schmitt, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Frontiers in medicine*, 7:233, 2020

The start set combined with the selected papers resulted in a total of 31 papers to be included in the literature review. From these papers the relevant aspects have been extracted and summarized below.

2.2 | AI Applications in Healthcare

Through the new technologies built with AI support, a discussion has been fueled on whether these technologies might deliver groundbreaking innovations in healthcare that could even replace physicians [44]. A paper by Jiang et al. [44] reflects on the AI applications in medicine by investigating the motivation behind the drive for AI, analysed data types, AI mechanisms to generate results, and disease types that AI, at the moment of writing, was commonly tackling. The primary drive behind AI research in healthcare is explained to be the fact that AI can use high volume data to generate insights based on learned features that may assist in medical practice. In addition, through the way these algorithms are designed, it is possible to constantly improve by learning from new cases. The major diseases that literature predominantly concentrates around are cancer, nervous system disease, and cardiovascular disease. A reason for the fact that research is mainly done into these diseases is first and foremost that these diseases have a high mortality rate and thus have a high interest from many researchers. Early diagnosis of these diseases may elevate chances of survival significantly. On top of that, improvement in the early detection can be achieved through better analysis of images, genetics or electronic medical records (EMR), which are all within the strength of AI [44]. Even so, implementations in other areas of healthcare are researched as well. These areas are all closely related in the sense that the diagnostic methods used for these diseases is based on one or more of the aforementioned analysis methods. For example, research in the diagnosis of autoimmune diseases makes use of AI in genetics and MRI analysis [81], and research into the diagnosis of ophthalmology diseases explores image classification algorithms and AI in EMR analysis [47]. Similarly, in the prediction of sepsis at the ICU, research is done into how an AI can make use of the live tracking of a high resolution time series of vital signs to predict the occurrence of sepsis within the upcoming 4-12 hours [62]. The performance of these algorithms is often measured against similar metrics. These include specificity, sensitivity, accuracy, and area under curve (AUC) or area under receiver operating characteristic curve (AUROC) [81]. The specificity and sensitivity of an algorithm relate to the true negative and true positive rate respectively, the accuracy relates to the ratio of predictions being correct, and the AUC/AUROC incorporates both the specificity and sensitivity in the sense that if the AUC/AUROC is 1, both specificity and sensitivity are 1 (all predictions correct) [59]. This section will further set out to evaluate the different methods of AI diagnosis and analyse the strengths and weaknesses as well as its threats and opportunities.

2.2.1 | Image classification

Some of the AI applications within skin cancer healthcare have been briefly mentioned in section 1.3.5. When looking through published research, it is evident that most papers delve into image classification tasks when it comes to AI in skin cancer diagnosis. In a paper by Goyal et al. [29], research into three areas of image classification have been listed and discussed. These include classification of dermoscopic images (images captured with a digital single-lens reflex or smartphone and an illumination system), clinical images (images taken with any type of camera), and histopathology images (microscopic images of tissue). The difference between these three areas thus comes from the type of input they use, even though all imagery, these images look very different hence different features need to be extracted by the AI. In order to develop an algorithm for a specific type of images, a specific database must be chosen for the training of the algorithm. Databases may contain different types of images and classify these into different categories [29]. In case an algorithm is trained against a database of dermoscopic images that are classified in four classes of lesion, algorithm will only function with dermoscopic images and output one of these four classes. It is thus important to decide on what algorithm to design and what database can be used for that goal. Many researchers make use of datasets of the International Skin Imaging Collaboration (ISIC), who provide open-source datasets containing images with ground-truth diagnosis and clinical metadata [41].

Brinker et al. [5] tested an algorithm that merely classifies dermoscopic images as being either melanoma or nevi. As an evaluation metric they used sensitivity and specificity. The advantage of using only two classes is the fact that they were able to set an operational value that altered the sensitivity and specificity in the sense that they were able to tweak the algorithm in such a way that the desirable trade-off between sensitivity and specificity could be adapted to the requirements. As in a screening setting a high sensitivity is required, the algorithm could be adapted to achieve this goal [5]. The designed algorithm achieved a sensitivity and specificity of 84.2% and 69.2% respectively outperforming dermatologists who achieved 74.1% and 60.0% respectively. Similar results were achieved by Codella et al. [10] and Haenssle et al. [33] who also used dermoscopic images as input and a binary classifier of melanoma vs nevi as output. In both studies, an AI was tested against a group of dermatologists and in both studies the AI reached a higher sensitivity for a similar level of specificity. Opposed to these binary classification algorithms, researchers have examined the possibilities of multiclass classification. Maron et al. [52] performed a study where the classification was both done by deciding on whether a lesion was benign or malignant, and afterwards classifying as either nevi, melanoma, SCC and AK, BCC or benign keratosis. Due to this, a more specified output is given that is closer to the reality of a dermatologist screening [52]. The output is again measured by looking at the sensitivity and specificity of the predictions and taking the mean value. In the case of multiclass classification, these results must be interpreted slightly different as a lower specificity is expected for a binary classification due to the fact that in the multiclass classification there is the one-vs-all approach. This means that there will be a class imbalance where the number of true positives and false negatives is much lower then the number of true negatives and false positives [52].

In applications where patients themselves are to make pictures, dermoscopic images are not available. Instead, these images from regular cellphones are called clinical images and may have different backgrounds, lighting, angles, and colours that give a new dimension to the analysis [29]. Applications that can make use of clinical images have the potential of being used for self-screening by patients that have suspicious lesions as is done by e.g. SkinVision [51]. In a paper by Esteva et al. [19] an algorithm is trained against a dataset including around 120000 of these clinical images, which was an extensive amount at that time. Two binary classifications were performed, benign nevi vs malignant melanoma, and keratinocyte carcinomas versus benign seborrheic keratoses. Both classification tasks performed on par with dermatologists in terms of accuracy (around 55% in both classifications). In addition to this, they showed that their partitioning algorithm performed better then a multiclass algorithm did. As such, if a diagnosis can be partitioned into either one of two cases this would deem more accurate. In another research performed by Fujisawa et al. [24], an algorithm was build that could classify clinical images into 14 different diagnoses consisting of both benign and malignant lesions with a dataset of only around 5000 images. The algorithm outperformed dermatologists with an accuracy of 74.5% versus 59.7% showing that even with a smaller dataset and more classes, accurate results can be achieved. Their research also tested the accuarcy of classifying the lesions into only 4 diagnoses: benign epithelial or melanocytic, or malignant epithelial or melanocytic. Similar to the results of Esteva et al. [19], this resulted in an even higher accuracy of 92.4% versus 85.3%.

The final area of classification tasks is done within histopathological images. These are images that are made during the microscopic evaluation of a tissue biopsy [29]. In specific, the images used as input of the algorithm are digitalized conventional histology glass slides, or in other words whole-slide images (WSI) [90]. An application where the fast detection of cancer tissue in WSIs is specifically useful is during Mohs micrographic surgery (MMS). MMS is a surgery procedure during which the excision is immediately histologically mapped to check for remaining cancer tissue [57]. This procedure is costly due to long operation times, therefore, an accelerated check of the WSIs may reduce these costs [90]. van Zon et al. [90] performed research into the accuracy of an AI in the detection of BCC in these WSIs. Their algorithm was created using 171 WSIs and achieved an AUC of 0.90. It must be stated that the ground-truth of the slides used for learning was established by only one dermatologist and might thus contain flaws. As a tissue biopsy is only taken when a certain type of malignant lesion is expected, the used algorithms can be used as a binary classifier to check whether the lesions is cancerous or not. As such, research is e.g. done into algorithms that distinguish between melanoma and nevi [38, 97], or between BCC and normal tissue [45, 90].

2.2.2 | Patient clinical data

The papers that are discussed above display high potential with metric scores often similar or better than the those of dermatologists. However, these papers only examine the applications when it comes to image classification without including patients' medical history and clinical data [29]. In addition, the papers do mention the fact that the AI outperforming dermatologists can be related to the fact that dermatologists normally diagnose a lesion based not only on a single image but on clinical data as well. As such, Pacheco and Krohling [67] performed research into the effects of the inclusion of patients' clinical features in classification algorithms and found that the number of correct predictions increased while the number of incorrect predictions decreased. Their paper presented a new database containing images from smartphones combined with patient clinical data containing eight clinical features: the patient's age, the part of the body where the lesion is located, if the lesion itches, bleeds or has bled, hurts, has recently increased, has changed its pattern, and if it has an elevation [67]. Followingly, these clinical features are aggregated into an AI resulting in an increase of approximately 7% of the balanced accuracy.

After the paper by Pacheco and Krohling [67], more research was published that followed the same principle of including patient data into the AI algorithm. A reason for an increase of research into this topic is that in 2020 a new database (PAD-UFES-20) was released that contains images taken by smartphone cameras combined with patient data. With this database new algorithms can be trained against these parameters [69]. As such, Pacheco and Krohling [68] enhanced their classification even further while Krohling et al. [49] elaborated on the same research by creating a smartphone app to assist clinicians with the lesion classification. With this app, clinicians were able to upload an image taken with the smartphone and enter details on the aforementioned 8 lesion characteristics. With a balanced accuracy and recall of 85.5% and 96.42% respectively in their best scenario, the method shows great potential.

In a different approach, Höhn et al. [40] used patient clinical data together with the results of their AI for the classification of their WSIs. Their research did in general not show a significant increase in accuracy when patient data was included in the algorithm. However, in case the image classification algorithm output score was low, the accuracy could be enhanced by replacing the image classification algorithm with the patient data classification algorithm. This method was introduced as the naive strategy [40]. That it is possible to create an AI that is able to correctly diagnose a lesion based on solely patient clinical data had been proven before [27, 75], research performed by Geer-Rutten et al. [27] made use of patient data collected through surveys at MohsA. This data was very extensive and contained both questions related to patient characteristics (e.g. skin type, history of UV exposure) and lesion characteristics. Two algorithms, using the patient characteristics as input, were able to classify BCC versus normal tissue, and AK versus normal tissue. The input of the questionnaire answers was bundled into 10 and 11 characteristics respectively. The resulting predictors and predictions can be found in appendix A. From the results in the appendix can be seen that 91.4% respectively 84.7% of the cases were predicted correct for BCC and AK when a cut-off value of 0.5 was used, which was similar to the results achieved by trained nurses. An advantage of the algorithm is that the cut-off value can be altered to balance the trade-off between false negatives and false positives.

2.2.3 | Discussion

The papers above all show the increasing potential of AI applications in healthcare, however, there are limitations to these results. One of these limitations arisen by He et al. [36] is the lack of transparency of the dataset. This relates to the quality of the dataset used for training and whether this dataset can be analyzed by others. A dataset that contains flaws can be detrimental as the performance of an AI is closer related to the quality of the dataset than to the tuning of parameters [29]. In addition, even though the performance of the papers discussed above is high, these algorithms are only tested in a controlled setting where the testing data is often uniform to the training data. This enhances the performance as the training and testing data have similar characteristics, research that uses testing data from a different cohort often displays a drop in accuracy [92]. As such, many papers do mention that real-world trials are needed to test the performance against the full spectrum of lesions that are encountered in practice [19, 33, 49, 52]. Besides this, it is important that the dataset is a true representation of society. As skin cancer is most common in fair-skinned people, datasets tend to over represent these skin types neglecting darker skin types. On top of that, populations that have an advanced data-infrastructure can contribute more easily towards datasets [92]. As a result, an algorithm may develop an unintended bias towards certain skin types or ethnicities [20]. In case the dataset is not transparent, such biases might be difficult to detect.

2.3 | Ethics of AI implementations

Besides the fact that AI paves the way for interesting opportunities, there are ethical points to consider. First off, there is the issue that was already shortly touched upon in the discussion above, namely the possibility of an unintended bias [20]. As mentioned before, the AI may develop a bias towards certain racial groups due to an unbalanced training set, however, there is also the possibility of an automation bias in medical practitioners [20]. This automation bias refers to the cognitive bias where practitioners over-rely on the outcome of an AI. In such a case, a faulty prediction from an AI may mislead practitioners resulting in wrongful diagnoses [85]. Because of this, the placing of AI within the healthcare process should be taken into consideration. Janda and Soyer [43] explains that in order to accelerate the triage process an AI is best placed before the decision of a clinician. However, the possibility over-reliance on AI and regulatory issues must be prevented.

Secondly, there is the fact that parts of the decision making process of an AI may be unclear to clinicians and be considered as a black-box [29, 74]. As a result, decisions that are made cannot be explained, from this you can reason that if you are unable to explain whether a result is valid, you cannot be certain. Especially in a setting like healthcare where it often is essential that a decision is correct, clinicians need to be able to justify the diagnosis and treatment plan [36]. With the increasing number of papers published into AI applications [6], a similar trend can be seen for explainable AI research [87]. Explainable AI (or XAI) refers to algorithms that can be interpreted and understood by humans and may include methods that allow visual explanation by saliency maps or through the inspection of learned features and feature importance [87]. These methods might help not only clinicians in diagnosing but may also provide patients with a satisfactory explanation. Besides the fact that the algorithms might need to provide methods to explain the made decisions, the clinicians that use these algorithms need to be educated in the aspects of AI. Especially its benefits and limitations should be known in order to be able to interpret the produced results in context [36].

With the implication of bias and threat of errors comes the matter of liability. When decisions are made by an AI, the question arises who is liable in case of malpractice. From the introduction of an AI may result a shift in feeling of responsibility of a clinician [36]. Within the current line of practice, it might in some cases already be unclear who is responsible, therefore it is essential that the liability is determined before AI is used in practice [74]. However, can you hold the developer of the AI accountable, or the provider of the dataset, in most cases the accountability shifts away from the care giver towards a more abstract entity [36]. In addition, these entities might not be able to cover these liabilities with insurance while medical professionals can be covered against malplractice [98]. Insurers would need to create policies on AI in order to set legislation towards such possible scenarios.

From the point of view of patients, other concerns can be raised. Nelson et al. [61] performed a study in which these concerns were quantitatively analysed. The main aspects raised by patients follow from the fear of losing human interaction and replacing this by a machine. As a result, most patients (94%) desired a symbioses between AI and a clinician that could extract the best of both. Similar results were found by Jutzi et al. [46] where most patients (94%) desired the AI as a support system for clinicians while less than half (41%) of the patients would use an AI as stand alone system. Again, one of the main concerns that was outed was the impersonal interaction of an AI, being unable to act on human emotion.

Finally, as the technological advancements in AI are data-driven, requiring large high quality datasets preferably from heterogeneous sources that are shared with each other, new privacy concerns are raised [36]. Sharon [78] went into detail on these concerns, especially in the context of existing big data companies that are proceeding towards healthcare data. They define a new research model as the Googlization of health research (GHR) in which companies like Facebook, Google, or Amazon, expand their research towards healthcare in order to exploit their data driven knowledge in a new market. Their purpose is to integrate and combine data from different contexts for future research meaning sharing of more and more patient data. Unforeseen use of patient might thus become more difficult to regulate. With that comes the fact that anonymized data can be re-identified with increasing ease losing the sense of anonymity [78]. As these big tech concerns and data scientist that are responsible for the creation and training of the algorithms often may not have received clear ethics education, a lack of awareness of the privacy concerns may emerge [60]. Therefore it is a necessity to establish regulation that guarantees the privacy of patients.

2.4 | State-of-practice Solutions

In order to decide on what (AI) technologies are best implementable in the context of MohsA, the possibilities that are available and can already be practiced are be mapped. Applications that are already present at MohsA and can be helpful for the implementation of the new process will be discussed. First off, the research performed by Geer-Rutten et al. [27] as discussed in section 2.2.2 was executed at MohsA. The questionnaire used in this research is available and can be amended to the current needs.

As stated in section 1.3.5, in an earlier research performed by Hoepel [39] at MohsA, several deep learning/machine learning models were developed for the classification of NMSC. These models aimed to classify images of lesions into one of 6 categories: Verrucae, ID, Naevi, SCC, BCC & AK. The results showed an accuracy of 67.5% for the best model when combining all 6 categories. The accuracies of naevi and AK were the highest, correctly classifying 79.5% and 85.4% respectively for the best model. These accuracies are promising and show similar results to papers discussed in section 2.2.1. The model itself used images from a database at MohsA together with another database of the internet in the training process. The fact that the model was trained with MohsA images enhances the chances that new images will be comparable to the training and validation data in terms of lighting, quality, angle etc.

When disregarding AI, another available practice at MohsA can be seen as skin therapists that are trained in NMSC classification. These skin therapists are educated in order to be able to classify skin lesions as NMSC. Similarly, they know what lesion characteristics can be related to NMSC and which not. In this way, the skin therapist can act as the AI classifier by using available images and data from the questionnaire in order to diagnose the lesion. Earlier research has already shown the effectiveness of educated nurses in the screening of lesions [64]. Geer-Rutten [26] researched the accuracy with which these skin therapists were able to classify AK and BCC. This resulted in an overall percentage of correct classifications of 88.3% and 90.4% respectively (appendix A). This accuracy is thus higher then the accuracy of the model made by Hoepel [39], however, it must be noted that this accuracy was only tested for AK and BCC and not for other lesion types while the model by Hoepel [39] classified the lesions into one of 6 categories.

Finally, Van der Geer et al. [86] proposed a one-stop-shop treatment as part of a disease management strategy that should reduce the number of visits patients have to undergo. This one-stop-shop treatment was developed at MohsA and is currently being executed. The idea behind the one-stop-shop treatment is that the consult, diagnosis, and treatment are all executed on the same day. One of the treatments that facilitates the one-stop-shop model is Mohs Micrographic Surgery (MMS). In case an excision needs to be done, a pathologist checks whether the excision has removed all cancer cells which may take some time. During MMS, this is checked directly during the procedure so that there can be immediate action if there is malignant tissue remaining. As an addition to this procedure, the research performed by van Zon et al. [90] has the potential to accelerate the check of the excised tissue.

2.5 | SWOT analysis

In order to summarize the literature in terms of the potential opportunities and concerns, a SWOT analysis is performed. In a SWOT analysis the strengths, weaknesses, opportunities, and threats are analysed. These have been extracted from the literature above. The goal of this analysis is to function as a guide towards proposing relevant questions, options, and criteria. The different aspects of the SWOT are explained and mapped to literature in table 2.1. In addition, a diagram visualizing a summary of the SWOT analysis can be found in figure 2.1

2.5.1 | Strengths

The strengths of AI flow from its ability to process high volume data in order (to assist) to diagnose lesions [44]. This accelerates the process in order to overcome the increasing pressure on the healthcare (section 1.1). On top of this, the strength of an AI can be found in its performance. Many papers have displayed that an AI is able to perform on par with diagnoses performed by dermatologists or even excel them [5, 19, 10, 33, 24]. Especially when combining patient clinical data with images, which is data dermatologists regular have access to, AI has proven to perform according to the standards. In addition, AI has the possibility of improving constantly due to new training data being fed into the system and learning from made mistakes [44]. Finally, with AI being able to provide accurate results based on clinical images, applications can be built that are available remotely to patients increasing the accessibility [29].

2.5.2 | Weaknesses

Opposed to its strengths, there are weaknesses to consider. The first weakness of AI is the dependence on quality input. An AI that is built and trained for a specific type of input, requires this same input in the same format in order to perform. Due to this, an AI loses flexibility and might lose performance in case the input differs from the training data [29]. A similar weakness is the fact that the performance itself is only as good as the dataset used for training. This means that even though the algorithm might

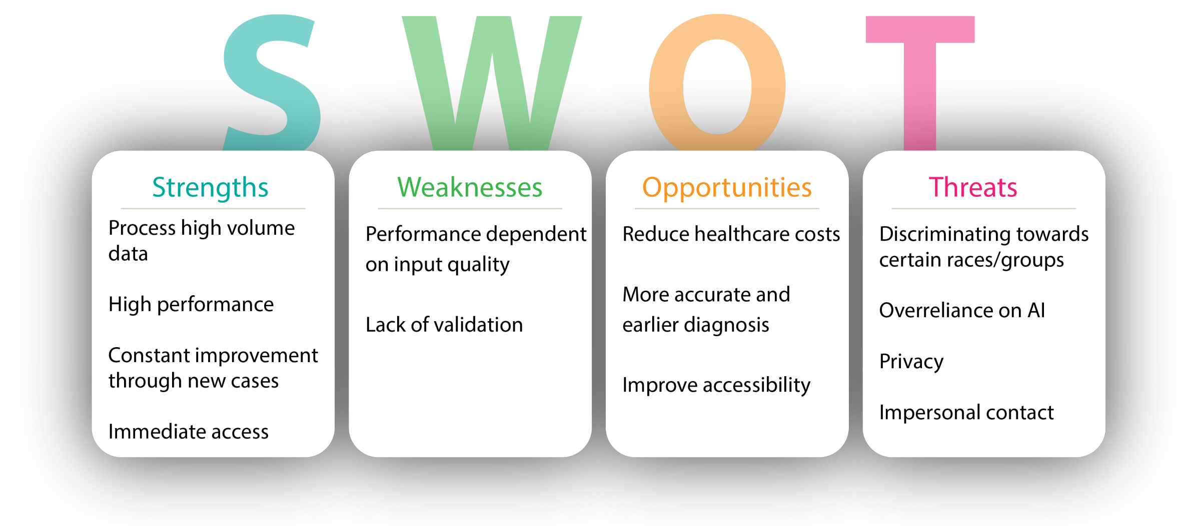


Figure 2.1: SWOT diagram of the AI applications found in literature

Table 2.1: Elements of SWOT analysis

S/W/O/T	Explanation	Litarature
Strength	Ability to process high volume data	Jiang et al. [44]
Strength	Performance in terms of accuracy that exceeds or is on par with dermatologists	Brinker et al. [5], Codella et al. [10], Esteva et al. [19], Fujisawa et al. [24], Haenssle et al. [33]
Strength	Ability to be accessible through remote applications	Goyal et al. [29], Krohling et al. [49]
Weakness	Performance depends on quality and type of data used for training	Goyal et al. [29], Wen et al. [92]
Weakness	Little real-world validation supporting evidence from controlled settings	Esteva et al. [19], Felmingham et al. [20], Goyal et al. [29], Yu et al. [98]
Opportunity	Reduced costs by more efficient treatment through assisted decision making	Tschandl et al. [85], Roffman et al. [75]
Opportunity	Earlier detection of malignant lesions	Kapoor et al. [47], Tschandl et al. [85]
Opportunity	Better accessibility to healthcare	Esteva et al. [19]
Threat	Risk of discriminating against races or groups	Wen et al. [92]
Threat	Over-reliance on AI outcome due to cognitive bias of dermatologists	Felmingham et al. [20]
Threat	Risk of privacy breach due to large data sharing	He et al. [36], Nebeker et al. [60]
Threat	Loss of personal contact between patient and dermatologist	Jutzi et al. [46]

Figure 2.2: Quality model for software products by ISO/IEC 25000 [42]. The quality indicators that will be used for the evaluation of the final approach of the QOC are indicated by a red border.

be flawless, low quality datasets may deem an AI useless. Furthermore, AI applications in skin cancer healthcare have not yet been validated in practice. Therefore, the performance in a real-world healthcare setting is unknown and might turn out to underperform compared to expectations [98, 29, 19]. The applications are still in early stages of development resulting in a weakness that unforeseen problems might emerge during real-world testing where unusual cases might emerge.

2.5.3 | Opportunities

If AI lives up to its potential, it might potentially reduce healthcare costs as earlier intervention can reduce the complexity and necessity of operations [85]. With an AI as a detection mechanism, dermatologists can work more efficiently [47] and support the dermatologists in decision making [75]. This can help to counteract the increasing burden on the healthcare. Besides this, AI has the opportunity to make healthcare more accessible to people [19]. Especially since AI algorithms are specific to the input data used. In case of other diseases, if enough data is available extra applications can be created by the same algorithm [19].

2.5.4 | Threats

The threats mainly arise from the ethical point of view. First off, due to an unintended bias, AI has the risk of being discriminating against certain races or other types of groups. This threat emerges from training against datasets that are not well balanced and representative of the population, and from the black-box character AI currently possesses [92]. With that, there is the threat that clinicians will rely on the outcome of AI too much losing sense of their own expertise [20]. Another aspect is the risk of loss of privacy. With the need for large datasets and sharing of data in order to improve performance, the privacy of patients might be difficult to ensure [36, 60]. Finally, the implementation of AI may have the effect of making healthcare less personal. For patients this may result in feeling less understood due to no emotional interaction. In addition, patients' trust might reduce as patients express their doubts regarding AI as a stand-alone system [46].

2.6 | Quality model

In order to be able to evaluate the final decision space, it is necessary to specify the criteria against which it will be evaluated. In order to do so, a quality model will be created. A quality model specifies which properties are important for an artefact. ISO/IEC 25000 [42] created a quality model to be used for the evaluation of software product quality. As this thesis focuses on the optimization of information science, this is closely related to the evaluation of a software product. The quality model as described by ISO/IEC 25000 [42] will therefore be used as a reference model. This model includes 8 quality indicators that can be used as a quality evaluation system: Functional Suitability, Performance Efficiency, Compatibility, Usability, Reliability, Security, Maintainability, and Portability. These 8 quality indicators are split up further into subcharacteristics. The quality model of ISO/IEC 25000 [42] can be found in figure 2.2. Not all of the quality indicators and subcharacteristics of the reference model are relevant for the focus of this thesis. The quality indicators that are deemed relevant and will be used for the evaluation based on the literature will be explained below and are identified by a red border in figure 2.2. In addition, a summary of the subcharacteristics with its mapping to literature can be found in table 2.2.

Performance Efficiency
Performance efficiency refers to the effectiveness of the system with regards to the used resources. As the implementation of an AI could enhance the efficiency of the healthcare [47], this quality indicator can be used as an evaluation tool. For this quality indicator the following subcharacteristics are included: **1. Time behaviour**: how do the response and processing times improve with respect to the current scenario. In other words, is the gain in efficiency significant. This gain in efficiency can result from human tasks being fulfilled by AI or speeding up processes [36]. **2. Resource utilization**: how many resources are required in the sense of personnel. AI has the potential of replacing clinicians in their tasks which could increase the efficiency [47, 98]. However, an AI literate workforce might be needed in order to carry out AI related tasks [36]. **3. Capacity**: what are the maximum limits of the system in sense of number of patients that can be processed with the available resources. As AI have the potential of processing high volume data [3, 44], this characteristic evaluates the implications in a real-life scenario.

Security
As the healthcare sector deals with lots of personal details, it is of importance that the security of this information and thus the systems is guaranteed. Especially as the AI is highly dependable on input data [29], the concern of privacy is raised [36]. The quality indicator security evaluates both data integrity and system accessibility. The subcharacteristics that are related to this are: **1. Confidentiality**: the accessibility of data by authorized people. The focus is here on what data is accessible and how easily accessible this is. Only those who are authorized to access the data should be able to access it. **2. Integrity**: in what way does the system prevent unauthorized access to data. As patient data will need to be analyzed, it is of importance for the privacy that this data can only be accessed by authorized personnel. In order to achieve this, the data structures must be designed to protect patients [60]. **3. Accountability**: degree in which actions can be traced back to certain individuals and who is responsible for what task. It should be clear who made what decision to be able to justify a certain approach and manage the liability. By definition of rules and regulations may be identified who is accountable in the unclear scenario where AI supports healthcare services [74]. The subcharacteristics *Non-repudation* and *Authenticity* are not included as these do not evaluate relevant characteristics based on the literature.

Table 2.2: Quality model indicators

Subcharacteristic	Explanation	Literature
Time behaviour	Gain in efficiency	He et al. [36]
Resource utilization	Human labour required to perform healthcare tasks	Kapoor et al. [47], He et al. [36], Yu et al. [98]
Capacity	Possibility to process high volume data	Ben-Israel et al. [3], Jiang et al. [44]
Confidentiality	What data is available to which users of the system	Felmingham et al. [20], He et al. [36]
Integrity	The system should be designed to prevent unauthorized access	Nebeker et al. [60]
Accountability	Regulations with respect to liability of actions	Reddy et al. [74], Yu et al. [98]

2.7 | Questions, Options, Criteria

By evaluating the literature and SWOT analysis it is possible to establish a draft for a question, option, criteria (QOC) model that can shape the decision space. The goal of this decision space is to place the artefact in the space of possibilities and use the model to explain why a certain artefact was chosen from these possibilities [50]. This is done by proposing: questions, which represent the key issues that occur within the decision space; options, which consist out of different approaches to solve these questions; and criteria, against which the approaches will be evaluated in order to select the best outcome [50]. Within a QOC model, multiple levels of depth can be reached by attaining new questions after an option has been selected for a higher level question. The QOC elements can be depicted as shown in figure 2.3. The line between an option and criterion shows a positive assessment whereas the dotted line shows a negative assessment.

The primary question that is raised is *How can AI be applied in the skin cancer health care process to enhance the efficiency?* Based on the literature, several options are available. First off, the current process can be enhanced by accelerating the treatment procedure in the one-stop-shop model as was introduced by

Van der Geer et al. [86]. This can be done by making use of an AI to perform the diagnose in the WSIs in MSS to shorten the total time of operation. In this case, the pathological slides that are obtained during MMS can immediately be evaluated for cancerous tissue instead of having this done by a pathologist. Secondly, an AI can be used in order to assist clinicians to diagnose a lesion as was proposed by Krohling et al. [49]. Through this, inexperienced clinicians can be enabled to accurately diagnose a lesion or difficult cases can be assessed with more certainty. Finally, AI can be used to give an evaluation of risk in the screening process of NMSC [51]. In order to decide on what option to choose, these options are measured against a set of criteria. The first criterion that all papers evaluate AI against is the accuracy. This is often compared to the accuracy of dermatologists in order to state whether the AI performs sufficiently. As such, the criterion can be described as performing on par with dermatologists. In addition, the drive behind most papers is to assist in clinical decision making and be able to process high volume data [44]. This can be translated into a criterion as increasing the efficiency.

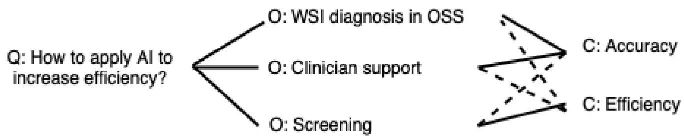


Figure 2.3: QOC representation of the primary question of the design space based on literature

In case the choice is made that AI is applied in screening, the question can be raised whether this screening application should promote self-screening as proposed by Maier et al. [51], replace the clinicians in a clinic or as an extra tool to support clinicians. In this case a criterion arises from the papers by Nelson et al. [61] and Jutzi et al. [46]. These papers show that patients desire that there is a symbiosis between human and AI instead of AI as a stand alone system. Another criterion is the fact that by sharing data, the risk of a privacy breach increases as unforeseen use of patient data may occur [78]. In case the option is selected where the AI support screening by clinicians, an extra question is raised whether the AI should be placed before or after the clinician's decision as was discussed by Janda and Soyer [43]. In this case criteria include the increase in efficiency as well as the accuracy. Janda and Soyer [43] discuss that in over-reliance on AI might be an issue if the AI is placed before the clinician, this is confirmed by Felmingham et al. [20] who explained the possibility of a cognitive bias from humans towards AI.

Another question that arises from the discussion of the papers is how the AI should be trained. There are different publicly available datasets with different sizes and properties out of which the dataset must be chosen that fits the classification task best. The range of databases that can be chosen from differs per task as the database must contain the same type of data that will be used for the input of the algorithm. For the classification of WSIs only databases that contain WSI can be used, however, for screening and diagnostics tasks a choice can be made between dermoscopic and clinical images and/or patient clinical data. For the decision between multiple datatypes, it is again necessary to evaluate against the achievable accuracy. Literature shows that the addition of patient clinical data to images can enhance the achievable accuracy deeming this a well argumented option. Followingly, it must be decided which database to choose. Most papers either use a database from ISIC or the PAD-UFES-20 database when considering both images and patient clinical data. These options must again be evaluated against a set of criteria. The criteria are shaped through the ethical concerns that were raised before. In order to prevent bias, the used dataset must be balanced to be a true representation of society. In addition, the dataset must

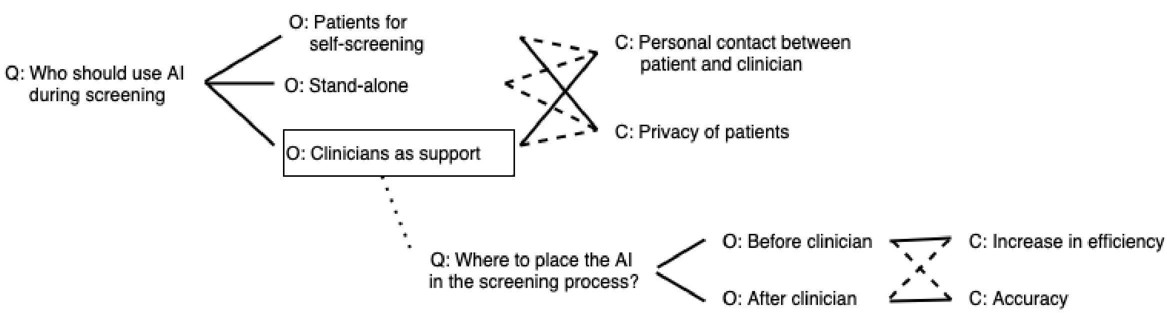


Figure 2.4: QOC representation of the decision space with respect to the screening by AI.

contain the full spectrum of cases that may occur in practice. Finally, as the input data of the algorithm must be the same as the data it is trained against, the dataset should contain datatypes that are flexible. This results in the QOC elements in figure 2.5.

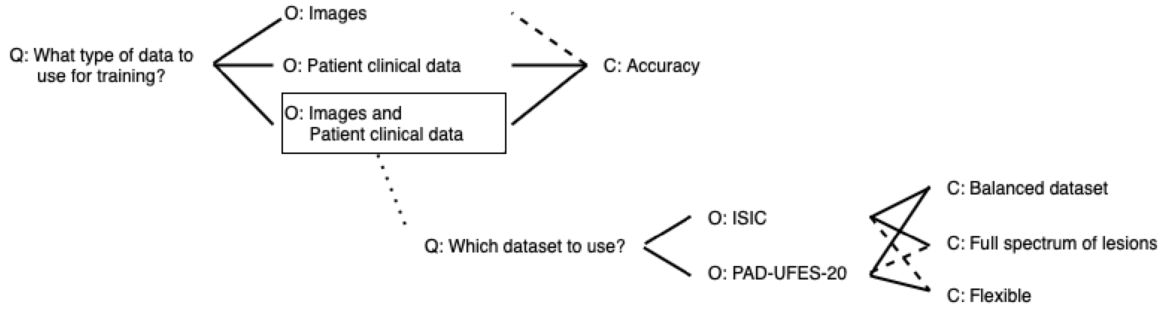


Figure 2.5: QOC representation of design space with respect to the dataset.

3 | Phase I: Problem Investigation

During the problem investigation phase the main goal is to describe the stakeholders, the current situation, and investigate possible treatment options. The literature review that was done before, is used as a backbone during this phase. In order to generate extra insights, interviews were conducted. These interviews served as source of information for the stakeholder analysis and in order to map the current process through business process model notation (BPMN).

3.1 | Interviews

In order to collect the stakeholder information and details on the current situation, interviews will be held. Interviews are one of the most used methods of obtaining qualitative data [15]. The advantages an interview may have in the collection of data come forth from its flexibility. Questions can be rephrased effortlessly in order to provide greater understanding of the data. In addition, an interviewee may uncover information that could not have been stumbled upon within existing literature, or which might have been overlooked by the researcher. Finally, as an interview is a ultimately an interaction between the interviewee and researcher, the attitude of the interviewee during the interview may be observed to present a deeper meaning [76]. In order to conduct a valid interview, it is important for the interviewer not to conduct any bias that may influence the answers of the interviewee. Also, by using techniques like summarizing the given answers, and asking the same question in different contexts, it should be made sure that there are no interpretation errors [76].

Within interviews, a distinction can be made between multiple types of interview. DiCicco-Bloom and Crabtree [15] defined these types as unstructured, semi-structured, and structured interviews. As the goal of the interviews in this research is to explore the problem definition, stakeholders, and solution space, the focus will be on unstructured and semi-structured interviews as structured interviews tend to be more quantitative in nature [15]. In an unstructured interview, the interviewer tries to uncover information on observations by letting a guided conversation unveil truths and let questions emerge over time. As there are no predefined questions, the interview can be seen more as a guided conversation [30]. A semi-structured interview is not as intertwined with observations as an unstructured interview. Within a semi-structured interview, the collection of qualitative data flows is done by the interview itself. The interview is often led by a interview guide that is developer by the interviewer. This guide contains a set of predefined questions and sub-questions related to the topic that can evolve into a more open ended conversation [15].

In an earlier case that was performed at MohsA, the implementation of a smartphone app that made use of patient data to generate a risk analysis was examined. As part of this, semi-structured interviews were held with care takers and insurance companies. There were three interviews with care takers, of which 2 male-female couples and 1 individual woman. In addition, a focus group workshop was held with the care givers at MohsA in which multiple statements were treated with regards to e-health and an NMSC detection app. As this research set out to explore the usage of e-health, in specific an NMSC detection app, in NMSC healthcare, and thus touches closely upon the topic in this paper, the findings and transcripts from these interviews have been used in the stakeholder analysis of this thesis. The transcripts can be provided upon request.

3.2 | Stakeholder Analysis

Stakeholders can be defined as a person, group of persons, or institution, that is affected by the designed treatment [93]. Additionally, in healthcare a common definition is ”a person or group with a vested interest in a particular clinical decision and the evidence that supports that decision” [8, 12]. In design science, the goal is to treat the design in such a way that stakeholders are better off. However, in practice, situations may occur where some stakeholders might benefit in a certain aspect but be disadvantageous in others. In the worst case scenario some stakeholders might end up in a worse situation than before [93].

To prevent the last scenario from occurring, a stakeholder analysis is conducted. This stakeholder analysis describes what stakeholders there are and in specific how the implementation of the artefact may influence these stakeholders. This stakeholder analysis is an approach to improve understanding of the “behaviour, intentions, interrelations and interests” [91]. Within the current project, stakeholders may be identified within multiple contexts. These contexts can be derived from the problem definition. First off, from the problem definition is evident that the ultimate goal of the project is to reduce the burden on the healthcare and create a more efficient process. Therefore, we will define one context on which to analyse the stakeholders as the cost-efficiency context. Within this context all stakeholders are analysed that have a vested interest in the cost-efficiency of the problem. When trying to generate requirements for the different options that are available within the decision making process, we can define the second context in which to analyse the stakeholders: the process context. Here, the purpose is to define all users in the process as well as stakeholders that may be affected by the process but are no direct users. By looking at the problem from multiple contexts, different stakeholders may be identified that could otherwise be overlooked. Similarly, the same stakeholder may have different interests within a different context. Per context, the findings are modeled using the goal, question, metric (GQM) method. This method specifies the goals for the project at a conceptual level, provides questions to characterize how these goals may be achieved, and defines against what metric this will be evaluated [95]. Finally, all stakeholders are combined in table 3.1 to present an overview.

The data for this stakeholder analysis was acquired through the problem definition, performing unstructured interviews in the preliminary stages of the project, and through document analysis. Documents were made available by MohsA in which a student performed semi-structured interviews with stakeholders regarding the implementation of an app used for monitoring odd-spots. As this implementation regards an AI-enabled process in the detection of NMSC, the requirements and concerns of the stakeholders in that research are addressed here as well. The quotations presented in this section have been translated from Dutch to English.

3.2.1 | Cost-Efficiency Context

In the preliminary stages of this project, multiple meetings were held with the project leader at MohsA in order to describe the problem definition. From these meetings the first few stakeholders were identified within the cost-efficiency context.

MohsA
The main stakeholder, and problem owner, is MohsA itself. Within MohsA this can be translated to the persons that cover the board of the clinic. The interests of this stakeholder are essentially what is found during the problem definition phase and described in section 1. They are interested in increasing the efficiency. During the exploratory interviews was mentioned that to do so, they aim to reduce the number of patient visits to the clinic and thus indirectly to reduce the costs. They would like to achieve this increase in efficiency through optimising the current process. In the ideal case, an AI centered process helps with the diagnostics, taking away a part of the dermatologists burden. Questions that arise revolve around the efficiency in the current situation so that this can be compared to a new scenario.

Insurance company
An additional stakeholder in this context is identified through the treatments' financial details made available by MohsA. From the this information became evident that the treatments at MohsA get invoiced to insurance companies under a certain billing code that is treatment specific, depending on the type, number, and difficulty of operations. As MohsA does not charge patients directly, the insurance companies form a stakeholder in the sense that they are financially responsible and thus must be willing to cover the treatment costs in some way. From the interviews performed by the student, the interests of the insurance companies were highlighted. The general interest of these companies is the accessibility and affordability of high quality care.

> "Our most important job is that people who need care have access to it ... we commit ourselves to make care as qualitative and effective as possible." - Healthcare buyer

Accessibility relates to the access to good health care on the short or long term, tailoring the care to patients needs, and accessible by different population groups [60]. With this the insurance companies try to provide enough care at hospitals or care givers that are in close proximity of care receivers. They insinuate that due to changing events like corona, or an ageing population, the accessibility might be in jeopardy. The most recognisable example of this is the postponement of treatment that occurred during the corona pandemic [89]. Due to this postponement, healthcare was not accessible and patients had to deal with the consequences. To overcome this, the healthcare insurance companies aim for a situation where it is possible to effectively deploy care givers where needed. The accessibility and affordability translates into several interests when it comes to AI implementations in NMSC healthcare. It was primarily mentioned that the accuracy of the implementation should maintain at the current level. The app should 'do what it should do' in the sense that implementation does not incur an increased number of visits due to false positives but also not reduce the number of false negatives, which is related to quality of care. Furthermore, the insurance companies mentioned privacy and safe data transfer of patient information in their requirements.

> "... one thing is the transfer of information. Furthermore the safety of information, ... The trustworthiness must be present." - Healthcare buyer

> "With digital care, you quickly talk about the sense in which privacy is guaranteed." - Projectleader Skinvision

In this sense the privacy relates to the privacy of the insured patients. Since insurers can decide what institutions to fund, they act like a quality guarantee of care. As such, they do not only value accessibility but also safety of patients including patient privacy. One of the interviewees displayed concerns regarding the liability of an AI. In case something goes wrong because of an AI, what happens then? Who is in charge of the predictions that are made by the technology and the way the results are interpreted. This implies that regulations need to be in place to cover issues like the liability in order to generate a safer environment for patients.

> "Besides, what if the app is wrong and someone does seem to have skin cancer. Oh boy, what then..." - Healthcare buyer

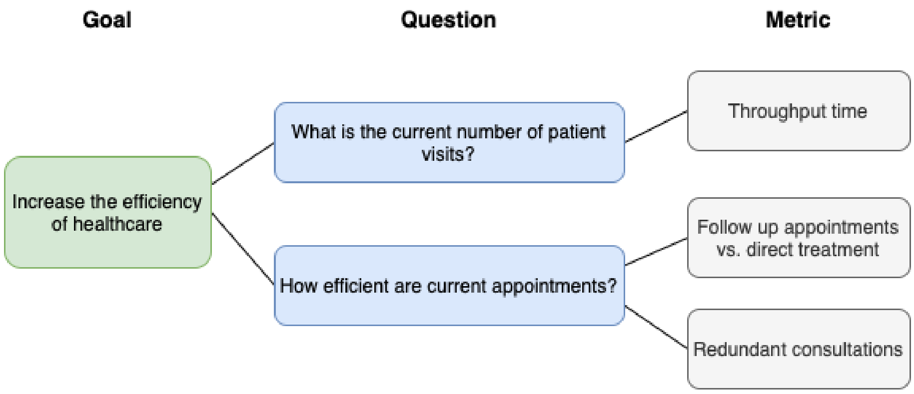
3.2.2 | Goal Question Metric - Cost-Efficiency

The first goal for the GQM model is defined by MohsA. Their aim is to increase the efficiency of care, which is defined by the number of visits, and throughput time of patients. This goal follows from the fact that currently there is a high burden on dermatologists and a rapid increase in skin cancer cases as was explained in section 1.1. To evaluate this, the current number of visits and efficiency of appointments should be know. A scale of measurement that can be used is the patient throughput time which can be expressed in a number of days. Related to this is the number of redundant consultations. If there are many appointments that could have been postponed, the efficiency is low. Automatically, the patient throughput time will decrease if these appointments are skipped. Finally, the ratio between the number of appointments that lead to an extra appointment for treatment versus appointments during which the treatment is directly executed is used as a metric to display how much could be improved. The second and third goal that are defined follow from what is mentioned by the insurance agents in combination with the primary goal to maintain patient satisfaction. The accessibility and quality of care are noted to be important by the insurance agents. These can be related to the accuracy and availability which will be measured through the accuracy of diagnoses, number of follow-up treatments and appointment lead time for new patients. Finally, the safety of patients is set as goal as the insurance companies want to guarantee the quality of the care patients receive and can act in this by deciding which instances to cover. The questions that follow refer to the privacy of data and accountability of malpractice. As privacy and accountability are not measurable, the metrics used relate to in what extent regulations are set and followed. For privacy these regulations have been set by the government (aglemene verordening gegevensbescherming), however, for accountability these regulations may not yet exist. If that is the case, this metric measures to what extent regulations have been set by the company. A visualization of the GQM method can be found in figure 3.1.

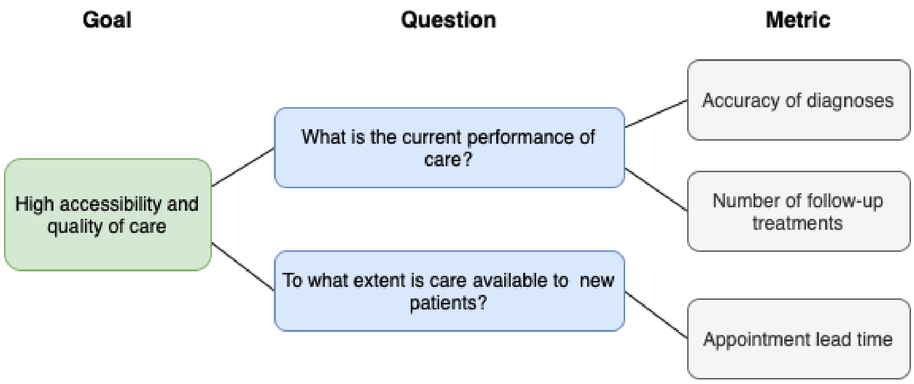
3.2.3 | Process Context

Within the process context, another set of stakeholders is defined. These stakeholders are identified by looking at the changes that occur within the organisation in case an AI-enabled process would be implemented. As the new process is not yet defined, an exploratory approach is taken in which, based on potential process directions, the stakeholders are defined. First off, by performing an unstructured interview with the project owner and an office manager of MohsA, several stakeholders can be described. These are the the data supplier that will provide data, the patient, the data analyst who will make use of the data for analysis, and the dermatologist. These stakeholders have been described below. In figure 3.2 the results of the GQM method are displayed.

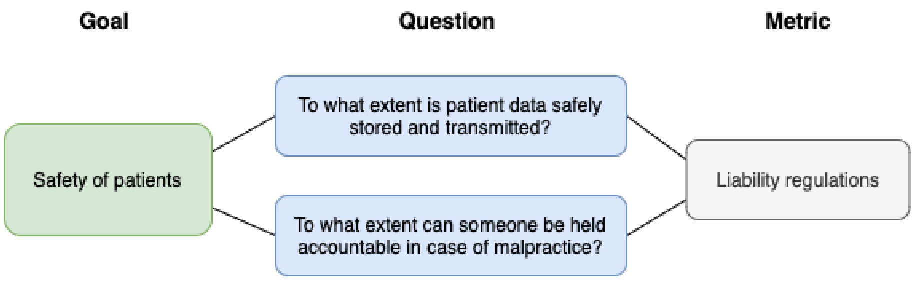
Data supplier
The first stakeholder was identified by looking at the supply of data and can be seen as the person that is



(a)



(b)



(c)

Figure 3.1: The goal, question, metric method applied to the cost-efficiency context of the stakeholder analysis

in charge of collecting and uploading the data that is needed for the AI applications. In the context of MohsA, this can either be the patient him/herself, a nurse, the clinician etc.

> "The patient could upload this data him/herself, or (if economically feasible) a nurse or dermatologist could take the pictures and enter the questionnaire."
> - Board member MohsA

For the sake of analysis this stakeholder will for now be defined as the data supplier. The main interest of the data supplier is the ease of use of the system. Specifically, data suppliers prefer a ready to use system that does not require an additional learning process. Downloading new applications, or requiring to learn how new applications work is unfavorable as the usage of the system for most data suppliers will be sporadically. Another aspect that is of importance for the data supplier are the requirements of data that needs to be uploaded. To keep the process as simple as possible, it is desirable that the data can be provided from all different platforms. This includes computers, tablets, and mobile phones. Images and textual data must be uploadable through the system directly from the device, without the necessity to convert data. In other words, the system must be flexible.

Patient
Even though the patient may in some cases not be the data supplier, they are still an important stakeholder due to the fact that the process changes affect their care. The healthcare system is in the end designed for the patient, therefore, their needs and concerns must be heard. Changes in this process, that may lead to an alteration in the scheduling of appointments or consultations, will eventually impact the patient. Jutzi et al. [46] and Nelson et al. [61] looked into the patient perspective of AI in SC diagnostics. In specific, the attitude of patients towards the usage of an AI for assessing their lesions was tested. Their papers displayed that patients have a high level of confidence towards AI decision makers if these AIs had been proven effective. Jutzi et al. [46] even pointed out that when an AI had proven to be more reliable than a physician, more people would trust the AI over the physician in case of judgement. However, both papers showed that the vast majority of the participants was determined that the AI should be used as a support tool for physicians and not replace them. The reasons for the desire of conjoint operation were related to the patient-physician relationship. A few key factors could be extracted from both papers in combination with interviews of MohsA patients performed by the student: (1) empathetic abilities, the participants desired someone that could deal with emotions and use (non-)verbal communication to cross the information, (2) possibility of asking questions, the participants stated that through personal consultations they are able to ask questions and that they fear that with the implementation of an AI these possibilities would diminish, (3) reliability, even though most participants stated they have faith in an AI detection algorithm, many of them also expressed concerns regarding reliability due to the factors like deviations in photos, unreliable training set, many false-negatives/false-positives.

Aside from these aspects, a high priority of the patients is the privacy of their health records. Some participants of the paper of Jutzi et al. [46] were afraid that insurance companies might misuse their information to their disadvantage. In addition, the participants interviewed by the student stated that they would not want anyone but the physicians to have insight in their information.

Final interests of the patients come forth from their demographic. As explained in section 1.1, elderly have the highest risk of becoming a chronic NMSC patient. A report by the European Commission [11] describes the challenges that occur within long-term care of an ageing population. As the patients with chronic NMSC are also in need of long-term care, the remarks made in this report apply to the current situation at MohsA. For starters, the report states that the quality of care is essential. They define this not only as quality of actions performed during healthcare, but in the more general sense towards quality of life. Some dimensions of quality of life that are mentioned and are applicable to this situation are the control over daily life, safety, and social participation. These aspects have been mentioned before in the form of reliability of the AI system and personal consultations for social participation. Added to that is the fact that patients should be able to have access to the care when they need it and must be given the decision on how to participate.

Data analyst
A third stakeholder that can be identified in the process context, is the person who is going to be using the provided data in order to diagnose the lesion of the patient and use this diagnosis to decide on the following steps in the process. From an unstructured interview performed in the preliminary stages of the thesis came forward that this person can either be the dermatologist, a skin therapist, or a nurse that is trained on NMSC recognition. As this person will be working directly with the delivered data and

needs to understand the process and AI application, it is of importance that the needs of this stakeholder are considered. During a focus group conducted by the student with MohsA employees, their interests were collected. One of the major aspects that was mentioned is the fact that new technologies are often designed primarily from the patient perspective while the healthcare users are overlooked.

> "It is primarily focussed on the patient ..." - Care giver
> "Yes, it is mostly client driven and not from the point of view of the care giver" - Care giver

Due to this, problems are not solved but only extra trouble arises if technologies do not work easily. Furthermore, they express that the technologies should not increase the workload or require an extensive learning process. In case a lot of extra work is involved, new technologies are not considered an improvement.

> "... I went completely crazy. It took me more time to fill out records than I was performing surgery, that must not be the case." - Dermatologist
> "It must be efficient. The time that it takes you must be little so that you think, this is faster than picking up the phone to do it." - Care giver

Dermatologist
The final stakeholder that is identified is the dermatologist. The dermatologist plays a major role as their way of performing diagnostics, as well as their contact to patients might change. A primary interest of the dermatologists is the quality of care. The artefact should ensure that the patient well-being is similar as in the current scenario. This includes the fact that patients should receive an accurate diagnosis, but also receive the care they desire in the sense of personal contact and aftercare. As the artefact will include options for AI implementation, this means that a requirement for the AI (with or without human interaction) is that it must be able to deliver comparable diagnoses as a physician would do. Furthermore, the dermatologist ultimately carries the burden of an overstressed sector. Therefore, their interest is similar to that of MohsA in the sense that they desire a reduction in workload.

3.2.4 | Goal Question Metric - Process

Within the process context, similar goals as to the ones in the cost-efficiency context were defined. The goal of increasing the efficiency was mentioned by the dermatologists and quality of care was mentioned both by the patients as the dermatologists. The mentioned aspects of increasing the efficiency correlate between the two contexts and will thus not be treated again, regarding the quality of care however, another point of view was introduced. From the patients' point of view, the quality of care is more related to personal contact and self control. As such these aspects have been included as questions. First off, personal contact is measured through the time patients can be in direct contact with a MohsA employee. This can for example be through consultations either live or via telephone. Self-control is related to the way patients can influence their own care process. The number of decisions that patients can make indicate how much control they have. Decisions can be deciding on whether to postpone a treatment or not, whether to perform a certain procedure, or whether to make use of remote screening or not. The last question is derived from the fact that possibly care may be delivered by an AI or other clinician than a dermatologist. In order to guarantee the quality, the accuracy of the care can be measured through looking at the specificity and sensitivity of the diagnosed lesions. The second additional goal follows from the fact that both the data supplier and the data analyst value a system that is easy to use. This is broken down into several questions. Since they indicate that it should not result in additional work, the question can be asked how fast the system is in use, which can be measured by throughput time per case. Additionally, it is essential that users can upload their data from any device without converting, so that the question is asked how adaptable is the system in different scenarios? Here a metric is used to check for the compatibility of files and operating systems. A visualization of the GQM method can be seen in figure 3.2.

3.2.5 | Users

In a paper by McLeod Jr and Clark [54] the key aspects of users in an IS stakeholder analysis were described within the health care setting. They implicate that within health care, not all stakeholders are automatically considered as users within an IS. When applying this way of looking to the stakeholders that are defined in table 3.1, we can identify 2 users of the system: the data supplier, and the data analyst. The patient in this scenario is no user but merely a data source. Only if the patient ultimately takes the

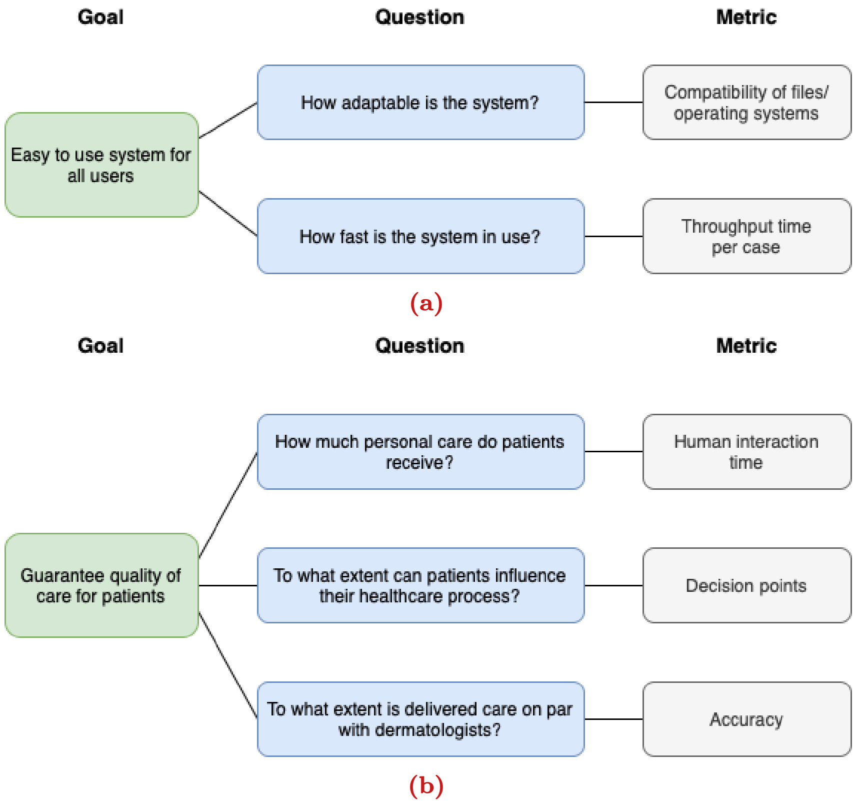


Figure 3.2: The goal, question, metric method applied to the process context of the stakeholder analysis

Table 3.1: Stakeholders and their interests

Stakeholder	Interests
MohsA	Increase efficiency; decrease number of patient visits, reduce costs
Dermatologist	Decrease burden, maintain healthcare quality, receive substantiated information on lesion and preliminary diagnose
Insurance company	Quality of care, accessibility of care, privacy of data, liability regulation
Data supplier	Ease of use, ready to use system that is compatible to all devices, wide range of accepted data types
Patient	Personal consultations: possibility to ask questions & empathetic involvement, involvement of dermatologist to enhance reliability, privacy of data, accessibility to care
Data analyst	Design from a health provider point of view, time efficient, easily executable

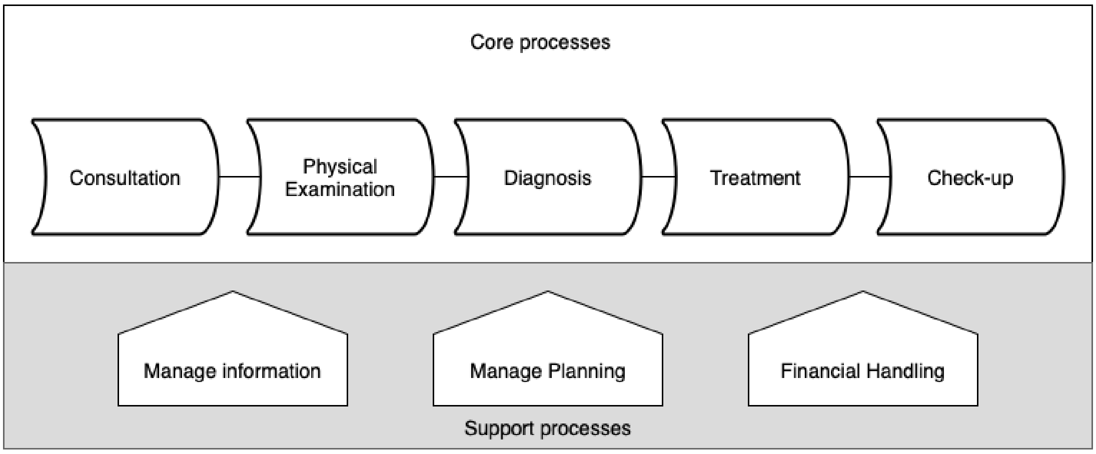


Figure 3.3: Process landscape showing the process categories divided in the core and support processes

role of the data supplier will they become a user. Following from this, is that the system itself should be designed keeping in mind the interests of the data supplier and data analyst.

3.3 | BPMN of Current Process

In order to display business processes within the organization, and to define the new processes, the business process model and notation (BPMN) is used as described by OMG [65]. The reason that this notation is chosen is that it is a standardized notation method that can be understood by business users. In this way, the notation cannot only be used to describe the design by the creator, but can later also be managed and monitored in real life by the business process owners. In order to generate the BPMN process, the stages of process identification and process discovery are performed. It is necessary to perform a process discovery in order to generate a process in case there is no defined process or process description available [18]. Since at MohsA the current process has not yet been defined explicitly, this approach is chosen.

3.3.1 | Process Identification

Before process discovery can lead to the description of a model, the processes need to be identified. This translates into identifying the processes that occur, and the relation between them. Accordingly, these processes are mapped in a process architecture. This is done by defining the categories of processes, and describing different relationships between these processes [18]. This process architecture is divided into core processes and support processes. The core processes comprise the 'production' tasks in the clinic. From literature, five main processes of healthcare have been identified. These include prevention, detecting health problems, diagnosing diseases, treating diseases, and providing for a good end of life [4]. Within the context of this research, detecting health problems, diagnosing disease, and treating diseases are directly related. MohsA actively detects health problems by scheduling regular check-ups even if there are no direct concerns of the patient. The diagnosing and treating of diseases are done at MohsA during the process. To go into more depth, the detection of health problems is subdivided into 'check-up' and 'consultation', and the diagnosis is divided in 'physical examination' and 'diagnosis'. The support processes enable these core processes to be executed and comprise of IT management, administration and finance, and planning. The process categories are displayed in the simple process landscape in figure 3.3.

3.3.2 | Process Discovery

The next step is process discovery [18] in which the processes and the related resources are defined. This is done through interviews with employees, observations in real-life, and through document analysis. MohsA uses an electronic health record in which not only the procedures are mentioned that were executed, but also the procedures that could be executed within a certain clinical picture. This information can be used to define the possible processes that occur within consultation, diagnosis and treatment. The observations are done by shadowing a dermatologist during some of their appointments with NMSC patients that came for their regular check-up. For the modeling of the process, 5 steps are followed as defined by Dumas et al. [18]: 1. identify the process boundaries, 2. identify activities and events, 3. identify resources and their hand-offs, 4. identify the control flow, 5. identify additional elements.

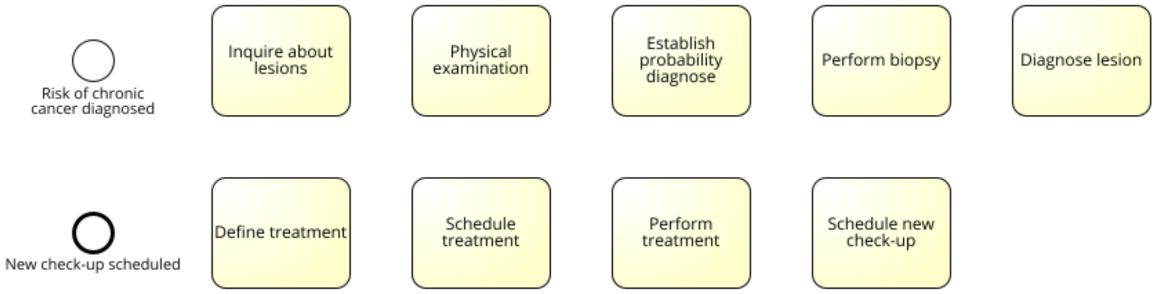


Figure 3.4: BPMN activities and events

To start off, the process boundaries were identified. In other words, what activity signals the beginning and what activity signals the end. In this case, the process starts when a NMSC patient is diagnosed with 'chronic' skin cancer. This can be the case when a patient has been physically examined, and other lesions that could potentially develop into skin cancer are found. Another option is that a patient has an elevated risk on skin cancer based on age, skin type, UV-exposer etc. We define this start event as *Risk of chronic cancer diagnosed*. The process ends, as soon as the next check-up is scheduled. Whether this is done after the check-up in case no alarming lesions are found, or after a full treatment process has taken place is unimportant as a chronic patient remains under therapy and will thus need to schedule a new check-up at the end. We define this as *new check-up scheduled*.

Secondly, all occurring activities and events were identified. During the observations that were performed while shadowing the dermatologist at patient consultations, became evident that all patients are asked whether they have any lesions they are concerned about themselves. Following, independent of whether they say yes or no, the patients are still checked for any odd spots during a physical examination. Additionally, if there are risky lesions, the patients are asked a few questions about the lesion regarding its characteristics. During the observations this was enough to diagnose the lesion as these cases only showed low risk lesions. However, from interviews and data in the electronic health record became evident that not all lesions are those low risk types. In some cases, a higher risk of skin cancer is assumed, and a biopsy must be performed to check the cell composition. This biopsy is sent to a pathologist for examination. The results of this examination assist in the final diagnosis. From the diagnosis a treatment plan follows. During the observations, the treatment was easily executed and could be performed immediately. This situation, in which consultation and diagnosis (and sometimes treatment) can be performed on the same day, is defined as the one-stop-shop model at MohsA [56]. From the data and interviews is evident that some treatments are more enhanced and a new appointment must be scheduled to perform the treatment, or that the one-stop-shop must be scheduled on another day when they perform Mohs surgery. After the treatment is performed, the next check up is scheduled at which point the process starts over. These processes can be found in the overview in figure 3.4. Note that there are more activities performed at MohsA and that the treatments can consist out of different procedures, however, these are outside the scope of this research and therefore these activities have been grouped together as 'treatment'.

Thirdly, the resources were identified that are responsible for the activities. In addition, the points during the process are identified where the task is handed off to another resource. The resources are identified through interviewing the project owner about these activities. The activities that are performed during the consultation are done by a dermatologist, this includes inquiring about lesions and checking for lesions. In case a biopsy is needed, the dermatologist will also perform this procedure. Afterwards, the biopsy is checked by a pathologist. As this is not being done at MohsA but externally, the pathologist is shown as a collapsed pool. The initial examination with possibly the results from the biopt lead to a diagnosis that is made by the dermatologist. After the lesion is diagnosed, the treatment is defined by the dermatologist. If the treatment is not directly executable, the patient will schedule an appointment for the treatment with an administrative employee. The treatment itself is again executed by the dermatologist. Finally, the next check-up is scheduled by the administrative employee. In this scenario, the patient is considered to play a passive role and is only included through message flows as advised by Pufahl et al. [71], therefore, the patient is displayed as a collapsed pool. An overview, showing the resources and their hand-offs, is displayed in figure 3.5.

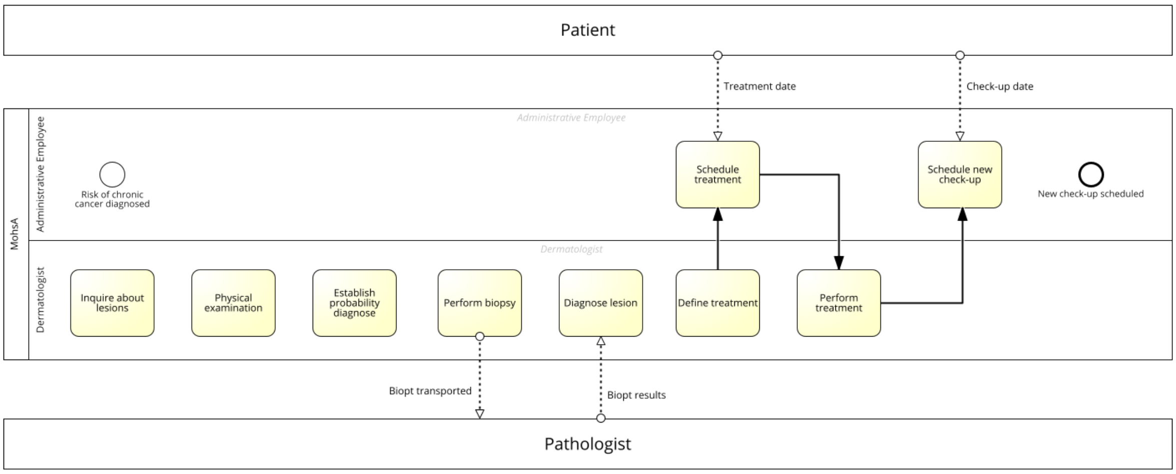


Figure 3.5: BPMN resources and their hand-offs

The final steps are to identify the control flow and any additional elements. These additional elements include gateways and intermediate events. The gateways include the possible directions the process can take as explained earlier. This depends on the probability diagnose, and whether the treatment is directly executable. The additional intermediate events consist out of the interaction with the pathologist where the tissue is sent and results are received, and a timer event in case a treatment cannot be performed directly. Finally, since after a new check-up is scheduled the patient loops back to the beginning of the process, it is chosen to add an additional timer at the end that indicates the time until the next check-up. When all is put together, the final business process is modeled in figure 3.6.

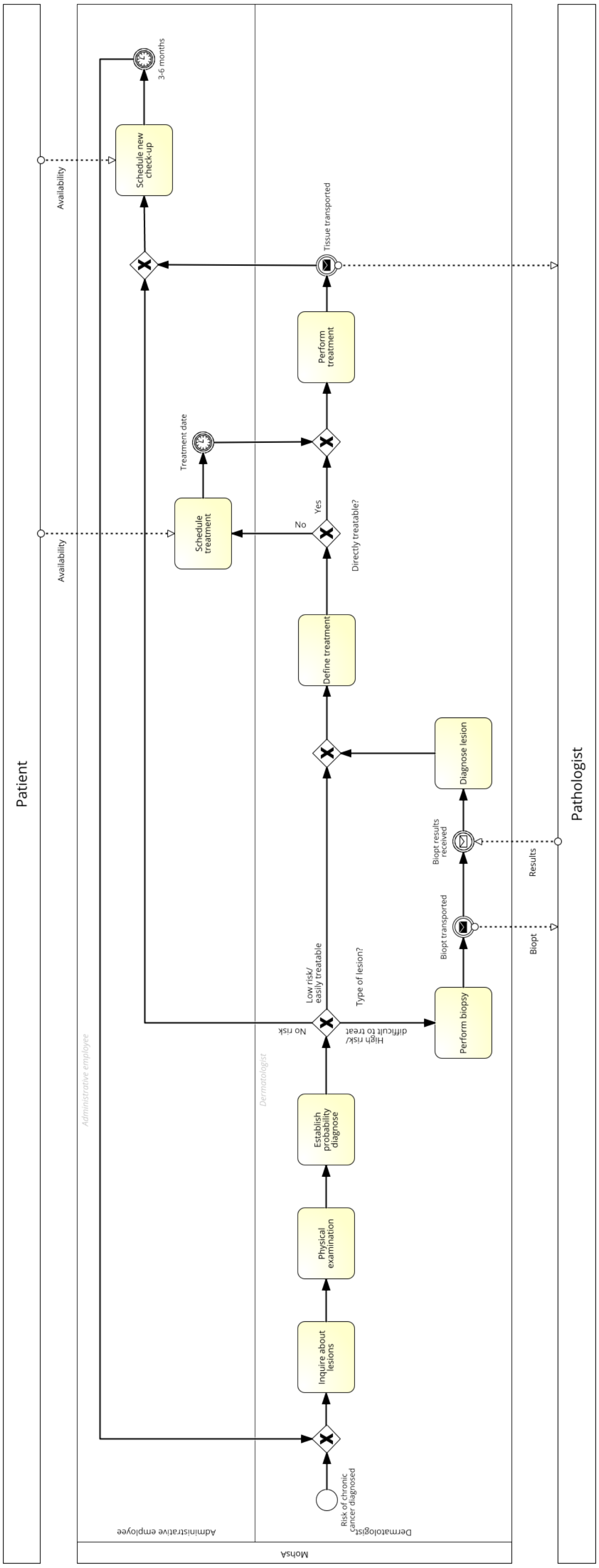


Figure 3.6: BPMN diagram of current process at MohsA

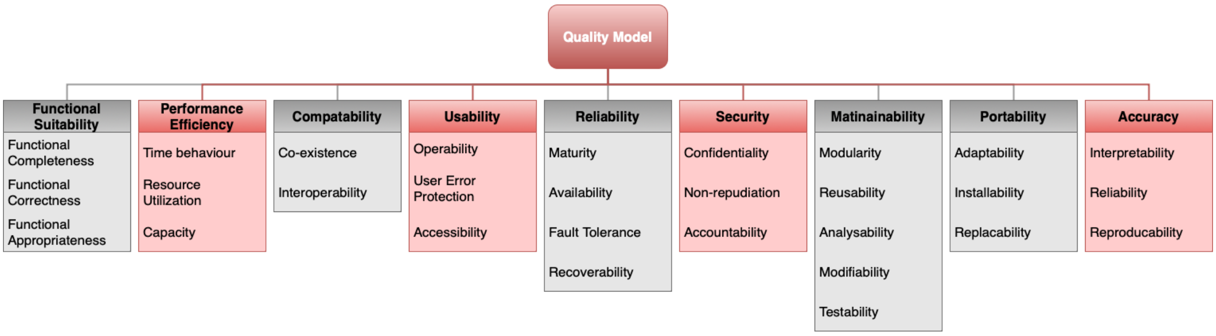


Figure 3.7: Quality Model

3.4 | Quality model

Based on the findings form the stakeholder analysis, the quality model from section 2.6 is extended with extra quality indicators. First off, the stakeholder analysis has shown that the usability of the system is of importance. Therefore, the quality indicator *Usability*, as described in the quality model from ISO/IEC 25000 [42], has been included. In addition, an extra quality indicator *Accuracy* is introduced that is based on both the literature and the stakeholder analysis. The literature showed that the evaluation of how well an AI performs is measured by different metrics related to accuracy. Based on these metrics, it is possible to determine whether an AI has a similar performance as a dermatologist. In addition, the stakeholder analysis showed that the reliability and healthcare quality are of importance which is indirectly related to the accuracy. The added quality indicators are described below. The final quality model can be found in figure 3.7. A summary of the subcharacteristics and their definition can be found in table 3.2.

Usability
The usability is an indicator for the sense in which specific users can use the system in a way that satisfies their needs. This is thus related to all users on both ends of the system. As subcharacteristics are included: **1. Operability**: how easily can the system be learned by the users in order to effectively, safely, and satisfactory achieve the goals. This will be evaluated not only in the sense of the digital applications, but also as how easily the process can be understood and what steps need to be taken. **2. User error protection**: how well does the system protect users from making errors. Think of data errors that could be the result of unclear instructions or errors within the process steps. **3. Accessibility**: the way in which the system is accessible by all types of different users. The accessibility is especially important with respect to the patients as this is a varied group of different ages and backgrounds. The system should keep in mind that e.g. elderly might have difficulty accessing certain digital systems.

Accuracy
Finally, one extra quality indicator has been added to the reference model. As the application is to be used in a healthcare setting, it is essential that the accuracy of the system measures up to the expectations. These expectations might differ per stakeholder, therefore, multiple aspects need to be included for consideration. For this the following subcharacteristics are identified: **1. Interpretability**: degree to which the output of the system can be interpreted by the care givers. This will help care givers in being more certain about their decision, as well as providing a substantiated basis for informing patients. **2. Reliability**: degree to which the output is reliable in sense of sensitivity and specificity. This subcharacteristic measures the way in which the artefact can produce results that are on par or better then previous set standards. **3. Reproducability**: to what extent can results that are outputted be reproduced in the same scenario. This measurement is necessary to make sure an application uses the correct inputs to define its output and does not use secondary information due to which output may drastically change.

Table 3.2: Quality model indicators

Subcharacteristic	Definition
Time behaviour	The extent to which the efficiency increases
Resource utilization	The amount of human labour required to perform healthcare tasks
Capacity	The possibility with which high volume data can be processed
Operability	Ease with which the system can be operated or learned
User Error Protection	The extent to which the system prevents errors made by users
Accessibility	The accessibility of the system by users regardless their technological background
Confidentiality	The extent to which the system ensures that data is only accessible by authorized people
Integrity	Degree to which the system is designed to prevent unauthorized access
Accountability	The extent to which regulations have been defined with respect to liability of actions
Interpretability	The interpretability of the output without ambiguity
Reliability	The accuracy of the system in terms of sensitivity and specificity
Reproducability	The extent to which rerunning the system will produce similar results

4 | Phase II: Artefact Design

The artefact will be designed in accordance with the goals and requirements that were set at the beginning of the project (section 1). The goal of the project is to ultimately decrease the number of patient visits and reduce the costs while maintaining patient satisfaction. To achieve this, it was decided to provide a decision space that identifies the questions, options, and criteria (QOC) surrounding the implementation of an AI-enabled process. This decision space is required to provide context and insights in the decision making activities and provide the best practices for an AI-enabled process. This section will set out to develop the decision space by using a QOC model representation as was introduced in section 2.7.

4.1 | Methodology

The questions, options, and criteria are designed in accordance with all stakeholders. The stakeholder interests established in section 3.2 allowed for the development of the GQM. The insights that came from the GQM method, together with the QOC elements that were developed based on the literature, will provide the basis and substantiation for the elements of the decision space that are explored in this section. To develop the artefact, several stages are performed. First off, a set of draft questions is formulated based on the stakeholder analysis and literature to serve as a starting point. Afterwards, by performing interviews with different stakeholders, these draft elements are evaluated and more QOC elements are retrieved in order to develop a final QOC model. This is an iterative process where the newly introduced questions, options, and criteria are discussed in later interviews to gain additional insights. The data that was collected from the interviews through coding was validated by doing cross-validation between the interviews and through comparing it against existing literature. After the concepts were validated the data could be used for the substantiation of different options based on the established criteria. As a QOC model is the product of design, the elements are substantiated by arguments from the stakeholders and the designer. As such, the QOC may contain flaws that can lead to the re-design of the QOC model in the future.

4.2 | Draft QOC Elements

The goal of the artefact is to provide a decision making tool to assist with the decision making activities surrounding the implementation of an AI-enabled process in the NMSC healthcare setting to improve efficiency. As such, the QOC model must provide a guide that helps decision makers with building an argument for their decision making activities and present a complete overview of the options and criteria to be considered. During the exploration phase, the context was specified to only include patients that have already had NMSC before and are thus considered chronic patients. A primary question that is raised when examining the possibilities of improving the efficiency, is what activity/activities to enhance of the healthcare process. The BPMN that was created in section 3.3 shows the activities that occur and thus what options can be considered. From the discussions with the problem owner became evident that ways to enhance the efficiency can either be to reduce the number of patient visits, or fasten the throughput time of the process. First off, in order to know how to reduce the number of patient visits, it is important to know what visits are superfluous. These visits have primarily been identified by the problem owner as being consultations where patients have no or very low-risk lesions, and consultations where it is evident what the diagnosis is but the treatment cannot be performed immediately. Enhancing the screening process may prevent these superfluous consultations. Secondly, the one-stop-shop model used at MohsA (section 2.4) has proven efficient in enhancing their healthcare process. Enhancing this model further might increase the efficiency even more. These aspects lead to the definition of the primary question and options of the QOC model. The general criteria that are defined are based on the goals of this thesis as defined in the problem definition. These are cost reduction, decreased number of patient visits and patient satisfaction. From these criteria, more specific criteria can be derived for deeper levels of the QOC. These more specific criteria are defined as bridging criteria [50]. The first element of the QOC is depicted in figure 4.1.

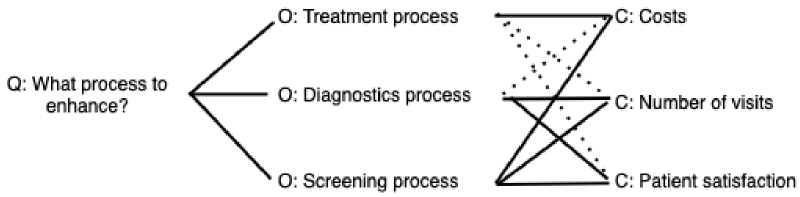


Figure 4.1: Primary element of QOC model

The depicted relations between the options and criteria are a draft that has been deduced from literature. In literature it was found that the treatment process during MMS can be accelerated by using an AI to detect malignant tissue in the WSIs [90]. This application will reduce costs through reducing the operation time, however, the number of visits stays equal. Patients may experience shorter operation times. The diagnostics process can be enhanced by using an AI to assist a dermatologist in diagnosing a lesion. This can enhance the precision of diagnoses through an extra check so that less errors occur in the diagnostics process. Less errors promote patient satisfaction as treatments may be more effective. Through more effective treatment, it is possible that less patient visits are required. However, to acquire both a dermatologist's diagnosis in combination with an AI assessment might increase the workload as two separate assessments have to be made and thus the costs may increase. Finally, enhancing the screening process may positively influence all three criteria. By making use of AI in screening, less patient visits may be needed as chronic patients currently need to revisit regularly for a physical examination. Due to less patient visits, the costs will also reduce. It is expected that the patient satisfaction will also increase as they receive more autonomy [46] and need to visit less frequently.

The next question that follows is how to enhance the screening process. From the interview that was performed during the development of the BPMN process, two potential improvements were derived. Both of these improvements require the patients to provide data on their lesion(s) in the form of an image and/or answers to a questionnaire. The first improvement that was introduced by making use of this data, is to use this data for prioritizing the appointments of patients. In case it is possible to determine, based on the patient data, that an appointment has a low priority, the appointment can be rescheduled. If the priority is high or there is uncertainty about it, the appointment is unaltered. The second improvement that was introduced, is to make use of the data to decide on a probable treatment. In that way, more time can be scheduled for appointments where more extensive treatment is expected so that separate follow up appointments to perform these treatments become obsolete. The criteria against which these options will be measured follow from the GQM method in the stakeholder analysis. The goals that resulted from this GQM method are the basis for these criteria. The first goal is to enhance the efficiency, therefore, the options must prove to be time efficient. Secondly, high accessibility and quality of care are strived for. This is translated into the criterion of accuracy as this is an aspect of quality of care and patient satisfaction. It is assumed that patient satisfaction flows from a high quality of care which relates to high accuracy. The QOC element is depicted in figure 4.2. More criteria may follow from the interviews performed in the data collection stage, this has been indicated in the figure by *C:...*.

In each of these two options for improvements of the screening process, data is received by MohsA and needs to be analysed. This raises the next question that was also introduced during the interview for the development of the BPMN process: who/what is going to analyse the data? Three options are considered. The first option is rather straightforward, namely, the dermatologist. The second option that was introduced is the usage of an AI to perform the analysis. In this scenario the data will be fed into an algorithm in order to receive a risk analysis of the types of lesions. The third option is that a nurse trained in recognition of malignant and benign lesions will perform the analysis. It was stated by the project owner that at MohsA there are nurses/skin therapists that are able to perform this analysis. Earlier research by Hoepel [39] performed at MohsA confirms this. The first criterion against which this will be measured follows from the goal to have a reduction in costs, as well as from the GQM method in the stakeholder analysis. As the data analyst in the stakeholder analysis mentioned the fact that a solution must be time efficient, this goal was set similarly to that of the previous QOC element. From a financial point of view, it is also required to be time efficient in order to reduce the costs. The second criterion that is defined is accuracy which can again be seen as a bridging criterion of quality of care and patient satisfaction. Other criteria might follow from the data collection stage.

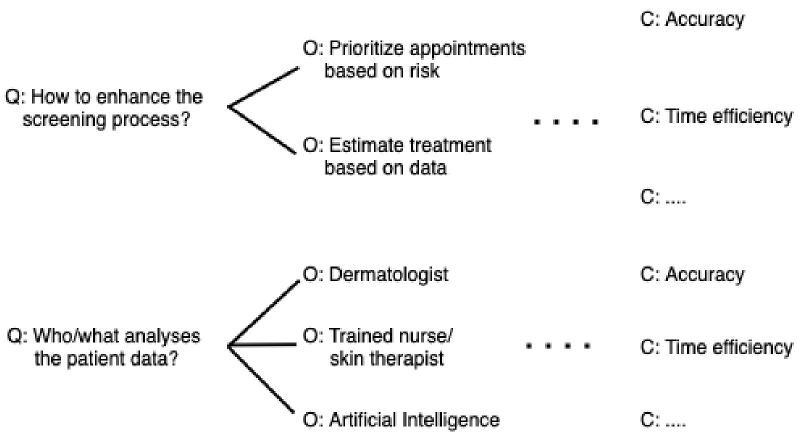


Figure 4.2: Draft QOC elements related to the improvements of the screening process. The dots indicate the relations and potential criteria that still need to be uncovered.

By introducing a new process to improve the healthcare efficiency, decisions have to be made on how this process will be implemented. The process revolves around the data stream of patient data going from the patients to MohsA, and of information from MohsA going back to the patients. As such, several questions are raised. First off, the question is asked through what medium this data will be collected. This question follows from the requirements that followed from the stakeholder analysis. These requirements include privacy of data, and ease of use. Several options were proposed by an IT employee at MohsA that take these requirements into consideration. The first option is to use a secure email server called Zivver that is already being used at MohsA. By using Zivver, patients are required to log-in to an online portal using two factor authentication in orer to be able to read the message. Through this extra layer of security, the privacy is ensured. The downside that was mentioned, is that even though patients are able to upload images through Zivver, it is not possible to implement a questionnaire template. Patients will therefore need to write the answers to a questionnaire and cannot simply select the correct option. Other options might be presented during the data collection phase that are not currently available at MohsA.

In addition to the collection of data, it must be decided how data will be communicated to the patients. Again, criteria from the patients perspective are privacy and ease of use but also include personal involvement. As was stated in the stakeholder analysis patients value the empathetic involvement of a dermatologist as well as the possibility to ask questions. From the perspective of the data analyst however, the criterion is that it is time efficient. Several options are proposed that are possible. First off, the information can be returned to patients through Zivver, keeping in mind the privacy of patients. Another option, is that patients receive a call from a MohsA employee to convey the message. This option is both secure, as does it allow for personal questions. However, it will take more time for the MohsA employees. Again, more options may be presented throughout the data collection phase.

One final concern that was raised by one of the insurance agents, is the accountability in case of malpractice as a result of an AI application. In order to guarantee the safety of patients, the insurance agent raised the question what happens when something goes wrong and no one is liable. As such, this question is included in the QOC. The options that are proposed have been described by He et al. [36] who describes that through AI the responsibility may shift from individual bad actors towards issues that are system-wide

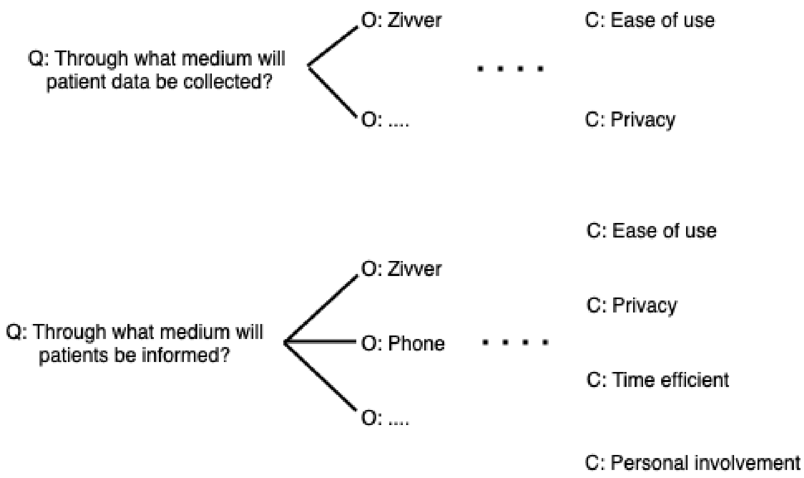


Figure 4.3: Draft QOC elements related to the data collection and communication medium. The dots indicate the relations and potential options that still need to be uncovered.

and can potentially be improved. These external options include the vendor of the system, the algorithm developer, and the source for training data. Finally the clinician is added as an option as these are currently accountable for the treatment procedure. The primary criterion against which these options are measured is patient safety. Additionally, it must be able to take action against a party that is held accountable in order to make sure that improvements will follow. This is translated in the criterion actionable.

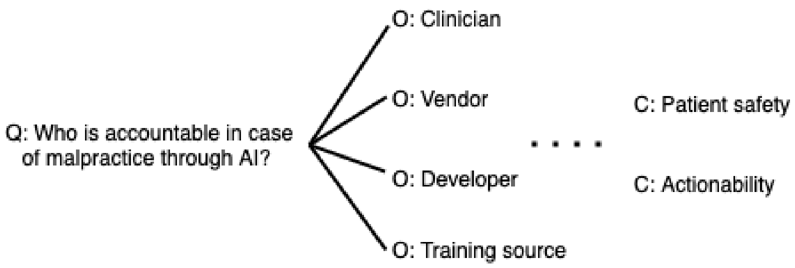


Figure 4.4: Draft QOC element related to the accountability of AI implementation. The dots indicate the relations that still need to be uncovered

4.3 | Data Collection and Analyses

In order to receive input on the options and criteria data was collected. This data was collected by performing interviews with stakeholders as defined in section 3.2. The conducted interviews were semi-structured as this method facilitates in-depth interviews to uncover new ideas in the exploration stage of a research domain [7]. In order to select participants, a few requirements were established. As the research is conducted as a case study for MohsA, it is required that all participants are related to MohsA. Furthermore, patients that are interviewed must have had a form of NMSC before so that they qualify as chronic patients. Finally, in case the participant is a dermatologist, they must have experience with NMSC healthcare and have knowledge about the one-stop-shop model that is practiced at MohsA. These criteria are listed in table 4.1. The interview guides have been added to appendix C.

The goal was to receive input on the draft elements of the QOC above and to acquire input for additional elements. The interview questions differed per stakeholder to specify for the healthcare processes that were of interest to them. To structure the interviews, an interview guide was created that contained explanatory text to inform the participants on the subjects, and the questions itself. The participants are listed in table 4.2. Comments are included in case specifics about the participants influenced the interview in some form. All interviews were recorded and transcribed in order to perform the analysis. As the respondents were Dutch, the interviews were also conducted in Dutch. The transcripts and recordings

Table 4.1: Interviewee requirements

Requirements for interview participants	
R1	All participants must be related to MohsA
R2	Patients must have had a form of NMSC before
R3	Clinicians must have experience with NMSC healthcare
R4	Clinicians must be familiar with the one-stop-shop model

Table 4.2: Participants QOC interviews

No.	Participant	Comments
1	Patient_01	-
2	Patient_02	Patient accompanied by daughter, daughter included in interview
3	Patient_03	Patient accompanied by partner, partner included in interview
4	Patient_04	Patient made use of one-stop-shop but not of regular physical examinations
5	Dermatologist	Familiar with one-stop-shop but no personal experience
6	Skin therapist	-
7	Administrative employee	-

can be provided upon request.

4.3.1 | Coding

In order to extract the relevant information from in-depth interviews, a qualitative data analysis (QDA) was performed. In order to perform the QDA, the interview transcripts have been coded. The coding strategy partly used the approach as presented by Deterding and Waters [14]. As this approach focuses on QDA with data on over 30 interviews, it was decided to only follow the first two steps of their approach. This includes data exploration and preparation, and the data analysis. The third stage of their approach makes use of QDA software to build models and test data-based theory, this will be replaced by performing conceptual validation of the earlier developed arguments based on literature. Besides using QDA software, the coding was performed using word.

First off, the transcripts are structured using indexing codes. These codes are applied to large chunks of text that help to pin comments to certain topics of the interview protocol. In this way, it is possible to reduce the data per topic, and easier for subsequent reading to be more focused. In addition, memos are made at a respondent-level and between the cases to start to develop an argument for hypothesized relations [14]. The indexing codes that are used can be found in the first column of table 4.3. These codes have been translated from Dutch to English as the coding process itself was performed in Dutch. The indexing codes have been established beforehand based on the literature and GQM model. The interview was structured around the concepts that came forward in these sections, therefore, the indexing codes could be applied easily. Secondly, by using the index codes to reduce the data, the resulting data per index can be analysed making use of analytical codes. These codes are generated based on concepts that may relate to known literature, or could be explored in literature. The analytical codes are grouped per index code and can be found in the second column of table 4.3. After having coded the interviews using the predetermined list of codes, another iteration of coding was performed during which quotes that introduced new concepts were coded using new codes. These newly introduced codes have been marked with purple in the table.

4.3.2 | Conceptual Validation

The coding of the interviews uncovered several concepts related to the draft QOC that were mentioned by the interviewees. These concepts in combination with the memos that were retrieved at a respondent-level and between cases, were compared to the literature described in section 2. In case new concepts were mentioned, these were validated by researching new papers. The concepts that were uncovered include 1) attitude towards physical examination, 2) attitude towards self-examination, 3) requirements uploading medium, and 4) attitude towards AI. From these concepts, 1, 3 and 4 have already been discussed

Table 4.3: Codes and their explanations that were used during the qualitative data analysis of the interviews.

Index code	Analytical code	Description
Self examination	Taking pictures	Possibility to take pictures, quality of images
	Filing questionnaire	Possibility to fill out questionnaire, relevant questions
	Assessment of risk	Assess what is benign or malignant for patients
Uploading medium for patients	Ease of system	How easily can it be used and is it prone to errors
	Privacy	Requirement due to personal data
	Examples	Mentioned examples are email, EMR, application, website
Medium to convey information	Ease of system	Easy access to information, easy to provide information
	Personal contact	Possibility to ask questions, empathetic relation
	Requirements	Requirements that the medium must adhere to
Prioritize appointments	Trust in assessment	How well can the risk be assessed based on images and a questionnaire
	Safety	Extra check after assessment, assessment criteria
	Expectations	What are the expectations regarding postponement, either positive or negative
Estimate treament	Trust in assessment	How well can the treatment be estimated based on images and a questionnaire
	Safety	Extra check after estimation
	Expectations	What are the expectations regarding estimation of treatment, either positive or negative
Artificial Intelligence	Knowledge	What is known about AI
	Trust	Participants trust in the assessment made by an AI
	Safety	Extra check after assessment by a human, verified algorithm
	Liability	Responsibility for assessment made by AI
Experience physical examination	Multi-purpose	In case multiple diseases are present, a dermatologist can asses these during consultation
	Dermatologist expertise	Understanding of what is benign or malignant, answer questions
	Familiar clinician	Personal contact
Other	Priority stakeholders	What stakeholder requirements should be most important
	Other	Any additional remarks made that were deemed of itnerest

before to some extent within the literature review. For the findings concerning patients' attitude towards self-examination new papers are introduced to support the uncovered concepts.

Physical examination

First off, the interviews inquired about the experience the interviewees had with the physical examinations during which patients that have had NMSC before are examined for malignant lesions. All of the interviewees, both on the care receiver as the care giver side, experienced these examinations as positive. The main reason that was presented is that dermatologists can assess effectively what lesions are benign or malignant where patients cannot and that patients can inquire about these lesions. Being able to inquire about lesions was also related to the personal contact patients had with their clinician. It was stated that the patients trust a clinician and they prefer to always have the same one who knows them as well. In specific, one patient who had multiple diseases was satisfied by the regular examinations as it allowed her to inquire about multiple issues at the same time with a dermatologist that knew what was going on. The report by the Commission et al. [11] describes that the number of elderly patients with multiple (chronic) diseases is increasing, and as it is know that patients value personal contact with a clinician [46, 61], the preferences described may become more frequent.

Self-examination

Another important aspect that was discussed with patients, is their attitude towards self examination. This relates to themselves, potentially with the help of others, examining their body for potential malignant lesions and taking pictures of these lesions. All patients that were interviewed initially mentioned that they are able to check for lesions themselves, however, some concerns were raised. First off, patients expressed that they expect difficulty with assessing what is benign or malignant. Due to a sometimes large number of potential lesions, they do not know what to assess and what to take pictures of. Literature on skin self-examination (SSE) to detect melanoma supports these notions. It is found that patients have difficulty in accurately determining the features of lesions and thus identify the malignancies, which is especially true for people with many lesions [35]. However, when SSE was performed by patients with the help of a partner and after learning about clinical detection rules, the number of identified malignancies increased [82].

Uploading medium
When examining the wishes of interviewees regarding the medium used for uploading data or receiving information, all interviewees mentioned that the medium should be easy to use as was expected based on the stakeholder analysis. What was observed, was that the older patients that were interviewed were less prone to using a new medium than younger patients. A difference in attitude between age groups is what is to be expected based on the article by Zhao et al. [99]. They show that ease of use, perceived vulnerability, and perceived severity are more important to middle-aged and older users. This perceived vulnerability can be translated into the risk of loss of privacy. This notion was however not supported by the interviewees. These expressed no concerns regarding their privacy as long as the regulations set by the government were followed. Furthermore, patients expressed that they valued personal contact when receiving information, which is again in accordance with literature [46, 61].

AI application
As the applications that were proposed from the draft QOC are new to healthcare, the concepts that were uncovered during the data collection phase could not be validated against literature. However, inquiries about the implementation of an AI solution uncovered concepts that were also seen in literature. First off, both the patients as the clinicians mentioned that they have trust in AI as a risk assessment tool as long as it was proven to be accurate. With this, patients imply that there should be no errors as a result of the AI application. They state that no application is 100% accurate but that they expect at least similar results as to when a dermatologist performs the examination. This can be related to the goal of many papers that were mentioned in literature, where the performance (or accuracy) of an AI application in terms of sensitivity and specificity is evaluated against dermatologists. To increase this sense of accuracy, two of the four patients and all clinicians mentioned the condition that a dermatologist should always perform an extra check on whether the assessments made by the AI are correct. One patient mentioned they do not mind whether a check was performed, and the last did not indicate whether this was preferred or not. These findings correspond with the paper of Jutzi et al. [46] where it was found that patients have a high sense of trust in AI but still prefer AI as support tool for clinicians. This is also mentioned by Felmingham et al. [20] where the approach of AI as a support tool prevents the loss of clinical expertise of dermatologists and prevent cognitive bias. Besides this, one of the clinicians mentioned that before an AI should be implemented as a tool, it should be validated against a certain standard. As mentioned by Ben-Israel et al. [3] such standards do not yet exist for healthcare, however, through the increase in accuracy and efficiency of AI, the use of these tools may become justified in the future.

4.4 | QOC Model

Based on the QOC model from literature, the draft QOC model, and the data collection, the 'final' QOC model can be proposed. As mentioned before, the artefact is a design product, meaning that alterations may be made in case these improve the artefact at that time. The evaluation of each QOC element is done through Architectural Knowledge design SPAce Modeling (AK-SPAM). This is a structured overview of the questions and criteria with the rationale for the different options. The overview consists out of the concern (question), ranking criteria (criteria), identifiers (options) with description, status (decided or rejected), relationship to other options/questions, an evaluation of the option against the criteria, and a decision rationale.

4.4.1 | QOC Element I - Process

The first element that is discussed is what process should be enhanced to achieve a higher overall efficiency. The options, criteria and rationale were already given in the draft section and have been summarized in table 4.4. During the data collection phase, clinicians were asked how they experienced the treatment process and one-stop-shop model and whether they saw any enhancements in this process. Both clinicians mentioned that they thought that the process worked particularly well and did not see any new improvements that could increase the efficiency. This opinion was shared by patients who were asked about the treatment process. Together with the notions in the draft section, this results in the treatment process option to be rejected. The rationale from the draft regarding the diagnostics process remains unchanged. The data collection phase did not provide new insights. The data collection phase did however provide new insights regarding the screening process. First off, patients indicated that they saw benefit in enhancing the screening process. However, it was very patient specific whether they thought the screening should be performed remotely or whether a consultation was always necessary. Through patients that

Table 4.4: AK-SPAM of QOC Element I

Concern (Identifier: Description)		*Con#1: What process to enhance?*
Ranking criteria (Identifier: Name)		*Cr#1: Costs* *Cr#2: Number of visits* *Cr#3: Patient satisfaction*
Options	***Identifier: Name***	*Con#1-Opt#1: Treatment process*
	Evaluation	*Cr#1: Operation times get reduced in case the treatments can be performed faster, however, the reduced costs are limited by the fact that MohsA already uses the one-stop-shop model that reduces treatment time significantly* *Cr#2: Patients will need to visit for their treatment anyhow so the number of visits is unaltered.* *Cr#3: Patients may need less time under operation which increases satisfaction.*
	Identifier: Name	*Con#1-Opt#2: Diagnostics process*
	Evaluation	*Cr#1: The clinician performing the diagnoses will still need to perform all regular steps, therefore, the same amount of resources are required and costs stay the same.* *Cr#2: Enhancing the diagnoses may result in less mistakes. As such, the treatments may be more effective resulting in less required visits of patietns* *Cr#3: In case patietnts receive better diagnoses, the satisfaction may increase*
	Identifier: Name	*Con#1-Opt#3: Screening process*
	Evaluation	*Cr#1: Costs are reduced as less patient visits might be needed through remote screening. In addition, the screening can be performed by other resources than the dermatologist that might be cheaper.* *Cr#2: Through enhancing the screening process less frequent visits might be necessary* *Cr#3: By having to visit less frequently or being helped more efficiently patient satisfaction may rise.*

prefer remote consultations, the number of patients will decrease. Furthermore, the patients indicated that they would like to be able to choose whether they visit the clinic or not based on the screening. Due to this, it is expected that the patient satisfaction will increase through the earlier mentioned notion of patient self-control [46].

4.4.2 | QOC Element II - Enhancements

After deciding to enhance the screening process, the question is raised how to enhance this process. The draft QOC already presented the options and several criteria that are considered for this enhancement. One more criterion has been proposed during the data collection phase. The extra criterion against which the options are evaluated is again the patient satisfaction. In order to reach adoption of patients it is necessary that they are satisfied with the changes. The first option that prioritizes and potentially reschedules appointments based on risk was proposed to the interviewees. As explained in the conceptual validation phase, concerns were outed regarding the possibility of assessing the correct lesions which decreases the accuracy. All interviewees, both patients as MohsA employees, indicated that such an improvement would increase the efficiency significantly as less consultations would be necessary. However, the skin therapist as well as two patients mentioned that they were not fond of postponing appointments based on remote data as they doubted the possibility of correct assessment. This means that the patient satisfaction would decrease in case this is enforced. If patients are given the possibility to choose whether to make use of this service the satisfaction may increase as one of the patients was very excited about being able to skip appointments that way. The second option, where the patient data is used to estimate a treatment in order to schedule the right amount of time, was received more positively by both patients as clinicians. A reason for this that was mentioned is the fact that clinicians will always perform an extra real-life check during the appointment, which increases the personal contact and the accuracy. A downside that was mentioned is that the efficiency will decrease in case an appointment is extended for a longer treatment which turns out to be unnecessary.

Table 4.5: AK-SPAM of QOC Element II

Concern (Identifier: Description)		*Con#2: How to enhance the screening process?*
Ranking criteria (Identifier: Name)		*Cr#1: Patient satisfaction* *Cr#2: Time efficiency* *Cr#3: Accuracy*
Options	***Identifier: Name***	*Con#2-Opt#1: Prioritize appointments based on risk*
	Evaluation	*Cr#1: It is patient dependent whether they prefer to regular consultations or whether they don't mind re-scheduling. In case the option is presented that either can be chosen, the patient satisfaction increases* *Cr#2: By rescheduling appointments, redundant consultations are skipped reuslting in an increase in efficiency* *Cr#3: As it is more difficult to assess a lesion by picture, accuracy will decrease*
	Identifier: Name	*Con#2-Opt#2: Estimate treatment based on data*
	Evaluation	*Cr#1: Treatments can be carried out immediately without having to schedule new appointments which increases patient satisfaction* *Cr#2: By scheduling enough time to perform the treatment directly, less appointments are needed. However, if the estimated treatment is unnecessary, time is lost.* *Cr#3: Accuracy remains the same as there is always an extra live check.*

4.4.3 | QOC Element III - Patient Data

The QOC elements that were established from literature (section 2.7) discussed what data to use for the assessment using an AI. As this question is also relevant when there is no AI application, it has been included in the final QOC. As was already seen in the literature section, in case an AI is used, it is best to perform the analyses with both patient clinical data as images in terms of accuracy. The same question was asked to the clinicians during the data collection phase. They indicated that an image is necessary to be able to evaluate a lesion, however, as images might be unclear and do not uncover relevant information, it is very beneficial to also have access to patient clinical data on the lesion. With this information, the clinicians mentioned they would be able to give a more accurate diagnoses. It must be said that they still did not think the accuracy would equal the accuracy achieved during real-life diagnoses. In addition to the clinicians, the administrative employee mentioned that by presenting the patients with a questionnaire, all information required to forward the data to the clinician would probably be collected. Currently they experienced that patients often forgot simple administrative data like the date of birth which lengthens the process. Besides accuracy, a criterion has been added that evaluates the ease with which the data can be collected. This criterion is a bridging criterion from patient satisfaction and is related to ease of use. The ease of use for all users was introduced as a goal during the GQM method. During the interviews, all patients indicated that they would be able to file a questionnaire. However, uploading images was expected to be more challenging. Especially because help would be needed from others to make a picture of lesions on difficult to reach areas of the body. The matter of compatibility was not mentioned explicitly during the interviews, however, different interviewees mentioned different systems like a smartphone or laptop to upload the data. This compares to the question of the GQM of adaptability of the system. In the scenario where both images and clinical patient data are collected, the achievable accuracy is highest. In addition, images are a very useful tool but in case patients are unable to upload these correctly the questionnaire will still provide valuable information.

4.4.4 | QOC Element IV - Data Analysis

After data has been collected, the question is raised who/what will analyse this data. As explained in the draft section, three options are proposed: a dermatologist, a trained nurse/skin therapist, or an AI. Two of the criteria have already been established to be accuracy and time efficiency, one final criterion is proposed by the designer as implementation cost. This criterion is a bridging criterion of the main goal 'costs'. Even though a solution may be accurate and efficient, in case the costs are too high it is not feasible. During the interviews became evident that in some cases, dermatologists or skin therapists already review images that patients send of lesions. As these checks take less amount of time then physical consultations, there will be a slight increase in efficiency if this happens in a standardized way. As dermatologists are already trained in diagnosing lesions and the process is already possible, there are no implementation

Table 4.6: AK-SPAM of QOC Element III

Concern (Identifier: Description)		*Con#3: What type of data to use for training?*
Ranking criteria (Identifier: Name)		*Cr#1: Accuracy* *Cr#2: Easily collectable*
Options	***Identifier: Name***	*Con#3-Opt#1: Images*
	Evaluation	*Cr#1: Can be high in case of good pictures, depends on camera quality* *Cr#2: Patients indicate concerns regarding the possibility to make pictures of the correct lesions.*
	Identifier: Name	*Con#3-Opt#2: Patient clinical data*
	Evaluation	*Cr#1:Merely clinical data can reach high accuracies in case many characteristics are collected* *Cr#2: Patients indicate to be able to file questionnaires. Prever shorter length*
	Identifier: Name	*Con#3-Opt#3: Images and patient clinical data*
	Evaluation	*Cr#1: Using both images and patient clinical data achieves the highest accuracies* *Cr#2: Will take more time as two things need to be uploaded by patients instead of only one.*

costs in this scenario. The skin therapist mentioned during the interview that they were able to analyse lesions since they already had experience working at MohsA for several years. In case more lesions will be screened, more skin therapists or nurses are needed that are able to do so. This results in training costs for these employees. However, in general these employees will be less expensive than dermatologists and will take away some of their burden. Finally, an AI would be most efficiency to analyse the lesions as this is able to analyse high volume data. The accuracy has been proven to be on par or even higher than that of dermatologists, but still there is distrust of patients and employees alike. Furthermore, the costs of building an AI, testing, making regulations, and adapting the process are high. In case a verified AI is available and regulations are developed, these costs will be drastically lower making this option more feasible.

4.4.5 | QOC Element V - Collection Medium

In the draft section, the question was already raised how the data should be collected. At that moment, only the possibility of Zivver was introduced as that is currently available at MohsA. During the data collection phase, both patients as employees were asked about their preferences regarding this matter. The most common option to collect data that was mentioned is through email. The reason for this is that all interviewees would be able to use email relatively easy. However, other options were introduced. One patient proposed the use of an app where the data could be uploaded as this was fast and easy in use. Others opposed this as they were less technically educated. One final option was proposed by the MohsA employees, they mentioned that for them it is easiest to receive the data directly into their EMR through Medicore. This would mean that patients will need to log-in to a website to upload the data and questionnaire. The employees mentioned that this was beneficial as currently, the data needs to be copied and pasted into the EMR manually. This results in extra work and is prone to errors as human mistakes can be made. Two criteria to evaluate these options against were already introduced in the draft section. Additionally, the criterion of implementation costs is introduced by the designer, again as a bridging criterion of the main goal 'costs'. The ease of use is similar in both cases as both Zivver as Medicore require patients to identify themselves. Zivver might be perceived easier as it is similar to a regular mail client as opposed to Medicore. When looking at the privacy of both options, Medicore has more guarantee of privacy. First off, the administrative employee mentioned that patients often reply through regular email instead of Zivver (even though it is indicated that they should use Zivver). This might the case due to patients not understanding the difference. Additionally, emails that are received through Zivver can be accessed by anyone who has access to the MohsA email account. Currently all MohsA employees can access this account. Within Medicore, only the people who are granted access to the patient records can access the data. When looking at the implementation costs, Zivver has no costs as this option is already being used by MohsA. Medicore on the other hand does have implementation costs. Primarily, because Medicore needs to build the page in which the data can be uploaded. For this Medicore will invoice MohsA. At the same time, it is not clear how fast such a page can be built.

Table 4.7: AK-SPAM of QOC Element IV

Concern (Identifier: Description)		*Con#4: What/who analyses the patient data?*
Ranking criteria (Identifier: Name)		*Cr#1: Accuracy* *Cr#2: Time efficiency* *Cr#3: Implementation costs*
Options	***Identifier: Name***	*Con#4-Opt#1: Dermatologist*
	Evaluation	*Cr#1: Dermatologists are highly trained in detecting malignant lesions in real life and are thus expected to make accurate diagnoses. However, pictures and a questionnaire may not show all required features.* *Cr#2: In case dermatologists still need to review all lesions, there is a limited increase in efficiency* *Cr#3: There are no implementation costs as a dermatologist already performs this work*
	Identifier: Name	*Con#4-Opt#2: Trained nurse/skin therapist*
	Evaluation	*Cr#1: It was shown that a trained nurse can achieve a high accuracy when diagnosing lesions based on pictures.* *Cr#2: Since a nurse/skin therapist may be cheaper than a dermatologist, the resources in terms of costs are reduced. In addition it reduces the burden on dermatologists.* *Cr#3: There are implementation costs as the nurses need to be trained.*
	Identifier: Name	*Con#4-Opt#3: Artificial Intelligence*
	Evaluation	*Cr#1: Accuracy has proven to be on par or even better than dermatologists. However, still partial distrust in the system by both clinicians as patients* *Cr#2: No human interference means high efficiency increase. Fast processing* *Cr#3: Building, testing and implementing an AI is time consuming and process disrupting so high implementation costs*

Table 4.8: AK-SPAM of QOC Element V

Concern (Identifier: Description)		*Con#5: Through what medium will patient data be collected?*
Ranking criteria (Identifier: Name)		*Cr#1: Ease of use* *Cr#2: Privacy* *Cr#3: Implementation costs*
Options	***Identifier: Name***	*Con#5-Opt#1: Zivver*
	Evaluation	*Cr#1: As these messages are received through email it is easy to use. However, they need to be forwarded to the right clinician which is human labour and thus prone to error. To reply takes a little more effort for patients as they need to log-in through an online portal* *Cr#2: As Zivver uses two factor authentication the messages can only be read by the recipient and sender. If send to MohsA's general mail everyone with access can view this. Patients tend to reply without Zivver as they don't always understand.* *Cr#3: Zivver is already being used at MohsA and requires no extra implementation costs*
	Identifier: Name	*Con#5-Opt#2: Using the EMR Medicore*
	Evaluation	*Cr#1: This option makes it easier for clinicians and administrative employees to receive the information and less prone to errors as the info is directly in the right place without interference. Patients may need to learn how to log-in and enter the data in medicore as this is different than regular email* *Cr#2: The EMR is developed according to the privacy rules and regulations.* *Cr#3: The page for the uploading of data needs to be developed by Medicore which may be costy. Also, a new process should be described how the incoming data is processed*

Table 4.9: AK-SPAM of QOC Element VI

Concern (Identifier: Description)		*Con#6: Through what medium will patients be informed?*
Ranking criteria (Identifier: Name)		*Cr#1: Ease of use* *Cr#2: Time efficient* *Cr#3: Privacy* *Cr#4: Personal involvement*
Options	***Identifier: Name***	*Con#4-Opt#1: Email*
	Evaluation	*Cr#1: Zivver is easy to use for clinicians/administrative employees as it can be send directly from the mail client. Patients need to use 2FA in order to read message* *Cr#2: Zivver is just as fast as sending an email thus very time efficient.* *Cr#3: Since Zivver uses 2FA and is secure, only patients will be able to read their message* *Cr#4: Email does not allow for asking direct questions or speaking to an expert and is thus not personal*
	Identifier: Name	*Con#4-Opt#2: Phone*
	Evaluation	*Cr#1: For both sides it is easy to use as all people posses a phone and know how to use it. Slight inconvenience if patient does not answer* *Cr#2:Phone calls take relatively long. As such it is not time efficient* *Cr#3: After confirming you are speaking to the correct person it is private* *Cr#4: A phone call allows for personal contact and asking direct questions.*
	Identifier: Name	*Con#4-Opt#3: Patient dependent: phone/zivver*
	Evaluation	*Cr#1: MohsA employees will need to check what is preferred by the patient which reduces ease of use. However, patients can decide what is easiest for them.* *Cr#2: Patients that do not indicate to want a call can just receive an email which is time efficient. By calling the patients that do want to receive a call you can directly answer questions and prevent having to send extra emails to answer these questions* *Cr#3: Both options adhere to privacy regulations* *Cr#4: Patients have self-control as they can decide and get the opportunity of personal contact*

4.4.6 | QOC Element VI - Information Medium

The second question that was introduced regarding the data transfer, is how patients should be informed on the outcomes of the analysis. Initially, two options were proposed, through Zivver or by phone. Since Zivver in itself is not a communication medium but a means of secure email, this option is changed to email. The reason for this is that it might also be possible to send an email from Medicore. In both cases the patients will receive a textual message for which they need to log-in. During the interviews, the interviewees mentioned the criteria that were already proposed in the draft. Therefore, no new criteria are added. However, during the discussions, all interviewees had difficulty with expressing what option would be best as they mentioned that it depended heavily on the situation. One of the reasons that was given by patients, is that even though they prefer personal contact and the possibility to ask questions, information that is straight forward can be communicated by email. There were exceptions, one elderly patient preferred to be called in all cases as this was easier. Another patient was afraid that the emails would not be read as they did not regularly check whether new emails had arrived. From the employee point of view, email was preferred as this is the fastest way of communication. Again, exceptions were mentioned. The administrative employee noted that they prefer to call new patients as these tend to have more questions which can then be answered. In addition, they were aware of the fact that not all patients are technically able to read their emails. Therefore, all interviewees concluded that it would be best if per case was decided whether to call, or send an email. Therefore, this option has been added to the QOC.

4.4.7 | QOC Element VII - Accountability

Finally, the question of accountability as proposed in the draft section has been researched. The accountability element is included as this might help with future development of an AI enabled process.

The four options that were stated in the draft have been evaluated against the two criteria. To do so, the clinicians that were interviewed were asked for their view on the matter. When looking at patient safety, both stated that in their opinion the clinician should always be the one responsible. Primarily, because they have the most knowledge on the subject and should be able to judge what is safe or not. In addition, the clinicians will act as a safety net, by doing extra checks the safety of the system will increase. Furthermore, the accountability of clinicians is already set out. This was explained by the dermatologist who mentioned that in case a clinician can present what considerations were made, and based on what information the final course of action was determined, malpractice can be ruled out. With an AI system that is validated, the outcome of such a system may become sufficient to substantiate for a certain course of action. The other options were not discussed during the interviews, therefore, no data could be collected on these options. As the literature also gives no additional insights, the AK-SPAM of this element was not included.

After all QOC elements have been described and the relations between the options and criteria have been indicated, a final representation of the QOC model can be drawn. This representation shows the tree-like structure of the QOC model. Due to this structure, the model is easily interpretable and shows clearly what next steps follow from certain decisions. The model can be found in figure 4.5.

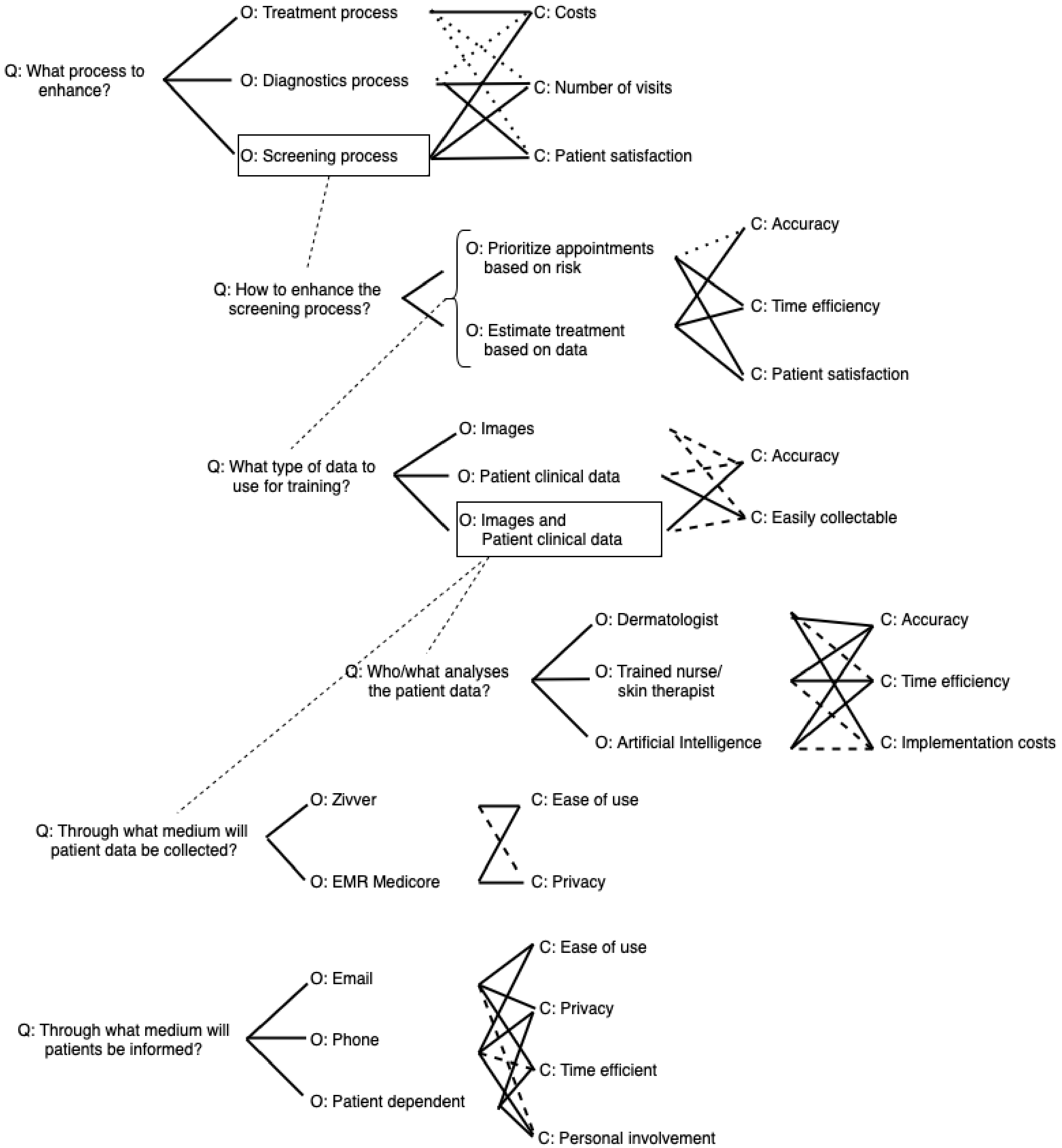


Figure 4.5: Tree-like representation of the QOC model showing relations between options and criteria, and links between options and related questions.

5 | Phase III - Validation

After the artefact design has been completed, it must be validated against the set requirements that were used for the development. The requirements were set in the research design and state 1) "provide context and insight in the decision making activities" and 2) "provide best practices for an AI-enabled process". This section sets out to validate whether the QOC model from the previous section does indeed provide the user with these requirements. In order to do so, a focus-group interview is conducted and the results of this interview are analyzed. These results are then used to propose a redesign of the artefact where needed.

5.1 | Methodology

The validation is performed through conducting a focus-group interview. During a focus-group interview, the setting of the interview allow for a highly interactive group dynamic. Different perspectives and ideas that individuals have can be brought to light through social interaction between participants. This can result in deeper and richer data than that obtained during one-to-one interviews [73]. The participants of a focus group are selected through purposive sampling of a specific population [73]. In this case, this is done by selecting representatives of the different stakeholder groups at MohsA. Since the artefact is used as a support tool for decision making, and the decisions are ultimately made by the caregiver (in this case MohsA), no representatives of the patients or insurance agents were present in the focus group. A list of the participants is given in table 5.1. As can be seen, the focus group consisted out of four participants. Initially, an administrative employee was invited to partake in the focus-group interview, resulting in five participant, however, they were unable to attend at the given moment. Due to this, the focus group is less diverse which may result in more one-sided results. Also, in general a group size of 6-8 people is recommended for a focus group as this is large enough to obtain a variety of perspectives yet small enough to stay manageable [73]. Since MohsA is a relatively small clinic, there were limited possibilities to obtain a focus group of this size. Therefore, it was decided to continue with the focus group of only four participants.

Table 5.1: Participants Focus Group

No.	Stakeholder group	Participant description
1	Dermatologist	Physician at MohsA, not yet specialised dermatologist
2	Dermatologist	Physician at MohsA, not yet specialised dermatologist
3	Data analyst	Skin therapist at MohsA trained in analysing lesions
4	MohsA / Dermatologist	Dermatologist and board member at MohsA

During the focus-group interview, the elements of the QOC were presented to the participants one by one. For each element a set of questions was asked to explore whether in their opinion all options and requirements had been covered. In addition, it was asked whether they would indeed be able to decide for one of the options based on the given information. During the discussions, new insights and relevant remarks were noted down by the researcher. In specific, the researcher looked for the reasoning process of participants and whether they showed a good understanding of the topics covered in the QOC model. This display in understanding and the development of an argument based on the model would indicate that the participants utilise the model in their decision making activities. The process of thinking about all possible options and taking into consideration what is included shows that the model does provide insights in the decision making activities and is thus in line with requirement 1. The focus group was recorded and transcribed in its entirety for further analysis. In order to analyse the transcript efficiently, the data was ordered and reduced. This was done by going through all the statements and linking them to the relevant QOC element. In addition, remarks that were not related to any of the elements were neglected. The resulting data, together with the notes made during the focus-group interview, formed the document used for further analysis. This resulting document was coded using codes that represent themes that are relevant for validation based on the two requirements.

The first theme is the *context* of the model and follows directly from the first requirement. The users of the artefact should be able to use the model at the correct decision moments, to do so the context should be made clear by the model. In addition, the relevance of the decision elements within the problem context should be evident. This relates to whether users are able to know in what scenario the decision model can provide insights for their decision making activities. Secondly, the model should have a high degree of *completeness*. This is related to the second requirement as a solution can only be considered a best practice in case all alternatives are considered. Within the problem context, all relevant questions, options and criteria should be addressed in order to substantiate a decision. Through the omission of elements, the decisions may lead to sub-optimal implementations. The third validation theme is the *contribution* of the model in the decision making activities linking back to requirement 1. This validates whether users feel like they are able to decide on an option based on the information obtained from the model. By looking at the arguments given within the decision model and the relations between different options and criteria, users should be able to come to a decision. In case these arguments and relations are not well substantiated or vague, the contribution of the model decreases. The final validation theme is the *focus* of the model. This represents the actual problem the model tackles, i.e. is it clear that the decision model should be used to provide a solution towards more efficient NMSC healthcare. This is closely related to context in the sense that both themes place the artefact in the real-world scenario. However, the context specifically refers to the current situation while the focus refers to the potential solutions and future problems. As such, this theme follows from both requirement 1 and 2. The themes and their definitions are listed in table 5.2.

After the document was coded for these four themes, another iteration of coding has been performed during which additional validation themes were added that had not yet been covered. These themes followed from statements and questions of the participants that were not placed within one of the validation themes but felt like important to the researcher. During this iteration, one extra theme was defined. From the document analysis became evident that some elements of the model lacked clarity leading to misunderstanding by the participants. Therefore, *clarity* was added as a validation theme. This theme focuses on whether it is clear what the elements represent and what their effects will be in order to provide better context and insights (requirement 1). This theme is indicated in purple in table 5.2.

Table 5.2: Validation themes

Theme	Definition	Requirement
Context	Degree in which the model can be placed in context of the decision	1
Completeness	Degree in which all relevant elements are included in the model	2
Contribution	Degree to which the model contributes in decision making	1
Focus	Focus of the model (i.e.the problem that is tackled)	1 & 2
Clarity	Degree to which it is clear what the elements represent	1

5.2 | Results

During the qualitative data analysis performed, five validation themes were revealed. The results will address these five themes and discuss whether the focus-group interview supports the validity in their context.

Context
Throughout the focus-group interview, this theme was raised multiple times by several participants. Especially the representative of MohsA showed understanding of the context in which the artefact should be used based on the given information. This was evident as at instances where any of the other participants wrongfully interpreted the context or posed a question regarding this theme, the representative of MohsA was able to explain the context such that the others were able to understand this. However, this does show that this extra explanation was necessary in order for other participants to understand the context of the issue. In specific, the artefact should be applied in the context of chronic non-melanoma skin cancer which was unclear to some of the participants. Instead, they interpreted it in the context of all skin cancer cases. Therefore, clearly defining this context might increase the validity. Furthermore, all participants were able to place the decision elements into the context of their daily work. They occasionally linked the decision elements to examples of their daily work thereby also displaying the relevance of the artefact in the MohsA context.

Completeness
During the focus-group interview the question was asked for every element of the decision model whether the participants felt like all relevant options and criteria were included. In addition, participants were asked whether they missed any decisions making activities they imagined would be relevant for coming to an overall solution design. During the discussion of the first element of the QOC model, the participants proposed to look into the usage of data for postponing patient consults in order to improve the healthcare

efficiency. However, these options were already included in the next element of the model. This validates that these options are indeed relevant and show that the model was already complete in that sense. When these options were discussed in more detail during the discussion of the second element of the decision model, two of the participants stated that the criterion of costs should be added. They argued that in order to select the best option, it would be relevant to know the difference in the gain in costs. The representative of MohsA extended this by stating the societal costs should be considered as elderly patients often require the assistance of a caregiver to visit appointments.

During the discussion of the fifth and sixth element of the decision model, which focus on the data collection and information media, additional options were proposed that had not been included in the model. First off, one of the participants proposed the development of a smartphone app in order to collect data. The argument was that this would be easy and fast to use. Others argued that an app might pose problems to eldery that are less tech savvy. During the discussion of the sixth element another additional option was proposed. One of the participants proposed to use video messages that provide a clear explanation on the decisions and frequently asked questions. All of the participants agreed that this would indeed be a valid option to be considered and positively responded to the inclusion of this option in the decision model. Still, issues were raised regarding the relation of this option with the criteria. For example on how elderly will experience this with regards to ease of use. For both of these additional options, research must be done into the relation between the criteria and the options in order to increase the completeness. The participants had no additional remarks regarding the other elements and did not introduce new questions that should be answered. Therefore, with the addition of the aforementioned, the model represents a complete overview of elements.

Contribution
During the focus-group interview, participants were asked whether the information given would contribute to their decision making process. Additionally, attention was paid towards remarks that displayed that the participants favoured certain options based on the decision model. For most options, the model was able to contribute to the decision making of the participants. They used arguments from the model to explain what decision they would make. These arguments did however differentiate per participant. It became evident that some participants weighed certain criteria heavier than others. This led to a different approach of the decision problem and in some cases to a different outcome. Additionally, the model helped the participants in understanding the different perspectives of stakeholders and relate to their requirements. This validates that the model contributes to balanced decision making.

At the same time, participants indicated that for some of the elements of the decision model more evidence was needed for them to make a substantiated decision. The elements that included the costs as a criteria raised the discussion on how these costs would turn out in practice. One participant mentioned for multiple elements that the decided option should be feasible within a reasonable amount of time. As the model currently only presents a raw overview of the difference in costs between options, they mentioned that it is difficult to decide whether an investment is worth it. Therefore, in order to provide a better contribution in the decision making process, more evidence should be provided. However, as the decision model is to be used as a guiding tool towards substantiated decision making, and not as a performance evaluation tool, it is impossible to provide the full evidence for all options considered. Additional testing of potential options may provide the user with the insights needed for the final decision making. Even though the participant preferred to have extra information, the discussion that took place and their approach show that the model helps them in considering all aspects that might be of importance. In this sense, the decision model does contribute to the decision making process.

Focus
The focus of the model was validated by analysing the big picture of the focus group. By going through all arguments and questions that are raised throughout the focus-group interview, the overall conception should be that the participants understand what the purpose of the model is and why it is relevant for future implementation. This is mainly the case for the MohsA representative, which is to be expected as they are also the problem owner and thus raised the issue. In several occasions, they expressed the ultimate goal of the project and what was to be achieved by the options that were discussed. The other participants displayed understanding in the sense that they gave arguments that were relevant to the focus of the project and proposed options that ultimately had the goal of increasing the healthcare efficiency. In addition, the participants were able to present each other with arguments that support the goal of the project in case any questions were raised by others. Overall, these aspects validate that the focus of the model was clear to the participants.

Clarity
Finally, it was validated whether the model showed enough clarity to be used without additional aid. This validation theme was added as it became evident during the discussion of the first two elements of the decision model that the participants required extra information in order to correctly interpret the options. In specific, for the first element it was unclear how the different processes were made up. They understood the concept but could not clearly define what different tasks belonged to which processes as these overlapped at some stages. In addition, the options that were proposed in the second element in order to enhance the screening process lacked specificity. One participant mentioned that they believed that estimating the treatment based on data would be part of the treatment process. However, since it will ultimately be the screening process that alters whilst the treatment process remains the same, this was refuted. The participants had additional questions regarding the practical implementation. In order to enhance the clarity of the process, additional information should be provided that clearly defines what is being meant. After clarification of the first two elements of the decision model, the participants of the focus group understood what was meant and no extra clarifications were needed for later elements.

5.2.1 | Conclusion

Based on the results that were found during the QDA, it can be concluded that the decision model overall does comply with both requirements. The first requirement was satisfied due to the fact that the participants were able to place the decision model within their daily context. Also, their reasoning and arguments displayed that they used the insights provided by the decision model in their decision process. Even so, the model did lack clarity in some aspects which must be resolved to fully provide the correct context and satisfy the requirement. The second requirement was satisfied by the fact that overall the model provided a complete overview of the possibilities and thus what the best practices might be. Additionally, the focus of the decision model was clear which helps the user with their decision to what might be the best solution to their problem. Still, as was mentioned, several additions to the model must be made to cover the full range of options and increase the completeness. Only when all options and criteria are taken into consideration can the best practice be revealed.

5.3 | Artefact Redesign

After validation of the artefact based on the five themes, the results are used to propose a redesign in order to improve the decision model. First off, it is proposed that the additional options and criteria that were mentioned should be added to the decision model. This will result in the following changes: 1) the second element of the decision model (“How to enhance the screening process?”) should include the criterion ’costs’. This criterion should evaluate the potential financial gain of the two options. Currently, only the gain in efficiency is considered, however, an option may prove to be very efficient whilst being utterly expensive deeming it unfeasible. In order to evaluate this criterion, the costs of a consult should be examined and with that how much money can be saved by reducing this number of consults. 2) The fifth element of the decision model (“Through what medium will patient data be collected?”) should include the option ’Smartphone app’. This option relates to the development of a MohsA specific application that smartphone users can use to upload their data. Even though the focus group already outed their concerns regarding the ease of use for elderly, the option may prove to score high on the other criteria. Therefore, the exploration of this option will provide a more complete decision model. 3) The sixth element of the decision model (“Through what medium will patients be informed?”) should include the option ’video message’. This option obtained positive response from all participants of the focus group. Therefore, in order to generate a complete decision model, the option should be evaluated against the set criteria with representatives of all stakeholder groups. Since eventually the patients will be influenced most by this change, their opinions should be considered heavily. A redesigned image of the decision model (as was displayed in figure 4.5) can be found in figure 5.1.

In addition to the additional options and criteria of the decision model, an extra redesign is proposed. In order to clarify the definition of the questions and options, it is proposed to extend the information that is presented with the rationales of the decision model. In specific, this information should define what tasks are considered in the different processes. Currently, these tasks are considered to be the physical examination and inquiry about lesions for the screening process, the diagnosing and performing a biopsy for the diagnostics process, and performing treatment and after care for the treatment process. These clarifications have been added in the QOC elements through adding a ’description’ to the options. The resulting AK-SPAM element can be found in table 5.3. Similarly, it should be made clear that the option

Table 5.3: Redesigned AK-SPAM of QOC Element I

Concern (Identifier: Description)		*Con#1: What process to enhance?*
Ranking criteria (Identifier: Name)		*Cr#1: Costs* *Cr#2: Number of visits* *Cr#3: Patient satisfaction*
Options	***Identifier: Name***	*Con#1-Opt#1: Treatment process*
	Description	*The process that includes the treatment tasks and after-care*
	Evaluation	*Cr#1: Operation times get reduced in case the treatments can be performed faster, however, the reduced costs are limited by the fact that MohsA already uses the one-stop-shop model that reduces treatment time significantly* *Cr#2: Patients will need to visit for their treatment anyhow so the number of visits is unaltered.* *Cr#3: Patients may need less time under operation which increases satisfaction.*
	Identifier: Name	*Con#1-Opt#2: Diagnostics process*
	Description	*The process that includes the tasks that concern the dianosis i.e. biopsy and diagnosing lesion*
	Evaluation	*Cr#1: The clinician performing the diagnoses will still need to perform all regular steps, therefore, the same amount of resources are required and costs stay the same.* *Cr#2: Enhancing the diagnoses may result in less mistakes. As such, the treatments may be more effective resulting in less required visits of patietns* *Cr#3: In case patietnts receive better diagnoses, the satisfaction may increase*
	Identifier: Name	*Con#1-Opt#3: Screening process*
	Description	*The process that includes physical examination and inquiry about lesions*
	Evaluation	*Cr#1: Costs are reduced as less patient visits might be needed through remote screening. In addition, the screening can be performed by other resources than the dermatologist that might be cheaper.* *Cr#2: Through enhancing the screening process less frequent visits might be necessary* *Cr#3: By having to visit less frequently or being helped more efficiently patient satisfaction may rise.*

'Estimate treatment based on data' is part of the screening process. This option incorporates a change in the screening process as the estimation is done using data that was collected in the screening process. At the same time, the treatment process remains the same. The results of the focus group that there was confusion about the process this option belongs to. Additionally, it should be made clearer in the model that the options proposed for enhancing the treatment process both incorporate patient data. This results in the AK-SPAM element that can be found in table 5.4.

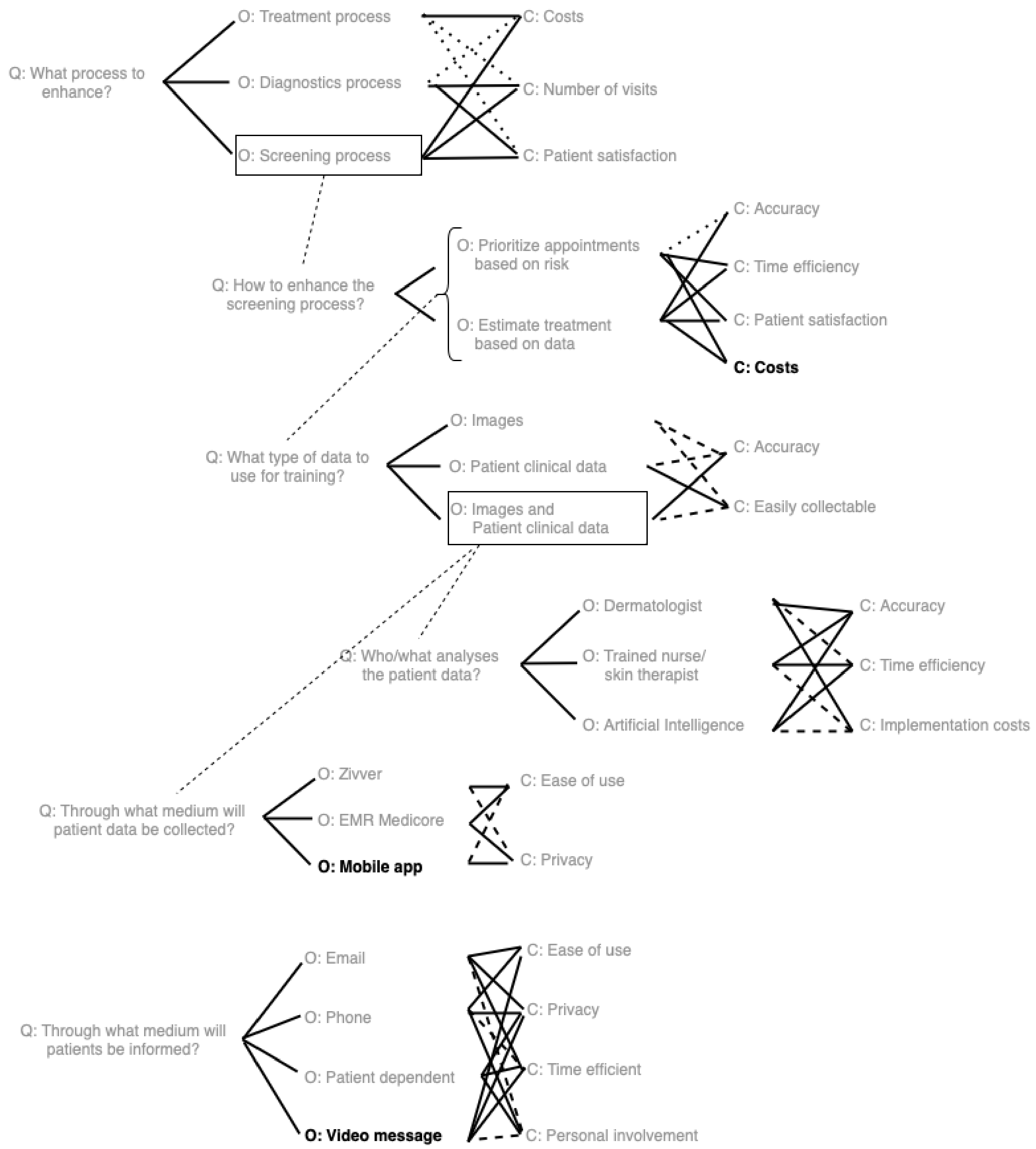


Figure 5.1: Redesign of the tree-like representation of the QOC model based on the findings of the validation. Added elements are shown in bold.

Table 5.4: Redesigned AK-SPAM of QOC Element II

Concern (Identifier: Description)		*Con#2: How to enhance the screening process using patient data?*
Ranking criteria (Identifier: Name)		*Cr#1: Patient satisfaction* *Cr#2: Time efficiency* *Cr#3: Accuracy*
Options	**Identifier: Name**	*Con#2-Opt#1: Prioritize appointments based on risk analysis of data*
	Evaluation	*Cr#1: It is patient dependent whether they prefer to regular consultations or whether they don't mind re-scheduling. In case the option is presented that either can be chosen, the patient satisfaction increases* *Cr#2: By rescheduling appointments, redundant consultations are skipped reuslting in an increase in efficiency* *Cr#3: As it is more difficult to assess a lesion by picture, accuracy will decrease*
	Identifier: Name	*Con#2-Opt#2: Estimate and schedule treatment based on risk analysis of data*
	Evaluation	*Cr#1: Treatments can be carried out immediately without having to schedule new appointments which increases patient satisfaction* *Cr#2: By scheduling enough time to perform the treatment directly, less appointments are needed. However, if the estimated treatment is unnecessary, time is lost.* *Cr#3: Accuracy remains the same as there is always an extra live check.*

6 | Testing

The validation of the artefact has given insight in the contribution of the decision model in the decision making activities. The implementation of the decision model should lead to effective decision making and a clear reasoning process. To assist with this reasoning process, it is beneficial to gain insights in the expected results that the options in the decision model might produce. This is in line with the second requirement of the artefact: "provide best practices for an AI-enables process". When an indication can be given of the expected outcomes, the possibilities can be compared and the best practice can be identified. To provide these insights, this section sets out to combine the decision model and results from the focus group together with the earlier developed BPM. First off, the decision model is presented to representatives of MohsA in order to decide on the specific options that should be implemented in the process redesign in section 6.1. For every element, the selected options are used for the redesign decisions made in section 6.2. After the redesigned process is complete, a simulation is performed of both the redesigned and the current process (as defined in section 3.3). Finally, the results show whether the redesigned process can indeed enhance their current process and it is discussed how this could additionally assist the decision making process.

6.1 | Decision Process

The original QOC model from section 4.4 was presented to the problem owner and a dermatologist at MohsA. This session was held separately from the focus group intentionally to be able to focus on the outcome of the decision model instead of on the decision making process. The findings of the decision model were presented and based on these findings the options that deemed most feasible by the problem owner and dermatologist have been selected for the simulation. The first element, regarding what process to enhance, resulted in little discussion. From the start, the problem owner expected to benefit the most by enhancing the screening process. The findings supported this notion as this option displayed the most support for decreasing the number of visits and decreasing the costs. In addition, it was found that the patient satisfaction may also benefit from this option. As a result, this option was decided and the next option was presented.

The second element, on how to enhance the screening process, resulted in a more vivid decision. The goal to increase the efficiency was considered most important. Earlier on it was already mentioned that the accuracy with which NMSC is detected is often not that relevant as most lesions pose little health threat. Postponing the treatment of a lesion like AK or BCC a few months will not have disastrous results. Following this reasoning, the option where appointments are prioritized based on risk would be most beneficial. The other option, however, was also supported. Most gain was expected if a combination of the two options could be implemented. What was indicated is that patients should always be able to decide whether or not they prefer to make use of this process. This is in accordance with the patients wishes so that the patient satisfaction does not drop.

The data type that should be used for training was decided upon easily. Since at MohsA earlier research had already indicated that the usage of images as well as the usage of patient clinical data separately provides useful information for predicting the type of lesion, their idea was that by using both of these types of information would provide even more insight. The downside that collection of both types of data requires more patient effort is disputed by the fact that there are no costs related to this whereas longer decision making due to a lack of data does inflict more costs. Therefore, the option of collecting both types of data is decided.

For the fourth element, what/who analyses the patient data, there was again one option that raised the most interest. This was the option of an AI analysing the data. The reason this option received the most interest was primarily since there was a lot of curiosity into how well this could potentially function when implemented. Also, as the other two options would require significantly more human labour, the option of an AI would present the most beneficial outcome for MohsA. The downside of the implementation costs that are present at this time were considered, however, the implementation of the new process was not due in very short time and therefore they preferred to explore the option further to see whether it is feasible. A trained nurse/skin therapist was proposed to assist in the early stages of exploration.

The fifth and sixth element of the QOC were discussed shortly. The collection of data was preferred through Zivver as this is already being used and thus easily implementable. However, the designer recommended to delve deeper into the option of using the EMR as the advantages in the long run were expected to outweigh the implementation costs. Regarding the information medium to contact patients a consensus was easily reached that it should be patient specific. As this is also the current way of working this would require no alterations in the process.

6.2 | Process Redesign

The options that were deemed most interesting during the decision process have been used to propose a new business process. This process results in the redesign of the process that was defined in section 3.3. In order to do so, several design decisions have been made that can be substantiated by the reasoning of the decision process. The process will be redesigned in order to implement the enhancement of the screening process by usage of an AI. The decisions have been described below and have been implemented in the BPMN diagram. The first alteration of the process is the addition of the task to request patient data which starts two weeks prior to the appointment. This is the first task that happens in the redesigned process. Afterwards, on of two things can happen, 1) patients send their data within one week and the process continues with this data or 2) a week passes in which no data is received. In the second case the process will be carried out further in the same way as the current process. This redesign has been displayed in figure 6.1.

After the data has been received by MohsA, the administrative employee will need to schedule the check of the data. Currently, when data comes in, a 'work meeting' of around 10 minutes is scheduled with a skin therapist or dermatologist to carry out the analysis. As an AI is able to analyze large chunks of data, multiple cases can be processed together in a short amount of time. After analysis, the patient is informed on the outcome and further steps. Since it was decided to contact patients in the way they prefer, the process is split for people who want to be contacted by phone and people who want to be emailed. In case patients do not answer the phone, they will still receive an email with the information.

Based on how the lesions have been assessed, either of three paths can be chosen. These are based on both of the options that were discussed in the QOC, the prioritizing of appointments based on risk, and the estimation of the treatment. First off, in case the risk is determined to be low, the appointment can be canceled and a new check-up can be scheduled, in which case the process starts over. In the second scenario, there is a relative certainty with which the lesion can be classified as being malignant. In that case, a treatment plan can be estimated and the appointment can be rescheduled to ensure that the treatment can be executed immediately. Finally, in case the data does not reveal what type of lesion it is with enough certainty, the appointment can remain and the tasks continue as in the current process.

When the process continues as in the second scenario (established treatment), the tasks of inquiring about lesions and physical examination have already been performed and documented. Therefore, only an

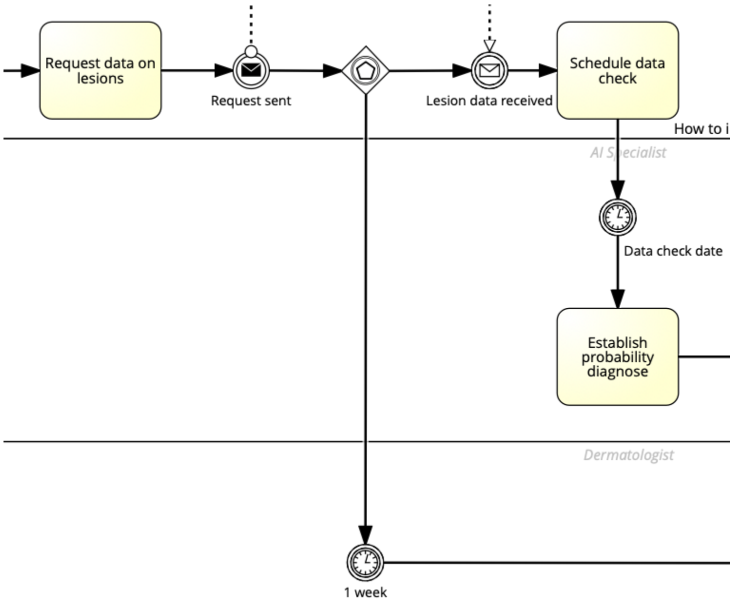


Figure 6.1: BPMN process redesign implementing the data request and analysis

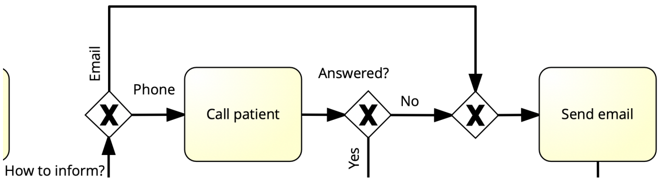


Figure 6.2: BPMN process redesign implementing the way patients are informed

additional check to establish a probability diagnose in real life needs to be performed. In some cases, it might turn out that no treatment is necessary and the initial assessment was incorrect resulting in lost time. The patient will then again schedule a new check-up and restart the process. In case the other cases, the process continues as usual but with the exception that now all treatments can be carried out immediately. This ensures shorter throughput times as the patients will not need to schedule a new appointment for the treatment. In order to achieve reliable results during the simulation, the treatments have been split into treatments where excision is performed and others. These other treatments can include for example cryotherapy or the presctiption of a medical ointment. The reason that a distinction is made between these treatment tasks, is that these processes differ significantly in duration. Therefore, it is necessary that they are performed in separate tasks.

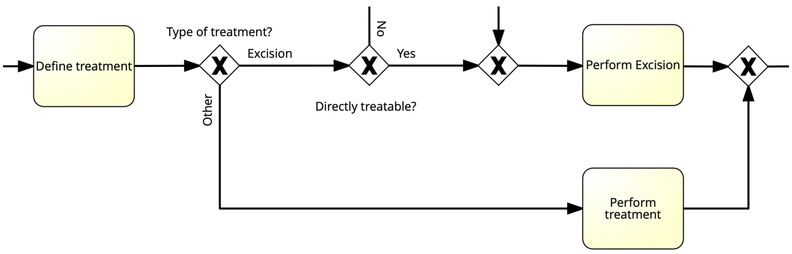


Figure 6.3: BPMN process redesign differentiation between treatments

One final redesign follows from the fact that patients that have not had NMSC for five consecutive years, are discharged from the regular checks. Therefore, the path has been added where patients that fulfil this criteria end the process. In this way, patients are not stuck in an infinite loop. The full BPMN diagram can be found in appendix B.

6.3 | Simulation

The redesigned process can be used for performing the simulation of the proposed scenario. This simulation is performed through the usage of the BPMN software Signavio. Signavio has been chosen as a tool as it is simple and adequate to the task at hand. There are some limitations to the extensiveness of Signavio, one of these limitations is that it is not possible to model the queues. Queues form when a patient arrives at a task but there is no resource to fulfil that task. In real-life, there would be guides for how to handle long queues or what processes to prioritize over others in case of queues, in Signavio this is not possible. In addition, any waiting times that occur in the process must be modeled as being constant. As some of the waiting times vary in real-life, this will reduce the accuracy of the simulation. However, as the goal of the simulation in this project is to provide additional insight into the potential gain in efficiency, these cons are considered negligible. In addition, the parameters that are used for the simulation (e.g. task durations, number of cases per week, number of cases per split) have been estimated by the employees at MohsA and will thus not be 100% accurate.

Finally, simulations in Signavio produce the exact same results every run. Normally a simulation is used to simulate the randomness of a real-life situation. Multiple runs can be used to approach the true value of the mean μ. In addition, by performing multiple runs a standard deviation s of the observed mean $\bar{x}$ can be calculated. In this way a confidence interval can be provided in which μ lies with a certainty $1-\alpha$. By obtaining the exact same observations of the simulation every run, these statistics cannot be applied. On the other side, due to the low computing capacity Signavio requires, a single run can be performed very fast regardless of the length of the simulation. Since in general a longer simulation results in a better estimate of the true mean, it is decided to perform simulations of 365 days. In this way, 52 weeks pass during which patients enter the simulation. To further improve the significance of the results of the simulations, it is decided to simulate three scenarios for both the redesigned process as for the current process. These scenarios constitue of a lower-bound, in which the optimal scenario is estimated, an upper-bound, in which the worst-case scenario is estimated, and an average that balances between the two.

6.3.1 | Performance Indicators

In order to compare the redesigned process to the current process, a set of performance indicators have been selected to test the performance on. As the goal is to test whether the efficiency of the healthcare increases in the redesigned process, the performance is tested against the resource consumption and the cycle time. The resource consumption relates to the percentage of time that the resources (employees) are performing tasks. This indicator has been selected to test whether the burden that is on dermatologists will decrease in case the new process is implemented. The cycle time refers to the total time it takes for a patient to go from the start of the process (request data on lesions) to the end of the process (schedule new check-up). This indicator is selected to test whether the number of patient visits decreases in the redesigned process. It is expected that through fewer visit, the total cycle time will reduce.

6.3.2 | Simulation Parameters Redesigned Process

Before performing the simulation, the waiting times, task durations, resources and frequencies have been estimated. The waiting times are the durations of events that require patients or employees to wait before the process can continue. This can be for example due to a scheduled treatment in the future. Task durations are simply the time it takes for a certain task to be completed. The resources are the employees that perform these tasks. These are measured in number of hours per week based on a daily working schedule. The frequencies refer to the number of cases that come in and the percentage of cases passing through each option at the splits. Different simulations have been performed in order to generate an upper-bound, lower-bound, and average based on different values of the simulation parameters. In the lower-bound simulation, the 'ideal' scenario is approached where the waiting times and task durations are as short as possible whereas the upper-bound simulation does the opposite. The estimates for the redesigned process are explained below. Afterwards, the estimates for the current process are given.

Tasks
The process consists out of 16 tasks. These tasks can be modeled in three ways: as being constant, as following a uniform distribution or as following a normal distribution. First off, the tasks of the administrative employee are discussed. These tasks are generally of shorter nature as they mainly involve sending emails or scheduling appointments. The duration of scheduling an appointment is depending

on the speed with which patients can provide their availability. Most new appointments at MohsA are scheduled at the clinic directly after an appointment. In that case this roughly takes 5 minutes. In case patients need to be contacted about their availability, it is estimated that these tasks take around 10 minutes. Both these tasks are modeled using a normal distribution with a standard deviation of 2 minutes in order to simulate the randomness of a real-world scenario. This results in the task 'Reschedule appointment' to take 10 minutes with a s.d. of 2 minutes as in this scenario all patients need to be contacted for their availability, 'Schedule treatment' to be 5 minutes with a s.d. of 2 minutes as in this scenario all patients come from an appointment, and 'Schedule new check-up' consists out of both scenarios resulting in the average of 7.5 minutes with a standard deviation of 2 minutes. No distinction is made between the lower-bound and upper-bound since these times are considered to be constant and the randomness is already incorporated by using a normal distribution. The task 'schedule data check' can be fast as this only requires the schedule of the employees which is available to the administrative employee. Therefore, it is expected that this can be done in 2 minutes.

The other tasks of the administrative employee are related to contacting or informing the patients. The first task, 'Request lesion data', can be performed through standard emails and in batch. Therefore, it is expected that this task will only take around 2 minutes per patient. Next, the task 'Call patient' will probably take longer as this call will not only inform patients, but provide them with an opportunity to ask some questions about the given information. The lower-bound of this task is set to 10 minutes with a standard deviation of 2 minutes, the upper bound is set to 15 minutes with a standard deviation of 2 minutes and the average thus to 12:30 minutes with a standard deviation of 2 minutes. Again, this task is currently not performed and therefore estimated based on the current time it takes to contact patients. The final task 'Send email' can again be done through standardised messages that can be generated for the most common cases. Still, since every case needs to be reviewed individually, it is expected that this will take around 5 minutes. The emailing tasks do not have lower and upper bounds as these can be considered as standardised tasks where deviations from the average are considered negligible.

The AI specialist only performs one task which is 'Establish probability diagnosis'. As the AI will perform the true risk analysis, the AI specialist is only required to feed the data into the algorithm, and interpret the results. The analysis of the cases can be performed in batch as an AI is able to handle large amounts of data. Currently, a ten minute 'work meeting' is scheduled when patient data needs to be analysed by a dermatologist or skin therapist. Since it is expected that the use of an AI will accelerate this process, a lower-bound of 4 minutes and an upper-bound of 8 minutes. This results in an average of 6 minutes.

The remaining 8 tasks are performed by the dermatologist. The first three tasks are performed during the consult. For these consults 10 minutes are scheduled, therefore the task times of these three tasks should add up to 10 minutes. These 10 minute consultations are scheduled consecutively, therefore, if a consult needs to be extended this needs to be made up by shortening another. Due to this, there is no difference between the lower-bound, upper-bound and average as there is no wiggle room. In some cases, it is necessary to perform a biopsy. In these cases, the biopsy is performed during the consult. Since this will not require extra time but not all patients require a biopsy, it has been decided that to model the task 'Perform biopsy' as being instantaneous. In this way, the 10 minute consults are not exceeded. Based on the results of the biopsy a diagnosis is made. This diagnosis is modeled using a normal distribution with a mean of 7 minutes and a standard deviation of 2 minutes. This is based on the current 'work meetings' that are scheduled of 10 minutes to diagnose lesions based on data. The major part of this task will thus be able to be performed within those 10 minutes. Since dermatologists will be able to continue with another task early if these diagnoses are performed within these 10 minutes, it is decided to model the exact time instead of the scheduled time.

The final tasks are related to the actual treatment of the patients. The first task, 'Define treatment', is performed directly after the lesion has been diagnosed or the probability diagnosis has been established. This task is performed during the time that is scheduled for the consults or lesion diagnosis. Therefore, it is decided to model this task as being instantaneous. Following, there are two options for the treatment. In case no excision is necessary but another treatment is performed (e.g. cryotherapy, ointment, curetage etc.), this treatment can again be performed during the 10 minute consult. Therefore, 'Perform treatment' is modeled as being instantaneous. When an excision needs to be performed, there are two different cases. Either the excision is in the face, in which case an appointment of 30 minutes is scheduled, or the excision is somewhere else on the body, in which case an appointment of 20 minutes is scheduled. This results in a lower-bound of 20 minutes, upper-bound of 30 minutes and average of 25 minutes assuming a uniform distribution.

Table 6.1: Estimates of the lower-bound, upper-bound, and average of the tasks *(hh:mm)*. In case the task is modeled using a normal distribution, the standard deviation is indicated by σ

Event	Lower-bound	Upper-bound	Average
Reschedule appointment	00:10 (σ=00:02)	00:10 (σ=00:02)	00:10 (σ=00:02)
Schedule treatment	00:05 (σ=00:02)	00:05 (σ=00:02)	00:05 (σ=00:02)
Schedule new check-up	00:07:30 (σ=00:02)	00:07:30 (σ=00:02)	00:07:30 (σ=00:02)
Schedule data check	00:02	00:02	00:02
Request lesion data	00:02	00:02	00:02
Call patient	00:10 (σ=00:02)	00:15 (σ=00:02)	00:12:30 (σ=00:02)
Send email	00:05	00:05	00:05
Establish probability diagnosis (AI)	00:04	00:08	00:06
Inquire about lesions	00:01	00:01	00:01
Physical examination	00:04	00:04	00:04
Establish probability diagnosis (dermatologist)	00:05	00:05	00:05
Perform biopsy	00:00	00:00	00:00
Diagnose lesion	00:07 (σ = 00:02)	00:07 (σ = 00:02)	00:07 (σ = 00:02)
Define treatment	00:00	00:00	00:00
Perform treatment	00:00	00:00	00:00
Perform excision	00:20	00:30	00:25

Frequencies

The frequencies relate to the number of cases that pass through every loop of the process expressed in percentages. This is initiated by the number of cases that come in. In order to represent the busy schedule at MohsA, this number has been set to 30 patients per day. The first split that occurs relates to the number of patients that send data on their lesions versus the number that does not. The results from the interviews with patients show that whether patients are willing to do so is related to their ability to do so technically, their trust in the system, and whether they can get help with performing self-examination. In the worst-case scenario (upper-bound) it is assumed that only 30% of the patients will make use of the service. In the best-case scenario (lower-bound) 70% is assumed giving an average of 50%. This average of 50% is also found in a paper by van Elburg et al. [88] where the intention to use mHealth applications was researched among elderly people. Over the years, this number of people that is willing to make use of the service is expected to increase due to elderly people becoming more tech savvy and gaining knowledge on AI applications.

Afterwards, there is a split in how people would like to be informed. Since currently the patients at MohsA get informed primarily per email without any issues being raised, it is assumed that this will continue especially if some sort of video message with extended information is included. Therefore, an estimation is made that on average only 20% of people will prefer to be called. An upper-bound and lower-bound of 10% and 30% are taken to simulate the best and worst case scenario respectively. These same percentages are used for the split of the people that answered the phone where on average 20% did answer and 80% did not.

The following splits are related to the risk of the lesions. The estimates for these splits are based on a dataset that was provided by MohsA in which the results of 100 oncological consults were documented. This includes the type of diagnosis, the treatment and the anamnesis. The first split being discussed occurs based on the received patient data. From the interviews with the dermatologists came forward that a part of the received data is not accurate enough to base a decision on. It is assumed that this is 20%. The other 80% is split based on the type of lesions in the dataset. From this dataset is evident that currently in around 60% of the cases, the treatment is small and could have been postponed to a later moment or no treatment is needed at all. The other 40% is thus related to more risky lesions needing attention. This results in a split of 20% uncertain, 48% no risk, and 32% defined risk. The second split regarding treatment occurs after the probability diagnosis has been established in real-life. Again, the decision is made between no risk, high risk requiring surgery, or low risk resulting in another type of treatment. Since a part of the patients that arrive here are already expected to have a lesion that requires immediate attention, these patients do not require an additional biopsy and can continue through the 'easily treatable' path even though excision is necessary. By combining the cases from the dataset with the earlier defined split percentages, the percentages for this split have been estimated.

Finally, the splits occur whether an excision is necessary or not and whether this can be performed immediately. In both the lower-bound and upper-bound scenario the number of required excisions should be the same as even though the path might be different, the high risk cases will still arrive there. However, in the lower-bound scenario more people will have had their appointment rescheduled meaning that their excision can be performed immediately as opposed to the upper-bound scenario. A small part of the excisions that had not been scheduled beforehand can be performed immediately. This results in a small percentage of the high risk group together with the part that needs excision of the low risk group to be able to be treated immediately. The last split is the split where a part of the patients is discharged from the yearly check-ups as they have not had NMSC for 5 years. It is decided to disregard this split as the simulation is only run for a time period of 1 year meaning that the next cycle of check-ups will not start.

Table 6.2: Estimates of the lower-bound, upper-bound, and average of the frequencies

Decision	Lower-bound	Upper-bound	Average
No. of incoming patients per day	30	30	30
Lesion data received? (Yes)	70%	30%	50%
How to inform? (Email)	90%	70%	80%
Answered? (Yes)	90%	70%	80%
Lesion risk? (No risk)	48%	48%	48%
Lesion risk? (Defined risk)	32%	32%	32%
Lesion risk? (Uncertain)	20%	20%	20%
Type of lesion? (No risk)	30%	50%	40%
Type of lesion? (High risk)	10%	20%	15%
Type of lesion? (Low risk)	60%	30%	45%
Type of treatment? (Excision)	55%	45%	50%
Directly treatable? (Yes)	80%	50%	65%
Five years without NMSC?	-	-	-

Waiting times

The waiting events are all events where the process is paused until the waiting time has passed. Two types of waiting events have been modeled, the first type is a timer event where a constant time is set to wait e.g. for a treatment date, the second type is a message event where there is a constant waiting time for receipt of a message. Normally, these times would not be constant but be case depended. However, as explained above, Signavio does not allow modeling variable waiting times. The first waiting event in the model is the receipt of lesion data from the patient after a request sent by MohsA. As this event is newly introduced in the redesigned process, an estimate must be made based on reasoning. The patients will receive one week for sending their information, therefore, the upper bound of this event is just before this setting it as 6 days. It can be reasoned that patients will not send this data instantaneous but rather in the evening when they are at home, therefore the lower bound is set to one day. The average is set to 3.5 days. After receiving the patient data, the check is scheduled. In the new scenario can be expected that this check will be performed daily say around 12.00 PM. Therefore, the maximum waiting time is 1 day. Provided that the employees cannot check the lesion data instantaneous, the minimum waiting time is set to 1 hour giving an average of 12:30 hours when a uniform distribution is expected. In case no data is received before 1 week, the other path is chosen which results in 1 week waiting time no matter the scenario.

After the analysis of data has been completed or no data is received, the appointments take place. There are two waiting events that take place before the appointments. The first event 'appointment date' is the remainder of the waiting time from the start, which is 2 weeks prior to the appointment, until the actual appointment. If no data is received, this is exactly one week no matter the scenario. However, in case data is received but the patients still need to visit the appointment, the remainder of the waiting time is 2 weeks - lesion data received - data check date - intermediate tasks. The total duration of these intermediate tasks is neglected. These values need to be averaged using the split frequencies that have been defined above to calculate the final waiting time. The second event 'Rescheduled date' is waiting time before the treatment can be scheduled. This is dependent on the type of lesion at MohsA. In case of a BCC, this is within 6 weeks, in case of a more urgent lesion, this is within 2 weeks. Therefore, upper-bounds are taken as 6 weeks and 2 weeks respectively. For the lower-bounds, an estimation is made of 4 weeks and 1 week respectively. Again, these values are averaged over the frequencies of BCC (60%)

and more urgent lesions (40%) that have been estimated from the database on oncological consults that was provided. The third event 'Treatment date' uses the same waiting times as the event 'Rescheduled date' as both events schedule the treatment of the same types of lesions. Finally, the waiting event 'Biopsy results received' is modeled. These results are received within two weeks setting this as the upper-bound. The lower-bound is then set at 1 week resulting in an average of 1.5 week.

Table 6.3: Estimates of the lower-bound, upper-bound, and average of the waiting times *(hh:mm)*

Event	Lower-bound	Upper-bound	Average
Lesion data received	24:00	144:00	84:00
Data check date	01:00	24:00	12:30
1 week	168:00	168:00	168:00
Appointment date	318:30	285:40	287:45
Rescheduled date	470:00	739:20	604:40
Treatment date	470:00	739:20	604:40
Biopsy results received	168:00	336:00	252:00

Resources

The last simulation parameters are the resources. These are hours of availability of the administrative employee, AI specialist, and dermatologist per week. Since the process that is modeled is only a part of the variety of activities that the employees perform, the resources are only used to compare the redesigned scenario to the current scenario. Therefore, it has been decided to assign those resources such that the maximum capacity is not reached but the number of resources is still reasonable. The AI specialist is only performing a single task. As was mentioned previously this will be a daily activity of expectedly 2 hours each day between 12:00 and 14:00. As an administrative employee is required during the entire day, a single resource is assigned between 09:00 and 17:00. Finally, since the dermatologists have the most demanding tasks time wise, it is decided to assign two resources between 09:00 and 17:00 resulting in 120 hours per week.

6.3.3 | Simulation Parameters Current Process

The estimates of the current process are similar to the redesigned process in many ways. The major difference is that in the current process the loop of the AI specialist is disgarded. This is done through modelling the process as if 100% of the patients did not send their lesion data. As a result, the waiting time of 'Appointment date' is set to 1 week to adjust for the total waiting time of 2 weeks. Also, the task 'Request data on lesions' is not necessary and thus modeled as being instantaneous. Since this loop is not performed, the frequencies change for the type of lesions that are found after establishing the probability diagnosis. In the current scenario these follow the frequencies that are found in the provided dataset. As mentioned earlier, this dataset showed that in around 40% of the cases a excision is necessary, 50% requires some other form of treatment, and for 10% no treatment is performed. A reason that for only 10% no treatment is performed is the fact that currently, small lesions that could have been treated at the next check-up are already removed to ease the patient. Another alteration is the frequency of the appointments that can be treated directly. In the current scenario, patients first attend the consults of 10 minutes and so the excisions will need to be scheduled before they are able to perform it. Only in very few cases where there was already the expectation of an excision this is not the case. Therefore the split 'Directly treatable' has been altered to 95% no and 5% yes. The task durations and resources are assumed to be equal in the current process and the redesigned process and thus remain unaltered. The list with relevant estimates of the current process can be found in table 6.4.

6.4 | Results

The results were obtained by initiating the simulation parameters in Signavio and performing the runs. As explained before, the duration of the runs was set to 365 days in order to provide the most reliable estimate of the true mean. After running the simulation, the results provide details on the cycle time and resource consumption. The cycle times are provided for the total process and per individual task. Additionally, a distinction is made between the cycle time per task including possible waiting times and the availability of resources. In this way a clear overview is given of how long the actual task has taken and how long patients had to queue before the task could be executed. First off, the results are presented concerning the resource consumption and afterwards the cycle time is discussed.

Table 6.4: Estimates of the current process

Event	Lower-bound	Upper-bound	Average
1 week	168:00	168:00	168:00
Appointment date	168:00	168:00	168:00
Treatment date	470:00	739:20	604:40
Biopsy results received	168:00	336:00	252:00
No. of incoming patients per day	30	30	30
Lesion data received? (Yes)	0%	0%	0%
Type of lesion? (No risk)	10%	10%	10%
Type of lesion? (High risk)	40%	40%	40%
Type of lesion? (Low risk)	50%	50%	50%
Type of treatment? (Excision)	44%	44%	44%
Directly treatable? (Yes)	5%	5%	5%
Five years without NMSC?	-	-	-
Schedule treatment	00:05 (σ=00:02)	00:05 (σ=00:02)	00:05 (σ=00:02)
Schedule new check-up	00:07:30 (σ=00:02)	00:07:30 (σ=00:02)	00:07:30 (σ=00:02)
Request lesion data	00:00	00:00	00:00
Inquire about lesions	00:01	00:01	00:01
Physical examination	00:04	00:04	00:04
Establish probability diagnosis	00:05	00:05	00:05
Perform biopsy	00:00	00:00	00:00
Diagnose lesion	00:07 (σ = 00:02)	00:07 (σ = 00:02)	00:07 (σ = 00:02)
Define treatment	00:00	00:00	00:00
Perform treatment	00:00	00:00	00:00
Perform excision	00:20	00:30	00:25

6.4.1 | Resource Consumption

The resource consumption is measured per resource and shows two different indicators: the consumed time, and the workload in percentages. These numbers represent the totals for all allocated resources. The consumed time indicates how many hours the resource has spent performing tasks. In case there are more then one resource of a specific type, the total time is displayed. If there are no tasks to be performed, the resources are idle. The ratio of the working time with respect to the total allocated time is represented by the workload. The results of both indicators have been summarized in table 6.5 per resource for the six different simulations.

Insight 1: Workload Distribution
From the results, two conclusions can be drawn regarding the workload differences between the current process and the redesigned process. First off, the redesigned process shows a drastic increase in the workload of the administrative employees. The differences in the lower- and upper-bound of the individual processes display large differences. This increase in workload can be linked to the increased number of tasks that the administrative employee has to perform in the redesigned process. Specifically, 5 tasks are added to the 2 tasks that are currently being performed. This increase in number of tasks, and thus in workload, was expected to happen as during the interviews that were held with administrative employees, they indicated that they expect to require extra workforce to fulfill all of the extra activities they will have to perform.

Secondly, when looking at the workload of the dermatologists, a drastic decrease can be seen. Since the number and types of tasks the dermatologists perform are equal in both processes and no alterations were made in the the durations of these tasks, these results must follow from a reduction in the number of patients that pass through these tasks. Where this exact drop is coming from will be presented with the cycle time results.

Insight 2: Consumed Time
One key point can be seen when looking at the results of the consumed time. What is shown by the results is that the total consumed time of all resources does not decrease for the redesigned process. Instead, the lower-bound and average of the both processes display a slight increase in total consumed time for the redesigned process (±2813 vs. ±3026 and ±3018 vs. ±3041 respectively). This shows that whilst the burden on the dermatologists drops, this must be compensated elsewhere. This follows the expectations

Table 6.5: Summary of the simulation results regarding resource consumption

Scenario	Consumed time	Workload
Current process lower-bound		
Administrative Employee	240h:30m 31s	11,52%
Dermatologist	2573h:00m 43s	41,08%
Current process upper-bound		
Administrative Employee	239h:07m 19s	11,45%
Dermatologist	2985h:06m 18s	47,66%
Current process average		
Administrative employee	241h:06m 05s	11,55%
Dermatologist	2776h:45m 38s	44,33%
Redesigned process lower-bound		
Administrative employee	1294h:09m 21s	61,98%
Dermatologist	1362h:52m 57s	21,76%
AI Specialist	369h:12m 00s	70,73%
Redesigned process upper-bound		
Administrative employee	860h:11m 08s	41,20%
Dermatologist	1795h:21m 51s	28,66%
AI Specialist	318h:16m 00s	60,97%
Redesigned process average		
Administrative employee	1139h:50m 17s	54,59%
Dermatologist	1607h:26m 15s	25,66%
AI Specialist	394h:12m 00s	75,52%

of the problem owner that were outed during the interviews. It was expected that a larger administrative workforce would be needed (in line with what was said during the interview with the administrative employee) and that there should be an extra employee in charge of the data analysis. Even though there is an increase in the consumed time, the total costs are expected to decrease as the costs of dermatologists are the highest compared to the other workforces.

6.4.2 | Cycle Time

The results of the cycle times show the maximum, minimum, average, and total cycle time of the processes. In addition, the execution times per task are given. These execution times show the number of cases that was performed by a task and again the maximum, minimum, average, and total execution time. By analysing these results per tasks, the differences in cycle time between the two processes can be traced back to specific tasks. As explained before, a distinction can be made between the actual execution time, and the execution time including waiting times. Since the full list of results is too extensive and in many cases irrelevant for the discussion, a summary of the average and total cycle time of the different scenarios is given in table 6.6. Other relevant indicators will be presented throughout the discussion of the results.

Insight 3: Average and Total Cycle Time Reduction
The first conclusion that can be drawn from the summary of the simulation results is that the average and total cycle time are lower in the redesigned process than in the current process. This means that the throughput time of patients through the process is lower. This decrease in average and total cycle time can be related to the fact that less patients are being scheduled for a treatment in the redesigned process. Due to this, the total waiting time before these treatments is reduced leading to a lower throughput time. This simultaneously shows that the goal of reducing the number of patient visits is obtained in the redesigned process, which is confirmed when looking at the absolute number of cases. The number of patients that receive a 'fast' treatment during the consult drops drastically from around 3500 in the current process to around 1500 in the redesigned process on average. Similarly, the number of patients that need to have a treatment scheduled drops drastically. It must be stated that these numbers are the direct result of the definition of the simulation parameters and were thus to be expected.

Simulation Limitations
However, the results also display a flaw. It is found that in the redesigned process the number of excisions is much lower than in the current process (1636 versus 2816 on average). As patients that require an excision should always be treated, regardless of the scenario, this indicated that an error occurs. This drop

Table 6.6: Summary of the simulation results regarding cycle time

Scenario	Average cycle time	Total cycle time
Current process lower-bound	588h:14m 43s	4289485h:35m 00s
Current process upper-bound	743h:08m 22s	5299328h:56m 46s
Current process average	691h:27m 50s	4961254h:12m 38s
Redesigned process lower-bound	368h:41m 01s	2762546h:19m 46s
Redesigned process upper-bound	590h:33m 25s	4313428h:49m 46s
Redesigned process average	526h:18m 04s	3881998h:11m 42s

in excisions can be related to two factors. First off, it is possible that the simulation parameters used in the redesigned process are not a true representation of the real-world scenario. Altering these parameters till the same number of excisions is reached might show a more representative scenario. Another possibility is that a part of the patients that is now considered as 'low risk' (and receives no treatment) will be considered as 'high risk' (and must undergo excision) at their next consult. Since the process is simulated for only one loop to maintain simplicity and stay within the abilities of Signavio, this group is currently not considered. By performing a more detailed simulation that includes the fact that patients who were in the 'low risk' group the first year have a higher chance of being in the 'high risk' group next year, the number of excisions may stabilize over time. This reduction in the number of excisions that is performed also explains the drop in the workload of dermatologists that was described by the first insight. As this task has the longest duration of the tasks performed by the dermatologist, a decrease in the number of cases will have a significant impact on the workload of this workforce.

Secondly, it is important to note that the average and total cycle time of the lower-bound of the current process are lower than those of the upper-bound of the redesigned process. Since the simulation parameters were initialized in such a way that the actual scenario will be between the upper-bound and lower-bound, it is possible that the true values of both processes are not significantly different. However, since the interval between the upper-bound and lower-bound within a process are very large, and the average of both processes show large differences, it is probably safe to say that there will be a significant difference in the true values.

7 | Conclusion and Discussion

This project set out to research the applicability and value of a decision model for the decision making activities that take place for the improvement of the efficiency of non-melanoma skin cancer healthcare. In specific, the focus was put on the implementation of modern artificial intelligence technologies to achieve a decrease in the number of patient visits and decrease the burden on dermatologists while maintaining patient satisfaction and a feasible process in terms of costs. Furthermore, the project examined the potential of combining business process modeling with a decision model in order to assist in the decision making process. To do so, the research was performed following the design science methodology. This section presents the conclusions of both the individual elements of the project as well as their interaction. A discussion is presented on the limitations of these conclusions and finally recommendations for future projects are given.

7.1 | Conclusions

State-of-the-Art / State-of-Practice AI Solutions

One of the research questions that was answered during the project was what state-of-the-art and state-of-practice AI solutions are available to increase the healthcare efficiency and how they interact with the problem at hand. This research question was answered through performing a systematic literature review. From this review became evident that there is great potential in the existing AI algorithms built for lesion classification. Research has shown that the performance of these algorithms is able to perform on par with, or even surpass dermatologists. The usage of both lesion images and patient clinical data has proven to be beneficial for the performance. In practice, these algorithms still need to prove their potential in a real-world scenario as multiple issues remain that can hamper the road to implementation. These issues include the threat of a discriminating algorithm, lack of validation, patient privacy, and a lack of regulation. In the context of NMSC healthcare, these issues translate themselves into real-world problems. First off, since NMSC occurs more in fair-skinned people, databases predominantly represent these people. Therefore, an AI algorithm might not be able to classify dark skinned people correctly leading to a discriminating algorithm. Furthermore, healthcare is heavily subjected to privacy concerns. This may result in resistance for sharing medical information for the purpose of improving AI applications. Finally, due to a lack of regulation there is no standard on the minimal performance of AI or who is liable in case of malpractice.

What became evident from the literature review is that the vast majority of research is focused on the performance of an AI classifier. The implications and issues are widely mentioned throughout the papers, however, little research is being done that tries to solve these implications. To achieve an effective implementation of AI, it is essential that these issues are addressed, therefore, a shift in the focus of the research being performed should occur. By performing real-world trials, the literature gap might be filled while simultaneously providing regulators with gaining insights in the implications of these AI applications.

Stakeholders

Secondly, the project researched how the perspectives of different stakeholders may influence decision making activities. From interviews and document analysis became evident that requirements of different stakeholders may lead to different outcomes when these are included in decision making. Most stakeholder groups indicated that the requirements of the patients should eventually be most important as their well-being is at stake. However, while there is many research being done into the technical aspects of AI applications for skin cancer healthcare, only a few papers delve into the patients' perspective. Alignment of the goals of both patients and care givers is necessary to provide a practical solution that will lead to acceptance of all users. By approaching the problem from the different perspectives might help the decision maker to not only focus on the quantitative results, but also consider easily overlooked aspects like social well-being and adaptability of elderly. By making use of a goal, question, metric model, a clear overview of all aspects can be given which helps in the development of the decision space.

Since the decision makers are representatives of one of the stakeholder groups, the interests and requirements of other stakeholder groups can easily be overlooked. From this project can be learned that the performance of a stakeholder analysis may provide decision makers with different perspectives that can influence their decision making. Within research it is therefore necessary that a researcher is able to take all perspectives when deciding on their approach.

Decision Tool

The main research question that was answered by this thesis is how the modeling of a decision space influences the decision making activities in the implementation of an AI enabled process. In order to do so, a question, option, and criteria model was developed based on the results from the earlier answered research questions and semi-structured interviews with representatives from the stakeholder groups. When validating the QOC model through a focus-group interview, it became evident that the QOC model assisted with the decision making in multiple ways. First off, the model provided context for the decisions that were to be made. The participants of the focus group showed understanding of the problem and could place it in their daily work activities. Additionally, the model clearly contributed to the decision making by providing the decision makers with the variety of decisions that have to be made and the different perspectives at play. Decision makers were able to provide a line of argument for their approach that resulted form the decision model. From this we can conclude that a decision model is indeed an adequate tool to be used in the decision making activities of NMSC healthcare.

Whether this decision tool is also applicable in other healthcare settings can be disputed. The reason for its success in this setting can be related to the fact that a large part of the process is executable by patients. Due to their contribution in the process, it is essential that their wishes are incorporated in the decisions. The success of the redesigned process is highly dependent of the cooperation of the patients. Therefore, the usage of a simple, complete overview of the options and criteria of all stakeholders can prove effective. Another healthcare setting where this type of decision making is ought to be effective is in a general practice as the patients play a similar role in these processes and can autonomously act in requesting care. In case a similar approach is applied in a healthcare setting where patients make no contribution to the process, their requirements are expected to be less relevant and might even collide with effective decision making. E.g. it is to be expected that for improving a surgical procedure, a surgeon can best decide based on factual data what is the best procedure and not on a patient's wishes.

Business Process Simulation

Finally, as the ultimate goal was to enhance a business process using a decision space, the contribution of business process modeling to the decision model was researched. First off, it was shown that modeling of the decision process can assist in developing the options for the decision model as it presents an overview of the activities that take place within a process. More importantly, business process modeling can assist with the decision making activities by showing what effects the decisions will have through redesigning the business process and performing simulations. The usage of simple simulation software can provide quick and easy insights into the effects of certain decisions. To our knowledge, no research has been done in the additional value of business process modeling for decision making. While the simulation and business process modeling were a relative small part of this project, the results did seem promising. Therefore, new research might expand on this by looking into the additional value in other (healthcare) settings.

7.2 | Managerial Implications

The results of this thesis can be translated into managerial implications. Managers that are tasked with the redesign of one of their business processes in order to increase the overall efficiency, can use these practical implications to successfully achieve their goals. First off, this project has shown that it is essential for managers to maintain an unbiased perspective and consider the wishes of all different stakeholders. To manage this, it is advised to perform in-depth interviews with representatives of all stakeholder groups in order to unveil these perspectives and requirements. With this it must be noted that this project has shown that there can exist large differences in perspective within a stakeholder group. Therefore, using a random selection of a number of representatives is necessary to acquire these perspectives. By appointing an interviewer that is not part of one of the stakeholder groups, the possibility of acquiring biased results drops.

Furthermore, during the development of the decision model, managers should understand that the creation of such a model is an iterative process. It was shown during this project that throughout the development and validation phases, new insights and perspectives might be brought to light that could result in the redesign of the decision model. By performing multiple iterations of going back and forth in the design process, the model will eventually result in a better outcome. Even when the decision model has been implemented and is being used in the decision process, further development of the model to include new questions that might come to light during the implementation will increase its value.

To effectively apply BP simulation in the decision making activities, managers should maintain data on their current process. The more data there is available that can be used to substantiate process redesign decisions and initiate simulation parameters, the more valuable the simulations will be. This project showed that by using a small sample of cases some insights can be given but slight flaws occurred. When there is an extensive overview of the current processes, throughput time, and workload, these insights might be more beneficial in deciding on the best practice. Building a database containing these indicators may be time-consuming, however, the benefits of possessing such a database are expected to pay off.

Finally, managers should take the context of their decision making activities into consideration when deciding on whether to use the approach present throughout this project. As stated earlier, this approach is expected to be beneficial in scenarios where many stakeholders are at play and where the contributions of these stakeholders is essential for the success of the process. Therefore it is advised to start by a thorough analysis of the current process and stakeholders to see whether these criteria are met before starting with the redesign cycle.

7.3 | Limitations

During this thesis some limitations occurred as a result of lack in time and restriction of resources. First off, the design science methodology suggests iterating through the design cycle to keep improving on the artefact. Due to limited time, this thesis only performed three of the four stages of the design cycle. Due to this, points of improvement that were derived from the validation of the artefact have not been implemented. Additionally, the implementation stage was performed limiting the available results. By performing another cycle, a more final version of the artefact could have been developed which would have given more insight into the effects on the decision making. Another limitation that was the result of a lack in time was the fact that the validation of the artefact was performed with a focus group of only four people. The possibilities to schedule a more extensive focus group within a short time period were limited. A broader range of perspectives could have given a better understanding of how the decision model was received.

Another limitation of the thesis is the fact that all qualitative data that was retrieved and analyzed was related to a single case study. Due to this, the external validity decreases. As the decision model was designed specifically for the decision making activities at MohsA, the model cannot be applied at different clinics. Therefore, in case the structure and types of decisions at other clinics are very different, the results may not be generalized. Still, it is expected that the usage of a decision model can assist in many similar occasions where there are multiple perspectives to be considered carefully. Besides the limitations with respect to the external validity, there is the matter of internal validity. Since the research was performed by only one researcher, subjective elements like the qualitative data analysis may be subject to prejudice. To increase the internal validity, the interviews, coding, and analysis should be performed by multiple researchers.

Finally, the data on which the simulation was based was limited in the sense that many estimations had to be made. This was mainly due to the fact that at MohsA there was no extensive documentation available on throughput times, waiting times, case types, and performed treatments. Due to this, the conclusions that can be drawn from the simulation are limited. More specific estimations may specify the results even further and decrease the interval gap between the lower- an upper-bound.

7.4 | Further Recommendations

Based on this thesis, multiple directions for future research can be proposed. First off, the research can be extended to other skin clinics or skin cancer healthcare settings. By applying the same concepts within another setting may display whether the implementation of a decision model is applicable in other contexts as well. In specific it is proposed to perform a multiple case study where the applicability of a decision model is tested for these different cases. For this, the decision model should be altered slightly to fit the processes at these clinics but maintain the same tree-like structure and format. Cross-validation of these cases can show whether the results of a decision model can be generalized for skin cancer healthcare. Additionally, it is proposed to include skin clinics and hospitals in the study that have a less developed healthcare process, to research whether the implementation of an AI-enabled process is overall the best practice for improving efficiency or whether in other settings the efficiency may benefit more from other process redesigns (e.g. one-stop-shop model). To identify the latter, the decision model should be extended to include these options as well.

Furthermore, the contribution of business process simulation towards the decision making activities can be explored in more detail. For this, two different approaches are proposed. First off, it is proposed to perform multiple basic simulations, like the one performed in this report, for all scenarios that are considered in the decision process. As during the validation of the decision model in this project was shown that one of the main difficulties was to decide on an option without knowing the potential drop in costs or gain in efficiency, it is expected that including these simulation results may further assist decision makers in their decision activities. The second approach is to perform a more advanced simulation of the final outcome of the decision model and compare this to a similar simulation of the current process. As the simulation in this project provided general insights but lacked details, it was difficult to identify the limiting processes and possible jams of the system. A more detailed simulation could help with identifying the feasibility of the process and uncover potential drawbacks.

Lastly, it is proposed to validate the real-world performance of AI trained against both patient clinical data and images. As was explained before, the current research is too focused on the performance of AI in a controlled environment. By performing real-world tests with the randomness and (lack of) quality of actual patient data, it can be tested whether these AI applications are ready for real-world implementations. In specific, it is proposed to implement a model that has already proven its performance in a clinical setting, in multiple healthcare clinics in different regions so that there is a wide variety of cases. These tests may determine whether the limitations that currently withhold the implementation of an AI enabled process, like algorithm bias, can be overcome and the road to a clinician-AI synergy can continue.

8 | References

[1] Adekanmi A. Adegun and Serestina Viriri. Deep learning-based system for automatic melanoma detection. *IEEE Access*, 8:7160–7172, 2020. doi: 10.1109/ACCESS.2019.2962812.

[2] Zoe Apalla, Dorothée Nashan, Richard B Weller, and Xavier Castellsagué. Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches. *Dermatology and therapy*, 7 (1):5–19, 2017.

[3] David Ben-Israel, W. Bradley Jacobs, Steve Casha, Stefan Lang, Won Hyung A. Ryu, Madeleine de Lotbiniere-Bassett, and David W. Cadotte. The impact of machine learning on patient care: A systematic review. *Artificial Intelligence in Medicine*, 103:101785, 2020. ISSN 0933-3657. doi: https://doi.org/10.1016/j.artmed.2019.101785.

[4] Bo Bergman, Duncan Neuhauser, and Lloyd Provost. Five main processes in healthcare: a citizen perspective. *BMJ quality & safety*, 20(Suppl 1):i41–i42, 2011.

[5] Titus J Brinker, Achim Hekler, Alexander H Enk, Joachim Klode, Axel Hauschild, Carola Berking, Bastian Schilling, Sebastian Haferkamp, Dirk Schadendorf, Tim Holland-Letz, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113:47–54, 2019.

[6] Federico Cabitza, Andrea Campagner, and Clara Balsano. Bridging the "last mile" gap between ai implementation and operation: "data awareness" that matters. *Annals of Translational Medicine*, 8 (7), 2020. ISSN 2305-5847.

[7] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological methods & research*, 42(3):294–320, 2013.

[8] Insook Cho, Minyoung Lee, and Yeonjin Kim. What are the main patient safety concerns of healthcare stakeholders: a mixed-method study of web-based text. *International journal of medical informatics*, 140:104162, 2020.

[9] Magdalena Ciażyńska, Grażyna Kamińska-Winciorek, Dariusz Lange, Bogumił Lewandowski, Adam Reich, Martyna Sławińska, Marta Pabianek, Katarzyna Szczepaniak, Adam Hankiewicz, Małgorzata Ułańska, et al. The incidence and clinical analysis of non-melanoma skin cancer. *Scientific reports*, 11(1):1–10, 2021.

[10] Noel CF Codella, Q-B Nguyen, Sharath Pankanti, David A Gutman, Brian Helba, Allan C Halpern, and John R Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4/5):5–1, 2017.

[11] European Commission, Social Affairs Directorate-General for Employment, and Inclusion. *Long-term care report : trends, challenges and opportunities in an ageing society. Volume I.* Publications Office, 2021. doi: doi/10.2767/677726.

[12] Erika Cottrell, Evelyn Whitlock, Elisabeth Kato, Stacey Uhl, Suzanne Belinson, Christine Chang, Ties Hoomans, David Meltzer, Hussein Noorani, Karen Robinson, et al. *Defining the benefits of stakeholder engagement in systematic reviews.* AHRQ Methods for Effective Health Care. Agency for Healthcare Research and Quality (US), Rockville (MD), 2014.

[13] Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2):94, 2019.

[14] Nicole M Deterding and Mary C Waters. Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research*, 50(2):708–739, 2021.

[15] Barbara DiCicco-Bloom and Benjamin F Crabtree. The qualitative research interview. *Medical education*, 40(4):314–321, 2006.

[16] Brian Diffey. Climate change, ozone depletion and the impact on ultraviolet exposure of human skin. *Physics in medicine & biology*, 49(1):R1, 2003.

[17] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Abdul Hakeem M Saeed, Mohammed Olaythah

Alraddadi, and Mater Hussen Mahnashi. Skin cancer detection: a review using deep learning techniques. *International journal of environmental research and public health*, 18(10):5479, 2021.

[18] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A Reijers. *Fundamentals of business process management*, volume 2. Springer, 2018. ISBN 978-3-662-56509-4.

[19] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[20] Claire M Felmingham, Nikki R Adler, Zongyuan Ge, Rachael L Morton, Monika Janda, and Victoria J Mar. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. *American Journal of Clinical Dermatology*, 22 (2):233–242, 2021.

[21] International Agency for Research on Cancer. *IARC Monographs of the Evaluation of Carcinogenic Risks to Humans*, volume 55. World Health Organisation, 2 1992. ISBN 92 832 1255 X.

[22] International Agency for Research on Cancer. Estimated number of new cases in 2020, world, both sexes, all ages. https://gco.iarc.fr/today/online-analysis-table, 2020.

[23] The Skin Cancer Foundation. Skin cancer information. https://www.skincancer.org/skin-cancer-information/, 2022.

[24] Y Fujisawa, Y Otomo, Y Ogata, Y Nakamura, R Fujita, Y Ishitsuka, R Watanabe, N Okiyama, K Ohara, and M Fujimoto. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *British Journal of Dermatology*, 180(2):373–381, 2019.

[25] Alena Furdova, Karolina Kapitanova, Alexandra Kollarova, and Juraj Sekac. Periocular basal cell carcinoma-clinical perspectives. *Oncology Reviews*, 14(1), 2020.

[26] Simone van der Geer-Rutten. *Disease Management for Chronic Skin Cancer.* PhD thesis, Erasmus University Rotterdam, 2012.

[27] Simone van der Geer-Rutten, P Ad M Kleingeld, Chris CP Snijders, Frank JCH Rinkens, Geert AE Jansen, HA Martino Neumann, and Gertruud AM Krekels. Development of a non-melanoma skin cancer detection model. *Dermatology*, 230(2):161–169, 2015.

[28] Karen Glanz, Elinor Schoenfeld, Martin A Weinstock, Gabriela Layi, Jeanne Kidd, and Dorothy M Shigaki. Development and reliability of a brief skin cancer risk assessment tool. *Cancer detection and prevention*, 27(4):311–315, 2003.

[29] Manu Goyal, Thomas Knackstedt, Shaofeng Yan, and Saeed Hassanpour. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine*, 127:104065, 2020.

[30] David E Gray. *Doing research in the real world*, volume 5. sage, 2021.

[31] Adele Green, Graeme Beardmore, Veronica Hart, David Leslie, Robin Marks, and Donald Staines. Skin cancer in a queensland population. *Journal of the American Academy of Dermatology*, 19(6): 1045–1052, 1988.

[32] Gery P Guy Jr, Steven R Machlin, Donatus U Ekwueme, and K Robin Yabroff. Prevalence and costs of skin cancer treatment in the us, 2002- 2006 and 2007- 2011. *American journal of preventive medicine*, 48(2):183–187, 2015.

[33] Holger A Haenssle, Christine Fink, Roland Schneiderbauer, Ferdinand Toberer, Timo Buhl, Andreas Blum, A Kalloo, A Ben Hadj Hassen, Luc Thomas, A Enk, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology*, 29(8):1836–1842, 2018.

[34] Pavel Hamet and Johanne Tremblay. Artificial intelligence in medicine. *Metabolism*, 69:S36–S40, 2017.

[35] Reyhaneh Hamidi, David Peng, and Myles Cockburn. Efficacy of skin self-examination for the early detection of melanoma. *International journal of dermatology*, 49(2):126–134, 2010.

[36] Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1):30–36, 2019.

[37] Carolyn J Heckman, Diane B Wilson, and Karen S Ingersoll. The influence of appearance, health, and future orientations on tanning behavior. *American Journal of Health Behavior*, 33(3):238, 2009.

[38] Achim Hekler, Jochen Sven Utikal, Alexander H Enk, Carola Berking, Joachim Klode, Dirk Schadendorf, Philipp Jansen, Cindy Franklin, Tim Holland-Letz, Dieter Krahl, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *European Journal of Cancer*, 115:79–83, 2019.

[39] Frank Hoepel. Integration of heterogeneous data for the classification of non-melanoma skin cancer. Master's thesis, Eindhoven University of Technology, 2020.

[40] Julia Höhn, Eva Krieghoff-Henning, Tanja B Jutzi, Christof von Kalle, Jochen S Utikal, Friedegund Meier, Frank F Gellrich, Sarah Hobelsberger, Axel Hauschild, Justin G Schlager, et al. Combining cnn-based histologic whole slide image analysis and patient data to improve skin cancer classification. *European Journal of Cancer*, 149:94–101, 2021.

[41] International Skin Imaging Collaboration. Goals of the archive, 2022. URL https://www.isic-archive.com/#!/topWithHeader/tightContentTop/about/isicArchiveGoals.

[42] ISO/IEC 25000. System and software quality requirements and evaluation, 2022. URL https://iso25000.com/index.php/en/.

[43] M Janda and HP Soyer. Can clinical decision making be enhanced by artificial intelligence? *The British journal of dermatology*, 180(2):247–248, 2019.

[44] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017.

[45] YQ Jiang, JH Xiong, HY Li, XH Yang, WT Yu, M Gao, X Zhao, YP Ma, W Zhang, YF Guan, et al. Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network. *British Journal of Dermatology*, 182(3):754–762, 2020.

[46] Tanja B Jutzi, Eva I Krieghoff-Henning, Tim Holland-Letz, Jochen Sven Utikal, Axel Hauschild, Dirk Schadendorf, Wiebke Sondermann, Stefan Fröhling, Achim Hekler, Max Schmitt, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Frontiers in medicine*, 7:233, 2020.

[47] Rahul Kapoor, Stephen P Walters, and Lama A Al-Aswad. The current state of artificial intelligence in ophthalmology. *Survey of ophthalmology*, 64(2):233–240, 2019.

[48] Bita Arbab Kash, Aaron Spaulding, Christopher E Johnson, and Larry Gamm. Success factors for strategic change initiatives: A qualitative study of healthcare administrators' perspectives. *Journal of Healthcare Management*, 59(1):65–81, 2014.

[49] Breno Krohling, Pedro BC Castro, Andre GC Pacheco, and Renato A Krohling. A smartphone based application for skin cancer classification using deep learning with clinical images and lesion information. *arXiv preprint arXiv:2104.14353*, 2021.

[50] Allan Maclean, Richard M. Young, Victoria M. E. Bellotti, and Thomas P. Moran. Questions, options and criteria: Clements of a design space analysis. *Human-Computer Interaction*, 6:201–250, 1991.

[51] Tanja Maier, D Kulichova, K Schotten, R Astrid, T Ruzicka, C Berking, and A Udrea. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *Journal of the European Academy of Dermatology and Venereology*, 29(4):663–667, 2015.

[52] Roman C Maron, Michael Weichenthal, Jochen S Utikal, Achim Hekler, Carola Berking, Axel Hauschild, Alexander H Enk, Sebastian Haferkamp, Joachim Klode, Dirk Schadendorf, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *European Journal of Cancer*, 119:57–65, 2019.

[53] Camilla Mattiuzzi and Giuseppe Lippi. Current cancer epidemiology. *Journal of epidemiology and global health*, 9(4):217, 2019.

[54] Alexander J McLeod Jr and Jan Guynes Clark. Using stakeholder analysis to identify users in healthcare information systems research: who is the real user? *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, 4(3):1–15, 2009.

[55] MohsA. Over mohsa. https://www.mohsa.nl/over-mohsa/, 10 2022.

[56] MohsA. One-stop-shop voor huidkanker. https://www.mohsa.nl/huidkanker/behandelingen/one-stop-shop-voor-huidkanker/, 10 2022.

[57] Christian Murray, D Sivajohanathan, TP Hanna, S Bradshaw, N Solish, B Moran, R Hekkenberg, AC Wei, and T Petrella. Patient indications for mohs micrographic surgery: A clinical practice guideline. *Current Oncology*, 26(1):94–99, 2019.

[58] Hardik Nahata and Satya P. Singh. *Deep Learning Solutions for Skin Cancer Detection and Diagnosis*. Springer International Publishing, 2020. ISBN 978-3-030-40850-3.

[59] Francis Sahngun Nahm. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75(1):25–36, 2022.

[60] Camille Nebeker, John Torous, and Rebecca J Bartlett Ellis. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC medicine*, 17(1):1–7, 2019.

[61] Caroline A Nelson, Lourdes Maria Pérez-Chada, Andrew Creadore, Sara Jiayang Li, Kelly Lo, Priya Manjaly, Ashley Bahareh Pournamdari, Elizabeth Tkachenko, John S Barbieri, Justin M Ko, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA dermatology*, 156(5):501–512, 2020.

[62] Shamim Nemati, Andre Holder, Fereshteh Razmi, Matthew D Stanley, Gari D Clifford, and Timothy G Buchman. An interpretable machine learning model for accurate prediction of sepsis in the icu. *Critical care medicine*, 46(4):547, 2018.

[63] Eline Noels, Loes Hollestein, Kees Luijkx, Marieke Louwman, Carin de Uyl-de Groot, Renate van den Bos, Astrid van der Veldt, Dirk Grünhagen, and Marlies Wakkee. Increasing costs of skin cancer due to increasing incidence and introduction of pharmaceuticals, 2007–2017. *Acta Dermato-Venereologica*, 100(10):1–6, 2020.

[64] Susan A Oliveria, Kishwer S Nehal, Paul J Christos, Neeta Sharma, Jennifer S Tromberg, and Allan C Halpern. Using nurse practitioners for skin cancer screening: a pilot study. *American journal of preventive medicine*, 21(3):214–217, 2001.

[65] OMG. Business Process Model and Notation (BPMN), Version 2.0.2, December 2013. URL http://www.omg.org/spec/BPMN/2.0.2.

[66] World Health Organisation. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, 10 2022.

[67] Andre GC Pacheco and Renato A Krohling. The impact of patient clinical information on automated skin cancer detection. *Computers in biology and medicine*, 116:103545, 2020.

[68] Andre GC Pacheco and Renato A Krohling. An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, 25(9):3554–3563, 2021.

[69] Andre G.C. Pacheco, Gustavo R. Lima, Amanda S. Salomão, Breno Krohling, Igor P. Biral, Gabriel G. de Angelo, Fábio C.R. Alves Jr, José G.M. Esgario, Alana C. Simora, Pedro B.C. Castro, Felipe B. Rodrigues, Patricia H.L. Frasson, Renato A. Krohling, Helder Knidel, Maria C.S. Santos, Rachel B. do Espírito Santo, Telma L.S.G. Macedo, Tania R.P. Canuto, and Luíz F.S. de Barros. Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, 32:106221, 2020. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2020.106221.

[70] Rosario Emanuele Perrotta, Maria Giordano, and Mariano Malaguarnera. Non-melanoma skin cancers in elderly patients. *Critical reviews in oncology/hematology*, 80(3):474–480, 2011.

[71] Luise Pufahl, Francesca Zerbato, Barbara Weber, and Ingo Weber. Bpmn in healthcare: Challenges and best practices. *Information Systems*, 107:102013, 2022.

[72] Gaëlle Quéreux, Jean-Michel Nguyen, Christelle Volteau, Yves Lequeux, and Brigitte Dréno. Creation and test of a questionnaire for self-assessment of melanoma risk factors. *European Journal of Cancer Prevention*, 19(1):48–54, 2010. ISSN 09598278, 14735709.

[73] Fatemeh Rabiee. Focus-group interview and data analysis. *Proceedings of the nutrition society*, 63(4): 655–660, 2004.

[74] Sandeep Reddy, John Fox, and Maulik P Purohit. Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, 112(1):22–28, 2019.

[75] David Roffman, Gregory Hart, Michael Girardi, Christine J Ko, and Jun Deng. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Scientific reports*, 8(1):1–7, 2018.

[76] Mark Saunders, Philip Lewis, and Adrian Thornhill. *Research methods for business students*, volume 6. Pearson, 2012. ISBN 9781292208787.

[77] William B Schwartz, Ramesh S Patil, and Peter Szolovits. Artificial intelligence in medicine. *New England Journal of Medicine*, 316(11):685–688, 1987.

[78] Tamar Sharon. *Data-driven decision making, AI and the Googlization of health research.* Pavia: Pavia University Press, 2020.

[79] Doaa A Shoieb, Sherin M Youssef, and Walid M Aly. Computer-aided model for skin diagnosis using deep learning. *Journal of Image and Graphics*, 4(2):122–129, 2016.

[80] H. Slaper, A. van Dijk, P. den Outer, H. van Kranen, and L. Slobbe. Uv-straling en gezondheid. Technical report, Rijksinstituut voor volksgezondheid en milieu, 2017.

[81] IS Stafford, M Kellermann, E Mossotto, Robert Mark Beattie, Ben D MacArthur, and Sarah Ennis. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ digital medicine*, 3(1):1–11, 2020.

[82] Jerod L Stapleton, Rob Turrisi, Kimberly A Mallett, and June K Robinson. Correspondence between pigmented lesions identified by melanoma patients trained to perform partner-assisted skin self-examination and dermatological examinationaccuracy of partner-assisted skin self-examination. *Cancer Epidemiology, Biomarkers & Prevention*, 24(8):1247–1253, 2015.

[83] Simon M. Thomas, James G. Lefevre, Glenn Baxter, and Nicholas A. Hamilton. Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. *Medical Image Analysis*, 68:101915, 2021. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2020.101915.

[84] Kristina Todorova and Anna Mandinova. Novel approaches for managing aged skin and nonmelanoma skin cancer. *Advanced Drug Delivery Reviews*, 153:18–27, 2020.

[85] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, et al. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, 2020.

[86] S Van der Geer, M Frunt, HL Romero, NP Dellaert, MH Jansen-Vullers, TBJ Demeyere, HAM Neumann, and GAM Krekels. One-stop-shop treatment for basal cell carcinoma, part of a new disease management strategy. *Journal of the European Academy of Dermatology and Venereology*, 26(9): 1154–1157, 2012.

[87] Bas HM van der Velden, Hugo J Kuijf, Kenneth GA Gilhuijs, and Max A Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, page 102470, 2022.

[88] Floris Ruben Tobias van Elburg, Nicky Sabine Klaver, Anna Petra Nieboer, and Marjan Askari. Gender differences regarding intention to use mhealth applications in the dutch elderly population: a cross-sectional study. *BMC geriatrics*, 22(1):1–11, 2022.

[89] Olivier J Van Not, Jesper van Breeschoten, Alfonsus JM van den Eertwegh, Doranne L Hilarius, Melissa M De Meza, John B Haanen, Christian U Blank, Maureen JB Aarts, Franchette WPJ van den Berkmortel, Jan Willem B de Groot, et al. The unfavorable effects of covid-19 on dutch advanced melanoma care. *International journal of cancer*, 150(5):816–824, 2022.

[90] Mike CM van Zon, José D van der Waa, Mitko Veta, and Gertruud AM Krekels. Whole-slide margin control through deep learning in mohs micrographic surgery for basal cell carcinoma. *Experimental Dermatology*, 30(5):733–738, 2021.

[91] Zsuzsa Varvasovszky and Ruairí Brugha. A stakeholder analysis. *Health policy and planning*, 15(3): 338–345, 2000.

[92] David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. *The Lancet Digital Health*, 2021.

[93] Roel J Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014. ISBN 978-3-662-43838-1.

[94] Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[95] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[96] Yifan Xiang, Lanqin Zhao, Zhenzhen Liu, Xiaohang Wu, Jingjing Chen, Erping Long, Duoru Lin, Yi Zhu, Chuan Chen, Zhuoling Lin, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. *Artificial intelligence in medicine*, 102:101780, 2020.

[97] Peizhen Xie, Ke Zuo, Yu Zhang, Fangfang Li, Mingzhu Yin, and Kai Lu. Interpretable classification from skin cancer histology slides using deep learning: A retrospective multicenter study. *arXiv preprint arXiv:1904.06156*, 2019.

[98] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10):719–731, 2018.

[99] Yang Zhao, Qi Ni, and Ruoxin Zhou. What factors influence the mobile health service adoption? a meta-analysis and the moderating role of age. *International Journal of Information Management*, 43: 342–350, 2018.

Appendix

A | Model predictors by Geer-Rutten

	B	S.E.	Sig.	Exp(B)	95% C.I.forExp(B)		Marginal effects (dy/dx)
					lower	upper	
Shape: elevated	-2,642	,616	,000	,071	,021	,238	-.234
Sun exposure < 65 years of age	1,458	,521	,005	4,297	1,549	11,924	.129
Ulceration	-1,303	,659	,048	,272	,075	,988	-.116
Shininess	-1,935	,893	,030	,144	,025	,830	-.172
Keratotic	2,801	,761	,000	16,467	3,704	73,200	.248
Frequent holidays with sun exposure	-,479	,182	,008	,619	,434	,884	-.042
Light red	1,902	,647	,003	6,700	1,886	23,804	.161
Located at sun exposed area	1,825	,695	,009	6,202	1,589	24,204	.162
Widened bloodvessels	1,471	,956	,124	4,352	,668	28,346	.130
Itching: yes	1,404	,941	,136	,246	,039	1,553	.079
Keratotic X Light red	-1,610	,774	,037	,200	,044	,911	-.143
Constant	-6,709	2,700	,013	,001			

B= Regression coefficient. These are the values for the logistic regression equation for predicting the dependent variable (outcome) from the independent variable (predictor).

S.E= Standard Error.
Sig= Significance. Two-tailed p-value used in testing the null hypothesis that the coefficient B is zero.
Exp(B) = Odds ratio. This is an indicator of the change in odds resulting from a unit change in the predictor value. If this value is greater than 1, it indicates that an increase in the predictor variable will result in an increase of the probability of the outcome occurring. Equally, if the exp(b) is less than 1, an increase in the predictor variable will result in a decrease of the probability of the outcome occurring.
C.I. = Confidence Interval

Figure A.1: Predictors for AK in the model by Geer-Rutten [26]

Observed	Prediction model AK negative (N)	Prediction model AK positive (N)	Correct percentage (%)	Diagnosis nurses AK negative (N)	Diagnosis nurses AK positive (N)	Correct percentage (%)
AK negative (N)	126	13	90.6	123	16	88.5
AK positive (N)	17	40	70.2	7	50	87.7
Overall percentage (%)			84.7			88.3

N= number AK = actinic keratosis

Figure A.2: Predictions AK by model and by nurse (cut-off value=0.5)[26]

	B	S.E.	Sig.	Exp(B)	95% C.I.forExp(B)		Marginal effects (dy/dx)
					Lower	Upper	
Dark-red	4,212	1,831	,021	67,516	1,866	2442,345	.625
Light-red	3,601	1,077	,001	36,638	4,439	302,389	.122
Shininess	1,145	,677	,091	3,141	,833	11,851	.033
Age (years)	,098	,030	,001	1,103	1,040	1,170	.003
Induration	-1,439	,870	,098	,237	,043	1,304	-.041
Frequent holidays with sun exposure	,516	,225	,022	1,675	1,079	2,603	.015
Ulceration	2,639	,857	,002	13,997	2,611	75,028	.075
Keratotic	-1,916	,481	,000	,147	,057	,378	-.054
Bleeds easily	,876	,631	,165	2,402	,698	8,271	.025
Shape: elevated	2,124	,766	,006	8,362	1,865	37,498	.060
Constant	-16,332	3,483	,000	,000			

B= Regression coefficient.
S.E= Standard Error.
Exp(B) = Odds ratio.
C.I. = Confidence Interval

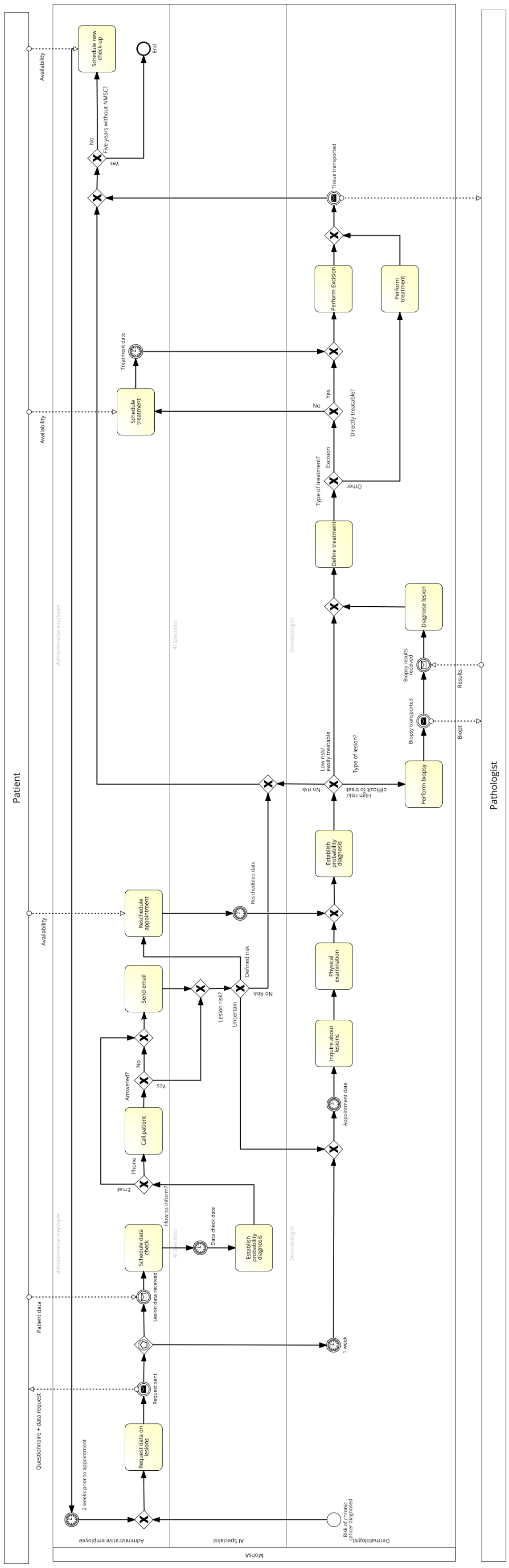
Figure A.3: Predictors for BCC in the model by Geer-Rutten [26]

Observed	Prediction model BCC negative (N)	Prediction model BCC positive (N)	Correct percentage (%)	Diagnosis nurses BCC negative (N)	Diagnosis nurses BCC positive (N)	Correct percentage (%)
BCC negative (N)	161	4	97.6	153	12	92.7
BCC positive (N)	13	20	60.6	7	26	78.8
Overall percentage (%)			91.4			90.4

N=number BCC= basal cell carcinoma

Figure A.4: Predictions BCC by model and by nurse (cut-off value=0.5)[26]

B Redesigned BPMN diagram



C | Interview Guide QOC

Patiënten:

Bedankt voor jullie deelname aan mijn onderzoek. Zoals jullie weten gaat mijn onderzoek over het implementeren van innovatieve technieken in de huidkanker zorg met als doel om de efficiëntie te verbeteren. In het specifiek proberen we een manier te vinden waarbij patiënten die al eerder huidkanker hebben gehad, alleen nog op controle hoeven te komen wanneer er ook daadwerkelijk sprake is van huidkanker. Het idee achter de implementatie is dat patiënten gevraagde zullen worden om voorgaand aan een controle foto's te maken van plekjes die ze verdacht vinden en daarbij een korte vragenlijst in te vullen. Om deze implementatie voor zowel de arts als patiënt een verbetering te laten zijn, ontvang ik graag jullie input over een aantal onderwerpen.

1. Komen jullie vast op controle bij MohsA? a. Hoe ervaren jullie/ervaar je het nu om op controle te komen, is het eenvoudig/lastig, snel/duurt lang, vaak/te weinig?

2. Zijn er dingen die volgens jullie beter/sneller zouden kunnen verlopen?

3. Hoe zouden jullie het vinden om vooraf aan een controle een foto te maken/laten maken van plekje waar je je zorgen om maakt en deze op te sturen?

4. Hoe zouden jullie het vinden om vooraf aan een controle een vragenlijst in te vullen over plekjes waar je je zorgen om maakt en deze op te sturen? Waarom wel/niet

5. Op welke manier zouden jullie het fijn vinden als je dit kan insturen (denk aan mail, app, website)

6. Wat zijn voor jullie eisen aan de manier waarop je je gegevens zou moeten insturen? (denk aan privacy, makkelijk te gebruiken, snel)

7. Wie zouden de gegevens in moeten kunnen zien? (Zuster, dermatoloog, administratief medewerker)

Aan de hand van de data die jullie insturen zal worden bepaald wat het risico is op huidkanker. Aan de hand daarvan kunnen verschillende acties worden ondernomen. Een optie is bijvoorbeeld om de controle uit te stellen wanneer er weinig risico is.

8. Hoe zouden jullie het vinden wanneer dit zou gebeuren? a. Wat zou je fijner vinden op deze manier b. Wat zou je vervelender vinden op deze manier

9. Zouden jullie liever willen dat je zelf kan kiezen om de controle uit te stellen bij weinig risico of dat MohsA dit inschat? Waarom?

10. Hoe zouden jullie hierover gecommuniceerd willen worden?

11. Welke informatie zou je willen ontvangen indien je afspraak wordt verplaatst of wanneer deze juist niet wordt verplaatst?

Een andere situatie zou kunnen zijn dat aan de hand van de data alvast een inschatting kan worden gemaakt van welke behandeling er plaats zou moeten vinden. Patiënten worden dan vooraf geïnformeerd over welke behandeling er waarschijnlijk plaats zal gaan vinden en of ze daarmee akkoord gaan.

12. Hoe zou je het vinden dat je meteen behandeld zou kunnen worden op de eerste afspraak zonder eerst de fysieke controle te hebben gehad? a. Wat zou je fijner vinden op deze manier b. Wat zou je vervelender vinden

13. Hoe zou je geïnformeerd willen worden over de behandeling? (Welk middel, door wie)

14. Welke informatie zou je willen ontvangen?

15. Zijn er nog dingen waarvan je wilt dat er rekening mee gehouden wordt wanneer deze situatie zou plaatsvinden?

De risico inschatting die wordt gedaan zou geautomatiseerd kunnen worden met behulp van een kunstmatige intelligentie. In dat geval worden de gegevens die jullie aanleveren door een computer geanalyseerd en kan die aangeven wat het risico is dat het wel of niet huidkanker is.

16. Hoe zou je het vinden wanneer een kunstmatige intelligentie de risico inschatting maakt? Welke voor/nadelen zie je?

17. Aan welke eisen zou een KI moeten voldoen om ervoor te zorgen dat je het wel ziet zitten? (Bijvoorbeeld dat de resultaten altijd nog gecontroleerd worden door een zuster)

18. Hoe zou je in dit geval geïnformeerd willen worden over de resultaten?

19. Stel jullie zouden zelf een verbetering mogen voorstellen voor het zorgproces, wat zou dat dan zijn?

20. Wie denken jullie dat het meest beïnvloed worden door deze nieuwe toepassingen? (Arts, patiënt)

a. Wie heeft er het meeste baat bij? b. Welke eisen wegen het zwaarst?

Bedankt voor jullie deelname! Hebben jullie nog vragen?

Artsen/verpleegster:

Bedankt voor jullie deelname aan mijn onderzoek. Zoals jullie weten gaat mijn onderzoek over het implementeren van innovatieve technieken in de huidkanker zorg met als doel om de efficiëntie te verbeteren. In het specifiek proberen we een manier te vinden waarbij patiënten die al eerder huidkanker hebben gehad, alleen nog op controle hoeven te komen wanneer er ook daadwerkelijk sprake is van huidkanker. Het idee achter de implementatie is dat patiënten gevraagde zullen worden om voorgaand aan een controle foto's te maken van plekjes die ze verdacht vinden en daarbij een korte vragenlijst in te vullen. Om deze implementatie voor zowel de arts als patiënt een verbetering te laten zijn, ontvang ik graag jullie input over een aantal onderwerpen.

1. Hoe vinden jullie momenteel dat het zorgproces verloopt voor patiënten die al eens eerder huidkanker hebben gehad en dus geregeld op controle moeten komen? a. Wat loopt er goed b. Wat zou er beter kunnen

2. Wat zijn jullie ervaringen met het one-stop-shop systeem? +/-

3. Denken jullie dat er nog winst valt te behalen binnen het one-stop-shop systeem, op welke manier?

4. Hoe zouden jullie het vinden om data van patiënten te ontvangen en op basis daarvan een risico inschatting te maken? a. Wat zouden de voordelen daarvan zijn b. Wat zouden de nadelen daarvan zijn

5. Wat voor data zou je allemaal nodig hebben om een risico inschatting te kunnen maken? (Foto, vragenlijst en welke vragen dan)

6. Verwacht je dat patiënten zelf in staat zijn om deze data correct aan te leveren?

7. Op welke manier zou je de data het liefst ontvangen? (E-mail, via een website, gedeelde map etc.)

8. Wat voor toepassing zie je zelf voor je die mogelijk wordt gemaakt door het ontvangen van data van patiënten?

Een van de toepassingen die ik onderzoek is het uitstellen van controles wanneer het risico aan de hand van data laag blijkt.

9. Wat zijn jouw opvattingen over deze toepassing? a. Welke voordelen zie je? b. Welke nadelen?

10. Heb je een idee over hoe de patiënt geïnformeerd zou moeten worden over het wel of niet uitstellen van een controle? 11. Welke eisen vind je van belang wanneer deze toepassing wordt ingesteld? (Denk aan snel in gebruik, privacy gericht, etc.)

Een andere situatie zou kunnen zijn dat aan de hand van de data alvast een inschatting kan worden gemaakt van welke behandeling er plaats zou moeten vinden.

12. Wat zijn jouw opvattingen over deze toepassing? a. Welke voordelen zie je? b. Welke nadelen?

13. Welke eisen vind je van belang wanneer deze toepassing wordt ingesteld? (Denk aan genoeg zekerheid, genoeg bedenktijd voor patiënt etc.)

De risico inschatting die wordt gedaan zou geautomatiseerd kunnen worden met behulp van een kunstmatige intelligentie. In dat geval worden de gegevens door een computer geanalyseerd en kan die aangeven wat het risico is dat het wel of niet huidkanker is.

14. Wat zijn jouw opvattingen over deze toepassing? a. Welke voordelen zie je? b. Welke nadelen?

15. Welke eisen vind je van belang wanneer deze toepassing wordt ingesteld? (Denk aan genoeg zekerheid, privacy, uitlegbaarheid etc.)

16. Waar zou een KI aan moeten voldoen in jouw optiek om de plaats van een zuster/dermatoloog over te nemen?

17. Hoe zie je zelf de samenwerking tussen dermatoloog/zuster en KI voor je? a. In hoeverre heb je er vertrouwen in dat een KI de goede keuze maakt? b. Wie zou je zien als verantwoordelijke voor de uitkomst van een KI?

18. Wie denken jullie dat het meest beïnvloed worden door deze nieuwe toepassingen? (Arts, patiënt, MohsA, verzekeraar) a. Wie heeft er het meeste baat bij? b. Welke eisen wegen het zwaarst?

Bedankt voor je deelname! Hebben jullie nog vragen?

Administratief medewerker:

Bedankt voor jullie deelname aan mijn onderzoek. Zoals jullie weten gaat mijn onderzoek over het implementeren van innovatieve technieken in de huidkanker zorg met als doel om de efficiëntie te verbeteren. In het specifiek proberen we een manier te vinden waarbij patiënten die al eerder huidkanker hebben gehad, alleen nog op controle hoeven te komen wanneer er ook daadwerkelijk sprake is van huidkanker. Het idee achter de implementatie is dat patiënten gevraagde zullen worden om voorgaand aan een controle foto's te maken van plekjes die ze verdacht vinden en daarbij een korte vragenlijst in te vullen. Om deze implementatie voor zowel de arts als patiënt een verbetering te laten zijn, ontvang ik graag jullie input over een aantal onderwerpen.

1. Hoe vind je dat het huidige proces verloopt met betrekking tot patiënten die regelmatig op controle komen? a. Wat gaat er goed b. Wat kan er beter

2. Wat zou je ervan vinden wanneer patiënten vooraf data insturen die door artsen/verpleegsters moet worden behandeld? a. Voordelen/nadelen

3. Hoe zou je dit proces voor je zien?

4. Via welke kanalen zou je de data het liefst ontvangen?

5. Vanuit jouw ervaring, zijn er valkuilen of dingen waar expliciet rekening mee gehouden moet worden voor zo'n toepassing?

6. Hoe vind je dat patiënten geïnformeerd moeten worden over eventuele uitstel van controle?

7. Zou je zelf een toepassing kunnen bedenken die gebruik maakt van de data van patië om het zorgproces efficiënter te maken?

8. Wie denk je dat het meest beïnvloed worden door deze nieuwe toepassingen? (Arts, patiënt, MohsA, verzekeraar) a. Wie heeft er het meeste baat bij? b. Welke eisen wegen het zwaarst?