

Digital Health History Taking and Data-driven Information Delivery

Improving Patients Satisfaction with Orthopedic Outpatient
Consultations

Tristan Warren
STUDENT NUMBER: 2043905

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & ENTREPRENEURSHIP

Thesis committee:
Dr. Laura Genga
Dr. Florian Böing-Messing

Tilburg University
Eindhoven University of Technology
Jheronimus Academy of Data Science
's-Hertogenbosch, The Netherlands
June 2021

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to the people who made this thesis possible.

Thank you, Dr. Laura Genga, for your guidance and advice during the past year, but most of all, thank you for approving the unconventional setup of my work. I would also like to thank my second supervisor Dr. Florian Böing-Messing for reviewing my work.

I owe a deep sense of gratitude to Walter van der Weegen, a researcher at St. Anna's Knowledge Center for Orthopedic Surgery and my supervisor at the hospital. Your excitement about AI is influencing many in the world of orthopaedics to hop on the hype train, and your influence on this thesis is undeniable. I probably learned as much from you as you did from me, which in my eyes, is more valuable than anything else.

I would like to pay my special regards to Michiel Siebelt and Dirk Das, orthopaedic surgeons at St. Anna. You have shown me every side of orthopaedic care, from consultation to surgery, to ensure that I knew what I was doing it all for. I would also like to thank you for your contributions in setting up the interventions, which likely would not have been of the same quality without.

I am incredibly thankful to Thomas Timmers, Co-Founder of Interactive Studios, for granting me access to OnlinePROMS and the Patient Journey App and for his help in setting it all up. I can not wait to discover how our future collaborations will unfold.

Last but definitely not least, I am indebted to all the orthopaedic surgeons of St. Anna, the employees of the outpatient clinic, and the participants of the trial who I have never met but in one way or another have been an essential piece of the puzzle.

Thank you all,

Tristan Warren

Contents

1	Introduction	2
1.1	Research Problem	3
1.2	Objectives	8
1.3	Outline	8
2	Background and Related Work	9
2.1	Measuring Patient Satisfaction	9
2.2	Health History Questionnaires	12
2.3	Educational eHealth applications	13
2.4	Machine Learning in Orthopedics	15
2.5	Ethical, Legal, Policy, and Practice Considerations	19
3	Methodology: Model Development	22
3.1	Data Source and Patient Selection	22
3.2	Data Safety	23
3.3	Outcome of Interest	23
3.4	Candidate Variables of Interest	26
3.5	Missing Data	30
3.6	Pre-processing	31
3.7	Model Selection	32
3.8	Quality Metrics	34
3.9	Modelling Process	35
3.10	Software	40
4	Results: Model Development	41
4.1	Validation Performance	43
4.2	Test Performance	45
4.3	Prediction Tool	49
5	Discussion: Model Development	51
5.1	Clinical Implications	51
5.2	Limitations	52
5.3	Future Work	52
6	Methodology: Randomised Controlled Trial	54
6.1	Objectives	54
6.2	Trial Design	54
6.3	Informed Consent and Ethical Consideration	55
6.4	Participant Selection	56
6.5	Randomization	56
6.6	Intervention	57
6.7	Study Outcomes	60
6.8	Statistical Analysis & Sample Size	63
6.9	Software	64
7	Pilot results: Randomized Controlled Trial	65
7.1	Validation performance	65
7.2	Participants	66
7.3	Outcomes and Estimation	69
8	Discussion: Randomised Controlled Trial	76
8.1	Strengths and limitations	76
8.2	Future Work	78
9	Conclusion	79

Appendices	93
A Original Distribution of Knee Diagnoses	93
B Original Distribution of Hip Diagnoses	93
C Original Distribution of Knee Treatments	94
D Original Distribution of Hip Treatments	94
E Selected hyperparameters	95
F METC n-WMO Applicaton form	96
G Study Protocol	102
H METC Study Approval Letter	106
I MEC St. Anna Study Approval Letter	107
J RCT Invitation Email	108
K Study Information and Consent Form	109
K.1 Introduction and Eligiblity (first page)	109
K.2 Study information (second page)	110
K.3 Informed Consent Form (third page)	112
L Study Exclusion Email	113
M Randomisation and Prediction Registration	114
N Group Assignment Email	115
O Timeline Gonarthrosis Surgical	116
P Timeline Gonarthrosis Conservative	131
Q Timeline Degenerative Meniscus Tear	136
R Real- and Perceived Knowledge Questionnaires	142
R.1 Gonarthrosis Surgical	143
R.2 Gonarthrosis Conservative	145
R.3 Degenerative Meniscus Tear	147
S Patient Satisfaction Questionnaire	149
T Subjective Experience Questionnaire	151

Digital Health History Taking and Data-driven Information Delivery

Tristan Warren

Abstract. As the need for orthopaedic care continues to increase, trying to keep patients satisfied with their consultations is becoming an increasingly challenging task. The increasing administrative burden of orthopaedic surgeons is impairing their ability to prepare for their consultations properly, which causes them to be less able to properly communicate with, empathise with and relate to their patients. Furthermore, a lack of medical knowledge in patients and limitations in their memory and understanding are causing orthopaedic surgeons to face patients with unrealistic expectations and poor patient-doctor communication increasingly often. These factors have been reported to negatively influence patients' satisfaction with their consultation.

Hence, the primary objective of the current research was to make orthopaedic consultations future-proof by supporting orthopaedic surgeons in preparing for their consultations and increasing patients' medical knowledge and, by extension, increase patient satisfaction. A digital Health History Questionnaire was introduced to support orthopaedic surgeons in preparing for their consultations. Simultaneously, an informational eHealth application was introduced to provide information to patients before their consultation regarding a prediction of their diagnosis and appropriate treatment to increase their a priori medical knowledge.

The study was split into two separate phases. In the first phase, Machine Learning methods were applied Health History data of patients with hip and knee complaints to develop predictive models that can predict diagnoses of the knee and hip and the appropriate way to treat the complaints. The developed models' performance was moderate, but hip-related models performed slightly better than knee-related models. The predictive models were developed to provide information to the patients in the subsequent research phase.

In the second phase, a randomised controlled trial was performed. The control group received standard consultations, whereas the experimental group received the interventions. Unfortunately, the study was not finished within the timeframe of this thesis, and as a result, the trial is still ongoing at time of writing. However, the pilot results were analysed to identify early trends. First, participants in the experimental group ($n=14$) seemed to have a higher level of real knowledge than participants in the control group ($n=16$), but the difference was not statistically significant. Second, participants in the experimental group seemed to have a higher level of perceived knowledge than participants in the control group. In contrast to real knowledge, this difference was significant. Finally, participants in the experimental group seemed to be more satisfied with their consultation than participants in the control group, but the difference was not statistically significant.

With the trial still ongoing, it was too early to draw definitive conclusions regarding patients satisfaction with their consultations or patients medical knowledge. The final results of the trial will have to show if the study achieves its primary objective of improving patients satisfaction with orthopedic outpatient consultations.

1 Introduction

The St. Anna Hospital is an orthopaedic hospital in Geldrop, the Netherlands, focused on the care of patients with complaints related to the musculoskeletal system. The number of patients visiting the hospital has steadily increased over the years. The rise in the number of patients is caused by increasing demand for orthopaedic care in the Netherlands, driven by an ageing population (Smits et al., 2014).

One of the departments at St Anna that has noticed the increase in the number of patients is the outpatient clinic, of which medical consultations are an essential activity. During an outpatient consultation, the goal of the orthopaedic surgeon is to determine the differential diagnosis and appropriate treatment. The orthopaedic surgeon begins a consultation by taking the patient's medical history, which is traditionally a major part of the orthopaedic consultation. It is well-known that the diagnosis of a patient is often revealed by the patient's history (Tidy, 2019). Taking an accurate history of the patient can provide eighty percent or more of the information that would be required for a correct diagnosis (Epstein, Perkin, and Cookson, 2008). An aphorism that is often quoted is 'Listen to your patient; they are telling you the diagnosis' (Smith, 2003).

After taking the patient's history, the orthopaedic surgeon summarizes and reviews what the patient has said. If necessary, the orthopaedic surgeon performs a physical examination or evaluates medical imaging results to confirm his or her findings. Based on the collected information, the orthopaedic surgeon determines the differential diagnosis and discusses possible treatments together with the patient. If all goes well, the patient has received a correct diagnosis and a suitable treatment by the end of the consultation.

As the need for orthopaedic care will continue to increase, trying to keep patients satisfied with their consultations becomes an increasingly challenging task. Patient satisfaction is an emotion-driven cognitive evaluation of the care that is received (Urden, 2002; Crow et al., 2002) and the extent to which a patient is satisfied with a consultation is dependent on a variety of factors (Waters et al., 2016).

A significant influence on patients satisfaction with consultations is the quality of the communication between the physician and the patient, also called physician-patient communication. Good physician-patient communication is characterized by effective interaction between the physician and the patient. Patients consider good communication to be vital for feeling included and connected during the consultation.

A different aspect that patients consider to be important is the clinical contact time. Patients tend to dislike short consultations because they can make them feel like they are being managed and rushed through the consultation, which can cause disappointment in patients.

The expectations that patients have are of importance as well. For example, patients frequently have expectations regarding their consultation that the physician is unable to meet. If the expectations of a patient match with reality, it is more likely that the patient will be satisfied with the consultation.

Likewise, that consultations are led by physicians that can empathize and relate to their patients can adopt the patient's perspective, understand the patient's beliefs and feelings, and express their understanding to the patient in a positive manner. Physicians with the ability to relate to their patients can make their patients feel connected, respected, and understood. Patients tend to be more satisfied with consultations led by physicians who possess these abilities than with consultations led by physicians who lack these abilities.

A survey developed by St. Anna revealed that their patients are currently quite satisfied with their consultation ($n=143$). For example, patients rated the time that the orthopaedic surgeon had for them with an average score of 3.64/5 (95% CI [3.45, 3.82]). Furthermore, patients rated the empathy of their orthopaedic surgeon with an average score of 3.96/5 (95% CI [3.79, 4.13]). Patients were least satisfied with the extent to which their expectations were met, which they rated with an average score of 2.88/5 (95% CI [2.61, 3.15]). Finally, patients rated the overall care at the outpatient clinic with an average score of 3.65/5 (95% CI [3.5, 3.8]).

Despite the fact that patients at St. Anna are currently quite satisfied with their consultations, continuously improving their patient's satisfaction is a primary ambition of St. Anna. Moreover, the increasing demand for orthopaedic care makes it evident that there are no guarantees for the future. Several problems that might severely harm patients satisfaction with their consultation are easily identified and necessitate examination to ensure that orthopaedic outpatient consultations remain future-proof for many years to come.

1.1 Research Problem

A significant source of potential problems affecting patients' satisfaction is the poor medical knowledge and understanding of the average patient. Medical knowledge refers to knowledge about the patient's possible condition(s) and the treatments associated with these conditions. Although patients usually receive this information during a consultation, possessing medical knowledge before the consultation takes place has several advantages.

First, knowledge about conditions and treatment possibilities is crucial for effective doctor-patient communication (Kyle and Shaw, 2014). For example, patient participation in the shared decision-making process is strongly dependent on the patient's knowledge and understanding of the benefits and drawbacks of the potential treatments for his or her condition (du Long et al., 2016).

A high level of medical knowledge has also been found to reduce the anxiety that patients experience (Beddoes, 1997; Devine, 1992), which is reported in the literature as a cause of poor physician-patient communication (Fischer and Ereat, 2012). For that reason, patients with a high level of medical knowledge are more likely to communicate effectively during their consultation.

Finally, patients with a high level of medical knowledge often have more realistic expectations than patients with a low level of medical knowledge (Thomas and Sethares, 2008). Patients with realistic expectations are more likely to be satisfied with the care that they receive because it is more likely that their needs are met (Greene et al., 2013).

On top of having a poor level of a priori medical knowledge, research has shown that patients have difficulty remembering the information presented to them during their consultation. This limitation was found to be particularly strong for the retention of medical information (Mcguire, 1996). Of the mere 40% to 80% of what patients do remember, about half turns out to be inaccurate (Kessels, 2003). Moreover, the information presented during a consultation is often poorly understood (Wills, 2008). Most patients have a high need for information about their condition and treatment (Attfield, Adams, and Blandford, 2006). However, the limitations in their retention and understanding of medical knowledge are interfering with their ability to obtain and remember the desired information.

There is a wide range of factors that contribute to these limitations in patients' memory and understanding. Patient-specific factors include the age of the patient (Ekdahl, Andersson, and Friedrichsen, 2010; Kawabata et al., 2009) and their level of education (Marks et al., 2010). What is more, communication skills are less emphasised within orthopaedic surgery than diagnostic and surgical skills, contributing to the often unsatisfactory communication skills of orthopaedic surgeons. The lack of emphasis on communication skills is amplified by the busy schedules of physicians who, as a consequence, provide too much information in too little time (Seli et al., 2011) whilst using language and jargon that is difficult for the average patient to understand (Say, Murtagh, and Thomson, 2006).

These busy schedules are caused by a considerable increase in the administrative burden of orthopaedic surgeons in recent years, while the duration of a standard consultation remains relatively short. One element of a consultation that is known to add to the administrative burden of orthopaedic surgeons is taking the patient's health history. Orthopaedic surgeons currently perform health history taking by looking at the general practitioner's referral (GP) and taking the patient's medical history. Referrals by GP's are generally not well-standardised, and the information that they contain is often not sufficiently complete to sketch a good picture of the patient's situation (Tobin-Schnittger et al., 2018).

Furthermore, the quality of the information that the surgeon is able to collect is highly dependent on factors such as preparation and communication between the physician and the patient (Gask and Usherwood, 2002). One example related to communication is that many patients find it difficult to express themselves about their complaints and clinical histories (While, 2002). Hence, they often need the guidance of the physician to support them in this process.

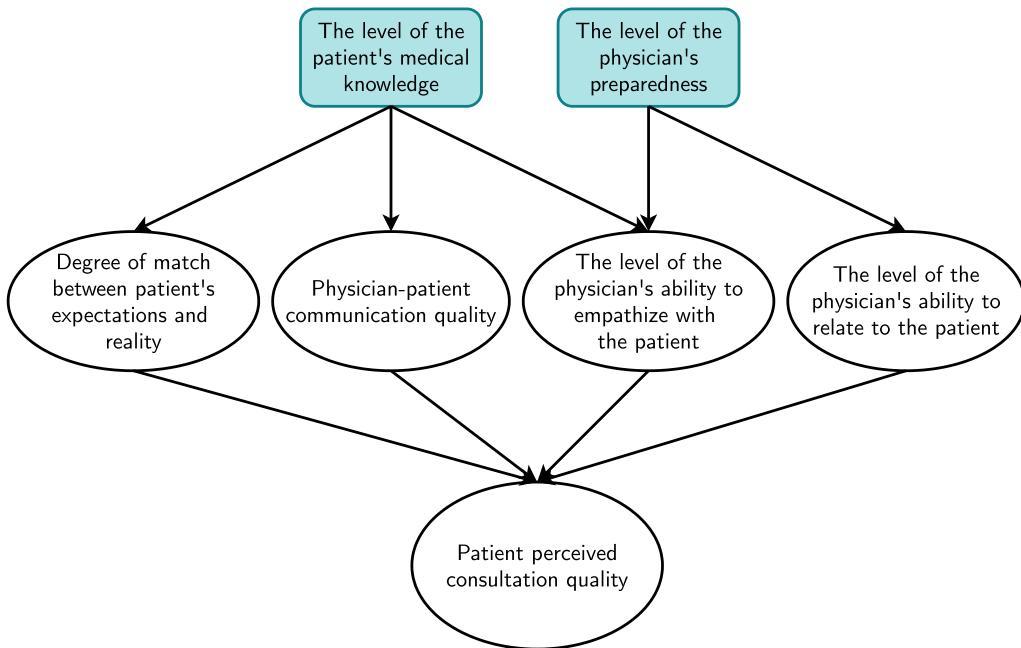
All things considered, taking a patient's health history is a complex process that can easily result in dissatisfied patients. For a physician, the process of history taking can be compared to that of a detective: 'searching for clues, collecting information without bias, yet staying on track to solve the puzzle' (Kaufman, 2008).

In summary, two problems have been identified that are indirectly affecting patients satisfaction with their consultation. First, a lack of a priori medical knowledge in patients, and limitations in their memory and understanding of medical knowledge are causing orthopaedic surgeons to face patients with unrealistic expectations increasingly often and are causing the communication between the surgeon and the patient to be far from optimal frequently. Furthermore, the increasing administrative burden of orthopaedic surgeons is impairing their ability to prepare for their consultations properly. In turn, inadequate preparation can cause orthopaedic surgeons to be unable to properly communicate with, empathise with and relate to their patients.

Patient expectations, doctor-patient communication, and the ability of orthopaedic surgeons to empathise with and relate to their patients have all been reported to influence patients' satisfaction with their consultation. Based on these interactions, a summary model was drawn by the author, which is presented in figure 1. Please note that this model is by no means a complete model of all the factors that influence a patient's satisfaction with a consultation. The presented model is a simplification that aims to provide an overview of the identified factors for this research's purposes.

The main objective of the current research is to make orthopaedic consultations future-proof by supporting orthopaedic surgeons in preparing for their consultations and increasing patients' medical knowledge and, most importantly, increase patients satisfaction with their consultations.

Figure 1: Summary model of factors and their interactions that influence patient satisfaction with consultations. The blue boxes represent the factors that the current study aims to improve.



First, the introduction of digital health history questionnaires might support orthopaedic surgeons in preparing for their consultations. The St. Anna hospital has already implemented a validated health history questionnaire, but the potential effect of the questionnaire on the satisfaction of patients with their consultation has yet to be measured. A health history questionnaire contains questions regarding information that the orthopaedic surgeon would commonly gather by taking the patient's health history. Such information includes demographic information, complaints, and the relevant medical history of the patient. Patients would be asked to fill out a health history questionnaire online before visiting the outpatient clinic. After that, the questionnaire becomes available in the Electronic Patient Record (EPR) of the hospital and can be accessed by the orthopaedic surgeon before the consultation.

In early conversations with orthopaedic surgeons at St. Anna, allowing the orthopaedic surgeon to prepare a consultation with a health history questionnaire was claimed to affect their preparedness positively. They asserted that the positive effect is caused by the questionnaire's unambiguous questions and the standardised representation of the information. They remark that GP referrals become less important and that they can spend more time on 'social talk' because the emphasis of the health history taking part in the consultation shifts from gathering as much information as possible to discussing the information gathered from the questionnaire. As a result, orthopaedic surgeons at St. Anna believe that they become more capable of tuning

their communication to individual patients, become more capable of empathising with their patients, and become better able to relate to their patients. However, it is currently unknown whether these claims are true and whether the introduction of health history has increased patients' satisfaction with their consultations.

Second, increasing patients' medical knowledge can potentially be achieved by educating patients through an educational eHealth application. eHealth has many definitions (Oh et al., 2005) but the definition by Pretlow (2000) summarises the concept perfectly: "E-health is the process of providing health care via electronic means, in particular over the Internet. It can include teaching, monitoring (e.g. physiologic data), and interaction with healthcare providers, as well as interaction with other patients afflicted with the same condition". With the technological advances of the past years, eHealth interventions have been gaining traction as a tool to, among other things, increase patient engagement (Barello et al., 2016), to address the limited capacity of the health care system (Ahern, Kreslake, and Phalen, 2006), and to enable health behaviour change (Ahern, Kreslake, and Phalen, 2006).

In the field of orthopaedics, the application of eHealth tools has also been investigated in recent years. An educational eHealth application enables the caregiver to provide information to patients in an easily accessible way and allows the caregiver to tailor the information to the specific preferences and needs of the patient without the interference of a physician (Bental, Cawsey, and Jones, 1999). In contrast to information provided during a consultation, information provided through an app is complete, consistent, and continually available to the patient (Adsit, 1996). A positive side effect is that patients are free to determine their own pace of education (Kahn, 1993).

An earlier study conducted at St. Anna hospital investigated whether the medical knowledge of osteoarthritis patients can be increased by offering information regarding their diagnosis and possible treatments by means of a mobile eHealth application (Timmers et al., 2018). The results of this study showed that offering information to osteoarthritis patients before the consultation takes place has a positive influence on both the real and the self-reported medical knowledge of patients.

Similar educational eHealth tools aimed at improving shared-decision making have been shown to yield high satisfaction rates. Moreover, they have a positive influence on the knowledge of patients, their involvement in decision-making, and the number of questions that they ask during a consultation (Belkora et al., 2015; Shue et al., 2016). Based on the main objective and the previously discussed options to support orthopaedic surgeons and increase patients medical knowledge, the following main research question was formulated:

Main RQ : Is the satisfaction of patients whose consultations are prepared by the orthopaedic surgeon by examining their digital health history questionnaire and prepared by the patient with an educational eHealth application higher than the satisfaction of patients that receive a standard consultation?

To investigate the direct effect of an educational eHealth application on patients medical knowledge, the following sub-questions were formulated:

Sub RQ 1 : Is the actual and perceived medical knowledge of orthopaedic patients increased by providing them with information before their consultation by means of an educational eHealth application compared to patients who do not receive this information?

Providing information to patients before their consultation is a challenge because their diagnosis and treatment are commonly determined during the consultation. Providing information to patients before their consultation requires that there is some information available about their possible diagnoses and treatment before the consultation takes place. One possibility is to predict what the diagnosis and treatment will be, and Machine Learning offers a set of techniques that might be able to achieve that.

Machine Learning (ML) is a sub-field of Artificial Intelligence (AI), which broadly refers to the "mimicking of human cognition by computers" (Jha and Topol, 2016). ML itself involves computational algorithms that can learn from experience, where experience comes in the form of data. ML has been gaining traction in the field of orthopaedics in recent years (Cabitza, Locoro, and Banfi, 2018) and is able to solve prediction problems in areas such as prosthesis control (LeMoyne et al., 2015) and spine pathology detection (Jamaludin et al., 2017a). Fortunately, the previously introduced Health History Questionnaire might provide a valuable data source for developing algorithms that can predict diagnoses and treatments. Taking into account that the correctness of the information provided to patients depends on the correctness of the predictions on which the information is based, the following sub-question was formulated:

Sub RQ 2 : How accurately can orthopaedic diagnoses be predicted by machine learning algorithms trained on health history questionnaire data?

Finally, to find out what research has been done on these topics in the past and to identify what research gaps still exist, the following sub-questions were formulated:

Sub RQ 3 : How can patient satisfaction be measured?

Sub RQ 4 : What research regarding the utility of health history questionnaires has been done in the past, and what is still missing?

Sub RQ 5 : What research has been done on the utility of education eHealth applications for improving the medical knowledge and satisfaction of patients, and what is still missing?

Sub RQ 6 : What research on Machine Learning in orthopaedics has been done in the past, and what is still missing?

Sub RQ 7 : Are there any ethical, legal, policy, or practical considerations that have to be considered in the context of the current research?

Sub RQ 3, 4, 5, 6 and 7 are answered by reviewing previous research on the topics in question. After that, the current research is split into two subsequent phases, each with its own methodology, results, and discussion section. First, an attempt is made to answer **Sub RQ 2** by applying machine learning techniques to health history questionnaire data in an effort to predict orthopaedic diagnoses and treatments accurately. In this research phase, the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research ([Luo et al., 2016](#)) as well as the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) ([Moons et al., 2015](#)) were followed.

After that, **main RQ** and **Sub RQ 2** are answered by performing a randomised controlled trial (RCT) at the St. Anna hospital according to CONSORT guidelines ([Schulz, Altma, and Moher, 2010](#)). The control group receives standard consultations, whereas the patients in the experimental group receive information regarding a prediction of their diagnosis and treatment by means of an educational eHealth application. The models that are developed in the first phase of the study are used to provide predictions.

1.2 Objectives

In summary, the aim of the present study is twofold. The primary objective is to improve patients' satisfaction with orthopaedic consultations by supporting orthopaedic surgeons in preparing for their consultations with the already implemented health history questionnaire and increasing patients' medical knowledge through an educational eHealth tool. The secondary objective is to apply machine learning techniques to health history data to develop models that can predict orthopaedic diagnoses and treatments sufficiently accurate that they can be used to provide correct information to patients.

1.3 Outline

The outline of the current paper is as follows. First, an overview of the previous research related to this research is presented in section [2](#). After that, the paper is split into two phases. The first phase covers the methodology (section [3](#)) and results (section [4](#)) of the model development process and concludes with a discussion of the results (section [5](#)). The second phase covers the methodology (section [6](#)) and results (section [7](#)) of the randomized controlled trial that is performed at St. Anna to investigate if the introduction of the interventions achieves the primary objective of this study. The second phase concludes with a discussion of the trial results (section [8](#)). The paper is wrapped up with section [9](#), in which the two research phases are unified and their main findings summarised.

2 Background and Related Work

In this section, the study background and review previous work on the topics related to this research are described. First, **sub RQ 3** is answered by reviewing research related to measuring patient satisfaction in section [2.1](#). After that, a deep dive into the history and application of health history questionnaires is provided in section [2.2](#) to answer **sub RQ 4**. In section [2.3](#), **sub RQ 5** is answered by discussing research related to educational eHealth tools and their applications in orthopaedics, which is followed with an examination of the state of the art research related to machine learning in orthopaedics in section [2.4](#) to answer **sub RQ 6**. Finally, to answer **sub RQ 7**, ethical, legal, policy, and practice considerations are identified in section [2.5](#).

2.1 Measuring Patient Satisfaction

To evaluate how patient satisfaction is affected by the introduction of the intervention, it is necessary to find a proper way to measure patient satisfaction. There has been plenty of research on this topic, but a paper by [Shirley, Josephson, and Sanders \(2016\)](#) provides a comprehensive overview of the fundamentals of patient satisfaction measurement. For that reason, the main points from that paper and the research that it references is summarised.

The purpose of measuring patient satisfaction is to identify the strengths and limitations of the department of care under investigation and to improve the quality of care accordingly ([Espinel et al., 2014](#)). Patient satisfaction can be measured through both quantitative and qualitative methods ([Shirley, Josephson, and Sanders, 2016](#)).

2.1.1 Measurement Instruments

Qualitative methods measure patient satisfaction by observing or by interviewing patients one-on-one or in focus groups. Although such methods potentially result in higher response rates ([Burroughs et al., 2001](#)), disadvantages of qualitative methods are frequently used as arguments for opting for a quantitative method instead ([Fottler, Ford, and Bach, 1997](#)). Disadvantages of qualitative methods include the lack of statistical validity and reliability, the possibility of observer bias and high labour intensity.

Quantitative methods measure patient satisfaction more accurately, have higher statistical validity and reliability, and can often be automated. These methods often come in the form of surveys. They tend to score patient satisfaction on multiple dimensions because a global measure of patient satisfaction is unlikely to capture the specifics of why a patient was satisfied or not ([Shirley, Josephson, and Sanders, 2016](#)). Dimensions that are often included in quantitative surveys are caregiver characteristics, the technical quality of care, the physical environment, and outcomes of care ([Ware et al., 1983](#)). In any case, the number of dimensions in a survey should be limited to an amount that is acceptable to patients in terms of the time required to complete it ([Beattie et al., 2014](#)).

The content of a survey should consist of meticulously worded, multiple response questions. These questions include both simple yes-no questions and multipoint scales such as the 5-point Likert scale (Clark, 2003), which asks respondents to specify their level of agreement or disagreement according to the following scale: 1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly agree. Respondents may have additional remarks that the survey does not cover. Including a section for comments is recommended to capture this information (Shirley, Josephson, and Sanders, 2016).

2.1.2 Validity and Reliability

Patient satisfaction surveys are only helpful for improving care if they are reliable and valid (Urdan, 2002). Validity refers to the extent to which a survey represents what it is trying to measure. There are various types of validity such as face, criterion, and content validity (Shirley, Josephson, and Sanders, 2016), but in the context of measuring patient satisfaction, content validity is the most critical. Content validity refers to the extent to which the items in a survey represent the entire concept that it seeks to measure. In other words, for a patient satisfaction survey to have content validity, the questions in the survey should cover all the factors that ultimately influence patient satisfaction. Unfortunately, there is no direct method to measure content validity beforehand, so evaluation of content validity relies on expert knowledge and research (Barnett et al., 2013). However, it is possible to get an indication of the validity of the survey by evaluating the responsiveness of the survey (Shirley, Josephson, and Sanders, 2016). For a patient satisfaction survey to be responsive, the satisfaction should increase if interventions are implemented to improve it. If this is not the case, the survey might have measured factors not related to patient satisfaction or that the survey failed to include critical factors that influence patient satisfaction.

Reliability refers to the extent to which respondents would answer consistently if the survey is administered more than once under the same conditions (Shirley, Josephson, and Sanders, 2016). In contrast to validity, there are three strategies to estimate reliability: test-retest reliability, equivalent forms reliability, and internal consistency reliability (Brown, 1997). Test-retest and equivalent forms reliability are unfeasible for the current research because it requires the survey to be administered at least twice or two versions of the survey. An example of a popular internal consistency reliability measurement is Cronbach Alpha (Cronbach, 1960), which indicates if each specific question would be answered consistently by a respondent if the questions are presented repeatedly to him or her.

2.1.3 Biases

Surveys are prone to biases (Shirley, Josephson, and Sanders, 2016). Failing to take them into account can damage the validity of the survey. The most common biases are:

1. **Optimizing bias** can arise when questions are difficult to understand or interpret. Answers to these questions are unreliable. Therefore the formulation of questions must be simple and unambiguous.
2. **Recall bias** can arise when questions are asked that relate to the past of the respondent. The further back in time a question is asked, the higher the chance that the patients' memory is incomplete or incorrect. Hence it is best if questions are asked about recent events.
3. **Satisficing bias** can arise when surveys are too long. As a consequence, respondents might try to rush through the questions by giving arbitrary answers that do not represent the true experience of the patient. This can be prevented by keeping the survey short.
4. **Social desirability bias** can arise if patients consciously or subconsciously give socially desirable or socially undesirable answers that do not reflect the genuine experience of the respondent. There are several methods to prevent this, including protecting the confidentiality of the patient and administering self-report surveys instead of surveys taken in person. (Durmaz, Dursun, and Kabadayi, 2020).
5. **Central tendency bias** can arise if respondents have a tendency to avoid extreme answers. This is a major problem for Likert scales. It is not easy to prevent but providing clear and unambiguous questions without requiring respondents to justify their answers can help to mitigate the chance of it to occur(KyLeads, n.d.).
6. **Halo bias** can arise if respondents have a tendency to unjustifiably generalize their positive experience in one area to their experience of other areas.

2.1.4 Existing instruments

Thousands of patient satisfaction surveys were developed over time (Morris et al., 2014). However, very few of these surveys have been verified in terms of validity and reliability (Sitzia, 1999; Barnett et al., 2013). Four popular verified surveys include the Client Satisfaction Questionnaire (CSQ-18) (Larsen et al., 1979), the Short Assessment of Patient Satisfaction (SAPS) (Hawthorne et al., 2014) questionnaire, the Consultation Satisfaction Questionnaire (ConsultSQ) (Baker, 1990), and the Patient Satisfaction Index (PSI) questionnaire (Guyatt et al., 1995). A golden standard survey does not exist (Espinel et al., 2014) so the choice of the survey depends on the specific requirements of the context in which it will be used.

ConsultSQ seems to be the most appropriate choice for this research given its specific focus on clinician consultations. ConsultQ assigns 18 questions to three different factors: professional care, depth of the relationship between doctor and patient, and perceived time. However, it does not touch upon all the specific factors that addressed in this research. For that reason a custom survey was developed in collaboration with researchers from St. Anna. The survey combines relevant questions from the most popular and validated surveys with several new questions constructed specifically for my research.

2.2 Health History Questionnaires

The classical approach to history taking is a labour-intensive process that often takes up a significant chunk of the clinical contact time. According to the general framework for taking the medical history of a patient in orthopaedics ([Epomedicine, 2014](#)), the orthopaedic surgeon generally collects information about the patient by questioning him or her about a range of topics such as demographic information, complaints and history of the presenting illness, and other possibly relevant medical and personal histories.

Clearly, the patient's history in the classical approach of health history taking is expressed through subjective narrative descriptions by the patient. Consequently, health history questionnaires have been advocated as a complement to the classical health history taking interview. A health history questionnaire is a survey that patients complete before visiting their consultation. It contains questions related to information that the orthopaedic surgeon would normally collect while taking the patient's history during the consultation.

In contrast to classical health history taking, health history questionnaires allow for the collection of health history data in a quantitative and structured manner ([Pincus, Schmukler, and Castrejón, 2019](#)). What is more, health history questionnaires improve the quantity and quality of the information that is collected ([Hall, 1972](#)) and optimize the time spent with the patient ([Gumpel and Mason, 1974](#)). For instance, name([Hall, 1972](#)) found that health history questionnaires have a positive effect on the ability of physicians to concentrate on their patients' presenting symptoms, whereas name([Gumpel and Mason, 1974](#)) reported that patients provide information on paper that they might not have mentioned during a consultation.

Research by [Inui et al. \(1979\)](#) supports the suggestion that the implementation of health history questionnaires results in more complete patient records. They also add that health history questionnaires contribute to "earlier and more active patient participation during the history-taking process" and that their implementation supports physicians in identifying additional medical problems that the physician would not have noticed otherwise. Other theorized benefits include the standardization of data collection, reduced interview bias, and uniformity of administration ([Collen et al., 1969](#))

The validity and reliability of health history questionnaires have been evaluated and confirmed in a variety of medical settings (see for example [Gilkison, Fenton, and Lester \(1992\)](#), [Pecoraro et al. \(1979\)](#), and [Katz et al. \(1996\)](#)). A concrete example is a study carried out by [Boissonnault and Badke \(2005\)](#), who found that the mean percentage of agreement between health history items recorded by orthopaedic surgeons during an outpatient consultation and patients' self-reported health history items across questionnaire items was 96%. Additionally, they found that their results are in line with an earlier study performed by [Pecoraro et al. \(1979\)](#), which implies that the orthopaedic surgeon can be confident that self-reports by patients are correct.

However, research into the effect of their application on patients' satisfaction with their consultation is lacking. Furthermore, existing research related to health history questionnaires is often quite dated, and these studies always provided the questionnaire to patients on paper. Hence, new research into the utility of health history questionnaires in a digital environment and their effect on patient's satisfaction with their consultation is warranted.

2.3 Educational eHealth applications

The introduction established that patients might be educated by means of an educational eHealth tool. In the current section, previous research on educational eHealth tools introduced to educate patients and/or improve patient satisfaction is reviewed to get an idea of the utility of educational eHealth tools for increasing patient's knowledge and satisfaction and to identify gaps in the literature.

[Claassen et al. \(2020\)](#) investigated the effect of using a stand-alone mobile and web-based educational eHealth tool to prepare for orthopaedic consultations on patient satisfaction. They found that the intervention did not result in higher patient satisfaction with the consultations, but it did increase patients' medical knowledge. They also provided multiple explanations for the lack of success in achieving an improvement in patient satisfaction.

First, they argued that their intervention was possibly not comprehensive enough to positively influence patient satisfaction due to their decision to focus primarily on changing patients' expectations. Instead, they referred to the fact that patient satisfaction is multifaceted and that targeting additional factors such as communication and empathy might yield better results. Second, they used the Dutch standard for measuring patient experience: the Consumer Quality Index (CQI) ([Delnoij, Rademakers, and Groenewegen, 2010](#)), which did not properly reflect their primary outcome. Accordingly, they proposed that future research should use a validated questionnaire that incorporates the specific satisfaction-influencing factors targeted by the intervention.

In earlier research by [Timmers et al. \(2018\)](#), the efficacy of using an educational smartphone app to increase patients' medical knowledge was investigated. They found that patients who received the intervention had a 52% increase in their actual level of medical knowledge compared to patients who received standard care.

The authors argued that their tool seems relatively effective compared to decision aids introduced in earlier studies. They specifically cite the systematic review by [Stacey et al. \(2017\)](#), in which a total of 52 studies was reviewed. From the 13,316 participants in these studies, the medical knowledge of the average patient increased by a mere 13,27%. They theorised that the difference in knowledge gain was because they used a smartphone app as a decision aid, in contrast to earlier studies that provided decision aids in the form of booklets, phone calls, websites, one-on-one conversations, DVDs, or group sessions.

Additionally, the app integrated a combination of validated information retention mechanisms that earlier studies did not. First, the app provided small bits of explicitly categorised ([Kessels, 2003](#)) information at a time instead of providing all information at once ([Jansen et al., 2008; Ferguson et al., 2015](#)). Furthermore, they provided patients with the information both by video and in text. Previous research showed that people remember up to 80% if these modalities are offered together, while if only one of both is offered, people remember only up to 20% or 40% respectively. Finally, they implemented quiz-like questions with instant feedback in an attempt to reinforce the retention of information ([Jansen et al., 2008; Tait, Teig, and Voepel-Lewis, 2014](#)).

The application of smartphone or tablet applications as a means to provide education to patients seemingly started to gain interest around the time of the study by Timmers et al. In a recent study by Timmers et al. (2020), randomised controlled trials that were performed in the period 2015-2019 and which investigated education of patients by means of a mobile or tablet app were systematically reviewed. They included only research from this period because earlier research generally provided education through other means. They assessed the effects that these interventions had on several outcomes. However, only interventions that aimed to increase patients' medical knowledge or that aimed to improve patient's satisfaction with the information or care they received are relevant for the current research.

All four studies that aimed to improve the medical knowledge of patients were successful. One of these studies was the study by Timmers et al. (2018), which was also the only reported research in the context of orthopaedics. The other studies were aimed at increasing knowledge about atrial fibrillation (Guo et al., 2017), knowledge about breast cancer and screening options (Lee et al., 2017), or knowledge about diabetes mellitus (Alanzi et al., 2018). Although none of these studies aimed to increase patients' medical knowledge to improve their satisfaction with consultations, they provide evidence that mobile educational eHealth applications effectively increase patients' medical knowledge.

In contrast, mobile educational eHealth applications seem less effective at improving patients' satisfaction with the overall care they received. From the eight studies that investigated the effect, only four reported being successful at improving patient satisfaction. However, neither successful nor unsuccessful studies specifically aimed to improve patient satisfaction with consultations. For example, Meij et al. (2018) aimed to improve patients' satisfaction with abdominal surgery while Li et al. (2019) aimed to improve patient's satisfaction with the recovery process after receiving pediatric surgery.

2.3.1 Research gaps

Pursuant to the overview of educational eHealth applications in orthopaedics, the following research gaps were identified.

First, whereas the effectiveness of educational eHealth apps in improving patients' knowledge has been consistently confirmed in previous research, only two studies aimed to improve the medical knowledge of patients' before their consultation (Timmers et al., 2018; Claassen et al., 2020). All other studies focused on perio- and postoperative education or self-management. On top of that, not a single study incorporated personalised information based on algorithmic predictions.

For example, a limitation of the study by Timmers et al. was that possible osteoarthritis patients had to be manually pre-selected according to their pathology because the diagnosis of patients is unknown before the consultation. Algorithmic-based predictions of the diagnosis and treatment allow for the automation of this process, enabling application of the intervention in practice without requiring the intervention of a researcher that has to determine the diagnosis manually. As a consequence, it also becomes feasible to offer the intervention to a larger population.

Another interesting difference that warrants new research is the fact that Machine Learning algorithms are never flawless. Whereas existing educational eHealth applications always offer consistent and correct information, an educational eHealth application that provides personalised information based on algorithmic predictions will occasionally provide wrong information to patients. It is entirely possible that providing wrong information to patients changes the effectiveness of the application in terms of improving the medical knowledge of patients.

What is more, it might also have an impact on patient satisfaction. It is possible that patients who receive wrong information would have been more satisfied if they did not receive any information at all. On the other hand, many diagnoses have some common elements. As a result, information about different diagnoses will have common elements as well. Perhaps these common elements are sufficient to increase satisfaction, even if the information is not entirely correct.

Past research undoubtedly begs whether educational eHealth applications are a suitable approach to improving patient satisfaction at all, with about half of the studies reporting to have failed in that task. Moreover, there are only two studies that investigated their effect in improving patients satisfaction with orthopaedic consultations (Timmers et al., 2018; Claassen et al., 2020), and both reported that satisfaction did not increase. However, the current study takes their limitations into account and aims to improve on them. Aside from building on top of previous work, the current study adds health history questionnaires to the equation. It is conceivable that the combination of both interventions would lead to patient satisfaction that is greater than the sum of its parts.

2.4 Machine Learning in Orthopedics

In addition to the implementation of health history questionnaires to support orthopaedic surgeons in preparing their consultations, these questionnaires might be a valuable data source for developing machine learning models that can predict orthopaedic diagnoses and treatments.

Machine Learning (ML) is a sub-field of Artificial Intelligence (AI), which broadly refers to the "mimicking of human cognition by computers" (Jha and Topol, 2016). ML itself involves computational algorithms that can learn from experience, where experience comes in the form of data. ML can extract and learn patterns from data and build models that describe the behaviour of a system by linking covariates in the data to a target variable (Gupta et al., 2020; Deo, 2015). In order to build ML models, the input data is split into a training and test set. The training set is used to build the mathematical model, while the test set is used to assess the model's performance.

The learning that takes place is achieved by incrementally updating the internal parameters of the model through mathematical functions such that the predicted outputs of the model match increasingly well with the desired outputs (Cabitzza, Locoro, and Banfi, 2018). ML is often compared to the human ability to learn from experience, which is realised by continually assessing new inputs and updating the analyses that are performed accordingly (Myers et al., 2020). As the expertise of an orthopaedic surgeon, the performance of ML algorithms is strongly dependent on the quality and volume of their training data (Makhni, Makhni, and Ramkumar, 2020).

ML algorithms are typically divided into supervised and unsupervised algorithms (Galbusera, Casaroli, and Bassani, 2019). The basis for supervised ML is labelled data sets. Labelled data sets are data sets with data that has been labelled with its desired or true classification. For example, suppose the task is to detect arthritis on medical imaging data of knees. In that case, the images need to be labelled with whether the image in question concerns a patient with knee arthritis or a normal knee. With labelled data as input, supervised ML algorithms attempt to learn a function that maps the input images to their correct labels.

In contrast, rather than predicting an output, unsupervised algorithms attempt to learn the underlying structure of the data and uncover patterns or groupings within the data (Choy et al., 2018). In this case, the data is not labelled. Instead, the algorithm looks for clusters or self-organisation in the data.

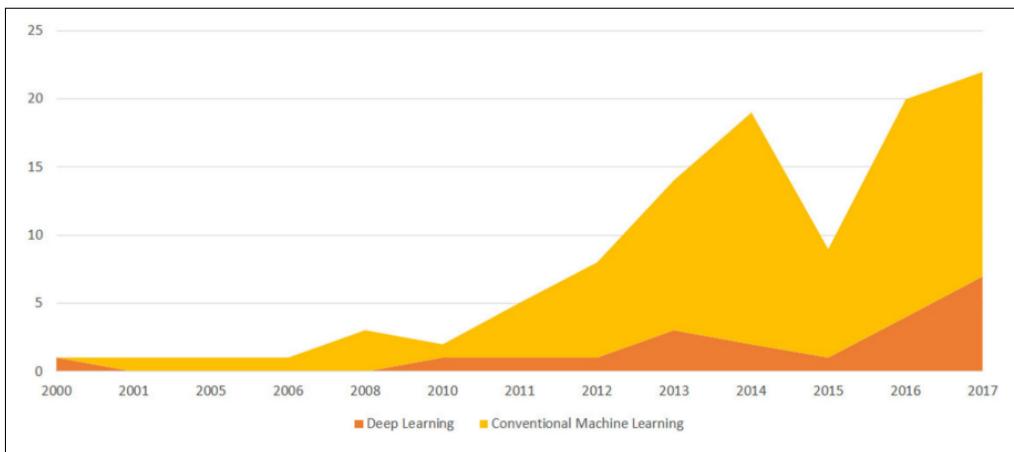
2.4.1 Overview of Machine Learning in Orthopedics

The last few decades have enabled rapid adoption of new technologies into orthopaedic practice, including new types of implants, 3D printing, and CAD-CAM design (Poduval et al., 2020). However, whereas the application of Machine Learning is already popular in medical fields such as mental health (Shatte, Hutchinson, and Teague, 2019), dermatology (Thomsen et al., 2019), and cardiovascular diseases (Krittanawong et al., 2020), Machine Learning research in the field of orthopaedics is still relatively under-represented (Cabitza, Locoro, and Banfi, 2018). The popularity of ML in other fields of medicine is due to their success, with some studies concluding that ML models can outperform physicians themselves (Brynjolfsson and Mitchell, 2017).

Luckily, there has been a 10-fold increase in the amount of research in recent years (see figure 2), which reveals that there is a growing interest in the application of Machine Learning in the field of orthopaedics. Maffulli et al. (2020) posit that this growing interest is fueled by "growing datasets, integration of medical records with images and financial transactions, systematic recording of patient outcomes, increasing storage capacity, and a steep rise in affordable computational power". Despite being a laggard, orthopaedics is one of the most technologically innovative fields in medicine (Maffulli et al., 2020) and has even been posited to be "uniquely suited to harness the power of big data" compared to other fields of medicine (Helm et al., 2020). Accordingly, Machine Learning is slowly becoming the new disruptive force in orthopaedics.

Machine Learning can be a valuable tool for a wide range of orthopaedic applications, including prediction of post-operative complications, clinical decision making, and pathology detection (Han and Tian, 2019). In general, everyday tasks that surgeons find repetitive, such as history taking, are especially suited for Machine Learning. Repetitive tasks can suffer from human error, and supporting the surgeon in these tasks may alleviate some of the time pressure that they experience and, as a consequence, provide them with valuable time to focus on their patients (Bien et al., 2018; DeVries et al., 2019). For instance, several studies found that surgeons supported by a ML tool when interpreting imaging results allocate less time to this task than surgeons who perform this task without support (Gilbert et al., 2008; Hollon et al., 2019; Lindsey et al., 2018). Moreover, ML tools offer increased patient safety and efficiency of care. For example, a study by Ghomrawi et al. (2017) found that it is challenging for surgeons to predict whether a patient will benefit from total knee replacement, with expert surgeons barely able to predict better than chance. In contrast, Fontana et al. (2019) found that a ML

Figure 2: Machine learning in orthopedics research over time. Adapted from "Machine Learning in Orthopedics: A Literature Review," by Cabitza F, Locoro A and Banfi G, 2018 , Front. Bioeng. Biotechnol, 6:75, Copyright 2018 by "Cabitza, Locoro and Banfi"



algorithm was able to predict whether a patient will benefit from total knee replacement relatively well.

An overview of the number of papers that have been published for the application of specific ML techniques to particular orthopaedic tasks was constructed in a literature review by [Cabitza, Locoro, and Banfi \(2018\)](#). According to that review, the most prevalent application of machine learning in orthopaedics is the detection and prediction of osteoarthritis. The machine learning technique is applied the most is the Support Vector Machine. Table 1 shows an overview of the best performance measures achieved per field of application. It shows that overall, high performance was achieved in each of these fields of application. An overview of the main data sources that have been used for the predictive modelling of these orthopaedic problems can be found in table 2. This table shows that the majority of existing research has used medical imaging as a data source.

The literature review by [Cabitza, Locoro, and Banfi](#) is the most complete literature review related to ML in orthopaedics to date. However, [Maffulli et al. \(2020\)](#) found that there are over 3300 published articles related to AI and ML in orthopaedics, of which 1100 were published in the previous two years alone.

Whereas much of the earlier research primarily focused on imaging-based pathology detection and prediction, as can be observed in figure ?? and table 2, the focus of more recent research has shifted to patient-specific, value-based care models ([Ramkumar et al., 2019](#)), to using preoperative patient data for the identification of risk factors for complications ([Haeblerle et al., 2019](#)), and to support clinicians and patients in the process of shared decision-making ([Sambare, Uhler, and Bozic, 2017](#)). Moreover, hospitals are increasingly often partnering up with software companies to establish more data-driven and personalised healthcare (see for example ([Clinic, 2016](#))).

Table 1: State of the art metric scores achieved per field of application

Field of application	Metric score
Spine pathology detection	95% accuracy (Jamaludin et al., 2017b)
Bone age assessment	1.15 MAE (Spampinato et al., 2017)
Prosthesis control	100% accuracy (LeMoyne et al., 2015)
Gait classification	95% accuracy (Pogorelc and Gams, 2010)
Osteoarthritis detection	98% accuracy (Phinyomark et al., 2016)
Fracture detection	0.81 AUC (Atkinson et al., 2012)
Image segmentation	99% accuracy (Prasoon et al., 2013)

Table 2: Data sources for machine learning in orthopedics

Data Source	Number of papers
Biochemical data	3
Bioelectrical data	4
Medical Imaging data	56
Patient data	17
Sensor data	16
Video data	1

It seems that the application of Machine Learning is maturing in the field of orthopaedics. However, we are also at a point where these technologies need to be tested and validated extensively in real clinical workflows, and their impact measured on actual patient-specific clinical outcomes ([Parikh, Obermeyer, and Navathe, 2019](#)), patient experiences, costs, and patient limitations ([Jayakumar, Moore, and Bozic, 2019](#)). Additionally, performing such experimental research in the form of cohort studies and randomised controlled trial will be crucial for ML to become widely accepted and embraced in orthopaedics ([Cabitza, Locoro, and Banfi, 2018](#)).

2.4.2 Research Gaps

The current research attempts to address the following research gaps that were identified in the existing literature. First, medical imaging data is the most popular data source for ML in orthopaedics (see table 2) and is also the data source used in most of the state of art applications. In comparison, patient data is less popular, and studies that used patient data as a data source found that performance was often relatively poor. For example, [Zupan et al. \(2001\)](#) attempted to predict the long-term clinical status of patient's after hip arthroplasty using hierarchical decision modelling, but their results revealed an unsatisfactory classification accuracy of 56%. Better results were found by [Yoo et al. \(2013\)](#), who reported an AUC of 0.827 with a Support Vector Machine trained on medical records with the purpose to predict the risk of osteoporosis in postmenopausal women.

To my knowledge, there are no studies that attempt to differentiate between multiple orthopaedic diagnoses or pathologies. Researchers tend to focus on predicting the presence or the progression of a single diagnosis or pathology (see for example [Kotti et al. \(2017\)](#) or [Ashinsky et al. \(2017\)](#)).

It seems that patient data is not often considered as a source of data for orthopaedic ML problems. It also seems that multiclass or multidiagnosis problems are generally not considered in the existing literature. Hence, the current study fills that research gap by attempting to solve the problem of predicting a patient's diagnosis from a range of known diagnoses and applying ML algorithms to a novel type of patient data.

The patient data used in this study can be considered novel because it is collected pre-consultation and related to data that the orthopaedic surgeon would typically collect when taking the patient's history. In contrast, patient data in the existing literature is typically preoperative or postoperative data. On top of that, applying ML to history taking questionnaire data is unprecedented, warranting an investigation into its value. Finally, no studies were identified that attempt to predict which orthopaedic treatment a patient will receive or whether the treatment will be conservative or operative.

2.5 Ethical, Legal, Policy, and Practice Considerations

Despite the potential of health history questionnaires, Machine Learning and eHealth applications, there are ethical, legal, policy, and practice considerations that need to be taken into account, especially in the context of healthcare ([Matheny et al., 2019](#)).

2.5.1 Acceptance

According to several studies, the acceptance of AI and ML by both the general public and physicians is an issue. A study from the UK has reported that 63% of the population that they questioned disapproves replacement of physicians by ML for standard tasks and, what is worse, they are fundamentally opposed to the use of personal data for the improvement of healthcare in general ([Fenech and Bouston, 2018](#)). A more recent study from Germany reported similar findings. They found that patients generally prefer that clinical tasks are performed by physicians over clinical tasks that are performed purely with AI, except for treatment planning ([Lennartz et al., 2021](#)). They also observed that patients are more likely to trust the physician's opinion when the physician and the

AI disagree on the diagnosis or treatment. However, the same study also found that patients are more favourable towards scenarios where the physician supervises the AI. This sentiment is substantiated by [Jayakumar, Moore, and Bozic \(2019\)](#), who suggest that given the complexity of clinical care, technologies like ML should preferably enhance and complement the physician rather than act as a substitute. Physicians seem to agree as well, with 44% of the consulted medical students in a study by [Santos et al. \(2018\)](#) being sceptical about AI making conclusive diagnoses. In any case, communicating with and educating patients and physicians about ML is a first step towards increasing the acceptance of both ([Vayena, Blasimme, and Cohen, 2018](#)).

2.5.2 Physician Understanding

According to the studies discussed in the previous section, ML and AI will not replace physicians anytime soon. Instead, physicians are more likely to use these tools as a complement to their clinical tasks. It is, therefore, imperative for physicians to be involved in the development of algorithms and to understand both the characteristics of the population used to develop the algorithm and the algorithm itself ([Liu et al., 2019](#)). Physicians that adopt ML tools in their practice without understanding the data that goes in, how the tool comes to a conclusion, and how they should interpret its conclusions are at a potential risk to "legal, medical malpractice, and product-liability issues" ([Jayakumar, Moore, and Bozic, 2019](#)).

2.5.3 Data Privacy and Ownership

The development of MI models for healthcare requires increasingly large datasets. Not merely does the collection, storage, and processing of these datasets call for the implementation of high-quality standards but also for laws and regulations regarding data privacy and data ownership ([Price and Cohen, 2019](#); [Racine, Boehlen, and Sample, 2019](#); [Gruson et al., 2018](#)). Unfortunately, the pace of the development and implementation of new regulatory frameworks and ethical guidelines does not seem to match with the pace of technological progress in healthcare ([Parikh, Obermeyer, and Navathe, 2019](#)).

A recent attempt by the European Union to tackle privacy concerns is the EU-wide General Data Protection Regulation (GDPR). The GDPR imposes requirements on the use of personal data and protects data subjects by granting them rights on concepts such as data portability, transparency, rectification, and erasure that must be honoured by the parties that process their data ([Mccall, 2018](#)).

The GDPR strictly defines personal data as "any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." On top of that, the GDPR introduces additional definitions that pertain to health data of which one applies to the current research: "Data concerning health is defined as personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status."

In case personal data concerns health data, the GDPR prohibits the collection, storage and processing of this data unless one of the following three conditions applies ([Rohatgi, 2018](#)):

1. "Processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services ..."
2. "Processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices ..."
3. The data subject has given "informed, specific, freely given, unambiguous and explicit consent." ([Mostert et al., 2016](#))

The constituents for consent are defined as follows. For consent to be informed and specific and to prevent 'function creep', the individual (data subject) must be made aware of the entity that handles the data (the data controller), which data will be processed, and for what purposes the data will be processed ([Intersoft Consulting, 2021](#)). The processing purposes must be sufficiently explained and need to be understandable by the data subject. For consent to qualify as freely given, the data subject must have given consent voluntarily. Moreover, the data subject must be able to refuse or withdraw consent without any negative consequences. Finally, for consent to be unambiguous, the data subject must consent through a statement or a clear affirmative act. There can be no misunderstanding that the data subject has given consent to process his or her data.

Regardless of which condition is applicable, the data subject has additional rights that the data processor has to respect. First, patients must be guaranteed their right of access to and erasure of the data that is gathered about them whenever they want. Second, patients have the right to data portability, which means that patients are free to share their data with other health care providers. Finally, the processing of the data should be limited to what is strictly necessary for processing purposes.

3 Methodology: Model Development

In this section, the methodology that was applied for the development of Machine Learning models to predict diagnoses and treatments related to the hip and knee is described. The methodology is reported according to the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research (Luo et al., 2016) as well as the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (Moons et al., 2015). The purpose of these models is to provide personalized information to patients regarding their predicted diagnosis and treatment.

In section 3.1 the data collection process is explained and the data sets are described. After that, details on data safety and how the privacy of the patients whose data was collected was protected is elaborated in section 3.2. Next, the prediction problems, their limitations, and how these limitations were handled is explained in section 3.3. This is followed by sections 3.4, 3.5, 3.6. In these sections, the variables of the data sets, how missing data was handled, and how the data sets were preprocessed is described. These sections are followed with a description of the algorithms that were selected (section 3.7) and the metrics that were used to evaluate the algorithms (section 3.8). The methodology is finished with an explanation of the model development process in section 3.9 and the software that was used in section 3.10.

3.1 Data Source and Patient Selection

Patient-Reported Outcome Measures are standardized and validated questionnaires that “seek to ascertain patients’ views of their symptoms, their functional status, and their health-related quality of life” (Black, 2013). OnlinePROMS (Interactive-Studios, 2020) is a web-based software that enables the implementation of PROMs in practice through web-based surveys. The care provider can manually set up measurement processes that consist of one or more surveys that aim to track some patient outcome, for example, recovery after surgery.

The St. Anna hospital (Geldrop, the Netherlands) has been using OnlinePROMS to administer a health history questionnaire. Patients that make an appointment for an outpatient consultation after being referred by their general practitioner are assigned to this measurement process by the employee who makes the appointment. The patient then receives an invitation for the questionnaire by mail. The questionnaire first asks patients which part of the body is the source of their complaints. After that, the questionnaire asks patients questions that relate to that specific extremity. The survey questions have been composed by orthopaedic surgeons based on their professional experience and relate to topics that they often address during the health history taking of the patient.

The data produced by this survey forms the basis for the development of the predictive models. Therefore the current research is a retrospective study. The data has been collected between March 2018 and December 2020. The outcome diagnoses and treatments were added to the dataset post-consultation by the physician. However, they only did this for diagnoses and treatments of complaints related to the hip and knee. These two extremities are the most common sources of complaints within St. Anna. For these reasons, the author made the decision to restrict the research to developing models that predict diagnoses and treatments of complaints related to the hip and knee.

A total of 1037 filled out health history questionnaires related to the knee, and a total of 972 filled out health history questionnaires related to the hip were labelled with a diagnosis and treatment and were considered for inclusion.

3.2 Data Safety

According to the General Data Protection Regulation definitions, the current research deals with personal health data because the health history data relates to personal data that reveals the physical health of natural persons. For that reason, it is necessary to find a valid ground on which to process the data.

The decision was made to process the data on the following ground: "Processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services....". This study aims to improve the quality of care that patients receive. One might argue that improving the quality of care is not strictly necessary, but this is up for debate.

St. Anna implemented the following safeguards to prevent unlawful processing of the data. First, patients are asked by default to consent to the following statement before filling in the health history questionnaire: "The answers you enter will be used for monitoring your treatment and for internal quality improvement of our care. The anonymised data is provided to client organisations, government and regulators. In addition, if you are receiving a prosthesis, your answers are stored in the national prosthesis database (LROI). Participation in these questionnaires is voluntary." This statement refers to quality improvement as a processing purpose. Unfortunately, this statement is not sufficient for explicit consent because it is not specific enough. Second, the data was anonymised. The anonymisation process combined with the general nature of the data makes it very unlikely that the data can be related to a specific patient without considerable effort. Finally, the developed models based on this data will strictly be used to provide information to patients. The models will remain property of the St. Anna hospital and will never be shared with third parties.

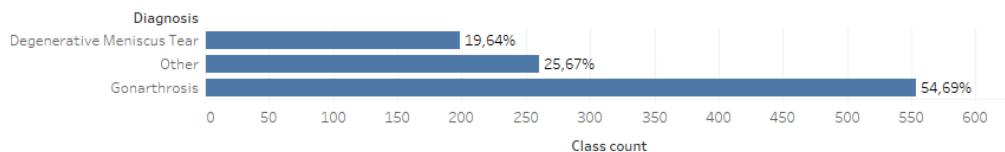
3.3 Outcome of Interest

There were four different outcomes of interest. The first outcome of interest was the clinical diagnosis of patients with knee-related complaints. Orthopaedic surgeons at St. Anna use a classification system that includes 37 diagnoses of knee complaints. Unfortunately, most of these diagnoses occur infrequently by nature. This natural imbalance in the prevalence of diagnoses also influenced the distribution of diagnoses in the data set.

Until a sufficient number of records is collected for each class, including each diagnosis independently was not an option. On the other hand, excluding all infrequently occurring diagnoses was also not an option because there had to be a way to intercept these patients. Otherwise, any patient with an illness that deviated from the included illnesses would have been diagnosed with one of the included diagnoses, introducing wrong predictions by default. Therefore the choice was made to include infrequently occurring diagnoses by combining all diagnoses in a single "other" class except for the two most occurring diagnoses.

The original distribution of knee diagnoses in the data set is presented in appendix A. Figure 4 shows the class distribution after rearranging the diagnoses. The 36 least occurring diagnoses combined accounted for 25.7% (n=231) of the diagnoses in the data set. The two most occurring diagnoses were degenerative meniscus tear (n=193 [19.6%]) and gonarthrosis (n=539 [54.7%]). Degenerative meniscus tear is the degeneration of cartilage in the meniscus, which causes the meniscus to tear. It is often seen as an early sign of the most-occurring diagnosis: gonarthrosis. Gonarthrosis is the progressive wear of the knee joint. Together, they accounted for nearly 75% of all diagnoses in the dataset. Hence, successful differentiation between these two diagnoses and the remaining diagnoses would make it possible to provide information to 3/4 patients, and it would allow for the interception of the remaining 25% of patients.

Figure 3: Class distribution of knee diagnoses



The second outcome of interest was the clinical diagnosis of patients with hip-related complaints. Orthopaedic surgeons at St. Anna use a classification system that includes 19 diagnoses of hip complaints. Like the knee diagnoses, a large imbalance in the prevalence of diagnoses was present in the hip data set. Therefore the diagnoses in this data set were rearranged according to the same method as the knee data set. The original distribution of hip diagnoses in the data set can be found in appendix B. Figure 4 shows the class distribution of hip diagnoses in the hip data set after rearranging the diagnoses. The 17 least occurring diagnoses combined accounted for 27% (n=96) of the diagnoses in the data set. The two most occurring diagnoses were trochanteric bursitis (n=49 [13.8%]) and coxarthrosis (n=212 [59.2%]). Trochanteric bursitis is the inflammation of the trochanteric bursa, a fluid-filled sac at the outside point of the hip known as the greater trochanter. Coxarthrosis is the progressive wear of cartilage in the hip joint. Together they accounted for over 70% of all diagnoses in the hip data set. Hence, successful differentiation between these two diagnoses and the remaining diagnoses would make it possible to provide information to 7/10 patients, and it would allow the interception the 3/10 remaining patients.

Figure 4: Class distribution of hip diagnoses

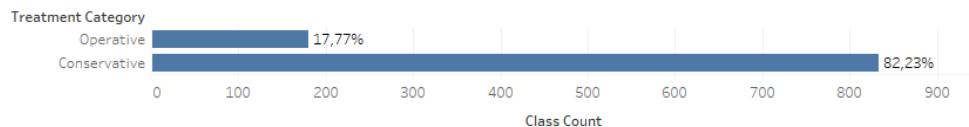


The third outcome of interest was the choice of treatment for patients with knee-related complaints. The same data set that was used to develop the knee diagnosis prediction models was used to develop the knee treatment models. However, the two models were developed independently of each other. Accordingly, the treatments were excluded from the data set during the development of the diagnosis prediction models and diagnoses during the development of the treatment prediction models. The reason for this is a practical one: the diagnosis and treatment are unknown at the time of prediction. As a consequence, it was not possible to take them into account.

Orthopaedic surgeons at St. Anna use a classification system that includes 38 possible treatments of knee complaints. Most of these treatments are chosen infrequently, with the top five most chosen treatments accounting for more than 50% of all chosen treatments. Treatments generally belong to one of two categories: conservative and surgical treatments. Conservative treatments include, for example, physiotherapy and pain medication. Surgical treatments include total knee replacement surgery and arthroscopy. Although the data did not allow for predicting specific treatments due to the low prevalence of most treatments, it might allow for accurate prediction of these two treatment categories. Hence, successful differentiation between these two treatment categories would make it possible to determine which treatment category is most appropriate, and it would allow patients to consider their options beforehand.

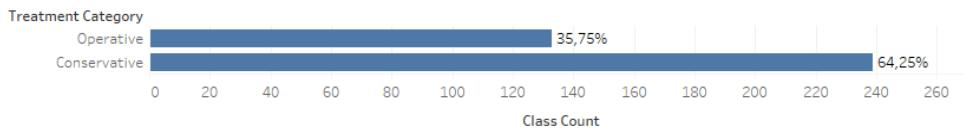
The original distribution of knee treatments in the data set can be found in appendix C. Figure 5 shows the distribution of surgical and conservative treatments after assigning each treatment to either the conservative or surgical class. Although there was still an imbalance in favour of conservative treatments ($n=793$ [82.2%]) over surgical treatments ($n=170$ [17.8%]), differentiating between two imbalanced treatment categories is easier than differentiating between 38 specific treatments.

Figure 5: Class distribution of knee treatment categories



The last outcome of interest was the choice of treatment for patients with hip-related complaints. Identical to the knee diagnosis and treatment models, the hip diagnosis and treatment models were developed independently. Orthopaedic surgeons at St. Anna use a classification system that includes 20 possible treatments of hip complaints. Like the knee data set, the top four most occurring treatments accounted for almost 50% of all chosen treatments. Therefore the choice was made to apply the same categorization strategy that was applied to knee treatments. The original distribution of hip treatments in the data set can be found in appendix D. Figure 6 shows the distribution of surgical and conservative treatments after assigning each treatment to either the conservative or the surgical class. Although there was still an imbalance in favour of conservative treatments ($n=227$ [64.25%]) over surgical treatments ($n=130$ [35.75%]), differentiating between two imbalanced treatment categories is easier than differentiating between 38 specific treatments.

Figure 6: Class distribution of hip treatment categories



Combining the infrequently occurring diagnoses and by classifying the specific treatments as either conservative or surgical improved the balance of the classes up to a point. However, the distributions after rearranging the diagnoses and treatments show that there was still a strong imbalance between the classes for each outcome of interest. The methods that were applied to take class imbalance into account and to mitigate its problems are described in the sections to follow.

3.4 Candidate Variables of Interest

All questions included in the health history questionnaires were considered candidate variables. The knee and hip health history questionnaires included standard history-taking questions for suspected knee or hip pathologies such as the location of pain, trauma, and severity and duration of symptoms. These questions were supplemented with questions from validated PROM questionnaires such as the Oxford Hip Score (OHS) (De Groot et al., 2007), the Oxford Knee Score (OKS) (Haverkamp et al., 2005), and Numeric Rating Scale (NRS) (Salaffi et al., 2004) for measuring the severity of pain. The knee health history questionnaire and the hip health history questionnaire had overlapping questions, which are presented in table 3. The candidate input variables that were unique to the knee data set are described in table 4. The candidate input variables that were unique to the hip data set are described in table 5.

Table 3: Shared Candidate Input Variables

Variable	Details
Age	Years
Length	Centimeters
Weight	Kilograms
Do you work?	Yes/No
Did you have to stop working/exercising/hobby's due to your complaints?	(a) No, I don't experience any complaints during these activities; (b) No, but I do experience complaints during these activities; (c) I had to stop doing my hobby's and/or sport but I am still working; (d) Yes, I had to stop all activities
Are you allergic to antibiotics?	Yes/No
Do you smoke	Yes/No

Table 3 continued from previous page

Variable	Details
For how long have you been experiencing complaints?	(a) A few days, (b) A few weeks, (c) A few months, (d) A few years
Did your complaints deteriorate or did they improve since your complaints started?	(a) Improved, (b) Stayed the same, (c) Varying in severity, (d) Detoriated
Do you experience pain or are you restricted in your mobility?	(a) No pain or mobility restriction, (b) Predominantly pain, (c) Predominantly mobility restriction, (d) Both pain and mobility restriction
How long can you walk before the pain starts to increase?	(a) Longer than 30 minutes, (b) 16 to 30 minutes, (c) 5 to 15 minutes, (d) Only in and around the house, (e) Not at all
How much pain do you experience when you are not moving?	Scale of 1 to 10
How much pain do you experience when you are moving?	Scale of 1 to 10
Have you tried muscle strengthening exercises for a period of 4 to 6 weeks?	(a) No, I have not yet tried muscle strengthening exercises; (b) Yes, I have recently started muscle strengthening exercises (less than 4 weeks of training); (c) Yes, I tried muscle strengthening exercises for at least four weeks but without results; (d) Yes, I tried muscle strengthening exercises for at least four weeks and it reduced my complaints
Have you used any pain medication for your complaints?	(a) No, I don't use pain medication; (b) Yes, paracetemol; (c) Yes, NSAIDs (ibuprofen, diclofenac, naproxen, celecoxib, etoricoxib, meloxicam); (d) Yes, Tramadol; (e) Yes, opioid (oxycodon, oxynorm, morphine)
If you used pain medication, did they reduce your complaints?	Yes/No
With which treatment do you think you can get rid of your complaints?	(a) A non-surgical treatment (conservative), (b) Arthroscopy, (c) A prosthesis surgery, (d) I do not know
Would you want a surgery if it is necessary?	1. No; 2. Maybe; 3. Yes

Table 4: Candidate Input Variables Knee Dataset

Variable	Details
Which knee is the cause of your complaints?	(a) left; (b) both, left more than right; (c) right; (d) right more than left
Is your knee swollen and / or thick?	(a) not at all, (b) a little bit, (c) very
Where does you knee hurt the most?	(a) On the outside, (b) On the inside, (c) On top of the kneecap, (d) On top of the knee tendon, (e) Behind the kneecap
Have you every had surgery of the left knee?	Yes/No
What kind of knee surgery did you have?	(a) Arthroscopy, (b) Anterior cruciate ligament surgery, (c) Posterior cruciate ligament surgery, (d) Position correction, (e) Bursa surgery, (f) Partial knee replacement, (g) Complete knee replacement, (h) Revision knee prosthesis, (i) Other, (f) Not applicable
Have you ever had surgery of the right knee?	Yes/No
Did your knee complaint arise after a twisting, accident or other acute moment?	Yes/No
Did your kneecap dislocate from this trauma?	(a) No, this did not happen; (b) This has happened in the past but is not the cause of my complaints; (c) Yes; (d) Not applicable
How long ago did the accident and / or twisting happen?	(a) 0 to 2 days, (b) 2 to 7 days, (c) 1 to 2 weeks, (d) More than two weeks ago, (e) Not applicable
Are you able to stand on the affected knee with a straitghened knee?	(a) Yes, (b) No, (c) Not applicable
Do you suffer from an unstable feeling in your knee?	(a) No, my knee does not feel unstable; (b) No, my knee is too painful to move; (c) Yes, my knee feels unstable
Can you bend you knee all the way?	(a) Yes; (b) Yes, but it hurts; (c) No
Can you straighten your knee completely?	(a) Yes; (b) Yes, but it hurts; (c) No
Does your knee feel stiff during the first steps of walking?	(a) Yes, (b) Sometimes, (c) No
Have you started walking differently?	(a) No; (b) Yes, someone told me; (c) Yes, I noticed myself
Do you currently have lower back complaints?	Yes/No
Have you ever had surgery on your back?	Yes/No

Table 4 continued from previous page

Variable	Details
Do you ever have groin pain?	(a) No; (b) Yes, on the same side as my affected knee; (c) Yes, but on the other side of my affected knee; (d) Yes, in both my groins
Have you ever had an injection (infiltration) in your knee?	(a) No; (b) Yes, but this has not helped me enough; (c) Yes, this reduced my complaints for a few weeks

Table 5: Candidate Input Variables Hip Dataset

Variable	Details
Which hip causes most complaints?	(a) Left, (b) Right, (c) Both
Where does your hip hurt the most?	(a) At the front of the thigh, (b) In the groin, (c) Over the buttock / back of the thigh, (d) On the side of the thigh
Can you put on your socks, stockings or tights yourself?	(a) Yes, easily; (b) With a little effort; (c) With some effort; (d) With a lot of effort; (e) I can't
Can you climb stairs?	(a) Yes, easily; (b) With a little effort; (c) With some effort; (d) With a lot of effort; (e) I can't
Does your knee feel stiff during the first steps of walking?	(a) Yes, (b) Sometimes, (c) No
Have you started walking differently?	(a) No; (b) Yes, someone told me; (c) Yes, I noticed myself
Have you previously had surgery on your left hip?	Yes/No
Have you previously had surgery on your right hip?	Yes/No
What kind of surgery did you have?	(a) Complete hip replacement, (b) Operated for a fracture of the hip, (c) Hip dysplasia surgery, (d) Arthroscopy, (e) Bursa, (f) Other, (g) Not applicable
Do you currently have lower back complaints?	Yes/No
Have you ever had surgery on your back?	Yes/No
Do you ever have knee pain?	(a) No; (b) Yes, on the same side as my affected hip; (c) Yes, but on the opposite side of my affected hip; (d) Yes, in both knees

3.5 Missing Data

An overview of data that was missing in the knee data set is presented in figure 7. An overview of data that was missing in the hip data set is presented in figure 8. All variables with less than 1% missing values were grouped as a single variable that shows the maximum number of missing values within that group of variables. All rows from the data set containing a missing value for one or more of these variables were removed. The small percentage limit of missingness ensured that removing these rows did not significantly decrease the size of the data set.

Figure 7 also shows five variables with an equal number of missing values. These variables are related to trauma. Patients with knee complaints were only shown questions related to these variables when they indicated that their injury was due to trauma. The same is true for the effect of painkillers, which was included in both data sets. The effect of painkillers was only inquired if patients indicated that they had tried painkillers. In other words, these values were missing by design. Given that these variables are nominal and their values were systematically missing, their missing values were replaced with 0.

Finally, figure 8 shows three variables of which more than 1% of the values was missing and whose values were missing at random. Considering that these are only three variables and their fraction of missing values is small, the author decided to remove rows containing a missing value for one or more of these variables. Imputation was also considered for these variables, but deemed unnecessary and unpractical due to the categorical nature of the variables.

Figure 7: Bar chart of missing values in the knee dataset

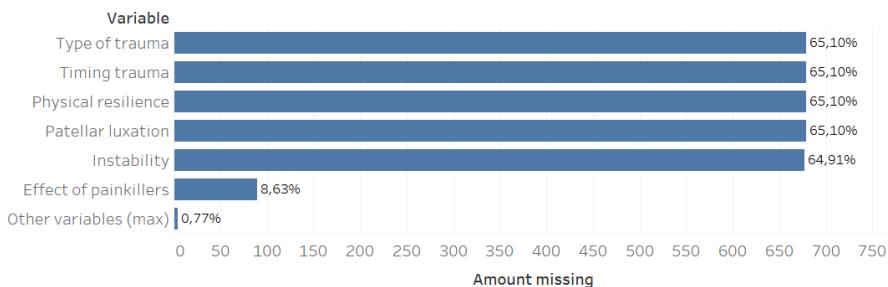
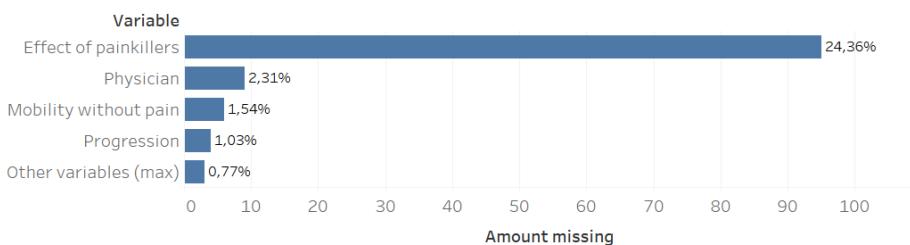


Figure 8: Bar chart of missing values in the hip dataset



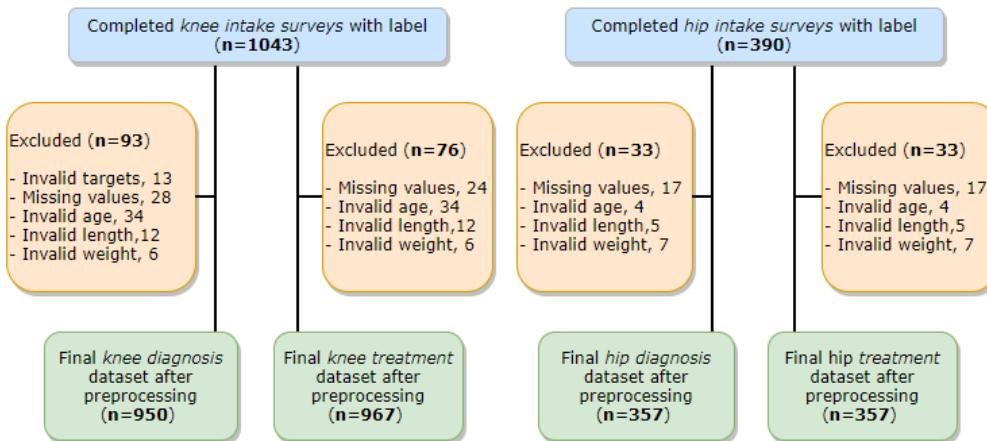
3.6 Pre-processing

Both the knee and hip dataset were pre-processed prior to model development. First, missing values in both data sets were handled according to the methods described in section 3.5. Second, the knee data set contained some invalid (i.e. typos and unrelated information such as dates) and missing diagnosis labels. The typos were fixed by hand and rows with missing labels or labels that could not be identified were removed. Typos were also present in numerical columns such as age and weight. Invalid values were identified by checking them against a logical range of possible values and removed from the data sets to prevent outliers. Non-numerical columns could not contain invalid or outlier values because their answers were a fixed multiple choice in the health history questionnaire.

After that, the targets were rearranged for each prediction problem according to the methods described in section 3.3. Finally, categorical, binary, and target columns were encoded to represent them numerically. Categorical columns in both data sets were one-hot-encoded. Columns in both data sets with values that could be represented in binary (i.e. yes/no, 1/2) were remapped to a binary representation (1/0). Lastly, the targets were encoded with a value between 0 and the number of classes –1.

The study flow from the raw data sets to the pre-processed data sets is shown in figure 9

Figure 9: Study flow from raw to pre-processed data



3.7 Model Selection

Based on previous research (??) and their specific advantages, the following algorithms were selected: (1) Logistic Regression, (2) the Support Vector Machine (SVM), (3) The k-nearest neighbour algorithm (k-NN), (4) the Decision Tree (DT), (5), the Random Forest (RF), (6) the Adaptive Boosting (AdaBoost) algorithm, (7) and the Gradient Boosting (GB) algorithm. More recent Deep Learning techniques such as Neural Networks were investigated in the early stages of the study as well. However, unsatisfactory performance (barely better than chance) compared to traditional Machine Learning techniques is why they were not included in the current research. Their unsatisfactory performance might be explained by the volume of the data sets, which is small for deep learning standards.

Explaining each of the selected algorithms in-depth is out of the scope of this paper, but a general description of them is provided below. First, logistic regression is a simple linear algorithm that is a derivative of linear regression, which attempts to fit the following linear equation ([Hastie, Tibshirani, and Friedman, 2009](#)):

$$\hat{y} = wx + b = \sum_{i=1}^K w_i * x_i + b \quad (1)$$

Where x are the input features, K is the number of classes, w are weights of the input features, b is a bias, and \hat{y} is the prediction of the function. b and w are the parameters that have to be learned. However, classification requires a prediction with a value between 0 and 1, which is why logistic regression forces the output of the linear function between 0 and 1 by incorporating the sigmoid function (formula 2). The parameters are optimized by finding the set of parameters that minimize the logistic loss error between predictions and the true regression targets on the training set. After learning the parameters, logistic regression classifies a new data point based on the decision boundary, which is customarily set at 0.5.

$$\hat{y} = \frac{1}{1 + \exp(-(wx + b))} = \frac{1}{1 + \exp(-(\sum_{i=1}^p w_i * x_i + b))} \quad (2)$$

The Support Vector Machine is another linear algorithm, which attempts to construct an N-dimensional (with N = the number of features) hyperplane that separates the classes by maximizing the margin between the classes ([Bishop, 2006](#)). A hyperplane is also called the decision boundary, which assigns data points to a class based on which side of the decision boundary they fall. Maximizing the margin with respect to the hyperplane is achieved by maximizing the distance of the hyperplane to the nearest data points that lie on either side of the hyperplane. These nearest points are also called the support vectors, where the algorithm gets its name.

In contrast to the linear algorithms before, the k-nearest neighbour algorithm is a simple non-linear algorithm that predicts the class of a new data point according to its distance (i.e. Euclidean distance) to neighbouring data points in the training set (Zhang, 2016). The k in k-NN denotes the number of nearest neighbouring data points taken into account when classifying a new data point. For example, if k is selected to be 6, the algorithm will classify a new data point according to the most prevalent class in the set of 6 data points that are closest to the new data point in terms of distance.

The decision tree is another relatively simple algorithm. It recursively splits the data on a single feature until all features have come to pass (Loh, 2011). The decision of the variable to split on is based on information gain, a function that measures how much entropy would be reduced by splitting on a particular attribute. The entropy of an attribute measures its unpredictability. While constructing the decision tree, the tree chooses the attribute that returns the highest information gain as the decision node. This process results in a tree-like structure with each feature represented as a decision node in the tree and leaf nodes that correspond to the class labels. A new data point is predicted by evaluating each feature's value at its respective decision node and taking the correct path until a leaf node is reached.

Random forests are built on top of the concept of a decision tree. Random forests consist of a large number of individual decision trees trained on random bootstraps of the data that together form a bagging ensemble (Breiman, 2001). In the case of classification, each tree in the random forest makes a class prediction, and the class with the most votes determines the prediction. A random forest is often seen as a variance (overfitting) reduction method because it averages predictions.

Similar to the random forest, AdaBoost is a method that combines multiple machine learning methods (weak learners) to create more powerful models (Freund and Schapire, 1997). AdaBoost builds an ensemble of weighted weak learners, often consisting of decision trees. AdaBoost trains base models successively, and each base model tries to correct the mistakes of the previous one by giving misclassified samples a higher weight value, thereby decreasing bias. Adaboost forces the next model to get the misclassified points accurate by passing on the weight to the loss or sampling important data multiple times. Adaboost predicts a new data point by a weighted vote over all learners incorporated in the model.

Gradient boosting is similar to Adaboosts in that it builds an ensemble of weighted weak learners by successively adding new learners that attempt to correct the mistakes of the previous one (Friedman, 2002). In contrast to AdaBoost, Gradient boosting starts by making an initial prediction by averaging the target values. After that, the pseudo-residual, the difference between the actual target value and the target value predicted by the learner, is computed for each sample. After that, each new learner is fit to predict the pseudo-residuals of the model before decreasing the pseudo-residuals until they do not decrease anymore. Gradient boosting predicts a new data point by summing the individual predictions, turning it into a probability by taking its logarithm and assign it to a class based on the probability threshold.

3.8 Quality Metrics

The data imbalance and binary and multi-class nature of the prediction problems need to be taken into account in selecting evaluation metrics. [Grandini, Bagli, and Visani \(2020\)](#) and [Döring \(2018\)](#) describe the following metrics that take multi-class imbalance into account while working for imbalanced binary problems as well.

Note that in the formulas, K denotes the number of classes, k denotes a class, n_k denotes the number of samples in class k , tp denotes the number of true positives, tn denotes the number of true negatives, fp denotes the number of false-positives, fn denotes false negatives, n_k denotes the number of samples in class k , and w_k denotes the weight according to the frequency of class k .

First, classification accuracy is the fraction of predictions that were predicted correctly. Balanced accuracy is an alternative to regular accuracy that takes class imbalance into account by taking the average of recall on each class, where each recall is weighted according to the frequency of the class in the complete dataset ([Grandini, Bagli, and Visani, 2020](#)). This is also called a weighted macro average. Its computation is presented by formula 3.

$$\text{Balanced Accuracy} = \frac{\sum_{k=1}^K \frac{tp_k}{n_k \cdot w_k}}{K} \quad (3)$$

Second, precision and recall can also be adapted to unbalanced multi-class problems by computing their weighted macro averages ([Grandini, Bagli, and Visani, 2020](#)), denoted by WMAP and WMAR, respectively. Their computations are as follows:

$$WMAP = \frac{\sum_{k=1}^K (Precision_k \cdot w_k)}{K} \quad (4)$$

$$WMAR = \frac{\sum_{k=1}^K (Recall_k \cdot w_k)}{K} \quad (5)$$

where

$$Precision_k = \frac{tp_k}{tp_k + fp_k} \quad (6)$$

$$Recall_k = \frac{tp_k}{tp_k + fn_k} \quad (7)$$

The weighted macro averaged precision and recall can also be used to compute the weighted macro averaged F1-score (WMA-F1) (Döring, 2018), which computes a harmonic mean between the two:

$$WMA - F1 = 2 * \left(\frac{WMAP \cdot WMAC}{WMAP + WMAR} \right) \quad (8)$$

Finally, the receiver operating characteristic (ROC) curve is a graphical plot that plots the true positive rate of a binary classifier against the false positive rate at different discrimination thresholds (Fawcett, 2006). The area under the resulting curve, also called the Area Under the Curve (AUC), can be used as a metric to evaluate the classifier. It can be extended to unbalanced multi-class problems by computing the average AUC of all possible pairwise combinations of classes and weighting each metric according to the frequency of the class in the complete dataset, which translates to the following formula (Hand and Till, 2001):

$$WMA - AUC = \frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j>i}^K p(i \cup j)(AUC(i|j) + AUC(j|i)) \quad (9)$$

where K denotes the number of classes, i denotes the positive class and j denotes the negative class.

Rather than a single metric, the choice was made to use all the previously discussed metrics to evaluate the quality of the classifiers and to see if consistent performance is achieved across metrics.

3.9 Modelling Process

First, the data sets were randomly split into a train (80%) and test set (20%), stratified with respect to the target distribution. Thereafter, each algorithm was evaluated on the four training data sets in a pipeline consisting of four steps: (1) optimization of a selection of additional data preprocessing steps and evaluation strategies, (2) selection of the optimal number of features in terms of performance, (3) tuning of the model hyperparameters, (4) evaluation on the training set in a repeated K-fold cross-validation loop. The steps were performed successively and each step included the results of the previous steps.

3.9.1 Optimization

A wide range of well-known data pre-processing techniques might improve model performance under certain conditions. Additionally, multi-class classification and imbalanced problems require specific techniques to be solved. A selection of these techniques was considered with the intention to improve the models' performance.

First, some of the previously described algorithms inherently support multi-class classification, such as K-Nearest Neighbours, logistic regression, the decision tree, and the random forest. k-NN supports multi-class by nature because it simply considers the k data points that are closest to the new data point in terms of some distance

measure, regardless of their class. The decision tree also handles multi-class by nature because each leaf node can refer to any included classes. This also means that the ensemble methods based on decision trees inherently support multi-class classification, namely the random forest, Adaboost, and gradient boosting. Logistic regression is easily adapted to multi-class problems by swapping out the sigmoid function for the softmax function.

For the remaining models, the multiclass classification problem needs to be deconstructed to a binary one. There are two well-known strategies for doing so: the one-vs-one strategy and the one-vs-all strategy (Bishop, 2006). The one-vs-rest strategy involves training a single classifier per class to predict that class against all other classes. The classifier that predicts the highest probability for its positive class determines the prediction. In contrast, the one-vs one strategy involves training a classifier for each pair of classes. A prediction is made by summing the output classification probabilities of each classifier and selecting the class with the highest sum.

Similarly to multi-class classification problems, some of previously described algorithms inherently take imbalanced data into account, while others have to take this into account explicitly. Boosting techniques like AdaBoost and gradient boosting naturally take unbalanced data into account by assigning more weight to misclassified samples (which are most likely to be those in the minority classes) in each iteration. This is an example of cost-sensitive learning, where instead of deeming all misclassification errors equally severe, some misclassifications, for example in the minority class, are given a higher cost (He and Ma, 2013). All other models except for k-NN allow for a form of cost-sensitive learning where classes can be given a weight. Higher class weights result in higher penalties in the cost function and thus putting more emphasis on these classes (King and Zeng, 2001). A commonly applied scheme is to adjust weights inversely proportional to class frequencies automatically.

Second, features are often represented by different scales. Differing scales can significantly influence the performance of distance-based algorithms such as k-NN, and the SVM (Roy, 2020). The performance of these algorithms can improve by representing all features on the same scale. Two methods that can transform all features to a similar scale are standardization and normalization. Normalization can be performed in several ways, but the min-max scaler is appropriate in this case because it responds well to non-Gaussian data (Roy, 2020). The data sets were tested for normality by performing the Shapiro-Wilk test (Shapiro and Wilk, 1965), but for most variables, the hypothesis was rejected. The min-max scaler transforms each feature by shrinking its values to a range between zero and one by applying the following formula:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (10)$$

Where x denotes the feature value, x_{min} denotes the minimal feature value, and x_{max} denotes the maximal feature value.

In contrast, standardisation is performed by removing the mean (μ) of the feature from the feature values (x) and then scaling them according to the standard deviation (σ) of the feature (formula 11). Standardisation does not restrict the values to a particular

range but instead centres them to have zero-mean and a standard deviation of 1. However, this method works best for normally distributed data.

$$x_{standardized} = \frac{x - \mu}{\sigma} \quad (11)$$

Third, the dimensionality of the data may also influence model performance. A high amount of features is linked to the so-called curse of dimensionality ([Oseledets and Tyrtysnikov, 2009](#)). The curse of dimensionality means that as the number of feature dimensions grows, the amount of data that needs to be generalised grows exponentially. As a result, exponentially more data is needed for every feature that is added to avoid sparseness. Again, this affects any distance-based algorithm. The curse of dimensionality can be countered by selecting features or using dimensionality reduction techniques such as Principal Component Analysis (PCA) ([Abdi and Williams, 2010](#)). PCA projects data to a lower-dimensional space through Singular Value Decomposition, which decomposes the dataset into principal components that preserve as much of the variance of the original dataset as possible.

Meanwhile, increasing the dimensionality of the data might also yield positive results. Specifically, interactions between features often occur and explicitly engineering new features that capture these interactions might improve model performance ([Brownlee, 2020](#)). Engineering new features by taking the polynomials of the original features is one way to capture these interactions. Polynomial feature engineering involves raising features to an exponent and multiplying features with each other to a certain degree. PCA can be applied to decrease the number of features to deter the curse of dimensionality.

Returning to the problem of class imbalance: several sampling techniques exist to balance the classes to an equal number of samples. Undersampling strategies decrease the number of samples in the majority class to match the number of samples in the minority class. In contrast, oversampling strategies increase the number of samples in the minority class to match the number of samples in the majority class.

Simple sampling strategies are random oversampling and random undersampling. Random oversampling randomly duplicates instances from the minority class until the number of instances in the minority class is equal to the number of instances in the majority class. Random undersampling randomly removes instances from the majority class until the number of instances in the majority class is equal to the number of instances in the minority class.

More clever sampling techniques exist as well. A popular oversampling method is Synthetic Minority Over-sampling Technique (SMOTE) ([Bowyer et al., 2011](#)). SMOTE generates synthetic samples by finding minority samples close to each other in the feature space and generating new samples along the lines that connect these minority samples. In contrast, Tomek Links removes data points from the majority class rather than generating new data points ([He and Ma, 2013](#)). Tomek links are pairs of data points with opposing labels that are each other's closest neighbour. These data points tend to overlap with the decision boundary, introducing class separation difficulties. The algorithm removes the majority instances from the Tomek Links to increase the separation between the classes. A hybrid between SMOTE and Tomek Links exists as well. First, oversampling is performed with SMOTE, and the resulting data is cleaned using Tomek Links.

To investigate which of the described techniques positively influence model performance, the first step of the pipeline was a grid search in which every unique combination of techniques was applied to the data and evaluated in a 3-fold stratified cross-validation loop with balanced accuracy as the metric to maximise. This process was performed for each data set and model independently. The unique combination of techniques that maximised performance was returned by the grid search and applied to the data in all subsequent steps of the pipeline. A summary of the techniques that were tested is presented in table 6.

Table 6: Options for data pre-processing, multi-class classification, and imbalanced learning techniques that were evaluated in a grid search

Technique	Goal	How
One-vs-Rest multiclass strategy	Deconstructing the multiclass problems into a binary ones	Train single classifier per class to predict that class against all other classes
One-vs-One multiclass strategy	Deconstructing the multiclass problems into binary ones	Train classifier for each pair of classes
Cost-sensitive learning	Taking class imbalance into account	Adjust weights inversely proportional to class frequencies
Feature Normalization	Scaling the features to the same range and improve the performance of distance-based algorithms	See equation 10
Feature Standardization	Scaling the features to the same range and improve the performance of distance-based algorithms	See equation 11
PCA Dimensionality Reduction	Reducing the dimensions of the data to prevent the curse of dimensionality	Project data to a lower-dimensional space through Singular Value Decomposition
Polynomial Features	Engineering new features and capture their interactions	Raise features to an exponent and multiply features with each other
Random Undersampling	Balancing the classes to have an equal number of samples	Randomly duplicate instances from the minority class to match the number of instances in the majority class
Random Oversampling	Balancing the classes to have an equal number of samples	Randomly remove instances from the majority class to match with the number of instances in the minority class
SMOTE Oversampling	Balancing the classes to have an equal number of samples	Find minority samples close to each other and generate new samples between these minority samples
Tomek Links Undersampling	Balancing the classes to have an equal number of samples	Remove the majority instances from the Tomek Links
SMOTE + Tomek Links	Balancing the classes to have an equal number of samples	Perform SMOTE and Tomek Links successively

3.9.2 Feature selection

After optimising the data preprocessing techniques, the next step of the pipeline was feature selection to select the most valuable features and limit the data's dimensionality. There is a multitude of feature selection methods available, each with its limitations. Therefore, the choice was made to employ several feature selection methods simultaneously and combine their results to average out their limitations.

Two algorithms have inherent methods to determine feature importance. First, the coefficients of logistic regression can be used as a measure of feature importance. The coefficients are obtained by fitting a logistic regression to the training data. Second, a random forest can be fit to the training data. The trained random forest can assign feature importances by averaging how much each feature decreases the impurity or entropy in every tree in the forest.

A different method that uses algorithms with inherent methods to select features is Recursive Feature Elimination (RFE). Rather than assigning feature importances by training a single model on all training data, RFE recursively considers increasingly smaller sets of features (Kuhn, Johnson et al., 2013). RFE starts by fitting a model on all features. The model then ranks the features, eliminating those from the data that it considers least important. This process repeats until a predetermined number of features remain. For the purpose of this study, the random forest was selected as the base estimator. Furthermore, rather than stopping at a predetermined number of features, the choice was made not to terminate the algorithm until all features were removed. This decision allowed for extracting the rankings of the complete feature sets.

Finally, a univariate feature selection method was applied to compute feature importances. Univariate feature selection methods compute feature importances based on univariate statistical tests that compute the strength of the relation between features and the target variable (Jović, Brkić, and Bogunović, 2015). Several tests are available, but Mutual Information (MI) was picked because it can handle both numerical and categorical features well (Ross, 2014).

The four methods described above were combined to select features as follows. First, each feature selection method was repeated ten times to account for variance in their results. After that, their results were averaged to obtain a single set of feature importances per method. Taking into account that each method computes feature importances on different scales, each set of feature importances was normalised according to formula 10, where x denotes the feature importance score, x_{min} denotes the minimal feature importance score, and x_{max} denotes the maximal feature importance score. Thereafter, the normalised feature importance vectors were combined by averaging them, resulting in a single vector of feature importances. Finally, the optimal number of features for a particular model was determined by training and evaluating the model several times on an increasing number of features (sorted by importance) in a 3-fold stratified cross-validation loop with balanced accuracy as the metric to maximise. The first N features that resulted in the highest score were selected for the next step.

3.9.3 Hyperparameter tuning

Machine learning algorithms customarily have a set of hyperparameters that can be adjusted to change their behaviour and improve their performance. They can either be adjusted manually or tuned with an optimisation algorithm. Manually selecting the hyperparameters is a cumbersome task, so the hyperparameters for each model were optimised algorithmically with a grid search.

Grid search exhaustively evaluates all hyperparameter combinations from a predetermined grid of hyperparameter values in a stratified k-fold cross-validation loop. It returns the set of hyperparameter values that resulted in the best performance in terms of the chosen performance metric. In this case, balanced accuracy was picked as the performance metric to maximise. The sets of hyperparameters to tune and their values to evaluate were based on the documentation of the algorithms. The hyperparameter values of each algorithm were automatically set to the optimal values returned by the grid search.

3.9.4 Evaluation

After optimising data pre-processing techniques, evaluation strategies, selection of features, and hyperparameter values for each algorithm, their performance on the training data was evaluated. The evaluation was performed in a 5-fold stratified cross-validation loop that was repeated 100 times. Considering the small size of the data sets, five folds ensured sufficient data for both training and validation. Moreover, cross-validation was repeated 100 times to compute confidence intervals with a confidence level of 95%. The metric scores were computed by averaging the results from all cross-validation repetitions.

After that, the model that performed best across all metrics for each prediction problem was selected for further evaluation on the test sets. Then, these algorithms were re-trained on the complete training data sets and evaluated again according to the performance metrics. Finally, these algorithms were incorporated into a prediction tool that was used for providing information to patients in the second phase of the study.

3.10 Software

The entire pipeline was written in python ([Van Rossum and Drake, 2009](#)). The algorithms, metrics, grid-search, multi-class strategies, feature selection methods, and most of the options for pre-processing were implemented with off-the-shelf code from the scikit-learn package ([Pedregosa et al., 2011](#)). The sampling methods were implemented with off-the-shelf code from the imbalanced-learn package ([Lemaître, Nogueira, and Aridas, 2017](#)). The code that brought it all together and the code for the prediction tool was written by myself. Finally, Tableau ([Tableau, 2021](#)) was used for most of the visualisations.

4 Results: Model Development

The knee dataset and hip dataset included records of 967 and 357 patients, respectively. Important demographic characteristics of patients with complaints related to the knee are shown in tables 7a and 8. Important demographic characteristics of patients with complaints related to the hip are shown in tables 7b and 9. Patients with knee complaints were 55 years old on average, whereas patients with hip complaints were 64 years old on average. Both groups had a nearly equal distribution between complaints to the left and the right extremities, with a small percentage suffering from complaints on both sides. Most patients in both groups were experiencing complaints for at least a few months. Finally, most patients in both groups indicated that their complaints varied in severity or deteriorated.

Table 7: Numeric demographics

Variable	Mean	Median	σ	Variable	Mean	Median	σ
Age	55	57	15	Age	64	66	13
Length	174	174	10	Length	171	170	9
Weight	84	83	17	Weight	79	78	15
Pain in rest	4	4	3	Pain in rest	5	5	3
Pain while moving	7	7	2	Pain while moving	7	7	2

(a) Knee patients

(b) Hip patients

Table 8: Nominal demographics knee patients

Variable	Category	Count	Percentage
Side of complaints	Both, left more than right	82	8.5%
	Both, right more than left	103	10.7%
	Left	346	36.0%
	Right	430	44.7%
Duration of complaints	A few days	41	4.3%
	A few months	361	37.6%
	A few weeks	201	20.9%
	A few years	358	37.3%

Table 8 continued from previous page

Variable	Category	Count	Percentage
Progression of complaints	Deteriorated	501	52.1%
	Improved	52	5.4%
	Remained the same	109	11.3%
	Varying in severity	299	31.1%
Complaints due to trauma	No	631	65.7%
	Yes	330	34.3%

Table 9: Nominal demographics hip patients

Variable	Category	Count	Percentage
Side of complaints	Both left and right	31	8.7%
	Left	161	45.2%
	Right	164	46.1%
Duration of complaints	A few days	3	0.8%
	A few months	190	53.4%
	A few weeks	35	9.8%
	A few years	128	36.0%
Progression of complaints	Deteriorated	223	62.6%
	Improved	4	1.1%
	Unchanged	29	8.1%
	Varying in severity	100	28.1%

4.1 Validation Performance

The validation performance of each model that was evaluated on the knee diagnosis, hip diagnosis, knee treatment, and hip treatment training sets are shown in tables 10, 11, 12, and 13 respectively. The highest performances metric-wise are marked with an asterisk (*). Models that performed best across metrics are marked with a double asterisk (**). The ranges behind each score represent the confidence intervals with a confidence level of 95%. The pre-processing steps, evaluation strategies, features, and model hyperparameters selected in the pipeline are not reported to prevent an overkill of information. However, they are reported for the final selection of models in the next section.

In general, knee diagnosis prediction models fared better (max 0.6 accuracy, 0.67 F1) than their hip counterparts (max 0.52 accuracy, 0.59 F1). Unfortunately, while better than the poor performance of the hip diagnosis prediction models, the knee diagnosis models failed to exceed moderate performance. Overall performance within each group of models is comparable, with most scores falling within a .1 range of each other across metrics. Hence, it was not possible to determine a superior algorithm for either problem. Based purely on raw comparisons of scores across metrics, AdaBoost was the best algorithm for predicting knee diagnoses, and the random forest was the best algorithm for predicting hip diagnoses.

Table 10: Validation performance of knee diagnosis prediction models

	Accuracy	AUC	Recall	Precision	F1
k-NN	0.5 (0.49- 0.51)	0.71 (0.71 - 0.72)	0.56 (0.55 - 0.57)	0.62 (0.61 - 0.63)	0.56 (0.55 - 0.57)
DT	0.53 (0.53 - 0.54)	0.71 (0.71 - 0.72)	0.62 (0.61 - 0.62)	0.62 (0.612 - 0.63)	0.61 (0.61 - 0.62)
RF	0.6 (0.59 - 0.6) *	0.79 (0.78 - 0.79) *	0.63 (0.62 - 0.63)	0.7 (0.7 - 0.7) *	0.64 (0.64 - 0.65)
AdaBoost**	0.6 (0.59 - 0.6) *	0.79 (0.78 - 0.79) *	0.68 (0.67 - 0.68) *	0.67 (0.67 - 0.67)	0.67 (0.67 - 0.67) *
SVM	0.56 (0.56 - 0.57)	0.79 (0.78 - 0.79) *	0.61 (0.61 - 0.61)	0.69 (0.68 - 0.7)	0.61 (0.61 - 0.62)
GB	0.59 (0.58 - 0.59)	0.78 (0.77 - 0.78)	0.66 (0.65 - 0.66)	0.66 (0.66 - 0.67)	0.66 (0.65 - 0.66)

Table 11: Validation performance of hip diagnosis prediction models

	Accuracy	AUC	Recall	Precision	F1
k-NN	0.46 (0.45 - 0.47)	0.65 (0.64 - 0.66)	0.53 (0.52 - 0.54)	0.54 (0.52 - 0.56)	0.49 (0.49 - 0.5)
DT	0.42 (0.41 - 0.43)	0.64 (0.63 - 0.65)	0.56 (0.56 - 0.57)	0.53 (0.52 - 0.54)	0.53 (0.52 - 0.53)
RF **	0.52 (0.51 - 0.53) *	0.72 (0.71 - 0.72) *	0.58 (0.57 - 0.59)	0.63 (0.62 - 0.63) *	0.59 (0.58 - 0.6) *
AdaBoost	0.46 (0.45- 0.47)	0.66 (0.65 - 0.67)	0.59 (0.58 - 0.6)	0.59 (0.58 - 0.6)	0.56 (0.55 - 0.56)
SVM	0.48 (0.48 - 0.49)	0.68 (0.67 - 0.69)	0.51 (0.5 - 0.53)	0.56 (0.54 - 0.57)	0.49 (0.48 - 0.49)
GB	0.43 (0.42 - 0.44)	0.68 (0.67 - 0.69)	0.62 (0.61 - 0.63) *	0.58 (0.57 - 0.59)	0.57 (0.56 - 0.58)

The performance of knee and hip treatment prediction models (max 0.66 accuracy, 0.78 F1 and 0.71 accuracy, 0.72 F1 respectively) surpass the performance of the diagnosis prediction models. Moreover, in contrast to the diagnoses prediction models, both groups of models have similar performance. Additionally, the margins of within-group metric scores are relatively substantial. Based on performance across metrics, both the Random Forest and Support Vector Machine were best for predicting knee treatments, although the Random Forest performed slightly better in terms of F1 score. The same models were best for the prediction of hip treatments. Although the Support Vector Machine had a slight advantage across metrics, the difference was negligible.

Table 12: Validation performance of knee treatment prediction models

	Accuracy	AUC	Recall	Precision	F1
k-NN	0.54 (0.54 - 0.55)	0.54 (0.54 - 0.55)	0.8 (0.8 - 0.81) *	0.75 (0.74 - 0.76)	0.77 (0.76 - 0.77)
DT	0.58 (0.58 - 0.59)	0.58 (0.58 - 0.59)	0.72 (0.72 - 0.73)	0.75 (0.75 - 0.76)	0.74 (0.73 - 0.74)
RF **	0.66 (0.65 - 0.67) *	0.66 (0.66 - 0.67) *	0.75 (0.74 - 0.75)	0.79 (0.79 - 0.8) *	0.77 (0.76 - 0.77)
AdaBoost	0.58 (0.57 - 0.58)	0.58 (0.57 - 0.58)	0.8 (0.8 - 0.8) *	0.77 (0.77 - 0.77)	0.78 (0.77 - 0.779) *
SVM **	0.66 (0.65 - 0.67) *	0.66 (0.65 - 0.66) *	0.69 (0.69 - 0.7)	0.79 (0.79 - 0.79) *	0.73 (0.72 - 0.73)
GB	0.6 (0.59 - 0.6)	0.6 (0.59 - 0.6)	0.79 (0.79 - 0.79)	0.77 (0.77 - 0.77)	0.78 (0.78 - 0.78) *

Table 13: Validation performance of hip treatment prediction models

	Accuracy	AUC	Recall	Precision	F1
k-NN	0.66 (0.65 - 0.67)	0.66 (0.65 - 0.67)	0.69 (0.68 - 0.7)	0.7 (0.69 - 0.71)	0.68 (0.67 - 0.7)
DT	0.52 (0.51 - 0.53)	0.52 (0.51 - 0.53)	0.65 (0.65 - 0.66)	0.51 (0.48 - 0.54)	0.54 (0.53 - 0.55)
RF	0.7 (0.69 - 0.71)	0.7 (0.69 - 0.71)	0.7 (0.69 - 0.71)	0.72 (0.71 - 0.73)	0.7 (0.69 - 0.71)
AdaBoost	0.65 (0.64 - 0.66)	0.65 (0.65 - 0.66)	0.7 (0.69 - 0.71)	0.69 (0.69 - 0.7)	0.69 (0.68 - 0.7)
SVM **	0.71 (0.7 - 0.72) *	0.71 (0.7 - 0.72) *	0.71 (0.7 - 0.72) *	0.73 (0.72 - 0.74) *	0.72 (0.71 - 0.73) *
GB	0.66 (0.65 - 0.67)	0.66 (0.65 - 0.67)	0.7 (0.69 - 0.71)	0.69 (0.68 - 0.7)	0.69 (0.68 - 0.7)

Based on the validation results, four models were selected for training on the complete training sets, one for each prediction problem. First, Adaboost was selected for the prediction of knee diagnoses. Next, the Random Forest was selected for the prediction of hip diagnoses and knee treatments. Finally, the Support Vector Machine was selected for the prediction of hip treatments. Then, the pre-processing steps, evaluation strategies, features, and model hyperparameters for each of the four models were optimized again according to the first three steps of the pipeline described in section 3.9. After that, the models were trained on the complete training sets and evaluated on the test sets.

4.2 Test Performance

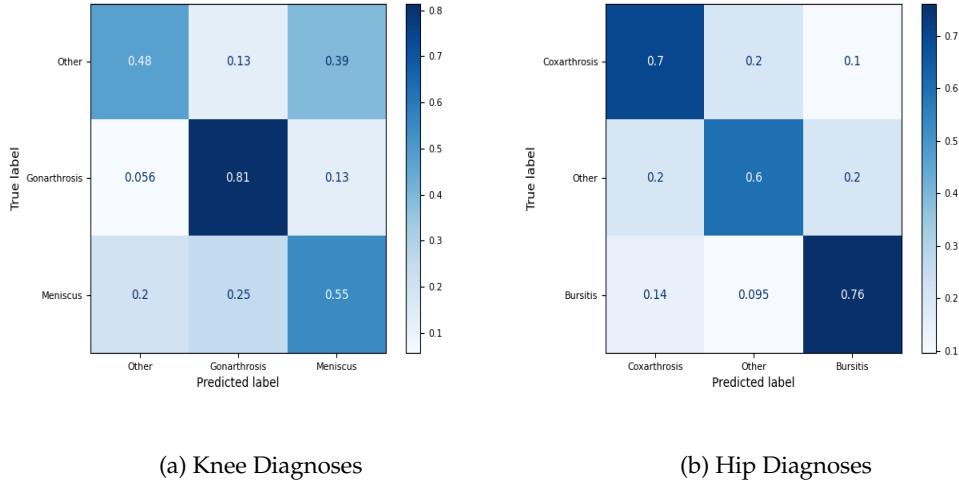
Table 14 shows the test performance of the models that were selected for each prediction problem. The hyperparameters that were selected for each model are presented in appendix E. Except for the knee diagnosis prediction model, the test performance of the prediction models was better than their validation performance. The knee diagnosis prediction model remained poor (0.61 accuracy, 0.69 F1). In contrast, the performance of the hip diagnosis prediction model significantly improved (0.69 accuracy, 0.73 F1). Whereas the knee treatment prediction model only slightly improved (0.7 accuracy, 0.77 F1), the performance of the hip treatment prediction model improved significantly (0.84 accuracy, 0.86 F1).

Table 14: Test performance of selected model for each prediction problem

	Accuracy	AUC	Recall	Precision	F1
Knee Diagnosis - AdaBoost	0.61	0.82	0.68	0.7	0.69
Hip Diagnosis - RF	0.69	0.8	0.72	0.75	0.73
Knee Treatment - RF	0.7	0.7	0.74	0.82	0.77
Hip Treatment - SVM	0.84	0.84	0.86	0.86	0.86

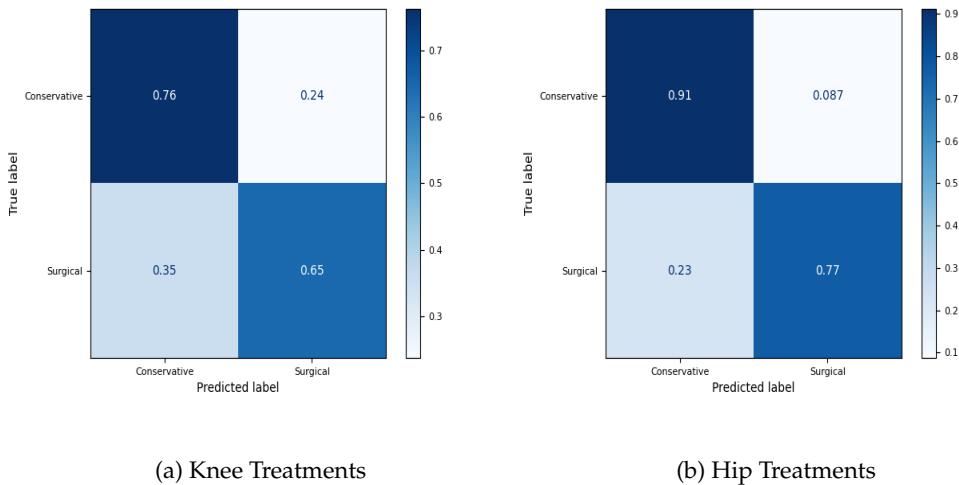
The difference in performance between the knee diagnosis prediction model and the hip diagnosis prediction model is also reflected by their confusion matrices (figure 10). The number of predictions for each diagnosis was normalized according to its relative frequency in the test set. The diagonal of the hip diagnosis confusion matrix shows that the model could predict each of the three diagnoses somewhat accurately. Bursitis was predicted correctly most often, while the aggregate "other" class was predicted correctly the least. In contrast, the knee diagnosis confusion matrix shows a more scattered pattern. While gonarthrosis was predicted correctly considerably often, degenerative meniscus tear and the aggregate "other" class were misclassified around half the time. In addition, the aggregate "other" class tended to be misclassified as a degenerative meniscus tear, whereas degenerative meniscus tear was approximately equally often misclassified as either "other" or gonarthrosis.

Figure 10: Normalized Diagnosis Confusion Matrices



Similar to the diagnosis prediction models, the normalized confusion matrices of the treatment prediction models (figure 11) reflect their difference in performance. The hip treatment prediction model is superior in classifying both conservative and surgical treatments correctly. However, the difference is less evident since both matrices show similar behaviour for both models. Both models could predict conservative treatments most accurately and surgical treatments least accurately.

Figure 11: Normalized Treatment Confusion Matrices



4.2.1 Selected Data Transformations and Evaluation Strategies

Table 15 shows the additional data transformations and evaluation strategies that were selected in the first step of the pipeline. Feature normalization was preferred for the diagnosis prediction models, while feature standardization was preferred for treatment prediction models. Neither reducing the dimensions of the data with Principal Component Analysis nor increasing the dimensions of the data by applying polynomial transformations to the data increased performance. Random oversampling was selected as the superior sampling technique for every prediction problem. Furthermore, the optimal sampling scheme was oversampling the minority class(es) to match the majority class in every case. Finally, the One-vs-Rest multi-class strategy was selected over the One-vs-One multi-class strategy for each prediction problem.

Table 15: Selected data transformations and evaluation strategies for each prediction problem

	KD	HD	KT	HT
Feature Normalization	✓	✓	-	-
Feature Standardization	-	-	✓	✓
PCA Dimensionality Reduction	-	-	-	-
Polynomial Features	-	-	-	-
Random Undersampling	-	-	-	-
Random Oversampling	✓	✓	✓	✓
SMOTE Oversampling	-	-	-	-
Tomek Links Undersampling	-	-	-	-
SMOTE + Tomek Links	-	-	-	-
One-vs-Rest multiclass strategy	✓	✓	✓	✓
One-vs-One multiclass strategy	-	-	-	-

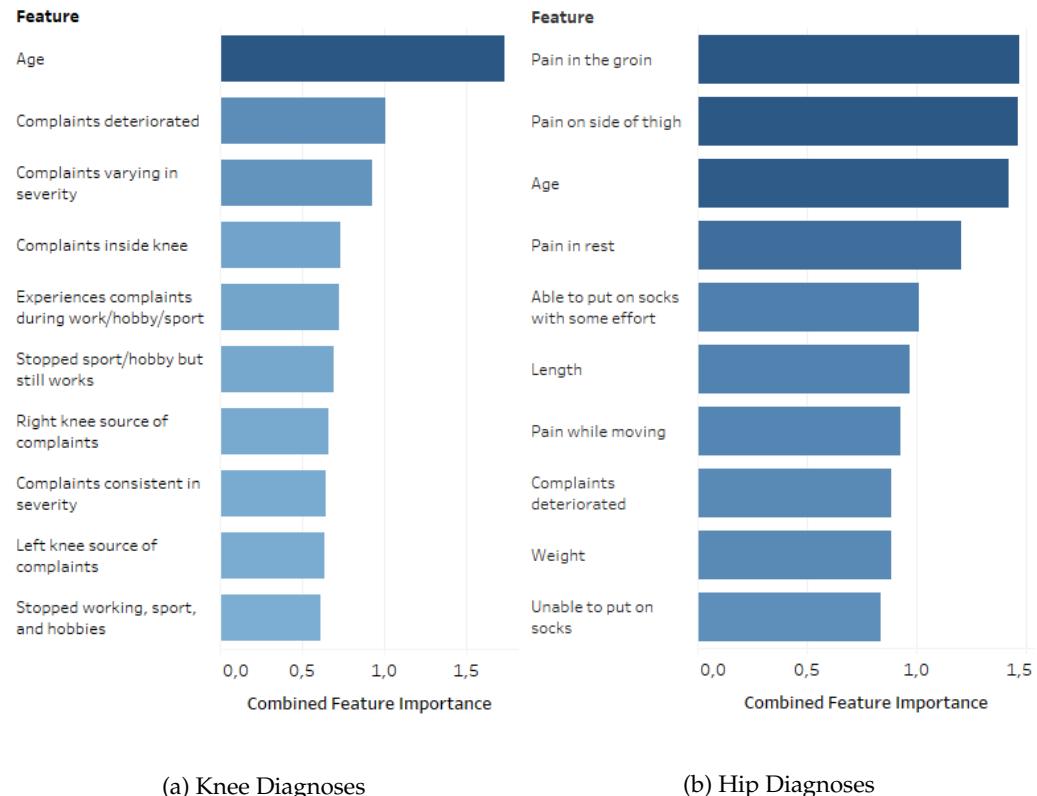
4.2.2 Feature Selection and Feature Importance

Feature selection by selecting a subset of features did not result in increased performance. Each model seemed to benefit from the inclusion of all features rather than including only a subset of features deemed important by the feature selection methods. Nevertheless, the addition of features in decreasing importance did expose diminishing returns, showing that some features add more information to the models than others.

The feature importance scores of the top 10 most important features for both diagnosis prediction models are shown in figure 12. Please note that the feature importance scores are aggregated and arbitrary by nature; they merely show the importance of features relative to each other. Unfortunately, this also means that direct comparison of scores between models is not possible.

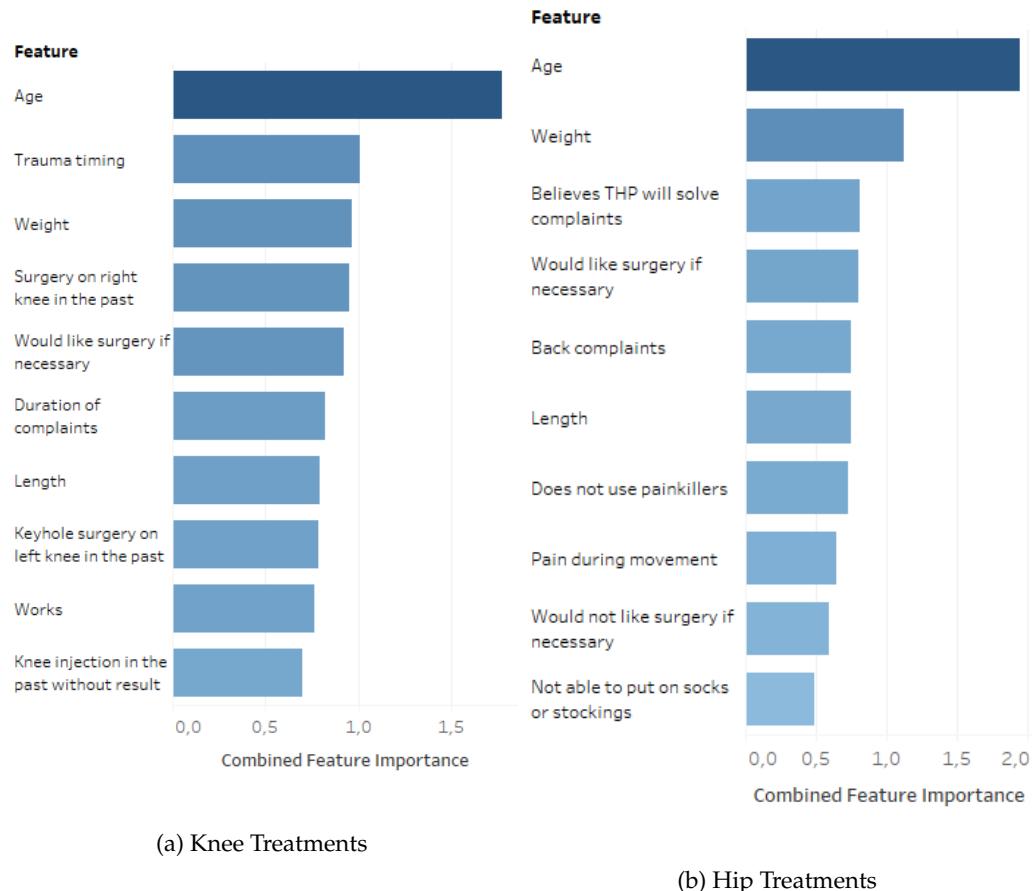
Age and deteriorating complaints seemed to be important features for both models. Both models also considered the location of the complaints to be important. The difference between the two models is that the knee diagnosis prediction model considers features related to the patient's work/hobby/sport to be of importance. In contrast, the hip diagnosis model emphasized the pain that the patient experiences, his/her ability to put on socks and demographic information such as weight and length.

Figure 12: Combined feature importance scores diagnoses for top 10 most important features



The feature importance scores for the top 10 most important features for both treatment prediction models are shown in figure 13. An interesting difference was observed between the features considered to be important for predicting diagnoses and treatments. Whereas complaint-related features primarily defined the selection of most important features for predicting diagnoses, treatment-related features primarily defined the selection of most important features for predicting treatments. Examples include whether patients have had surgeries or other treatments in the past, if so, which ones, and if they would like surgery if necessary. In contrast, some features that are considered important for the prediction of diagnoses were also considered important for the prediction of treatments, such as the duration of complaints and whether the patient can put on socks or stockings. Finally, all demographic features (age/weight/length) were considered to be important.

Figure 13: Combined feature importance scores treatments



4.3 Prediction Tool

The final selection of models was included in an offline prediction tool that was developed to provide information to patients regarding their predicted diagnosis and

treatments in the next section of the study. A screenshot of the tool is provided in figure 14.

The text box at the bottom of the interface provides feedback for every step that the user performs. First, the user can load the dataset with the health history questionnaire of the patient by clicking on the "browse" button. Second, the patient's data is filtered from the dataset and pre-processed according to the data transformations selected for the model in question by entering the patient identifier and clicking on the "pre-process" button. There are separate buttons for predicting the diagnosis and the treatment, given that the required transformations are different for both models. After pre-processing, the user can predict the diagnosis and treatment of the patient by clicking the "predict diagnosis" and "predict treatment" buttons. The prediction appears in the text box at the bottom. Furthermore, the predicted probabilities for each class are shown to indicate the certainty of the model.

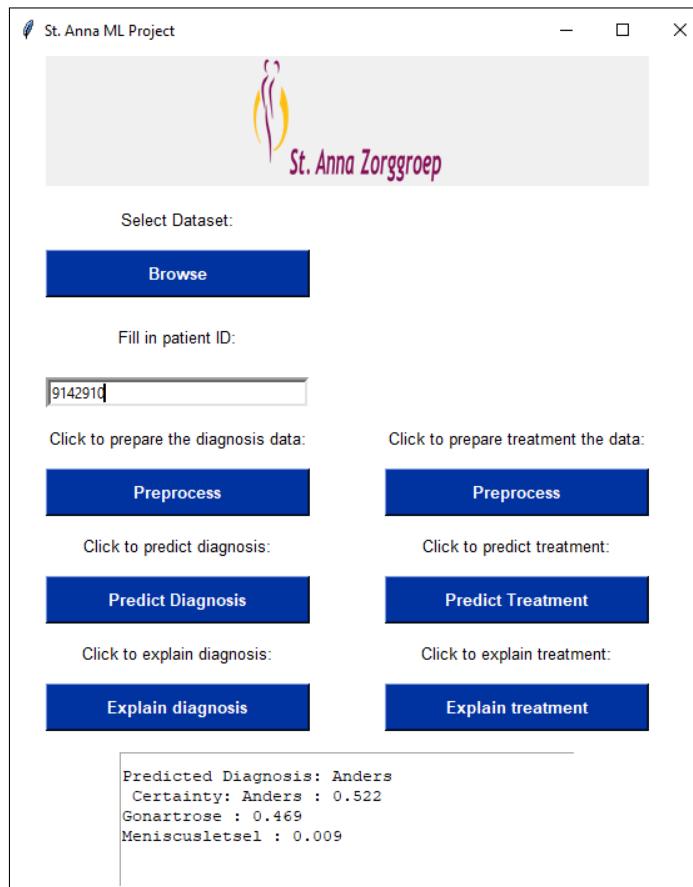


Figure 14: Diagnosis and Treatment Prediction Tool

5 Discussion: Model Development

The aim of the current research phase was to develop predictive models that can accurately predict orthopaedic diagnoses and treatments of the knee and hip by applying machine learning techniques to health history questionnaire data. To the knowledge of the author, the current study is the first to attempt to develop predictive models for the prediction of orthopaedic diagnoses and treatment pre-consultation. Moreover, it is, to the knowledge of the author, the first study to develop predictive models on health history questionnaire data. According to the results, predictive models of diagnoses and treatments of the hip can achieve moderate performance on a limited amount of health history questionnaire data.

Even though more health history questionnaire data was available for patients with knee complaints, predictive models of diagnoses and treatments of the knee underperformed compared to their hip counterparts. This especially true for the knee diagnosis prediction model, which was the worst model in terms of performance. However, it can be argued that the knee diagnosis prediction task was inherently more challenging. First, there are more knee diagnoses than hip diagnoses aggregated into the "other" class (35 and 17, respectively). Hence, there was likely to be more variance in the aggregated knee diagnosis class than in the aggregated hip diagnosis class. Second, the two explicitly included knee diagnoses, gonarthrosis and degenerative meniscus tear are closely related. Degenerative meniscus tear is commonly regarded as a pre-stage of gonarthrosis, and their pathologies tend to be very similar. In contrast, the two explicitly included hip diagnoses are relatively unrelated and might have been easier to discriminate.

In summary, the current study has revealed that Machine Learning can achieve somewhat promising results for predicting orthopaedic diagnoses and treatments on a limited amount of health history questionnaire data.

5.1 Clinical Implications

First and foremost, the models allow care providers to provide information to patients about the diagnosis and treatment that the model predicts. Supported by the implemented prediction tool, this information could be provided to patients employing an eHealth tool such as a mobile application, which would make the information easily accessible to a significant portion of the patient population.

Although the developed models are moderate at best, patients may accept a certain margin of error in terms of receiving wrong information. Even if the information is not always entirely correct, encouraging patients to think about their consultation beforehand and teaching them about orthopaedics, in general, might be valuable additions by themselves.

Second, they models could enable the grouping of patients according to their predicting treatment category. The grouping of patients according to their predicted treatment category would allow the outpatient clinic to adapt the consultations of each group. For example, consultations of patients predicted to receive a conservative treatment could be joined by other specialised healthcare providers such as physical therapists. Moreover, the number of consultations of the group of patients predicted to receive a surgical treatment could be reduced by directly scheduling them for preoperative screening. Unfortunately, misclassifications are more costly in these scenarios, so applying the models for these purposes is unfeasible at this time.

5.2 Limitations

Several limitations of the current study were identified by the author and a researcher of St. Anna. First, while the amount of data collected for each problem is considered substantial in healthcare, it is considered somewhat limited for the application of Machine Learning. This limitation might contribute to the moderate performance of the developed models. The limited number of samples in each data set also contributes to the second limitation; the imbalance in diagnoses and treatments. Although naturally, some diagnoses and treatments are more prevalent than others, it is detrimental to have sufficient samples in each class to develop usable models. Without sufficient samples, not a single model can separate minority classes from classes with plentiful records.

With the top few classes representing a substantial share of each data set, there were no other choices than removing all the samples of infrequently occurring diagnoses and treatments or aggregating all these diagnoses and treatments into a single class. Of course, removing all samples from infrequently occurring classes would have led to even smaller data sets. What is worse, patients with a different diagnosis or treatment than those included in the models would be predicted wrong by default. Including aggregation classes at least offers the opportunity to isolate patients with infrequently occurring diagnoses and treatments.

Moreover, the questionnaire questions themselves are also biased towards frequently occurring diagnoses and treatments. For example, many questions in both data sets are directly related to osteoarthritis, whereas few or none of the questions are directly related to infrequently occurring diagnoses. Biassing the questions by favouring frequently occurring diagnoses makes sense from the point of view of a physician, who mainly uses the history taking questionnaire to separate the wheat from the chaff quickly. In contrast, an ML algorithm requires that the data represents each class.

Finally, the current study did not suggest an algorithm that is superior in terms of performance within the set of evaluated algorithms. Performance differences between algorithms were small, and the selection of algorithms for production based on a comparison between raw metric scores might not have been entirely appropriate. In other words, the selected algorithms for production might not be optimal in a different setting.

5.3 Future Work

The current study points to several exciting directions that future research could take. First, the study showed that health history questionnaire data has discriminative value. However, due to the limited amount of data available for this study, it is unclear how much collecting more data could improve model performance. Did the current study already hit the ceiling due to the nature of the data, or was there simply not enough data for the development of excellent models? Apart from collecting more data overall, it would be particularly interesting to collect more data on the minority classes and evaluate if an improved balance between the classes improves model performance. It would match with the observation that the models could discriminate majority classes from minority classes relatively well.

Collecting more data for minority classes has an additional benefit. It would allow for the inclusion of diagnoses and treatments that do not frequently occur by nature. Even though merely predicting the frequently occurring diagnoses correctly covers a substantial share of the patient population, the ability to pick out the rare cases has the additional benefit of enabling truly personalised information to patients. The same benefit is achieved with the ability to predict specific treatments compared to the current dichotomy of conservative versus surgical treatments.

If the previous suggestions do not lead to significant improvements, one or more of the following options is worth considering. First, it could be interesting to change or add questions to the questionnaires, especially more specific questions about infrequently occurring diagnoses and treatments. Additionally, the formulation of the questions could be changed. The current nature of the questionnaires is such that it consists predominantly of categorical questions. Interestingly, the feature selection techniques classified numeric features such as weight and length as important for almost every prediction problem. In contrast, these are generally not considered import predictors by orthopaedic surgeons. This contrast suggests that numerical features might allow for higher discriminability are preferred.

The second option is to enrich the health history questionnaire data with other data sources. These data sources could include data sources as Electronic Health Records (EHR) and medical imaging data. These three data sources are generally available to orthopaedic surgeons, and providing models with the same level of information might significantly improve performance. Both collecting more health history questionnaire data and including new data sources would also make it possible to apply deep learning rather than the more traditional algorithms applied in the current research. Deep learning algorithms ordinarily require substantially more data than traditional machine learning algorithms, which was not available at this point.

In conclusion, the developed models must be validated externally, both within the St. Anna hospital and other hospitals, to investigate whether the models generalise to real-world settings. Therefore, external validation is part of the next phase of this study: a randomised controlled trial at the St. Anna Hospital.

6 Methodology: Randomised Controlled Trial

In the previous research phase, it was concluded that models that aim to predict hip diagnoses and treatments performed better than models that aim to predict knee diagnoses and treatments with the available data. However, the two phases of this research partially overlapped due to time constraints. In an early stage of the study, the decision was made, in collaboration with a researcher of St. Anna, to focus the trial on patients with knee complaints because there are substantially more patients with knee complaints, both in the data sets and in practice. Accordingly, it was expected that models predicting knee diagnoses and treatments would be more accurate than models that aim to predict hip diagnoses and treatments. To ensure that the trial could commence in time, the decision was made to develop the knee diagnoses and treatment prediction models first. By the time the hip diagnoses and treatments models were developed, the trial had already commenced. The trial was performed and reported according to CONSORT guidelines ([Schulz, Altma, and Moher, 2010](#)).

First, section [6.1](#) delineates the objectives of the trial. In section [\(6.2\)](#), the trial design is described. After that, the process for obtaining the consent of participants (section [6.3](#)), how participants were selected (section [6.4](#)), and how the participants were randomised (section [6.5](#)) are discussed. Next, the two interventions and how the the study outcomes were measured is described in sections [6.6](#) and [6.7](#), respectively. Finally, the statistical analyses performed to obtain interim results are specified in section [6.8](#) and the software that was used in section [6.9](#).

6.1 Objectives

The main objective of this trial was to investigate whether the satisfaction of patients whose consultations prepared by orthopaedic surgeons with a medical history questionnaire of the patient and of patients who are offered information through an education eHealth application is higher than the satisfaction of patients who receive a standard consultation. The secondary objective of this trial was to investigate if the real- and perceived medical knowledge of patients who are offered information through an education eHealth application is higher than the satisfaction of patients who receive a standard consultation.

6.2 Trial Design

Between March and June 2021, patients referred to the St. Anna Hospital by their general practitioner with knee complaints were asked to participate in a randomised controlled trial. This study investigated the effect of orthopaedic surgeons preparing a consultation with a medical history questionnaire and offering patients information through an educational eHealth application on patient's satisfaction with their consultation. These interventions were compared with standard consultations in a parallel-group design with an equal allocation ratio.

Blinding was not possible for either group, introducing the risk of performance bias. Performance bias refers to differences that might occur because the patient or the surgeon is aware of the intervention allocation (Probst et al., 2016). Participants were required to be aware of which group they were assigned to because complete information was required for explicit consent and approval by the ethical board. Additionally, the complexity of the study could confuse participants. Providing complete information increased the likelihood that patients would understand and complete the study. Orthopaedic surgeons could not be blinded due to their involvement in the intervention. We considered the risk of performance bias to be of less importance than the benefits of not blinding.

6.3 Informed Consent and Ethical Consideration

Eligible patients were asked to consider participating in the study after scheduling an appointment for a consultation at the outpatient clinic. Usual care at the hospital includes that patients who schedule an appointment at the hospital are asked to fill in an online health history questionnaire. The software used for creating and distributing the health history questionnaires is onlinePROMs (Interactive-Studios, 2020).

Patients that marked their knee as the source of their complaints in the health history questionnaire received an invitation email (appendix J) with a link to an additional survey that contained all required information regarding the study, additional eligibility questions, and an explicit consent form to indicate their preference (appendix K). Patients that clicked on the link were shown basic information and were asked if they would be interested in participating. Patients that were not interested were able to close the survey. Patients that were interested were asked additional eligibility questions. Patients that were not eligible were thanked for their interest and were able to close the survey.

Interested and eligible patients were informed about the goals of the research and what it means to participate. Additionally, patients were informed about the purposes of the collection, processing, and storage of their data and that their data would be kept confidential and protected at all times. Patients were offered at least two days to reflect on the information and to sign the consent form. The information letter also contained my contact details in case patients had any questions. Finally, those who decided to participate were offered to receive all information again by email and receive the research paper after publication. There were no indicators of substantial risk as a function of participating in this study. The study was approved by both the regional Medical Ethical Board of the Maxima Medisch Centrum (Eindhoven, The Netherlands), reference number N21.001 (appendix H), and the Medical Ethical Commission of the St. Anna Hospital (appendix I) after handing in the required documents to both (appendix ?? and G). Processing of the data that has been collected was carried out on the ground of explicit consent under the General Data Protection Regulation. Finally, it is attested that the appropriate permissions were obtained and that any required fees for the use of copyright-protected materials were paid.

6.4 Participant Selection

Eligible participants had to be at least 18 years of age and referred by their general practitioner because of knee complaints. Furthermore, eligible participants were required to speak Dutch and be in possession of an email address and a smartphone or tablet. Lastly, there had to be at least six days between scheduling the appointment and the visit to the outpatient clinic to give patients enough time to consider participation, experience the intervention, and fill in the questionnaires.

A question in the medical history questionnaire assessed the initial eligibility of patients, which asked patients which part(s) of the body is the source of their complaints. Patients who answered that their knee was the source of their complaints received the invitation email. Further eligibility was assessed by asking the patient's age, whether the patient was in possession of a smartphone or tablet, and whether there were at least six days between the date of giving consent and the day of their consultation. Patients that did not meet the inclusion criteria were thanked for their interest and could close the survey.

Additional exclusion criteria applied to eligible patients that decided to participate. These additional criteria were determined in consultation with a researcher from St. Anna. First, participants whose predicted diagnoses was "other" were excluded from the study. Providing information to these patients was not possible because their specific diagnosis was unknown. Patients that were excluded from the study due to this reason received an email informing them of their exclusion and the reason why (appendix L). Second, participants were excluded if, for any reason, they did not entirely complete the study. Reasons include missed surveys, cancelled consultations, and technical issues. Finally, patients from the experimental group were excluded if they indicate that they did not use the patient Journey app.

6.5 Randomization

After an eligible patient signed the consent form, the researcher was automatically informed by e-mail. New participants were processed by the author at the end of each day by controlling and randomly assigning them to either the control group or the experimental group according to an automatically generated randomization scheme ([Urbanik and Plous, 2013](#)). Randomisation was performed without block or stratification restrictions.

After that, participants' diagnosis and treatment were predicted based on their completed health history questionnaire through the prediction tool described in section [4.3](#). Randomisation and prediction results were recorded by filling in a survey that linked to the participant (appendix M). Completing this survey triggered the release of the remaining content of the study.

Following randomisation, participants allocated to the experimental group were automatically informed of their allocation and that they would receive a separate e-mail with instructions on how to install the Patient Journey app ([Interactive-Studios, 2021](#)) by mail. These instructions were sent after the patient was created and assigned the patient to the correct information path in the online environment of the Patient Journey App.

Participants assigned to the experimental group began receiving information about their predicted diagnosis and treatment precisely five days before their consultation took place. Furthermore, both groups received the same questionnaires to measure this

study's primary and secondary outcomes at identical times. All questionnaires were implemented in the onlinePROMs environment of the health history questionnaire. The baseline perceived and actual medical knowledge questionnaires were made available directly after randomisation and prediction. Thereafter, all participants received these questionnaires at two other moments: one day before their consultation took place and one day after their consultation. Finally, all participants received a questionnaire regarding their satisfaction with the consultation one day after their consultation. If necessary, participants received a maximum of two e-mails as a reminder to complete the questionnaires.

6.6 Intervention

The current study investigated the utility of two interventions. First, patients that allocated to the experimental group received information on the Patient Journey App ([Interactive-Studios, 2021](#)) regarding a diagnosis and treatment that was predicted based on information from their health history questionnaire. Second, the health history questionnaires were studied by the orthopaedic surgeons to prepare consultations.

6.6.1 Educational eHealth application

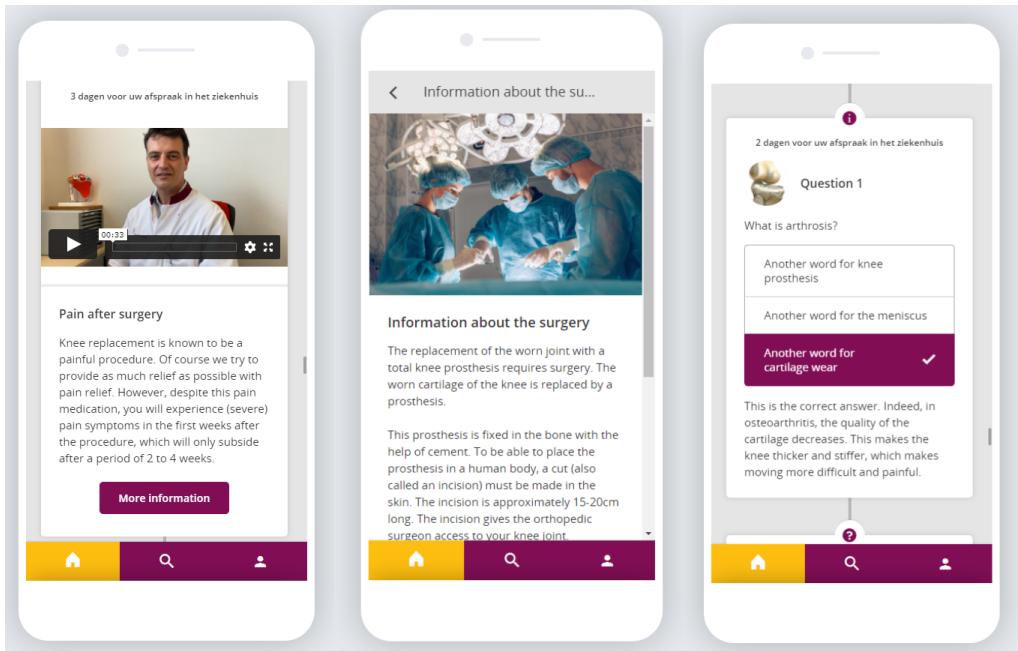
The Patient Journey App (PJA) is a mobile application that allows care providers to provide information to patients and monitor their progress. Access to the online environment of the app was granted to the author for the current research by Interactive Studios, the company behind the PJA.

Three interactive timelines were composed in collaboration with orthopaedic surgeons of St. Anna and an employee of Interactive Studios to offer patients information. One timeline was developed for each combination of diagnosis and treatment category: (1) gonarthrosis and conservative treatment, (2) gonarthrosis and surgical treatment, and (3) degenerative meniscus tear and either type of treatment. We chose to combine conservative and operative treatment information for degenerative meniscus tear because, in practice, degenerative meniscus tears are usually treated conservatively, except for rare cases in which the patient suffers from a locked knee. The complete timelines are shown in appendix O, P, and Q respectively. Translations of the content are provided next to the items. In case of an optional information item that the patient can click on, the items are placed next to each other with translations below the items. Duplicate items in the timelines are only displayed for the first timeline.

Each timeline offered subdivided and categorized pieces of information to patients over time. That is to say; patients were presented daily with a piece of information related to a particular theme such as their predicted diagnosis, their predicted treatment category, or general information about their consultation. Each piece of information was presented in several modalities, including text, image, and video content. The amount of information presented was limited initially to avert attention loss, but patients were offered the choice to obtain more detailed information by clicking on a button. Finally, patients were offered quiz-like questions to test and reinforce the knowledge that was gained. Example screenshots are shown in figure 15.

We composed the content based on concepts of information letters that orthopaedic surgeons wrote and literature. The following topics were addressed in each timeline: (1) information about the app, orthopaedics in general, and the study; (2) knee anatomy;

Figure 15: Examples screenshots of the gonarthrosis with total knee replacement surgery timeline. From left to right: the timeline, information item, and quiz question.



(3) the predicted diagnosis; (4) the predicted treatment; and (5) expectations and risks of the predicted treatment. Patients were reminded several times that the information is based on a prediction and that the diagnosis and/or treatment might differ in reality. Reminding patients was critical to reassure patients for whom surgery was predicted due to the impact that surgery can have. Additionally, we also expected that it would contribute to patients' understanding and acceptance of models' imperfections.

Patients started to receive information five days before their consultation. Thereafter, a part of the content was made available to the participants each day until their consultation. Patients were informed of new content with a notification at 19:00. Five days before their consultation, patients received information about the first three topics. On day 2 and 3, patients received information about topics 4 and 5, respectively. The final two days were reserved for the short quiz, a summary of the content, and practical information on preparing for the consultation.

6.6.2 Digital Health History Questionnaire

Orthopaedic surgeons and outpatient clinic assistants were informed of the study beforehand. The consultation dates for participants were tracked by the author. Each day, a list of all participants (if there were any) and their consultation times for the day after was sent to the outpatient clinic. The assistants let the orthopaedic surgeons know

when their next patient was a participant and if they were allocated to the control or intervention group.

The orthopaedic surgeon was instructed to perform a regular consultation for patients in the control group. In contrast, the orthopaedic surgeon was instructed to study the health history questionnaire of that participant before letting the patient in for patients in the experimental group.

6.7 Study Outcomes

Measurement of the study outcomes was performed at three moments in time: baseline, two days before the consultation, and one day after the consultation (Figure 16). All outcomes were measured with patient-reported questionnaires that were implemented in OnlinePROMs. The baseline measurement was unlocked directly after randomization and prediction of the participant. The timing of the follow-up measurement allowed participants in the experimental group to experience all the content in the Patient Journey App while providing enough time to complete the questionnaires before their consultation. The final follow-up measurement was planned one day after the consultation to evaluate recall and patient satisfaction with the consultation. Additionally, participants in the experimental group received an additional questionnaire to evaluate their subjective experience with the intervention.

Table 16: Outcomes measured at each measuring moment

Baseline	2 days before consultation	1 day after consultation
Perceived knowledge	Perceived knowledge	Perceived knowledge
Actual knowledge	Actual knowledge	Actual knowledge
n.a.	n.a.	Satisfaction with consultation
n.a.	n.a.	Subjective experience intervention

6.7.1 Real and perceived knowledge

The perceived- and real knowledge questionnaires (appendix R) were based on a similar questionnaires by Timmers et al. (2020). Although they did not validate the questionnaires in a patient population, they were carefully composed by orthopaedic surgeons and researchers of the hospitals that participated in that study.

To at least estimate the content validity, it is possible to evaluate the responsiveness of the survey afterwards (Shirley, Josephson, and Sanders, 2016). Considering that the study reported a significant difference in real- and perceived medical knowledge between the control and experimental group, the questionnaires were assumed to be responsive and thus content valid. Unfortunately, the reliability of the survey could not be estimated due to the inaccessibility of the raw results from that research.

The current research required a unique real- and perceived knowledge questionnaire for each diagnosis and treatment prediction combination. The content of these questionnaires was based on the five topics that were addressed in the PJA. The topics were similar to those in the research by Timmers et al. (2020), whose questionnaires tested the knowledge of patients about gonarthrosis and the options for treatment.

The perceived knowledge questionnaires included five Likert-scale questions, with answers that range from 1 (patient believes that he or she has no knowledge about the subject at all) to 5 (patient believes that he or she has complete knowledge of the subject). In total, patients could achieve a sum score between 5 and 25. The questionnaires aimed to evaluate patients' perceived knowledge on the topics that the PJA addressed. However, the questions slightly differed based on the predictions of the diagnosis and treatment.

The real knowledge questionnaires included twelve multiple-choice questions. A correct answer was rewarded with a score of 3, while wrong answers were rewarded with a score of 0. In total, patients could achieve a sum score between 0 and 36. Similar to the perceived knowledge questionnaire, the questions were based on the topics addressed in the PJA. Again, there was some variability in the questions based on the predictions of the diagnosis and treatment.

Patients automatically received the correct questionnaires based on their predicted diagnosis and treatment at baseline, two days before their consultation and one day after their consultation. An orthopaedic surgeon and two other experts reviewed all questionnaires on correctness, language, and simplicity.

6.7.2 Patient Satisfaction

The patient satisfaction questionnaire (appendix S) was specifically developed for this study and was based on validated questionnaires such as ConsultSQ (Baker, 1990). Unfortunately, none of the reviewed questionnaires captured all the elements addressed by the interventions, which is why the choice was made to introduce additional questions in consultation with a researcher of St. Anna.

The developed patient satisfaction questionnaire included eleven Likert-scale questions, with answers that ranged from 1 (unsatisfactory) to 5 (excellent). In total, patients could achieve a sum score between 11 and 55. Seven questions were selected to evaluate the general satisfaction of the participant with the consultation, and four questions were selected to evaluate the satisfaction of the patient with behavioural characteristics of the orthopaedic surgeon. The questions were composed based on the variables that previous studies had determined to influence patient satisfaction and that the proposed interventions addressed. In other words, the questionnaire aimed not to measure every aspect of patient satisfaction but rather those that the interventions attempted to improve.

A researcher of St. Anna collected pilot data ($n=322$) by administering the questionnaire to patients that visited the hospital for a regular consultation. Based on this data, the reliability of the questionnaire was evaluated by computing Cronbach's Alpha (Cronbach, 1960) according to formula 12. Here, N is equal to the number of questions, \bar{c} is the average inter-questions covariance among the questions, and v equals the average variance. With $\alpha = 0,93$, the questionnaire was demonstrated to be internally consistent.

$$\alpha = \frac{N\bar{c}}{v + (N - 1)\bar{(c)}} \quad (12)$$

6.7.3 Subjective experience of intervention

Participants allocated to the experimental group received a final questionnaire one day after their consultation to analyse their experience with the information offered in the PJA (appendix T). The questionnaire was specifically developed for the current research. The questionnaire was composed of multiple-choice questions and Numeric Rating Scales (1 to 10).

Patients were asked whether the predictions of their diagnosis and treatment were correct. Patients were also asked the extent to which they believe the computer is able to predict diagnoses and treatments correctly and whether they would like to receive information again in the future, even if the information is not entirely correct because of a wrong prediction. These questions were included to gain insight into the acceptance of and trust in AI techniques. Additionally, the latter question was included to obtain an indication of the extent to which wrong predictions are accepted or not. Finally, participants were asked if they liked receiving information in general, how much the information contributed to their feeling of preparedness, and if they believed that the information positively influenced their experience with the consultation. Comment fields were included in case participants had any additional feedback.

6.7.4 Preventing Bias

In section 2.1.3 several biases that might plague questionnaires were discussed. Measures that were taken to prevent these biases are discussed here. First, all questions were formulated as simple as possible to minimize the *optimizing bias*. Several non-experts that were not involved in the study reviewed the questionnaires. Their feedback made sure that technical jargon was replaced with language that is easy to understand. Another intention of the simplified formulation of the question was to limit the *central tendency bias*. Furthermore, participants were explicitly informed that comment fields were optional as an additional measure to prevent this bias.

The number of questions per questionnaire was kept as low as possible to prevent the *satisficing bias* with a maximum of 12 questions for the real knowledge questionnaires. However, participants received this questionnaire three times. This may have caused some participants to rush questions in one of the subsequent questionnaires. Unfortunately, this was a necessary evil to measure the outcomes.

Recall bias was naturally minimized because the questionnaires quickly follow after presenting the information and the consultation itself. Similarly, the chances of the *social desirability bias* to occur were automatically reduced because the questionnaires were self-reported instead of taken in person. Extra caution was taken by reassuring participants that there are no negative consequences if they do not get an answer right or honestly answer if they had a bad experience.

6.8 Statistical Analysis & Sample Size

The difference in patient satisfaction between the control and experimental groups was statistically analysed with three different representations of the concept of satisfaction in increasing granularity. The first representation of satisfaction was the summed score of the individual satisfaction items, where a higher score indicated higher satisfaction. The summed score provided a direct measure of overall satisfaction. However, it does not provide insight into the specific distribution of answers.

A Independent Samples T-test was performed to determine whether the mean summed satisfaction score of participants in the experimental group μ_E was higher than the mean summed satisfaction score of participants in the control group μ_C . The hypothesis that was tested is presented in equation 13.

$$\begin{aligned} H_0 : \mu_E &= \mu_C \\ H_1 : \mu_E &< \mu_C \end{aligned} \tag{13}$$

The second representation of satisfaction was the proportion of participants that rated all satisfaction items higher than a specific value. A statistical analysis of the difference in proportions was performed at three different levels: (1) the proportion of participants that rated all items higher than or equal to "satisfactory", (2) the proportion of participants that rated all items higher than or equal to "good", and (3) the proportion of participants that rated all items higher than or equal to "very good". This representation provided a more specific measure of satisfaction at different levels. Z-tests for the difference between proportions were performed to determine whether the proportion of participants that rated all items higher than or equal to a certain level of satisfaction in the experimental group p_E was higher than the proportion of participants that rated all items higher than or equal to a certain level of satisfaction in the control group p_C . The hypothesis that was tested for each level of analysis is presented in equation 14:

$$\begin{aligned} H_0 : p_E &= p_C \\ H_1 : p_E &> p_C \end{aligned} \tag{14}$$

Finally, the between-group population means between the experimental group μ_E and control group μ_C were compared for individual patient satisfaction survey question with the Independent Samples T-Test to gain insight into specific differences between items. The hypothesis that was tested for each survey question is presented by equation 15.

$$\begin{aligned} H_0 : \mu_E &= \mu_C \\ H_1 : \mu_E &\neq \mu_C \end{aligned} \tag{15}$$

Real and perceived knowledge was represented by the summed score of individual survey items. Between-group differences in real- and perceived knowledge at baseline, two days before the consultation, and one day after the consultation were compared with Independent-Samples T-Tests. The hypothesis that was tested determine whether the mean real- and perceived knowledge levels of participants in the experimental group μ_E was higher than the mean real- and perceived knowledge levels of participants in the control group μ_C is presented in equation 13.

Moreover, within-group differences in real- and perceived knowledge were statistically compared using Repeated Measures ANOVA to determine whether there were any differences between the population means over all time points in general and with the Bonferroni post hoc test to determine whether there were differences in the population means between every two consecutive points in time. The hypothesis that was tested to determine whether there was any difference in the within-group real- and perceived knowledge means at baseline μ_{baseline} , two days before the consultation μ_{T1} , and one day after the consultation μ_{T2} is presented in equation 16.

$$\begin{aligned} H_0 : \mu_{\text{baseline}} &= \mu_{T1} = \mu_{T2} \\ H_1 : \text{at least two means are significantly different} \end{aligned} \quad (16)$$

The hypothesis that was tested to determine whether the real- and perceived knowledge means between any two consecutive points in time μ_i and μ_j were different is presented in equation 17.

$$\begin{aligned} H_0 : \mu_i &= \mu_j \\ H_1 : \mu_i &\neq \mu_j \end{aligned} \quad (17)$$

Finally, the intervention experience was not analysed statistically. Alternatively, numeric questions were represented by their mean value and range, and categorical questions were represented by counts of the answers and percentages.

The minimally required sample size was computed based on the proportion representation of satisfaction with an online calculator¹, which implements formulas from Chow, Wang, and Shao (2007). Additionally, the proportion of participants that rated all items higher than or equal to "good" was selected as the basis for the calculation. Based on the pilot data ($n=322$), of which a proportion of 0,45 (45%) patients rated all items higher than or equal to "good", and assuming a positively estimated increase to a proportion of 0,65 (65%) after introducing the interventions, the minimally required sample size was equal to 102 participants per study arm, or 204 participants in total.

6.9 Software

Statistical analyses were performed in Python (Van Rossum and Drake, 2009) with the Statsmodels package (Seabold and Perktold, 2010) and with IBM SPSS Statistics (IBM Corp., 2019). Visualisations were made with the python package Plotly (Plotly Technologies Inc., 2015)

¹ <http://powerandsamplesize.com/Calculators/Compare-2-Proportions/2-Sample-1-Sided>

7 Pilot results: Randomized Controlled Trial

In the current section, the pilot results of the randomised controlled trial are discussed. At the time of writing, the trial is still ongoing because the minimally required sample size was not met within the timeframe of the thesis. Hence, the results that are presented here can be regarded as pilot results. The analysis methodology remains as described.

First, the performance of the knee diagnosis and treatment models in practice is validated. After that, the flow of participants from eligibility assessment to analysis and their baseline demographics is presented. Lastly, the interim results of the statistical analyses that were performed on the pilot data are presented.

7.1 Validation performance

Table 17 shows the metric scores for the knee diagnosis and treatment models based on the predictions that were generated for participants of the trial ($n=68$). The predictions were compared with the true labels of the data, were manually added to the data after the consultations. Both models seemed to have higher metric scores than during testing. Looking at the confusion matrices (figure 16, classification accuracy was better for all diagnoses and both treatment categories. Hence, it seems that the models generalise to a real-world setting.

Figure 16: Normalized Diagnosis Confusion Matrices

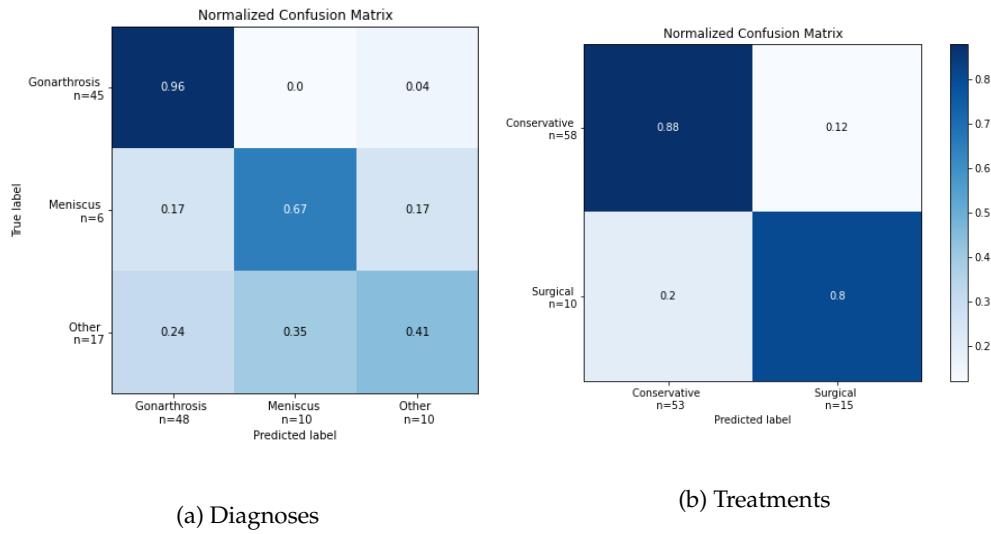


Table 17: External model validation

	Accuracy	Precision	Recall	F1 Score
Diagnoses	0.68	0.8	0.79	0.79
Treatments	0.84	0.9	0.87	0.88

7.2 Participants

Figure 17 shows the flow of participants from eligibility assessment to analysis. Between April and June 2021, 169 patients were assessed for eligibility after opening the invitation email. 113 patients were excluded because they either declined to participate (49 patients) or failed to meet the inclusion criteria (64 patients). Specifically, 3 patients did not own a smartphone or tablet, 55 patients had their consultation within 6 days after receiving the invitation, and for the remaining 9 patients, the predicted diagnoses was "other".

As a consequence, 56 patients remained that met the inclusion criteria and consented to participation. These patients were randomly assigned to either the intervention group (28 patients) or the control group (28 patients). 10 participants in the intervention group were excluded because they never started the study (2 participants), did not use the intervention (7 participants), or did not show up at their consultation (1 participant). In addition, 4 participants in the control group were excluded because they never started the study (1 participant), or their consultation date is sometime in the future after the time of writing.

Moreover, participants who started the study but missed one or more of the follow-up questionnaires were excluded as well. 4 participants in the experimental group and 8 participants in the control group were excluded for this reason. In the end, 14 participants in the experimental group and 16 patients in the experimental group remained for analysis. Figure 18 shows when participants were recruited and their course through the study. The length of the bars related to the baseline or one of the follow-up questionnaires represents the amount of time between the time that the invitation email was sent and the participant finishing the questionnaire. The recruitment rate was somewhat stable, and participants tended to fill out questionnaires within appropriate timeframes after being invited.

Table 18 summarises the demographics of participants in the control group (8 male (50%), mean age = 62.7 (7.55 SD) years) and demographics of participant in the experimental group (8 male (57.1%), mean age 62.7 (11.76 SD) years). Both groups reported a score of around 4 for pain in rest and 6.5 for during movement. Patients in the experimental groups had higher real- and perceived knowledge scores at baseline, but the differences were not significant.

Figure 17: Participant flow diagram

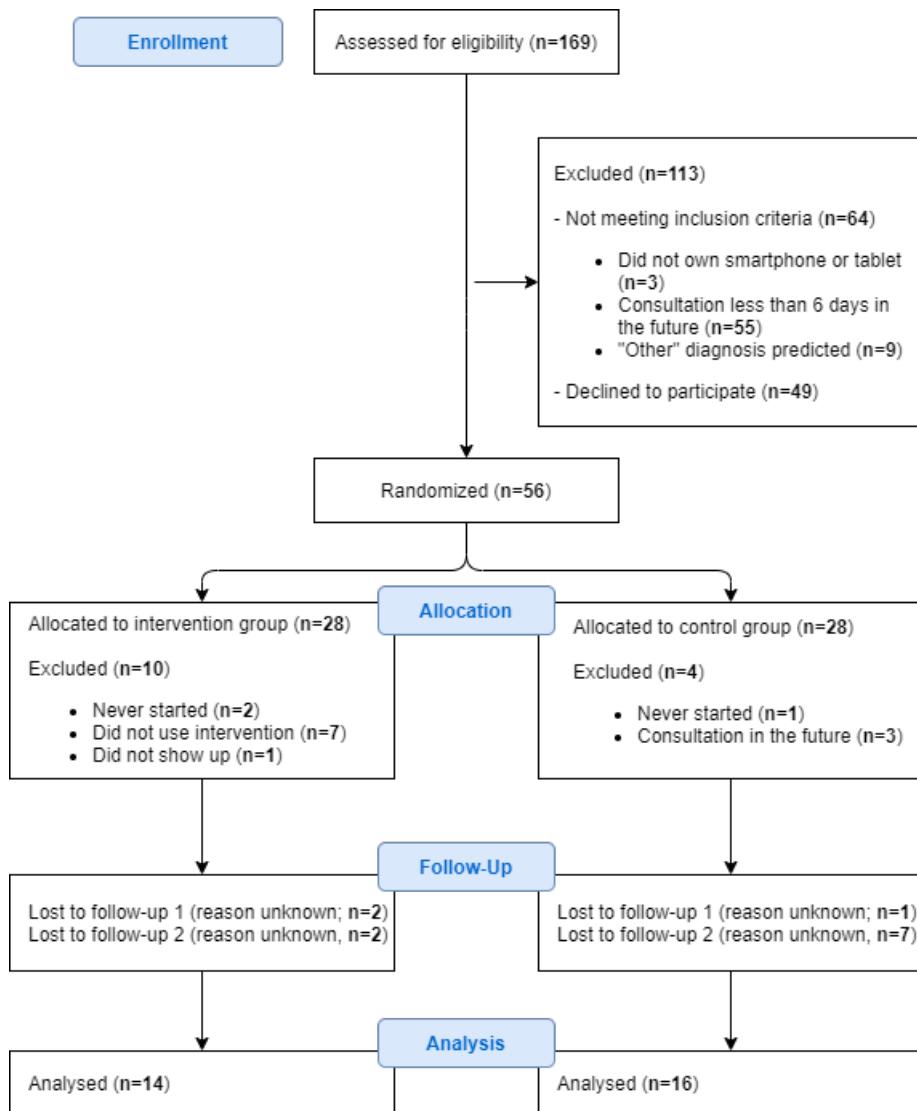


Table 18: Participant demographics

Numeric demographic		Experimental (n=14)	Control (n=16)
		mean (SD)	mean (SD)
Age		62.69 (11.76)	62.69 (7.55)
Length		173.85 (8.26)	172.38 (9.25)
Weight		81.0 (20.04)	81.38 (13.66)
Pain in rest ^a		4.0 (2.92)	4.13 (2.66)
Pain during movement ^a		6.54 (2.63)	6.63 (2.28)
Real knowledge score ^b		6.54 (1.39)	5.81 (1.64)
Perceived knowledge score ^c		15.85 (3.02)	13.38 (3.22)

Categorical demographic	Category	Experimental (n=14)	Control (n=16)
		count (%)	count (%)
Gender	Male	8 (57.1%)	8 (50%)
	Female	6 (42.9%)	8 (50%)
Duration of complaints	Days	0	2 (12.5%)
	Weeks	0	2 (12.5%)
	Months	5 (38.5%)	3 (18.75%)
	Years	8 (61.5%)	9 (56.25%)
Complaints due to trauma	Yes	1 (7.7%)	7 (43.75%)
	No	12 (92.3%)	9 (56.25%)
Able to walk	>15 minutes	8 (61.5%)	6 (37.5%)
	<15 minutes	5 (38.5%)	10 (62.5%)

^a Numeric Rating Scale from 1 to 10^b Based on 12 questions, 1 point per correct answer (max 12 points, higher is better)^c Based on 5 Likert scale questions, 1 to 5 points per answer (max 25 points, higher is better).

Figure 18: Recruitment



7.3 Outcomes and Estimation

In the current section, the statistical analyses that were performed on the pilot data are presented. First, the statistical analysis of between-group differences in patient satisfaction for each of the previously defined definitions of patient satisfaction is presented. After that, the statistical analyses of between-group and within-group differences in medical knowledge two days before the consultation and one day after the consultation are presented. Finally, the analysis of the subjective experience survey results is provided.

7.3.1 Patient satisfaction

Between-group comparison of summed patient satisfaction scores seemed to demonstrate a trend that patients in the control group are less satisfied with their consultation (mean 29.06 [11.45 SD]) than patients in the experimental group (mean 32.08 [10.23 SD]), but an Independent-Samples T-Test showed that this difference was not significant ($P=.466$; table 19).

In addition, between-group comparisons of proportions showed that there seemed to be a trend that patients in the experimental group tended to rate every item on the survey with satisfactory or better (proportion=1 [100%]) more often than patients in the control group (proportion=.875 [87.5%]), but a z-test for the difference between proportions showed that this difference was not statistically significant ($P=.093$; table 20). In contrast, the proportions of patients that rated every item on the survey with good or better, or with very good or better were nearly identical for both the experimental (proportion 0.688 [68.8%] and 0.385 [38.5%] respectively) and the control group (proportion 0.692 [69.2%] and 0.375 [37.5%] respectively). Z-tests for the difference

between proportions confirmed that these differences were not statistically significant ($P=.489$ & $P=.479$; table 20).

Table 19: Between-group differences of main outcomes

Outcome	2 days before consultation			1 day after consultation		
	App (n=14), mean (SD)	Control (n=16), mean (SD)	P value	App (n=14), mean (SD)	Control (n=16), mean (SD)	P value
Actual knowledge ^a	7.31 (2.62)	5.75 (2.49)	.114	7.15 (2.38)	6.69 (1.78)	.550
Perceived knowledge ^b	16.54 (5.97)	12.19 (4.78)	.038	17.46 (5.95)	15.31 (3.54)	.237
Satisfaction score ^c	n.a.	n.a.	n.a.	32.08 (10.23)	29.06 (11.45)	0.466

^a Based on 12 questions, 1 point per correct answer (max 12 points, higher is better)

^b Based on 5 Likert scale questions, 1 to 5 points per answer (max 25 points, higher is better).

^c Based on 11 Likert scale questions, 1 to 5 points per answer (max 55 points, higher is better)

Table 20: Proportion analysis of patient satisfaction

Outcome	Proportion experimental	Proportion control	P value
Every item rated with satisfactory or better	1	0.875	.093
Every item rated with good or better	0.688	0.692	.489
Every item rated with very good or better	0.385	0.375	.479

Lastly, the between-group comparisons of specific survey item with the Independent Samples T-Test did not demonstrate any significant differences (table 21). The biggest differences were found for the patients' belief of how good the medical care at the outpatient clinic is (experimental: mean 2.25 [.62 SD], control: 2.38 [.96 SD], $P=.680$), the politeness of the doctor (experimental: mean 1.67 [.78 SD], control: mean 2.06 [1.06 SD], $P=.235$), and the helpfulness of the doctor (experimental: mean 1.67 [.65 SD], control: mean 2.34 [1.20 SD], $P=.088$).

7.3.2 Medical Knowledge

Between-group comparison showed that there was a trend that participants in the experimental group had a higher level of real knowledge two days before the consultation (mean 7.31 [SD 2.62]) than participants in the control group (mean 5.75 [SD 2.49]), but the difference was not statistically significant ($P=.114$; table 19 and figure 19). Furthermore, the level of real knowledge of participants in the control group (mean 6.69 [1.78 SD]) was closer to that of participants in the experimental group (mean 17.46 [5.95 SD]) one day after the consultation ($P=.550$).

A similar trend was observed for perceived knowledge. Participants in the experimental group had a higher level of perceived knowledge two days before the consultation (mean 17.46 [5.95 SD]) than participants in the control group (mean 12.19 [4.78 SD]).

Table 21: Mean analysis of individual satisfaction items

Question	App (n=14)	Control (n=16)	P value
	mean (SD)	mean (SD)	
How do you feel about the amount of time the doctor had for you?	2.69 (1.01)	2.75 (.97)	.870
How thorough/comprehensive was the care you received from the doctor?	2.67 (.78)	2.62 (1.03)	.907
How did you experience the explanations/instructions that the doctor gave you about medication and aftercare?	2.17 (1.33)	2.31 (1.45)	.787
What do you think of the advice the doctor gave you about preventing complaints and staying healthy?	2.03 (1.44)	2.25 (1.44)	.765
How satisfied are you with the doctor's explanation of what he / she could do for you (examinations, diagnosis, treatment)?	2.67 (.78)	2.62 (1.03)	.907
What do you think of the results of the medical care that you received (to what extent were you helped as expected)?	2.42 (1.08)	2.50 (1.55)	.875
How good do you think the medical care at the outpatient clinic is?	2.75 (.62)	2.62 (.96)	.680
What did you think of the helpfulness of the doctor?	3.33 (.78)	2.62 (1.20)	.088
What did you think of the politeness of the doctor?	3.33 (.65)	2.94 (1.06)	.235
What did you think of the empathy of the doctor?	3.0 (.74)	2.87 (1.20)	.754
What did you think of the professionalism of the doctor?	3.0 (.74)	3.0 (.97)	1

In contrast to real knowledge, this difference was statistically significant ($P=.038$; table 19 and figure 20). However, this difference again seemed to become smaller one day after the consultation because the difference was no longer significant (experimental: mean 17.46 [5.95 SD], control: mean 15.31 [3.54 SD], $P=.237$).

Figure 19: Level of real knowledge at baseline and follow-up moments

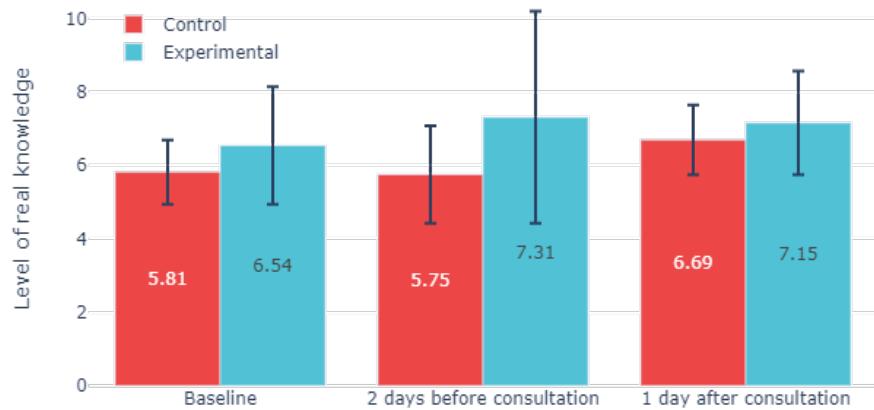
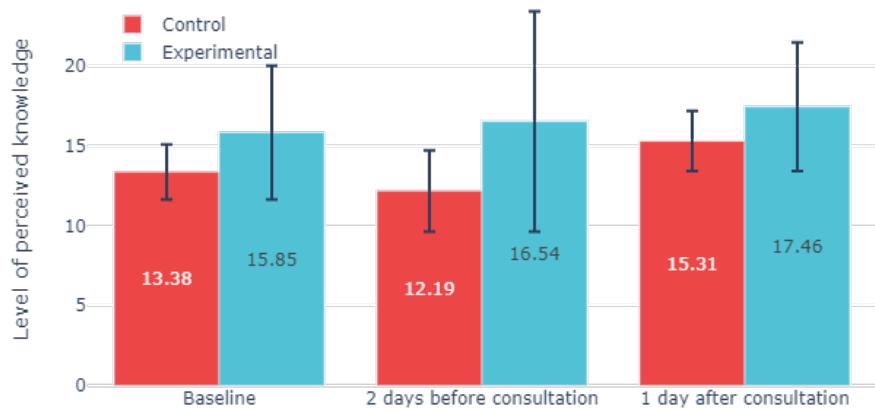


Figure 20: Level of perceived knowledge at baseline and follow-up moments



Repeated measures ANOVA analysis did not show a statistically significant difference in within-group real knowledge for either the control ($P=.147$) or the experimental group ($P=.302$). Insignificance was also reported for the within-group difference in perceived knowledge of the experimental group ($P=.461$). In contrast, a statistically significant difference was found for the within-group difference in perceived knowledge of the control group ($P=.015$). Comparisons of within-group pairwise means with the Bonferroni post hoc test did not show any significant differences (tables 22 and 23).

Table 22: Within-group differences in real knowledge

Time (i)	Time (j)	Group	Mean (SD) (i)	Mean (SD) (j)	P value
Baseline	2 days before consultation	Experimental	6.54 (1.39)	7.31 (2.6)	.719
		Control	5.81 (1.64)	5.75 (2.49)	1
2 days before consultation	1 day after consultation	Experimental	7.31 (2.63)	7.15 (2.38)	1
		Control	5.75 (2.49)	6.69 (1.78)	.395

Table 23: within-group differences in perceived knowledge

Time (i)	Time (j)	Group	Mean (SD) (i)	Mean (SD) (j)	P value
Baseline	2 days before consultation	Experimental	15.85 (3.02)	16.54 (5.97)	1
		Control	13.38 (3.22)	12.19 (4.78)	.455
2 days before consultation	1 day after consultation	Experimental	16.54 (5.97)	17.46 (5.95)	.656
		Control	12.19 (4.78)	15.31 (3.54)	.060

7.3.3 Intervention Experience

In the intervention experience survey, patients in the experimental group were asked about their experience with the information that they received through the Patient Journey App. Table 24 presents the interim results from the survey. Participants were divided on their trust in the ability of a computer to predict diagnoses and treatments (mean 6 [range 3-7]) and on whether they liked receiving information (mean 6 [range 4-10]). Participants were somewhat more optimistic about how well-prepared they felt for the consultation (mean 7 [range 6-8]) and the extent to which the information contributed to their feeling of preparedness (mean 7 [range 5-10]).

Table 24: Analysis of intervention experience survey

Question	Mean	Minimum	Maximum
To what extend do you believe a computer is able to correctly predict diagnoses and treatments?	6	3	7
Did you like receiving information before your consultation?	6	4	10
How well-prepared did you feel for your consultation?	7	6	8
To what extend did the information contribute to your feeling of preparedness?	7	5	10

Question	Answer	Count	Percentage
How did the information that you received influence your consultation?	Positively	3	23.1%
	Neutral	7	53.8%
	Negatively	1	7.7%
	No answer	2	15.4%
Would you like to receive information again next time?	Yes, but only if it is certain that the predictions are correct	1	7.7%
	Yes, even if the predictions might be wrong	7	53.8%
	No	3	23.1%
	No answer	2	15.4%
Do you think that the way the information was presented pleasant and clear?	Yes	11	84.6%
	No	0	0%

Table 24 continued from previous page

Question	Mean	Minimum	Maximum
	No answer	2	15.4%

Participants also seemed to have reservations about how the information had influenced their consultation, with most participants (7/13 [53.8%]) expressing that it had a neutral influence on their consultation. 3/13 (23.1%) of the participants expressed a positive influence, and 1 participant expressed a negative influence. Given the previously observed reservations, it was interesting to observe that most participants would like to receive information again before their next consultation, even if the predictions, and thus the information, might be wrong (7/13 [53.8%]). Nevertheless, 3 participants expressed that they would not like to receive information again at all (3/13 [23.1%]), or only if it is certain that the predictions are correct (1/13 [7.7%]). Another aspect that participants considered positive is how the information was presented, with all participants who answered the question expressing that the information was presented was pleasant and straightforward (11/13 [84.6%]).

Finally, some participants wrote explicit remarks in one of the open text fields. For example, one patient whose treatment was wrongfully predicted remarked that the prediction was not entirely wrong but that his home situation strongly influenced his treatment choice. A different patient was frustrated because he already had a diagnosis and received information based on a wrong prediction. On a positive note, two patients explicitly expressed that they very much enjoyed the intervention. They believed that it is an excellent addition to regular care and that they believe it will help many people in the future.

8 Discussion: Randomised Controlled Trial

The phase of the study aimed to improve patients' satisfaction with consultations and patients' medical knowledge by offering patients information through an educational eHealth tool and having orthopaedic surgeons prepare their consultations with health history questionnaires. At the time of writing, the trial is still ongoing. It was therefore not possible to draw valid conclusions from the pilot results. However, the pilot results allowed for identifying some early trends that will hopefully generalise to a larger population and shown to be statistically significant once the minimally required sample size is reached.

First, there seemed to be a trend that patients who received information by means of an educational eHealth tool and whose consultations were prepared by the orthopaedic surgeon by examining the patient's health history questionnaire were slightly more satisfied with their consultation than patients who received a regular consultation. A similar trend was observed for both real- and perceived medical knowledge, which are slightly higher for patients in the experimental group two days before their consultation. If these trends are shown to be statistically significant in the future, the interventions would be a valuable addition to regular care.

The value of the educational eHealth tool seemed to be corroborated by the pilot results of the intervention experience survey. A majority of patients expressed that they would like to receive information again in the future, even if the predictions might be incorrect. The generalisation of this trend to a larger population and a statistically significant difference in patient satisfaction would mean that a certain margin of error is acceptable to patients and would mean that even though imperfect at this time, an educational eHealth tool would be a valuable addition to the regular care of orthopaedic hospitals.

8.1 Strengths and limitations

Despite the fact that the trial is still ongoing, several strengths and limitations of the current study were already identified. To the knowledge of the author, the current study is the first to provide information to patients pre-consultation about predictions of their diagnosis and treatment. It is also the first study to investigate its effect, combined with having orthopaedic surgeons prepare consultations by examining health history questionnaires on patients satisfaction with their consultation. Moreover, the mechanisms used to provide information to patients are grounded in literature, and previous studies have shown that they improve patients' medical knowledge (Timmers et al., 2018, 2020). The information was composed in collaboration with orthopaedic surgeons and experts, and understandability was improved by involving non-experts in the development process.

A substantial limitation of the study is that the models used to predict diagnoses and treatments are moderate at best. Although performance seems somewhat better in practice compared to testing, there is still a margin of error of 30-40% in for diagnoses and 20-30% for the treatment categories. In consequence, at least 30-40% of the participants in the study receive wrongful information. The margins of error may have a significant adverse effect on the outcomes of the study. However, the difference in satisfaction and knowledge between participants that receive correct information and participants that receive wrongful information will be evaluated after the termination of the trial to check for the influence of wrongful information on the outcomes.

The second limitation of the study is that a considerable fraction of the patients that are assessed for eligibility do not meet the inclusion criteria (64/169 [37.9%]) or decline to participate (49/169 [29%]). As a result, more than half of patients that were assessed for eligibility were lost before randomization. Most notably, most patients who failed to meet the inclusion criteria did so because their consultation took place within six days from the day they were invited to participate (55/62 [85.9%]). This is primarily due to COVID-19, which has caused a decrease in the number of patients that visit the outpatient clinic, ensuring that consultations of new patients are planned relatively soon. What is more, half of the patients that met the inclusion criteria and consented to participate were excluded from analysis in the end. For participants in the experimental group, the major reasons for exclusion were that they did not use the Patient Journey App (7/14 [50%]) or failed to complete one or more of the follow-up questionnaires (4/14 [28.6%]). Several reasons might explain why a participant failed to use the Patient Journey App, such as failing to install the app, missing the email with instructions to download the app, or failing to understand how the app works. For participants in the control group, the major reason for exclusion was failing to complete one or more questionnaires (8/12 [0.67%]). Especially the final follow-up moment was often missing (7/8 [0.875%]). This might have happened because participants in the control group receive the same medical knowledge questionnaire three times without knowing exactly why. It is conceivable that participants do not want to fill in the same questionnaire three times or that they believe that it was a mistake by the researcher.

Another limitation is the strict selection of diagnoses included in the study, caused by the necessity to merge a wide range of infrequently occurring diagnoses under a single denominator during model development. Although, in practice, the two diagnoses that were included account for more than 80% of the patient at the St. Anna hospital, a significant portion of the patient population is excluded from the study, possibly introducing bias as a result. Other possible sources of bias include the single-centre design of the study, the inability to blind neither patients nor the orthopaedic surgeons participating in the study, and the repeated administration of the medical knowledge questionnaires. In other words, the trial is performed at a single hospital, calling the generalisability of the results into question. Furthermore, the inability to blind neither the patients nor the orthopaedic surgeons participating in the study might result in performance bias. Finally, the medical knowledge questionnaire's repeated administration might have primed patients for the questionnaires that followed after the baseline measurement. However, the possible influence of priming on the results are mitigated by the fact that it would occur in both groups.

Additionally, the decision to investigate the effects of both interventions simultaneously has a disadvantage. There is no way to tell the individual contribution of each intervention to changes in patients satisfaction with their consultation. The contribution of one intervention may be more significant than that of the other. Adding separate experimental groups that receive one of the two interventions would have been preferred. However, it would also have increased the complexity of the study and the minimally required sample size. The option was considered, but deemed infeasible for now.

Finally, the patient satisfaction survey was explicitly developed for the current study. Internal consistency was demonstrated, but its validity is unknown. Moreover, the questionnaire contains items that are not directly addressed by the interventions, possibly influencing the results in unforeseen ways. The development of the questionnaire itself was a challenging task because patient satisfaction is such a multifaceted concept that it is difficult to cover in a single questionnaire.

8.2 Future Work

In the first instance, the ongoing trial will continue until the minimally required sample size is reached. At that point, the analyses will be repeated with the intention to publish the results in a scientific journal. Regardless of the trial's outcome, there are a few exciting paths that future research could take.

First, future research could investigate ways to improve upon the existing models to improve the accuracy of the information that is provided to patients. Moreover, future research could investigate the inclusion of additional diagnoses and specific treatment options in the predictive models to expand the size of the population that the information can be offered to. Possibilities for improving upon the existing models and the inclusion of additional diagnoses and specific treatment options were previously discussed in section 5.3. Another way to increase the size of the patient population that the information can be offered to is to add new types of models such as the previously developed models for the prediction of diagnoses and treatments of the hip, or models that aim to predict diagnoses and treatments for currently unexplored extremities such as the shoulder.

Second, it is critical to validate if the results of the trial generalise to other hospitals. Repeating the study in other hospitals will have to show whether the performance of the developed models is comparable to their performance in the current study and whether the interventions achieve similar effects. If the trial results are positive and generalisation is demonstrated, implementing the interventions on a large scale could be an exciting business opportunity. Of course, it must be certain that the instrument that was used to measure patients' satisfaction is valid, necessitating the validation of the patient satisfaction questionnaire developed for this trial.

9 Conclusion

As the need for orthopaedic care continues to increase, trying to keep patients satisfied becomes an increasingly challenging task. Although patients at St. Anna are currently quite satisfied with their consultations, increasing administrative burdens for orthopaedic surgeons make it evident that there are no guarantees for the future. The increasing administrative burden of orthopaedic surgeons is impairing their ability to prepare for their consultations properly. In turn, inadequate preparation can cause orthopaedic surgeons to be unable to properly communicate with, empathise with and relate to their patients. Furthermore, a general lack of medical knowledge in patients and limitations in their memory and understanding are causing orthopaedic surgeons to face patients with unrealistic expectations increasingly often and are causing the communication between the surgeon and the patient to be far from optimal frequently.

Patient expectations, doctor-patient communication, and the ability of orthopaedic surgeons to empathise with and relate to the patient have all been shown to influence patients satisfaction with their consultations. Therefore, it is evident that it is necessary to keep improving care to prevent these challenges from harming patient satisfaction. Hence, the primary objective of the current research was to make orthopaedic consultations future-proof by supporting orthopaedic surgeons in preparing for their consultations and increasing patients' medical knowledge and, by extension, increase patients satisfaction with their consultations.

The introduction of a digital Health History Questionnaire was introduced to support orthopaedic surgeons in preparing for their consultations. Simultaneously, an informational eHealth tool was introduced to provide information regarding a prediction of their diagnosis and appropriate treatment to patients before their consultation to increase their a priori medical knowledge.

Based on the main objective and the proposed interventions, the main research question was formulated: "is the satisfaction of patients whose consultations are prepared by the orthopaedic surgeon by examining their digital health history questionnaire and prepared by the patient by reading personalised information regarding his or her predicted diagnosis and treatment higher than the satisfaction of patients that receive a standard consultation?". The first sub-question was formulated as follows: "is the actual and perceived medical knowledge of orthopaedic patients increased by providing them with information before their consultation by means of an educational eHealth application compared to patients who do not receive this information?". The fact that providing information to patients beforehand required a prediction of their diagnosis and treatment, the second sub-question was formulated: "How accurately can orthopaedic diagnoses be predicted by machine learning algorithms trained on health history questionnaire data?"

The current research was split into two separate phases with independent methodology, results, and discussion sections. The first phase aimed to answer the main research question and the first sub-question, while the second phase aimed to answer the first and the second sub-question, respectively. In the first phase, Machine Learning methods were applied to Health History data of patients with hip and knee complaints to develop predictive models that can predict diagnoses of the knee and hip and the appropriate way to treat the complaints. The predictive models were developed to provide information to the patients in the subsequent research phase. However, the characteristics of the datasets necessitated a few concessions. For example, infrequently

occurring diagnoses had to be combined into a single class, and treatments had to be categorised into a surgical and conservative class.

To answer the second sub-question: the developed models' performance was shown to be moderate, but hip-related models performed slightly better than knee-related models. Differences between the performance of the evaluated models were found to be small for each prediction problem, so it was not possible to determine a superior algorithm. The most successful algorithm for predicting knee diagnoses was found to be AdaBoost (0.61 balanced accuracy). The most successful algorithm for predicting treatments of knee complaints was found to be the Random Forest (0.7 balanced accuracy). The Random Forest was also found to be the most successful algorithm for the prediction of hip diagnoses (0.69 balanced accuracy). Finally, the Support Vector Machine was the most successful algorithm for predicting treatments of knee complaints (0.84 balanced accuracy).

Regardless of the moderate performance, the choice was made to perform a randomised controlled trial with the models that predict knee diagnoses and the appropriate treatment. This trial was the subject of the second phase of the study. The control group received standard consultations, whereas the experimental group received the interventions. Unfortunately, the minimally required sample size was not reached within the timeframe of this thesis, and as a result, the trial is still ongoing. However, the pilot results that were gathered were analysed to identify early trends.

First, there seemed to be a trend that participants in the experimental group ($n=14$) have a higher level of real knowledge two days before the consultation (mean 7.31 [SD 2.62]) than participants in the control group ($n=14$, mean 5.75 [SD 2.49]), but the difference was not statistically significant ($P=.114$). Second, it also appeared that participants in the experimental group have a higher level of perceived knowledge two days before the consultation (mean 17.46 [5.95 SD]) than participants in the control group (mean 17.46 [5.95 SD]). In contrast to real knowledge, this difference was found to be significant ($P=.038$). To provide a premature answer to the first sub-question: the real knowledge of orthopaedic patients seems to be higher for patients that received information by mean of an educational eHealth application compared to patients who do not receive this information, but the difference was not found to be statistically significant and therefore not conclusive. The same difference was observed for perceived knowledge, but this difference was significant in contrast to real knowledge. Hence, it is likely that patients who receive information before their consultation believe that they are more knowledgeable than patients who do not receive information before their consultation.

Finally, a positive trend was also found for patients satisfaction with their consultation, which seems to be higher for participants in the experimental group (mean 32.08 [10.23 SD]) than for participants in the control group (mean 29.06 [11.45 SD]). However, again, this difference was not found to be statistically significant ($P=.466$). Therefore, to provide a premature answer to the main research question: the satisfaction of patients whose consultations were prepared by the orthopaedic surgeon by examining their digital health history questionnaire and prepared by the patient with an educational eHealth application seems to be higher than the satisfaction of patients that received a standard consultation. Again, however, the difference was not statistically significant and, therefore, not conclusive.

With the trial still ongoing, it is too early to draw definitive conclusions regarding patients satisfaction with their consultations and patients medical knowledge. However, if the identified trends generalise to a larger population and proven to be statistically significant, this study would achieve its primary objective of improving patients satisfaction with orthopaedic outpatient consultations.

References

- Abdi, Hervé and Lynne J Williams. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459.
- Adsit, K. 1996. Multimedia in nursing and patient education. *Orthopaedic Nursing*, 15:59–63.
- Ahern, D., Jennifer M Kreslake, and Judith M. Phalen. 2006. What is ehealth (6): Perspectives on the evolution of ehealth research. *Journal of Medical Internet Research*, 8.
- Alanzi, T., S. Bah, Sarah Alzahrani, Sirah Alshammary, and Fatima Almunsef. 2018. Evaluation of a mobile social networking application for improving diabetes type 2 knowledge: an intervention study using whatsapp. *Journal of comparative effectiveness research*, 7 9:891–899.
- Ashinsky, B., M. Bouhrara, C. Coletta, B. Lehallier, K. Urish, Ping-Chang Lin, I. Goldberg, and R. Spencer. 2017. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *Journal of Orthopaedic Research*, 35.
- Atkinson, Elizabeth J., Terry M. Therneau, L. Melton, J. Camp, Sara J Achenbach, S. Amin, and S. Khosla. 2012. Assessing fracture risk using gradient boosting machine (gbm) models. *Journal of Bone and Mineral Research*, 27.
- Attfield, S., Anne Adams, and A. Blandford. 2006. Patient information needs: pre- and post-consultation. *Health Informatics Journal*, 12:165 – 177.
- Baker, R. 1990. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *The British journal of general practice : the journal of the Royal College of General Practitioners*, 40 341:487–90.
- Barello, S., S. Triberti, G. Graffigna, C. Libreri, S. Serino, J. Hibbard, and G. Riva. 2016. ehealth for patient engagement: A systematic review. *Frontiers in Psychology*, 6.
- Barnett, Sarah F, Ravi K Alagar, M. P. Grocott, S. Giannaris, John R. Dick, and S. R. Moonesinghe. 2013. Patient-satisfaction measures in anesthesia: Qualitative systematic review. *Anesthesiology*, 119:452–478.
- Beattie, M., W. Lauder, I. Atherton, and D. Murphy. 2014. Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. *Systematic Reviews*, 3:4 – 4.
- Beddows, J. 1997. Alleviating pre-operative anxiety in patients: a study. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, 11 37:35–8.
- Belkora, J., S. Volz, Meredith K. Loth, A. Teng, M. Zarin-Pass, D. Moore, and L. Esserman. 2015. Coaching patients in the use of decision and communication aids: Re-aim evaluation of a patient support program. *BMC Health Services Research*, 15.
- Bental, D., A. Cawsey, and R. Jones. 1999. Patient information systems that tailor to the individual. *Patient education and counseling*, 36 2:171–80.
- Bien, Nicholas, Pranav Rajpurkar, R. Ball, J. Irvin, Allison Park, Erik Jones, Michael Bereket, B. Patel, K. Yeom, K. Shpanskaya, S. Halabi, E. Zucker, G. Fanton, D. Amanatullah, C. Beaulieu, Geoffrey M. Riley, R. Stewart, F. Blankenberg, D. Larson, Ricky Jones, C. Langlotz, A. Ng, and M. Lungren. 2018.

- Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of mrnet. *PLoS Medicine*, 15.
- Bishop, Christopher M. 2006. *Pattern recognition and machine learning*. Springer.
- Black, Nick. 2013. Patient reported outcome measures could help transform healthcare. *Bmj*, 346.
- Boissonnault, W. and M. Badke. 2005. Collecting health history information: the accuracy of a patient self-administered questionnaire in an orthopedic outpatient setting. *Physical therapy*, 85 6:531–43.
- Bowyer, Kevin W., Nitesh V. Chawla, Lawrence O. Hall, and W. Philip Kegelmeyer. 2011. SMOTE: synthetic minority over-sampling technique. *CoRR*, abs/1106.1813.
- Breiman, Leo. 2001. Random forests. *Machine learning*, 45(1):5–32.
- Brown, J. D. 1997. Statistics corner: Questions and answers about language testing statistics: Reliability of surveys. *Shiken: JALT Testing Evaluation SIG Newsletter*, 1 (2):17 – 19.
- Brownlee, J. 2020. How to Use Polynomial Feature Transforms for Machine Learning.
- Brynjolfsson, E. and Tom Mitchell. 2017. What can machine learning do? workforce implications. *Science*, 358:1530 – 1534.
- Burroughs, T., B. Waterman, J. Cira, R. Desikan, and W. Claiborne Dunagan. 2001. Patient satisfaction measurement strategies: a comparison of phone and mail methods. *The Joint Commission journal on quality improvement*, 27 7:349–61.
- Cabitzia, Federico, Angela Locoro, and Giuseppe Banfi. 2018. Machine learning in orthopedics: A literature review. *Frontiers in Bioengineering and Biotechnology*, 6:75.
- Chow, Shein-Chung, Hansheng Wang, and Jun Shao. 2007. *Sample size calculations in clinical research*. Chapman and Hall/CRC.
- Choy, G., O. Khalilzadeh, M. Michalski, Synho Do, Anthony E. Samir, O. Pianykh, J. Geis, P. Pandharipande, J. Brink, and K. Dreyer. 2018. Current applications and future impact of machine learning in radiology. *Radiology*, 288 2:318–328.
- Claassen, Aniek A. O. M., H. Schers, V. Busch, P. Heesterbeek, F. V. D. van den Hoogen, T. V. Vliet Vlieland, and C. H. M. van den Ende. 2020. Preparing for an orthopedic consultation using an ehealth tool: a randomized controlled trial in patients with hip and knee osteoarthritis. *BMC Medical Informatics and Decision Making*, 20.
- Clark, P. A. 2003. Medical practices' sensitivity to patients' needs: Opportunities and practices for improvement. *Journal of Ambulatory Care Management*, 26:110–123.
- Clinic, Cleveland. 2016. Cleveland clinic, ibm collaborate to establish model for cognitive population health management and data-driven personalized healthcare. from <https://newsroom.clevelandclinic.org/2016/12/22/cleveland-clinic-ibm-collaborate-establish-model-cognitive-population-health-management-data-driven-personalized-healthcare/>. Retrieved on 2021-03-13.
- Collen, M., J. L. Cutler, A. Siegelaub, and R. L. Cella. 1969. Reliability of a self-administered medical questionnaire. *Archives of internal medicine*, 123 6:664–81.
- Cronbach, L. 1960. Essentials of psychological testing, 2nd ed.
- Crow, R., H. Gage, S. Hampson, J. Hart, A. Kimber, L. Storey, and H. Thomas. 2002. The measurement of satisfaction with healthcare: implications for practice from a

- systematic review of the literature. *Health technology assessment*, 6 32:1–244.
- De Groot, IB, Max Reijman, CB Terwee, SMA Bierma-Zeinstra, Marein Favejee, EM Roos, and JAN Verhaar. 2007. Validation of the dutch version of the hip disability and osteoarthritis outcome score. *Osteoarthritis and cartilage*, 15(1):104–109.
- Delnoij, D., J. Rademakers, and P. Groenewegen. 2010. The dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Services Research*, 10:88 – 88.
- Deo, Rahul C. 2015. Machine learning in medicine. *Circulation*, 132 20:1920–30.
- Devine, E. 1992. Effects of psychoeducational care for adult surgical patients: a meta-analysis of 191 studies. *Patient education and counseling*, 19 2:129–42.
- DeVries, Zachary, Mohamad Hoda, C. Rivers, Audrey Maher, E. Wai, Dita Moravek, Alexandra Stratton, S. Kingwell, Nader Fallah, J. Paquet, and P. Phan. 2019. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *The spine journal : official journal of the North American Spine Society*.
- Döring, Matthias. 2018. Performance measures for multi-class problems. *Data Science Blog*, December, 4.
- Durmaz, Ahmet, Inci Dursun, and Ebru Tümer Kabadayi. 2020. Mitigating the effects of social desirability bias in self-report surveys: Classical and new techniques.
- Ekdahl, A., L. Andersson, and M. Friedrichsen. 2010. "they do what they think is the best for me." frail elderly patients' preferences for participation in their care during hospitalization. *Patient education and counseling*, 80 2:233–40.
- Epomedicine. 2014. History taking in orthopedics. from <https://epomedicine.com/clinical-medicine/history-taking-in-orthopedics/>. Retrieved on 2021-01-18.
- Epstein, O, GD Perkin, and J Cookson. 2008. *Clinical Examination. Fourth edition*. Mosby, Edinburgh.
- Espinel, Alexandra G., R. Shah, M. McCormick, P. Krakovitz, and E. Boss. 2014. Patient satisfaction in pediatric surgical care. *Otolaryngology—Head and Neck Surgery*, 150:739 – 749.
- Fawcett, Tom. 2006. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.
- Fenech, Strukelj N., M. and O. Buston. 2018. Ehtical, social, and political challenges of artificial intelligence in health. Technical report, Future Advocacy. Retrieved on 2021-03-18.
- Ferguson, M., M. Brandreth, W. Brassington, and H. Wharrad. 2015. Information retention and overload in first-time hearing aid users: An interactive multimedia educational solution. *American journal of audiology*, 24 3:329–32.
- Fischer, Martin and Gill Ereaut. 2012. When doctors and patients talk: making sense of the consultation. Technical report, The Health Foundation.
- Fontana, M., S. Lyman, G. K. Sarker, D. Padgett, and C. MacLean. 2019. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clinical Orthopaedics and Related Research*, 477:1267–1279.

- Fottler, M., R. Ford, and S. Bach. 1997. Measuring patient satisfaction in healthcare organizations: qualitative and quantitative approaches. *Best practices and benchmarking in healthcare : a practical journal for clinical and management application*, 2 6:227–39.
- Freund, Yoav and Robert E Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.
- Friedman, Jerome H. 2002. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378.
- Galbusera, Fabio, Gloria Casaroli, and T. Bassani. 2019. Artificial intelligence and machine learning in spine research. *JOR Spine*, 2.
- Gask, L. and T. Usherwood. 2002. The consultation. *BMJ : British Medical Journal*, 324:1567 – 1569.
- Ghomrawi, H., C. Mancuso, A. Dunning, A. González Della Valle, M. Alexiades, Charles Cornell, T. Sculco, Matthias Bostrom, D. Mayman, R. Marx, G. Westrich, M. O'dell, and A. Mushlin. 2017. Do surgeon expectations predict clinically important improvements in womac scores after tha and tka? *Clinical Orthopaedics and Related Research®*, 475:2150–2158.
- Gilbert, F., S. Astley, M. G. Gillan, O. Agbaje, M. Wallis, Jonathan J James, C. Boggis, and S. Duffy. 2008. Single reading with computer-aided detection for screening mammography. *The New England journal of medicine*, 359 16:1675–84.
- Gilkison, C., M. Fenton, and J. Lester. 1992. Getting the story straight: evaluating the test-retest reliability of a university health history questionnaire. *Journal of American college health : J of ACH*, 40 6:247–52.
- Grandini, Margherita, Enrico Bagli, and Giorgio Visani. 2020. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.
- Greene, J., J. Hibbard, R. Sacks, and Valerie Overton. 2013. When seeing the same physician, highly activated patients have better care experiences than less activated patients. *Health affairs*, 32 7:1299–305.
- Gruson, D., Julia Petrelluzzi, Judith Mehl, A. Burgun, and N. Garcelon. 2018. [ethical, legal and operational issues of artificial intelligence]. *La Revue du praticien*, 68 10:1145–1148.
- Gumpel, J. and A. M. Mason. 1974. Self-administered clinical questionnaire for outpatients. *British Medical Journal*, 2:209 – 212.
- Guo, Yutao, Y. Chen, D. Lane, L. Liu, Y. Wang, and G. Lip. 2017. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: maf app trial. *The American journal of medicine*, 130 12:1388–1396.e6.
- Gupta, A., A. Potty, D. Ganta, R. J. Mistovich, Sreeram Penna, C. Cady, and Anish G R Potty. 2020. Streamlining the koos activities of daily living subscale using machine learning. *Orthopaedic Journal of Sports Medicine*, 8.
- Guyatt, G., A. Mitchell, D. Molloy, R. Capretta, J. Horsman, and L. Griffith. 1995. Measuring patient and relative satisfaction with level or aggressiveness of care and involvement in care decisions in the context of life threatening illness. *Journal of clinical epidemiology*, 48 10:1215–24.

- Haeberle, Heather S., J. M. Helm, S. Navarro, J. Karnuta, J. Schaffer, J. Callaghan, Michael A Mont, A. Kamath, V. Krebs, and P. Ramkumar. 2019. Artificial intelligence and machine learning in lower extremity arthroplasty: A review. *The Journal of arthroplasty*.
- Hall, G. 1972. Experiences with outpatient medical questionnaires. *British Medical Journal*, 1:42 – 45.
- Han, Xiao-Guang and W. Tian. 2019. Artificial intelligence in orthopedic surgery: current state and future perspective. *Chinese Medical Journal*, 132:2521 – 2523.
- Hand, David J and Robert J Till. 2001. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine learning*, 45(2):171–186.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Haverkamp, Daniël, Stefan JM Breugem, Inger N Sierevelt, Leendert Blankevoort, and C Nick van Dijk. 2005. Translation and validation of the dutch version of the oxford 12-item knee questionnaire for knee arthroplasty. *Acta orthopaedica*, 76(3):347–352.
- Hawthorne, G., J. Sansoni, L. Hayes, N. Marosszky, and E. Sansoni. 2014. Measuring patient satisfaction with health care treatment using the short assessment of patient satisfaction measure delivered superior and robust satisfaction estimates. *Journal of clinical epidemiology*, 67 5:527–37.
- He, Haibo and Yunqian Ma. 2013. Imbalanced learning: foundations, algorithms, and applications.
- Helm, J. M., Andrew M. Swiergosz, Heather S. Haeberle, J. Karnuta, J. Schaffer, V. Krebs, A. Spitzer, and P. Ramkumar. 2020. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, 13:69–76.
- Hollon, Todd C., B. Pandian, S. S. Khalsa, R. D'Amico, Michael B. Sisti, J. Bruce, K. Muraszko, B. Thompson, C. Freudiger, Shawn L. Hervey-Jumper, G. McKhann, S. Camelo-Piragua, P. Canoll, H. Lee, and D. Orringer. 2019. Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks. *Neurosurgery*, 66.
- IBM Corp. 2019. Ibm spss statistics for windows [computer software].
- Interactive-Studios. 2020. Online proms. <https://onlineproms.nl/>. Retrieved on 2021-01-29.
- Interactive-Studios. 2021. Patient journey app: Improve healthcare patient engagement. <https://patientjourneyapp.com/>. Retrieved on 2021-01-18.
- Intersoft Consulting. 2021. Gdpr consent. <https://gdpr-info.eu/issues/consent/>. Retrieved on 2021-03-17.
- Inui, T., R. A. Jared, W. Carter, D. Plorde, R. Pecoraro, M. Chen, and J. J. Dohan. 1979. Effects of a self-administered health history on new-patient visits in a general medical clinic. *Medical Care*, 17:1221–1228.
- Jamaludin, A., Meelis Lootus, T. Kadir, Andrew Zisserman, J. Urban, M. C. Battié, J. Fairbank, I. McCall, and The Genodisc Consortium. 2017a. Issls prize in bioengineering science 2017: Automation of reading of radiological features from magnetic resonance images (mrис) of the lumbar spine without human intervention is comparable with an expert radiologist. *European Spine Journal*, 26:1374–1383.

- Jamaludin, A., Meelis Lootus, T. Kadir, Andrew Zisserman, J. Urban, M. C. Battié, J. Fairbank, I. Mccall, and The Genodisc Consortium. 2017b. Issls prize in bioengineering science 2017: Automation of reading of radiological features from magnetic resonance images (mrис) of the lumbar spine without human intervention is comparable with an expert radiologist. *European Spine Journal*, 26:1374–1383.
- Jansen, J., J. V. van Weert, Nienke van der Meulen, S. van Dulmen, T. Heeren, and J. Bensing. 2008. Recall in older cancer patients: measuring memory for medical information. *The Gerontologist*, 48 2:149–57.
- Jayakumar, P., M. Moore, and K. Bozic. 2019. Value-based healthcare: Can artificial intelligence provide value in orthopaedic surgery? *Clinical orthopaedics and related research*, 477 8:1777–1780.
- Jha, S. and E. Topol. 2016. Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA*, 316 22:2353–2354.
- Jović, Alan, Karla Brkić, and Nikola Bogunović. 2015. A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 1200–1205, Ieee.
- Kahn, G. 1993. Computer-based patient education: a progress report. *M.D. computing : computers in medical practice*, 10 2:93–9.
- Katz, J., L. Chang, O. Sangha, A. Fossel, and D. Bates. 1996. Can comorbidity be measured by questionnaire rather than medical record review? *Medical care*, 34 1:73–84.
- Kaufman, G. 2008. Patient assessment: effective consultation and history taking. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, 23 4:50–6, quiz 58, 60.
- Kawabata, H., K. Konishi, Manabu Murakami, K. Kisa, and M. Maezawa. 2009. Factors affecting the physician-patient relationship regarding patient participation in medical encounters in primary care. *[Hokkaido igaku zasshi] The Hokkaido journal of medical science*, 84 3:171–5.
- Kessels, R. 2003. Patients' memory for medical information. *Journal of the Royal Society of Medicine*, 96 5:219–22.
- King, Gary and Langche Zeng. 2001. Logistic regression in rare events data. *Political analysis*, 9(2):137–163.
- Kotti, M., Lynsey D. Duffell, Aldo A. Faisal, and A. McGregor. 2017. Detecting knee osteoarthritis and its discriminating parameters using random forests. *Medical Engineering Physics*, 43:19 – 29.
- Krittanawong, Chayakrit, H. Virk, S. Bangalore, Z. Wang, K. Johnson, Rachel Pinotti, Hongju Zhang, S. Kaplin, B. Narasimhan, T. Kitai, U. Baber, J. Halperin, and W. Tang. 2020. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Scientific Reports*, 10.
- Kuhn, Max, Kjell Johnson, et al. 2013. *Applied predictive modeling*, volume 26. Springer.
- Kyle, S. and D. Shaw. 2014. Doctor–patient communication, patient knowledge and health literacy: how difficult can it all be? *Bulletin of The Royal College of Surgeons of England*, 96.
- KyLeads. n. d. What is the central tendency bias?
<https://www.kyleads.com/surveys/central-tendency-bias/>. Retrieved

- on 2021-04-16.
- Larsen, D., C. Attkisson, W. Hargreaves, and T. D. Nguyen. 1979. Assessment of client/patient satisfaction: development of a general scale. *Evaluation and program planning*, 2 3:197–207.
- Lee, H., R. Ghebre, C. Le, Y. Jang, M. Sharratt, and D. Yee. 2017. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: Pilot randomized controlled trial. *JMIR mHealth and uHealth*, 5.
- Lemaître, Guillaume, Fernando Nogueira, and Christos K. Aridas. 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5.
- LeMoine, R., T. Mastroianni, A. Hessel, and K. Nishikawa. 2015. Implementation of machine learning for classifying prosthesis type through conventional gait analysis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 202–205.
- LeMoine, R., T. Mastroianni, Anthony L. Hessel, and K. Nishikawa. 2015. Implementation of machine learning for classifying prosthesis type through conventional gait analysis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 202–205.
- Lennartz, S., Thomas Dratsch, D. Zopfs, T. Persigehl, D. Maintz, N. Große Hokamp, and D. Pinto Dos Santos. 2021. Use and control of artificial intelligence in patients across the medical workflow: Single-center questionnaire study of patient perspectives. *Journal of Medical Internet Research*, 23.
- Li, C., Shou jiang Huang, X. Su, Tingting Zhang, and Ke dong Jiang. 2019. Monitoring of home recovery using the 317-nursing mobile application following day-case surgery in children. *Medicine*, 98.
- Lindsey, Robert, A. Daluiski, S. Chopra, A. Lachapelle, M. Mozer, S. Sicular, D. Hanel, M. Gardner, A. Gupta, R. Hotchkiss, and H. Potter. 2018. Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences of the United States of America*, 115:11591 – 11596.
- Liu, Y., Po-Hsuan Cameron Chen, Jonathan Krause, and Lily Peng. 2019. How to read articles that use machine learning: Users' guides to the medical literature. *JAMA*, 322 18:1806–1816.
- Loh, Wei-Yin. 2011. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23.
- du Long, Jasmijn, Michiel Hageman, Dick Vuijk, A. Rakic, and D. Haverkamp. 2016. Facing the decision about the treatment of hip or knee osteoarthritis: What are patients' needs? *Knee Surgery, Sports Traumatology, Arthroscopy*, 24:1710–1716.
- Luo, Wei, Dinh Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, Alistair Shilton, J. Yearwood, N. Dimitrova, T. Ho, S. Venkatesh, and M. Berk. 2016. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *Journal of Medical Internet Research*, 18.
- Maffulli, N., Hugo C. Rodriguez, I. Stone, Andrew Nam, Albert P. Song, M. Gupta, Rebecca Alvarado, David Ramón, and Ashim Gupta. 2020. Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol. *Journal of Orthopaedic Surgery and Research*, 15.

- Makhni, E., Sonya Makhni, and P. Ramkumar. 2020. Artificial intelligence for the orthopaedic surgeon: An overview of potential benefits, limitations, and clinical applications. *The Journal of the American Academy of Orthopaedic Surgeons*.
- Marks, Jennifer R, J. Schechtman, H. Groninger, and Margaret Plews-Ogan. 2010. The association of health literacy and socio-demographic factors with medication knowledge. *Patient education and counseling*, 78 3:372–6.
- Matheny, M., ST. Israni, M. Ahmed, and D. Whicher. 2019. Artificial intelligence in health care: The hope, the hype, the promise, the peril. <https://nam.edu/artificial-intelligence-special-publication/>. Retrieved on 2021-03-16.
- Mccall, Becky. 2018. What does the gdpr mean for the medical community? *The Lancet*, 391:1249–1250.
- McGuire, L. 1996. Remembering what the doctor said: organization and adults' memory for medical information. *Experimental aging research*, 22 4:403–28.
- Meij, E. V. D., J. Anema, W. Leclercq, M. Bongers, and J. Huirne. 2018. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. *The Lancet*, 392:51–59.
- Moons, K., D. Altman, J. Reitsma, J. Ioannidis, P. Macaskill, E. Steyerberg, A. Vickers, D. Ransohoff, and G. Collins. 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. *Annals of Internal Medicine*, 162:W1–W73.
- Morris, B., J. Richards, K. Archer, Melissa Lasater, Denise Rabalais, M. Sethi, and A. A. Jahangir. 2014. Improving patient satisfaction in the orthopaedic trauma population. *Journal of Orthopaedic Trauma*, 28:e80–e84.
- Mostert, M., A. Bredenoord, M. Biesaart, and J. Delden. 2016. Big data in medical research and eu data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics*, 24:1096–1096.
- Myers, T. G., P. Ramkumar, Benjamin F Ricciardi, K. Urish, J. Kipper, and C. Ketonis. 2020. Artificial intelligence and orthopaedics: An introduction for clinicians. *The Journal of bone and joint surgery. American volume*.
- Oh, H., C. Rizo, M. Enkin, and A. Jadad. 2005. What is ehealth (3): A systematic review of published definitions. *Journal of Medical Internet Research*, 7.
- Oseledets, Ivan V and Eugene E Tyrtyshnikov. 2009. Breaking the curse of dimensionality, or how to use svd in many dimensions. *SIAM Journal on Scientific Computing*, 31(5):3744–3759.
- Parikh, Ravi B, Z. Obermeyer, and A. Navathe. 2019. Regulation of predictive analytics in medicine. *Science*, 363:810 – 812.
- Pecoraro, R., T. Inui, M. Chen, D. K. Plorde, and J. L. Heller. 1979. Validity and reliability of a self-administered health history questionnaire. *Public health reports*, 94 3:231–8.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

- Phinyomark, A., S. Osis, B. Hettinga, D. Kobsar, and R. Ferber. 2016. Gender differences in gait kinematics for patients with knee osteoarthritis. *BMC Musculoskeletal Disorders*, 17.
- Pincus, T., Juan Schmukler, and I. Castrejón. 2019. Patient questionnaires in osteoarthritis: what patients teach doctors about their osteoarthritis on a multidimensional health assessment questionnaire (mdhaq) in clinical trials and clinical care. *Clinical and experimental rheumatology*, 37 Suppl 120 5:100–111.
- Plotly Technologies Inc. 2015. Collaborative data science.
- Poduval, M., A. Ghose, Sanjeev Manchanda, V. Bagaria, and A. Sinha. 2020. Artificial intelligence and machine learning: A new disruptive force in orthopaedics. *Indian journal of orthopaedics*, 54 2:109–122.
- Pogorelc, B. and M. Gams. 2010. Diagnosing health problems from gait patterns of elderly. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2238–2241.
- Prasoon, A., Kersten Petersen, C. Igel, F. Lauze, E. Dam, and Mads Nielsen. 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16 Pt 2:246–53.
- Pretlow, Robert. 2000. ehealth international: A cutting edge company for a new age in health care. <http://www.ehealthnurse.com/ehealthi.html>. Retrieved on 2021-04-12.
- Price, W. and I. Cohen. 2019. Privacy in the age of medical big data. *Nature Medicine*, 25:37–43.
- Probst, Pascal, Kathrin Grummich, Patrick Heger, Steffen Zaschke, Phillip Knebel, Alexis Ulrich, Markus W Büchler, and Markus K Diener. 2016. Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. *Systematic reviews*, 5(1):1–6.
- Racine, E., Wren Boehlen, and M. Sample. 2019. Healthcare uses of artificial intelligence: Challenges and opportunities for growth. *Healthcare Management Forum*, 32:272 – 275.
- Ramkumar, P., J. Karnuta, S. Navarro, Heather S. Haeberle, R. Iorio, Michael A. Mont, B. Patterson, and V. Krebs. 2019. Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: Development and validation of a deep learning model. *The Journal of arthroplasty*.
- Rohatgi, Jitesh. 2018. Gdpr and healthcare: Understanding health data and consent. <https://www.pega.com/insights/articles/gdpr-and-healthcare-understanding-health-data-and-consent>. Retrieved on 2021-03-17.
- Ross, Brian C. 2014. Mutual information between discrete and continuous data sets. *PloS one*, 9(2):e87357.
- Roy, B. 2020. All about Feature Scaling - Towards Data Science.
- Salaffi, Fausto, Andrea Stancati, Carlo Alberto Silvestri, Alessandro Ciapetti, and Walter Grassi. 2004. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *European journal of pain*, 8(4):283–291.

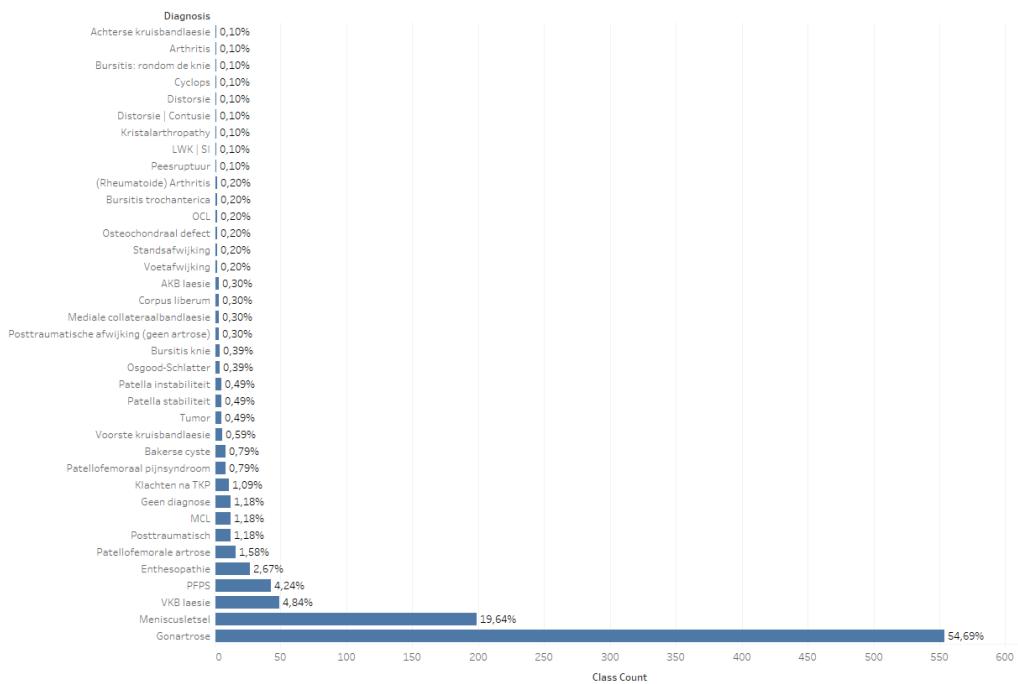
- Sambare, Tanmaya D., L. Uhler, and K. Bozic. 2017. Shared decision making: Time to get personal.
- Santos, D. Pinto Dos, D. Giese, S. Brodehl, S. Chon, W. Staab, R. Kleinert, D. Maintz, and B. Baessler. 2018. Medical students' attitude towards artificial intelligence: a multicentre survey. *European Radiology*, 29:1640–1646.
- Say, R., M. Murtagh, and R. Thomson. 2006. Patients' preference for involvement in medical decision making: a narrative review. *Patient education and counseling*, 60 2:102–14.
- Schulz, KF., DG. Altma, and D. Moher. 2010. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *Ann Int Med*, 2010(152).
- Seabold, Skipper and Josef Perktold. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.
- Seli, P., I. Svab, Marija Repolusk, and N. K. Guek. 2011. What factors affect patients' recall of general practitioners' advice? *BMC Family Practice*, 12:141 – 141.
- Shapiro, Samuel Sanford and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Shatte, Adrian B R, D. Hutchinson, and Samantha J Teague. 2019. Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, 49 9:1426–1448.
- Shirley, E., G. Josephson, and J. Sanders. 2016. Fundamentals of patient satisfaction measurement. *Physician leadership journal*, 3 1:12–7.
- Shue, J., R. Karia, Dennis A. Cardone, J. Samuels, M. Shah, and James D. Slover. 2016. A randomized controlled trial of two distinct shared decision-making aids for hip and knee osteoarthritis in an ethnically diverse patient population. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 19 4:487–93.
- Sitzia, J. 1999. How valid and reliable are patient satisfaction data? an analysis of 195 studies. *International journal for quality in health care : journal of the International Society for Quality in Health Care*, 11 4:319–28.
- Smith, R. 2003. Thoughts for new medical students at a new medical school. *BMJ : British Medical Journal*, 327:1430 – 1433.
- Smits, C., Hugo K van den Beld, M. Aartsen, and J. Schroots. 2014. Aging in the netherlands: state of the art and science. *The Gerontologist*, 54 3:335–43.
- Spampinato, C., S. Palazzo, D. Giordano, Marco Aldinucci, and Rosalia Leonardi. 2017. Deep learning for automated skeletal bone age assessment in x-ray images. *Medical Image Analysis*, 36:41–51.
- Stacey, D., F. Légaré, K. Lewis, M. Barry, C. Bennett, Karen B Eden, M. Holmes-Rovner, H. Llewellyn-Thomas, A. Lyddiatt, R. Thomson, and L. Trevena. 2017. Decision aids for people facing health treatment or screening decisions. *The Cochrane database of systematic reviews*, 4:CD001431.
- Tableau, Software. 2021. Business intelligence and analytics software. <https://www.tableau.com/>. Retrieved on 2021-06-19.
- Tait, A., Magnus K. Teig, and T. Voepel-Lewis. 2014. Informed consent for anesthesia: a review of practice and strategies for optimizing the consent process. *Canadian Journal*

- of Anesthesia/Journal canadien d'anesthésie*, 61:832–842.
- Thomas, Karen M and K. Sethares. 2008. An investigation of the effects of preoperative interdisciplinary patient education on understanding postoperative expectations following a total joint arthroplasty. *Orthopaedic Nursing*, 27:374–381.
- Thomsen, K., L. Iversen, Therese Louise Titlestad, and O. Winther. 2019. Systematic review of machine learning for diagnosis and prognosis in dermatology. *Journal of Dermatological Treatment*, 31:496 – 510.
- Tidy, C. 2019. History taking. from <https://patient.info/doctor/history-taking#ref-4>. Retrieved on 2021-01-28.
- Timmers, Thomas, L. Janssen, R. Kool, and J. Kremer. 2020. Educating patients by providing timely information using smartphone and tablet apps: Systematic review. *Journal of Medical Internet Research*, 22.
- Timmers, Thomas, L. Janssen, Y. Pronk, B. C. van der Zwaard, S. Koëter, D. van Oostveen, Stefan de Boer, Keetie Kremers, S. Rutten, D. Das, R. V. van Geenen, K. Koenraadt, Rob Kusters, and W. van der Weegen. 2018. Assessing the efficacy of an educational smartphone or tablet app with subdivided and interactive content to increase patients' medical knowledge: Randomized controlled trial. *JMIR mHealth and uHealth*, 6.
- Tobin-Schnittger, Patrick, J. O'Doherty, R. O'Connor, and A. O'Regan. 2018. Improving quality of referral letters from primary to secondary care: a literature review and discussion paper. *Primary health care research development*, 19 3:211–222.
- Urbaniak, G. C. and S. Plous. 2013. Research randomizer (version 4.0) [computer software]. from <http://www.randomizer.org/>. Retrieved on 2021-01-18.
- Urden, L. 2002. Patient satisfaction measurement: current issues and implications. *Outcomes management*, 6 3:125–31.
- Van Rossum, Guido and Fred L. Drake. 2009. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Vayena, E., A. Blasimme, and I. Cohen. 2018. Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, 15.
- Ware, J., M. K. Snyder, W. R. Wright, and A. Davies. 1983. Defining and measuring patient satisfaction with medical care. *Evaluation and program planning*, 6 3-4:247–63.
- Waters, Stuart, S. Edmondston, P. Yates, and D. Gucciardi. 2016. Identification of factors influencing patient satisfaction with orthopaedic outpatient clinic consultation: A qualitative study. *Manual therapy*, 25:48–55.
- While, A. 2002. Practical skills: prescribing consultation in practice. *British journal of community nursing*, 7 9:469–73.
- Wills, Jane. 2008. Health literacy: new packaging for health education or radical movement? *International Journal of Public Health*, 54:3–4.
- Yoo, T., S. K. Kim, D. Kim, J. Y. Choi, W. Lee, E. Oh, and E. Park. 2013. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. *Yonsei Medical Journal*, 54:1321 – 1330.
- Zhang, Zhongheng. 2016. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11).

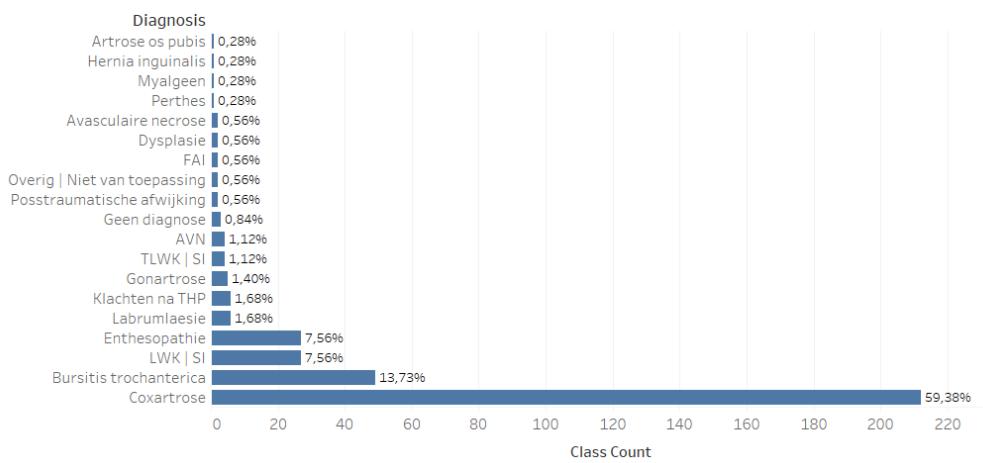
Zupan, B., J. Demsar, D. Smrke, K. Boikov, V. Stankovski, I. Bratko, and J. R. Beck. 2001. Predicting patient's long-term clinical status after hip arthroplasty using hierarchical decision modelling and data mining. *Methods of information in medicine*, 40 1:25–31.

Appendices

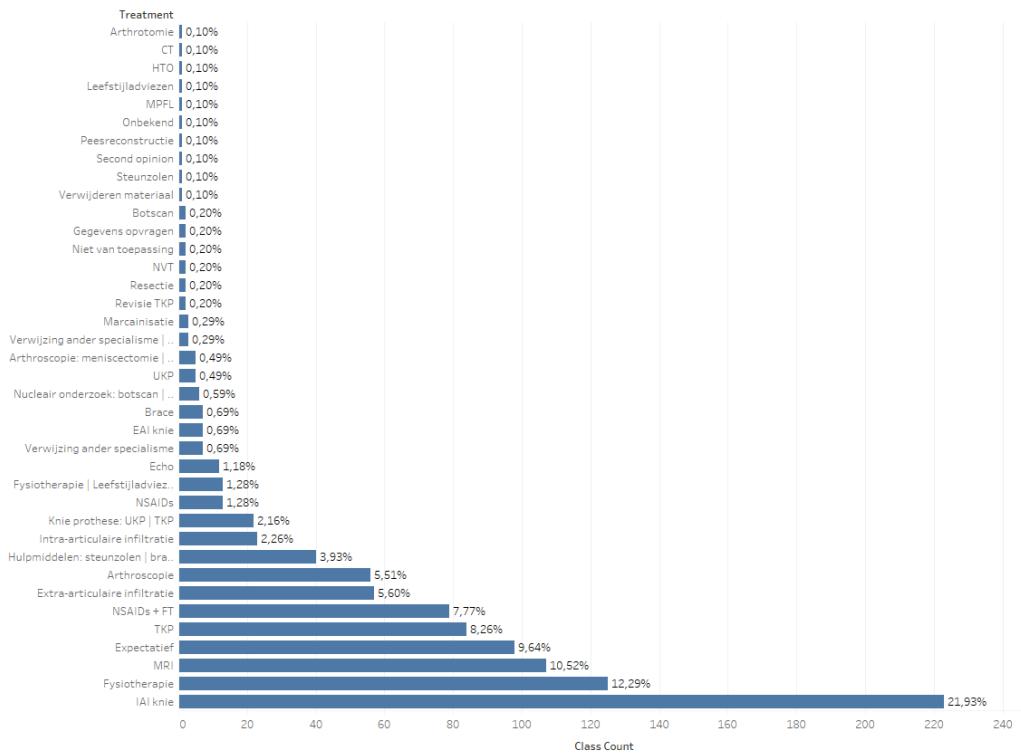
A Original Distribution of Knee Diagnoses



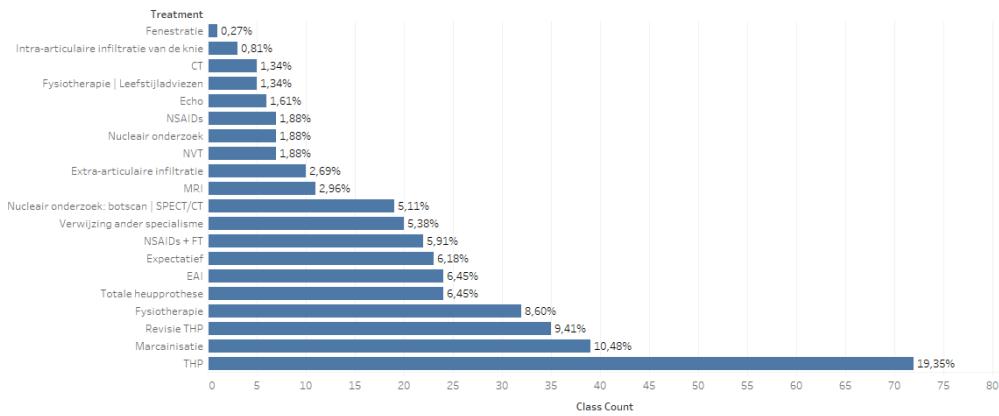
B Original Distribution of Hip Diagnoses



C Original Distribution of Knee Treatments



D Original Distribution of Hip Treatments



E Selected hyperparameters

Algorithm	Hyperparameter	Value
SVM - Hip Treatment	kernel	linear
	degree	1
	decision_function_shape	ovo
	C	0.1
Random Forest - Hip diagnosis and Knee Treatment	n_estimators	100
	min_samples_split	2
	min_samples_leaf	1
	max_features	auto
	max_depth	3
Knee Diagnosis - Adaboost	n_estimators	800
	learning_rate	0.1

F METC n-WMO Applicaton form



Application form nWMO METC Máxima MC

With the help of the information provided on this application form, the METC Máxima MC determines whether the intended research is covered by the Medical-Scientific Research with Humans Act (WMO). No substantive review is carried out; the METC does not assess the study for relevance, quality and conformity with other applicable laws and regulations.

Send the fully completed and signed form (please in PDF format) by e-mail to the METC Máxima MC: metc@mmc.nl. Don't forget to send any attachments.

Please note that a *non-WMO declaration* is not a permission for the execution of the study. To do this, the institution must follow the appropriate procedures (within MMC, the study is reported to the Committee on [Local Feasibility Research](#)).

1. Contact details

Removed due to privacy

2. General research

- a. Title research: Experienced quality of an orthopaedic consultation prepared with digital history and prediction of diagnosis and treatment
- b. acronym: N/a
- c. Brief summary: During an outpatient consultation, the differential diagnosis and appropriate treatment are primarily determined on the basis of the history taking and confirmed with physical examination and any evaluation of medical imaging Results. The performing orthopaedic surgeon must perform these tasks within the short duration of a consultation and thus communicate it to the patient as clearly as possible. The experienced quality of the consultation by the patient depends on many factors such as the duration of the consultation, confidence in the orthopaedic surgeon, empathy of the orthopedic surgeon, communication, and the expectations of the patient (Waters et al., 2016).

For a patient, a consultation can be intimidating while they receive a lot of information in a short time. Research has shown that patients have difficulty remembering the information offered to them during consultations, and in particular medical information (Mcguire, 1996). In addition, about half of what patients remember appears to be inaccurate (Kessels, 2003). Another problem faced by patients is that the information presented during



a consultation is often poorly understood (Wills, 2009). Given the number of patients who need orthopaedic care, there is little to no possibility to extend the duration of the consultation with the orthopedic surgeon. However, according to orthopedists within St. Anna, the above factors have a negative impact on the ability to provide information to the patient and meet the patient's expectations. In addition, as mentioned earlier, the orthopaedic surgeon has the complicated task of determining the differential diagnosis and appropriate treatment within the short duration of a consultation based on the history and physical examination, and the results from here to the patient. According to orthopedists within St. Anna, the time pressure ensures that they cannot always prepare well for a consultation, which can compromise factors such as empathy and good communication.

In other words, the quality of an orthopaedic consultation can be improved. A previous study conducted within St. Anna hospital has examined whether patients' medical knowledge can be increased by offering personalized information to osteoarthritis patients before a consultation takes place (Timmers et al., 2018). This has shown that the provision of personalized information to osteoarthritis patients before the consultation takes place has a positive impact on both the real and the self-reported medical knowledge of the patient. It is expected that increasing medical knowledge will also increase the experienced quality of the clinical consultation by the patient because the expectations of the patient are more realistic and because the patient is better able to communicate with the orthopedic surgeon.

In addition to the study of Timmers et al. (2018) in which patients with a high risk of osteoarthritis of the knee were selected in advance, personalized information will be provided to patients in the current study by applying machine learning techniques based on the most likely diagnosis. These machine learning techniques can be used to predict diagnoses and treatments and these predictions allow personalized information to be offered to the patient via a mobile application. The models are created based on already implemented digital history taking questionnaires (usual care) which are filled in by patients before they visit the outpatient clinic. In addition to providing personalized information to patients, the digital history questionnaire and the predictions can be used to support the orthopedic surgeon. They can use the information to prepare the consultation and it is expected that the orthopaedic surgeon will be able to give a better consultation as a result.

In conclusion, the goal of St. Anna is to improve the patient-experienced quality of outpatient consultations. For this we use new machine learning techniques that we apply to an already existing working method, namely the taking of digital history questionnaires prior to the first consultation. With this we provide less general but more personalized information to patients and support the orthopedic surgeon.

- d. Research objectives: With this study we want to investigate whether taking digital history questionnaires and providing personalized information to patients improves the experienced quality of an outpatient consultation.
- e. Intended start date: 1-4-2021
- f. Intended end date: 1-10-2021
- g. In your opinion, the research concerns a medical-scientific study according to [the definition of the CCMO](#) ('Medical-scientific research is research that aims to answer a question in the field of disease and health (*etiology, pathogenesis, symptoms/symptoms, diagnosis,*



prevention, outcome or treatment of disease), by systematically collecting and studying data. The research aims to contribute to medical knowledge that also applies to populations outside the direct research population.'

yes

Doubt, because: To enter text, click here.

No, because:

h. What kind of research is it?

- Data research *patient-related*
 - Retrospective data from status research/SPD
 - Retrospective research footage
 - Prospective data from status/EPD (also phase IV medicines)
 - Data from questionnaires, interviews
- Data research *volunteer-related*
 - Testing laboratory determinations
 - Testing equipment
 - Validate methods of research
- Data research *non-personal*
 - Quality research practice
 - Epidemiological studies
 - National registration systems
- Biobank research
 - Residual material
 - Material extra purchased in diagnostics
 - Material specially collected from patients or volunteers
 - Material made available by data subjects themselves (donor banks)
- Otherwise, namely: To enter text, click here.

3. Subjects / participants

a. Are the subjects themselves involved in this study?

yes

No, the subject does not have to do anything himself, for example retrospective dossier research)

b. Are individuals subjected to acts or are they subjected to rules of conduct (including taking questionnaires/interviews, please add them as an attachment)?

No, because: To enter text, click here.

Yes, namely the following actions/rules of conduct that are not done in the context of the standard treatment (indicate what the actions/rules of conduct best and how much time the participants spend with this in the total study): This is a patient-related data research that matches the 'usual care' process. In St. Anna's, the usual care around the first consultation is that patients fill in a digital history and that they prepare for the first consultation with the help of an app. This study extends this with questionnaires related to the actual and self-reported medical knowledge of the patient. It is estimated that the patient spends a total of



15 minutes on this. The general preparatory information shall be replaced by personalised information in the intervention group. In addition, a questionnaire about the experienced quality regarding the consultation is taken 1 day after the consultation. It is estimated that the patient spends 5 minutes on this.

- c. What's blood taken before this test?
 - no
 - Yes, from an existing line or planned vena puncture (indicate how often and how many milliliters):
 - Yes, from an extra prick action: times milliliters (indicate how often and how many milliliters):
- d. What are the characteristics of the research population (disease picture, target group, inclusion / exclusion criteria, type of condition, etc.)?

Patients over 18 years of age with complaints to the knee. In addition, they must speak Dutch, be in possession of a e-mail address, own a smartphone or tablet, and want to participate in the study.
- e. What is the age of the target group?
 - Children under 12 years
 - Children between 12 and 15 years
 - Young people between the age of 16 and 18
 - Adults 18 years and older
- f. Are all subjects capable of will?
 - yes
 - No, because:
- g. How many subjects are included in the overall study? For estimating/calculating the required sample Size we assume a two-sample one-sided proportion hypothesis in which we compare whether the proportion of subjects in the experimental group who have all quality-related questions with answering well, very well or excellently is significantly greater than the proportion of subjects in the control group that all quality-related questions with good, very good or excellent answer. Based on pilot data (n=322) is the current proportion patients who are all quality-related questions with good, very good or excellent answer 45%. We estimate that this proportion is increased by 20% in the intervention group. With a sampling ratio of 1, the minimum sample size is equal to 204. Based on an estimated 10% drop-out rate, we arrive at a sample size of 230 patients.

4. Recruitment / consent

- a. Is *informed consent* requested from the participants?
 - Yes, see the attachment for the information letter and the corresponding consent form
 - No*, because:
*without the participant's consent, scientific research or research in the field of public health can be carried out if 1) asking for consent is not reasonably possible or can be required and that there are guarantees that the patient's privacy is not harmed EN 2) that the study serves a public interest AND 3)



that the examination cannot be carried out without the patient concerned AND 4) than the patient has not objected. In all other cases, participants must be asked for permission.

- b. Briefly describe how the recruitment is going (how and by whom the participant is informed about the research and who asks the participant for permission to participate in the study):

Eligible patients will be asked to participate in this research after they have made an appointment for an outpatient consultation. The suitability of patients will be assessed during the patient's first contact with the hospital for planning a clinical consultation. Patients who are interested will receive an email with all the necessary information about the study for informed Consent. Patients will have a period of two days to process this information. When patients have questions, they can contact the lead or co-investigator. Patients who no longer have questions can indicate their preference in an online consent form. This form will also emphasize that the patient's data will be kept confidential and secure.

5. Medical Device

- a. Is it an examination with a medical device?
- No (continue with part 6)
- Yes, the medical device has a CE marking and is used in this study for the intended use for which the CE marking was issued (continue with question 6)
- Yes, the medical device has a CE marking but is **NOT** used in this study for the intended use for which the CE marking has been issued
- Yes, the medical device does not (yet) have a CE marking
- b. My research must be reported to the Health and Youth Inspectorate (IGJ).
- Yes, so I ask the METC Máxima MC to give a positive opinion.
- no

6. Additional information

- a. Here you can provide additional information that may be important for the METC to assess the study: Because there is only a slight burden for the patient and little to no risk for the patient when participating in this study, we ask the METC to exempt this study from the WMO obligation and that there should be no additional insurance necessary for carrying out this examination.



7. Checklist and signature

- a. The following Annexes* have been added to this application:
 The protocol of the intended study with version/date
 The subject information plus the consent form
 Questionnaires, diaries, interview schedule, topic list, etc., namely:

* These annexes are not assessed in substance.

- b. This application form has been completed carefully and truthfully.

Applicant name: Tristan Warren

date: 14-12-2020

Signature submitter:

Removed due to privacy

Send your request to metc@mmc.nl

G Study Protocol

Experienced quality of an orthopaedic consultation prepared with digital history taking and prediction of diagnosis and treatment

Study title: Experienced quality of an orthopaedic consultation prepared with digital history taking and prediction of diagnosis and treatment

Principal investigator
Bsc. T. Warren, Trainee Knowledge Centre Sport and Exercise, St. Anna hospital, Geldrop.

Co-researcher:
Dr. W. van Weegen, Knowledge Centre Sport and Exercise, St. Anna hospital, Geldrop.

Date: 14 December 2020

Introduction

During an outpatient consultation, the differential diagnosis and appropriate treatment are primarily determined based on the history taking of the patient and confirmed with physical examination and possible evaluation of medical imaging results. The orthopedic surgeon must perform these tasks within the short duration of a consultation and communicate the results to the patient as clearly as possible. The experienced quality of the consultation by the patient depends on many factors such as the duration of the consultation, confidence in the orthopaedic surgeon, empathy of the orthopaedic surgeon, communication and the expectations of the patient (Waters et al., 2016).

For a patient, a consultation can be intimidating while receiving a lot of information in a short time. Research has shown that patients have difficulty remembering the information offered to them during consultations, and in particular medical information (McGuire, 1996). In addition, about half of what patients remember appears to be inaccurate (Kessels, 2003). Another problem faced by patients is that the information presented during a consultation is often poorly understood (Wills, 2009). Given the number of patients who need

orthopaedic care, there is little to no possibility to increase the duration of the consultation with the orthopedic surgeon. However, according to orthopedists within St. Anna, the above factors have a negative influence on the ability to provide information to the patient in an understandable way and thus to meet the expectations of the patient. In addition, as mentioned earlier, the orthopaedic surgeon has the complicated task of determining the differential diagnosis and appropriate treatment within the short duration of a consultation on the basis of the history and physical examination and transferring the results to the patient. According to orthopedists within St. Anna, the time pressure ensures that they cannot always prepare well for a consultation, which can compromise factors such as empathy and good communication.

In other words, the quality of an orthopaedic consultation can be improved. A previous study conducted within the St. Anna hospital investigated whether patients' medical knowledge can be increased by offering personalized information to osteoarthritis patients before a consultation takes place (Timmers et al., 2018). This has shown that offering personalized information to osteoarthritis patients before the consultation takes place has a positive impact on both the real and self-reported medical knowledge of the patient. It is expected that increasing medical knowledge will also increase the experienced quality of the clinical consultation by the patient because the expectations of the patient are more realistic and because the patient is better able to communicate with the orthopedic surgeon.

In addition to the study by Timmers et al. (2018) in which patients with a high risk of osteoarthritis of the knee were selected in advance, the current study will provide personalized information to patients by applying machine learning techniques based on the most likely diagnosis. These machine learning techniques can be used to predict diagnoses and treatments and in turn these predictions allow personalized information to be offered to the patient via a mobile application. The models are made based on already implemented digital history questionnaires (usual care) that are completed by patients before they visit the outpatient clinic. In addition to providing personalized information to patients, the digital history taking questionnaire and the predictions can be used to support the orthopedic surgeon. The orthopedic surgeon can use the information to prepare the consultation and it is expected that the orthopaedic surgeon will be able to give a better consultation as a result.

In conclusion, the purpose of St. Anna is to improve the patient-experienced quality of outpatient consultations. For this we use machine learning techniques that we apply to an already existing method, namely the taking of digital history questionnaires prior to the first consultation. With this we provide less general but more personalized information to patients and we support the orthopedic surgeon.

Research goal:

With this study we want to investigate whether taking digital history

questionnaires and providing personalized information to patients improves the experienced quality of an outpatient consultation.

Method

Between April and June 2021, patients with complaints to the knee, hip, or shoulder and over 18 years of age will be asked to participate in a randomized controlled examination. This study investigates the effect of providing personalized information to orthopaedic patients prior to outpatient consultation and conducting digital history questionnaires on the quality of outpatient consultation.

Informed consent and ethical considerations

Patients over the age of 18 and referred due to complaints to the knee, hip or shoulder will be asked to participate in this examination after they have made an appointment for an outpatient consultation. Patients who are interested will receive an email with all the necessary information about the research for informed consent. Patients will have a period of two days to process this information. If patients have any questions, they will be able to contact the lead or co-investigator. Patients will indicate their informed consent by signing an online informed consent form. This form will also emphasize that the patient's data will be kept confidential and secure.

Selection of participants

The suitability of patients will be assessed during the patient's first contact with the hospital for planning a clinical consultation. Patients must speak Dutch, be in possession of an e-mail address and have a smartphone or tablet. In addition, there should be at least 6 days between making the appointment and the consultation to allow patients in the experimental group to use the intervention.

Randomization

Selected patients will be assigned to either the control group or the experimental group immediately after giving informed consent. Randomization will be carried out without block or stratification restrictions. Participants only receive information about the study by email. In addition, participants can immediately complete the standard baseline questionnaires (usual care) via this e-mail. One questionnaire is added for the study (baseline knowledge questionnaire, 13 questions). Patients assigned to the experimental group will receive an additional email with instructions on how to download the app and the patient's personal code. After this, both groups receive the same questionnaires related to the patient's actual and self-reported medical knowledge at the same time at three times: immediately after registration, two days before the outpatient consultation takes place and 1 day after the outpatient consultation has taken place. In addition, both groups receive a questionnaire about the experienced quality of the consultation 1 day after the outpatient consultation. If necessary, participants will receive a maximum of two emails to remind them to complete the questionnaires.

Burden and risk to the patient

This is a patient-related data study that is in line with the usual care process. In St. Anna's, the usual care for the first consultation is that patients fill in a digital history questionnaire and that they prepare for the first consultation with the help of an app. This extends this by taking the same questionnaires related to the patient's actual and self-reported medical knowledge at three times in the experimental and control group. The general preparatory information shall be replaced by personalised information in the intervention group. In addition, a questionnaire about the experienced quality regarding the consultation is taken 1 day after the consultation. There is therefore a slight burden on the patient with little to no chance of negative consequences.

Motivation exemption WMO duty

Because there is light load for the patient and little to no risk when conducting this study, we ask the METC to exempt this study from the WMO obligation and that no additional insurance is required to carry out this study.

Contact details

Removed due to privacy

References

Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [Medline: 12724430]

Ley P, Communicating with Patients: Improving Communication, Satisfaction and Compliance. New York: Croom Helm, 1988

McGuire LC. Remembering what the doctor said: organization and adults' memory for medical information. Exp Aging Res 1996;22(4):403-428. [doi: 10.1080/03610739608254020] [Medline: 8968711]

Waters S, Edmondston SJ, Yates PJ, Gucciardi DF. Identification of factors influencing patient satisfaction with orthopaedic outpatient clinic consultation: A qualitative study. Man Ther. 2016 Sep;25:48-55. Doi: 10.1016/j.math.2016.05.334. Epub 2016 Jun 4. PMID: 27422597.

Wills J. Health literacy: new packaging for health education or radical movement? Int J Pub Health. 2009;54:3-4.

H METC Study Approval Letter



medisch ethische toetsingscommissie
secretariaat
telefoonnummer: 040888 9528
e-mail:metc@mmc.nl

St. Anna Ziekenhuis Geldrop
Afdeling Orthopedie
t.a.v. dhr. dr. W. van der Weegen
Postbus 90
5660 AB Geldrop

Datum: 11 januari 2021
 Brief nummer: 2021-002
 Betreft: WMO-plichtigheid
 Studie: Ervaren kwaliteit van een orthopedisch consult voorbereid met digitale anamnese en voorspelling van de diagnose en behandeling
 METC nummer N21.001

Wij verzoeken u om bij verdere correspondentie bovenstaand METC nummer te gebruiken.

Geachte heer Van der Weegen,

De medische ethische toetsingscommissie (METC) van Máxima MC heeft bovengenoemd onderzoeksvoorstel compleet ontvangen op 08-01-2021. Het dagelijks bestuur van de commissie is tot de conclusie gekomen dat het onderzoek niet onder de werkingssfeer van de *Wet medisch wetenschappelijk onderzoek met mensen* (WMO) valt.

Deze beoordeling is gebaseerd op de volgende documenten:

- Aanvraagformulier nWMO METC Máxima MC met datum 14-12-2020
- Onderzoeksprotocol met datum 14-12-2020
- Informatiebrief voor deelnemer plus toestemmingsverklaring versie 1.0, 13-11-2020
- Vragenlijsten

Volledigheidshalve benadrukt de METC dat de studie niet is beoordeeld op relevantie, kwaliteit en conformiteit met overige mogelijk van toepassing zijnde wet- en regelgeving (zoals de WGBO, de AVG en de code 'Goed Gebruik').

Alle wijzigingen in deze studie dienen opnieuw aan de METC te worden voorgelegd.

Volledigheidshalve maken wij u erop attent dat het onderzoek pas mag worden uitgevoerd nadat u schriftelijk toestemming heeft gekregen van de instelling.

Vertrouwende u hiermee voldoende geïnformeerd te hebben.

Met vriendelijke groet,

Mevr. Y.I.C. (Yolanda) de Haan
Ambtelijk secretaris METC

To whom it may concern,

The Daily Board of the Medical Ethics Committee Máxima MC (hereafter the Committee), has reviewed the above mentioned research proposal. As a result of this review, the Committee informs you that the rules laid down in the Medical Research Involving Human Subjects Act (also known by its Dutch abbreviation WMO), do not apply to this research proposal.

I MEC St. Anna Study Approval Letter



De heren W. v.d. Weegen, projectleider orthopedie en
T. Warren, stagiair polikliniek orthopedie

St. Anna Zorggroep

HUIS

Geldrop, 22 februari 2021
 Betreft: Ervaren kwaliteit van een orthopedisch consult voorbereid met digitale anamnese en voorspelling van de diagnose en behandeling, intern nummer 21.002
 Kenmerk: MEC/PP/2021-004

Geachte heren V.d. Wegen en Warren,

In uw brief van 12 januari 2021 vraagt u toestemming aan de MEC voor deelname aan het onderzoek Ervaren kwaliteit van een orthopedisch consult voorbereid met digitale anamnese en voorspelling van de diagnose en behandeling. Het betreft een niet WMO-plichtige studie.

De MEC van de St. Anna Zorggroep is akkoord met uitvoering van dit onderzoek.

Wij wensen u veel succes toe.

Met vriendelijke groet,

P. Petra van Slout

M.J.J. Gels
 Voorzitter Medisch Ethische Commissie

St. Annaziekenhuis Bogardeind 2 5664 EH Geldrop	Centrum voor verpleegzorg Berkenheuvel Grote Bos 6 5666 AZ Geldrop	Zorgcentrum Akert Appelaar 35 5664 TZ Geldrop	Zorgcentrum Josephinehof Josephinehof 2 5664 AM Geldrop	Zorgcentrum Nicasiushuis Ds. Kremerstraat 7 5591 GH Heeze	www.st-anna.nl
---	--	--	--	--	--

J RCT Invitation Email

Dear sir / Madam,

You have just completed the intake questionnaire for your visit to St. Anna. Based on your answers, we have determined that you may be able to participate in a scientific study with the aim of improving your experience of the consultation. For this research we would ask you to complete a few more questionnaires and possibly read some information on your mobile phone or tablet that is offered by means of the Patient Journey. If you are interested, click on the link below. You will then receive more information about the study and the option to confirm your participation or to opt out.

Link button

Sincerely,

Tristan Warren

K Study Information and Consent Form

K.1 Introduction and Eligibility (first page)

Thank you for completing all the questions so far. Based on your answers, we have determined that you may be able to participate in a scientific study. In total, the entire survey would take you about half an hour of your time. On the other hand, you might contribute to the improvement of the quality of future consultations, of which many will benefit!

Below you can indicate whether you are interested. If you indicate that you are interested, we will ask you a few more questions to check whether you can really participate. If you turn out to be eligible, you will receive more information about the study on the next page and the possibility to give your consent for participation. If you indicate that you are not interested or if it turns out that you are not eligible then you will be able to close the questionnaire.

Are you interested in participating in a scientific study?

- Yes
- No

Are you older than 18?

- Yes
- No

Do you own a smartphone or tablet?

- Yes
- No

Does your consultation take place at least six days in the future?

- Yes
- No

K.2 Study information (second page)

Dear Sir / Madam,

We kindly ask you to participate in a scientific study. Participation is voluntary. Your written permission is required to participate. Before you decide whether you want to participate in this study, you will receive an explanation of what the study entails. Please read this information carefully and ask the researchers for an explanation if you have any questions. The contact details of the researchers can be found at the bottom of this page. You can also talk about it with your partner, friends, or family. If you decide to participate, you can choose to receive the information again by e-mail so that you can always go through this again.

Research goals

Since quality care is of paramount importance at St. Anna, we strive to keep improving our care. The aim of this study is therefore to improve the quality of our outpatient consultations. The results of this research will be published as a scientific article. You can indicate later whether you want to receive the publication when the research has been completed. This way you can see for yourself how you hopefully have been able to make a positive contribution to improving our consultations!

What participation entails

If you decide to participate in this research, we will try to improve your experience with the consultation for which you will be visiting us shortly. To achieve this, we are going to investigate two changes.

Firstly, the orthopedic surgeon with whom you have a consultation will go through the health history questionnaire that you have just completed to prepare for your consultation. Secondly, based on your answers to the health history questionnaire, a prediction will be made about the origin of your complaints and what the best course for treatment might be. We use artificial intelligence for this. You can find a short explanation in the video below.

[Click here to view the video](#)

We will then provide you with information about the diagnosis and treatment that the computer predicts for you. We use an app that you can install on your mobile to provide this information. You can then use the information in the app to prepare for the consultation.

Not everyone has to use the app. You will receive an invitation with instructions on how to install the app by email if you are supposed to. When you receive that invitation, remember that the diagnosis and / or treatment that are predicted by the computer may be incorrect. This means that the information you see on the app may not match what the orthopedic surgeon tells you during the consultation. So don't be alarmed if the computer predicts a surgery for you. Your final treatment will always be discussed and determined together with your orthopedic surgeon.

To measure the effect of these changes, we will send you a total of 4 short questionnaires. You will automatically receive an invitation in your mail when a questionnaire is ready so that you do not have to think about it yourself. A brief summary of what to expect and when to expect it is shown in the figure below.



Do you have questions?

That is possible. We have given you a brief summary of the information about the study here. If you want more information you can click on the following link:

[Click here for more information about the study](#)

If this does not answer all your questions or if you prefer to get in touch directly, please contact the research team:

** contact details of the research team **

Thank you for your attention. On the next page you can indicate whether you want to participate. If you are not sure yet, please close the page and come back to it later. All your answers are automatically saved. You can come back using the same link you used to open this questionnaire.

K.3 Informed Consent Form (third page)

I am aware of the purpose of the study. I was able to ask additional questions. My questions, if any, have been answered to my satisfaction. I had enough time to decide whether to participate.

I know participation is voluntary. I agree to participate in the study.

- Yes
- No

Would you like to receive all the information on your email again so that you can always go through it again?

- Yes
- No

Would you like to receive the published paper by e-mail when the research has been completed?

- Yes
- No

Does your consultation take place at least six days in the future?

- Yes
- No

L Study Exclusion Email

Dear sir / Madam,

You recently indicated that you wish to participate in a study at St. Anna. Unfortunately, we have to inform you on the basis of the computer's prediction that we cannot include you in the study. This is because we only include a limited number of predicted diagnoses in this study. The computer has predicted that your diagnosis is not one of the diagnoses that we include in this research. This means that you will no longer be asked to do anything. In addition, we will no longer keep your data and destroy it once the study is completed. If you have indicated that you would like to receive the results of the research when it has been completed, you will still receive them. We apologize for the inconvenience and thank you for your participation.

If you have any questions, please contact the research team:

Removed due to privacy

M Randomisation and Prediction Registration

Has the patient been assigned to the control group?

- Yes
- No

Has the patient been assigned to the experimental group?

- Yes
- No

What is the predicted diagnosis?

- Gonarthrosis
- Degenerative Meniscus Tear
- Other

What is the predicted probability of this diagnosis?

What is the predicted treatment category?

- Surgical
- Conservative

What is the predicted probability of this treatment category?

N Group Assignment Email

Dear sir / Madam,

You have been assigned to the experimental group of the study. This means that you will receive information via the Patient Journey App. You will receive a separate email with instructions for this. If you cannot find an answer or if you do not receive an email with instructions within two days, please contact the research team.

Sincerely,

Removed due to privacy

O Timeline Gonarthrosis Surgical

5 days before visit



Optimal preparation for your visit

You will soon have an appointment at the orthopedics clinic of our hospital.

With this app we would like to prepare you as well as possible for this appointment.

Enter your personal code

The use of this app is currently only possible for patients participating in the study.

If you have received a personal code, you can enter it below.

5 dagen voor uw afspraak in het ziekenhuis

Fijn dat u mee wilt doen aan ons project!

Dit project richt zich op de voorbereiding op uw bezoek aan uw orthopedisch chirurg. Daarbij maken we gebruik van kunstmatige intelligentie.

Met de antwoorden die u recent via de online vragenlijst aan ons heeft gegeven, kunnen wij een voorspelling doen over de mogelijke oorzaak van uw knieklachten én de mogelijke behandeling hiervan.

[Leer meer over het onderzoek](#)

Fijn dat u mee wilt doen aan ons project!

Dit project richt zich op de voorbereiding op uw bezoek aan uw orthopedisch chirurg. Daarbij maken we gebruik van kunstmatige intelligentie.

Met de antwoorden die u recent via de online vragenlijst aan ons heeft gegeven, proberen wij een voorspelling doen over een mogelijke oorzaak van uw knieklachten én de mogelijke behandeling hiervan.

Op basis van deze voorspelling bieden wij u ook alvast de nodige informatie aan. Tijdens de afspraak met uw orthopedisch chirurg wordt de definitieve, best passende behandeling bepaald. Ook ontvangt u dan de informatie nogmaals.

Thank you for participating in our project!

This project focuses on preparing for your visit to your orthopedic surgeon. We use artificial intelligence for this.

With the answers you recently provided to us via the online questionnaire, we can make a prediction about the possible cause of your knee complaints and the possible treatment thereof.

Thank you for participating in our project!

This project focuses on preparing for your visit to your orthopedic surgeon. We use artificial intelligence for this.

With the answers you recently provided to us via the online questionnaire, we try to make a prediction about a possible cause of your knee complaints and the possible treatment thereof.

Based on this prediction, we also offer you the necessary information. The definitive, most appropriate treatment will be determined during the appointment with your orthopedic surgeon.

⚠

5 dagen voor uw afspraak in het ziekenhuis

 **Belangrijk**

Dit project is **géén vervanging** van het consult of van de informatie die u normaal van ons zou ontvangen.

Het doel van het project is om te kijken in hoeverre kunstmatige intelligentie de voorbereiding op uw afspraak kan ondersteunen.

5 dagen voor uw afspraak in het ziekenhuis

 **Over de afdeling orthopedie**

Het vakgebied orthopedie houdt zich bezig met uw gewrichten, botten en pezen.

We behandelen bijvoorbeeld patiënten met klachten aan hun knie, heup, schouder, rug of voet.

Important

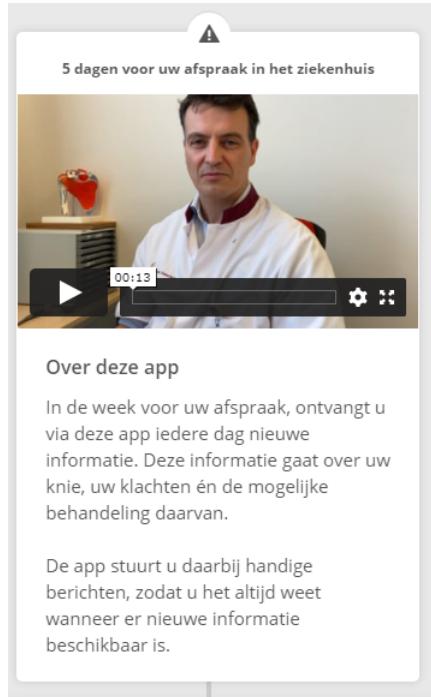
This project is not a substitute for the consultation or the information that you would normally receive from us.

The aim of the project is to see to what extent artificial intelligence can support the preparation for your appointment.

About the orthopedics department

The field of orthopedics deals with your joints, bones and tendons.

For example, we treat patients with knee, hip, shoulder, back or foot complaints.



About this app

In the week before your appointment, you will receive new information via this app every day. This information is about your knee, your complaints and the possible treatment thereof.

The app sends you useful messages, so that you always know when new information is available.

5 dagen voor uw afspraak in het ziekenhuis



Schakel push notificaties in

Op deze telefoon of tablet staan de push notificaties helaas niet aan voor deze app. We adviseren om dit wel te doen. Dit kan vaak aangepast worden via de instellingen van het apparaat.

Door de notificaties in te schakelen, worden er automatisch berichten naar dit apparaat gestuurd wanneer er nieuwe informatie beschikbaar is.

Enable push notifications

Unfortunately, the push notifications on this phone or tablet are not enabled for this app. We recommend that you do this. This can often be adjusted through the device settings.

Enabling the notifications will automatically send messages to this device when new information is available.

Push notificaties niet ingesteld

5 dagen voor uw afspraak in het ziekenhuis

De voorspelling door de computer

Zoals aangegeven, heeft een computer gekeken naar uw antwoorden op de intake vragenlijst.

De computer heeft op basis van deze antwoorden onderstaande voorspelling gedaan over de herkomst van uw klachten.

5 dagen voor uw afspraak in het ziekenhuis

Uw voorspelling: slijtage (artrose) van de knie

De klachten aan uw knie worden waarschijnlijk veroorzaakt door artrose. Dit wordt ook wel artrose genoemd.

Houd u er rekening mee dat de werkelijke diagnose door uw arts, anders kan zijn dan de voorspelling van de computer.

Prediction by the computer

As indicated, a computer looked at your answers to the intake questionnaire.

Based on these answers, the computer has made the following prediction about the origin of your complaints.

Your predicted diagnosis: degeneration (arthrosis) of the knee

The complaints to your knee are probably caused by osteoarthritis. This is also called osteoarthritis.

Please note that the actual diagnosis by your doctor may differ from the computer's prediction.

5 dagen voor uw afspraak in het ziekenhuis

Informatie over uw knie

Het **gezonde** kniegewricht bestaat uit de uiteinden van het dijbeen en scheenbeen die precies op elkaar passen.

Over deze botten in het kniegewricht zit een laagje kraakbeen wat ervoor zorgt dat de kniepeel soepel kan bewegen.

Meer informatie over uw knie

Over uw knie

Het **gezonde** kniegewricht bestaat uit de uiteinden van het dijbeen en scheenbeen die precies op elkaar passen. Door de ronde vorm van het dijbeen, kan de knie goed buigen en strekken.

Over deze botten in het kniegewricht zit een laagje kraakbeen wat ervoor zorgt dat deze bewegingen vrijwel zonder enige weerstand kunnen plaats vinden.

Ter vergelijking, gezond kraakbeen dat glijdt over gezond kraakbeen is wel 1000 maal gladder dan twee ijsklontjes die over elkaar heen glijden.

Information about your knee

A healthy knee joint consists of the ends of the thigh and shin that fit together exactly.

There is a layer of cartilage over these bones in the knee joint, which ensures that the knee can move smoothly.

About your knee

A healthy knee joint consists of the ends of the thigh and shin that fit together exactly. Due to the round shape of the thigh, the knee can bend and stretch well.

There is a layer of cartilage over these bones in the knee joint, which ensures that these movements can take place almost without any resistance.

In comparison, healthy cartilage sliding over healthy cartilage is 1000 times smoother than two ice cubes sliding over each other.

5 dagen voor uw afspraak in het ziekenhuis

 **Informatie over artrose**

Bij mensen die artrose ontwikkelen gaat de kwaliteit van het kraakbeen in de knie langzaam achteruit, dit noemen we degeneratie.

Het kraakbeen is niet meer zo glad als voorheen en er kunnen scheuren in ontstaan. Dit kan uiteindelijk leiden tot het volledig verdwijnen van het kraakbeen uit de knie.

Leer meer over artrose

Informatie over artrose

Bij mensen die **artrose** ontwikkelen gaat de kwaliteit van het kraakbeen in de knie langzaam achteruit, dit noemen we degeneratie.

Het kraakbeen is niet meer zo glad als voorheen en er kunnen scheuren in ontstaan. Dit kan uiteindelijk leiden tot het volledig verdwijnen van het kraakbeen uit de knie.

Tijdens dit ziekteproces wordt het gewrichtskapsel dikker en stijver. Dit alles zorgt ervoor dat een kniegewricht met artrose steeds moeizamer beweegt en uiteindelijk ervaart de patiënt veel pijn.

Information about arthrosis

In people who develop osteoarthritis, the quality of the cartilage in the knee slowly deteriorates, we call this degeneration.

The cartilage is no longer as smooth as before and cracks can occur. This can eventually lead to the complete disappearance of the cartilage from the knee.

Information about arthrosis

In people who develop osteoarthritis, the quality of the cartilage in the knee slowly deteriorates, we call this degeneration.

The cartilage is no longer as smooth as before and cracks can occur. This can eventually lead to the complete disappearance of the cartilage from the knee.

During this disease process, the joint capsule becomes thicker and stiffer. All this makes a knee joint with osteoarthritis increasingly difficult to move and eventually the patient experiences a lot of pain.

4 days before visit

4 dagen voor uw afspraak in het ziekenhuis

 De voorspelling door de computer

De computer heeft niet alleen een voorspelling gedaan naar de herkomst van de klachten.

Er is ook gekeken naar de best passende behandeling voor uw klachten. Dit onderdeel van de app geeft u daar meer informatie over.

The prediction by the computer

The computer did not only predict the origin of the complaints.

We also looked at the most appropriate treatment for your complaints. This part of the app gives you more information about this.

4 dagen voor uw afspraak in het ziekenhuis

 00:50

Uw voorspelling: een knieprothese operatie

Voor u is een knieprothese operatie mogelijk de best passende behandeling.

Houd u er rekening mee dat de behandeling die door uw arts wordt voorgesteld anders kan zijn dan de voorspelling van de computer.

Your predicted treatment: a knee prosthesis surgery

Knee replacement surgery may be the most suitable treatment for you.

Please note that the treatment suggested by your doctor may differ from the computer's prediction.

Informatie over de operatie

De vervanging van het versleten gewricht door een totale knieprothese gaat door een operatie.

Hierbij wordt het versleten kraakbeen van de knie vervangen door een prothese. Een prothese wordt ook wel kunstknie of kunstgewricht genoemd.

Meer informatie over de operatie

Information about the surgery

The replacement of the worn joint with a total knee prosthesis requires surgery.

The worn cartilage of the knee is replaced by a prosthesis. A prosthesis is also called an artificial knee or artificial joint

Informatie over de operatie

De vervanging van het versleten gewricht door een totale knieprothese gaat door een operatie. Hierbij wordt het versleten kraakbeen van de knie vervangen door een prothese.

Deze prothese wordt met behulp van **cement** in het bot worden bevestigd. Om de prothese in een menselijk lichaam te kunnen plaatsen moet er een snee (ook wel incisie genoemd) in de huid gezet worden.

De incisie bij een totale knieprothese loopt midden over de knie en is ongeveer 15-20cm lang. Bij de operatie wordt een snee in de knie gemaakt. Deze snee, ook wel incisie genoemd, geeft de orthopedisch chirurg toegang tot uw kniegewricht.

Het versleten kraakbeen en een dun laagje bot wordt van de botten gezaagd, waardoor een mooi recht oppervlak ontstaat.

Vervolgens wordt de prothese in uw bovenbeen en onderbeen geplaatst en vastgezet met cement. Tussen deze twee metalen delen, wordt een plastic plaatje geplaatst, om de prothese delen soepel te laten bewegen.

Information about the surgery

The replacement of the worn joint with a total knee prosthesis requires surgery. The worn cartilage of the knee is replaced by a prosthesis.

This prosthesis is fixed in the bone with the help of cement. In order to place the prosthesis in a human body, a cut (also called an incision) must be made in the skin.

The incision for a total knee prosthesis is mid-knee and is approximately 15-20cm long. A cut is made in the knee during the operation. Also called an incision, this cut gives the orthopedic surgeon access to your knee joint.

.....

A digital health card template for a knee replacement consultation. It features a header with a person icon and the text "4 dagen voor uw afspraak in het ziekenhuis". Below this is a circular icon showing two hands holding a knee joint. The text "Waarom is dit de best passende behandeling?" is next to it. The main content area contains two paragraphs: one about knee joint replacement as a good option for pain relief and mobility improvement, and another about the goal of a total knee prosthesis. The design is clean with a white background and a light gray border.

3 days before visit

A digital health card template for knee replacement information, specifically focusing on the benefits. It has a header with an info icon and the text "3 dagen voor uw afspraak in het ziekenhuis". Below this is a circular icon of a doctor writing. The text "Minder pijn en meer beweging" is next to it. The main content area contains two paragraphs: one about the reduction of pain and increased movement after surgery, and another statistic about the number of knee replacements performed in the Netherlands each year. The design is clean with a white background and a light gray border.

Why could this be the most suitable treatment?

Knee joint replacement may be a good option to resolve your symptoms.

The aim of a total knee prosthesis is to reduce pain, restore mobility and improve quality of life.

Less pain and more movement

Implanting a knee prosthesis is a commonly used procedure that has proven to be safe and effective. After this, patients have less pain and more movement in the operated knee.

In total, approximately 25,000 knee prostheses are placed each year in the Netherlands, 700 of which at St. Anna.

Pijn na de operatie

Het is bekend dat een knieprothese een pijnlijke ingreep is.

Natuurlijk proberen we met pijnstilling zo veel mogelijk verlichting te bieden. Echter, ondanks deze pijnmedicatie zult u in de eerste weken na de ingreep (forse) pijnklachten ervaren, die pas na een periode van 2 tot 4 weken af zullen nemen.

De duur van het totale herstel neemt 6 tot 12 maanden in beslag.

[Meer informatie over pijn na de operatie](#)

Pijn na de operatie

Het is bekend dat een knieprothese een pijnlijke ingreep is.

Natuurlijk proberen we met pijnstilling zo veel mogelijk verlichting te bieden. Echter, ondanks deze pijnmedicatie zult u in de eerste weken na de ingreep (forse) pijnklachten ervaren, die pas na een periode van 2 tot 4 weken af zullen nemen.

Dit ervaren patiënten vaak als een moeilijke periode.

Deze waarschuwing is niet om u van een operatie te houden, maar dient als mentale voorbereiding. Onderzoek heeft laten zien, dat wanneer de patiënt van tevoren goed op de hoogte is van deze pijn, vaak veel beter met de pijn kan omgaan na de operatie.

Pain after surgery

Knee replacement is known to be a painful procedure.

Of course, we try to provide as much relief as possible with pain relief. However, despite this pain medication, you will experience (severe) pain symptoms in the first weeks after the procedure, which will only subside after a period of 2 to 4 weeks.

The duration of total recovery takes 6 to 12 months.

Pain after surgery

Knee replacement is known to be a painful procedure.

Of course we try to provide as much relief as possible with pain relief. However, despite this pain medication, you will experience (severe) pain symptoms in the first weeks after the procedure, which will only subside after a period of 2 to 4 weeks.

Patients often experience this as a difficult period.

This warning is not to keep you from surgery, but serves as a mental preparation. Research has shown that if the patient is well aware of this pain in advance, it is often much better to deal with the pain after the operation.



The image shows a digital health interface. At the top, it says "3 dagen voor uw afspraak in het ziekenhuis". Below that is a section titled "Risico's" with a circular icon of hands performing surgery. The text explains that implanting a knee prosthesis is a safe and effective procedure, with patients experiencing less pain and more movement. It notes that approximately 25,000 such operations are performed annually in the Netherlands. The text also mentions the low risk of complications, around 1 to 2%, and lists potential complications like blood vessel or nerve injuries, bruising, swelling, infection, loosening, thrombosis, and residual pain. A "Lees verder" button is at the bottom.

Risico's

Het implanteren van een knieprothese is een vaak toegepaste procedure die **veilig en effectief** is gebleken. Patiënten hebben hierna minder pijn en meer beweging in de geopereerde knie. Jaarlijks worden in Nederland meer dan 25000 knieprothese operaties uitgevoerd.

De kans dat er iets mis gaat is erg klein: ongeveer 1 tot 2%. Bekende, maar weinig voorkomende, complicaties bij deze ingreep zijn letsls van bloedvaten of zenuwen, bloeduitstorting of zwelling, infectie van de prothese, loslating van de prothese, een trombose been of longembolie en mogelijke restpijnklachten.

Lees verder

Risks

Implanting a knee prosthesis is a commonly used procedure that has proven to be safe and effective. After this, patients have less pain and more movement in the operated knee. Every year, more than 25,000 knee prosthesis operations are performed in the Netherlands.

The chance that something will go wrong is very small: about 1 to 2%. Known, but uncommon, complications with this procedure are blood vessel or nerve injuries, bruising or swelling, infection of the prosthesis, loosening of the prosthesis, thrombosis of the leg or pulmonary embolism and possible residual pain.

Risks

....

As with other operations, there are of course also risks. There is a chance that something will go wrong, we call this a complication. However, the chance that something will go wrong is very small, about 1 to 2%.

....

To reiterate, the likelihood of a patient experiencing any of these complications is very small, but it is important that a patient is aware of the risks, no matter how small the likelihood. If during the consultation it appears that this is indeed the best treatment for your complaints, we will discuss this further.

2 days before visit

2 dagen voor uw afspraak in het ziekenhuis



Een korte quiz

In de afgelopen dagen hebben we u veel informatie gegeven over de klachten aan uw knie en de mogelijke behandeling hier.

Via deze korte quiz willen we samen met u nog eens terugkijken op al deze informatie.

De antwoorden die u geeft, hebben op geen enkele manier invloed op het verloop van uw afspraak.

A short quiz

In the past few days we have given you a lot of information about the complaints to your knee and the possible treatment here.

Through this short quiz we would like to review all this information together with you.

The answers you provide will in no way affect the course of your appointment.

2 dagen voor uw afspraak in het ziekenhuis



Vraag 1

Wat is artrose?

Een ander woord voor knieprothese
Een ander woord voor de meniscus
Een ander woord voor slijtage van het kraakbeen

Questions 1

What is arthrosis?

- A different word for knee prosthesis
- A different word for the meniscus
- A different word for the degeneration of cartilage

... and two other questions that are not included

2 days before visit



Your consultation is tomorrow

Tomorrow is your appointment with us at the hospital. Of course, we hope that this app has prepared you as well as possible.

We advise you to read the information again. In this way, you will come to the hospital optimally prepared.



Important to know

We would like to emphasize again that in this study we use the prediction of the computer.

The information in this app is based on this prediction and may therefore differ from what you will hear in the hospital tomorrow.

The image shows three vertical panels from a mobile application designed to guide users through their hospital appointment process. Each panel features a small icon at the top and a title at the bottom.

- Schrijf uw vragen op**
Heeft u vragen? Schrijf ze dan op.
Het is belangrijk dat uw vragen beantwoord worden. De behandeling gaat immers niet om uw knie, maar om u als mens.
- Wat neemt u mee naar de afspraak?**
Binnen het ziekenhuis is het altijd belangrijk om de volgende zaken bij te hebben:
 - uw legitimatiebewijs
 - uw ziekenhuispas
 - uw pas van de zorgverzekeraar
 - een overzicht van de medicatie die u gebruikt
- Succes met uw afspraak**
Wij willen u bedanken voor het gebruik van deze app en wensen u een prettige afspraak toe in ons ziekenhuis.

Write down any questions

Do you have questions? Then write them down.

It is important that your questions are answered. After all, the treatment is not about your knee, but about you as a person

What do you bring to your appointment?

Within the hospital it is always important to have the following items with you:

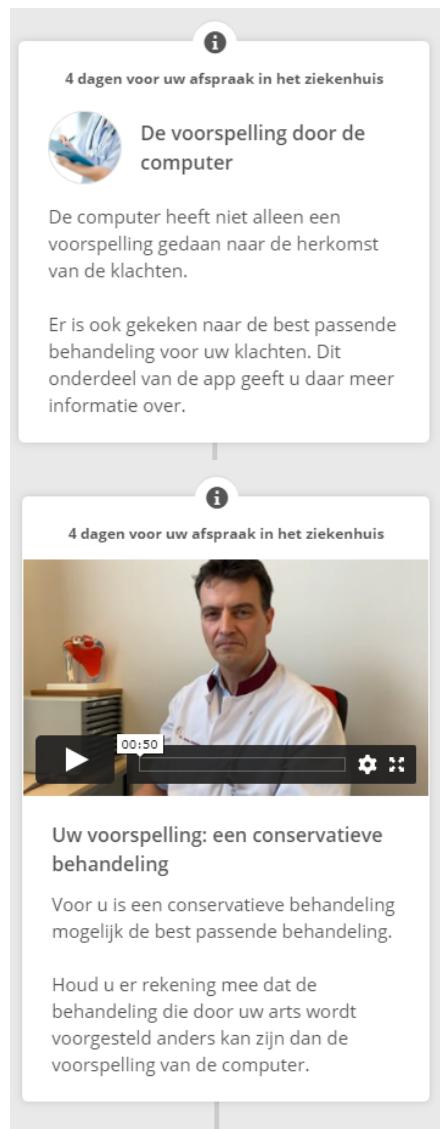
- your ID
- your hospital card
- your pass from the health insurance company*
- an overview of the medication you are taking

Good luck with your appointment

We would like to thank you for using this app and wish you a pleasant appointment in our hospital.

P Timeline Gonarthrosis Conservative

4 days before visit



4 dagen voor uw afspraak in het ziekenhuis

De voorspelling door de computer

De computer heeft niet alleen een voorspelling gedaan naar de herkomst van de klachten.

Er is ook gekeken naar de best passende behandeling voor uw klachten. Dit onderdeel van de app geeft u daar meer informatie over.

4 dagen voor uw afspraak in het ziekenhuis

Uw voorspelling: een conservatieve behandeling

Voor u is een conservatieve behandeling mogelijk de best passende behandeling.

Houd u er rekening mee dat de behandeling die door uw arts wordt voorgesteld anders kan zijn dan de voorspelling van de computer.

The prediction by the computer

The computer did not only predict the origin of the complaints.

We also looked at the most appropriate treatment for your complaints. This part of the app gives you more information about this.

Your predicted treatment: a conservative treatment

Conservative treatment may be the most appropriate treatment for you.

Please note that the treatment suggested by your doctor may differ from the computer's prediction.

The image shows two side-by-side mobile phone screens displaying a digital health application. Both screens have a header bar with a small icon and the text "4 dagen voor uw afspraak in het ziekenhuis".

Top Screen Content:

- Icon:** A circular icon showing a person's knee being examined.
- Section Title:** "Waarom is dit de best passende behandeling?" (Why is this the most suitable treatment?)
- Text:** "Een conservatieve behandeling is mogelijk een goede optie om uw klachten te verhelpen." (Conservative treatment may be a good option for resolving your symptoms.)
- Text:** "Het doel van deze behandelingen is namelijk het verminderen van de knieklachten en het herstellen van de beweeglijkheid van de pijnlijke knie." (The aim of these treatments is to reduce knee complaints and restore the mobility of the painful knee.)

Bottom Screen Content:

- Icon:** A circular icon showing a person's knee being examined.
- Section Title:** "Informatie over de conservatieve behandeling" (Information about conservative treatment)
- Text:** "Een conservatieve behandeling houdt in dat wij u mogelijk pijnstilling, oefentherapie en/of fysiotherapie zullen voorschrijven." (Conservative treatment means that we may prescribe pain relief, remedial therapy and / or physiotherapy.)
- Text:** "Lees verder" (Read more) - a button at the bottom of the screen.

Why could this be the most suitable treatment?

Conservative treatment may be a good option for resolving your symptoms.

The aim of these treatments is to reduce knee complaints and restore the mobility of the painful knee.

Information about conservative treatments

Conservative treatment means that we may prescribe pain relief, remedial therapy and / or physiotherapy.

Informatie conservatieve behandeling

Een patiënt met knieartrose heeft vaak al lang klachten van de knie. De vaak ernstige pijn heeft ervoor gezorgd dat een patiënt zijn 'slechte knie' als het ware gaat **vermijden**, iets waarbij ons brein een belangrijke rol speelt.

Naast de afgenomen beweeglijkheid van een versleten knie, zorgt ons bewustzijn van de 'slechte knie' ervoor dat er veranderingen optreden in de **houding en het looppatroon** van een patiënt met knieartrose. Patiënten krijgen daarom ook vaak klachten van andere gewrichten, bijvoorbeeld de rug of heup.

Door deze vermijding van het aangedane been ontstaat er **compensatiegedrag** waarbij het andere been wordt overbelast. Dit zijn uitingen dat de normale en gezonde **bewegingsketen verstoord** is geraakt, en het probleem van de knie verder verspreid over andere gewrichten.

Ons advies is daarom vaak om voor een periode van 4 weken **structureel pijnstilling** te gebruiken, gecombineerd met **spierversterkende oefentherapie**.

Een vaak gehoord argument, is dat een patiënt 'niet graag pillen slikt'. Dat begrijpen we maar al te goed. Zo maar pijnstilling gebruiken, moet iemand nooit doen. Maar zoals hierboven uitgelegd heeft artrose in een knie vaak geleid tot problemen in de hele bewegingsketen. Daarom geven we vaak het advies om te starten met **paracetamol 4 x per dag 1000mg** (dit zijn vaak 2 tabletten van 500mg die bij drogisterij of apotheek te halen zijn). Dit dient voor 4 weken geslikt te worden.

Onze specialist schrijft als het kan hier ook vaak een **NSAID** bij voor. Dit zijn middelen zoals ibuprofen, diclofenac, naproxen, etc. Deze middelen versterken het pijnstillende effect van de paracetamol en werken ook **ontstekingsremmend**. Ook deze ontstekingsremmer wordt voor 4 weken gebruikt, naast de paracetamol. Door deze adequate pijnstilling zal de pijn afnemen en kan de patiënt weer beter bewegen.

Maar dit is niet voldoende om ook op de lange termijn klachtenvrij te blijven. Zoals hierboven beschreven, is het bewegingspatroon van een patiënt met knieartrose vaak verstoord. Vandaar dat een verwijzing naar de **fysiotherapeut** wordt meegegeven om onder de begeleiding van deze bewegingsspecialist weer naar een goede kracht en normaal

Information about conservative treatments (optional information item)

A patient with knee osteoarthritis often has been having knee complaints about a long time. The often severe pain has caused a patient to avoid his "bad knee", something in which our brain plays an important role.

In addition to the reduced mobility of a worn knee, our awareness of the "bad knee" causes changes in the posture and gait pattern of a patient with knee OA. Patients therefore often get complaints from other joints, such as the back or hip.

This avoidance of the affected leg results in compensatory behavior in which the other leg is overloaded. These are indications that the normal and healthy movement chain has been disrupted, and that the problem of the knee has spread further to other joints.

Our advice is therefore often to use structural pain relief for a period of 4 weeks, combined with muscle-strengthening exercise therapy.

Our advice is therefore often to use structural pain relief for a period of 4 weeks, combined with muscle-strengthening exercise therapy.

A frequently heard argument is that a patient "does not like to take pills". We understand that all too well. A person should never use painkillers like that. But as explained above, osteoarthritis in a knee has often led to problems in the entire chain of motion. That is why we often advise you to start with paracetamol 1000mg 4 x per day (these are often 2 tablets of 500mg that can be obtained at the drugstore or pharmacy). This should be taken for 4 weeks.

If possible, our specialist often prescribes an NSAID for this. These are agents such as ibuprofen, diclofenac, naproxen, etc. These agents enhance the analgesic effect of paracetamol and also have an anti-inflammatory effect. This anti-inflammatory is also used for 4 weeks, in addition to paracetamol. Due to this adequate pain relief, the pain will decrease and the patient can move better again.

But this is not enough to remain free of complaints in the long term. As described above, the movement pattern of a patient with knee OA is often disturbed. That is why a referral to the physiotherapist is given to work towards good strength and normal gait under the guidance of this movement specialist. It is precisely this restoration of the movement chain and an increase in strength that are of great importance in order to remain complaints-free in the long term.

i

4 dagen voor uw afspraak in het ziekenhuis



Waarom geen operatie?



Bij nog relatief **milde** knieartrose, is het heel reëel om te verwachten dat de klachten door de conservatieve behandeling voor een langere periode goed verbeteren.

Daarbij speelt **leeftijd** ook een rol. Een knieprothese gaat gemiddeld 20 jaar mee. Bij jonge patiënten wordt de prothese vaak intensiever gebruikt en gaat deze minder lang mee. Als de prothese los gaat zitten gaat deze pijn doen en moet deze uiteindelijk vervangen worden.

[Lees verder](#)

Why not surgery?

In the case of relatively mild knee osteoarthritis, it is very realistic to expect that the complaints will improve well for a longer period as a result of conservative treatment.

Age also plays a role in this. A knee prosthesis lasts an average of 20 years. In young patients, the prosthesis is often used more intensively and lasts less. If the prosthesis becomes loose, it will hurt and will eventually need to be replaced.



Waarom geen operatie?

Sommige mensen hebben nog niet een totaal versleten knie, en zijn de slijtage klachten op de foto of de klachten nog relatief mild. Bij nog relatief **milde** knieartrose, is het heel reëel om te verwachten dat de klachten door de conservatieve behandeling voor een langere periode goed verbeteren.

Het is wél bekend van arrose, dat de klachten in de tijd verder **verslechtern** en weer terug kunnen komen. Gebeurt dit, dan kan de conservatieve behandeling met pijnstilling en oefentherapie nogmaals gestart worden. Blijkt dat het effect van deze behandeling voor de patiënt steeds minder goed werkt, kan dit een indicatie zijn om alsnog voor een operatie te kiezen.

Why not surgery? (optional information item)

Some people do not have a completely worn-out knee yet, and the wear and tear complaints in the photo or the complaints are still relatively mild. In the case of relatively mild knee osteoarthritis, it is very realistic to expect that the complaints will improve well for a longer period as a result of conservative treatment.

It is known for osteoarthritis that the complaints worsen over time and can return. If this happens, the conservative treatment with pain relief and exercise therapy can be started again. If it appears that the effect of this treatment is working less and less well for the patient, this may be an indication to opt for surgery after all.

Verder is het van belang dat ook de **leeftijd** van een patiënt een rol speelt. Het is bekend dat in Nederland een patiënt ongeveer 65 jaar oud is als die een knieprothese krijgt. Ook weten we dat een goed geplaatste prothese ongeveer **20 jaar** goed vast blijft zitten in de patiënt. Dit betekent dat een patiënt van 65 jaar na de operatie heel waarschijnlijk tot een leeftijd van 85 jaar veel plezier kan hebben van de knieprothese. In veel gevallen blijkt het dan niet noodzakelijk te zijn om een knieprothese ooit te moeten vervangen.

Is een patiënt nu veel jonger dan 65 jaar, dan weten we dat de knieprothese in deze patiënten vaak **intensiever** gebruikt worden en vaak minder dan 20 jaar lang meegaan. Als de prothese los gaat zitten gaat deze pijn doen en moet deze uiteindelijk **vervangen** worden (een revisie van de knieprothese). Dit is een ingreep die tegenwoordig goed kan worden uitgevoerd, maar de kans dat er iets misgaat bij deze revisie-ingreep (een complicatie) is hoger dan bij de eerste keer als iemand een knieprothese krijgt. Vandaar dat de orthopedisch chirurg bij de jongere patiënten wil proberen om een operatie uit te stellen. Op die manier loopt de patiënt het **minste** risico op een complicatie.

Why not surgery? (continued)

It is also important that the age of a patient also plays a role. It is known that in the Netherlands a patient is about 65 years old on average when he or she receives a knee prosthesis. We also know that a properly placed prosthesis remains firmly in place in the patient for about 20 years. This means that a 65-year-old patient will most likely be able to enjoy the knee prosthesis up to the age of 85 after surgery. In many cases, it turns out that it is not necessary to ever have to replace a knee prosthesis.

If a patient is now much younger than 65, we know that the knee prosthesis in these patients is often used more intensively and often lasts less than 20 years. If the prosthesis becomes loose, it will hurt and will eventually have to be replaced (a revision of the knee prosthesis). This is a procedure that can be performed well nowadays, but the chance that something will go wrong with this revision procedure (a complication) is higher than the first time someone gets a knee prosthesis. Hence, the orthopedic surgeon wants to try to postpone surgery in the younger patients. In this way, the patient has the least risk of a complication.

Q Timeline Degenerative Meniscus Tear

5 days before visit

The image shows two screenshots of a mobile application interface. The top screenshot displays a card titled "5 dagen voor uw afspraak in het ziekenhuis" (5 days before your appointment at the hospital). It features a circular icon with an "i" and the text "De voorspelling door de computer" (Prediction by the computer). Below this is a small image of a doctor. The text reads: "Zoals aangegeven, heeft een computer gekeken naar uw antwoorden op de intake vragenlijst. De computer heeft op basis van deze antwoorden onderstaande voorspelling gedaan over de herkomst van uw klachten." The bottom screenshot shows a video player interface with a play button, a progress bar at 00:46, and a video thumbnail of a doctor. The text below the thumbnail reads: "Uw voorspelling: slijtage van de meniscus. De klachten aan uw knie worden waarschijnlijk veroorzaakt door slijtage van de meniscus. Dit wordt ook wel degeneratief meniscusletsel genoemd. Houd u er rekening mee dat de werkelijke diagnose door uw arts, anders kan zijn dan de voorspelling van de computer."

Prediction by the computer

As indicated, a computer looked at your answers to the intake questionnaire.

Based on these answers, the computer has made the following prediction about the origin of your complaints.

Your predicted diagnosis: degenerative meniscus tear

The complaints that you experience are probably caused by wear and tear of the meniscus. This is also called a degenerative meniscus tear.

Please note that the actual diagnosis by your doctor may differ from the computer's prediction.

 5 dagen voor uw afspraak in het ziekenhuis

 Informatie over uw knie

Het **gezonde** kniegewricht bestaat uit de uiteinden van het dijbeen en scheenbeen die precies op elkaar passen. Door de ronde vorm van het dijbeen, kan de knie goed buigen en strekken.

Over deze botten in het kniegewricht zit een laagje kraakbeen wat ervoor zorgt dat deze bewegingen vrijwel zonder enige weerstand kunnen plaats vinden.

Naast kraakbeen wat op de botten zit, zit er in de knie nog een andere structuur van kraakbeen, de **meniscus**. De meniscus zorgt ervoor dat de botten uit het boven- en onderbeen goed op elkaar passen.

Meer informatie over uw knie

Information about your knee

A healthy knee joint consists of the ends of the thigh and shin that fit perfectly together.

There is a layer of cartilage over these bones in the knee joint, which ensures that the knee can move smoothly.

In addition to the cartilage on the bones, there is another structure of cartilage in the knee, the meniscus. The meniscus ensures that the bones of the upper and lower leg fit together properly.

Over uw knie

Het **gezonde** kniegewricht bestaat uit de uiteinden van het dijbeen en scheenbeen die precies op elkaar passen. Door de ronde vorm van het dijbeen, kan de knie goed buigen en strekken.

Over deze botten in het kniegewricht zit een laagje kraakbeen wat ervoor zorgt dat deze bewegingen vrijwel zonder enige weerstand kunnen plaats vinden.

Ter vergelijking, gezond kraakbeen dat glijdt over gezond kraakbeen is wel 1000 maal gladder dan twee ijsklontjes die over elkaar heen glijden.

Naast kraakbeen wat op de botten zit, zit er in de knie nog een andere structuur van kraakbeen, de **meniscus**. De meniscus zorgt ervoor dat de botten uit het boven- en onderbeen goed op elkaar passen. De onderkant van bovenbeen is namelijk rond van vorm, terwijl de bovenkant van het onderbeen recht is.

In een gezonde knie zijn zowel het kraakbeen over de botten als de meniscus van voldoende kwaliteit om de druk en de krachten op de knie op te vangen.

About your knee (information item)

...

There is a layer of cartilage over these bones in the knee joint, which ensures that these movements can take place almost without any resistance.

In comparison, healthy cartilage sliding over healthy cartilage is 1000 times smoother than two ice cubes sliding over each other.

...

In a healthy knee, both the cartilage over the bones and the meniscus are of sufficient quality to absorb the pressure and forces on the knee.



5 dagen voor uw afspraak in het ziekenhuis

Informatie over meniscusletsel

Bij mensen die **artrose** ontwikkelen gaat de kwaliteit van het kraakbeen in de knie langzaam achteruit, dit noemen we degeneratie. Artrose zorgt ervoor dat een kniegewricht steeds moeizamer beweegt en uiteindelijk ervaart de patiënt veel pijn.

Meniscusletsel is een voorstadium van knieartrose. Bij meniscusletsel verliest de meniscus, die ook van kraakbeen is gemaakt, aan kwaliteit. Hierdoor ontstaan scheurtjes in de meniscus waardoor klachten aan de knie ontstaan.

Lees verder

Informatie over meniscusletsel

Bij mensen die **artrose** ontwikkelen gaat de kwaliteit van het kraakbeen in de knie langzaam achteruit, dit noemen we degeneratie. Het kraakbeen is niet meer zo glad als voorheen en er kunnen scheuren in ontstaan.

Dit kan uiteindelijk leiden tot het volledig verdwijnen van het kraakbeen uit de knie. Tijdens dit ziekteproces wordt het gewrichtskapsel dikker en stijver. Dit alles zorgt ervoor dat een kniegewricht met artrose steeds moeizamer beweegt en uiteindelijk ervaart de patiënt veel pijn.

Een knie is niet in één tel helemaal versleten. Dit is een proces van jaren, vandaar dat artrose met name op hogere leeftijd voorkomt. In de eerste fasen van de knieartrose (**milde artrose**), neemt de kwaliteit van het kraakbeen al wel af. Het is minder goed bestand tegen hoge drukken en krachten zoals gezond kraakbeen en kan gemakkelijk beschadigen.

De meniscus die ook van kraakbeen is gemaakt verliest ook aan kwaliteit en kan daardoor makkelijker scheuren. Het krijgen van (degeneratieve) meniscusklachten is dan ook een eerste uiting van dit **voorstadium van knieartrose**.

Information about degenerative meniscus tear

In people who develop osteoarthritis, the quality of the cartilage in the knee slowly deteriorates. We call this degeneration. Osteoarthritis makes it increasingly difficult for the knee joint to move and eventually the patient will experience a lot of pain.

Degenerative meniscus tears are a precursor to knee osteoarthritis, in which the meniscus, also made of cartilage, loses quality. This causes tears in the meniscus, causing complaints to the knee.

Information item

In people who develop osteoarthritis, the quality of the cartilage in the knee slowly deteriorates, we call this degeneration. The cartilage is no longer as smooth as before and cracks can occur.

This can eventually lead to the complete disappearance of the cartilage from the knee. During this disease process, the joint capsule becomes thicker and stiffer. All this makes a knee joint with osteoarthritis increasingly difficult to move and eventually the patient experiences a lot of pain.

A knee is not completely worn out at once. This is a process that takes years, which is why osteoarthritis mainly occurs in old age. In the first phases of knee osteoarthritis (mild osteoarthritis), the quality of the cartilage decreases. It is less resistant to high pressures and forces such as healthy cartilage and can easily be damaged.

....

4 days before visit

4 dagen voor uw afspraak in het ziekenhuis

 De voorspelling door de computer

De computer heeft niet alleen een voorspelling gedaan naar de herkomst van de klachten.

Er is ook gekeken naar de best passende behandeling voor uw klachten. Dit onderdeel van de app geeft u daar meer informatie over.

The prediction by the computer

The computer did not only predict the origin of the complaints.

We also looked at the most appropriate treatment for your complaints. This part of the app gives you more information about this.

4 dagen voor uw afspraak in het ziekenhuis



Uw voorspelling: een conservatieve behandeling

Voor u is een conservatieve behandeling mogelijk de best passende behandeling.

Houd u er rekening mee dat de behandeling die door uw arts wordt voorgesteld anders kan zijn dan de voorspelling van de computer.

Your predicted treatment: a conservative treatment

Conservative treatment may be the most appropriate treatment for you.

Please note that the treatment suggested by your doctor may differ from the computer's prediction.

4 dagen voor uw afspraak in het ziekenhuis



Waarom is dit de best passende behandeling?



Een conservatieve behandeling is mogelijk een goede optie om uw klachten te verhelpen.

Het doel van deze behandelingen is namelijk het verminderen van de knieklachten en het herstellen van de beweeglijkheid van de pijnlijke knie.

Why could this be the most suitable treatment?

Conservative treatment may be a good option for resolving your symptoms.

The aim of these treatments is to reduce knee complaints and restore the mobility of the painful knee.

4 dagen voor uw afspraak in het ziekenhuis



Waarom geen operatie?



Een kijkoperatie vergroot de kans dat u in de toekomst ernstige knieartrose gaat ontwikkelen. Dit zou er toe leiden dat het plaatsen van een kunstgewricht noodzakelijk is.

Daarnaast laten onderzoeken zien dat patiënten zonder kijkoperatie voor hun meniscuscheur 6 tot 12 maanden na start van een niet-operatieve behandeling even goed presteren als patiënten die wel geopereerd werden.

Lees verder

Why not surgery?

Keyhole surgery increases the chance that you will develop serious knee osteoarthritis in the future. This would make it necessary to place an artificial joint.

In addition, studies show that patients without keyhole surgery perform just as well 6 to 12 months after starting non-operative treatment as patients who did undergo surgery.

Waarom geen operatie?

Vroeger werden patiënten vaak geopereerd aan hun meniscus. Via een kijkoperatie (arthroscopie) werd de meniscus dan deels of in zijn geheel uit de knie verwijderen.

Inmiddels weten we vanuit meerdere grote en zeer betrouwbare onderzoeken, dat de meniscus een **zeer belangrijke structuur** in de knie is. Hoe minder meniscus in knie, hoe groter de kans is dat er op korte termijn artrose gaat ontwikkelen.

Eerder legden we al uit, dat een (degeneratieve) meniscusscheur een kenmerk is van een voorstadium van knieartrose. Als in een knie met voorstadium artrose óók nog eens de meniscus (deels) wordt verwijderd, vergroten we de kans dat er een ernstige knieartrose gaat ontwikkelen waarvoor uiteindelijk een kunstgewricht (totale knieprothese) noodzakelijk is.

Deze onderzoeken laten ook zien, dat patiënten zonder kijkoperatie voor hun meniscusscheur 6 tot 12 maanden na start van een niet-operatieve behandeling even goed presteren als patiënten die wel geopereerd werden.

Dit is waarom wij als uw orthopedisch specialist adviseren om bij dit probleem een kijkoperatie als kan te vermijden. Hoe langer u met uw eigen knie en zo veel mogelijk meniscus kan wandelen, fietsen, leven, hoe beter!

Mocht er tijdens het oefenen of door een nieuw acute moment **een slotklacht** ontstaan, dan moet u dit direct laten weten. Onder een slotklacht wordt verstaan dat een patiënt de knie niet volledig kan strekken of niet meer kan buigen. Wiebelnen, masseren of andere trucs helpen niet. De stand van de knie zit als het ware "op slot". Een knie met een slotstand is niet functioneel te gebruiken in het dagelijks leven. En als de knie gedurende lange tijd in dezelfde houding blijft staan, kan dit leiden tot een levenslange forse bewegingsbeperking.

In een dergelijke situatie dient u contact op te nemen met uw orthopedisch chirurg, zodat **een kijkoperatie (arthroscopie)** gepland kan worden. Bij deze arthroscopie wordt een deel van de meniscus uit de knie verwijderd (**partiële meniscectomie**). De risico's van een mogelijke knieartrose wegen dan niet op tegen de gevolgen van het rondlopen met een knie in slotstand.

Why not surgery? (information item)

In the past, patients often underwent surgery on their meniscus. Through keyhole surgery (arthroscopy), the meniscus was partially or completely removed from the knee.

We now know from several large and very reliable studies that the meniscus is a very important structure in the knee. The less meniscus in the knee, the greater the chance that osteoarthritis will develop in the short term.

We explained earlier that a (degenerative) meniscal tear is a hallmark of early-stage knee osteoarthritis. If the meniscus is (partially) removed in a knee with pre-stage osteoarthritis, we increase the chance that serious knee osteoarthritis will develop for which an artificial joint (total knee prosthesis) is ultimately necessary.

These studies also show that patients without keyhole surgery for their meniscal tear perform as well as patients who did undergo surgery 6 to 12 months after starting a non-operative treatment.

This is why we, as your orthopedic specialist, advise you to avoid keyhole surgery for this problem. The longer you can walk, cycle, live with your own knee and as much meniscus as possible, the better!

If your knee suddenly locks during the practice or due to a new acute moment, you must let us know immediately. A knee lock means that a patient is unable to fully extend or bend the knee. Wobbling, massaging or other tricks will not help. The position of the knee is, as it were, "locked". A knee with a lock position cannot be used functionally in everyday life. And if the knee remains in the same position for a long time, it can lead to a lifelong severe movement restriction.

In such a situation, you should contact your orthopedic surgeon so that keyhole surgery (arthroscopy) can be scheduled. In this arthroscopy, part of the meniscus is removed from the knee (partial meniscectomy). The risk of getting knee osteoarthritis does not outweigh the consequences of walking with a locked knee.

R Real- and Perceived Knowledge Questionnaires

Hello and welcome to this questionnaire. You will receive this questionnaire three times in total. This questionnaire will ask you a number of questions about your knowledge of certain orthopedic diagnoses and treatments. Do not be alarmed if you receive a question about, for example, placing a prosthesis. This does not mean that you will receive a knee prosthesis. It is also okay if you do not know an answer, there are no negative consequences if you give an incorrect answer. We also ask that you do not look up answers on the Internet or discuss the questions with anyone else as this would adversely affect the results of our research. If you don't know an answer, simply enter 'I don't know'. On the second page we will ask you how much knowledge you think you have about certain topics. Please answer honestly here too, we do not expect you to be an expert!

R.1 Gonarthrosis Surgical

Real knowledge

1. *What is knee arthrosis?*

- A. Inflammation of the knee joint
- B. Wear of the knee joint
- C. Bruising of the knee joint
- D. I don't know

2. *How does osteoarthritis cause knee complaints?*

- A. Inflammation of the knee joint
- B. Degeneration of the quality of the cartilage in the knee
- C. The bone is getting thinner
- D. I don't know

3. *Which of the following treatments is a surgical treatment?*

- A. Walking with crutches or a stick
- B. Injection into the knee
- C. Placing a knee prosthesis
- D. I don't know

4. *What is the average lifespan of a knee prosthesis?*

- A. On average 5 to 10 years
- B. On average 10 to 15 years
- C. On average 15 to 20 years
- D. I don't know

5. *Which complication can **not** occur with a knee prosthesis?*

- A. Infection
- B. Injury to blood vessels or nerves
- C. Loosening of prosthesis
- D. Paralysis
- E. Thrombosis
- F. I don't know

6. *What is **not** a purpose of placing a knee prosthesis?*

- A. Improving quality of life
- B. Repairing cartilage
- C. Reducing pain
- D. Restoring mobility
- E. I don't know

7. *How is a knee prosthesis fixed in the bone?*

- A. With plastic
- B. With screws
- C. With cement
- D. I don't know

8. *How long on average do you have to stay in the hospital after knee replacement surgery?*

- A. 1 to 3 days
- B. 4 to 7 days
- C. 7 to 10 days
- D. I don't know

9. *How many months does the average patient receive physical therapy after a knee replacement?*

- A. Less than a month
- B. 1 to 3 months
- C. 3 to 6 months
- D. I don't know

10. *How long on average does it take to fully recover from knee replacement surgery?*

- A. 1 to 3 months
- B. 3 to 6 months
- C. 6 to 12 months
- D. I don't know

11. Which statement about a knee prosthesis is **correct**?

- A. For many patients, the pain decreases, making it easier for them to move
- B. It is safe to participate in activities such as basketball, football and volleyball
- C. Many patients can partially resume their daily activities after 2 weeks
- D. I don't know

12. How long does the average patient still have severe pain after a knee prosthesis surgery?

- A. Less than a week
- B. A week
- C. That varies greatly per patient
- D. I don't know

Perceived knowledge

In this section we would like to ask you to indicate how much you know about each of the topics below.

How much do you know about the following topics:

1. The functioning of your knee and the origin of your complaints

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

2. Knee replacement surgery

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

3. The possible risks and complications of Knee replacement surgery

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

4. Rehabilitation after Knee replacement surgery

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

5. What you can expect from a knee prosthesis

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

R.2 Gonarthrosis Conservative

Real knowledge

1. *What is knee arthrosis?*

- A. Inflammation of the knee joint
- B. Wear of the knee joint
- C. Bruising of the knee joint
- D. I don't know

2. *How does osteoarthritis cause knee complaints?*

- A. Inflammation of the knee joint
- B. Degeneration of the quality of the cartilage in the knee
- C. The bone is getting thinner
- D. I don't know

3. *Which of the following treatments is not a conservative treatment?*

- A. Physiotherapy
- B. Painkillers
- C. Placing a knee prosthesis
- D. I don't know

4. *How long should you take paracetamol each day if this is prescribed?*

- A. 1 week
- B. 2 weeks
- C. 4 weeks
- D. I don't know

5. *What is a reason for not opting for a surgery?*

- A. The damage to the cartilage can still recover
- B. Conservative treatment can improve the complaints for a long period of time
- C. You are relatively old
- D. I don't know

6. *What is not an effect of NSAID drugs?*

- A. Pain relief
- B. Cartilage strengthening
- C. Anti-inflammation
- D. I don't know

7. *Why is it initially tried to postpone surgery for a young osteoarthritis patient?*

- A. A young patient's cartilage can still recover
- B. Young people often do not rest enough after an operation, which causes complications
- C. Young people are more likely to require revision surgery
- D. I don't know

8. *How old is the average patient that receives a knee prosthesis?*

- A. 60 years old
- B. 65 years old
- C. 75 years old
- D. 80 years old
- E. I don't know

9. *What is the average lifespan of a knee prosthesis?*

- A. On average 5 to 10 years
- B. On average 10 to 15 years
- C. On average 15 to 20 years
- D. I don't know

10. *What is a sign of early-stage osteoarthritis?*

- A. Tears in the meniscus
- B. Sinking through your knee
- C. Moving with difficulty
- D. I don't know

11. What is not part of the knee joint?

- A. The meniscus
- B. Cartilage
- C. Thigh bone
- D. Hamstring
- E. Shin
- F. I don't know

12. What is the use of cartilage in the knee?

- A. Keeping the knee joint together
- B. It prevents your knee from cracking
- C. It ensures that the knee can move smoothly
- D. I don't know

Perceived knowledge

In this section we would like to ask you to indicate how much you know about each of the topics below.

How much do you know about the following topics:

1. The functioning of your knee and the origin of your complaints

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

4. The course of conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

2. The conservative treatment options for osteoarthritis

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

5. What results you can expect from conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

3. The potential risks and complications of conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

R.3 Degenerative Meniscus Tear

Real knowledge

1. *What is degenerative meniscus tear?*

- A. Inflammation of the knee joint
- B. Wear of the knee joint
- C. Bruising of the knee joint
- D. I don't know

2. *How does degenerative meniscus tear cause knee complaints?*

- A. Inflammation of the knee joint
- B. Tears in the meniscus
- C. The bone is getting thinner
- D. I don't know

3. *Which of the following treatments is not a conservative treatment?*

- A. Physiotherapy
- B. Painkillers
- C. Keyhole surgery
- D. I don't know

4. *How long should you take paracetamol each day if this is prescribed?*

- A. 1 week
- B. 2 weeks
- C. 4 weeks
- D. I don't know

5. *What is a reason for not opting for surgery?*

- A. The damage to the meniscus can still recover
- B. A surgery increases your chances of getting osteoarthritis in the future
- C. You are relatively old
- D. I don't know

6. *What is not an effect of NSAID drugs?*

- A. Pain relief
- B. Cartilage strengthening
- C. Anti-inflammation
- D. I don't know

7. *What is not a reason why patients with degenerative meniscus tear are often prescribed painkillers?*

- A. To reduce the pain
- B. To restore the patient's posture and gait
- C. To improve the patient's muscle strength
- D. I don't know

8. *Which of the following sports is recommended when you suffer from degenerative meniscus tear?*

- A. Cycling
- B. Running
- C. Soccer
- D. I don't know

9. *When is it wise to opt for surgery if you suffer from degenerative meniscus tear?*

- A. If the patient is in constant pain
- B. If the patient is hindered in his or her work
- C. If the patient is no longer able to fully extend or bend his or her leg
- D. I don't know

10. *What does keyhole surgery (arthroscopy) entail?*

- A. The meniscus is investigated to determine which treatment is most suitable
- B. The meniscus is investigated to determine the severity of the injury
- C. Part of the meniscus is removed
- D. I don't know

11. What is not part of the knee joint?

- A. The meniscus
- B. Cartilage
- C. Thigh bone
- D. Hamstring
- E. Shin
- F. I don't know

12. What is the use of cartilage in the knee?

- A. Keeping the knee joint together
- B. It prevents your knee from cracking
- C. It ensures that the knee can move smoothly
- D. I don't know

Perceived knowledge

In this section we would like to ask you to indicate how much you know about each of the topics below.

How much do you know about the following topics:

1. The functioning of your knee and the origin of your complaints

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

4. The course of conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

2. The conservative treatment options for degenerative meniscus tear

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

5. What results you can expect from conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

3. The potential risks and complications of conservative treatment options

- A. Very little knowledge
- B. Little knowledge
- C. Neutral
- D. Considerable knowledge
- E. A lot of knowledge

S Patient Satisfaction Questionnaire

Dear participant,

Welcome to the last questionnaire. Yesterday you had your consultation at the St. Anna hospital. Hopefully you have experienced this positively. We will ask you a few questions about how you experienced the consultation to find out. Your answers will not be shared with the healthcare provider, so do not hesitate to give an honest answer if you have had a negative experience.

General Care Questions

1. *How do you feel about the amount of time the doctor had for you?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

2. *How thorough / comprehensive was the care you received from the doctor?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

3. *How did you experience the explanations / instructions that the doctor gave you about medication and aftercare?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

4. *What do you think of the advice the doctor gave you about preventing complaints and staying healthy?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

5. *How satisfied are you with the doctor's explanation of what he / she could do for you (examinations, diagnosis, treatment)?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

6. *What do you think of the results of the medical care that you received (to what extent were you helped as you expected)?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

7. *How good do you think the medical care is at the outpatient clinic?*

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

Optional explanation:

Surgeon Specific Questions

What did you think of the doctor's attitude towards you regarding:

1. Helpfulness?

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

4. Professionalism?

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

2. Politeness?

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

Optional explanation:

3. Empathy?

- A. Excellent
- B. Very good
- C. Good
- D. Sufficient
- E. Insufficient

T Subjective Experience Questionnaire

1. Did the computer predict your diagnosis and treatment correctly?

- A. The diagnosis and treatment were both predicted correctly
- B. The diagnosis and treatment were both predicted incorrectly
- C. Only the diagnosis was predicted correctly
- D. Only the treatment was predicted correctly
- E. I don't know

1. To what extent do you think the computer is capable of correctly predicting diagnoses and treatments?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Perfectly

Optional explanation:

2. How pleased were you to receive information about your predicted diagnosis and treatment in advance?

Not pleased 0 1 2 3 4 5 6 7 8 9 10 Very pleased

Optional explanation:

3. How well-prepared did you feel for your first appointment with the orthopedist?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

Optional explanation:

4. To what extent did the information in the app contribute to your sense of preparation?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Entirely

Optional explanation:

5. How did the information in the app affect your experience of the consultation?

- A. Positively
- B. Negatively
- C. Neutral

Optional explanation:

6. Would you like to receive information again next time?

- A. Yes, even if the predictions may be incorrect
- B. Yes, but only if it is certain that the predictions are correct
- C. No

Optional explanation:

7. Did you find the way the information was presented pleasant and clear?

- A. Yes
- B. No

8. Is there anything that you would change? (optional)

9. Do you have any other comments? (optional)

