Improving Follow-Up Strategies for Soft Tissue Sarcoma Patients

A Personalised Approach to Post-Treatment Monitoring

MSc. Thesis

Nina van Staalduine

TU Delft | LUMC Leids Universitair Medisch Centrum

Improving Follow-Up Strategies for Soft Tissue Sarcoma Patients

A Personalised Approach to Post-Treatment Monitoring

Master thesis by
Nina van Staalduine

In partial fulfilment of the requirements for the degree of:

Master of Science
in Engineering and Policy Analysis

at the Delft University of Technology,
to be defended publicly on Friday August 8, 2025 at 11:00 AM

Thesis committee

First Supervisor and Chair:	Dr. P.J. (Perla) Marang-van de Mheen
Second Supervisor:	Dr. I. (Iulia) Lefter
Advisor:	Dr. R. (Rioshar) Yarveisy
External Supervisor:	Dr. R.J.P. (Robert) van der Wal (LUMC)

Preface

Writing a Master's Thesis is well-known for bringing its ups and downs to the student and for me, that was no different. It took months of hard work to get familiar with a new research method and I found myself falling into the same pitfalls over and over again. However, it was a very valuable journey, learning to use a new research method and getting familiar with the probabilistic mindset, even though it took me quite some time.

First of all, I would like to express my gratitude to Rioshar as my daily supervisor. You were always available to answer my questions and I am very thankful that you taught me to work with Bayesian networks. Thank you for your patience while I was trying to get into the probabilistic mindset, but kept falling back to the deterministic one.

Then, I would like to thank Perla, for jumping in on the project when I struggled most. Your input into our meetings was highly valuable and your knowledge about both the modelling and medical world has really helped me get a grip on the project.

This research would not have been possible without all the clinical input from Robert. So, I really want to thank you for always being open for questions or discussions. And I want to thank Iulia for your kind feedback that really helped me to communicate my findings clearer and bring my thesis to a higher level.

Lastly, I want to thank my sister, parents and friends for their continuous support and encouragement, but also bringing me a nice distraction or food when I needed it.

In the end, the ups of the thesis writing definitely outweigh the downs. I have learned a lot during this project, not only about research and data analysis, but also about the complexities of clinical decision-making and communication between the modelling world and the medical world. It also sparked my interest for a, to me, completely new world of science: the medical world. All in all, an extremely valuable experience.

Abstract

Soft tissue sarcomas (STS) are rare cancers with a high risk of recurrence, therefore requiring structured and effective follow-up strategies. Current follow-up guidelines, such as those provided by the NCCN and ESMO, use fixed time intervals and do not fully account for individual patient risk factors when scheduling follow-up visits. Additionally, adherence to these guidelines is inconsistent, with some patients receiving unnecessary follow-up visits while others do not receive sufficient monitoring. This inefficiency places a burden on healthcare resources while potentially impacting patient outcomes.

This research aimed to develop a decision-support tool, in the form of a Bayesian network, that helps to personalise follow-up strategies. It does so by combining insights from historical patient data and expert knowledge to estimate recurrence risk, where risk is based on both likelihood of a recurrence and the severity of a potential recurrence and its treatment. The tool shows how frequent similar patients were historically followed-up and what their outcomes were. Rather than giving recommendations, the model supports clinicians by providing insights into recurrence patterns and follow-up intensity among similar cases, helping clinicians reflect on what might be appropriate for a given case.

The study consisted of three phases: identifying key recurrence risk factors and follow-up patterns through data analysis and literature review, developing and validating the Bayesian model in consultation with clinical experts and finally analysing mortality and economic effects for different follow-up frequencies.

Tumour grade and surgical margin were found to be most important risk factors for recurrence, with grade being the strongest. Two main model structures were made, one including both tumour grade and surgical margin to estimate probability of recurrence, while the other included only tumour grade as a reflection of probability of recurrence. Both models contained the variables expected remaining lifespan and fitness of the patient to estimate consequence of recurrence. The models showed considerable variation in risk classification compared to expert opinion and tumour grade alone. Since there is not one ‘correct’ way to classify risk, this reflects the perceptive nature of risk and the different models offer different viewpoints. Additionally, the study found no clear survival benefit from more intensive follow-up in any of the models, while it did substantially increase costs, especially per survivor. This highlights the need for clinicians to reflect on and discuss optimal allocation of follow-up resources.

By developing a structured, evidence-based approach to follow-up, this study contributes to improving STS care. The proposed decision-support tool could help clinicians make more transparent, risk-based follow-up decisions, while reducing unnecessary burdens on healthcare systems.

Contents

List of Abbreviations

STS	Soft tissue sarcoma
LUMC	Leiden University Medical Centre
ESMO	European Society of Medical Oncology
NCCN	National Comprehensive Cancer Network
LR	Local recurrence
DM	Distant metastasis
OS	Overall survival
PRS	Post recurrence survival
DSS	Disease-specific survival

1. Introduction

1.1 Problem Statement

Soft tissue sarcomas (STS) are among the rarest forms of cancer, accounting for less than 1% of all cancer cases in the Netherlands, with an estimated 1200 yearly new cases (IKNL, 2025a; IKNL, 2025b). These tumours originate in soft tissues such as muscles and fat, and are known for their relatively high recurrence risk compared to other cancers, with a reported 5-year recurrence rate of 37% (Li et al., 2011). Gerrand et al. (2003) found a 10% local recurrence rate within five years, while 27% of patients developed metastases. This high recurrence risk highlights the need for effective follow-up care to enable early detection and treatment, ultimately improving patient outcomes.

The main goal of follow-up care is to allow clinicians to detect recurrences as early as possible. The idea behind it is that the sooner a recurrence is identified, the better the possibilities are for further treatment and survival. In contrast, if a recurrence is detected later, it may have had more time to grow or spread, which could reduce the effectiveness of further treatment options.

However, not all patients have the same risk of recurrence. This risk varies per patient, depending on several tumour characteristics and clinical factors. For instance, tumour size has been associated with the likelihood of distant metastases (Zhao & Yang, 2011), while tumour grade plays a central role in prognosis, with high-grade tumours linked to higher recurrence rates (Posch et al., 2017). Histological subtype affects local recurrence (Li et al., 2011), and surgical margins after treatment predict both local recurrence and metastasis (Li et al., 2011). These and other factors could serve as the basis for a more individualised approach to follow-up care, but they are not currently used in standard follow-up guidelines.

While some centres, including LUMC, have developed their own guidelines, both the National Comprehensive Cancer Network (NCCN) and the European Society of Medical Oncology (ESMO) currently provide guidelines for follow-up frequency after STS treatment (Gronchi et al., 2021; Von Mehren et al., 2022). These guidelines are, however, subject to debate and face two criticisms. First, they lack personalisation, as individual risk factors are not incorporated in the follow-up schedules, leaving clinicians to make their own assessment whether to deviate from these schedules (Tseng et al., 2015). Secondly, the guidelines are fixed and based on given calendar intervals, rather than dynamic and based on individual risk factors or evidence (Dammerer et al., 2020). Moreover, adherence to these guidelines is inconsistent. A study by Kruiswijk et al. (2023) found that only 24% of post-treatment STS patients actually received the suggested number of follow-up visits according to guidelines within the first year post-surgery. Some patients received too few follow-up visits, while others received too many.

Undertesting logically seems undesirable, as it risks delayed recurrence detection. However, overtesting is also not desirable. Firstly, unnecessary follow-up visits put both a financial and logistical burden on the already overloaded health system. Secondly, while more frequent follow-up imaging is associated with higher disease-specific survival for STS patients (Chou et al., 2012), it does increase radiation exposure which may cause harm and could cause extra psychological stress.

At this moment, known risk factors for recurrence are not systemically used to personalize follow-up strategies and existing guidelines are not consistently followed. Since the follow-up period can be stressful for patients (Lenze et al., 2019), a structured, but personalised approach is needed. A personalised approach that allows follow-up frequency based on individual recurrence risk rather than fixed schedules for all patients.

1.2 Research objective

This research aims to improve follow-up strategies for STS patients by balancing resource allocation and patient outcomes. It does so by integrating clinical data and expert opinion into a probabilistic Bayesian model, based on known recurrence risk factors and their impact on patient survival. Traditional follow-up schedules often rely on fixed time intervals, which may not accurately reflect individual patient risk. This research aims to move beyond this one-size-fits-all approach by understanding key predictors of recurrence and incorporating them into a risk-based follow-up model.

First, an analysis of historical patient data, in combination with findings from literature, will identify key predictors of recurrence. These insights are then used to build a Bayesian decision-support model that assigns patients to risk categories and shows what follow-up frequencies similar patients had. The model is clinically validated through expert input to ensure its relevance and reliability. Finally, the study explores how the model could support follow-up decisions by showing what follow-up frequencies and outcomes were observed among historically similar patients, highlighting potential patterns in survival and the associated healthcare costs.

By balancing patient well-being with hospital resource efficiency, this study aims to provide a more effective follow-up strategy that minimizes unnecessary hospital visits, while ensuring timely detection of recurrences. The results will offer a structured, evidence-based framework for optimising follow-up policies in clinical practice.

1.3 EPA Relevance

Managing STS follow-up care is part of a broader challenge in healthcare: finding the right balance between personalized care and safe and efficient use of resources. Rising healthcare costs and an overburdened system require smarter follow-up strategies. Patient outcomes should be optimised, while minimizing unnecessary hospital visits. This research aligns with the Engineering and Policy Analysis (EPA) program by applying quantitative data analysis, probabilistic modelling, and decision-making frameworks to this complex societal challenge.

The development of a decision-support model, using probabilistic methods, directly connects to EPA's focus on quantitative modelling for policy and decision-making. Moreover, the combination of insights from data, expert knowledge, and probabilistic modelling reflects the EPA approach of applying analytical methods to complex, real-world, multi-stakeholder problems. Examples of stakeholders are the patients, clinicians, hospitals and insurance companies. The aspect of efficient resource allocation also reflects the management perspective within the faculty of Technology, Policy and Management. This study will demonstrate how EPA methods can support decision-makers, in this case clinicians, to make risk-based follow-up decisions in a more structured and evidence-based way that may result in better logistics in the hospital and cost savings.

1.4 Thesis Structure

This thesis is structured to provide a clear and logical overview of the research. After the introduction, the literature review introduces key concepts, describes current literature and knowledge gap which leads to the main research question. Subsequently, the study design describes the main research approach, materials used and research methodology. Next, the results are presented and discussed, followed by a final conclusion.

2. Literature Review

This chapter aims to provide an overview of the current state-of-the-art in STS follow-up guidelines and strategies. The literature review will first define core concepts in the field and then discuss and synthesize the main findings from the existing literature. These findings will highlight the knowledge gap and lead to the formulation of the research question for this study.

2.1 Core Concepts

Soft tissue sarcomas (STS) are rare malignant tumours that develop in soft tissues, such as muscles, fat, nerves and blood vessels. Characteristics of STS can be described by the following variables:

- **Tumour size**: The diameter of the tumour.
- **Tumour location**: The location in the body where the tumour is found. Most STS tumours occur in the lower extremities (hips, legs and feet), upper extremities (arms and hands) or the pelvis (Kruiswijk et al., 2024).
- **Tumour grade**: The primary grading system for STS was developed by the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC), categorising the tumours into 3 grades (Trojani et al., 1984). Grade 1 is given to tumours that show little aggressive characteristics and with a relatively low chance of metastasis, while grade 3 tumours have very aggressive characteristics (e.g. high speed of cell division) and have an increased risk of fast spreading and metastasis. Grade 2 is in between these grades.
- **Histological subtype**: The histological subtype refers to the specific cell type of which the STS consists. Two examples are liposarcoma, which originates from fat cells, and leiomyosarcoma, which originates from smooth muscle cells.

Once a patient is diagnosed with STS, he or she can be treated. This typically involves a **surgery**, where the tumour is resected from the body. After the tumour is taken out at surgery, the pathologist checks whether there are still cancer cells remaining at the surgical margins, meaning the cut surfaces of the tissue that was removed. The **surgical margin** is usually classified as follows:

- R0: There are no cancer cells left
- R1: There are still cancer cells left, but only visible with a microscope
- R2: There are still cancer cells left and they are visible to the naked eye

In addition, **radiotherapy** or **chemotherapy** can be administered to the patient, both aiming to destroy (remaining) cancer cells. These therapies can be given **neoadjuvant** (before surgery) usually aiming to shrink the tumour for easier resection, or **adjuvant** (after surgery) to target and eliminate remaining tumour cells after resection.

It is good to note that the tumour and treatment characteristics are **interrelated**. For instance, low-grade tumours are usually treated with only surgery, high-grade tumours also have radiotherapy as part of the standard treatment (Gronchi et al., 2021). Furthermore, tumour size can affect the surgical margin that can be achieved, which in turn may influence the decision to administer radiotherapy beforehand to shrink the tumour and allow for a better resection.

The **recurrence risk** refers to the likelihood and consequence (severity) of a tumour returning after initial treatment. It can be a **local recurrence (LR)**, meaning that it recurs at the original site of the tumour, or it can be **distant metastasis (DM)**, recurring at another site in the body than the original tumour. The previously mentioned tumour and treatment characteristics are possible risk factors for recurrence.

After initial treatment of STS, patients receive **follow-up (FU) care**. The purpose of FU care is early detection of recurrence. The follow-up visit is usually a combination of physical examination and **imaging**. The imaging can be done using an MRI-scan for detection of local recurrence (Gronchi et al., 2021). However, as STS often recurs in the lungs, chest X-rays (Tseng et al., 2015) or CT-scans (Gronchi et al., 2021) can also be done for detection of metastasis in the lungs.

Both the European Society of Medical Oncology (ESMO) and the National Comprehensive Cancer Network (NCCN) currently provide **guidelines** for follow-up schedules (Gronchi et al., 2021; Von Mehren et al., 2022). An overview of these guidelines regarding follow-up schedules is given in Table 1 below. It is good to note that while these guidelines offer clear schedules, they are less strict on what should be done during these follow-up visits. The ESMO guideline suggests MRI for local recurrence detection and a CT-scan for lung metastases (Gronchi et al., 2021), while the NCCN guideline suggests a combination of physical examination with any imaging modality (Von Mehren et al., 2022)

Table 1: Overview of current ESMO and NCCN follow-up guidelines (Gronchi et al., 2021; Von Mehren et al., 2022)

		Year									
Guideline	Type of patient	1	2	3	4	5	6	7	8	9	10
ESMO	Low-grade	Every 6 months					Annually				
	Intermediate- and high-grade	Every 3 to 4 months			Every 6 months		Annually				
NCCN	All	Every 3 to 6 months			Every 6 months		Annually				

As Table 1 shows, the ESMO guidelines distinguish between low-grade and intermediate- or high-grade, while the NCCN does not. The ESMO suggests a follow-up interval of 6 months in the first five years for low-grade patients, while it suggests following-up every 3 to 4 months in the first 2 to 3 years and every 6 months up until the 5th year for high-grade patients. The third year is flexible with suggested follow-up every 3 to 6 months for high-grade patients. The NCCN also suggests a more frequent follow-up in the first 2 to 3 years, namely every 3 to 6 months, and every 6 months up until the 5th year. In this guideline, the third year is also flexible with suggested follow-up every 3 to 6 months. However, the guideline does not indicate on what this flexibility should be based. After the 5th year, both the ESMO and NCCN guideline suggest annual follow-up. Note that the table depicts follow-up until the 10th year, but there is no final year defined in the guidelines.

Table 1 also shows that the current guidelines give a fixed schedule for all patients, with little room for personalisation, except in the first 2 to 3 years. In the cases where some flexibility is indicated, the final decision on follow-up frequency is left to the clinician's judgement. However, by including patient-specific risk factors into the decision-making for the follow-up schedule, it will become more flexible and adaptive. This would allow for more **personalised follow-up schedules** to be created that are still evidence-based.

2.2 Literature Review Approach

For this literature review, a search string was used including terms that are relevant given the topic of this study. The search was kept quite broad on purpose, to find as many relevant publications as possible, in a field of science where a lot of different terms are used for similar concepts. The search string is as follows:

"(TI=("soft tissue sarcoma"OR "STS")) AND (TS=(recurrence OR "follow-up" OR surveillance OR monitoring OR guideline OR predict* OR risk)"*

Note that the search string contains 'TI 'and 'TS'. 'TI' means that the strings that follow it should be in the title, while 'TS' means that the strings that follow it should be in the topic, thus the title, abstract and keywords. The search string was used on Web of Science. Nowadays, many advancements are made in the medical world, so the search results were filtered on publications from 2014 or later to ensure the most recent and relevant insights are included. Furthermore, only articles and reviews having 15 citations or more were included as an indicator of relevance judged by other experts in the field who cite the paper. While screening the titles and keywords, relevant publications were divided into 3 categories: follow-up guidelines, predictive models (for recurrence and survival) to support decision-making and risk factors for recurrence. These categories reflect the different elements in the objective of this study. The entire process of literature selection can be found in Figure 1.

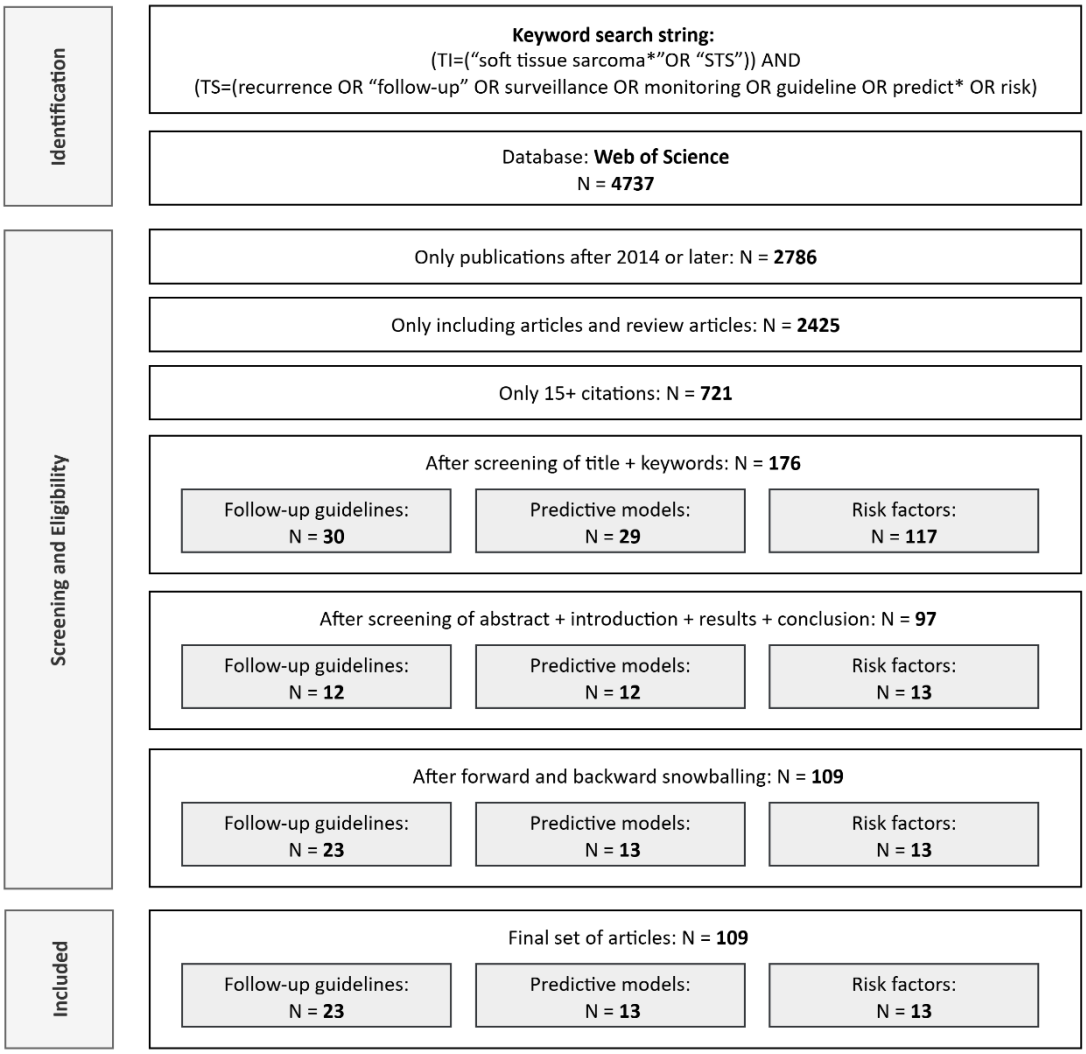


Figure 1: Selection process of literature

2.3 Main Findings and Synthesis

The total set of 49 articles were already grouped during the process of literature selection. For each of the 3 categories, an overview table was made and the findings synthesized. After these sections, the knowledge gap will be described, leading to the main research question.

2.3.1 Follow-up Guidelines

The entire table including the overview of the literature within the category 'follow-up guidelines', is shown in the appendix (A.1.1 Follow-up Guidelines). The articles within this category were divided into two subcategories: perspectives on current follow-up guidelines and strategies and effectiveness of follow-up strategies. As both subcategories are relatively large, they are synthesized separately.

Perspectives on current follow-up guidelines and strategies

The current literature shows considerable differences in follow-up strategies for STS patients. Even though national and international guidelines exist, there is no unified or universally accepted standard for follow-up care (Patel & Matcuk, 2018). Most guidelines recommend regular physical examinations and imaging, such as chest X-rays, CT scans or MRI. However, the way that these check-ups are scheduled is not individualised. Multiple studies call for an individualised approach based on tumour characteristics and recurrence risk (Cipriano et al., 2020; Dammerer et al., 2020; Ezuddin et al., 2018; Kruijswijk et al., 2024; Noebauer-Huhmann et al., 2024; Tseng et al., 2015; Rothermundt et al., 2023; Spunt et al., 2020; Whitaker et al., 2023; Zaidi et al., 2018). While multiple studies are advocating for an individualised approach, there is not yet a practical implementation.

Risk factors mentioned in guidelines are tumour grade, histological subtype, tumour size and margin status (Rutkowski & Ługowska, 2014; Spunt et al., 2020). Spunt et al. (2020) studied a risk-based treatment strategy for STS patients under 30 with non-rhabdomyosarcoma and divided the patients in their study into three risk groups: high, intermediate and low. These groups were based on tumour grade, tumour size, surgical margin and whether the tumour was metastatic or not. While the risk groups could predict event-free and overall survival, the criteria used to define them are rather fixed. For instance, patients would only be high risk if the tumour was metastatic.

Patel & Matcuk (2018) state that imaging is an essential part of STS follow-up and new imaging techniques (PET/MRI, DWI, DCE-MRI) have potential, but are costly and have limited availability. They also note that the costliness, limited availability and broad diversity of sarcomas are causing the big variations in follow-up strategies. Multiple studies also indicate the importance of taking the tumour site into account when deciding on follow-up strategies (Blay et al., 2022; Ezuddin et al., 2018; Rutkowski & Ługowska, 2014; Van Ewijk et al., 2021). Imaging should be site-specific, for instance MRI should only be used if the tumour site is not easily accessible (Ezuddin et al., 2018).

Clinical guidelines often suggest more intensive surveillance during the first few years post-treatment, with deceasing frequency over time (Von Mehren et al., 2022; Ferré et al., 2021; Gronchi et al., 2021). However, international surveys found that there is a significant variation in clinical decision-making in practice, indicating that follow-up strategies are influenced by regional and specialty-based differences (Acem et al., 2021). Moreover, Kruiswijk et al. (2024) showed that adherence to follow-up guidelines is inconsistent, with younger patients and those with high-grade tumours receiving more frequent surveillance. Also, many patients without recurrence received more visits than recommended by guidelines. This suggests that clinicians may struggle to accurately incorporate risk of recurrence in their decisions about follow-up frequency, which highlights a need for a decision-support tool that can integrate risk factors to develop individualised evidence-based follow-up schedules.

Effectiveness of follow-up strategies

Several studies indicate that more intensive surveillance does not necessarily improve survival rates (Puri et al., 2014; Glasbey et al., 2021). A randomized trial by Puri et al. (2014) found no significant survival benefit from more frequent visits or the use of CT scans over chest X-rays. A cost-

effectiveness analysis suggests that chest X-rays are sufficient for low-risk patients, while CT scans should be reserved for high-risk cases (Royce et al., 2017).

Sawamura et al. (2014) reported that the majority of local recurrences occur in the first 7.1 years and for metastasis in the first 4.2 years. Furthermore they state that high-grade tumours tend to show more recurrences earlier on, while low-grade tumours have a more constant rate over time. There are few events after 10 years, indicating that surveillance after 10 years might not be necessary.

Clinical symptoms are often the first indication of recurrence. Studies show that most local recurrences are detected by patients themselves through symptoms such as pain or palpable mass (Blaye et al., 2019; Cheney et al., 2014; England et al., 2020; Fujiki et al., 2016). In many cases, routine imaging fails to detect asymptomatic recurrences, questioning its utility in certain patient groups. For example, Cheney et al. (2014) found that only 1 out of 11 local recurrences in their study was detected by surveillance MRI, the rest was discovered through patient-reported symptoms. They emphasize the selective use of MRI for tumour sites that are hard to assess physically. England et al. (2020), however, stated that imaging still detected a substantial proportion of recurrences that were otherwise asymptomatic. The studies from both Cheney et al. (2014) and Fujiki et al. (2016) emphasize the importance of imaging at tumour sites that are not easily accessible.

Sawamura et al. (2014) states that the current guidelines for follow-up lack evidence. Other studies advocate for more modelling studies in this field, so that better tumour predictions and individualised follow-up strategies can be made (Blaye et al., 2019) and to optimise the imaging frequency (England et al., 2020). These findings emphasize the need for an evidence-based approach that optimizes follow-up frequency and modality based on individual patient risk. This would reduce unnecessary procedures while ensuring effective recurrence detection.

2.3.2 Predictive Models to Support Decision-making

The entire table including the overview of the literature within the category ‘predictive models to support decision making’, is shown in the appendix (A.1.2 Predictive Models).

Over the past years several prognostic models have been developed to predict outcomes and thereby support clinical decision-making for patients with STS. Well-known models like the Sarculator and PERSARC use retrospective clinical data to predict overall survival (OS), local recurrence (LR) and distant metastases (DM) (Callegaro et al., 2016; Callegaro et al., 2019; Pasquali et al., 2022; Van Praag et al., 2017; Willeumier et al., 2017; Rueten-Budde et al., 2018; Smolle et al., 2020; Rueten-Budde et al., 2021). These models are generally based on factors like age, tumour size, tumour grade, histological subtype and treatment-related variables such as surgical margin and radiotherapy or chemotherapy and are well-validated. Some of these models are static, meaning that they give one prediction for the entire follow-up, while others are dynamic and can update predictions during the follow-up period. Besides the Sarculator and PERSARC, several other types of models have been developed, including nomograms (Shuman et al., 2015; Li et al., 2020; Gu et al., 2020), gene expression-based models (Gu et al., 2020; van IJzendoorn et al., 2019), machine learning approaches (Van IJzendoorn et al., 2019) and multi-state models (Posch et al., 2017). Even though there are quite some developments in this field, there still remain some knowledge gaps.

Firstly, very few models translate risk predictions into actionable follow-up strategies. While several models can predict risk of local recurrence or distant metastasis, they rarely provide guidance on how follow-up schedules could be adjusted based on those risks. The study of Smolle et al. (2020) mentions the idea of individualised follow-up, but actual implementation or validation of follow-up

protocols is missing. Secondly, most models focus on extremity STS. This means that other patient groups, like retroperitoneal or trunk tumours, are often excluded. This limits the generalisability and clinical usefulness of the models in non-extremity tumours. Thirdly, the current models are purely clinical and do not include patient preferences or aspects of quality of life of the patients in their predictions. While survival and recurrence are important, factors like the impact on daily life imposed by more frequent follow-up or the willingness to undergo intensive follow-up also matter.

In short, the existing models are a useful starting point for risk prediction, but there is still a need for models that translate risk into actionable follow-up decisions, include broader patient groups and incorporate patient preferences.

2.3.3 Risk factors for Recurrence

The entire table including the overview of the literature within the category 'risk factors for recurrence', is shown in the appendix (A.1.3 Risk factors for Recurrence). As there were too many different risk factors in this literature review to put in one overview table, there is an overview table for the most commonly discussed risk factors: age, tumour size, histologic grade, tumour site, surgical margin, radiotherapy, chemotherapy, histological subtype and tumour depth. These frequently mentioned factors are likely the most relevant. The fact that they are widely discussed in literature indicates that they are well-known in the clinical field and therefore likely used by clinicians in their decision-making processes. For the risk factors that were discussed less often, a textual section is made, also found in the appendix (A.1.3 Risk factors for Recurrence). Furthermore, it is important to note that the term 'risk of recurrence' typically refers only to the probability of recurrence in literature, and does not take the potential consequences or severity of a recurrence into account.

Age

There is consensus in the literature about the influence of age in the context of STS recurrence and overall survival. Although there is some variability, a patient who is older than 45 to 60 years is generally considered to have a worse prognosis for recurrence and survival. Even though Daigeler et al. (2014) did not find an association between age and post recurrence survival, other studies found that being older than 45 is associated with higher disease-specific mortality (Maretty-Nielsen et al., 2014) and patients older than 60 (Park et al., 2015) or 74 (Smith et al., 2016) years are associated with worse disease-specific survival.

Furthermore, STS patients over 55 (Italiano et al., 2014; Maretty-Nielsen et al., 2014; Tsagozis et al., 2015), over 70 (Vodanovich et al., 2019) and over 74 (Smith et al., 2016) years of age are associated with higher rates of local recurrence. Lastly, Tsagozis et al. (2015) found that STS patients older than 55 were more likely to have distant metastasis.

Tumour size

Most studies indicate that a larger tumour size is associated with more disadvantageous outcomes for the patient. Multiple studies report a link between tumours larger than 4 or 5 centimetres and distant metastasis (Italiano et al., 2014; Vodanovich et al., 2019; Baroudi et al., 2014; Sugiura et al., 2014; Tsagozis et al., 2015). Smith et al. (2016), however, indicated that only tumour larger than 14 centimetres show an association with distant metastasis.

Toulmonde et al. (2014) found that tumours larger than 10 centimetres are linked to late local recurrences and Park et al. (2015) showed worse locoregional control in tumours above this size. Italiano et al. (2014), on the other hand, found no association between tumour size and local recurrence. Maretty-Nielsen et al. (2014) showed that tumours smaller than 4 centimetres are associated with better outcomes regarding local recurrence and disease-specific mortality. Smith et

al. (2016) also found that tumours larger than 6 centimetres are associated with worse disease-specific survival.

The association between tumour size and overall survival is less clear. Bonvalot et al. (2017) found no link between tumour size and overall survival, whereas Galy-Bernadoy & Garrel (2016), however, did find a link between tumour size and overall survival in the context of head and neck STS. Baroudi et al. (2014), however, reported that tumours smaller than 4 centimetres are associated with better overall survival.

Grade

The tumour grade seems a consistent predictor of patient outcomes, with high-grade tumours generally linked to poorer prognosis. All studies show that a higher tumour grade is associated with worse outcomes. The distinction is typically made between low-grade (grade 1) and high-grade (grades 2 and 3). High-grade tumours are associated with higher rates of local recurrence (Maretty-Nielsen et al., 2014; Italiano et al., 2014; Tsagozis et al., 2015), more distant metastasis (Italiano et al., 2014; Sugiura et al., 2014), late distant metastasis (Toulmonde et al., 2014), worse post-recurrence survival (Daigeler et al., 2014) and worse overall survival (Bonvalot et al., 2017).

Smith et al. (2016) indicate that low-grade tumours have better outcomes in terms of local recurrence, distant metastasis and disease-specific survival. Furthermore, more favourable results are reported for low-grade tumours regarding distant metastasis (Chou et al., 2016), disease specific survival (Park et al., 2015) and overall survival (Chou et al., 2016; Baroudi et al., 2014; Park et al., 2015).

Tumour site

Several studies suggest that tumour location might play a role in outcomes, but findings are not always consistent. Daigeler et al. (2014) point out that sarcomas located in the extremities were associated with better post-recurrence survival, while those in the trunk showed worse outcomes. Similarly, Vodanovich et al. (2019) reported better overall survival for sarcomas in the upper extremities, whereas sarcomas in the head and neck had worse overall survival (Galy-Bernadoy & Garrel, 2016). Maretty-Nielsen et al. (2014) found that both trunk and lower extremity sarcomas were associated with worse disease-specific mortality.

Regarding local recurrence, limb tumours had better outcomes, while tumours in the head, neck and trunk had worse outcomes (Italiano et al., 2014). Sugiura et al. (2014), however, found that both upper extremity and trunk tumours were associated with higher rates of local recurrence and Toulmonde at al. (2014) found internal trunk sarcomas to be associated with late local recurrence.

Surgical margin

Surgical margin appears strongly associated with local recurrence, but the findings on overall survival vary. While Bonvalot et al. (2017) indicate there is no association between surgical margin and overall survival, Galy-Bernadoy & Garrel (2016) indicate that there is. Several studies indicate an association between surgical margin and local recurrence, where positive margins (R1 and R2) are associated with worse outcomes (Maretty-Nielsen et al., 2014; Vodanovich et al., 2019; Smith et al., 2016). Similarly, an R0 margin was found to have better outcomes regarding local recurrence (Chou et al., 2016) and post-recurrence survival (Daigeler et al., 2014).

Radiotherapy

Radiotherapy is generally linked to better outcomes, particularly for local control and survival. Both Maretty-Nielsen et al. (2014) and Smith et al. (2016) indicate that radiotherapy has an advantageous effect on the patient outcomes regarding local recurrence. Maretty-Nielsen et al. (2014) also state

that radiotherapy is associated with lower disease-specific mortality. Furthermore, adjuvant radiotherapy is associated with better local control (Italiano et al., 2014) and less distant metastasis (Vodanovich et al., 2019).

Chemotherapy

The effect of chemotherapy seems inconsistent and might depend on tumour grade. A study from Daigeler et al. (2014) showed no association between chemotherapy and post-recurrence survival. Italiano et al. (2014), however, found that administering adjuvant chemotherapy is associated with better outcomes regarding distant metastasis. Though, adjuvant chemotherapy only appeared beneficial for grade 3 sarcomas.

Histological subtype

There is no clear consensus regarding histological subtype as a risk factor for recurrence and survival, although certain histological subtypes are associated with worse outcomes. While Smith et al. (2016) state that undifferentiated pleomorphic sarcoma had worse outcomes, they also state that histological subtype is no risk factor for local recurrence, distant metastasis or disease-specific survival. Italiano et al. (2014), on the contrary, indicate that the histological subtype is a risk factor for both local recurrence and distant metastasis. Bonvalot et al. (2017) state that only leiomyosarcoma has worse outcomes regarding overall survival.

Tumour depth

Some studies show that deep tumours are linked to worse outcomes, but others find no association at all. Multiple studies indicate that tumour depth is a risk factor for recurrence. Deep tumours are linked to more local recurrence (Smith et al., 2016; Sugiura et al., 2014) and distant metastasis (Italiano et al., 2014; Sugiura et al., 2014). However, two studies show no relation between tumour depth and local recurrence (Maretty-Nielsen et al., 2014; Italiano et al., 2014). Maretty-Nielsen et al. (2014) also indicates no relation between disease-specific mortality and tumour depth.

2.4 Knowledge gap and Research Question

Although a wide range of guidelines, predictive models, and clinical studies exist investigating the follow-up of soft tissue sarcoma (STS) patients, current follow-up strategies remain largely uniform and do not adequately incorporate individual patient risks. Multiple studies call for a more personalized approach based on recurrence risk, but there is not yet an established method to translate known risk factors into concrete, individualized follow-up schedules. Existing prognostic models, such as Sarculator and PERSARC, are valuable for predicting outcomes like survival and recurrence but do not offer guidance on how to adapt follow-up protocols accordingly. Moreover, these models often focus on extremity STS, limiting their relevance for patients with other tumour locations, and do not incorporate patient preferences or quality-of-life considerations that might be relevant in decisions around the number of follow-up visits.

Additionally, evidence suggests that clinicians currently seem to rely on their own judgement to decide on follow-up schedules, as decisions are inconsistent with current guidelines, which may not always be aligned with risk levels. This gap between predictive capability and practical application highlights the need for an integrated tool that supports clinicians in making evidence-based, patient-centred decisions. Such a tool should combine clinical risk factors, follow-up evidence, and expert knowledge to generate interpretable follow-up recommendations. Approaches that allow for this kind of integration, especially those that can handle uncertainty and support decision-making under variable conditions, are currently missing in the field. A decision-support system based on a Bayesian

network could fill this gap. The knowledge gap described in this section leads to the main research question of this study:

How can historical patient data and expert knowledge be integrated to inform a decision-support system to improve STS follow-up relying on Bayesian Networks, balancing hospital resource efficiency and patient well-being?

To answer the main research question, several sub-questions have been formulated. These sub-questions are given and explained briefly below.

SQ1: What risk factors for recurrence and follow-up patterns can be identified through the exploration of historical STS patient data?

This sub-question focuses on identifying risk factors associated with recurrence and follow-up patterns for STS patients by analysing historical patient data and literature. The goal is to determine which factors have the strongest conditional dependency with recurrence (and survival), as these factors will also influence clinician decision making regarding follow-up visits in daily practice, forming the foundation for the risk-based Bayesian model. First, an exploratory discussion with an STS expert allowed for creating a preliminary model structure, based on the available variables in the dataset. After cleaning and structuring the historical patient data, conditional dependencies are analysed between the available variables and recurrence to confirm or adapt the model structure. We examine which clinical variables often occur together e.g. some histological tumour types are always high grade tumours, as that means only one of these variables needs to be used and results in a simpler and easier to use model. Additionally, conditional dependencies between risk level, follow-up adherence, and survival were explored to understand how they relate. These analyses lay the foundation for developing a probabilistic model capable of predicting recurrence risk based on patient-specific factors.

SQ2: How can patient risk factors be integrated into a clinically validated Bayesian decision-support tool to guide risk-based FU-frequency?

The aim of this sub-question is to construct and validate a probabilistic model that supports risk-based decisions on follow-up frequency. The first step of this sub-question is to develop the structure of the model in the form of an influence diagram. The structure is based on findings from the first sub-question, where important risk factors are identified, combined with input from an STS clinician. The STS clinician not only validates the identified risk factors, but also helps to understand the underlying mechanism by which these factors influence the risk of recurrence, to develop a well-structured influence diagram that only contains the necessary risk factors. Once the structure of the model is settled, the focus lies on adding data to each part of the model. Conditional probability tables will be derived from the historical data to inform the Bayesian network.

In addition, a survey was conducted to validate the model structure and to complete the conditional probability tables for variables not present in the dataset, specifically, those related to the patient-centred aspect of the model. While the model considers the likelihood of recurrence, reducing the burden on the healthcare system also requires accounting for the severity or consequences of recurrence, which is captured through this patient-centred aspect. Although the survey collected a broad range of information, its primary focus was to better understand this patient-centred dimension, as it represents a critical but previously unmeasured factor in clinical decision-making. The survey also evaluated clinicians' willingness to potentially use the proposed model.

SQ3: To what extent could a risk-based follow-up approach guided by the model, improve care efficiency by reducing unnecessary follow-up visits and associated healthcare costs without compromising patient outcomes?

The goal of the third sub-question is to evaluate how follow-up frequency relates to mortality rates and to estimate the economic impact of different follow-up intensities for STS patients. We analyse whether earlier detection of recurrence through more frequent follow-up actually improves survival outcomes for patients. Using historical patient data, survival outcomes are compared across varying follow-up schedules, incorporating risk classifications from the model developed in the second sub-question and comparing them with other classifications. The findings contribute to understanding if higher follow-up frequency translates into better survival, which in turn informs whether the additional healthcare resources are justified.

3. Study design

3.1 Research approach

This research aims to develop a decision-support tool to make follow-up strategies individualised and more efficient for STS patients and improve utilisation of hospital resources. A probabilistic modelling approach will be taken, using Bayesian networks. Bayesian networks were primarily chosen because their main purpose is “to update beliefs in response to evidence” (Pearl, 1988) which is aligned with the problem faced by clinicians when deciding on follow-up strategies.

In the context of STS follow-up, Bayesian networks can help estimate a patient’s risk level by combining available information in a systematic way. As Pearl (1988) described, these models use evidence (known characteristics of the patient) to update prior beliefs about outcomes. If we had no specific information about a patient, our prediction would be based on an average of what happened to all patients in the past. However, once we do know certain characteristics (such as tumour grade, size, or surgical margins), we can update the model with this *evidence*. The model then identifies patients in the historical data who closely match the current patient, and gives us an estimate of what typically happened to similar patients, e.g. if they experienced a recurrence.

In other words, we use all known characteristics at the decision point to personalise follow-up care. It allows clinicians to systematically incorporate the most current understanding of a patient’s situation and estimated risk into determining the follow-up schedule, based on outcomes observed in similar cases.

There are several additional advantages to Bayesian networks, which make it a well-suited method for this research. Firstly, Bayesian networks are probabilistic models that are able to represent and incorporate uncertainty (Pearl, 1988). This is particularly relevant for this research, as patient outcomes (such as the effect of radiotherapy) can vary widely. Secondly, Bayesian networks allow for the integration of domain knowledge to supplement empirical data (Wang et al., 1998). Since data availability is limited due to the rarity of the STS, it is a valuable capability in this context. Lastly, Bayesian networks are easily interpretable graphical models. While the algorithm behind the model is complex, the visual structure clearly shows the variables that factor into the decision making, making it easier to understand and trust the model’s reasoning. This interpretability is crucial for gaining acceptance in clinical settings, where clinicians often do not have a background in modelling.

The research follows a structured methodology divided into three main parts, each addressing one of the sub-questions. Each sub-question, and how it contributes to answering the main research question, is described in section 3.2. However, to gain knowledge on the main approach of this research, the following section explains the foundations of Bayesian networks.

3.2 Materials

This section describes the materials used in this study. It introduces and explains the fundaments of the Bayesian modelling approach, describes the data sources and includes a brief statement on the use of AI tools.

3.2.1 Explanation of the Bayesian Approach

A Bayesian network consists of both a graphical structure and quantitative parameters that lie behind this structure. This section explains how both parts are constructed.

Graphical structure of the model

The structure of a Bayesian network can be visualized using an influence diagram, which illustrates how variables are connected and influence one another. The network consists of both nodes and edges, where nodes represent the variables and edges represent a conditional dependency between two variables (Cowell et al., 1999). Edges in Bayesian networks are always directed, meaning that each edge is an arrow pointing from one node to another. The node where the arrow starts is called the *parent* and the node it points towards is the *child*. A simple example, in the form of an influence diagram, is given in Figure 2. Each node is associated with a probability function, that takes as input a set of values for the node's parent variables, to give (as output) the probability of the variable represented by the node.

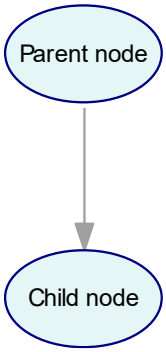


Figure 2: Parent and child node in Bayesian network

In addition to being directed, the network must also be acyclic, meaning there can be no loops or feedback cycles. In other words, it should never be possible to start at one node and follow a series of arrows that eventually lead back to that same node. This acyclic structure is essential because Bayesian networks rely on a clear hierarchical ordering of variables, where each node depends only on its defined parent nodes. If cycles were allowed, there would be no clear hierarchy of parent and child nodes, and the necessary probabilistic calculations would break down. For this reason, Bayesian networks are always directed acyclic graphs.

Parameters of the model

Now that we have explained how the graphical structure of Bayesian networks works, we can describe the quantitative parameters behind this structure. Bayesian networks are based on Bayes' Theorem (Korb & Nicholson, 2011), therefore a brief explanation is provided here using a simple example and a Venn diagram (Figure 3). Imagine two groups of people: those who have a fever (denoted by set A, shown as a yellow circle), and those who are coughing (denoted by set B, a blue circle). However, some individuals, belong to both groups, as they have both a fever and a cough. This is represented by the overlapping green area (A ∩ B) in the diagram.

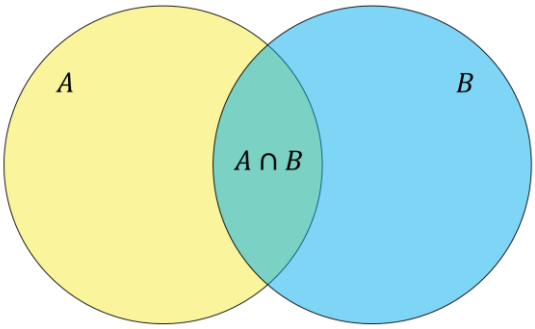


Figure 3: Illustrative Venn diagram

Suppose it is of interest to calculate the probability that someone coughs, given that we know they have a fever. Mathematically, this would be written as $P(Cough \mid Fever)$, or more generally $P(B|A)$. This conditional probability can be calculated by dividing the probability of having both a cough and a fever by the probability of having a fever. The equation would be the following:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{1}$$

Rearranging this equation gives:

$$P(B \cap A) = P(B|A) \cdot P(A) \tag{2}$$

Since the intersection of two sets is the same, regardless of the order ($A \cap B = B \cap A$), the following can be written:

$$P(A \cap B) = P(B \cap A) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B) \tag{3}$$

Substituting equation 3 into equation 1 gives the Bayes' theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{4}$$

Bayes' theorem provides the mathematical foundation for the conditional probability tables (CPTs) in Bayesian networks. CPTs define the probability of a variable given its parent variables in the network. For example, if node A has parent nodes B and C, the CPT specifies $P(A|B, C)$, which is directly in line with the conditional probabilities defined by Bayes' Theorem.

To understand how CPTs are constructed for a certain network structure and using Bayes' theorem, another example is given. A relatively simple Bayesian network is given in Figure 4, where *Smoking*

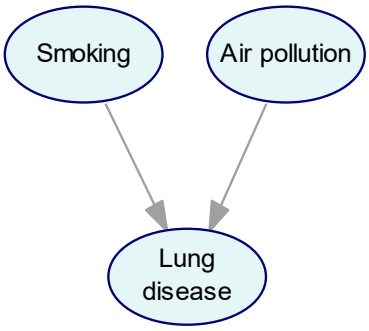


Figure 4: Bayesian network about lung disease and two risk factors

and *Air pollution* are modelled as risk factors for *Lung disease*. These risk factors are represented as parent nodes, with *Lung disease* as the child node. The arrows indicate that smoking and air pollution are both connected to lung disease. For instance, given that we know the person smokes and is exposed to high air pollution, we can compute the probability of lung disease.

For the Bayesian network in Figure 4, it is assumed that a person either smokes or does not smoke and that air pollution can be either high or low. The outcome variable, lung disease, has two possible states: present or absent. Based on these variables, a CPT for this network can be created. Table 2 shows an example CPT using imaginary values.

Table 2: Imaginary CPT for Bayesian network of lung disease and risk factors

Smoking	Air Pollution	P(Lung disease = Yes)	P(Lung disease = No)
Yes	High	0.90	0.10
Yes	Low	0.60	0.40
No	High	0.50	0.50
No	Low	0.05	0.95

The first row of the table can be interpreted as follows: the probability that a person has lung disease, given that they smoke and are exposed to high air pollution, is 0.90 (or 90%). This can be written in mathematical notation as:

$$P(Lung\ disease = yes \mid Smoking = yes,\ Air\ pollution = high\ ) = 0.90$$

Similarly, the probability that the same person does not have lung disease is 0.10 (or 10%). Every row in the table can be interpreted in this way. Note that the probabilities in each row sum to 1, as a person either does or does not have lung disease. These outcomes are mutually exclusive, meaning a person cannot be in multiple states at the same time, and exhaustive, meaning there are no other possible states.

In addition to CPTs, Bayesian networks also work with prior probability tables (PPTs) for the root nodes, meaning nodes without parents. The root nodes in Figure 4 are *Smoking* and *Air pollution*. If, for instance, 10% of the studied population appears to smoke and 90% does not smoke, the PPT as shown in table x can be constructed.

Table 3: Prior probability table for Smoking

Smoking	Prior probability
No	0.90
Yes	0.10

So, Bayesian networks use CPTs to represent the likelihood of each variable state given its parent variables. If a variable has no parents, a PPT is used to describe its initial distribution. Once certain information, or evidence, is entered into the network, such as a known tumour grade, the network updates the probabilities of all connected variables accordingly. This process, which is known as inference, allows the model to estimate the likelihood of outcomes, such as recurrence or survival, based on the given evidence.

3.2.2 Data Sources

Historical patient data, evidence from the literature in combination with input from an STS clinician will be the fundament for developing the Bayesian network. To validate the model, a survey will be conducted amongst STS specialists to seek their views on the way the model is defined. Whereas the historical patient data already exist, the survey data will be collected during this research. A more detailed description of the two data sources is written below.

Historical patient data

The historical patient data comes from an existing retrospective observational dataset from the Leiden University Medical Centre (LUMC), which is the largest tertiary referral centre in the Netherlands. The dataset includes data about all STS patients who underwent surgery at the LUMC between January 1st 2000 and January 1st 2020. The data was routinely collected and stored in a

database. The dataset was already cleaned and used by Kruiswijk et al. (2023). The dataset includes information about patients' follow-up in the first five years after surgery. Apart from demographic patient data, age and sex, the dataset includes tumour characteristics: size, location, grade and histological subtype as well as data on treatment (radiotherapy and surgical margin) and outcomes for patients (survival and recurrence). It also contains information about the number of follow-up visits the patient has received. The information of the dataset that will be used in this research is the following:

- Age (under 40 years, between 40 and 70 years, over 70 years of age)
- Gender (male or female)
- Tumour grade (low-grade: grade I, or high-grade: grade II or III)
- Tumour morphology (Myxofibrosarcoma, Liposarcoma, Leiomyosarcoma, Myxoid liposarcoma, Synovial sarcoma, Dedifferentiated liposarcoma, Hemangiosarcoma, Spindle cell sarcoma or Other (low incidence histologies grouped together))
- Tumour size (5 cm or smaller versus bigger than 5 cm)
- Radiotherapy (neoadjuvant, none or adjuvant)
- Surgical margin (R0 or R1/R2)
- Recurrence (yes/no, and if yes also time until recurrence)
- Survival (yes/no, and if no also time until death)
- Annual follow-up count for each of the five years of the follow-up period

Survey data

The survey is distributed amongst STS specialists in the Netherlands during this research. The survey responses contain perspectives on STS recurrence risk factors, follow-up and the model structure. A more detailed description is given in the section Research Methodology.

Costs of follow-up

For the third sub-question, it is necessary to have (an estimate of) the costs of follow-up visits. A follow-up visit usually consists of physical examination and imaging. Since the guidelines are not strict on what imaging modality should be used for a follow-up visit, we assume that an MRI-scan will be done during the visit. The costs of an out-patient visit and an MRI-scan within the LUMC are denoted in their annual price list (LUMC, 2025).

3.2.3 AI Statement

Artificial Intelligence was used to support the writing of this thesis. Specifically, ChatGPT (free-access version) was used to improve or shorten the writing. Prompts like 'Improve readability on this piece of text', 'Improve the flow of the following text' or 'Shorten the following paragraph' were inputted. The output was always checked for accuracy.

3.3 Research Methodology

This section will elaborate on how each sub-question will be answered, where each of the following sections describes a sub-question, in order to answer the main research question.

3.3.1 Identifying risk factors and follow-up patterns from historical STS patient data

To gain a preliminary understanding of relevant risk factors and clinical decision-making processes in STS follow-up, an exploratory discussion with an STS expert was held. Based on this discussion, a preliminary influence diagram was constructed to map out the potential relationships between key variables. This reflects the mental model of the STS expert we had contacted. However, not all variables in this diagram were also available in the dataset. Therefore, the influence diagram was

reduced to include only the variables available in the dataset. After settling the initial model structure, the conditional dependencies of all variables in the dataset with recurrence were analysed. The aim was to see whether the in- or exclusion of these variables in the model structure was justified.

While the dataset has already been used and cleaned for the research done by Kruiswijk et al. (2023), it had to go through the process of data preparation again, as this research serves another purpose, and thus requires a different structure.

The data analysis is conducted using Python. First of all, we analysed how patient, tumour, and treatment characteristics influence recurrence. Since recurrence and most of the variables are categorical variables, this was done using cross-tabulation to analyse conditional dependencies. To allow for this analysis to be performed, a few assumptions and transformations were made:

- Age and tumour diameter are continuous variables and if they were to be included as variables in the Bayesian network, they should be changed into categorical variables. After consulting an STS expert, they were converted into categorical variables. Age was divided into three groups: under 40, between 40 and 70, and over 70 years. Diameter was split into two categories: 5 cm or smaller, and larger than 5 cm.
- In the provided dataset, surgical margins R1 and R2 were already combined into a single category, limiting our ability to differentiate between these two margin types
- The dataset included 37 different histological subtypes, of which 27 had an incidence of 8 or fewer. These low-incidence types were grouped into the category ‘Other’.
- Similarly, for tumour topography, 12 of the 15 subtypes had an incidence of 5 or fewer and were also grouped under ‘Other’.
- Due to the dataset being relatively small (with only 367 patients after data cleaning), we combined local recurrence, regional recurrence, and distant metastasis into a single variable: recurrence.

The percentage of recurrence was calculated for each category of a variable, e.g. younger versus older patients, but also the proportion of total recurrences occurring within each category. A variable is considered more important when both the recurrence rate is high within a category and a substantial share of the total recurrences falls within that category. In that case, the variable is possibly influences recurrence and is therefore considered for the model. On the other hand, if a category has a high recurrence rate but accounts for only a small number of total recurrences (e.g. older patients have a higher recurrence percentage, but there are only few older overall), the variable may be less influential in practice. This variable is not likely to influence recurrence and therefore not further considered for the model.

Subsequently, in order to see whether the model can be simplified further, the interrelations between the predicting variables (patient, tumour, and treatment characteristics) are analysed, also using cross-tabulation. As many combinations of these variables are possible, this analysis was guided by an STS expert indicating which combinations would be clinically relevant. The goal was to analyse whether some variables are strongly interrelated, i.e. they almost always occur together, as then only one of the variables needs to be included in the model to simplify the model. For instance, if tumour size and tumour grade were to be almost always high together, they are strongly interrelated, and only one of them would be retained in the model. We assess which of the two variables captures more of the variation in recurrence rates. The one that explains more variation across patient groups is retained, while the other can be left out without compromising the model’s predictive value.

A simple and efficient model is important, in a Bayesian network, because including many variables could lead to large conditional probability tables (explained in the next sub-question), which requires sufficient data for each possible combination. The relatively small size of the historical dataset, due to STS being a rare tumour and availability of data from only one centre, could result in conditional probability tables with many empty fields and unreliable estimates. By retaining only those variables that are not strongly related, the model will be robust and interpretable.

Next, if we know the recurrence risk for a patient based on risk factors, we need to be able to translate that risk into a follow-up frequency. To that end, an independent expert orthopaedic surgeon will classify every patient as low-risk or high-risk based on a predefined set of variables: tumour grade, surgical margin, radiotherapy, tumour histology, tumour size and age group. To reduce bias, the surgeon will be blinded to actual patient outcomes.

However, one of the commonly used guidelines, The ESMO-EURACAN guideline, makes a distinction between patients with low- and high-grade tumours, where patients with a high-grade tumour are advised to have more frequent follow-up. This implies that high-grade tumours on their own are already seen as higher risk within this guideline. Therefore, we will do the following analyses both for the risk level as classified by the STS expert, and for tumour grade.

First, the relation between risk level/tumour grade and follow-up frequency will be assessed separately for each of the five years of the follow-up period. Histograms show the distributions of patients with different follow-up frequencies, for both high- and low-risk patients and patients with high- and low-grade tumours. This will give a general indication of the distribution. Subsequently, both high- and low-risk patients, are categorised into three groups of follow-up frequency based on the LUMC guideline: below, on or above the suggested amount by the guideline. This indicates guideline adherence regarding follow-up frequency. The guideline adherence regarding follow-up frequency is also determined for the ESMO-guideline, for both low- and high-grade tumours. The dataset contains the number of follow-up visits patients received for each of the first five years of their follow-up period. The research done by Kruiswijk et al. (2023) used the same dataset and they indicated that LUMC had used the guideline indicated in Table 4.

Table 4: Follow-up frequency guideline LUMC from 2000 to 2020 (Kruiswijk et al., 2023) and ESMO (Gronchi et al., 2021)

Year	Advised number of follow-ups per year		
	LUMC	ESMO	
		Low-grade	High-grade
1	4	2	3-4
2	4	2	3-4
3	4	2	2-4
4	2	2	2
5	2	2	2

Both the ESMO and NCCN guidelines indicate that, at least for high-grade patients, the follow-up frequency after the second year can decrease, suggesting that the risk of recurrence is highest in the first two years. Based on this clinical rationale, follow-up visits were grouped into two time periods: year 1–2 and year 3–5. To verify whether this split was appropriate for the studied population, the timing of recurrences in the dataset was analysed by year.

Guideline adherence was assessed by comparing the number of follow-up visits a patient received within these periods to the number recommended by the LUMC guideline and the ESMO guideline.

However, not all patients would receive follow-up throughout the entire five-year follow-up period, due to either death or a recurrence. For these patients, the suggested number of total follow-up visits, according to the guideline, is calculated based on the time until a patient experienced a recurrence or passed away. For instance, if a patient died halfway through the second year of follow-up, the total suggested number of visits according to the LUMC guideline would be six. Furthermore, a tolerance of 20% (rounded to the closest integer) was added to the guideline, meaning that patients would still be considered to have received on-guideline follow-up if their total number of visits would differ 20% or less compared to the number recommended in the guideline. Cross-tabulation is again used to point out the difference in number of follow-up visits between both high-risk and low-risk following the LUMC guideline, and high-grade and low-grade patients following the ESMO guideline over the five years of the follow-up period.

Lastly, we need to understand the impact of different follow-up strategies. Therefore, a survival analysis is conducted to analyse potential differences in patient survival within five years post-treatment between patients receiving different follow-up frequencies, separately for both high- and low-risk patients, and high- and low-grade tumours. The same categorisation as before will be used: below, on or above the suggested follow-up frequency by the guideline in years 1-2 and years 3-5. The survival analysis will be performed using conditional dependencies. It will, again, consider guideline adherence both during the first two years of follow-up and the third to fifth year of follow-up. This analysis will indicate if patient outcomes are associated with follow-up frequency within risk/grade groups.

3.3.2 Developing a Bayesian decision-support tool to guide risk-based FU-frequency

In this sub-question, the Bayesian model will be developed. As indicated in section 3.2.1 Explanation of the Bayesian Approach, a Bayesian network consists of both the graphical structure and the quantitative parameters. First, the graphical structure needs to be designed, after which the quantitative parameters can be filled in. Lastly, the model will be validated.

The aim of this Bayesian network is to support clinicians in their decision-making process on the follow-up frequency of STS patients. This Bayesian network will reflect how the STS expert we have contacted makes decisions on follow-up frequency. For this specific STS expert, the risk level of patients is the main factor in deciding the follow-up frequency. Since the medical world and the modelling world do not always have the same definition for risk, we will define it here, following Kaplan & Garrick (1981).

Kaplan & Garrick state that risk can be defined by a set of triplets: (s_i, p_i, c_i). It is a set with combinations of a certain scenario (s_i), the likelihood that this scenario happens (p_i) and the consequences (or severity) of the certain scenario if it were to happen (c_i). In the context of scheduling follow-up for STS patients, we will focus on a single scenario: the patient experiencing a recurrence. The risk of this scenario is therefore defined by a combination of the probability that a patient would experience a recurrence and the severity of its consequences. In this case, the ultimate consequence of a recurrence would be death. However, the consequence of a recurrence can be mitigated if it is found timely and treated. Therefore, the key concern regarding consequence is identifying which patients are likely to benefit from follow-up and treatment, and ensuring that the burden of follow-up is justified by a meaningful chance of improving survival. This view on risk, which weighs both the chance of a recurrence and whether treatment is justified, forms the starting point for designing the Bayesian network.

The structure of the Bayesian network is designed in collaboration with a clinical expert in STS. The design includes defining the nodes (e.g., patient and tumour characteristics) and their influence

(edges), resulting in an influence diagram. The exploratory data analysis that was done in sub-question 1, together with findings from the literature, highlighted which variables were important risk factors for recurrence. These factors influence the probability of recurrence. With the risk definition in mind, the probability side of the model was mainly constructed using data, but the consequence side was created in collaboration with an STS expert, to also include individual circumstances affecting the perceived value of timely recurrence detection, and thus the risk of recurrence.

The probability and consequence components were subsequently merged into a final model structure. By combining these perspectives, the network can reflect how both medical risk and patient context inform the utility of follow-up, which in turn guides the determination appropriate follow-up frequency. Two versions of the model were created: the Extensive risk model and Simplified risk model, both based on the findings from sub-question 1 and input from the STS expert. In addition, two other models were created, to allow for comparison. One of these is based on the risk classifications made by the STS expert and the other is based on the ESMO guidelines for follow-up and thus tumour grade. An overview of these four model versions is given in Table 5.

Table 5: Four model types and a description of their corresponding risk classifications

Model type	Description of risk classification
Extensive risk model	Based on risk definition by Kaplan & Garrick (1981), including probability and consequence of recurrence. Probability of recurrence is estimated by multiple important risk factors that were found in sub-question 1.
Simplified risk model	Based on risk definition by Kaplan & Garrick (1981), including probability and consequence of recurrence Probability of recurrence is estimated by the most important risk factor found in sub-question 1.
Clinician model	STS expert classified patients in the historical dataset as low- or high-risk
Grade model	Based on the ESMO guidelines: high-grade tumours are considered high-risk, while low-grade tumours are considered low-risk

After settling the structure of the models, the quantitative parameters of the model can be assigned. The model is created in the GeNIe Modeller software, developed by BayesFusion, LLC (n.d.). Within the model, a distinction is made between root nodes and other nodes. Root nodes, which have no parent nodes, require prior probability tables (PPTs). These are calculated or estimated, when the exact variable is not available in the dataset, using historical patient data processed in Python with the Pandas library.

Two variables are used to estimate the consequence of recurrence in the Extensive and Simplified risk model: expected remaining lifespan and fitness. These variables are not in the dataset, meaning that a few assumptions had to be made for their PPTs.

Firstly, the expected remaining lifespan is based on the average life expectancy in the Netherlands. The average life expectancy in the Netherlands has increased over the past decades, from 73.3 years for people born in 1960 to 81.4 years for those born in 2020 (Centraal Bureau voor de Statistiek, n.d.). Because the exact birth years of the patients in the dataset are unknown, an average life expectancy of 80 years was used as a basis for constructing the prior probability table. To estimate the expected remaining lifespan, we used the same age groupings defined by the STS expert: under 40, between 40

and 70, and over 70 years of age. These group names are converted to more than 40 years, between 10 and 40 years and less than 10 years respectively in terms of expected remaining lifespan.

Secondly, the fitness of the patient is based on the Performance Status Scale created by the Eastern Cooperative Oncology Group (ECOG) (Oken et al., 1982), also adopted by the World Health Organization (WHO). This scale consists of 5 grades of performance (or 6, if death is counted). To reduce complexity in the Bayesian network, the 5 levels of fitness will be reduced to 2, as shown in Table 6. The 2-level division was based on other studies in the cancer domain (Lilenbaum et al., 2008; Kenis et al., 2018; Yang et al., 2025). Since there was no information on fitness in the dataset, the prior probability table was estimated based on age. All patients under 60 were considered fit, while patients over 60 were considered unfit.

Table 6: ECOG/WHO Performance status and corresponding fitness levels used in Bayesian network (Oken et al., 1982)

ECOG Grade	ECOG Performance status	Fitness level in Bayesian network
0	Fully active, able to carry on all pre-disease performance without restriction	Fit
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work	Fit
2	Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours	Unfit
3	Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours	Unfit
4	Completely disabled; cannot carry on any selfcare; totally confined to bed or chair	Unfit

Intermediate and final nodes require conditional probability tables (CPTs) to define the probabilistic relationships between variables. In the Extensive and Simplified risk model, the CPT for risk of recurrence, a central node, is constructed based on the traditional risk matrix. While the traditional matrix contains five levels for probability and four for consequence, we use an adapted version, as shown in Table 7. (Note that risk of recurrence in the Clinician and Grade model is a root node and their PPTs can be directly obtained from the dataset)

Table 7: Adaptation of traditional risk matrix for this study

Probability of recurrence	Consequence of recurrence		
	Low	Moderate	High
Low	Low-risk	Low-risk	High-risk
High	Low-risk	High-risk	High-risk

For most variables, the CPT can be derived from the historical patient data, again using the Pandas library in Python. However, if the variable is not in the dataset, which is the case for the variable consequence of recurrence, we try to obtain an estimate for the probability distribution of the CPT through a survey. Similarly, since risk of recurrence is based on the adaptation of the risk matrix, we validate the CPT for this node through the survey. For these nodes, STS experts were asked to provide input. The questions described a combination of parent node states, and the expert had to assign the

most likely consequence or risk level for the child node. A screenshot of one such question, as used in the survey, is shown in Figure 5.

Question 10

What level of risk of recurrence would you assign to the following patients?

	Low risk	High risk
Patient with a high probability of recurrence and high consequence of recurrence?	○	○
Patient with a high probability of recurrence and moderate consequence of recurrence?	○	○
Patient with a high probability of recurrence and low consequence of recurrence?	○	○
Patient with a low probability of recurrence and high consequence of recurrence?	○	○
Patient with a low probability of recurrence and moderate consequence of recurrence?	○	○
Patient with a low probability of recurrence and low consequence of recurrence?	○	○

Figure 5: Screenshot of survey question to obtain an estimate for the CPT of a node in Bayesian network

The survey was conducted using Qualtrics and spread amongst STS experts in the Netherlands. It also served other purposes than filling in missing CPTs in the Bayesian network: gaining insights on follow-up scheduling in practice and validation of the model structure. The complete survey, including opening statement, can be found in the appendix, A.2 Survey setup, but the main purposes of the questions are mentioned below.

- Q1: Understand whether clinicians consider the risk of recurrence the main factor when deciding on follow-up frequency, and if not, which other factors play a role. This question also validates the model structure.
- Q2, Q3, Q4: Understand how clinicians make follow-up decisions in practice, which guidelines they follow (using what factors) and to what extent clinicians follow or deviate from the recommended frequency.

- Q5, Q6: Explore to what extent clinicians believe that more frequent follow-up or timely detection of recurrence improves survival
- Q7: Validation of model structure
- Q8, Q10: Obtaining estimates for CPTs of consequence of recurrence and risk of recurrence (to validate adaptation of risk matrix)
- Q9: Evaluate whether clinicians would consider using a decision-support tool aligned with the proposed model structure

The first function of the model is to give an estimate for the risk of recurrence for a patient with certain characteristics. Afterwards, we want to show what kind of follow-up (above, on or below guideline) patients with a similar risk level had, and what their survival rates were. Given the inputs for the root nodes, the model outputs the probability that a patient is low- or high-risk (which together always adds up to 100%). However, to define the CPTs for the follow-up and survival nodes, we first need to establish a threshold that separates low- from high-risk patients. This allows us to classify each patient in the dataset according to the model's risk estimation.

To determine an appropriate threshold for high-risk classification, four potential threshold values were tested (30%, 40%, 50% and 60%). For each threshold, all possible input combinations of the model were classified as either high-risk or low-risk, depending on whether the calculated probability of recurrence exceeded the threshold. These classifications were visualised and compared to the classifications made by the expert, using confusion matrices. Evaluation metrics (accuracy and the proportion of over- and under-classifications) were calculated for each threshold. The threshold that scored the best on these metrics was chosen to be used.

After completing the upper part of the model (up to and including the estimation of recurrence risk) and the determination of the threshold, the bottom part of the original model structure was adapted. This serves two purposes:

1. To show observed follow-up adherence and survival outcomes for patients with similar risk profiles
2. To allow clinicians to simulate follow-up decisions

The CPTs for the bottom part of the model were directly obtained from the historical patient data. Some columns in the conditional probability tables (CPTs) for the node *Survival in Year 3 to 5 (Year 1 and 2 recurrence-free)*, however, were empty. To avoid errors in the model, we filled these empty cells with neutral values. The cells of completely empty columns were set to 1 and then normalised, so each possible outcome had an equal probability (for example, 0.5 and 0.5). This ensures the model remains functional without one outcome being favoured over another.

To validate if the model behaves as expected, every possible combination of input variables was entered into both models. For each profile, the following outputs were computed: the probability of recurrence (available only in the Extensive risk model), the consequence of recurrence (defined similarly for both models), and the resulting high-risk percentage for both models. This allows a direct comparison between the classifications made by the Extensive and Simplified risk model.

Furthermore an agreement matrix was created, to indicate to what extent the different models classify patients in the same way. However, due to the lack of a variable that is used to estimate patient consequences, fitness, in the dataset, the Extensive and Simplified risk models are evaluated in two ways. First, by assuming fitness based on age: patients under 60 are classified as fit, and those over 60 as unfit. Second, by excluding patient consequences from the models, so that only recurrence probability determines risk classification. This results in five model versions used for validation:

- Extensive risk model including patient consequences (with fitness assumption)
- Extensive risk model excluding patient consequences
- Simplified risk model including patient consequences (with fitness assumption)
- Simplified risk model excluding patient consequences / Grade model
- Clinician model

3.3.3 Evaluating the impact of risk-based follow-up on patient outcomes and care efficiency

To better understand the implications of follow-up strategies in STS care, this part of the analysis focuses on the relationship between follow-up frequency, patient survival, and associated healthcare costs. For this sub-question, the same five model versions as mentioned at the end of sub-question 2 will be used, since the dataset lacks information on the fitness variable which is used to calculate the patient consequence.

First, each patient in the historical dataset is classified as either low- or high-risk according to each of the five model versions. Within these risk categories, patients are then divided into three groups based on their follow-up frequency: below, on, or above the LUMC guideline. This categorisation follows the same method as in sub-question 1, applying a 20% tolerance and analysing years 1–2 and 3–5 separately.

For each combination of risk level and follow-up frequency, the five-year mortality rate is calculated. This allows for a comparison of outcomes across different levels of follow-up intensity and patient risk. The five-year mortality rate was chosen because recurrences do not typically lead to immediate death. A recurrence identified near the end of year two, for example, may result in death in later years. Therefore, evaluating five-year mortality provides a better view of the potential effects of early follow-up strategies in the long-term.

Second, the economic impact of follow-up strategies is estimated. Firstly, the actual number of follow-up visits patients received, for every combination of risk level and follow-up frequency, is obtained from the historical dataset.

Since guidelines do not provide strict instructions on what imaging modality should be used for follow-up visits (Gronchi et al., 2021; Von Mehren et al., 2022), it is assumed all patients undergo an MRI-scan at the follow-up visit. According to LUMC's price list (LUMC, 2025), the price for an outpatient consultation for an STS tumour is 287.11 euros. Furthermore, the price list indicates that MRI's for the lower extremities, abdomen and pelvis cost between 358.47 and 476.53 euros (costs for upper extremities were not indicated). Since most tumours appeared in the extremities, leaning more towards lower costs, the price of 400 euros per MRI-scan and thus follow-up visit was used. For this calculation, one follow-up visit is estimated to cost 687.11 euros.

The total number of follow-up visits for each risk level and follow-up frequency combination is multiplied by 678.11 euros, to obtain an estimate for the total costs of follow-up for this patient group. Next, the average cost per patient and per survivor for each combination of risk level and follow-up category are calculated. This analysis is again performed separately for the early follow-up years (1 to 2) and the later years (3 to 5), to reflect potential differences in clinical decision-making and outcomes over time. This enabled a comparison of follow-up costs across models and patient groups, highlighting the economic burden of more intensive monitoring relative to survival outcomes.

Together, these analyses aim to provide insight into how follow-up practices are historically linked to patient outcomes and resource use, and to inform future decisions about balancing clinical benefit with cost-effectiveness in long-term STS care.

4. Results

In this chapter, the results are presented per sub-question. So, it starts with the identification of risk factors for recurrence and follow-up patterns from historical data. Then, it describes the development of the Bayesian decision-support tool. Lastly, it evaluates the impact of follow-up frequencies on mortality and its economic effects.

4.1: Identification of risk factors and follow-up patterns from historical STS patient data

Before we started with the actual data analysis, an exploratory discussion with an STS expert already allowed us to create a preliminary model structure. This model forms the basis for the analysis in this sub-question, where it is continuously explained why certain factors are retained or excluded from the model. Afterwards, the follow-up patterns found in the historical data are shown.

4.1.1 Mental model of STS expert

To gain a basic understanding of the relations between variables in the post-treatment context, an exploratory discussion was held with an STS expert. Based on this discussion a preliminary influence diagram was made, as shown in the appendix (A.3 Preliminary Influence Diagram Post-Treatment Context). However, since the dataset for this studies contains limited variables, a new version of the influence diagram regarding probability of recurrence was created. This diagram is shown in Figure 6.

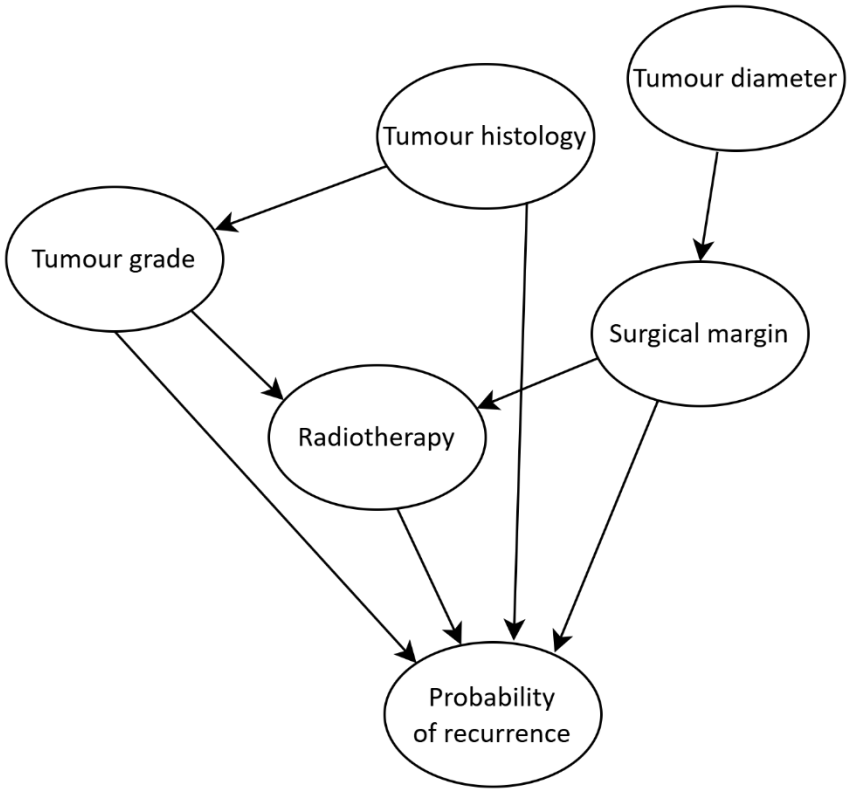


Figure 6: Mental model of the STS expert regarding probability of recurrence in the form of an influence diagram

The STS expert indicated that there could be an important association between tumour grade and tumour histology. The same holds for diameter and surgical margin, where a larger tumour is usually harder to resect fully, often leading to an intralesional resection, and thus remaining tumour cells after surgery. The surgeon stated that the influence of tumour diameter almost always goes through

surgical margin, which is reflected in figure 6 by no arrow going directly from tumour diameter to the probability of recurrence.

Furthermore, the expert indicated that the main indicator for radiotherapy is tumour grade, where it is often administered to high-grade patients. In some cases, the (expected) surgical margin could also play a role in the decision for radiotherapy. For example, if the surgeon expects to perform an intralesional resection (meaning that tumour cells remain at the cut surface after surgery) with a low-grade tumour, the surgeon might opt for neoadjuvant radiotherapy. Or, if the surgery unexpectedly turned out to be an intralesional resection, the surgeon might administer adjuvant radiotherapy to the patient.

4.1.2 Conditional Dependencies with Recurrence

With the initial model structure in place, the next step is to analyse the available variables in the dataset in order to assess whether the data supports this structure. The dataset includes the following variables: gender, age, diameter, radiotherapy, surgical margin, tumour grade, histology, and topography. Although not all of these were included in the initial model, each one is analysed to determine whether excluding it is justified. For every variable, a table was created showing the number and percentage of recurrences for each category within that variable.

Gender

As shown in Table 8, the total incidence of STS appears to be nearly equal between men and women. The shares of gender within the recurrence group are close to 50% and thus also almost similar to the total population. Furthermore, the recurrence rates among the two genders are also nearly identical. This indicates that gender likely has no influence on the probability of recurrence.

Table 8: Recurrence cases and gender distribution, including share within recurrence groups and total population and recurrence percentage

Gender	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
Male	132 (50.0%)	53 (51.5%)	185 (50.4%)	28.6%
Female	132 (50.0%)	50 (48.5%)	182 (49.6%)	27.5%
Total	264	103	367	28.1%

Age

Table 9 indicates that most STS patients in general are in the age group between 40 and 70 years (55.6%), and this is also the group where most recurrences happen (49.5%). While recurrence rates appear similar amongst under 40 and between 40 and 70 year olds, the recurrence rate is higher amongst patients over 70. This variability indicates that age could possibly influence the probability of recurrence.

Table 9: Recurrence cases and age group distribution, including share within recurrence groups and total population and recurrence percentage

Age group (years)	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
Under 40	45 (17.0%)	16 (15.5%)	61 (16.7%)	26.2%
Between 40 and 70	153 (58.0%)	51 (49.5%)	204 (55.6%)	25.0%
Over 70	66 (25.0%)	36 (35.0%)	102 (27.8%)	35.3%
Total	264	103	367	28.1%

Diameter

By looking at the recurrence percentages in Table 10, the recurrence rate is higher in tumours with a diameter bigger than 5 cm. Furthermore, the share of tumours bigger than 5 cm among patients with a recurrence is almost 80%, while this share is lower in the total patient group (67.6%). Most recurrences seem to happen in the group of larger tumours. These findings imply that diameter is likely to influence the probability of recurrence.

Table 10: Recurrence cases and diameter group distribution, including share within recurrence groups and total population and recurrence percentage

Diameter group	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
5 cm or smaller	98 (37.1%)	21 (20.4%)	119 (32.4%)	17.6%
Bigger than 5 cm	166 (62.9%)	82 (79.6%)	248 (67.6%)	33.1%
Total	264	103	367	28.1%

Radiotherapy

What immediately becomes clear from Table 11 is that the majority of STS patients (73.8%) did not receive radiotherapy. It also illustrates that recurrence rates are higher among patients who *did* receive radiotherapy. Furthermore, Table 11 shows that although most patients who experienced a recurrence had not received radiotherapy (64.1%), adjuvant and neoadjuvant radiotherapy make up a relatively larger share of the recurrence group compared to their share in the total patient population.

Radiotherapy is well-known for its ability to shrink tumours (if given before surgery) or eliminate remaining tumour cells (if given after surgery). One might therefore expect it to reduce the probability of recurrence. However, the data shows the opposite. This is probably because radiotherapy is given more frequently to patients that are considered high-risk by clinicians (e.g. due to tumour grade or histology). So, while the higher recurrence rate among patients with radiotherapy may seem contradictory at first, it likely reflects treatment selection based on clinical risk factors, rather than the effect of radiotherapy itself.

Table 11: Recurrence cases and radiotherapy distribution, including share within recurrence groups and total population and recurrence percentage

Radiotherapy	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
None	205 (77.7%)	66 (64.1%)	271 (73.8%)	24.4%
Neoadjuvant	20 (7.6%)	12 (11.7%)	32 (8.7%)	37.5%
Adjuvant	39 (14.8%)	25 (24.3%)	64 (17.4%)	39.1%
Total	264	103	367	28.1%

Surgical margin

Most surgeries (66.8%) result in an R0 margin, see Table 12. The recurrence percentages point out that an R0 margin has a lower recurrence rate than R1 or R2 margins. Table 12 also indicates that about 61% of all the recurrences were with an R0 margin, leaving 39% for the R1 or R2 margins. Although R0 margins were twice as common as R1/R2, the number of recurrences was only 1.5 times higher. Altogether, this suggests that surgical margin could possibly influence the recurrence rate.

Table 12: Recurrence cases and surgical margin distribution, including share within recurrence groups and total population and recurrence percentage

Surgical margin	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
R0	182 (68.9%)	63 (61.2%)	245 (66.8%)	25.7%
R1/R2	82 (31.1%)	40 (38.8%)	122 (33.2%)	32.8%
Total	264	103	367	28.1%

Tumour grade

Table 13 highlights that high-grade tumours appear to have a higher recurrence rate than low-grade tumours, differing about 26%. It also indicates that incidence of high-grade STS tumours in general is almost twice as high as the incidence of low-grade tumours. The share of high-grade tumours within patients with recurrence, however, is more than six times higher than the share of low-grade tumours. Among the recurrences, about 86% were high-grade tumours, leaving 14% low-grade tumours. Based on these findings, tumour grade is likely to influence the recurrence rate.

Table 13 Recurrence cases and tumour grade distribution, including share within recurrence groups and total population and recurrence percentage

Tumour grade	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
Low grade	113 (42.8%)	14 (13.6%)	127 (34.6%)	11.0%
High grade	151 (57.2%)	89 (86.4%)	240 (65.4%)	37.1%
Total	264	103	367	28.1%

Tumour histology

Table 14 indicates that myxofibrosarcoma is the histology with the highest incidence in total, followed by liposarcoma and leiomyosarcoma. The recurrence rates vary widely amongst the histologies, with Spindle cell sarcoma having the highest (70%) and Liposarcoma having the lowest (4%). Among patients with a recurrence, most of them had a myxofibrosarcoma, leiomyosarcoma or a sarcoma within the 'other' category. Liposarcoma only represents 1.9% of the recurrences. Table 14 also indicates there is a notable difference between the share among recurrence and the total patient group for some histologies, such as liposarcoma and myxoid liposarcoma. These findings indicate that tumour histology is likely to influence recurrence rates. It is, however, important to note that some histologies in Table 14 have rather low incidence, reducing the reliability of the observed recurrence rates for these groups.

Table 14: Recurrence cases and tumour histology distribution (low incidence histologies grouped into 'Other'), including share within recurrence groups and total population and recurrence percentage

Tumour histology	No Recurrence cases (n, %)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
Myxofibrosarcoma	59 (22.3%)	19 (18.4%)	78 (21.3%)	24.4%
Liposarcoma	51 (19.3%)	2 (1.9%)	53 (14.4%)	3.8%
Leiomyosarcoma	31 (11.7%)	16 (15.5%)	47 (12.8%)	34.0%
Myxoid liposarcoma	33 (12.5%)	3 (2.9%)	36 (9.8%)	8.3%
Synovial sarcoma	14 (5.3%)	10 (9.7%)	24 (6.5%)	41.7%

Dedifferentiated liposarcoma	8 (3.0%)	4 (3.9%)	12 (3.3%)	33.3%
Hemangiosarcoma	5 (1.9%)	5 (4.9%)	10 (2.7%)	50.0%
Spindle cell sarcoma	3 (1.1%)	7 (6.8%)	10 (2.7%)	70.0%
Other	60 (22.7%)	37 (35.9%)	97 (26.4%)	38.1%
Total	264	103	367	28.1%

Tumour topography

Table 15 points out that the lower extremities are the topography where most STS tumours appear. The recurrence percentages, however, show that there is not much difference in recurrence rates amongst the different topographies. The distribution of topographies within the group of patients who experienced a recurrence, is rather similar to their overall distribution within STS patients. These findings indicate that tumour topography is not likely to influence the probability of recurrence.

Table 15: Recurrence cases and tumour topography distribution, including share within recurrence groups and total population and recurrence percentage

Tumour topography	No Recurrence cases (n)	Recurrence cases (n, %)	Total (n, %)	Recurrence percentage
Lower extremities	193 (73.1%)	68 (66.0%)	261 (71.1%)	26.1%
Upper extremities	35 (13.3%)	19 (18.4%)	54 (14.7%)	35.2%
Pelvis	20 (7.6%)	9 (8.7%)	29 (7.9%)	31.0%
Other	16 (6.1%)	7 (6.8%)	23 (6.3%)	30.4%
Total	264 (100%)	103 (100%)	367 (100%)	28.1%

Initial Risk Factor Selection

After having analysed the conditional dependencies of the available variables with recurrence, the variables can be divided into two groups: possibly influential for recurrence and not influential. This division is as follows:

- **Possibly influential**: diameter, radiotherapy, tumour grade, tumour histology, age, surgical margin
- **Not influential**: gender, tumour topography

The STS expert confirmed the variables from this conditional dependency analysis that had a possible influence on recurrence. The discussion with the STS expert did, however, point out that the tumour- and treatment-related factors might, in fact, be more important than the variable age regarding the influence on probability of recurrence. Age could be more important in terms of the consequences of a recurrence, i.e. how severe one would be impacted by a recurrence and its potential treatment (which is probably different for a 95-year old, compared to a 20-year old). Therefore, age is not be considered a risk factor for the probability of recurrence, but will be considered for the consequences of a recurrence in this model, which is described in sub-question 2. Gender and tumour topography will also not be considered risk factors for recurrence. The current model structure is therefore still similar to the diagram in Figure 6.

4.1.3 Conditional Dependencies between Variables

To see if we can simplify the model, we look at how risk factors relate to each other, because if some are strongly linked or always appear together, we might be able to leave some out without losing important information. Therefore, the conditional dependencies between risk factors are analysed.

Based on the STS expert's input, as indicated in section 4.1.1 Mental model of STS expert, and the potential risk factors, specific pairs of variables were selected for further analysis. The following combinations were analysed:

- Histology and grade
- Diameter and surgical margin
- Grade and radiotherapy

Histology and Grade

Table 16 shows that most histology types mainly consist of high-grade tumours. For example, 100% of the synovial sarcomas are classified as high-grade. Other histology types, such as leiomyosarcoma, myxoid liposarcoma, dedifferentiated liposarcoma, hemangiosarcoma, spindle cell sarcoma and the 'other' category show high-grade proportions between 75% and 95%. Myxofibrosarcoma appears to be the only histology type with an almost even distribution between high- and low-grade, with 53.8% high-grade. Liposarcoma, however, is the only histology type where nearly all tumours are low grade (96.2%), also representing the largest part of low-grade tumours in Table 16.

Table 16: High- and low-grade cases and tumour histology distribution, including share within grade groups and total population and high-grade percentage

Tumour histology	Low-grade cases (n, %)	High-grade cases (n, %)	Total (n, %)	High-grade percentage
Myxofibrosarcoma	36 (28.3%)	42 (17.5%)	78 (21.3%)	53.8%
Liposarcoma	51 (40.2%)	2 (0.8%)	53 (14.4%)	3.8%
Leiomyosarcoma	11 (8.7%)	36 (15.0%)	47 (12.8%)	76.6%
Myxoid liposarcoma	9 (7.1%)	27 (11.2%)	36 (9.8%)	75.0%
Synovial sarcoma	0 (0%)	24 (10.0%)	24 (6.5%)	100.0%
Dedifferentiated liposarcoma	3 (2.4%)	9 (3.8%)	12 (3.3%)	75.0%
Hemangiosarcoma	1 (0.8%)	9 (3.8%)	10 (2.7%)	90.0%
Spindle cell sarcoma	1 (0.8%)	9 (3.8%)	10 (2.7%)	90.0%
Other	15 (11.8%)	82 (34.2%)	97 (26.4%)	84.5%
Total	127	240	367	65.4%

The results of Table 16 together with a discussion with an STS expert indicate that there is an association between tumour histology and grade. Therefore, a closer look was taken at combinations of histology and grade with their corresponding recurrence rates to determine whether the overall recurrence rate by grade would accurately reflect the recurrence rate for histology-grade combinations, as shown in Table 17.

Table 17: Recurrence counts and percentages by grade and histology

Grade	Histology	Recurrence counts		Recurrence percentage	
		No	Yes	By grade and histology	By grade
Low grade	Myxofibrosarcoma	32	4	11.1%	11.00%
	Liposarcoma	49	2	3.9%	
	Leiomyosarcoma	9	2	18.2%	
	Myxoid liposarcoma	9	0	0%	
	Synovial sarcoma	0	0		
	Dedifferentiated liposarcoma	3	0	0%	
	Hemangiosarcoma	0	1	100%	
	Spindle cell sarcoma	1	0	0%	
	Other	10	5	33.3%	
High grade	Myxofibrosarcoma	27	15	35.7%	37.10%
	Liposarcoma	2	0	0%	
	Leiomyosarcoma	22	14	38.9%	
	Myxoid liposarcoma	24	3	11.1%	
	Synovial sarcoma	14	10	41.7%	
	Dedifferentiated liposarcoma	5	4	44.4%	
	Hemangiosarcoma	5	4	44.4%	
	Spindle cell sarcoma	2	7	77.8%	
	Other	50	32	39.0%	

Table 17 can be read as follows: among patients with low-grade myxofibrosarcoma, there were 32 who had no recurrence, while 4 did, resulting in an 11.1% recurrence percentage. This is close to the overall recurrence rate for low-grade sarcomas, which is 11.0%.

Within the low-grade group, recurrence rates are generally low and closely align with the overall low-grade recurrence rate of 11.0%. For high-grade tumours, most histological subtypes show recurrence rates between 35% and 45%, aligning closely with the overall high-grade recurrence rate of 37.1%. This indicates that tumour grade alone already captures most of the variation in recurrence risk, and adding histology contributes little additional predictive value.

Some deviations do exist, such as myxoid liposarcoma and spindle cell sarcoma in the high-grade group, but these subtypes have relatively low case numbers and therefore limited impact on the model's accuracy. In total, 11 of the 18 histology-grade combinations have 15 cases or fewer, which limits their reliability on the distribution.

To take a closer look at tumour grade, Table 13 indicates that 14 low-grade and 89 high-grade tumours had a recurrence. When considering all patients in the dataset (367), 24.3% both had a recurrence and a high-grade tumour, while only 3.8% had a recurrence and a low-grade tumour (71.9% did not have recurrence). This also illustrates the strong association between tumour grade and recurrence.

Given the large number of histological subtypes with low incidence and the finding that high-grade tumours are clearly associated with a higher probability of recurrence, histology does not need to be considered to accurately predict recurrences. An updated version of the 'mental model' of the STS expert is shown in Figure 7. Tumour histology and its outgoing arrows are depicted in grey to show that they do not need to be taken into account as tumour grade alone will be sufficient to predict recurrence.

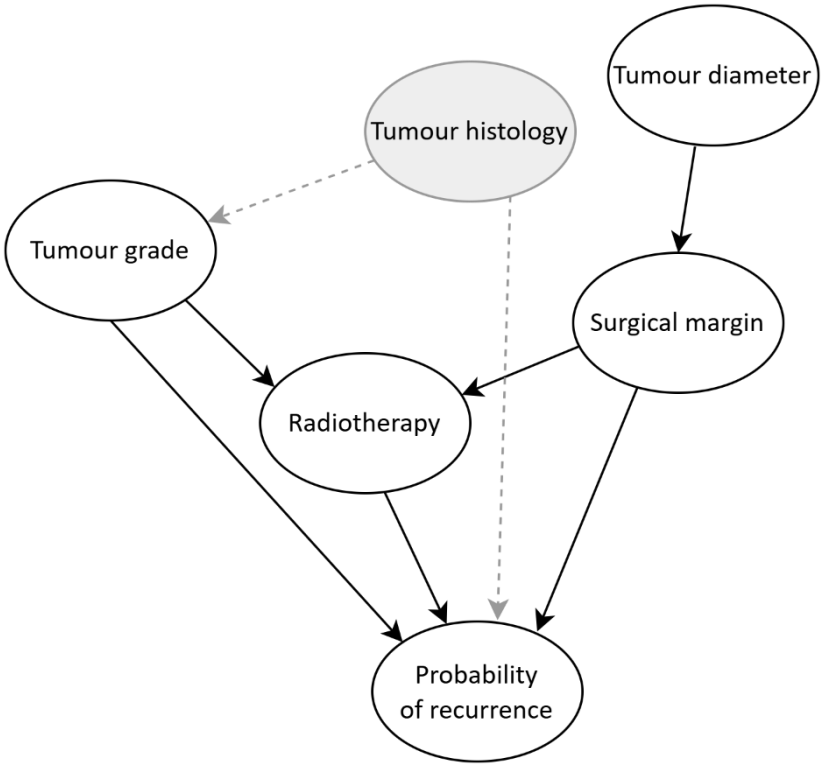


Figure 7: Mental model of the STS expert regarding probability of recurrence in the form of an influence diagram, without tumour histology

Diameter and Surgical margin

A consultation with an STS expert pointed out that there could be an important relation between diameter and surgical margin. The reason is as follows: if a tumour is larger, the chance that the tumour is deeper or lies closer to a critical structure (e.g. a nerve) increases. Deep tumours are usually harder to resect with an R0 margin, and the same goes for tumours close to a critical structure, as the structure should not be damaged. Therefore, there could be an association between diameter and surgical margin.

It is important to note that the surgical margin seemed less strongly related to recurrence than tumour diameter in the previous conditional dependency analysis with recurrence. This could, however, be due to the fact that the dataset combines R1 and R2 into one category. While they both indicate positive margins, expert consultations point out that they do differ in severity, with R2 usually being associated with worse outcomes. This merging may have caused the influence of surgical margin to be less pronounced, which is important to keep in mind.

Table 18 indicates that the share of tumours of different diameter groups do not differ much. Furthermore, the percentage of R1/R2 margins for the diameter groups also are similar, as shown in Table 18. Based on these results, there does not appear to be a strong or consistent association between tumour diameter and margin status.

Table 18: Margin R0 and R1/R2 cases and diameter distribution, including share within margin groups and total population

Diameter group	Margin R0	Margin R1/R2	Total	Margin R1/R2 percentage
5 cm or smaller	76 (31.0%)	43 (35.2%)	119 (32.4%)	36.1%
Bigger than 5 cm	169 (69.0%)	79 (64.8%)	248 (67.6%)	31.9%
Total	245 (100%)	122 (100%)	367 (100%)	33.2%

Table 19 is a cross-tabulation of recurrence for surgical margin – diameter combinations, showing that across both diameter groups, the recurrence rate is higher for R1/R2 margins than for R0. Both R1/R2 margins and tumours with a diameter larger than 5 cm have an increased probability of recurrence. However, the variation in recurrence across the combinations does not differ substantially from the individual effects of margin status and diameter alone. This suggests that while both factors influence recurrence risk, their interaction does not add much additional explanatory power.

Table 19: Recurrence cases and percentages by surgical margin and diameter

		Recurrence counts		Recurrence percentage	
Surgical margin	Diameter group	No	Yes	By surgical margin and diameter	By surgical margin
R0	5 cm or smaller	65	11	14.5%	25.7%
	Bigger than 5 cm	117	52	30.8%	
R1/R2	5 cm or smaller	33	10	23.3%	32.8%
	Bigger than 5 cm	49	30	38.0%	

While tumour diameter and surgical margin do not seem clearly or strongly related from the data, and they both seem to have a similar relation with the probability of recurrence, a consultation with an STS expert led to a final decision. The STS expert indicated that surgical margin is more directly related to recurrence risk, and that the influence of the diameter might always go through the surgical margin. Therefore, surgical margin is retained in the model. An updated version of the mental model is given in Figure 8.

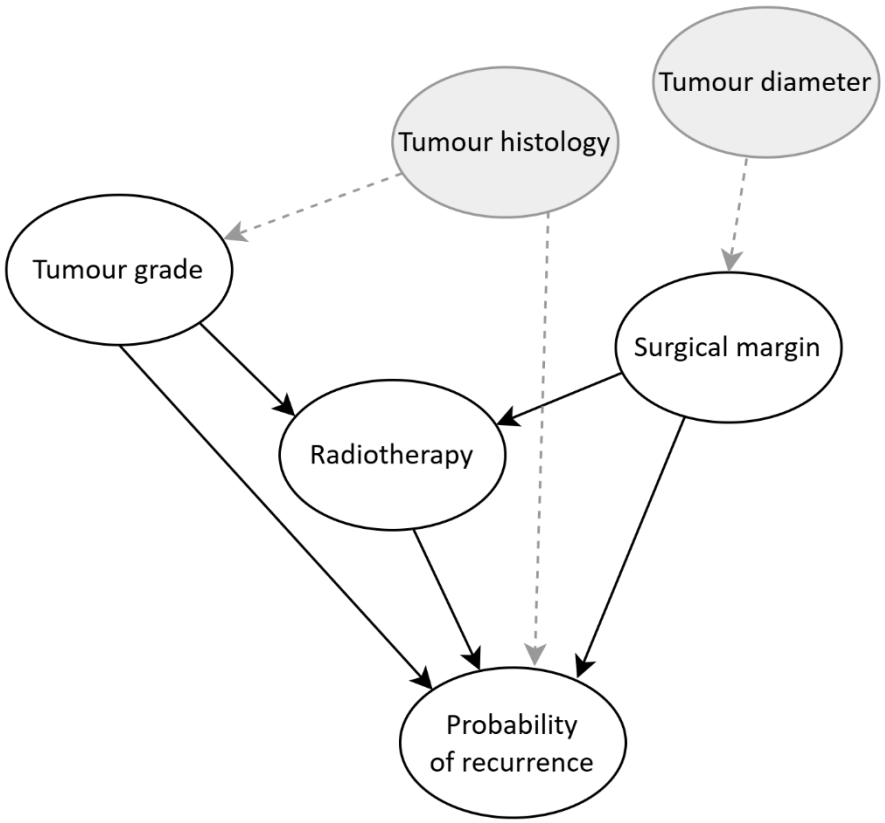


Figure 8: Mental model of the STS expert regarding probability of recurrence in the form of an influence diagram, without tumour histology and diameter

Grade and Radiotherapy

Since the results from the conditional dependency analysis between radiotherapy and recurrence pointed out that patients who received radiotherapy had a higher recurrence rate, these data were further analysed, particularly the relationship with grade. This was based on a consultation with an STS expert pointing out that tumour grade is the main indicator for radiotherapy, where patients with a high-grade tumour are usually given radiotherapy.

The ESMO guidelines indicate that radiotherapy can be considered based on histology and anatomical location (Gronchi et al., 2021). However, it also states that radiotherapy is often given to high-grade tumours, and it can be considered for both high- and low-grade tumours depending on risk factors for local recurrence: surgical margin, tumour size, tumour site, histological type and the potential consequences of a local recurrence. This implies that whether a patient received radiotherapy could be seen as a reflection of the estimated risk level of this patient.

Table 20 shows that the share of patients without radiotherapy is larger in low-grade cases (90.6%) than in high-grade cases (65.0%). Furthermore, it indicates that the percentage of high-grade tumours within the group of patients who received radiotherapy is 87.5%. The findings from both tables are in line with the fact that mostly high-grade patients receive radiotherapy, and the ESMO guidelines.

This would also explain why patients with radiotherapy had higher recurrence rates (as shown in Table 11): radiotherapy is usually given to high-grade tumours, which tend to have a higher

recurrence rate. If radiotherapy had not been given to part of these high-grade tumours, the overall recurrence rate for high-grade tumours would probably have been even higher.

Table 20: Low- and high-grade cases and radiotherapy distribution, including share within grade groups and total population

Radiotherapy	Low-grade cases (n, %)	High-grade cases (n, %)	Total (n, %)	High-grade percentage
None	115 (90.6%)	156 (65.0%)	271 (73.8%)	57.6%
Neoadjuvant	4 (3.1%)	28 (11.7%)	32 (8.7%)	87.5%
Adjuvant	8 (6.3%%)	56 (23.3%)	64 (17.4%)	87.5%
Total	127 (100%)	240 (100%)	367 (100%)	65.4%

Table 21 shows that the recurrence percentages within each grade are in a similar range: from 0% to 12.5% for low-grade tumours and from 34% to 43% for high-grade tumours. These ranges closely match the overall recurrence rates per grade. This suggests that grade by itself already captures most of the variation in recurrence rate, making radiotherapy less decisive for estimating the probability of recurrence.

Table 21: Recurrence cases and percentages by tumour grade and radiotherapy

		Recurrence counts		Recurrence percentage	
Grade	Radiotherapy	No	Yes	By grade and radiotherapy	By grade
Low-grade	Adjuvant	7	1	12.5%	11.0%
	Neoadjuvant	4	0	0.0%	
	None	102	13	11.3%	
High-grade	Adjuvant	32	24	42.9%	37.1%
	Neoadjuvant	16	12	42.9%	
	None	103	53	34.0%	

Tumour grade was found to be the most important indicator for radiotherapy, but since the STS expert indicated that surgical margin would sometimes be used in the decision-making process of radiotherapy, this variable was also analysed. The analysis can be found in the appendix (A.4 Conditional Dependency between Surgical Margin and Radiotherapy) and showed that tumour grade explained most of the variation. An updated version of the STS expert's mental model is shown in Figure 9.

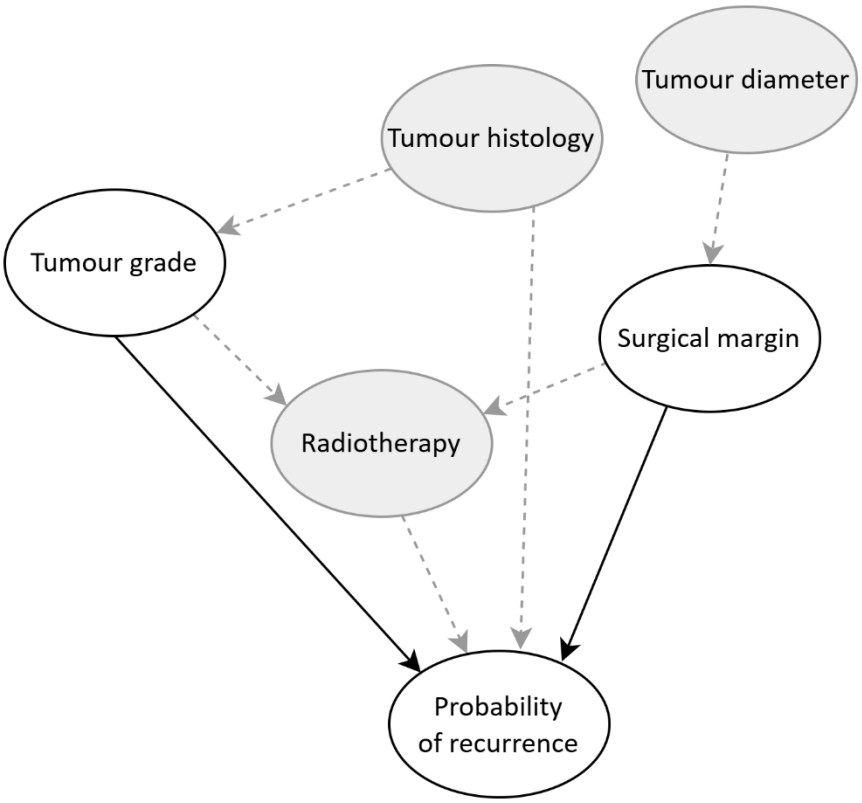


Figure 9: Mental model of the STS expert regarding probability of recurrence in the form of an influence diagram, without tumour histology, diameter and radiotherapy

4.1.4 Selection of risk factors

The following variables were identified as risk factors: tumour histology, tumour grade, tumour diameter, surgical margin and radiotherapy. Further analysis into certain combinations of these variables, based on expert input, highlighted potential dependencies between variables.

The analysis of histology and grade revealed an association between the two. However, due to the low incidence of many histology-grade combinations and the consistently strong predictive value of tumour grade on its own, grade was retained as a risk factor for in the model, while histology was excluded.

The combination of diameter and surgical margin showed no clear or consistent relationship. Since surgical margin has a more direct association with recurrence, according to an STS expert, it was prioritised over diameter.

Radiotherapy was strongly associated with tumour grade, aligning with clinical guidelines. The higher recurrence rates among patients who received radiotherapy likely reflect treatment selection based on the clinician's estimated patient risk, mainly based on tumour grade. A further analysis showed that tumour grade captures most of the variation in recurrence rate, also when taking into account radiotherapy. Therefore, radiotherapy is left out as a risk factor in the model.

This leaves tumour grade and surgical margin as the final important risk factors that are needed to predict recurrences. Since we saw that tumour grade seems to have a stronger association with recurrence than surgical margin, a cross-tabulation for these three factors is shown in Table 22.

Table 22: Recurrence cases and percentages by tumour grade and surgical margin

Grade	Surgical margin	Recurrence counts		Recurrence percentage	
		No	Yes	By grade and surgical margin	By grade
Low-grade	R0	71	6	7.8%	11.0%
	R1/R2	42	8	16.0%	
High-grade	R0	111	57	33.9%	37.1%
	R1/R2	40	32	44.4%	

The surgical margin does seem to make a difference within the grade-groups. The primary factor, however, is tumour grade. Since the STS clinician emphasized the importance of surgical margin (and the grouping of R1 and R2 could affect the strength of the found association), it will be retained in the model. The final model with risk factors for recurrence can therefore be simplified, to include only tumour grade and surgical margin as risk factors. This simplified model is shown in Figure 10.

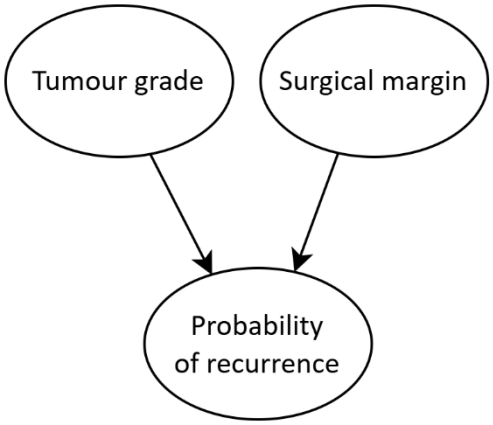


Figure 10: Influence diagram with tumour grade and surgical margin as risk factors for recurrence

4.1.5 Risk Classification by STS Expert

All the patients in the historical dataset from the LUMC were classified as either high- or low-risk patients by an STS expert, without knowing the outcomes for these patients. By analysing these classifications, a pattern in the expert's decision-making could be identified. Based on this pattern, a decision tree was constructed to reflect how the STS expert had classified the patients. This decision tree is shown in Figure 11.

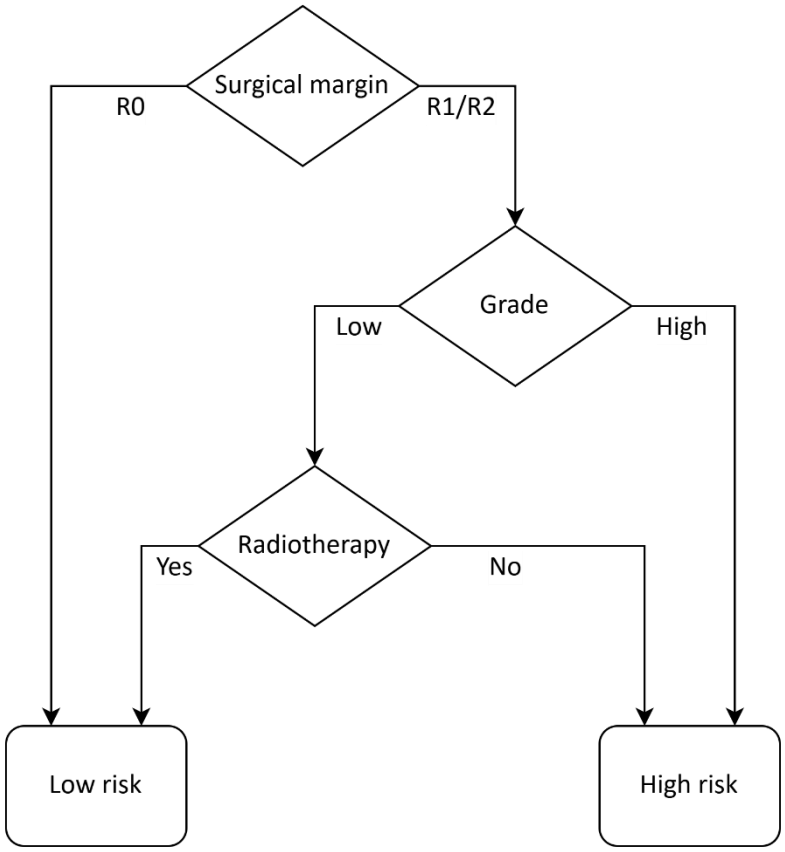


Figure 11: Decision tree of STS expert's risk classification

The decision tree shows that for the STS expert the surgical margin is an important indicator for risk. A patient with an R0-margin is immediately classified as low-risk. For a patient with an R1 or R2 margin, grade is also taken into account. If the tumour is high-grade, the patient is immediately classified as high-risk. If the tumour is low-grade, radiotherapy becomes the deciding factor: those who received radiotherapy (either neoadjuvant or adjuvant) are classified as low-risk, while those who did not receive radiotherapy are classified as high-risk. The decision tree shows that the two factors we have selected as the most important risk factors for the model, are also the most important for this STS expert in his risk classification.

4.1.6 Risk level/Tumour grade versus Recurrence and Survival

As a next step, cross-tabulations were made using the risk level assigned by the STS expert against both recurrence and survival. Table 13 shows the cross-tabulations for grade and recurrence, which will be referenced here given the importance of grade in previous analyses.

Table 23 indicates that the share of high-risk patients within the recurrence group is higher than in the no-recurrence group. Furthermore, the recurrence percentage is 8.2% higher for high-risk patients (33.6%) than for low-risk patients (25.4%). These findings indicate that high risk, as classified by the STS expert, is associated with a higher chance of recurrence. However, Table 13 showed a much larger difference of 26.1% between the recurrence percentages in high (37.1%) versus low tumour grade (11.0%) .

Table 23: Recurrence cases and risk distribution, including share within recurrence groups and total population and recurrence percentage

Risk	No recurrence cases (n, %)	Recurrence (n, %)	Total (n, %)	Recurrence percentage
Low	185 (70.1%)	63 (61.2%)	248 (67.6%)	25.4%
High	79 (29.9%)	40 (38.8%)	119 (32.4%)	33.6%
Total	264	103	367	28.1%

Therefore, tumour grade explains more variation in recurrence. Conducting a similar analysis for survival, Table 24 shows that the share of high- and low-risk patients is nearly identical within the different survival groups. Moreover, the percentages of patients who died within the two risk groups are also nearly similar, differing only 1%. These findings indicate that the risk level assigned by the STS expert does not seem to be associated with survival of patients. The difference in death percentages between tumour grades, on the other hand, is 6.3% of the low-grade patients and 32.9% of the high-grade patients (Table 25) so a difference of 26.6%. Furthermore, only 9.2% of the patients who died had a low-grade tumour, leaving the other 90.8% of the patients who died with a high-grade tumour. This indicates that survival is associated with tumour grade.

Table 24: Survival and risk distribution, including share within survival groups and total population and death percentage

Risk	Survived patients (n, %)	Died patients (n, %)	Total (n, %)	Death percentage
Low	190 (67.9%)	58 (66.7%)	248 (67.6%)	23.4%
High	90 (32.1%)	29 (33.3%)	119 (32.4%)	24.4%
Total	280	87	367	23.7%

Table 25: Survival and grade distribution, including share within survival groups and total population and death percentage

Grade	Survived patients (n, %)	Died patients (n, %)	Total (n, %)	Death percentage
Low	119 (42.5%)	8 (9.2%)	127 (34.6%)	6.3%
High	161 (57.5%)	79 (90.8%)	240 (65.4%)	32.9%
Total	280	87	367	23.7%

To further evaluate the relation between risk/grade and survival, two cross-tabulations were made that also included recurrence. These cross-tabulations shown in Table 26, indicate that there are big differences in death rates between the recurrence and no recurrence groups for both high- and low-risk patients. While the overall death rates for the high- and low-risk are nearly identical, there are clear differences for patients with and without recurrence: within the high-risk group about 50% and for the low-risk group about 56%. Therefore, where risk level does not seem associated with survival, recurrence does seem strongly associated with survival. Similarly, as shown in Table 27, the death percentage is considerably higher amongst patients with a recurrence in both high- and low-grade groups. However, grade also seems to be an important factor for survival, since the death percentage amongst patients without a recurrence is almost five times higher for high-grade cases than low-grade cases. Among patients with a recurrence, the death percentage is almost twice as high for high-grade cases compared to low-grade cases.

Table 26: Death cases and percentages by risk and recurrence

	Survival counts	Death percentage

Risk	Recurrence	Survived	Died	By risk and recurrence	By risk
Low	No	168	17	9.2%	
	Yes	22	41	65.1%	23.4%
High	No	73	6	7.6%	
	Yes	17	23	57.5%	24.4%

Table 27: Death cases and percentages by grade and recurrence

		Survival counts		Death percentage	
Grade	Recurrence	Survived	Died	By grade and recurrence	By grade
Low	No	110	3	2.7%	
	Yes	9	5	35.7%	6.3%
High	No	131	20	13.2%	
	Yes	30	59	66.3%	32.9%

4.1.7 Recurrence over the Years

Both the ESMO and NCCN guidelines indicate that follow-up frequency, at least for high-grade patients, can be lowered after the second year, implying that the highest risk of recurrence lies in the first two years of follow-up.

To confirm the rationale for dividing the follow-up period into the first two years and years three to five, the timing of recurrences was analysed annually. The results, as shown in Figure 13 and Figure 12, showed that most recurrences occur within the first two years, especially among patients with high-grade tumours, where recurrence percentages were notably higher in these years. In contrast, patients with low-grade tumours showed a more consistent, low recurrence risk throughout the entire follow-up period. When looking at the risk classification made by the STS expert, the difference in recurrence timing was less pronounced but still supported a focus on closer monitoring during the first two years. All these findings indicate that the first two years are the most crucial regarding recurrence and align with the clinical guidelines. Therefore, it is justified look separately at the first two years for the follow-up adherence.

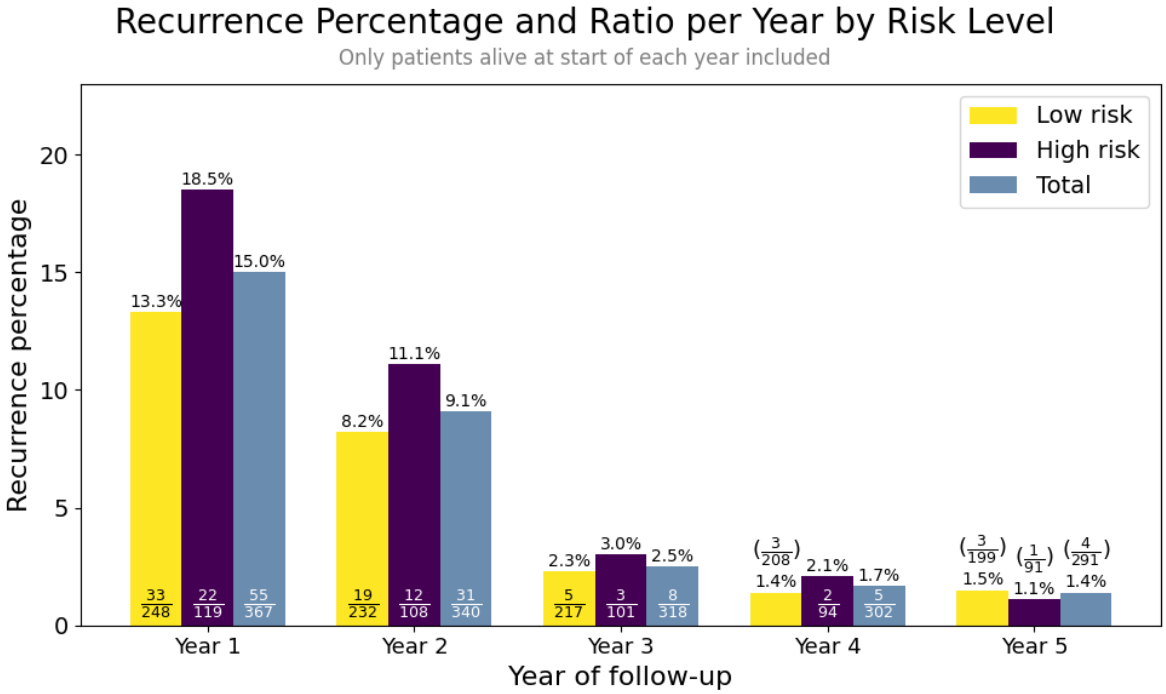


Figure 13: Recurrence percentage and ratio per year by risk level and total, calculated among patients alive at start of each year

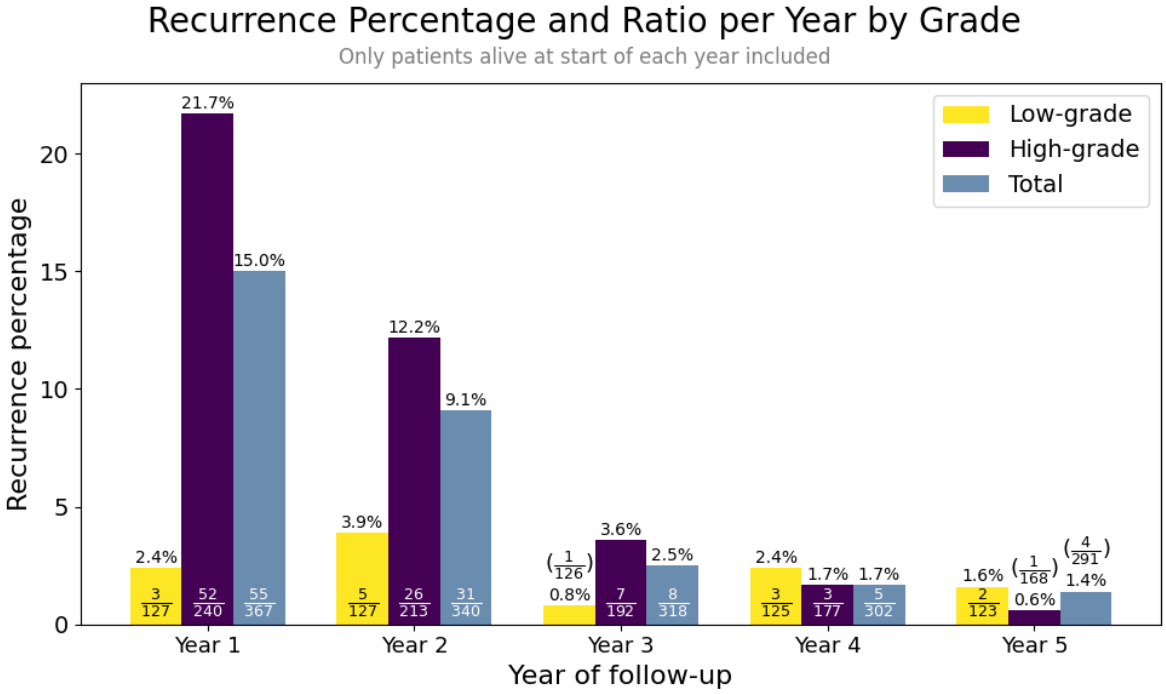


Figure 12: Recurrence percentage and ratio per year by grade and total, calculated among patients alive at start of each year

4.1.8 Follow-up

To get a general idea about the follow-up frequency patients received in each of the five years of recurrence, a histogram was made, as shown in Figure 14 and Figure 15. Firstly, these figures show that the total number of patients decreases over the years. This is due to the fact that patients die during the follow-up period or experience a recurrence, after which their follow-up frequency is no longer registered in the dataset. Secondly, the figures show that for both risk groups, as well as both grade groups, most patients received a follow-up frequency between 1 and 4 visits in the first two years. In the last three years, most patients get a follow-up frequency of 1 or 2 visits.

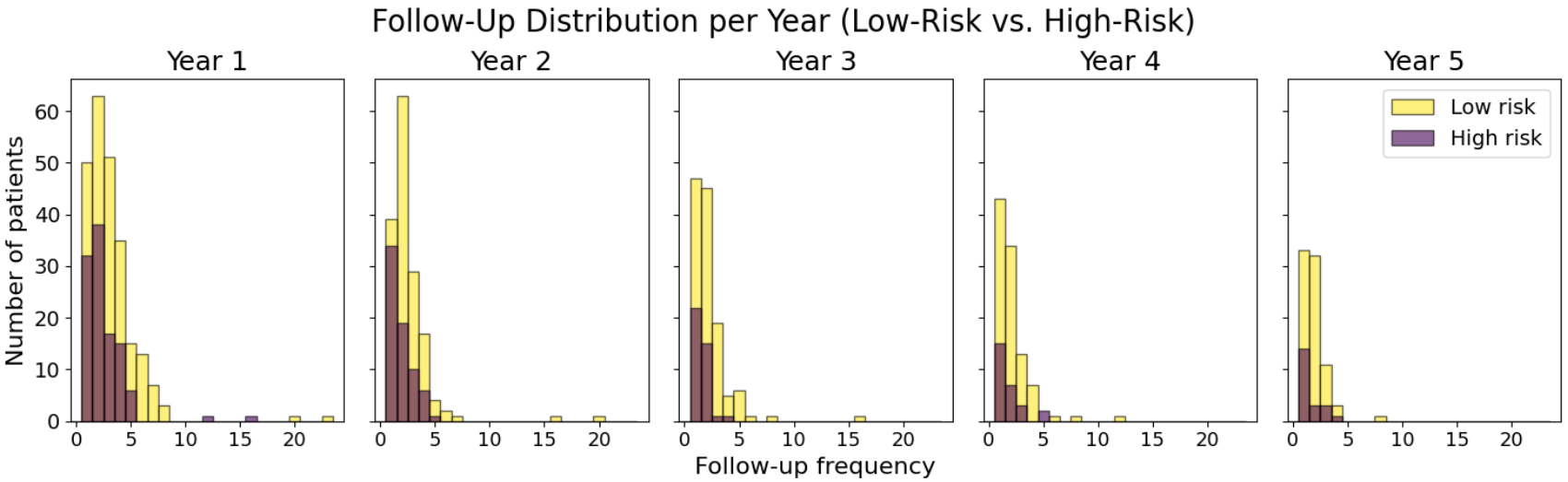


Figure 14: Histogram of follow-up frequency distribution per year, showing low-risk and high-risk (Note that it is not a stacked, but an overlapping bar chart)

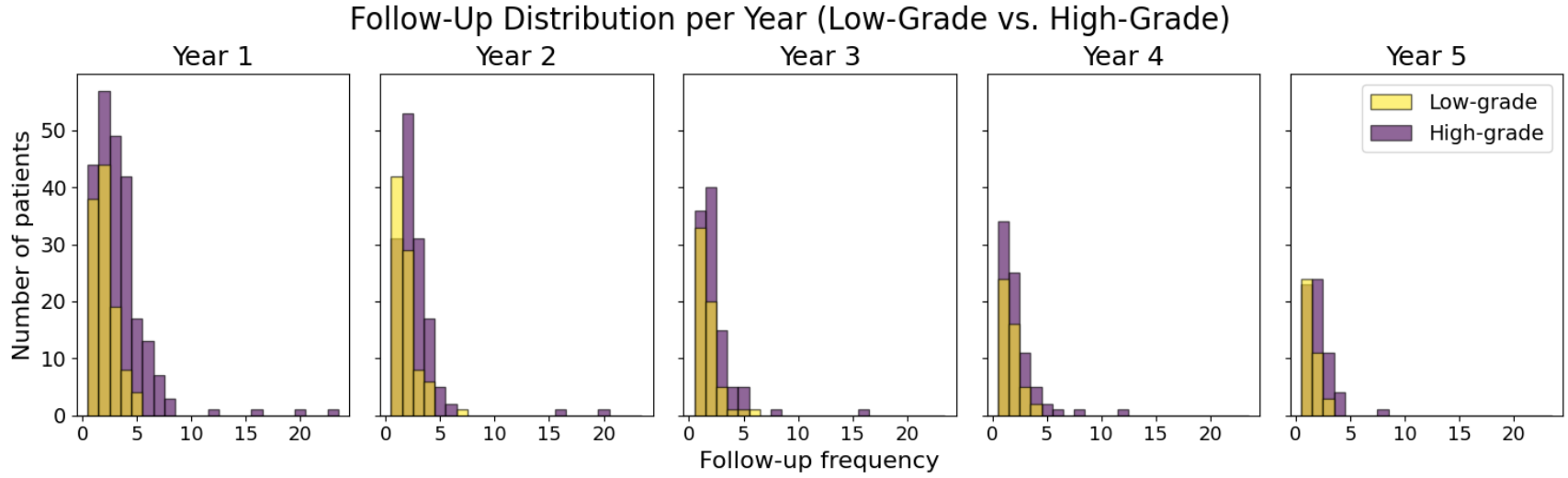


Figure 15: Histogram of follow-up frequency distribution per year, showing low-grade and high-grade (Note that it is not a stacked, but an overlapping bar chart)

Table 28 shows the distribution of guideline adherence regarding follow-up in the first two years per risk level. It shows that most patients in the total population (59.1%) receive a follow-up frequency that is below the suggestion of the guideline. A larger share of patients, 65.5%, receives a follow-up frequency lower than suggested by the guideline in the high-risk group. Furthermore, both the shares of patients within the high-risk group that receive on or above the suggested frequency are smaller (30.3% and 4.2% respectively) than the share within the total population (and thus also the low-risk group). These findings indicate that high-risk patients do not necessarily receive a follow-up frequency that is above the suggested frequency, they actually more often get a lower frequency than suggested by the guideline. The opposite is true for the low-risk patients, as they have a relatively large share of patients that receive a follow-up frequency that is higher than suggested the guidelines in the first two years.

Table 29 shows adherence to the ESMO guideline in the first two years, distinguishing by tumour grade. Most patients (56.7%) receive follow-up in line with the guideline, especially high-grade patients, since they have 63.8% on-guideline while it is only 43.3% for low-grade. A smaller share of high-grade patients receives either less (27.9%) or more (8.3%) follow-up than suggested compared to low-grade patients (41.7% and 15.0%, respectively). This suggests that high-grade patients are more likely to receive follow-up frequencies as suggested by the guideline in the first two years after treatment, while low-grade patients more often receive too little or too much follow-up.

Table 28: Year 1 and 2 LUMC follow-up guideline adherence and risk distribution, including share within risk groups and total population and high-risk percentage

Year 1 and 2 LUMC Follow-up Guideline Adherence	Low-risk patients (n, %)	High-risk patients (n, %)	Total (n, %)
Below	139 (56.0%)	78 (65.5%)	217 (59.1%)
On	93 (37.5%)	36 (30.3%)	129 (35.1%)
Above	16 (6.5%)	5 (4.2%)	21 (5.7%)
Total	248 (100%)	119 (100%)	367 (100%)

Table 29: Year 1 and 2 ESMO follow-up guideline adherence and grade distribution, including share within grade groups and total population and high-grade percentage

Year 1 and 2 ESMO Follow-up Guideline Adherence	Low-grade patients (n, %)	High-grade patients (n, %)	Total (n, %)
Below	53 (41.7%)	67 (27.9%)	120 (32.7%)
On	55 (43.3%)	153 (63.8%)	208 (56.7%)
Above	19 (15.0%)	20 (8.3%)	39 (10.6%)
Total	127 (100%)	240 (100%)	367 (100%)

For the third to fifth year of the follow-up period (Table 30), we see that the share of patients with a below-guideline frequency is notably higher in the high-risk group compared to the low-risk group and total population, while there are relatively few patients in the above-guideline group for high-risk. Table 31 shows that the share of below-guideline follow-up frequency within the high-grade cases has grown, compared to the first two years, as opposed to the on-guideline share, which shrunk. What becomes clear from both Table 30 and Table 31 is that the minority of patients receive on-guideline follow-up frequency between the third and fifth year after treatment.

Table 30: Year 3 to 5 LUMC follow-up guideline adherence and risk distribution, including share within risk groups and total population and high-risk percentage

Year 3 to 5 LUMC Follow-up Guideline Adherence	Low-risk patients (n, %)	High-risk patients (n, %)	Total (n, %)	High-risk percentage
Below	124 (71.7%)	58 (85.3%)	182 (75.5%)	31.9%
On	41 (23.7%)	9 (13.2%)	50 (20.7%)	18.0%
Above	8 (4.6%)	1 (1.5%)	9 (3.7%)	11.1%
Total	173	68	241	28.2%

Table 31: Year 3 to 5 ESMO follow-up guideline adherence and grade distribution, including share within grade groups and total population and high-grade percentage

Year 3 to 5 ESMO Follow-up Guideline Adherence	Low-grade patients (n, %)	High-grade patients (n, %)	Total (n, %)	High-grade percentage
Below	82 (73.9%)	60 (46.2%)	142 (58.9%)	42.3%
On	25 (22.5%)	62 (47.7%)	87 (36.1%)	71.3%
Above	4 (3.6%)	8 (6.2%)	12 (5.0%)	66.7%
Total	111	130	241	53.9%

While the death rates do not really differ between the two risk groups, there seem to be differences between the different guideline adherences in the first two years (Table 32). For both low- and high-risk groups, the death percentages are the lowest in the below-guideline group and the highest in the above-guideline group. Similarly, in Table 33, for both the low- and high-grade groups, the highest death percentage is found for the patients who received a follow-up frequency above the guideline. Of note, the highest death percentage in the low-grade group (26.3%) is still lower than the lowest death percentage in the high-grade group (26.9%)

Table 32: Death cases and percentages by risk and year 1 and 2 LUMC follow-up guideline adherence

Risk	Year 1 and 2 LUMC Follow-up Guideline Adherence	Survival counts		Death percentage	
		Survived	Died	By risk and recurrence	By risk
Low	Below	116	23	16.5%	
	On	67	26	28.0%	
	Above	7	9	56.3%	23.4%
High	Below	70	8	10.3%	
	On	19	17	47.2%	
	Above	1	4	80.0%	24.4%

Table 33: Death cases and percentages by grade and year 1 and 2 ESMO follow-up guideline adherence

Grade	Year 1 and 2 ESMO Follow-up Guideline Adherence	Survival counts		Death percentage	
		Survived	Died	By risk and recurrence	By risk
Low	Below	51	2	3.8%	
	On	54	1	1.8%	
	Above	14	5	26.3%	6.3%
High	Below	49	18	26.9%	
	On	104	49	32.0%	
	Above	8	12	60.0%	32.9%

Apart from looking into the first two years, we will also look into the latter three years of the five-year follow-up period. Table 34 shows that no high-risk patients died in the latter three years. For the low-risk group, patients that had below-guideline and on-guideline follow-up had about a 5% death rate in the latter three years, very similar to the overall low-risk death percentage in these years. The above-guideline group had a death percentage of 25%, but the total number of cases is only 8.

Table 35 shows that the overall death percentages in the latter three years of follow-up is 3.6% for low-grade and 5.4% for high-grade. The percentages for patients with below- and on-guideline follow-up are rather similar to the overall percentages in both the low- and high-grade group. The patients with above-guideline follow-up, however, have higher death percentages (both 25.0%). The low incidence (4 for low-grade, 8 for high-grade) does make these percentages unreliable.

Table 34: Death cases and percentages by risk and year 3 to 5 LUMC follow-up guideline adherence

Risk	Year 3 to 5 LUMC Follow-up Guideline Adherence	Survival counts		Death percentage	
		Survived	Died	By risk and recurrence	By risk
Low	Below	117	7	5.6%	
	On	39	2	4.9%	
	Above	6	2	25.0%	6.4%
High	Below	58	0	0.0%	
	On	9	0	0.0%	
	Above	1	0	0.0%	0.0%

Table 35: Death cases and percentages by grade and year 3 to 5 ESMO follow-up guideline adherence

Grade	Year 3 to 5ESMO Follow-up Guideline Adherence	Survival counts		Death percentage	
		Survived	Died	By risk and recurrence	By risk
Low	Below	80	2	2.4%	
	On	24	1	4.0%	
	Above	3	1	25.0%	3.6%
High	Below	56	4	6.7%	
	On	60	2	3.2%	
	Above	7	1	12.5%	5.4%

4.1.9 Summary SQ1

This sub-question started off with an exploratory discussion with an STS expert that led to an initial model structure including tumour grade, tumour histology, tumour diameter, radiotherapy and surgical margin. A conditional dependency analysis of the available variables in the dataset confirmed the inclusion of these variables, and exclusion of gender and tumour topography. While age seemed to have some conditional dependency with recurrence, the other tumour- and treatment-related factors were deemed more important for the probability of recurrence following an expert consultation. Age was considered to be more important regarding consequence of recurrence, which is treated in sub-question 2.

However, the STS expert mentioned the selected tumour- and treatment-related factors could be interrelated. Therefore, subgroup analyses with relevant combinations of different variables and recurrence were done.

Tumour histology and tumour grade appeared interrelated. Due to the large number of histologies, the relatively small dataset and therefore small incidence of many histologies, we chose to retain tumour grade in the model, because it has a strong and consistent association with recurrence.

The STS expert pointed out that diameter and surgical margin could be related, since larger tumours are often harder to fully remove. Although the data did not show a strong link between the two, this could be due to the grouping of R1 and R2 margins, which could make the differences less pronounced. However, because the STS expert indicated that diameter likely works through surgical margin, it was decided to retain surgical margin in the model.

Patients who received radiotherapy showed higher recurrence rates, but this is likely because radiotherapy was mostly given to high-grade tumours. So, the recurrence says more about the risk level of these patients than about the effect of radiotherapy itself. Tumour grade showed a clear and consistent relation with recurrence and is therefore kept in the model.

Tumour grade and surgical margin were selected as the most important risk factors. A subgroup analysis of these two variables with recurrence, shows that grade is the most strongly associated factor. Therefore, it was chosen to create two risk-models, one including tumour grade and surgical margin, both connected to probability of recurrence and the other just including tumour grade.

An STS expert classified all patients in the dataset as either high- or low-risk. While we have used a risk definition including both probability and consequence, the STS expert perceived risk as only the probability of recurrence, which is common in the medical world. The decision-pattern showed that

surgical margin was the most important factor in deciding, followed by tumour grade and then radiotherapy. Tumour grade, however, appeared to explain more variation in recurrence percentage than the risk level as classified by the STS expert. It appears that giving more information to the clinician, changes his decision. In addition, tumour grade was clearly associated with death, while the clinician's risk levels showed nearly identical death percentages.

The majority of recurrences were found to appear within the first two years of the follow-up. Especially high-grade tumours have the highest chance of a recurrence in the first two years, while low-grade tumours have a consistently lower chance over the entire five year period.

High-risk patients, as classified by the STS expert, were mostly given below-guideline follow-up frequency. Most of the low-risk patients were also given a below-guideline follow-up. The majority of patients with a high-grade tumour did get on-guideline follow-up, while the majority of low-grade patients received below-guideline follow-up.

Moreover, death rates were the highest among patients who received above-guideline follow-up, for both risk- and grade-groups. These higher death rates probably reflect risk level as perceived by the clinician at the time of treatment. The higher the perceived risk, the higher the number of follow-up visits for the patient. However, this suggests that a higher follow-up frequency does not necessarily improve survival chances.

4.2 Development of a Bayesian decision-support tool to guide risk-based FU-frequency

This section presents the results for sub-question 2. It starts with describing the model structure, which is followed by the quantitative parameters behind the model structure. Afterwards, an implementation of the model is shown and it is validated.

4.2.1 Model Structure

The structure of the Bayesian network, represented in the influence diagram shown in Figure 16, focuses on the period following a patient's initial treatment. This means that key tumour and treatment characteristics are already known to the clinicians at the time of decision-making about the frequency of follow-up visits.

As shown in Figure 16, risk of recurrence is the central node in this model, which is influenced by two main components. Firstly, there is the clinical aspect, where clinical factors influence the probability of recurrence. Secondly, there is the patient-centred aspect, where factors related to the patient's overall condition influence the consequence of a recurrence, indicating how important a timely recurrence detection would be for the patient e.g. to receive further treatment. This aspect is also referred to as 'patient consequences' in the rest of the studies. Together, these two components determine the expected benefit of early recurrence detection, which is captured in the node 'risk of recurrence'.

The clinical aspect of the model captures the probability of a recurrence. Note that the complete mental model of the STS expert, from sub-question 1, is depicted in Figure 16. After the analyses from sub-question 1, tumour diameter, tumour histology and radiotherapy were not retained in the model, and are therefore depicted in grey. Based on the findings from Sub-question 1, two variables were identified as potential risk factors for recurrence: tumour grade and surgical margin. These variables both have directed edges pointing towards the node 'probability of recurrence'. In other words, given that we know the tumour grade (high/low) and surgical margin (R0, R1/R2), we can compute the probability of recurrence. Since tumour diameter, tumour histology and radiotherapy

are excluded from the model, the final and simplified version of the influence diagram is given in Figure 17.

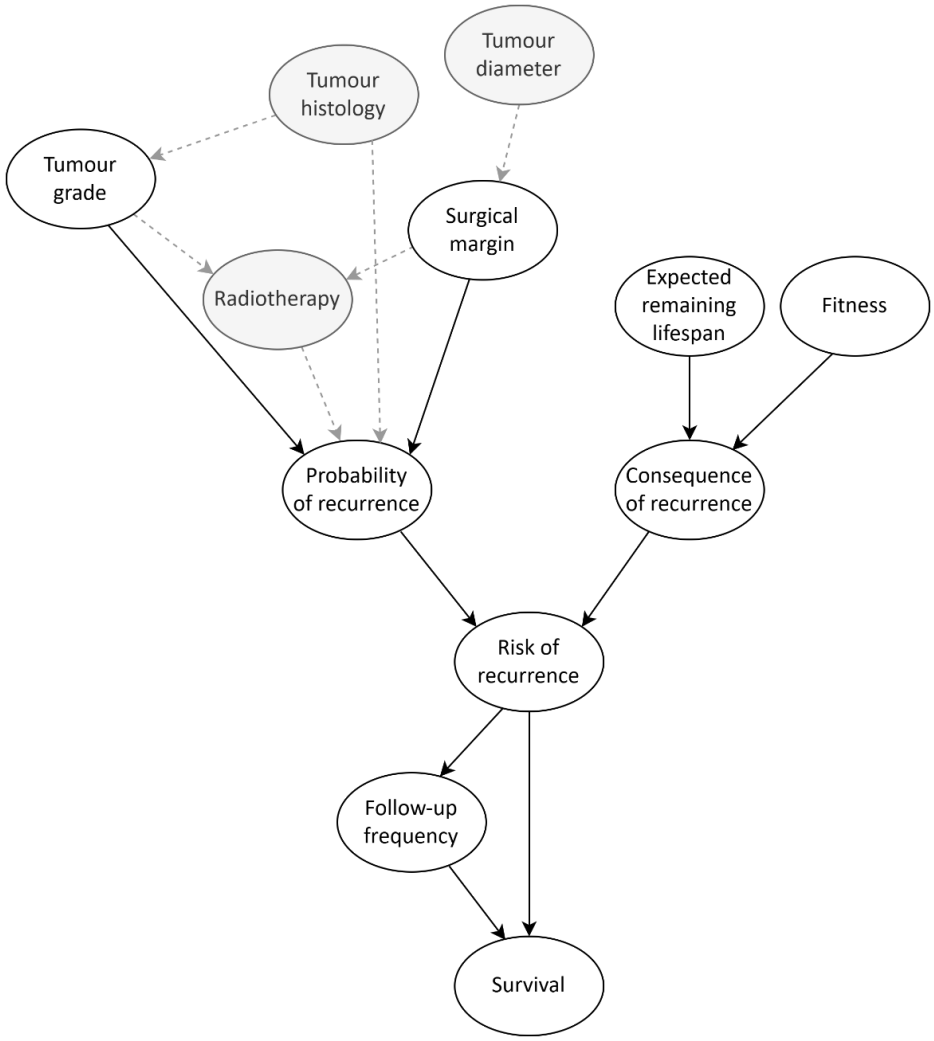


Figure 16: Influence diagram and structure for Bayesian network for risk of recurrence, follow-up frequency and survival (Factors influencing the probability of recurrence that were not retained after sub-question 1 are shown in grey)

In addition to the clinical aspect, the model also includes a patient-centred aspect, which addresses the potential consequences of a timely recurrence detection from the patient's perspective. The central node here is the 'consequence of recurrence'. This node is influenced by two factors: expected remaining lifespan (more than 40 years/between 10 and 40 years/less than 10 years) and fitness of the patient (fit/unfit). This aspect of the model captures that not all patients experience an equal burden from a recurrence, it depends on how much harm a recurrence and especially its potential treatment would cause. It reflects finding a balance between potential recurrence harm and potential treatment harm.

Sub-question 1 pointed out that age could possibly influence probability of recurrence, but the tumour- and treatment-related variables were deemed more important and age would be more important regarding the consequence of recurrences. It is reflected in the influence diagram by the node expected remaining lifespan. This node reflects the idea that a recurrence and its treatment will likely have a different impact on a younger patient (e.g. a 15-year-old with many potential life years remaining) than on an older patient (e.g. a 90-year-old nearing the end of his or her life). For younger patients, early recurrence detection could be important to improve further treatment options, to

preserve as many future life years. An older patient, however, might not benefit as much from early recurrence detection. He/She might opt out of further treatment, if the burden of the treatment outweighs the benefits for his/her remaining time. In other words, the consequence of a recurrence, and therefore also the benefit of detecting it early, tends to be lower for older patients than for younger patients.

Fitness of the patient is considered, because individuals in a better physical condition are, in general, assumed to better tolerate further treatment and recovery. Therefore, the consequence of recurrence may be higher for them, as they have more to lose in terms of quality of life and functional independence and would benefit more from a potential recurrence treatment.

The two aspects of the model (clinical and patient-centred) come together in the node 'risk of recurrence' (high/low). This node represents the overall expected benefit of early recurrence detection, and therefore also the expected benefit of conducting surveillance and treating a recurrence, if it is caught early. It is influenced by the probability of recurrence (meaning how likely a recurrence is to happen) and the consequence of a recurrence (meaning how severe a recurrence and potential treatment would be for the patient). It is important to note that the risk node is about an evaluation of benefit relevant for a decision: if a recurrence is both likely and the patient would benefit greatly from recurrence treatment, then the risk level of the patient high, and early recurrence detection would be valuable. If either the probability is low or the patient would not benefit greatly from recurrence treatment, the risk would be lower, and early recurrence detection would be less valuable.

Another node in the model is 'follow-up frequency' (below guideline/on guideline/above guideline), which represents how often a patient should be monitored for recurrence after initial treatment. This decision is influenced by the parent node 'risk of recurrence'. This connection is about a key principle: The higher the expected benefit of detecting a recurrence early (the higher the risk), the more reason there is to monitor the patient closely. In other words, if a recurrence is likely, and if treatment would be acceptable and beneficial to the patient, then it is worth investing in more frequent follow-up to catch the recurrence in time. On the other hand, if the risk is low, because recurrence is unlikely, or the patient would not benefit much from treatment, then less intensive follow-up may be justified. This node reflects the clinical reasoning behind personalised follow-up strategies. Based on what is known about a patient's situation, it is determined how valuable it is to detect recurrence timely and therefore how often it should be checked.

The final node in the model is 'survival', which is influenced by both follow-up frequency and risk of recurrence. The link from risk of recurrence reflects that recurrences, and especially metastasis, reduce the chances of survival. The arrow coming from follow-up frequency reflects the assumption that more frequent follow-up increases the chances of detecting a recurrence early which would in turn improve the likelihood of successful treatment and thus survival. While follow-up itself does not directly affect the disease, it enables timely intervention, making it an important factor in overall patient outcomes.

As indicated in sub-question 1, tumour grade seemed stronger associated with recurrence than surgical margin. Therefore, an extra model version is made, where the three nodes in the clinical aspect (tumour grade, surgical margin and probability of recurrence) are replaced by one node: tumour grade is directly linked to risk of recurrence. This version of the model, as shown in Figure 18, is referred to as the *Simplified risk model*, while the version in Figure 17 is referred to as the *Extensive risk model*. Two other model structures were also made to allow for comparison: the *Grade model* (Figure 19) and the *Clinician model*, where the risk classification made by the STS expert is used for risk of recurrence (Figure 20).

Extensive risk model

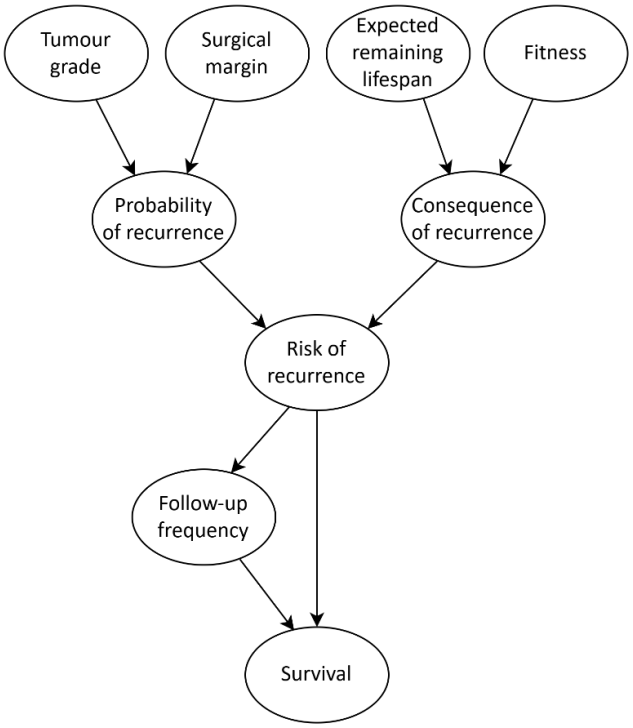


Figure 18: Influence diagram and structure for Bayesian network for risk of recurrence, follow-up frequency and survival (Extensive risk model)

Simplified risk model

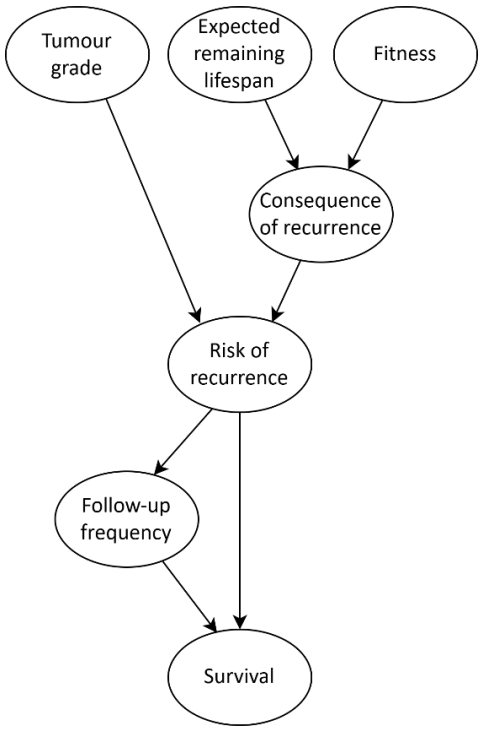


Figure 17: Influence diagram and structure for Bayesian network for risk of recurrence, follow-up frequency and survival (Simplified risk model)

Grade model

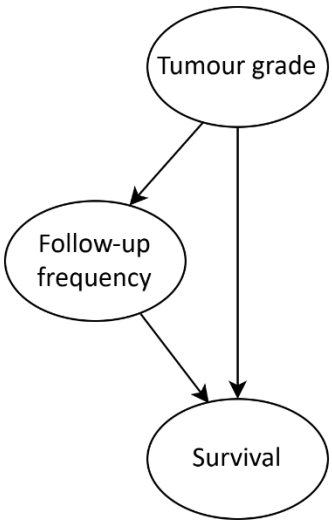


Figure 20: Influence diagram and structure for Bayesian network with tumour grade, follow-up frequency and survival (Tumour grade model)

Clinician model

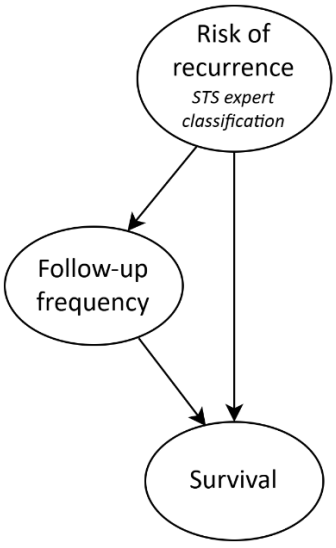


Figure 19: Influence diagram and structure for Bayesian network with risk of recurrence (classified by STS expert), follow-up frequency and survival (Clinician model)

4.2.2 Model Structure Validation

Parts of the model structure were validated using the survey. The complete survey responses can be found in the appendix (A.5 Survey Results). While the proposed model versions are based on the mental model of one STS expert, we wanted to check if other STS experts would agree with this form of decision-making. Important to note is that there are a small number of STS experts in the Netherlands, reflecting the fact that STS is a rare disease. There were seven respondents for the first six questions, while there were only six respondents for the last four questions.

Firstly, 4 out of 7 respondents answered that risk of recurrence is the main factor in deciding the follow-up frequency for your patients. The other 3 mentioned that 'it depends on a combination of tumour-related, treatment-related, and patient-related factors', or referred to both 'risk of recurrence and risk of metastasis'. One of them stated they followed a 'stratified protocol for follow-up, partly based on diagnosis (in some aspect this is recurrence)'. So, although only 4 respondents selected risk of recurrence directly, the others effectively described the same concept as defined in this study or indicated that it is a part of their decision-making process.

Secondly, 5 out of 6 respondents agreed that the consequences of finding a recurrence depend on both the patient's remaining lifespan and fitness level. The other respondent did not agree, but still emphasised that fitness level is an important factor: if the patient is unfit and the probability of recurrence is low, the risk is lower due to limited treatment benefit. This suggests that all respondents, recognise patient characteristics as a key factor of the consequence of recurrence.

Furthermore, all respondents indicated that they would consider using a decision-support tool where the probability of recurrence is estimated based on tumour grade and surgical margin, as proposed in the Extensive risk model.

While no direct questions about the bottom part of the model (below risk of recurrence) were asked, two questions in the survey implicitly covered this. 6 out of 7 respondents believed that a higher frequency of follow-up visits does not necessarily result in higher chances of survival. The other respondent noted that it would lead to higher chances of survival 'for some specific indications', e.g. high risk of lung metastasis. However, 4 out of 7 respondents do think that timely recurrence detection can improve survival chances. This also suggests that a higher follow-up frequency in itself does not necessarily improve survival, but only if follow-up visits leads to earlier detection of recurrence. Therefore, the current model structure (where follow-up frequency is linked to survival) is maintained.

4.2.3 Prior and Conditional Probability Tables

The prior probability tables (PPTs) for tumour grade, surgical margin, expected remaining lifespan and fitness are calculated using the historical patient data. These PPTs are found in the appendix (A.6 Prior Probability Tables for Root Nodes). Here the CPTs for the consequence of recurrence and probability of recurrence are presented.

The CPT for consequence of recurrence is based on question 8 from the survey, and shown in Table 36. It can be interpreted as follows: Given that a patient has more than 40 years of expected remaining lifespan and the patient is fit, then the consequence of recurrence of this patient is always high (hence the probability of 1, meaning 100%). However, if the patient has the same expected remaining lifespan but is unfit, then the consequence is considered moderate in 50% of the cases and high in the other 50%.

Table 36: Conditional probability table for the variable consequence of recurrence (rounded to 3 decimals)

Expected remaining lifespan	Fitness	Consequence of recurrence		
		Low	Moderate	High
More than 40 years	Fit	0	0	1
	Unfit	0	0.333	0.667
Between 10 and 40 years	Fit	0	0	1
	Unfit	0.167	0.500	0.333
Less than 10 years	Fit	0	1	0
	Unfit	0.667	0.333	0

The CPT for probability of recurrence comes directly from the historical patient data and is shown in Table 37.

Table 37: Conditional probability table for the variable Probability of recurrence (rounded to 3 decimals)

Tumour grade	Surgical margin	Probability of recurrence	
		No recurrence	Recurrence
Low	R0	0.922	0.078
	R1/R2	0.840	0.160
High	R0	0.661	0.339
	R1/R2	0.556	0.444

The CPT for risk of recurrence is based on the traditional risk matrix and shown in Table 38. This approach was validated through the survey, as respondents classified risk in a way that closely matched our adapted risk matrix.

Table 38: Conditional probability table for the variable Risk of recurrence (rounded to 3 decimals)

Consequence of recurrence	Probability of recurrence	Risk of recurrence	
		Low	High
Low	No recurrence	1	0
	Recurrence	1	0
Moderate	No recurrence	1	0
	Recurrence	0	1
High	No recurrence	0	1
	Recurrence	0	1

4.2.4 Model Implementation

After the upper part of the model (everything leading up to and including risk of recurrence) was finished, the threshold for risk classification was determined. The entire evaluation process is described in the appendix (A.7 Risk Percentage Threshold Determination). The 40% threshold scored the best and was therefore selected.

To enable both the display of observed follow-up adherence and survival distributions for patients with similar risk levels, and to allow clinicians to simulate decisions, a few adaptations were made to the previously presented model structure. The resulting extensive risk model is shown in Figure 21.

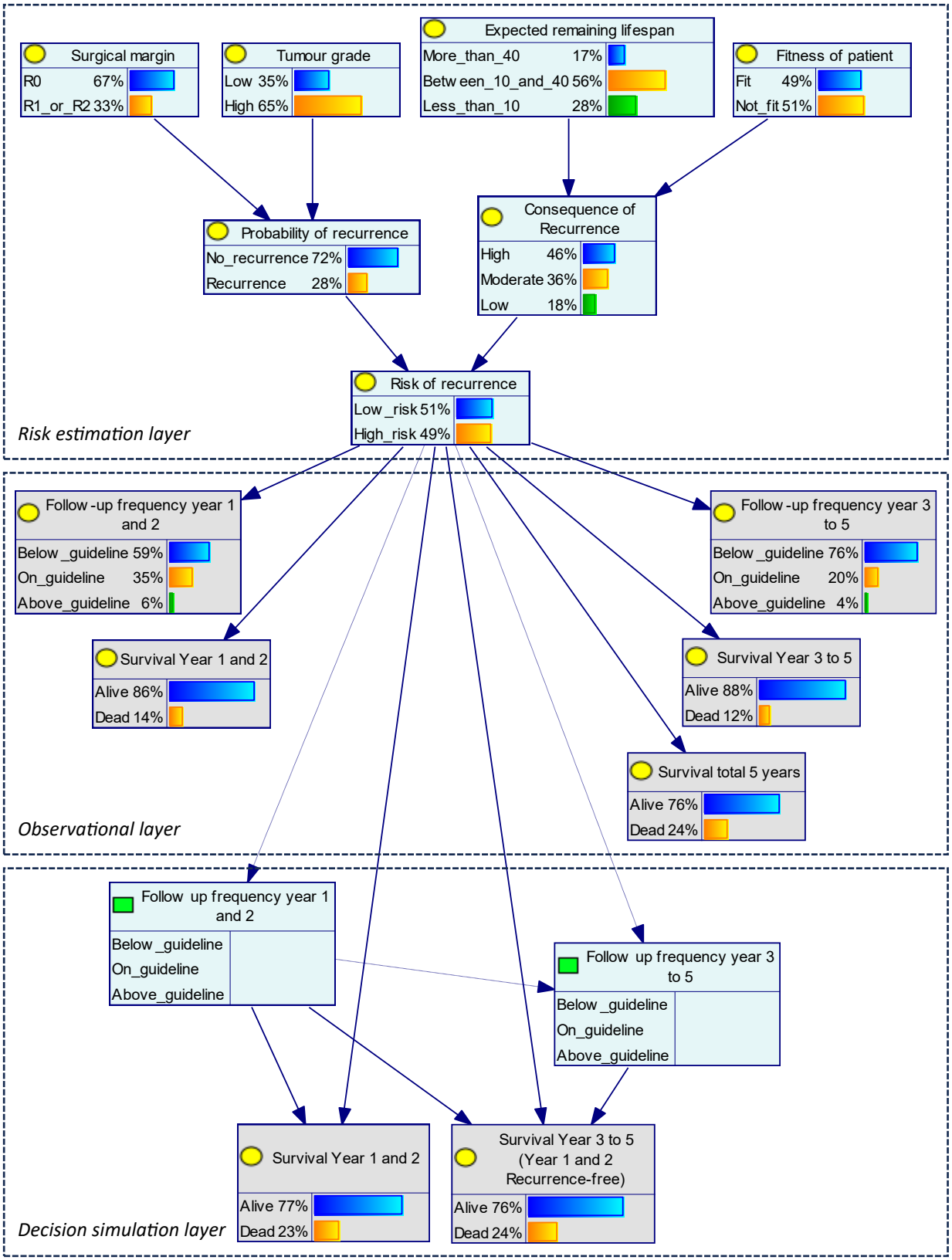


Figure 21: Implementation of Extensive Risk Model as Bayesian network, indicating the three different layers, and nodes for which evidence should never be set are depicted in grey

The Extensive Risk Model shown in Figure 21 was run in prior mode, meaning that no evidence is set for any of the nodes. It directly shows the PPTs for the root nodes. For all other (non-root) nodes, the model calculates the probability distributions through inference, based on the combination of PPTs and CPTs. The percentages in the node *risk of recurrence* indicate that, given the prior probability distributions, 51% of the patients are estimated to be low-risk while 49% are estimated to be high-risk.

The Extensive Risk Model consists of three conceptual layers:

1. Risk Estimation Layer
   This layer is similar to the original model structure. Evidence can be set for the root nodes, which then determines the distributions for probability, consequence and risk of recurrence are given. However, if the clinician, for instance, disagrees with the assigned *consequence of recurrence* for a certain patient, evidence can also be set directly for this node. The same applies to the *probability of recurrence* and *risk of recurrence* nodes.

2. Observational Layer
   This layer displays the distributions of follow-up frequency and survival outcomes, as observed in historical data for patients with a similar risk level. The nodes in this layer are shown in grey to indicate that no evidence should be entered for them. Setting evidence for these nodes would alter the inferred distributions in the risk estimation layer, which is undesirable. The left side of this layer (and the decision simulation layer) focuses on the first two years of follow-up, while the right side is about years three to five and the total follow-up period. The follow-up frequency nodes indicate the proportion of similar-risk patients who received a follow-up frequency below, on, or above guideline recommendations. The survival nodes show the proportion of patients with a similar risk-level who survived or died within each time period.

3. Decision Simulation Layer
   This layer allows clinicians to simulate decisions by selecting a follow-up frequency (below, on, or above guideline) and viewing the associated survival outcomes for patients with a similar risk-level in the dataset. The follow-up frequency nodes in this layer are decision nodes, indicated by a green square. Since they represent user-defined choices, they do not show bar charts. These nodes do not alter the inferred distributions in the risk estimation layer and they are just there to let the clinician simulate what survival chances are in combination with a certain follow-up strategy. The survival nodes are again grey, because evidence should not be set for these nodes.

   Note that there are arcs from the *follow-up frequency in year 1 and 2* to both the follow-up frequency and survival in years 3 to 5. These arcs are included because the decisions for the first two years and the following three years are made at different points in time. The follow-up strategy for years 1 and 2 is determined first. After those two years, clinicians reassess the situation and make a new decision for years 3 to 5, based on what occurred during the first two years.

   Furthermore, note that the node *Survival Year 1 and 2* is in both the observational layer and the decision simulation layer. However, they have different prior distributions, which is due to the fact that decision nodes have no prior probabilities, since they represent choices rather

than observed data. If a node has a decision node as a parent, GeNIe weighs all decision options equally (meaning that below, on, and above guideline each get 1/3 weight). This leads to a different prior distribution compared to the node in the observational layer (with only *risk of recurrence* as a parent), which reflects the actual data distribution.

The distributions of the *Survival in Years 3 to 5* nodes also differ between the observational and decision simulation layer. This is, however, not only due to the presence of the decision node *Follow-up Frequency in Years 3 to 5*, but also reflects an actual difference in the underlying data and CPTs. In the dataset, once a patient experiences a recurrence, follow-up visits after that point are no longer recorded. So, if a patient experienced a recurrence in the first two years, the follow-up frequency for the latter three years is unknown. Whilst we do know the survival outcomes for these patients, we do not know what influence follow-up frequency had on the survival outcomes. Therefore, the *Survival in Years 3 to 5* node in the decision simulation layer only includes outcomes for patients who did not experience a recurrence in the first two years.

A similar implementation in GeNIe Modeller was created for the Simplified Risk Model, as shown in Figure 22. The only structural difference compared to the implementation of the Extensive Risk Model lies in the risk estimation layer, as *tumour grade* is directly linked to *risk of recurrence*. Both the observational layer and decision simulation layer have identical structures in the two models, although their CPTs differ. For the Clinician model and the Tumour grade model, similar implementations were also made. Again, the only structural difference lies in the risk estimation layer, which only consists of one node in these models.

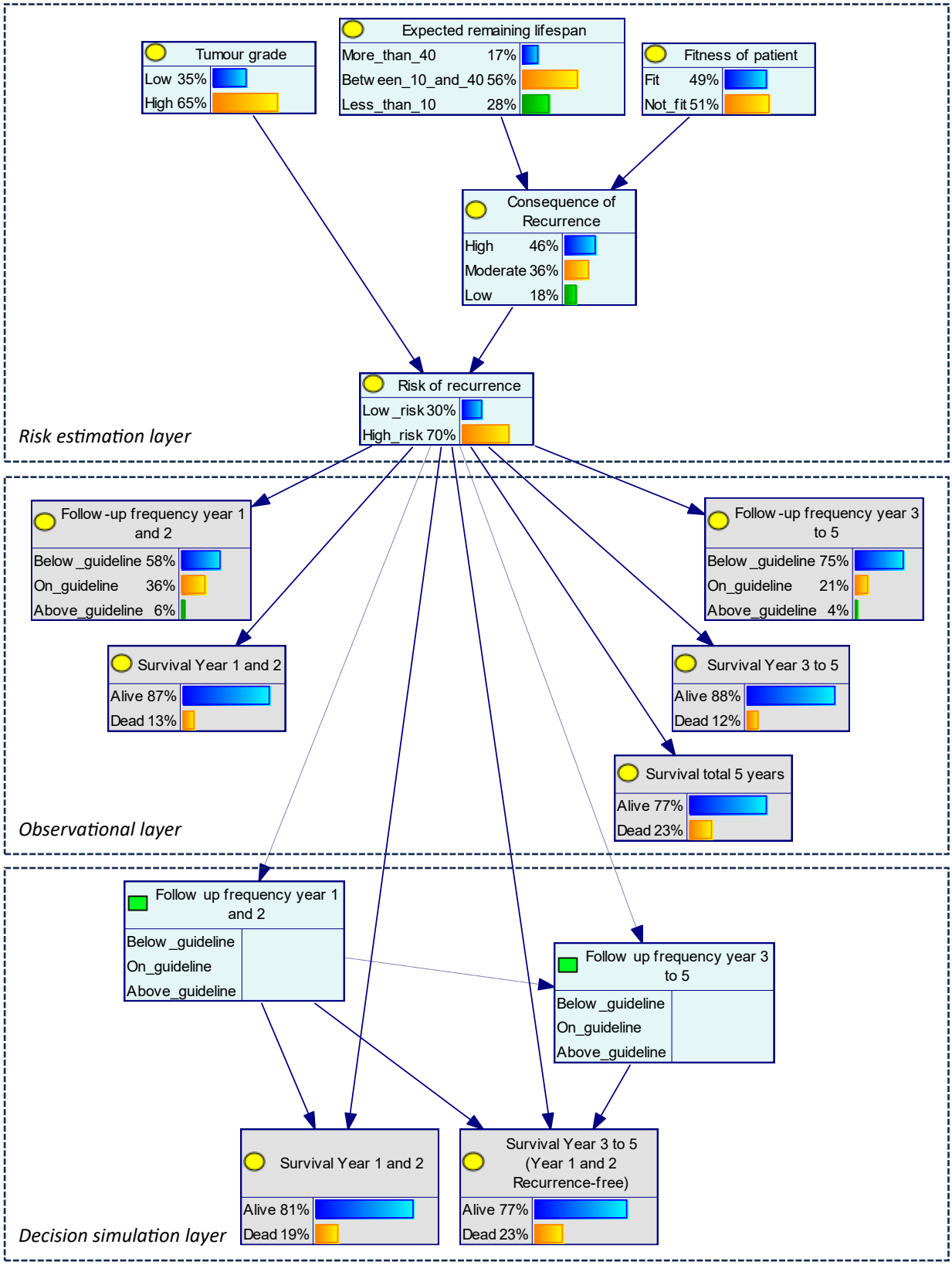


Figure 22: Implementation of Simplified Risk Model as Bayesian network, indicating the three different layers, and nodes for which evidence should never be set are depicted in grey

4.2.5 Model Validation

All possible patient profiles were input into the Extensive risk model and the Simplified risk model. Their corresponding probability of recurrence (only for the Extensive risk model), consequence of recurrence (which is similar for both models) and high-risk percentage are given in Table 39. All high-risk patient profiles are coloured red (their high-risk percentage is bigger than the 40% threshold).

Table 39: All possible patient profiles for both the Extensive and Simplified risk model, with their corresponding probability of recurrence (only for the Extensive risk model), consequence of recurrence and the high-risk percentage. High-risk patient profiles (with threshold 40%) are coloured red.

Fitness	Expected remaining lifespan	Tumour grade	Surgical margin	Probability of recurrence (only for extensive)	Consequence of recurrence			High-risk percentage	
					Low	Moderate	High	Extensive risk model	Simplified risk model
Fit	10-	High	R0	34	0	100	0	34	100
			R1/R2	44	0	100	0	44	
		Low	R0	8	0	100	0	8	0
			R1/R2	16	0	100	0	16	
	10 to 40	High	R0	34	0	0	100	100	100
			R1/R2	44	0	0	100	100	
		Low	R0	8	0	0	100	100	100
			R1/R2	16	0	0	100	100	
	40+	High	R0	34	0	0	100	100	100
			R1/R2	44	0	0	100	100	
		Low	R0	8	0	0	100	100	100
			R1/R2	16	0	0	100	100	
Unfit	10-	High	R0	34	67	33	0	11	33
			R1/R2	44	67	33	0	15	
		Low	R0	8	67	33	0	3	0
			R1/R2	16	67	33	0	5	
	10 to 40	High	R0	34	17	50	33	50	83
			R1/R2	44	17	50	33	56	
		Low	R0	8	17	50	33	37	33
			R1/R2	16	17	50	33	41	
	40+	High	R0	34	0	33	67	78	100
			R1/R2	44	0	33	67	81	
		Low	R0	8	0	33	67	69	67
			R1/R2	16	0	33	67	72	

We observe that the Extensive and Simplified risk models mostly classify patients in a similar way. If either the probability of recurrence or consequence of recurrence, or both, are considered high, the high-risk percentage is higher, following the risk matrix in Table 7.

All patients who are expected to still live for more than 40 years are classified high-risk. This is due to the fact that a recurrence for these patients would have severe consequences, as the patient could potentially lose many life years. Similarly, all fit patients with a remaining lifespan between 10 to 40 years are also considered high-risk. On the other hand all unfit patients with less than 10 expected life years are considered low-risk. A recurrence treatment would probably cause them more harm than the actual recurrence. All fit patients who are expected to live less than ten years are also considered low-risk, except for patients with high-grade tumours in the Simplified risk model, probably because the high tumour grade leads to a higher chance of recurrence and the fit patient

would be able to tolerate treatment. One the other hand, one could argue that a patient with less than 10 years remaining would not really benefit of recurrence treatment anyway, which is reflected by the Extensive risk model. Another difference between the models appears for unfit patients with 10 to 40 years of expected lifespan. In both models, patients with low-grade tumours in this group are classified as low-risk. However, for high-grade tumours, the Extensive model only classifies those with R1/R2 margins as high-risk, while the Simplified model considers all these patients with a high-grade tumour high-risk. Since R0 margins have lower probability of recurrence, it is reasonable that this leads to a difference.

Since the dataset lacks information on patient fitness, we cannot directly use the models that include fitness with the historical patient data. Therefore, both the Extensive risk model and the Simplified risk model are evaluated in two different ways for validation:

- With a fitness assumption: Patients under 60 are assumed to be fit; patients over 60 are assumed to be unfit.
- Without patient consequences: The models are simplified by removing the patient consequences aspect, meaning only the recurrence probability influences the risk classification. In the adjusted Extensive risk model, surgical margin and tumour grade directly determine recurrence risk. In the simplified version of the Simplified risk model, only tumour grade is used, making it identical to the Grade model.

The five model versions that are evaluated are:

- Extensive risk model including patient consequences (with fitness assumption)
- Extensive risk model excluding patient consequences
- Simplified risk model including patient consequences (with fitness assumption)
- Simplified risk model excluding patient consequences / Grade model
- Clinician model

To assess consistency between these models, an agreement matrix was created (

Table 40), showing how often each pair of models classified patients the same way.

The highest agreement percentage (95.1%) is found for the Simplified risk model and the Extensive risk model, both including patient consequences. Both use tumour grade to estimate recurrence probability and calculate consequences in a similar way, though the Extensive model also includes surgical margin, explaining the small difference. The Extensive risk model without patient consequences and the Clinician model also show relatively high agreement (87.2%), likely because both base classification on tumour grade and surgical margin.

In contrast, when comparing models with patient consequences to those without, agreement percentages drop to between 35.4% and 61.0%. This indicates that the inclusion of patient consequences appears important for classification. The Clinician and Grade models agree on only 41.4% of cases. This is explained by the fact that the Clinician model is driven mainly by surgical margin, while the Grade model considers only tumour grade. Lastly, the Extensive risk model without patient consequences and the Grade model agree on 54.2% of classifications. As shown in Table 37, using a 40% risk threshold in the Extensive model results in only patients with both a high-grade tumour and an R1/R2 margin being classified as high-risk, which explains the observed differences.

Table 40: Agreement matrix of the five model types

Model type		Extensive risk model		Simplified risk model		Clinician model
		including patient consequences	*excluding patient consequences*	*including patient consequences*	*excluding patient consequences, similar to* **Grade model**	
Extensive risk model	*including patient consequences*	100%	35.4%	95.1%	56.1%	41.1%
	excluding patient consequences	35.4%	100%	40.3%	54.2%	87.2%
Simplified risk model	*including patient consequences*	95.1%	40.3%	100%	61.0%	36.2%
	excluding patient consequences, similar to **Grade model**	56.1%	54.2%	61.0%	100%	41.4%
Clinician model		41.1%	87.2%	36.2%	41.4%	100%

4.2.6 Summary SQ2

Four model versions were implemented: the Extensive risk model, the Simplified risk model, the Clinician model and the Tumour grade model. In the Extensive risk model, both tumour grade and surgical margin influence probability of recurrence, but in the Simplified risk model, tumour grade on its own represents the probability of recurrence. Both of these models include a patient-centred side, which reflects the consequence of finding a recurrence. This aspect of the model reflects that a recurrence does not have the same impact on every patient. It is influenced by expected remaining lifespan and fitness of the patient. To allow for comparison, two other models were created, solely including risk of recurrence as classified by the STS expert or tumour grade, follow-up frequency and survival.

The Extensive and Simplified risk models nearly classify patients identically as low- and high-risk. All input combinations for the two models lead to reasonable and coherent risk classifications. The risk models were evaluated using historical data, accounting for missing fitness information by applying a fitness assumption or excluding patient consequences. This led to five model versions: two versions of the Extensive risk model, two of the Simplified risk model (of which one is identical to the Grade model), and one Clinician model. Models with similar input variables showed high agreement, particularly those using tumour grade and surgical margin. In contrast, agreement was lower when comparing models with and without patient consequences.

4.3 Evaluation the impact of risk-based follow-up on patient outcomes and care efficiency

In this sub-question, the relation between follow-up frequency and patient outcomes will first be analysed for each risk model. Secondly, the economic implications of these different follow-up strategies will be analysed.

The five model versions that were mentioned in section 4.2.5 Model Validation, will be used in this sub-question as well. They are as follows:

- Extensive risk model including patient consequences (with fitness assumption)
- Extensive risk model excluding patient consequences

- Simplified risk model including patient consequences (with fitness assumption)
- Simplified risk model excluding patient consequences / Grade model
- Clinician model

4.3.1 Patterns in Mortality and Follow-up Frequency

Using historical patient data, Table 41 was created to present how five-year death rates varied with follow-up frequency across models, based on historical patient data. It shows the total number of high- and low-risk patients as classified by each model, as well as how many received follow-up care below, on, or above the guideline frequency and the corresponding death rates.

For example, in the Extensive risk model including patient consequences, 80 of the 122 low-risk patients (65.6%) received below-guideline follow-up. Among them, 19 died within five years, resulting in a death rate of 23.8%. In contrast, 38 low-risk patients (31.1%) received on-guideline follow-up, with 17 (44.7%) dying during the follow-up period. The remaining 4 patients (3.3%) had above-guideline follow-up, of whom 3 (75.0%) died.

By looking at every row in Table 41 in a similar way, a clear trend is shown: higher follow-up frequency is associated with higher death rates, regardless of the risk model or risk level. In every model, patients with above-guideline follow-up had higher death rates than those with on-guideline follow-up. The difference in death rate is about 30% between on- and above-guideline follow-up for most models and risk levels. When comparing above-guideline to below-guideline groups, death rates are again consistently higher for the above-guideline group, with an average increase of 54.9%. Even on-guideline patients show higher death rates than those in the below-guideline group in every model and risk category.

The findings from this table show that regardless of how risk is classified, the death rate is higher amongst patients with increased follow-up frequency in the first two years of the follow-up period. This does not imply that a higher-follow-up frequency itself causes increased mortality. It can rather be seen as a reflection of clinical decision-making in the past. Patients who appeared to be at higher risk, and therefore more likely to die sooner, were assigned more intensive follow-up frequency in order to detect a recurrence earlier. However, more intensive follow-up does not necessarily seem to improve chances of survival. This was in line with responses to the survey, where STS experts indicated that higher-follow-up frequencies do not necessarily lead to higher chances of survival.

Looking within each model and follow-up frequency, we observe that high-risk patients consistently have higher death rates than low-risk patients in the Extensive risk model excluding patient consequences, the Grade model and the Clinician model. The only exception is in the below-guideline group of the Clinician model.

However, in both the Extensive and Simplified risk models that include patient consequences, the opposite pattern is observed: low-risk patients exhibit higher death rates across all follow-up frequencies. This is likely explained by the fact that expected remaining lifespan, based on age, is taken into account in these models. Older patients are more often considered low-risk, but older patients are more likely to die sooner due to age-related mortality.

Table 41: Patients and deaths (over entire 5-year follow-up period) per model and their risk classifications, per follow-up frequency in year 1-2

Model type	Patient risk level	Follow-up frequency year 1-2 (Death taken over entire 5-year follow-up period)					
		Below guideline		On guideline		Above guideline	
		Count (%)	Deaths (%)	Count (%)	Deaths (%)	Count (%)	Deaths (%)
Extensive risk model *(including patient consequences)*	Low-risk (n=122)	80 (65.6%)	19 (23.8%)	38 (31.1%)	17 (44.7%)	4 (3.3%)	3 (75.0%)
	High-risk (n=245)	137 (55.9%)	12 (8.8%)	91 (37.1%)	26 (28.6%)	17 (6.9%)	10 (58.8%)
Extensive risk model *(excluding patient consequences)*	Low-risk (n=295)	182 (61.7%)	24 (13.2%)	97 (32.9%)	26 (26.8%)	16 (5.4%)	9 (56.2%)
	High-risk (n=72)	35 (48.6%)	7 (20.0%)	32 (44.4%)	17 (53.1%)	5 (6.9%)	4 (80.0%)
Simplified risk model *(including patient consequences)*	Low-risk (n=140)	96 (68.6%)	19 (19.8%)	40 (28.6%)	17 (42.5%)	4 (2.9%)	3 (75.0%)
	High-risk (n=227)	121 (53.3%)	12 (9.9%)	89 (39.2%)	26 (29.2%)	17 (7.5%)	10 (58.8%)
Grade model	Low-grade (n=127)	108 (85.0%)	3 (2.8%)	18 (14.2%)	4 (22.2%)	1 (0.8%)	1 (100.0%)
	High-grade (n=240)	109 (45.4%)	28 (25.7%)	111 (46.2%)	39 (35.1%)	20 (8.3%)	12 (60.0%)
Clinician model	Low-risk (n=248)	139 (56.0%)	23 (16.5%)	93 (37.5%)	26 (28.0%)	16 (6.5%)	9 (56.2%)
	High-risk (n=119)	78 (65.5%)	8 (10.3%)	36 (30.3%)	17 (47.2%)	5 (4.2%)	4 (80.0%)

A similar table was created for years 3 to 5 of the follow-up period, shown in Table 42. In contrast to Table 41, this table does not show a clear overall trend. When comparing above-guideline follow-up frequency with on-guideline frequency, 8 out of 10 rows show a higher death rate for the above-guideline group. The remaining two rows (high-risk patients in both the extensive risk model excluding patient consequences and the clinician model) had no deaths in either the on-guideline or above-guideline groups. Across all models, the average difference in death rate between these two follow-up categories is 36.2%.

Similarly, when comparing above-guideline with below-guideline follow-up frequency, 8 out of 10 rows again show a higher death rate in the above-guideline group. The same two exceptions appear here as well, with no deaths in either of the groups.

However, when comparing on-guideline with below-guideline follow-up frequencies, only 4 out of 10 rows show a lower death rate in the on-guideline group. The colour coding in the table also suggests that the death rates between these two categories are relatively close. Therefore, in years 3 to 5, an above-guideline follow-up frequency is generally associated with higher death rates, but there is no clear difference between on- and below-guideline frequencies.

Additionally, unlike Table 41, Table 42 does not show a consistent pattern in death rates when comparing low- and high-risk patients within each model.

Table 42: Patients and deaths per model and their risk classifications, per follow-up frequency in year 3-5

Model type	Patient risk level	Follow-up frequency year 3-5					
		Below guideline		On guideline		Above guideline	
		Count (%)	Deaths (%)	Count (%)	Deaths (%)	Count (%)	Deaths (%)
Extensive risk model *(including patient consequences)*	Low-risk (n=69)	56 (81.2%)	6 (10.7%)	12 (17.4%)	0 (0.0%)	1 (1.4%)	1 (100.0%)
	High-risk (n=172)	126 (73.3%)	1 (0.8%)	38 (22.1%)	2 (5.3%)	8 (4.7%)	1 (12.5%)
Extensive risk model *(excluding patient consequences)*	Low-risk (n=212)	161 (75.9%)	7 (4.3%)	43 (20.3%)	2 (4.7%)	8 (3.8%)	2 (25.0%)
	High-risk (n=29)	21 (72.4%)	0 (0.0%)	7 (24.1%)	0 (0.0%)	1 (3.4%)	0 (0.0%)
Simplified risk model *(including patient consequences)*	Low-risk (n=83)	70 (84.3%)	6 (8.6%)	12 (14.5%)	0 (0.0%)	1 (1.2%)	1 (100.0%)
	High-risk (n=158)	112 (70.9%)	1 (0.9%)	38 (24.1%)	2 (5.3%)	8 (5.1%)	1 (12.5%)
Grade model	Low-grade (n=111)	97 (87.4%)	3 (3.1%)	13 (11.7%)	0 (0.0%)	1 (0.9%)	1 (100.0%)
	High-grade (n=130)	85 (65.4%)	4 (4.7%)	37 (28.5%)	2 (5.4%)	8 (6.2%)	1 (12.5%)
Clinician model	Low-risk (n=173)	124 (71.1%)	7 (5.6%)	41 (23.7%)	2 (4.9%)	8 (4.6%)	2 (25.0%)
	High-risk (n=68)	58 (85.3%)	0 (0.0%)	9 (13.2%)	0 (0.0%)	1 (1.5%)	0 (0.0%)

4.3.2 Economic Implications of Follow-Up Strategies

Table 41 showed that, across all risk classification models, patients who received more frequent follow-up in years 1 and 2 had higher death rates over the five-year follow-up period. Table 44 builds on this by estimating the economic implications of these follow-up strategies.

Similar to Table 41, Table 43 includes the number of patients, deaths, and death rates for each risk group and follow-up frequency. In addition, this table includes the number of follow-up visits that each patient risk level group received in each of the models for each of the follow-up categories. The number of follow-up visits was multiplied by 687.11, which is the estimated cost of one follow-up visit (including outpatient visit and MRI-scan) (LUMC, 2025), to give an estimate of the total follow-up costs for a certain patient group. Both the average follow-up cost for the first two years for every patient within the group and every survivor within the group are given.

For example, in the Extensive risk model including patient consequences, 80 low-risk patients received below-guideline follow-up. These patients had 202 follow-up visits in total according to the historical data, resulting in an estimated total cost of €138,796. Divided over 80 patients, this equals

€1,735 per patient. However, only 61 of these patients survived the five years, making the cost per survivor higher, at €2,275.

When comparing the average costs per patient across the three follow-up frequencies, a clear trend is shown:

- Below guideline: €1,711 to €2,061
- On-guideline: €3,135 to €4,008
- Above-guideline: €5,359 to €8,761

This shows that costs roughly double from below- to on-guideline, and double again from on- to above-guideline. Average costs per patient in the above-guideline group are three to five times the cost per patient compared to those in the below-guideline group.

The differences become even larger when looking at the average costs per survivor:

- Below-guideline: €1,786 to €2,363
- On-guideline: €5,153 to €6,688
- Above-guideline: €19,840 to €26,797

Here, the on-guideline group is two to three times more expensive than the below-guideline group, while the above-guideline group is about four times more expensive than the on-guideline group and eight to thirteen times more expensive than the below-guideline group.

These findings show that a more intensive follow-up in the first two years, which also leads to substantially higher costs, does not necessarily mean that survival chances are improved.

Table 43: Economic effects of follow-up strategies using different risk classifications for year 1 and 2 of follow-up period, but survival throughout the entire 5-year follow-up period

Model type	Patient risk level	Follow-up frequency year 1-2																		
		Below guideline							On guideline							Above guideline						
		Patients	Deaths (yr 1-5)	Visits (yr 1-2)	Total cost (yr 1-2)	Average cost (yr 1-2) per patient	Average cost (yr 1-2) per survivor (yr 1-5)	Death rate (yr 1-5)	Patients	Deaths (yr 1-5)	Visits (yr 1-2)	Total cost (yr 1-2)	Average cost (yr 1-2) per patient	Average cost (yr 1-2) per survivor (yr 1-5)	Death rate (yr 1-5)	Patients	Deaths (yr 1-5)	Visits (yr 1-2)	Total cost (yr 1-2)	Average cost (yr 1-2) per patient	Average cost (yr 1-2) per survivor (yr 1-5)	Death rate (yr 1-5)
Extensive risk model *(including patient consequences)*	Low-risk (n=122)	80	19	202	138796	1735	2275	23.8%	38	17	187	128490	3381	6119	44.7%	4	3	35	24049	6012	24049	75.0%
	High-risk (n=245)	137	12	398	273470	1996	2188	8.8%	91	26	512	351800	3866	5412	28.6%	17	10	208	142919	8407	20417	58.8%
Extensive risk model *(excluding patient consequences)*	Low-risk (n=295)	182	24	507	348365	1914	2205	13.2%	97	26	553	379972	3917	5352	26.8%	16	9	204	140170	8761	20024	56.3%
	High-risk (n=72)	35	7	93	63901	1826	2282	20.0%	32	17	146	100318	3135	6688	53.1%	5	4	39	26797	5359	26797	80.0%
Simplified risk model *(including patient consequences)*	Low-risk (n=140)	96	19	239	164219	1711	2133	19.8%	40	17	199	136735	3418	5945	42.5%	4	3	35	24049	6012	24049	75.0%
	High-risk (n=227)	121	12	361	248047	2050	2276	9.9%	89	26	500	343555	3860	5453	29.2%	17	10	208	142919	8407	20417	58.8%
Grade model	Low-grade (n=127)	108	3	273	187581	1737	1786	2.8%	18	4	105	72147	4008	5153	22.2%	1	1	12	8245	8245	-	100.0%
	High-grade (n=240)	109	28	327	224685	2061	2774	25.7%	111	39	594	408143	3677	5669	35.1%	20	12	231	158722	7936	19840	60.0%
Clinician model	Low-risk (n=248)	139	23	399	274157	1972	2363	16.5%	93	26	531	364855	3923	5446	28.0%	16	9	204	140170	8761	20024	56.3%
	High-risk (n=119)	78	8	201	138109	1771	1973	10.3%	36	17	168	115434	3207	6075	47.2%	5	4	39	26797	5359	26797	80.0%

Table 42 showed that, in years 3 to 5 of the follow-up period, a higher follow-up frequency is still associated with higher death rates, though less consistently than in years 1 and 2. Table 44 builds on this by presenting the estimated economic implications of these follow-up strategies during years 3 to 5.

When comparing the average cost per patient across the three follow-up frequency groups, we again observe a cost increase with higher follow-up intensity:

- Below-guideline: €960 to €1,455
- On-guideline: €3,321 to €4,647
- Above-guideline: €7,558 to €10,221

This shows that average follow-up costs per patient increase three to four times when moving from below- to on-guideline. The costs double when going from on-guideline to above-guideline follow-up. Above-guideline follow-up is about seven to eight times more expensive on average compared to below-guideline follow-up.

The differences are again even larger when looking at the average costs per survivor:

- Below-guideline: €960 to €1,527
- On-guideline: €3,321 to €4,905
- Above-guideline: €7,558 to €13,628

Here, the on-guideline group is roughly three times more expensive than the below-guideline group, and the above-guideline group is two to three times more expensive than the on-guideline group. Compared to the below-guideline group, the above-guideline group is six to ten times more expensive per survivor.

These findings show that the economic burden of more intensive follow-up in years 3 to 5 is substantial, particularly when accounting for survival outcomes. Intensive follow-up is expensive, while there might not be improved survival chances.

Table 44: Economic effects of follow-up strategies using different risk classifications for year 3 to 5 of follow-up period, survival is also taken for years 3 to 5 of the follow-up period

Model type	Patient risk level	Follow-up frequency year 3-5																						
		Below guideline							On guideline							Above guideline								
		Patients	Deaths (yr 3-5)	Visits (yr 3-5)	Total cost (yr 3-5)	Average cost (yr 3-5) per patient	Average cost (yr 3-5) per survivor (yr 3-5)	Death rate (yr 3-5)	Patients	Deaths (yr 3-5)	Visits (yr 3-5)	Total cost (yr 3-5)	Average cost (yr 3-5) per patient	Average cost (yr 3-5) per survivor (yr 3-5)	Death rate (yr 3-5)	Patients	Deaths (yr 3-5)	Visits (yr 3-5)	Total cost (yr 3-5)	Average cost (yr 3-5) per patient	Average cost (yr 3-5) per survivor (yr 3-5)	Death rate (yr 3-5)
Extensive risk model *(including patient-centred aspect)*	Low-risk (n=122)	56	6	82	56343	1006	1127	10.7%	12	0	58	39852	3321	3321	0.0%	1	1	11	7558	7558	-	100.0%
	High-risk (n=245)	126	1	241	165594	1314	1325	0.8%	38	2	257	176587	4647	4905	5.3%	8	1	119	81766	10221	11681	12.5%
Extensive risk model *(excluding patient-centred aspect)*	Low-risk (n=212)	161	7	285	195826	1216	1272	4.3%	43	2	275	188955	4394	4609	4.7%	8	2	119	81766	10221	13628	25.0%
	High-risk (n=29)	21	0	38	26110	1243	1243	0.0%	7	0	40	27484	3926	3926	0.0%	1	0	11	7558	7558	7558	0.0%
Simplified risk model *(including patient-centred aspect)*	Low-risk (n=140)	70	6	103	70772	1011	1106	8.6%	12	0	58	39852	3321	3321	0.0%	1	1	11	7558	7558	-	100.0%
	High-risk (n=227)	112	1	220	151164	1350	1362	0.9%	38	2	257	176587	4647	4905	5.3%	8	1	119	81766	10221	11681	12.5%
Grade model	Low-grade (n=111)	97	3	143	98257	1013	1045	3.1%	13	0	83	57030	4387	4387	0.0%	1	1	11	7558	7558	-	100.0%
	High-grade (n=130)	85	4	180	123680	1455	1527	4.7%	37	2	232	159410	4308	4555	5.4%	8	1	119	81766	10221	11681	12.5%
Clinician model	Low-risk (n=173)	124	7	242	166281	1341	1421	5.6%	41	2	267	183458	4475	4704	4.9%	8	2	119	81766	10221	13628	25.0%
	High-risk (n=68)	58	0	81	55656	960	960	0.0%	9	0	48	32981	3665	3665	0.0%	1	0	11	7558	7558	7558	0.0%

4.3.3 Summary SQ3

Firstly, this sub-question analysed the relationship between follow-up frequency and mortality, across five risk classification models using historical patient data.

In the first two years of follow-up, there was a clear pattern: patients who received more intensive follow-up had higher five-year death rates across all models, regardless of risk level. This does not suggest that higher follow-up causes increased mortality, but rather reflects historical clinical decision-making, meaning that patients who were perceived to be at higher risk likely received more intensive monitoring, but were also more likely to die sooner.

Interestingly, in models that include patient consequences (and thus account for remaining life expectancy and fitness), low-risk patients had higher death rates. This is likely because older patients were more often labelled low-risk, but they do have higher age-related mortality.

In years 3 to 5, the association between higher follow-up and increased death rates was still found, but less consistently. There was no clear pattern when comparing on- and below-guideline follow-up, but the above-guideline follow-up did show higher death rates.

Cost analyses showed that more frequent follow-up is significantly more expensive. In the first two years, average follow-up costs per patient approximately doubled when moving from below- to on-guideline follow-up, and doubled again from on- to above-guideline levels. When looking at costs per survivor, these differences become even larger, reflecting the higher mortality rates in groups receiving more intensive follow-up. Therefore costs per survivor can be about four times higher in the above-guideline group compared to on-guideline, while it can be eight to thirteen times higher compared to below-guideline.

In years 3 to 5, a similar pattern is found, with costs per patient rising substantially as follow-up intensity increases. Costs per survivor are clearly higher for patients with above-guideline follow-up. This emphasizes the significant economic burden of more intensive monitoring, particularly when it is uncertain whether this leads to better survival outcomes.

More intensive follow-up is strongly associated with higher mortality and substantially higher costs, particularly in the first two years after treatment. However, this does not reflect a causal relationship with mortality, but rather the prioritization of follow-up in patients who were perceived to be at higher risk by clinicians in the past. The findings highlight the considerable impact of intensive follow-up, while it does not necessarily appears to improve survival chances.

5. Discussion

5.1 Summary of Main Findings

It was first explored which tumour- and treatment-related factors are most strongly linked to recurrence and how they could be used to design risk-based follow-up strategies. Based on expert consultation, conditional dependency analysis, and subgroup comparisons, tumour grade and surgical margin were selected as the most important predictors of recurrence. Tumour grade showed the strongest and most consistent association with both recurrence and death, while other factors like tumour diameter, histology, and radiotherapy were excluded due to weak or indirect effects.

Two main risk models were developed, based on the previous findings, that incorporate both recurrence probability and patient-specific consequences (based on fitness and expected remaining lifespan). The difference between these two lies in how the probability of recurrence is estimated: one uses only tumour grade, while the other uses both tumour grade and surgical margin. In addition, two simpler models were created that only reflect recurrence probability: one based on tumour grade alone and the other based on risk classifications by a clinician. The consequence-based models showed strong internal consistency and coherent classifications, while the rest of the comparisons between models differed more in their classifications.

Historical data was used to compare follow-up intensity, survival, and cost across the different models. A clear pattern emerged: more intensive follow-up was associated with higher death rates and significantly higher costs, especially in the first two years. This is likely due to clinicians assigning more frequent follow-up to patients who were perceived as high-risk, rather than follow-up itself causing harm. A cost analysis showed that above-guideline follow-up can be up to 13 times more expensive per survivor than below-guideline follow-up, without clear evidence of improved survival. So, more intensive follow-up likely reflected how clinicians perceived a patient's risk at the time, but it did not seem to improve survival chances, while causing a substantial economic burden.

5.2 Interpretation of results

5.2.1 Risk Factors for Recurrence

First, tumour grade and surgical margin were selected as the most important factors for the probability of recurrence. While tumour grade showed a consistent and strong relation with recurrence and even survival, surgical margin did not show that as strong, and some other factors (like diameter) even appeared to have a stronger conditional dependency with recurrence. The difficult thing was that the available dataset already grouped R1 and R2 margins together, while they might 'behave' really differently, according to experts. Therefore, the effects of surgical margin might not show as pronounced as they could be. After discussions with an STS expert, we decided to retain surgical margin, even though the data did not show a very strong connection. We did this because surgical margin seemed more clinically relevant for the probability of recurrence, than e.g. diameter.

5.2.2 Alignment between Models

The agreement matrix (

Table 40) shows that risk classifications can vary substantially depending on which model for risk classification is used, particularly when patient consequences are included. While models that incorporated similar clinical variables (e.g. grade and margin) showed strong alignment, those that differed in their treatment of patient consequences had much lower agreement. Even with similar recurrence probabilities, including patient consequences in the decision process can shift patients between risk categories. This suggests that purely clinical classifications may not fully capture an important aspect of follow-up planning: potential benefits and burdens of timely recurrence detection and potential recurrence treatment for the patient.

The relatively low agreement between the Clinician model and the Grade model highlights the complexity of expert judgment, where incorporating multiple variables in the process of classification does not necessarily lead to more accurate predictions. In fact, tumour grade alone consistently outperformed the broader STS expert classification in predicting survival. This raises important questions about which variables should be prioritised in risk classifications. However, as stated before, the predictive value of surgical margin may have been underestimated in this study, as the dataset grouped R1 and R2 margins together.

It is important to realise that risk is a perceptive concept, meaning that everyone could interpret risk differently. Clinicians might weigh certain factors more heavily based on their own experience, the patient in front of them or the context they are working in. In that sense, it is not necessarily a problem that the models do not all agree on their classifications. Instead, the differences highlight that there is not just one 'correct' way to classify risk. Making these difference clear can actually help clinicians reflect on their own reasoning and compare it with other perspectives.

5.2.3 Mortality and Follow-up Patterns

Looking at patterns in mortality across follow-up frequencies, one of the most interesting observations comes from the below-guideline group in the first two years, where the Grade model stands out in its ability to distinguish between high- and low-risk patients (Table 41). In this group, where around two-thirds of patients belong, low-grade patients show substantially lower five-year mortality rates than high-grade patients. The Grade model has a strong predictive ability for the large group of patients with below-guideline follow-up in the first two years. In contrast, the other models show less pronounced differences in death rates between their risk categories within the below-guideline group, suggesting that their classifications are less informative in this context. Although in the on- and above-guideline groups the survival differences between risk levels are more comparable across the models, the fact that most patients fall into the below-guideline category makes this category especially relevant when evaluating overall model performance. In that sense, tumour grade alone appears to be a strong and valuable indicator of survival.

Interestingly, even the Clinician model, based on STS expert classifications, does not outperform the Grade model in this part. In fact, its agreement with actual survival appears lower. So, while the expert weighs several factors at once, it does not seem to improve how well patients are separated based on actual survival. However, it is again possible that surgical margin, which is a key variable in the Clinician model, is underrepresented here due to the grouping of R1 and R2 resections into one category, underestimating its actual impact.

At the same time, the data does suggest that clinicians historically did make some kind of risk-based decision when assigning follow-up schedules. Across all models, patients who received more intensive follow-up in the first two years of follow-up had significantly higher death rates. This pattern is probably not due to follow-up causing worse outcomes, but rather reflects how clinicians assigned

a higher frequency of follow-up to patients who were perceived to have a poorer prognosis. In that sense, the follow-up frequency can be viewed as a proxy for perceived risk at the time.

This brings us to an important point, because even though follow-up frequency was adjusted based on perceived risk in the past, this did not seem to lead to better survival outcomes. Patients who received above-guideline follow-up didn't live longer. In fact, they often even had higher mortality rates. The aim of follow-up care is to detect recurrences earlier, under the assumption that early detection would improve recurrence treatment options and thus chances of survival. These findings, however, challenge this assumption. Even though, on average, a higher follow-up frequency means that a potential recurrence is found earlier, it does not improve the outcomes regarding survival. The above-guideline group has a higher death rate than both the on- and below-guideline group, and the on-guideline group has a higher death rate than the below-guideline group. This holds for all different risk models in the first two years of the follow-up. In other words, more follow-up may lead to earlier detection, but not necessarily to better survival.

In years three to five of follow-up, the strong link between frequency and mortality mostly disappears. There is less consistency in how survival is related to risk level or follow-up intensity. This suggests that the first two years post-treatment are the most informative, for both clinicians and outcomes. This also aligns with the emphasis in current guideline on the early years of follow-up, when recurrence is more likely.

5.2.4 Economic Implications of Follow-Up Strategies

The data from Table 43 and Table 44 show that increasing follow-up frequency leads to a substantial increase in costs, while the survival benefit remains unclear. The difference in costs becomes particularly clear when looking at the cost per survivor. In years 1 and 2, above-guideline follow-up costs around €20,000 to nearly €27,000 per survivor, compared to just €1,700 to €2,300 in the below-guideline groups. In years 3 to 5, the differences in costs are smaller, due to fewer follow-up visits in general in these years, as the likelihood of a recurrence decreased. Below-guideline follow-up costs €900 to €1500 per survivor, but above-guideline follow-up costs up to about €13000 per survivor.

If the patients with above-guideline frequency were to have had on-guideline frequency, about €3000 to €5000 euros per patient could be saved in the first two years. Per survivor, this would be about €15000 to €20000 euros. Put differently, intensifying follow-up from on-guideline to above-guideline can cost an additional €15,000 to €20,000 per survivor, while survival rates do not necessarily seem to improve.

These findings call for a reflection on the value and cost-effectiveness of intensive follow-up strategies. If higher follow-up frequency clearly translated into higher survival rates, an investment of around €20,000 to save a life might be justified. However, our results suggest that the substantial additional costs associated with above-guideline follow-up do not clearly correspond to improved survival outcomes. Surveillance is an important part of post-treatment care, but simply increasing follow-up frequency, e.g. to higher than the guideline, may not lead to actual benefits and could lead to unnecessary costs.

5.2 Scientific Contribution

Firstly, our analysis showed that tumour grade had the most explanatory power for recurrence and survival, which aligns with the ESMO guideline that also uses grade as the main factor to determine follow-up frequency (Gronchi et al, 2021).

Multiple studies have advocated for a more individualised approach to STS follow-up, based on tumour characteristics and recurrence risk (Cipriano et al., 2020; Dammerer et al., 2020; Ezuddin et

al., 2018; Kruijswijk et al., 2024; Noebauer-Huhmann et al., 2024; Tseng et al., 2015; Rothermundt et al., 2023; Spunt et al., 2020; Whitaker et al., 2023; Zaidi et al., 2018). However, these studies typically define risk only in terms of recurrence probability. In this study, we extend that definition by also incorporating patient-specific consequences of recurrence. We define risk not only as the likelihood of recurrence (based on tumour grade and surgical margins, or tumour grade by itself), but also as the potential severity of recurrence and its treatment, which we estimate using patient fitness and expected remaining lifespan. This offers a broader and more patient-centred risk classification.

Currently existing models, like the Sarculator and PERSARC, can predict overall survival, local recurrence and distant metastasis. They help clinicians and patients understand what different treatment strategies might mean for their chances of recurrence and survival (Callegaro et al., 2016; Callegaro et al., 2019; Pasquali et al., 2022; Van Praag et al., 2017; Willeumier et al., 2017; Rueten-Budde et al., 2018; Smolle et al., 2020; Rueten-Budde et al., 2021). Our model, however, focuses on something different. Rather than predicting recurrence or survival directly, it shows how different follow-up strategies relate to survival outcomes for a certain patient, based on similar patients in historical data. Another key difference is that Sarculator and PERSARC treat all included patient-, tumour-, and treatment-related variables as independent. The Bayesian network in this study does not make that assumption, but instead it captures how these factors interrelate to influence recurrence risk. Furthermore, unlike the existing models, our model includes a patient-centred aspect by also considering potential consequences of recurrence.

Using the dataset from Kruiswijk et al. (2024), we also confirm their findings that clinicians do not consistently adhere to existing follow-up guidelines, even using a tolerance of 20%. Moreover, our findings show that even when a clinician assigns patients to low- or high-risk categories, the follow-up intensity often does not align with this classification. For instance, the majority of patients receiving more frequent follow-up than recommended by the guidelines was classified as low risk (Table 28). This suggests that follow-up decisions are not always systematically risk-based (with ‘risk’ here referring only to probability of recurrence).

The lack of a clear link between follow-up intensity and recurrence risk becomes even more relevant given that previous studies already questioned the benefits of intensive follow-up regarding survival chances. Previous research has shown that more frequent surveillance does not necessarily improve survival (Puri et al., 2014; Glasbey et al., 2021), and our analysis supports this finding. In our data, higher follow-up frequency was not associated with improved survival outcomes, regardless of how recurrence risk was defined. It rather reflected historical clinical decision-making, where patients perceived as higher risk at the time were given more frequent follow-up, but also died sooner.

While this study could not focus on how recurrences were actually detected as this was not in the dataset, several studies have suggested that many local recurrences are identified by patients themselves, often before imaging detects them (Blaye et al., 2019; Cheney et al., 2014; England et al., 2020; Fujiki et al., 2016). Cheney et al. (2014), for instance, found that only 1 out of 11 local recurrences was detected through routine MRI. This raises the question of how much added value intensive imaging provides, particularly for low-risk patients and tumours that are easy to monitor physically. This could also explain why increasing follow-up frequency does not necessarily seem to improve survival chances, since patients mostly find recurrences themselves.

Lastly, our findings support a risk-based approach to follow-up that aligns with previous studies on imaging and recurrence detection. Royce et al. (2017) demonstrated that chest X-rays are sufficient for low-risk patients, while CT scans should be reserved for high-risk cases from a cost-effectiveness

perspective. This suggests that our risk model could help guide imaging modality choices, not just frequency.

5.3 Implications for Clinical Practice

The findings from this study highlight two key takeaways for clinical practice. First, tumour grade consistently appeared as the factor with the most explanatory power for both recurrence and survival, far above the other commonly analysed variables. This supports the ESMO guideline's decision to base follow-up frequency primarily on grade, and highlights the need for clinicians to make this a more central factor in their risk assessments and follow-up decision-making process.

Second, our analysis found that more frequent follow-up does not necessarily improve survival, regardless of how risk was defined. In the historical data, patients who received more intensive follow-up had worse outcomes on average. Follow-up intensity appeared to reflect the clinicians' perception of risk. At the same time, more frequent follow-up came with substantially higher costs, up to 13 times more per survivor, especially in the first two years of the follow-up period. Together with the findings from literature indicating that patients mostly find recurrences themselves, even before imaging can detect it (Blaye et al., 2019; Cheney et al., 2014; England et al., 2020; Fujiki et al., 2016), this could start a discussion on how follow-up resources are currently allocated. Rather than assigning intensive follow-up schedules with imaging for many patients, clinicians could consider whether those resources might be more effectively used for tumour sites that are harder to monitor physically or where recurrences tend to be asymptomatic. At the same time, investing in patient education and self-monitoring could also improve early detection, without only relying on routine imaging. This highlights the need for clinicians to reflect on how follow-up resources are allocated, especially when the likelihood of a recurrence is low or when early detection may have limited benefit for the patient.

Together, these findings emphasise the need for a more structured and evidence-based approach to follow-up planning. This means putting more emphasis on tumour grade in clinical decision-making, being more critical about when frequent follow-up is truly necessary, and weighing the costs and burden of follow-up against its potential benefits. The model developed in this study could support such discussions by offering a more explicit and systematic way to classify risk, which is not just based on recurrence probability, but also on the consequences for the patient. While it is not meant to replace clinical judgment, it can help clinicians reflect on how they weigh different factors in their decisions. The results highlight the need for a more policy-level discussion about how to use follow-up capacity in a way that optimizes both patient benefit and efficiency of the healthcare system.

5.4 Research Limitations and Recommendations for Future Research

While the research aims to develop a robust Bayesian decision-support model for STS follow-up decisions, there are several limitations to this study that need to be considered. Multiple recommendations for future research were also made. The limitations and recommendations are grouped per topic.

5.4.1 Generalisability and Bias

Firstly, the research relies on a retrospective dataset from a cohort of a single centre. This may introduce selection bias and limit the ability to generalise outcomes to other hospitals or healthcare systems. Differences in clinical protocols, patient demographics or follow-up strategies across institutions could affect the applicability of the findings.

Secondly, the Bayesian network is constructed based on interrelations found in the historical patient data combined with expert input. The structure of the network relies heavily on the input from STS

specialists. The expert elicitation process does, however, introduce a subjective component. Clinician decision-making may vary based on experience, personal preferences, or institutional guidelines, leading to variability in the input used to inform the model.

5.4.2 Dataset Limitations

One main limitation of this study lies with the variable surgical margin. The dataset used in this study contains surgical margins, but instead of using three categories (R0, R1 and R2), it combines R1 and R2 into one. Even though discussions with experts point out that there could be an important difference between margins R1 and R2, a distinction between R1 and R2 cannot be made in this studies. The fact that R1 and R2 were already grouped in the dataset used for this study, means that the results regarding surgical margin might be skewed. The effect of the surgical margin in this study seems small compared to tumour grade, but it could be underestimated due to the grouping of R1 and R2. Therefore, in future studies, it is advised to look into R2 margins separate from the rest.

Besides, for certain variables, the dataset reflects choices that were made by clinicians in the past. These choices create a bias in the dataset for certain variables. This is the case for radiotherapy and follow-up frequency. The choices clinicians made regarding whether or not to administer radiotherapy and how often to schedule follow-up, reflect the risk level of the patient. Therefore, patient outcomes were worse for patients with radiotherapy or a more frequent follow-up. Future research could explore ways to separate patient risk from past clinical decisions, like giving radiotherapy or planning follow-up. Using prospective studies might help reduce this bias and better assess the actual effect of these choices.

Furthermore, since the dataset was relatively small (367 patients after data cleaning), there were also not many patients experiencing a recurrence. The dataset distinguished between local recurrence, regional recurrence and distant metastasis. However, these were all grouped to ‘recurrence’, because the low number of incidences would not create reliable CPTs for the Bayesian network. For future research, it is recommended to research the different types of recurrence separately, preferably with a larger dataset. Risk factors could, for instance, be different for local recurrence and distant metastasis, while metastasis usually leads to lower survival chances.

The small size of the dataset also caused empty columns in the CPTs for survival in years 3 to 5 in the decision simulation layer. The empty cells were now set to 1 and normalised, but that does not give valid results. It is recommended to do a similar study with a larger dataset, also to get a better estimate of, for instance, the effect of different tumour histologies. While the data seemed to indicate an association between tumour histology and recurrence, the low incidences made it unreliable.

Additionally, since fitness was included as a node in the model, but it was not included as a variable in the dataset, we had to make estimations based on age. This does not accurately reflect real-world clinical fitness assessments. Due to the age estimation, only two-thirds of the possible input combinations from the model were represented in the dataset. It limits the model’s reliability for guiding follow-up strategies in those cases and highlights the need for datasets including fitness (or ECOG performance status) as a variable.

Another limitation lies with the definition of survival in the dataset. The survival is not disease-specific, meaning that patients could just as well have died from older age or in a car crash, but the data does not indicate this. Therefore, we are unsure if the found associations with death are completely true and reliable. It is recommended to use disease-specific survival or death in future research.

Moreover, once a patient had a recurrence, the dataset no longer contains information on the follow-up frequency after a recurrence. As a result, it is not possible to evaluate how follow-up strategies may have changed after recurrence, or how these strategies influenced outcomes beyond that point.

Both the variables age and diameter are continuous variables. However, to be able to use them in a Bayesian network, they should be changed into a categorical one. With this conversion, detailed information is lost through the averaging of groups.

5.4.3Clinical decision-making

An interesting observation was that tumour grade consistently seemed to predict survival better than the risk classification made by the STS expert. Even within the same risk group, patients with high-grade tumours had noticeably worse outcomes. This suggests that, despite the expert's broader assessment of risk, grade on its own remains a strong and maybe even more reliable indicator of prognosis. Future research could study the clinical decision-making of multiple STS experts to better understand what factors they weigh most heavily and whether this varies amongst different centres or countries.

Even though STS experts do currently seem to know that higher-follow-up frequencies do not improve survival chances (as appeared from the survey), they did give higher follow-up frequencies in the period between 2000 and 2020. More research could be done to find out why clinicians kept giving frequent follow-up while they know it does not necessarily improve survival chances.

Following the findings of multiple studies in literature (Blaye et al., 2019; Cheney et al., 2014; England et al., 2020; Fujiki et al., 2016), where patients were found to mostly detect recurrences themselves, together with our findings that higher follow-up frequency does not necessarily improve survival, raises the question whether more focus should be put on patient education. Future research could study the role of patient self-monitoring, physical exams, and tumour location in personalising follow-up strategies, as this could potentially reduce unnecessary imaging without compromising outcomes.

Lastly, this research does not directly assess the impact of the Bayesian decision-support tool on clinical outcomes. While the model is designed to provide structured decision support, its real-world effectiveness in improving follow-up strategies, patient well-being and hospital resource efficiency would require further validation through prospective studies.

6. Conclusion

Soft tissue sarcoma (STS) follow-up care currently relies on generalised guidelines that apply the same fixed follow-up schedules to all patients, not taking into account individual risk. This 'one-size-fits-all' approach can be problematic, as it could lead to unnecessary hospital visits for low-risk patients, while those at higher risk could potentially not be adequately monitored. At the same time, clinicians often seem to deviate from these guidelines, resulting in inconsistent follow-up care. The current system could be seen as inefficient and possibly not optimal for the patients. There are models to predict outcomes for STS patients, based on certain tumour-, patient- and treatment-characteristics, but none of these models connected follow-up strategy to the predictions.

This study addressed the problem by developing a Bayesian decision-support model that uses clinical data and expert input to estimate personalised recurrence risk. Based on historical data and expert input, tumour grade and surgical margin were identified as the most important risk factors for recurrence. Tumour grade appeared to be most strongly and consistently associated with recurrence and death, even more than the risk classification made by the STS expert. This suggests that tumour grade alone might already be a reliable basis for personalising follow-up strategies.

To explore how risk could be translated into follow-up decisions, two main versions of the model were built: the Extensive and the Simplified risk model. These models not only estimated the probability of recurrence, but also incorporated patient-specific consequences, which is an important factor often not included in guidelines. The models were compared to both the STS expert classification and tumour grade alone, and although they were based on similar clinical inputs, they often classified patients quite differently. This reflects the fact that risk is a perceptive concept. How we define and weigh risk can vary, and having different models allows clinicians to approach the decision from a different perspective. Notably, tumour grade alone proved to be one of the most consistent and useful predictors of survival for the majority of patients, who received below-guideline follow-up in the beginning of follow-up in the past. This underlines the potential value of even relatively simple data-informed tools.

No clear survival benefit was found for patients who had received more intensive follow-up. In fact, higher follow-up frequency was associated with worse survival outcomes. In the first two years of follow-up, patients with above-guideline frequency had higher death rates than both on- and below-guideline. Similarly, patients with on-guideline frequency had higher death rates than below-guideline. This challenges the assumption that more follow-up necessarily leads to better outcomes. While we cannot know what would have happened to these patients under a different follow-up schedule, these findings do raise critical questions about how follow-up resources are currently allocated. The aim of follow-up is to detect recurrences early and while higher frequency means that recurrences are, on average, found sooner, the results suggest that it might not necessarily improve survival. Meanwhile, the cost differences between follow-up intensities are substantial, especially in the first two years, where more intensive follow-up could cost up to 13 times more per survivor than less intensive schedules.

These results also open the door for broader conversations about smarter resource allocation. Imaging may be more beneficial for tumour sites where recurrences often appear asymptomatic, while other patients may benefit more from education and guidance on recognising signs of recurrence themselves. As resources are limited, these trade-offs deserve more attention in clinical practice.

However, some limitations should be considered. The model and data analysis were based on data from a single hospital, which may limit generalisability. Also, surgical margins were grouped simply as R0 versus R1/R2, while differences between R1 and R2 could be important but were not captured. Future research should explore more detailed margin classifications and include larger, more diverse datasets to strengthen the findings.

Altogether, the findings from this study suggest that follow-up care for STS patients could be made more efficient and more personalised. Instead of treating all patients using a 'one-size-fits-all' schedule following the current guidelines, clinicians should take into account patient-specific factors. The structured decision-support tool could help clinicians weigh the probability and consequence of recurrence in a more transparent way, aiming to support more effective use of hospital resources while aligning care with individual patient needs.

Bibliography

Acem, I., Smit, M. M., Verhoef, C., Van Houdt, W. J., Haas, R. L., Van Der Hage, J. A., Grünhagen, D. J., & Van De Sande, M. a. J. (2021). Management of soft tissue sarcomas in extremities: variation in treatment recommendations and surveillance according to specialty and continent. *Annals of Surgical Oncology*, *28*(12), 7923–7936. https://doi.org/10.1245/s10434-021-09946-4

Alamanda, V. K., Song, Y., & Holt, G. E. (2014). Effect of marital status on treatment and survival of extremity soft tissue sarcoma. *Annals of Oncology*, *25*(3), 725–729. https://doi.org/10.1093/annonc/mdt583

Alamanda, V. K., Song, Y., Schwartz, H. S., & Holt, G. E. (2015). Racial disparities in extremity Soft-Tissue sarcoma outcomes. *American Journal of Clinical Oncology*, *38*(6), 595–599. https://doi.org/10.1097/coc.0000000000000004

Baroudi, M. R., Ferguson, P. C., Wunder, J. S., Isler, M. H., Mottard, S., Werier, J. A., & Turcotte, R. E. (2014). Forearm soft tissue sarcoma: Tumors characteristics and oncologic outcomes following limb salvage surgery. *Journal of Surgical Oncology*, *110*(6), 676–681. https://doi.org/10.1002/jso.23686

BayesFusion, LLC. (n.d.). GeNIe Modeler [Software]. Retrieved from http://www.bayesfusion.com

Bernstein, K. D., Bos, S. A., Veld, J., Lozano-Calderon, S. A., Torriani, M., & Bredella, M. A. (2018). Body composition predictors of therapy response in patients with primary extremity soft tissue sarcomas. *Acta Radiologica*, *59*(4), 478–484. https://doi.org/10.1177/0284185117723370

Blay, J., Hindi, N., Bollard, J., Aguiar, S., Angel, M., Araya, B., Badilla, R., Bernabeu, D., Campos, F., Caro-Sánchez, C., Carvajal, B., Montoya, A. C., Casavilca-Zambrano, S., Castro-Oliden, V., Chacón, M., Clara, M., Collini, P., Genoroso, R. C., Costa, F., . . . Martin-Broto, J. (2022). SELNET clinical practice guidelines for soft tissue sarcoma and GIST. *Cancer Treatment Reviews*, *102*, 102312. https://doi.org/10.1016/j.ctrv.2021.102312

Blaye, C., Kind, M., Stoeckle, E., Brouste, V., Kantor, G., Loarer, F. L., Italiano, A., & Toulmonde, M. (2019). Local and metastatic relapse features in patients after a primary soft tissue sarcoma: Advocating for a Better-Tailored Follow-Up. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00559

Bonvalot, S., Levy, A., Terrier, P., Tzanis, D., Bellefqih, S., Cesne, A. L., & Péchoux, C. L. (2017). Primary extremity soft tissue sarcomas: Does local control impact survival? *Annals of Surgical Oncology*, *24*(1), 194–201. https://doi.org/10.1245/s10434-016-5462-2

Callegaro, D., Miceli, R., Bonvalot, S., Ferguson, P. C., Strauss, D. C., Van Praag, V. V., Levy, A., Griffin, A. M., Hayes, A. J., Stacchiotti, S., Pèchoux, C. L., Smith, M. J., Fiore, M., Tos, A. P. D., Smith, H. G., Catton, C., Szkandera, J., Leithner, A., Van De Sande, M. A., . . . Gronchi, A. (2019). Development and external validation of a dynamic prognostic nomogram for primary extremity soft tissue sarcoma survivors. *EClinicalMedicine*, *17*, 100215. https://doi.org/10.1016/j.eclinm.2019.11.008

Callegaro, D., Miceli, R., Bonvalot, S., Ferguson, P., Strauss, D. C., Levy, A., Griffin, A., Hayes, A. J., Stacchiotti, S., Pechoux, C. L., Smith, M. J., Fiore, M., Tos, A. P. D., Smith, H. G., Mariani, L., Wunder, J. S., Pollock, R. E., Casali, P. G., & Gronchi, A. (2016). Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: a retrospective analysis. *The Lancet Oncology*, *17*(5), 671–680. https://doi.org/10.1016/s1470-2045(16)00010-3

Centraal Bureau voor de Statistiek. (n.d.). *Overlijden*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/levensloop/verweduwen

Cheney, M. D., Giraud, C., Goldberg, S. I., Rosenthal, D. I., Hornicek, F. J., Choy, E., Mullen, J. T., Chen, Y., & DeLaney, T. F. (2014). MRI surveillance following treatment of extremity soft tissue sarcoma. *Journal of Surgical Oncology*, *109*(6), 593–596. https://doi.org/10.1002/jso.23541

Chou, Y., Liu, C., Chang, Y., King, K., Chen, P. C., Pan, C., Shen, S., Liu, Y., Lin, A. T., Chen, K., Shyr, Y., Lee, R., Chao, T., Yang, M., Chan, C., You, J., & Yen, C. (2016). Prognostic factors of primary resected retroperitoneal soft tissue sarcoma: Analysis from a single asian tertiary center and external validation of gronchi's nomogram. *Journal of Surgical Oncology*, *113*(4), 355–360. https://doi.org/10.1002/jso.24155

Cipriano, C. A., Jang, E., & Tyler, W. (2020). Sarcoma Surveillance: A review of current evidence and guidelines. *Journal of the American Academy of Orthopaedic Surgeons*, *28*(4), 145–156. https://doi.org/10.5435/jaaos-d-19-00002

Cowell, R. G., Lauritzen, S. L., David, A. P., Spiegelhalter, D. J., Nair, V., Lawless, J., & Jordan, M. (1999). Probabilistic networks and expert systems. In *Information science and statistics*. https://doi.org/10.1007/b97670

Daigeler, A., Zmarsly, I., Hirsch, T., Goertz, O., Steinau, H., Lehnhardt, M., & Harati, K. (2014). Long-term outcome after local recurrence of soft tissue sarcoma: a retrospective analysis of factors predictive of survival in 135 patients with locally recurrent soft tissue sarcoma. *British Journal of Cancer*, *110*(6), 1456–1464. https://doi.org/10.1038/bjc.2014.21

Damery, S., Biswas, M., Billingham, L., Barton, P., Al-Janabi, H., & Grimer, R. (2014). Patient preferences for clinical follow-up after primary treatment for soft tissue sarcoma: A cross-sectional survey and discrete choice experiment. *European Journal of Surgical Oncology*, *40*(12), 1655–1661. https://doi.org/10.1016/j.ejso.2014.04.020

Dammerer, D., Van Beeck, A., Schneeweiss, V., & Schwabegger, A. (2020). Follow-up strategies for primary extremity soft-tissue sarcoma in Adults: A Systematic Review of the Published literature. *In Vivo*, *34*(6), 3057–3068. https://doi.org/10.21873/invivo.12140

Davidson, D., Barr, R. D., Riad, S., Griffin, A. M., Chung, P. W., Catton, C. N., O'Sullivan, B., Ferguson, P. C., Davis, A. M., & Wunder, J. S. (2016). Health-related quality of life following treatment for extremity soft tissue sarcoma. *Journal of Surgical Oncology*, *114*(7), 821–827. https://doi.org/10.1002/jso.24424

England, P., Hong, Z., Rhea, L., Hirbe, A., McDonald, D., & Cipriano, C. (2020). Does advanced imaging have a role in detecting local recurrence of soft-tissue sarcoma? *Clinical Orthopaedics and Related Research*, *478*(12), 2812–2820. https://doi.org/10.1097/corr.0000000000001351

Ezuddin, N. S., Pretell-Mazzini, J., Yechieli, R. L., Kerr, D. A., Wilky, B. A., & Subhawong, T. K. (2018). Local recurrence of soft-tissue sarcoma: issues in imaging surveillance strategy. *Skeletal Radiology*, *47*(12), 1595–1606. https://doi.org/10.1007/s00256-018-2965-x

Ferré, A. D., Álvarez, R. Á., Herráez, A. C., Jurado, J. C., González, A. E., Martín-Broto, J., Marín, V. M., Vega, A. M., García, A. S., & Morales, C. V. (2021). SEOM Clinical Guideline of management of soft-

tissue sarcoma (2020). *Clinical & Translational Oncology*, *23*(5), 922–930. https://doi.org/10.1007/s12094-020-02534-0

Fujiki, M., Miyamoto, S., Kobayashi, E., Sakuraba, M., & Chuman, H. (2016). Early detection of local recurrence after soft tissue sarcoma resection and flap reconstruction. *International Orthopaedics*, *40*(9), 1975–1980. https://doi.org/10.1007/s00264-016-3219-y

Galy-Bernadoy, C., & Garrel, R. (2016). Head and neck soft-tissue sarcoma in adults. *European Annals of Otorhinolaryngology Head and Neck Diseases*, *133*(1), 37–42. https://doi.org/10.1016/j.anorl.2015.09.003

Gannon, N. P., Stemm, M. H., King, D. M., & Bedi, M. (2019). Pathologic necrosis following neoadjuvant radiotherapy or chemoradiotherapy is prognostic of poor survival in soft tissue sarcoma. *Journal of Cancer Research and Clinical Oncology*, *145*(5), 1321–1330. https://doi.org/10.1007/s00432-019-02885-4

Gerrand, C. H., Bell, R. S., Wunder, J. S., Kandel, R. A., O'Sullivan, B., Catton, C. N., Griffin, A. M., & Davis, A. M. (2003). The influence of anatomic location on outcome in patients with soft tissue sarcoma of the extremity. *Cancer*, *97*(2), 485–492. https://doi.org/10.1002/cncr.11076

Glasbey, J., Bundred, J., Tyler, R., Hunt, J., Tattersall, H., Gourevitch, D., Almond, L., Desai, A., & Ford, S. (2021). The impact of postoperative radiological surveillance intensity on disease free and overall survival from primary retroperitoneal, abdominal and pelvic soft-tissue sarcoma. *European Journal of Surgical Oncology*, *47*(7), 1771–1777. https://doi.org/10.1016/j.ejso.2021.01.021

Gronchi, A., Miah, A., Tos, A. D., Abecassis, N., Bajpai, J., Bauer, S., Biagini, R., Bielack, S., Blay, J., Bolle, S., Bonvalot, S., Boukovinas, I., Bovee, J., Boye, K., Brennan, B., Brodowicz, T., Buonadonna, A., De Álava, E., Del Muro, X., . . . Stacchiotti, S. (2021). Soft tissue and visceral sarcomas: ESMO–EURACAN–GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, *32*(11), 1348–1365. https://doi.org/10.1016/j.annonc.2021.07.006

Gu, H., Zhang, C., Guo, J., Yang, M., Zhong, H., Jin, W., Liu, Y., Gao, L., & Wei, R. (2020). Risk score based on expression of five novel genes predicts survival in soft tissue sarcoma. *Aging*, *12*(4), 3807–3827. https://doi.org/10.18632/aging.102847

Houdek, M. T., Hevesi, M., Griffin, A. M., Wunder, J. S., & Ferguson, P. C. (2019). Morbid obesity is associated with an increased risk of wound complications and infection after lower extremity soft-tissue sarcoma resection. *Journal of the American Academy of Orthopaedic Surgeons*, *27*(21), 807–815. https://doi.org/10.5435/jaaos-d-18-00536

IKNL. (2025a, January 27). *Cijfers over bot- en wekedelenkanker*. https://iknl.nl/kankersoorten/bot-en-wekedelenkanker/cijfers

IKNL. (2025b, January 27). *NKR cijfers*. https://nkr-cijfers.iknl.nl/viewer/incidentie-per-jaar

Italiano, A., Cesne, A. L., Mendiboure, J., Blay, J., Piperno-Neumann, S., Chevreau, C., Delcambre, C., Penel, N., Terrier, P., Ranchere-Vince, D., Lae, M., Guellec, S. L., Michels, J., Robin, Y. M., Bellera, C., & Bonvalot, S. (2014). Prognostic factors and impact of adjuvant treatments on local and metastatic relapse of soft-tissue sarcoma patients in the competing risks setting. *Cancer*, *120*(21), 3361–3369. https://doi.org/10.1002/cncr.28885

Kang, S., Kim, H., Kim, W., Kim, J. H., Kang, S. H., & Han, I. (2015). Comorbidity is Independently Associated with Poor Outcome in Extremity Soft Tissue Sarcoma. *Clinics in Orthopedic Surgery*, *7*(1), 120. https://doi.org/10.4055/cios.2015.7.1.120

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, *1*(1), 11–27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x

Kenis, C., Baitar, A., Decoster, L., De Grève, J., Lobelle, J., Flamaing, J., Milisen, K., & Wildiers, H. (2018). The added value of geriatric screening and assessment for predicting overall survival in older patients with cancer. *Cancer*, *124*(18), 3753–3763. https://doi.org/10.1002/cncr.31581

Kruiswijk, A. A., Kuhrij, L. S., Dorleijn, D. M. J., Van De Sande, M. a. J., Van Bodegom-Vos, L., & De Mheen, P. J. M. (2023). Follow-Up after Curative Surgical Treatment of Soft-Tissue Sarcoma for Early Detection of Recurrence: Which Patients Have More or Fewer Visits than Advised in Guidelines? *Cancers*, *15*(18), 4617. https://doi.org/10.3390/cancers15184617

Korb, K. B., & Nicholson, A. E. (2011). Bayesian Artificial Intelligence. In *CRC Press eBooks* (2nd ed.). https://doi.org/10.1201/b10391

Lazarides, A. L., Visgauss, J. D., Nussbaum, D. P., Green, C. L., Blazer, D. G., Brigman, B. E., & Eward, W. C. (2018). Race is an independent predictor of survival in patients with soft tissue sarcoma of the extremities. *BMC Cancer*, *18*(1). https://doi.org/10.1186/s12885-018-4397-3

Lenze, F., Pohlig, F., Knebel, C., Mühlhofer, H., Rechl, H., Pichler, T., Herschbach, P., Von Eisenhart-Rothe, R., & Lenze, U. (2019). Psychosocial distress in follow-up care – results of a tablet-based routine screening in 202 patients with sarcoma. *Anticancer Research*, *39*(6), 3159–3165. https://doi.org/10.21873/anticanres.13453

Li, R., Zhou, Q., Li, A., Zhang, H., & Lin, Z. (2020). A nomogram to predict metastasis of soft tissue sarcoma of the extremities. *Medicine*, *99*(21), e20165. https://doi.org/10.1097/md.0000000000020165

Lilenbaum, R. C., Cashy, J., Hensing, T. A., Young, S., & Cella, D. (2008). Prevalence of poor performance status in lung cancer patients: Implications for research. *Journal of Thoracic Oncology*, *3*(2), 125–129. https://doi.org/10.1097/jto.0b013e3181622c17

LUMC. (2025). *Passanten prijslijst DBC-zorgproducten en overige zorgproducten jaar 2025*. https://www.lumc.nl/siteassets/patientenzorg/uw-bezoek-aan-het-lumc/uw-ervaring-rechten-en-zorgkosten/zorgkosten/vergoedingen-en-facturen/241107-passantenprijslijst-per-1-1-2025.pdf

Maretty-Nielsen, K., Aggerholm-Pedersen, N., Safwat, A., Jørgensen, P. H., Hansen, B. H., Baerentzen, S., Pedersen, A. B., & Keller, J. (2014). Prognostic factors for local recurrence and mortality in adult soft tissue sarcoma of the extremities and trunk wall. *Acta Orthopaedica*, *85*(3), 323–332. https://doi.org/10.3109/17453674.2014.908341

Mullen, J. T., Hornicek, F. J., Harmon, D. C., Raskin, K. A., Chen, Y., Szymonifka, J., Yeap, B. Y., Choy, E., DeLaney, T. F., & Nielsen, G. P. (2014). Prognostic significance of treatment-induced pathologic necrosis in extremity and truncal soft tissue sarcoma after neoadjuvant chemoradiotherapy. *Cancer*, *120*(23), 3676–3682. https://doi.org/10.1002/cncr.28945

Noebauer-Huhmann, I., Vilanova, J. C., Papakonstantinou, O., Weber, M., Lalam, R. K., Nikodinovska, V. V., Sanal, H. T., Lecouvet, F. E., Navas, A., Martel-Villagrán, J., De Rooy, J. W. J., Fritz, J., Verstraete, K., Grieser, T., Szomolanyi, P., Chaudhary, S., Sconfienza, L. M., Tagliafico, A. S., Afonso, P. D., . . .

Vanhoenacker, F. M. (2024). Soft tissue tumor imaging in adults: European Society of Musculoskeletal Radiology–Guidelines 2024: imaging immediately after neoadjuvant therapy in soft tissue sarcoma, soft tissue tumor surveillance, and the role of interventional radiology. *European Radiology*. https://doi.org/10.1007/s00330-024-11242-0

Oken, M. M., Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., & Carbone, P. P. (1982). Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology*, *5*(6), 649–656. https://doi.org/10.1097/00000421-198212000-00014

Park, J. T., Roh, J., Kim, S., Cho, K., Choi, S., Nam, S. Y., & Kim, S. Y. (2015). Prognostic factors and oncological outcomes of 122 head and neck soft tissue sarcoma patients treated at a single institution. *Annals of Surgical Oncology*, *22*(1), 248–255. https://doi.org/10.1245/s10434-014-3870-8

Pasquali, S., Palmerini, E., Quagliuolo, V., Martin-Broto, J., Lopez-Pousa, A., Grignani, G., Brunello, A., Blay, J., Tendero, O., Diaz-Beveridge, R., Ferraresi, V., Lugowska, I., Infante, G., Braglia, L., Merlo, D. F., Fontana, V., Marchesi, E., Donati, D. M., Palassini, E., . . . Gronchi, A. (2022). Neoadjuvant chemotherapy in high-risk soft tissue sarcomas: A Sarculator-based risk stratification analysis of the ISG-STS 1001 randomized trial. *Cancer*, *128*(1), 85–93. https://doi.org/10.1002/cncr.33895

Patel, D. B., & Matcuk, G. R., Jr. (2018). Imaging of soft tissue sarcomas. *Chinese Clinical Oncology*, *7*(4), 35. https://doi.org/10.21037/cco.2018.07.06

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* Elsevier. https://doi.org/10.1016/c2009-0-27609-4

Penumarthy, N. L., Goldsby, R. E., Shiboski, S. C., Wustrack, R., Murphy, P., & Winestone, L. E. (2020). Insurance impacts survival for children, adolescents, and young adults with bone and soft tissue sarcomas. *Cancer Medicine*, *9*(3), 951–958. https://doi.org/10.1002/cam4.2739

Posch, F., Leitner, L., Bergovec, M., Bezan, A., Stotz, M., Gerger, A., Pichler, M., Stöger, H., Liegl-Atzwanger, B., Leithner, A., & Szkandera, J. (2017). Can multistate modeling of local recurrence, distant metastasis, and death improve the prediction of outcome in patients with soft tissue sarcomas? *Clinical Orthopaedics and Related Research*, *475*(5), 1427–1435. https://doi.org/10.1007/s11999-017-5232-x

Puri, A., Gulia, A., Hawaldar, R., Ranganathan, P., & Badwe, R. A. (2014). Does intensity of surveillance affect survival after surgery for sarcomas? results of a randomized noninferiority trial. *Clinical Orthopaedics and Related Research*, *472*(5), 1568–1575. https://doi.org/10.1007/s11999-013-3385-9

Romano, A., Triarico, S., Rinninella, E., Natale, L., Brizi, M. G., Cintoni, M., Raoul, P., Maurizi, P., Attinà, G., Mastrangelo, S., Gasbarrini, A., Mele, M. C., & Ruggiero, A. (2022). Clinical Impact of Nutritional Status and Sarcopenia in Pediatric Patients with Bone and Soft Tissue Sarcomas: A Pilot Retrospective Study (SarcoPed). *Nutrients*, *14*(2), 383. https://doi.org/10.3390/nu14020383

Rothermundt, C., Andreou, D., Blay, J., Brodowicz, T., Desar, I. M., Dileo, P., Gelderblom, H., Haas, R., Jakob, J., Jones, R. L., Judson, I., Kunz, W. G., Liegl-Atzwanger, B., Lindner, L. H., Messiou, C., Miah, A. B., Reichardt, P., Szkandera, J., Van Der Graaf, W. T., . . . Cerny, T. (2023). Controversies in the management of patients with soft tissue sarcoma: Recommendations of the Conference on State of Science in Sarcoma 2022. *European Journal of Cancer*, *180*, 158–179. https://doi.org/10.1016/j.ejca.2022.11.008

Royce, T. J., Punglia, R. S., Chen, A. B., Patel, S. A., Thornton, K. A., Raut, C. P., & Baldini, E. H. (2017). Cost–Effectiveness of surveillance for distant recurrence in extremity soft tissue sarcoma. *Annals of Surgical Oncology*, *24*(11), 3264–3270. https://doi.org/10.1245/s10434-017-5996-y

Rueten-Budde, A. J., Van Praag, V. M., Van De Sande, M. a. J., & Fiocco, M. (2021). External validation and adaptation of a dynamic prediction model for patients with high-grade extremity soft tissue sarcoma. *Journal of Surgical Oncology*, *123*(4), 1050–1056. https://doi.org/10.1002/jso.26337

Rueten-Budde, A., Van Praag, V., Van De Sande, M., Fiocco, M., Jeys, L. M., Laitinen, M. K., Pollock, R., Aston, W., Van Der Hage, J. A., Dijkstra, P. S., Ferguson, P. C., Griffin, A. M., Willeumier, J. J., Wunder, J. S., Styring, E., Posch, F., Zaikova, O., Maretty-Kongstad, K., Keller, J., . . . Haas, R. L. (2018). Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma. *Surgical Oncology*, *27*(4), 695–701. https://doi.org/10.1016/j.suronc.2018.09.003

Rutkowski, P., & Ługowska, I. (2014). Follow-up in soft tissue sarcomas. *Memo - Magazine of European Medical Oncology*, *7*(2), 92–96. https://doi.org/10.1007/s12254-014-0146-8

Saebye, C., Fugloe, H. M., Nymark, T., Safwat, A., Petersen, M. M., Baad-Hansen, T., Krarup-Hansen, A., & Keller, J. (2017). Factors associated with reduced functional outcome and quality of life in patients having limb-sparing surgery for soft tissue sarcomas – a national multicenter study of 128 patients. *Acta Oncologica*, *56*(2), 239–244. https://doi.org/10.1080/0284186x.2016.1268267

Salah, S., Lewin, J., Amir, E., & Razak, A. A. (2018). Tumor necrosis and clinical outcomes following neoadjuvant therapy in soft tissue sarcoma: A systematic review and meta-analysis. *Cancer Treatment Reviews*, *69*, 1–10. https://doi.org/10.1016/j.ctrv.2018.05.007

Sawamura, C., Matsumoto, S., Shimoji, T., Okawa, A., & Ae, K. (2014). How long should we follow patients with soft tissue sarcomas? *Clinical Orthopaedics and Related Research*, *472*(3), 842–848. https://doi.org/10.1007/s11999-013-3076-6

Shuman, A. G., Brennan, M. F., Palmer, F. L., Kuk, D., Moraco, N., Singer, S., Shah, J. P., & Patel, S. G. (2015). Soft tissue sarcoma of the head & neck: Nomogram validation and analysis of staging systems. *Journal of Surgical Oncology*, *111*(6), 690–695. https://doi.org/10.1002/jso.23868

Smartt, A. A., Jang, E. S., & Tyler, W. K. (2020). Is there an association between insurance status and survival and treatment of primary bone and extremity soft-tissue sarcomas? a SEER database study. *Clinical Orthopaedics and Related Research*, *478*(3), 527–536. https://doi.org/10.1097/corr.0000000000000889

Smith, H. G., Memos, N., Thomas, J. M., Smith, M. J. F., Strauss, D. C., & Hayes, A. J. (2016). Patterns of disease relapse in primary extremity soft-tissue sarcoma. *British Journal of Surgery*, *103*(11), 1487–1496. https://doi.org/10.1002/bjs.10227

Smolle, M. A., Van De Sande, M., Callegaro, D., Wunder, J., Hayes, A., Leitner, L., Bergovec, M., Tunn, P., Van Praag, V., Fiocco, M., Panotopoulos, J., Willegger, M., Windhager, R., Dijkstra, S. P. D., Van Houdt, W. J., Riedl, J. M., Stotz, M., Gerger, A., Pichler, M., . . . Szkandera, J. (2020). Individualizing Follow-Up Strategies in High-Grade Soft Tissue Sarcoma with Flexible Parametric Competing Risk Regression Models. *Cancers*, *12*(1), 47. https://doi.org/10.3390/cancers12010047

Spunt, S. L., Million, L., Chi, Y., Anderson, J., Tian, J., Hibbitts, E., Coffin, C., McCarville, M. B., Randall, R. L., Parham, D. M., Black, J. O., Kao, S. C., Hayes-Jordan, A., Wolden, S., Laurie, F., Speights, R., Kawashima, E., Skapek, S. X., Meyer, W., . . . Hawkins, D. S. (2020). A risk-based treatment strategy for non-rhabdomyosarcoma soft-tissue sarcomas in patients younger than 30 years (ARST0332): a

Children's Oncology Group prospective study. *The Lancet Oncology*, *21*(1), 145–161. https://doi.org/10.1016/s1470-2045(19)30672-2

Sugiura, H., Nishida, Y., Nakashima, H., Yamada, Y., Tsukushi, S., & Yamada, K. (2014). Surgical procedures and prognostic factors for local recurrence of soft tissue sarcomas. *Journal of Orthopaedic Science*, *19*(1), 141–149. https://doi.org/10.1007/s00776-013-0469-z

Szkandera, J., Gerger, A., Liegl-Atzwanger, B., Stotz, M., Samonigg, H., Ploner, F., Stojakovic, T., Leithner, A., & Pichler, M. (2014). Pre-Treatment anemia is a poor prognostic factor in soft tissue sarcoma patients. *PLoS ONE*, *9*(9), e107297. https://doi.org/10.1371/journal.pone.0107297

Tepper, S. C., Holten, A. K., Jeffe, D. B., England, P. H., Hong, Z. L., Pérez, M., Ghert, M., Hirbe, A. C., & Cipriano, C. A. (2022). Examining patient perspectives on sarcoma surveillance: The Sarcoma Surveillance Survey. *Surgical Oncology*, *45*, 101861. https://doi.org/10.1016/j.suronc.2022.101861

Toulmonde, M., Cesne, A. L., Mendiboure, J., Blay, J., Piperno-Neumann, S., Chevreau, C., Delcambre, C., Penel, N., Terrier, P., Ranchère-Vince, D., Lae, M., Guellec, S. L., Michels, J., Robin, Y., Bellera, C., & Italiano, A. (2014). Long-term recurrence of soft tissue sarcomas: Prognostic factors and implications for prolonged follow-up. *Cancer*, *120*(19), 3003–3006. https://doi.org/10.1002/cncr.28836

Trojani, M., Contesso, G., Coindre, J. M., Rouesse, J., Bui, N. B., De Mascarel, A., Goussot, J. F., David, M., Bonichon, F., & Lagarde, C. (1984). Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *International Journal of Cancer*, *33*(1), 37–42. https://doi.org/10.1002/ijc.2910330108

Tsagozis, P., Bauer, H. C., Styring, E., Trovik, C. S., Zaikova, O., & Brosjö, O. (2015). Prognostic factors and follow-up strategy for superficial soft-tissue sarcomas: Analysis of 622 surgically treated patients from the scandinavian sarcoma group register. *Journal of Surgical Oncology*, *111*(8), 951–956. https://doi.org/10.1002/jso.23927

Tseng, W. W., Amini, B., & Madewell, J. E. (2015). Follow-up of the soft tissue sarcoma patient. *Journal of Surgical Oncology*, *111*(5), 641–645. https://doi.org/10.1002/jso.23814

Van Ewijk, R., Schoot, R. A., Sparber-Sauer, M., Ter Horst, S. a. J., Jehanno, N., Borgwardt, L., De Keizer, B., Merks, J. H. M., De Luca, A., McHugh, K., Von Kalle, T., Schäfer, J. F., Van Rijn, R. R., Bouhamama, A., Coma, A., Di Paolo, P. L., Fajardo, R. D., Franzius, C., Giraudo, C., . . . Tolboom, N. (2021). European guideline for imaging in paediatric and adolescent rhabdomyosarcoma — joint statement by the European Paediatric Soft Tissue Sarcoma Study Group, the Cooperative Weichteilsarkom Studiengruppe and the Oncology Task Force of the European Society of Paediatric Radiology. *Pediatric Radiology*, *51*(10), 1940–1951. https://doi.org/10.1007/s00247-021-05081-0

Van IJzendoorn, D. G. P., Szuhai, K., Bruijn, I. H. B., Kostine, M., Kuijjer, M. L., & Bovée, J. V. M. G. (2019). Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Computational Biology*, *15*(2), e1006826. https://doi.org/10.1371/journal.pcbi.1006826

Van Praag, V. M., Rueten-Budde, A. J., Jeys, L. M., Laitinen, M. K., Pollock, R., Aston, W., Van Der Hage, J. A., Dijkstra, P. S., Ferguson, P. C., Griffin, A. M., Willeumier, J. J., Wunder, J. S., Van De Sande, M. A., & Fiocco, M. (2017). A prediction model for treatment decisions in high-grade extremity soft-tissue sarcomas: Personalised sarcoma care (PERSARC). *European Journal of Cancer*, *83*, 313–323. https://doi.org/10.1016/j.ejca.2017.06.032

Vaynrub, M., Taheri, N., Ahlmann, E. R., Yao, C., Fedenko, A. N., Allison, D. C., Chawla, S. P., & Menendez, L. R. (2015). Prognostic value of necrosis after neoadjuvant therapy for soft tissue sarcoma. *Journal of Surgical Oncology*, *111*(2), 152–157. https://doi.org/10.1002/jso.23775

Vodanovich, D. A., Spelman, T., May, D., Slavin, J., & Choong, P. F. M. (2019). Predicting the prognosis of undifferentiated pleomorphic soft tissue sarcoma: a 20-year experience of 266 cases. *ANZ Journal of Surgery*, *89*(9), 1045–1050. https://doi.org/10.1111/ans.15348

Von Mehren, M., Kane, J. M., Agulnik, M., Bui, M. M., Carr-Ascher, J., Choy, E., Connelly, M., Dry, S., Ganjoo, K. N., Gonzalez, R. J., Holder, A., Homsi, J., Keedy, V., Kelly, C. M., Kim, E., Liebner, D., McCarter, M., McGarry, S. V., Mesko, N. W., . . . Bergman, M. A. (2022). Soft Tissue Sarcoma, Version 2.2022, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*, *20*(7), 815–833. https://doi.org/10.6004/jnccn.2022.0035

Wang, X., Zheng, B., Good, W., King, J., & Chang, Y. (1998). Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics*, *54*(2), 115–126. https://doi.org/10.1016/s1386-5056(98)00174-9

Whitaker, J., Nessim, C., Almond, M., & Ford, S. J. (2023). Surveillance post surgery for retroperitoneal soft tissue sarcoma. *Current Oncology*, *30*(3), 2781–2791. https://doi.org/10.3390/curroncol30030211

Willeumier, J. J., Rueten-Budde, A. J., Jeys, L. M., Laitinen, M., Pollock, R., Aston, W., Dijkstra, P. D. S., Ferguson, P. C., Griffin, A. M., Wunder, J. S., Fiocco, M., & Van De Sande, M. a. J. (2017). Individualised risk assessment for local recurrence and distant metastases in a retrospective transatlantic cohort of 687 patients with high-grade soft tissue sarcomas of the extremities: a multistate model. *BMJ Open*, *7*(2), e012930. https://doi.org/10.1136/bmjopen-2016-012930

Yang, T., Chen, I. A., Yu, C., Yang, C., Chen, W., Lin, P., Pang, S., Huang, S. K., Chen, Z., Tseng, J., Lin, W., Tsai, Y., Yu, C., Chen, P., Cheong, I., Jiang, Y., Wu, C., Huang, H., Tai, T., . . . Yeh, H. (2025). Preoperative ECOG performance status as a predictor of outcomes in upper tract urothelial cancer surgery. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-95128-1

Zaidi, M. Y., Canter, R., & Cardona, K. (2018). Post-operative surveillance in retroperitoneal soft tissue sarcoma: The importance of tumor histology in guiding strategy. *Journal of Surgical Oncology*, *117*(1), 99–104. https://doi.org/10.1002/jso.24927

Zhang, S., Wang, W., Liu, Z., & Wang, Z. (2019). Marital status and survival in patients with soft tissue sarcoma: A population-based, propensity-matched study. *Cancer Medicine*, *8*(2), 465–479. https://doi.org/10.1002/cam4.1802

Zhao, J., & Yang, Y. (2011). Retrospective analysis of relative prognostic factors in stage III soft tissue sarcomas with standard treatments. *Zhonghua Wai Ke Za Zhi*, *49*(11), PMID: 22333414.

Appendix

A.1 Literature Review tables

For the three categories in which the literature was divided, overview tables are given below.

A.1.1 Follow-up Guidelines

Table 45: Overview of literature review within the category follow-up guidelines

Category	Reference	Goal of study	Method	Main findings	View on STS follow-up
Perspectives on current follow-up guidelines and strategies	Rutkowski & Ługowska (2014)	To outline current follow-up recommendations for STS based on recurrence risk and site	Review of multiple guidelines and studies	Risk of recurrence depends on tumour grade, histological subtype, tumour size, surgery, and time from initial treatment. Risk factors should be included in follow-up strategy. As low-grade tumours on easily accessible sites have small chance of relapse, chest X-ray is suggested every 6 months in first 3 years, afterwards annually. High grade sarcomas require routine physical examination and chest imaging. Assessing lymph nodes is only necessary in certain subtypes	Follow-up should be adapted on risks and site-specific.
	Tseng et al. (2015)	To review surveillance strategies for STS patients after treatment and assess their effectiveness	Literature review of guidelines, surveys, and studies on follow-up strategies and outcomes	Regular check-ups and chest X-rays are standard, but surveillance should be personalized. More research is needed on its impact on survival.	Surveillance is essential but should be tailored to recurrence risk. Better studies, including randomized trials, are needed to refine strategies.
	Ezuddin et al. (2018)	To provide an updated overview of local recurrence risk in STS and discuss strategies for imaging surveillance	Literature review examining recurrence patterns in relation to histological type, grade, tumour size, and margin status	Most commonly suggested strategy: MRI and/or clinical examination every 3–4 months (years 1–3), every 6 months until year 5, then annually up to 10 years. Adherence varied across studies. Mixed: Some studies showed that recurrence was mostly detected clinically (e.g., 30 out of 31 cases), while others reported added value of routine MRI. PET/CT and ultrasound are suggested for specific indications	Imaging rarely detects asymptomatic local recurrence. Clinical examination is often sufficient. MRI is recommended when the tumour site is not easily accessible. Surveillance should be personalized based on patient risk factors and tumour location
	Patel & Matcuk (2018)	To review current imaging techniques and their role in the management of soft tissue sarcomas	Narrative literature review	Imaging is essential in staging, treatment response monitoring, and recurrence detection. New imaging techniques (e.g., PET/MRI, DWI, DCE-MRI) show promise but are costly or not widely available. MRI remains the primary modality, especially for local recurrence, but CT, US, or PET-CT can be alternatives. Advanced techniques like DWI and DCE-MRI improve specificity. No universal imaging protocol exists.	Imaging is essential in STS follow-up, but no consensus exists on how best to use emerging technologies. There is no guideline that is universally accepted and evidence-based. Due to costliness and broad diversity of sarcomas, imaging strategies differ a lot
	Zaidi et al. (2018)	To summarize recurrence patterns in retroperitoneal sarcomas (RPS) and	Review of international guidelines on Retroperitoneal sarcoma	Recurrence usually detected through routine imaging, as symptoms are often non-specific or late-stage. Surveillance intervals of current guidelines vary by tumour grade.	Recommends a risk-based and tumour-specific approach to surveillance rather than an approach based on triggers from patient

	propose histology-based surveillance		Recommendation: For high-grade/aggressive tumours: clinical assessment and CT of chest/abdomen/pelvis every 3–4 months for 2 years, then every 6 months for 3 years, then annually. For indolent tumours: clinical assessment and abdominal/pelvic CT every 6 months for 3 years, then annually just a clinical assessment.	symptoms. Emphasizes early detection for improved salvage opportunities.
Cipriano et al. (2020)	To explore current surveillance strategies for STS patients	A review of existing guidelines and literature on surveillance methods, with focus on possibility for individualised follow-up care.	Follow-up should be more intense in the first few years, with physical exams for local recurrence and imaging for pulmonary metastases. Extrapulmonary metastases are rare.	Follow-up should be individualised and based on sarcoma type and risk
Dammerer et al. (2020)	To evaluate follow-up strategies for primary eSTS in adults	Systematic review of literature, with 78 articles	No clear consensus on follow-up schedules. A variety of imaging methods are used, including chest X-rays, CT and MRI. Follow-up intervals vary, but a 6-month interval was suggested	Follow-up should be individualised based on tumour characteristics and patient factors. Calls for further investigation and standardization of individualised follow-up.
Spunt et al., 2020	To develop a risk-based treatment strategy for young patients with non-rhabdomyosarcoma soft-tissue sarcomas (NRSTS) and assess it in the context of risk-adapted therapy	A prospective study (ARST0332) enrolling 529 patients under 30 with non-rhabdomyosarcoma STS. Patients were categorized into low, intermediate, or high-risk groups based on tumour grade, tumour size, resection potential, and extent of disease. Treatment options included surgery alone, radiotherapy, chemoradiotherapy, and neoadjuvant chemoradiotherapy. Follow-up outcomes focused on event-free survival (EFS) and overall survival (OS).	Low-risk patients had high 5-year EFS (88.9%) and OS (96.2%). Intermediate-risk had 65.0% EFS and 79.2% OS, while high-risk had 21.2% EFS and 35.5 OS. The risk level is a predictor for EFS and OS. Adjuvant therapy is not necessary for low-risk patients.	The study emphasizes the importance of risk stratification for determining follow-up treatment. Low-risk patients may avoid adjuvant therapy and can be cured with surgery alone. Follow-up for high-risk and intermediate-risk patients is more complex, with a need for aggressive treatment and close monitoring.
Acem et al. (2021)	To analyse clinical decision-making and surveillance strategies for patients with eSTS	A web-based survey with 396 respondents from various continents and specialties	It found significant variation in perioperative treatment and follow-up strategies, influenced by specialty and continent.	Follow-up strategies show variation in imaging methods and frequency across regions.
Van Ewijk et al. (2021)	To outline a European imaging guideline for diagnosis, staging, response evaluation, and follow-up in paediatric/adolescent rhabdomyosarcoma	Consensus-based guideline developed by members of 3 expert committees, based on recent clinical and imaging insights	Recommendation: MRI of primary tumour site (or ultrasound if superficial) and chest X-ray every 4 months for 2 years. After 2 years, imaging only if symptoms arise. Over half of recurrences detected via clinical symptoms (pain or palpable mass). Asymptomatic recurrence found by imaging did not show improved survival. Estimated 178 MRIs and 178 chest X-rays needed to detect one asymptomatic recurrence.	Imaging-detected asymptomatic relapses did not improve survival. Clinical symptoms are often the first indication of recurrence. Routine long-term imaging has limited benefit and should be balanced against risks (e.g. radiation, stress). Prospective or randomized studies are lacking.

	Blay et al. (2022)	To outline an STS and GIST guideline for Latin American countries, with the Sarcoma European Latin-American Network (SELNET)	Review of multiple guidelines and studies.	Physical exam for primary site. MRI for limb/trunk/pelvis. CT for abdomen/lung. Chest X-ray or CT for metastasis, abdominal CT for some subtypes	Emphasizes site-specific imaging; chest imaging for metastasis
	Rothermundt et al. (2023)	To identify controversies in diagnosis, treatment, and follow-up of STS and provide recommendations for clinical practice and research	Delphi process with 62 European STS experts	16 items reached strong consensus; 30 additional items had >75% agreement, but several controversial topics remained unresolved	Follow-up should be individualized based on subtype, location, recurrence risk, and treatment. 82% consensus for local and metastatic surveillance every 3 months for the first 2 years, every 6 months until year 5, then annually. MRI is preferred for local surveillance in high-risk patients; CT of chest/abdomen commonly used. Survivorship care ideally managed by specialized clinics or oncologists.
	Whitaker et al. (2023)	To investigate optimal follow-up strategies for retroperitoneal sarcoma (RPS) patients	Literature review of current surveillance guidelines, retrospective data and ongoing research	Surveillance strategies vary a lot and evidence for intensive follow-up is limited. RPS patients would benefit from individualised follow-up based on risk stratification, but more research is needed.	Surveillance should be individualised based on risk stratification.
	Kruiswijk et al. (2024)	To evaluate adherence to follow-up guidelines and factors affecting follow-up frequency in STS	Retrospective cohort study of 394 STS patients surgically treated between 2000-2020, analysing radiological exams and follow-up visits compared to guidelines	Only 24% of patients received the advised follow-up visits in the first year. Younger patients and those with high-grade tumours had more visits. Many patients without recurrence received more visits than advised.	Follow-up should consider individual patient risk factors (e.g., tumour grade, age). Clinicians might need support in estimating recurrence risk, e.g. through prediction models.
	Noebauer-Huhmann (2024)	To update ESSR guidelines on soft tissue tumour imaging with new advancements and classifications	Delphi method with 46 radiologists across 12 countries.	Consensus reached on key points: post-neoadjuvant imaging is crucial, surveillance should be grade/type-dependent, MRI preferred for loco-regional monitoring, chest CT for metastasis, and interventional radiology for complex cases.	Advocates for standardized imaging and surveillance, to support radiologists' decisions. Standardization can improve the consistency of patient exams and help create databases for developing personalized strategies.
Effectiveness of follow-up strategies	Cheney et al. (2014)	To assess the utility of surveillance MRI in detecting asymptomatic local recurrences after treatment of extremity STS	Retrospective cohort study of 168 adult patients with extremity STS treated with limb-sparing surgery and radiotherapy; analysis of follow-up MRIs per NCCN guidelines, focusing on detection method of local recurrence	Patients had a median follow-up time of 4.7 years. Surveillance MRI used in 68% (114/168), 32% no MRI. 11/114 patients had LR. 1/11 of LR was detected by surveillance MRI, 9/11 were detected through symptoms reported by patients.	Routine MRI rarely detects asymptomatic LR. Selective use is recommended for tumour sites that are hard to assess physically
	Puri et al. (2014)	To compare intensive vs. less frequent follow-up after sarcoma surgery.	Randomized trial with 500 patients, comparing follow-up intervals (3 vs. 6 months) and imaging (X-ray vs. CT).	Survival rates were similar across groups, suggesting less frequent follow-up is sufficient.	A less intensive follow-up strategy with fewer visits and more cost-effective imaging appears to be sufficient without compromising survival. Patients play a crucial role in detecting local recurrences themselves.

	Sawamura et al. (2014)	To determine optimal follow-up duration by analysing timing of local recurrence and metastasis after STS surgery	Retrospective study of 867 STS patients. LR and DM were analysed by timing, grade and size	90% of local recurrences occurred within 7.1 years, and 90% of metastases within 4.2 years. Few events occurred after 10 years. High-grade tumours tend to show higher recurrence/metastasis early, while low-grade tumours have a more constant rate over time.	Current follow-up guidelines lack evidence. Surveillance after 10 years of the initial treatment shows very little benefit and might not be necessary, especially for high-grade tumours.
	Fujiki et al. (2016)	To evaluate patterns and detection methods of local recurrence after STS resection and flap reconstruction	Retrospective cohort study of 229 patients	33/229 (14.4%) developed local recurrence; 23 cases were found by clinical exam and were superficial, 6 were also found by clinical exam but were deep and 4 were detected by routine imaging only.	Combined clinical exam and imaging is recommended for effective surveillance after resection and flap reconstruction. Imaging is especially important to detect deep recurrences not clinically visible.
	Royce et al. (2017)	To evaluate cost-effectiveness of different imaging modalities for distant recurrence (DR) in stage II-III STS	A Markov model simulates outcomes using four surveillance strategies (WW, CXR, CCT, PET/CT) with data from literature and Medicare claims.	CXR is the most cost-effective, CCT is the most effective but costly, and PET/CT is never cost-effective.	Surveillance should be risk-adjusted, with CXR or infrequent CCT for low-risk patients and CCT for high-risk patients.
	Blaye et al. (2019)	To investigate recurrence features in STS patients to improve follow-up guidelines	Retrospective analysis of 359 patients with first local or metastatic recurrence after initial treatment at Institut Bergonie	Local recurrences were mostly self-detected, while metastatic recurrences were often found during planned imaging, with CT scans being more effective than X-rays	Patient education for self-examination is crucial for local recurrence detection, while optimized imaging strategies may improve metastatic recurrence outcomes. Modelling studies are needed to better predict tumour biology and optimize STS follow-up.
	England et al. (2020)	To assess the role of advanced imaging in detecting local recurrences of soft-tissue sarcoma	Retrospective cohort study of 366 STS patients over 20 years, comparing imaging vs clinical detection	66% of local recurrences were first detected by clinical signs/symptoms, 34% by imaging. No significant difference in tumour size at detection between the two methods. No clear factors (e.g. tumour size, depth, radiation) associated with detection method. Imaging still detected a substantial proportion of recurrences that were otherwise asymptomatic.	It supports current surveillance guidelines including advanced imaging, as it contributes meaningfully to early detection. However, it emphasizes the need for further study on optimizing imaging frequency, cost-effectiveness, and patient-centred factors.
	Glasbey et al. (2021)	To evaluate the impact of radiological surveillance on survival after retroperitoneal sarcoma resection	Retrospective cohort study at a high-volume center, comparing ESMO-compliant vs non-compliant follow-up intensity.	No survival benefit from more intense surveillance and thus earlier recurrence detection.	Intense surveillance doesn’t improve survival, but it could detect recurrences earlier.

A.1.2 Predictive Models to Support Decision-Making

The articles within this category were divided into five subcategories: Sarculator, PERSARC, Nomograms, Machine learning and Multistate model to reflect the different models types currently available.

Table 46: Overview of literature review within the category predictive models

Category	Reference	Goal of study	Model type	Method	Main findings	Risk factors mentioned
Sarculator	Callegaro et al., (2016)	To develop and externally validate two nomograms predicting overall survival and distant metastases in adults after resection of localized eSTS	Sarculator (nomograms)	Retrospective analysis of 1452 patients (development cohort) and 3 external validation cohorts (n=2300); Development of nomograms using Cox multivariable model for OS and Fine & Gray multivariable model for DM. Backward variable selection and external validation in three independent cohorts	Developed two nomograms based on age (not included for DM), tumour size, grade, and histological subtype showed good performance and were validated across French, Canadian and UK cohorts. They seem like reliable tools for personalized prognosis and decision-making after STS resection	Age, tumour size, grade and histological subtype for OS, and everything except age for DM
	Callegaro et al. (2019)	Develop and validate a dynamic nomogram to predict 5-year overall survival during follow-up of eSTS patients	Sarculator (dynamic nomogram)	Multicentre retrospective cohort (n=3740 for development, n=893 for validation). Landmark analysis and multivariable Cox regression for the development of the dynamic nomogram. Variable selection was done using Akaike Information Criterion	Nomogram allows for updated survival prediction during the first 3 years post-surgery (it can incorporate time-dependant interactions) and had strong calibration and discrimination	Age, tumour size (with time interaction), FNCLCC grade (with time interaction), histology, local recurrence (as event), distant metastasis (as event)
	Pasquali et al. (2022)	To assess whether patients with high-risk STS (according to Sarculator-predicted OS) benefit from neoadjuvant chemotherapy	Risk stratification using nomogram (Sarculator)	Post-hoc analysis of the ISG-STS 1001 randomised clinical trial. Patients were stratified by predicted 10-year OS (<60% = high-risk, ≥60% = low-risk) using the Sarculator nomogram. Comparison of observed OS across risk groups and treatment groups: anthracycline–ifosfamide (AI) vs histology-tailored (HT)	High-risk patients treated with AI chemotherapy had better observed survival than predicted, suggesting benefit of neoadjuvant AI in high-risk STS. No significant benefit in low-risk group. Sarculator enabled effective stratification to identify patients more likely to benefit.	Predicted 10-year OS from Sarculator; factors used in Sarculator include age, size, grade, histology, site, and margins (implied from model)

PERSARC	Van Praag et al. (2017)	To develop and internally validate a prediction model (PERSARC) for OS and LR in patients with high-grade ESTS, incorporating treatment characteristics (e.g. radiotherapy)	PERSARC (Risk prediction model)	Retrospective cohort of 766 patients from 5 international centres; Multivariate Cox regression for OS and Fine & Gray model for local recurrence; Internal cross-validation and calibration used for performance evaluation	The model predicts 3-, 5-, and 10-year OS and local recurrence risk. Radiotherapy significantly improved outcomes. The model showed good internal validation and can support personalized shared decision-making	Older age and larger tumours negatively influence OS. Patients who had radiotherapy had better outcomes on cumulative incidence of LR (CILR) and OS. Neoadjuvant RT had significant lower risk of CILR. Larger margins have lower CILR.
	Willeumier et al. (2017)	Assess risk of local recurrence and distant metastases/death using individualised multistate model in high-grade eSTS	PERSARC (multistate model)	Retrospective multicentre cohort (n=687). A multistate survival analysis with hazard ratios for transitions from disease-free to local recurrence or metastasis/death	Tumour size and radiotherapy are prognostic factors for LR. Neoadjuvant radiotherapy and wide surgical margins reduce risk of local recurrence. Tumour size and age predict distant metastasis/death. The model accounts for individual baseline and treatment effects	Tumour size, radiotherapy, surgical margin, age
	Rueten-Budde et al. (2018)	To develop a dynamic prediction model that updates overall survival probabilities during follow-up for patients with high-grade eSTS (dynamic PERSARC model)	PERSARC (dynamic risk prediction model)	Landmark analysis and Cox proportional hazards model using both baseline and time-dependent covariates	The model provides individualized, updated 5-year survival estimates. Surgical margin and histology had time-varying effects; local recurrence and distant metastases had strong prognostic impact and were integrated into updated predictions.	Surgical margin, histologic subtype, local recurrence, distant metastases
	Smolle et al. (2020)	To develop two flexible parametric competing risk regression models (FPCRRMs) for predicting local recurrence (LR) and distant metastasis (DM) risks in high-grade eSTS patients after curative surgery	PERSARC (flexible parametric competing risk regression models)	A retrospective cohort study of 3,016 eSTS patients was conducted, and FPCRRMs were applied using factors like tumour size and grade.	LR and DM risks vary significantly depending on individual factors, with good in- and external validation. Individualised follow-up strategies should be created based on patient risk profiles, instead of a “one-fits-all approach”	Gender, tumour size, histological subtype, surgical margin, (neo)adjuvant RT, adjuvant CT and grade (time-dependent) for LR (age and neoadjuvant CT were excluded). Gender, tumour grade, tumour size, surgical margin, histological

						subtype and neoadjuvant radiotherapy for DM (age, adjuvant RT and CT were excluded).
	Rueten-Budde et al. (2021)	To update and externally validate the dynamic prediction model (PERSARC) for overall survival in patients with high-grade eSTS	PERSARC (Dynamic prediction model)	Data from 3826 patients were used to update the previously developed PERSARC model, including grade as a new covariate. The model was externally validated using data from 1111 patients at a tertiary centre. The model uses patient-, tumour-, and treatment-related variables including time-dependent events (e.g. LR, DM). Calibration plots and dynamic C-indices at various timepoints post-surgery were calculated to assess predictive performance.	The updated model showed good calibration and strong discriminative ability across follow-up timepoints. The PERSARC model can accurately predict 5-year survival probabilities at multiple moments during follow-up and is implemented in a clinical app.	Age, tumour size, tumour depth, tumour grade, (neo)adjuvant radiotherapy, LR, DM, histological subtype, margin status
Nomograms	Shuman et al. (2015)	To assess the performance of general and site-specific staging systems, and to validate a nomogram for STS of the head and neck	Postoperative nomogram	A retrospective cohort study (n=319) was used for a survival prediction performance. Kaplan-Meier, Cox univariate regression model and concordance indices were used	The staging systems can both be used for outcome prediction. The nomogram was even better in predicting outcomes than the staging systems.	Tumour, size, depth of invasion (nodal disease or DM) and grade
	Li et al. (2020)	To develop a nomogram to predict the presence of distant metastasis at initial presentation in patients with eSTS	Nomogram	Retrospective analysis of 3884 patients with STS of the extremities or trunk from the SEER database (2010–2015). Patients were randomly assigned to a training and validation set (2:1 ratio). Univariate and binary logistic regression analyses identified significant predictors of metastasis. A nomogram was constructed using these predictors and validated using calibration plots, C-index, and decision curve analysis	The nomogram showed good discrimination and calibration. The model is clinically useful for assessing risk of DM at diagnosis and may support decision-making in initial patient evaluation	Age, histological subtype, tumour size, tumour grade, tumour depth, primary site of tumour

	Gu et al. (2020)	To identify gene expression-based biomarkers and build a risk score model to predict survival in STS	Gene expression-based nomogram	Gene expression data from TCGA and GTEx databases. Differentially expressed genes (DEGs) were identified and analysed via univariate Cox regression. Significant genes were selected using LASSO regression, followed by multivariate Cox regression to construct a 5-gene risk score model. Patients were stratified into high- and low-risk groups based on the median score. A nomogram was developed incorporating the risk score along with age, metastasis status, and margin status. Model performance was validated using test datasets, survival analyses, and ROC curves.	A 5-gene signature was used to calculate a risk score predicting OS. The score was an independent prognostic factor, and negatively associated with immune cell infiltration. A nomogram combining risk score with age, metastasis, and margin status showed good predictive accuracy for OS.	Gene expression, age, DM and margin status
Machine learning	van IJzendoorn et al. (2019)	To identify new diagnostic/prognostic biomarkers and therapeutic targets for STS using machine learning and gene expression data	Machine learning-based gene expression analysis	Used gene expression data from TCGA, GTEx, and the French Sarcoma Group. Applied unsupervised clustering (t-SNE), deep neural networks, random forest for diagnostic markers, k-nearest neighbour for prognostic gene prediction, and regulatory network reconstruction for therapeutic target discovery	Identified molecularly overlapping subtypes of STS and new diagnostic/prognostic markers. Machine learning approaches distinguished synovial sarcoma from MPNST, predicted disease outcome, and identified HDAC inhibitors as potential therapy for some subtypes	Genes
Multistate model	Posch et al. (2017)	To evaluate how baseline risk factors, LR and DM impact LR, DM and survival in patients with localized STS using multistate modelling	Multistate model	Retrospective cohort study (n=443) making use of multistate modelling	Local recurrence significantly increased the risk of distant metastasis and death. Distant metastasis was the strongest predictor of mortality. High-grade tumours had increased risk of recurrence and death. Multistate models offer a dynamic and individualized approach to predict outcomes in STS.	Tumour grade, LR, DM

A.1.3 Risk factors for Recurrence

For every study in the table below, the 'focus' column indicates the specific focus of the study and what the main measurements were. For the columns containing risk factors (age, tumour size, etc.), a plus sign (+) before a factor state means that it has a beneficial association with the outcome(s) in the parentheses behind it. A minus sign (-), however, means that it has a disadvantageous association with the outcome(s) in the parentheses behind it. 'No' means that the study mentioned that the factor had no association with the outcome, and a 'yes' means that it did have an association, but no direction was given.

Table 47: Overview of literature review within the category risk factors

Study	Focus	Age	Tumour size	Histologic grade	Tumour site	Surgical margin	Radio-therapy	Chemo-therapy	Histological subtype	Tumour depth	Notes about other risk factors
Daigeler et al. (2014)	Locally recurrent STS (PRS)	No (PRS)		+ G1 (PRS) - G2 (PRS) - G3 (PRS)	+ Extremity (PRS) - Trunk (PRS)	+ R0 (after last resection) (PRS) - R1/R2 (PRS)		No (PRS)			
Bonvalot et al. (2017)	OS	- Over 60 (OS)	No (OS)	- G3 (OS)		No (OS)			- Leiomyo-sarcoma (OS)		Male gender had worse OS LR was no risk factor for OS
Maretty-Nielsen et al. (2014)	LR and DSM	- Over 55 (LR) - Over 45 (DSM)	+ <4 cm (LR, DSM)	- G2 (DSM) - G3 (LR and DSM)	- Trunk (DSM) - Lower extremity (DSM)	- R1/R2 (LR, DSM)	+ Radiotherapy (LR, DSM)			No (LR, DSM)	Duration of symptoms: + 3 < months < 24 (LR) - 15 < months < 30 (DSM)
Italiano et al. (2014)	LR and DM	- Over 55 (LR)	No (LR) - >5 cm (DM)	+ G1 (LR, DM) - G2 (LR, DM) - G3 (LR, DM)	+ Limb (LR) - Head and neck (LR) - Trunk wall (LR) - Internal trunk (LR)		+ Adjuvant (LR)	+ Adjuvant (DM) Adj. CTX only beneficial in grade 3	Yes (LR, DM)	No (LR) - Deep (DM)	
Vodanovich et al. (2019)	Undifferentiated pleomorphic STS LR, DM and OS	- Over 70 (LR, OS)	- >5 cm (DM)		+ Upper extremity (OS)	- R1/R2 (LR)	+ Adjuvant (DM)				- Metastatic disease present at diagnosis (DM, OS)

Smith et al. (2016)	Extremity STS LR, DM, DSS	- Over 74 (LR, DSS)	- 6 to 14 cm (DSS) - >14 cm (DM, DSS)	+ G1 (LR, DM, DSS) + G2 (vs G3) (DM, DSS)		- R1 (LR) (R2 was not looked at)	+ Radiotherapy (LR)		No (LR, DM, DSS) Undifferentiated pleomorphic sarcoma had worse outcomes	- Deep (LR)	
Galy-Bernadoy & Garrel (2016)	Head and neck STS in adults OS		Yes (OS)	Yes (OS)	- Head and neck (OS)	Yes (OS)					Yes (OS)
Toulmonde et al. (2014)	Long-term recurrence Late LR and DM		- >10 cm (late LR)	- G2 and G3 (late DM)	- Internal trunk (late LR)						
Chou et al. (2016)	Retroperitoneal sarcoma (LR, DM, OS)			+ G1 (vs G3) (DM, OS)		+ R0 (LR)					
Baroudi et al. (2014)	Forearm STS and limb salvage surgery		+ <4 cm (OS) - >4 cm (DM)	+ G1 (OS)							- extra-compartmental site (DM) -extramuscular tumour (DM) - virgin tumor (DM)
Sugiura et al. (2014)	Locally recurrent STS (LR, DM)		- >10 cm (DM)	- high-grade (DM)	- Upper extremity (LR) - Trunk (LR)	- <1 cm (LR)				- Deep (LR, DM)	LR was a risk factor for DM
Tsagozis et al. (2015)	Superficial STS (LR, DM, OS)	- Over 55 (LR, DM, OS)	- >5 cm (LR, DM, OS)	- high-grade (LR)							
Park et al. (2015)	Head and Neck STS (LRC, DSS, OS)	- Over 60 (DSS, OS)	- >10 cm (LRC)	+ G1 (DSS, OS)							- overall stage (LRC, DSS, OS) - Nodal metastasis (LRC, DSS, OS)

Risk Factors (Less Often Discussed)

Necrosis

There are several studies about the prognostic significance of tumour necrosis following neoadjuvant therapy in STS, with varying conclusions on its impact on recurrence and survival. Mullen et al. (2014) found no significant correlation between the extent of pathologic necrosis and local control, disease-specific survival, or overall survival in STS patients treated with neoadjuvant chemoradiotherapy. Similarly, Vaynrub et al. (2015) reported that tumour necrosis of 90% or more was associated with improved disease-free survival in univariate analysis, but this did not get statistical significance in multivariate analysis, where age and tumour were the key independent predictors.

The role of tumour necrosis after neoadjuvant therapy in STS remains complex and appears to be dependent on its interpretation. Salah et al. (2018) found that necrosis of 90% or less was associated with higher risks of recurrence and death. This suggests that extensive necrosis may serve as a positive prognostic marker of treatment response. In contrast, Gannon et al. (2019) reported that higher levels of necrosis were linked to worse outcomes. They state that necrosis may reflect more aggressive tumour biology rather than treatment efficacy.

Demographic and Socioeconomic factors

The impact of race has been observed by two studies, mainly focusing on black or African American people. Both Alamanda et al. (2015) and Lazarides et al. (2018) found that African Americans with STS tend to come in with larger tumours and have worse overall survival compared to white or Asian people. However, both studies also note that this difference is not entirely explained by the race itself, but probably reflects differences in socioeconomic status and access to care. Furthermore, marital status has been found to be an independent prognostic factor, with unmarried patients showing worse overall and cancer-specific survival compared to married patients (Alamanda et al., 2014; Zhang et al., 2019). Research by Penumarthy et al. (2020) and Smartt et al. (2020) point out that insurance status, which is often related to income level, plays a role in survival. Patients with Medicaid, non-private, insurance had decreased disease-specific survival (Smartt et al., 2020) and low income patients had decreased overall survival irrespective of disease stage at diagnosis.

Body composition

Two studies highlight the prognostic value of body composition in sarcoma patients. Research by Romano et al. (2022) suggests that decreased prognostic nutritional index and more than 25% reduction in total psoas muscle area over 12 months are significantly associated with poorer overall survival in pediatric sarcoma patients. Sarcopenia, which is low muscle mass or muscle function, at diagnosis was not significantly associated with overall survival. Bernstein et al. (2018) found that in adults with extremity soft tissue sarcomas, higher subcutaneous adipose tissue and lower psoas muscle attenuation on CT were independently associated with increased tumour recurrence.

Comorbidity

Kang et al. (2015) retrospectively analysed the prognostic value of comorbidity (including, amongst other, cardiovascular diseases, diabetes and malignancies) for patients with STS in the extremities. They found that its presence independently relates to poor local recurrence-free survival and disease-specific survival. Whereas Kang et al. (2015) did not research morbid obesity, Houdek et al. (2019) did. They found that in patients with lower extremity soft tissue sarcoma, morbid obesity was significantly associated with higher rates of wound complications and infections after surgery, but showed no impact on local recurrence or overall oncologic outcomes. Furthermore, Szkandera et al. (2014) studied the impact of anemia. They found that low pre-operative hemoglobin levels were independently associated with significantly worse cancer-specific and overall survival, suggesting that hemoglobin may serve as a simple and cost-effective prognostic marker to identify high-risk patients.

A.2 Survey setup

Opening statement

Invitation to Participate

You are invited to participate in a research study titled "Improving Decisions About the Number of Post-Treatment Follow-Up Visits for Soft Tissue Sarcoma (STS) Patients Using Historical Data." This study is conducted by Nina van Staalduine at TU Delft, in collaboration with Leiden University Medical Centre (LUMC), as part of a Master's thesis.

Why Is This Study Important?

The current STS follow-up guidelines lack strong supporting evidence and do not account for individual patient risk factors. As a result, clinicians often deviate from these guidelines, with a previous study showing that this occurs in over 70% of cases.
This survey seeks your expert opinion to further refine a tool that aims to support decisions about STS follow-up scheduling. Your participation will help to ensure that this tool uses all relevant evidence regarding factors that influence such decisions.

What Should You Know Before Continuing?

- Duration: The survey contains 10 questions and will take approximately 15 minutes to complete.
- Voluntary Participation: Your participation is entirely voluntary, and you may withdraw at any time.
- Optional Questions: You are free to skip any questions.

We deeply appreciate your time and contribution to this research.

Privacy & Data Protection

While every effort will be made to protect your privacy, as with any online study, there is always a small risk of data breaches. To minimize this risk:

- Confidentiality: Your responses will remain confidential, and no personal data will be published.
- Anonymization: Results will be aggregated and anonymized before publication.
- Storage & Compliance: The raw survey data will be securely stored at TU Delft in accordance with GDPR regulations, accessible only by the research team.

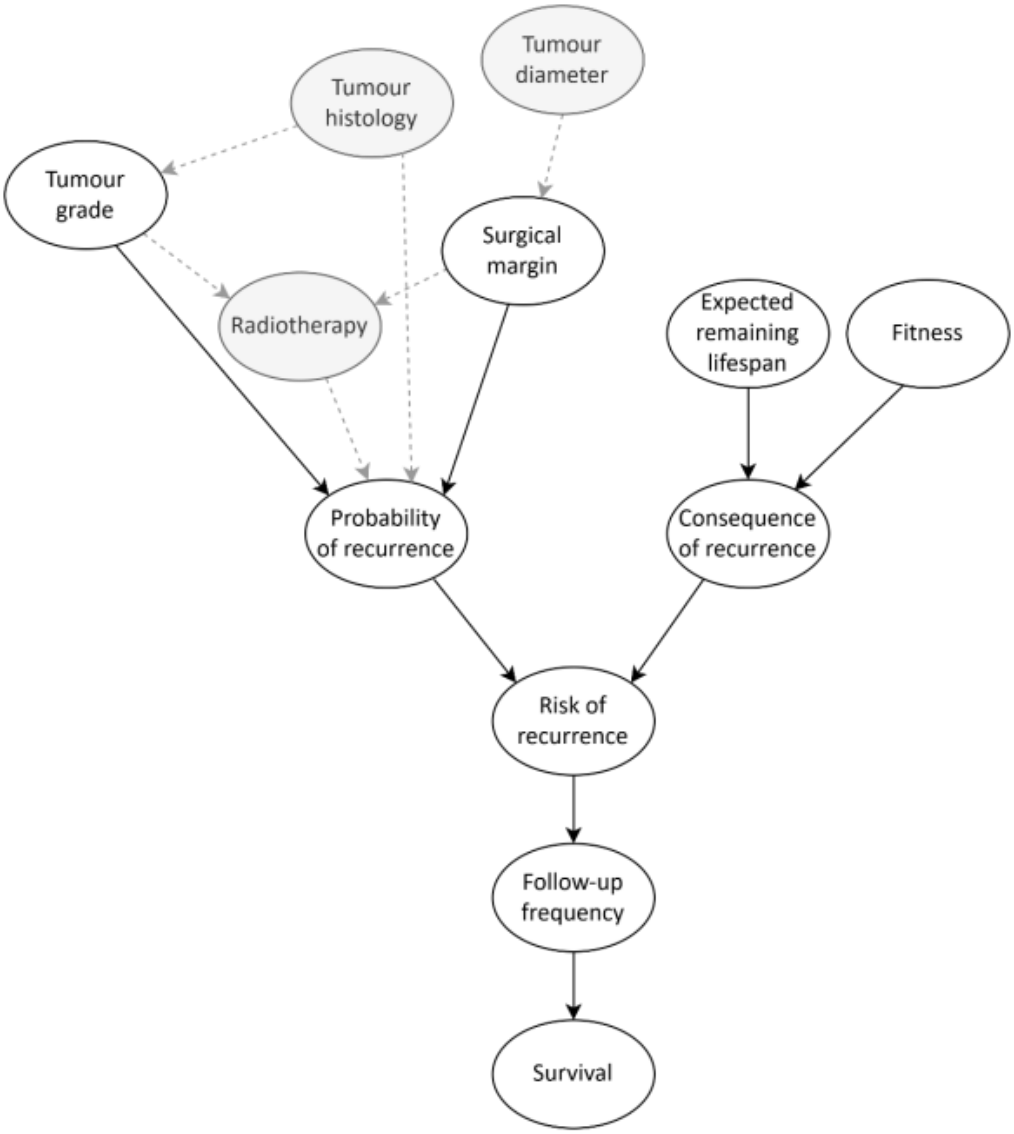
Your input is highly valuable, and we sincerely appreciate your expertise in shaping this research.
For any questions about this research, please contact:
Nina van Staalduine at C.M.C.vanStaalduine@student.tudelft.nl or Rioshar Yarveisy at r.yarveisy@tudelft.nl.
By proceeding to the survey, you confirm that you have read this information and agree to participate in the study.

Survey

Introduction into the Current Decision-Support Model

The illustration below shows the current decision-support model that is incorporated in the tool, and the evidence on which it is based. It is intended to be used in the post-treatment phase, when you as a clinician need to decide on the frequency of follow-up visits and should be read as follows:

- Follow-up visits are scheduled to ensure that recurrences are detected as soon as possible, so that we can offer treatment that will positively impact survival.
- The frequency of follow-up visits depends on the risk of recurrence for a particular patient: some patients may require higher frequency because recurrence is more likely, depending on tumour characteristics and surgical margin (left side), or because of the consequences of finding a recurrence (right side) depending on expected remaining life expectancy and fitness of the patient. For instance, the consequence of late detection for an older unfit patient may be low, as they may not be fit enough to undergo treatment and may prefer to enjoy their remaining days of life without the stress of more follow-up visits. Younger fit patients, however, are likely to undergo treatment, recover and continue with their lives, so that it will be worth the stress of more follow-up visits and allocation of resources to improve their chances of survival.



Section 1: Follow-up frequency and use of guidelines to support your decisions

Question 1

Is the risk of recurrence the main factor in deciding the follow-up frequency for your patients?

- Yes
- No

If respondent answered no:

Question 1B

Risk of recurrence is <u>not</u> your main factor in deciding follow-up frequency. Could you explain what factors you do use to decide on follow-up frequency?

Question 2

What guideline(s) do you follow when making follow-up decisions?

- ☐ ESMO
- ☐ EURACAN
- ☐ Other: ……

Question 3

Is there a distinction between high and low risk patients in your guideline?

- Yes
- No

If respondent answered yes:

Question 3B

Your guideline distinguishes between high and low risk patients, please tell us how it does that. The guideline uses:

- ☐ Tumour grade
- ☐ Tumur histology
- ☐ Surgical margin
- ☐ Other: (please elaborate) …….

If respondent answered yes in question 3:

Question 4

Your guideline distinguishes between high and low risk patients.

a. How do you usually apply the guideline for a patient you perceive to be high-risk?
   - I follow the guideline-recommended frequency
   - I schedule follow-up more frequently than the guideline advises
   - I schedule follow-up less frequently than the guideline advises
   - It depends on the specific case (please elaborate) …….

b. How do you usually apply the guideline for a patient you perceive to be low-risk?
   - I follow the guideline-recommended frequency
   - I schedule follow-up more frequently than the guideline advises
   - I schedule follow-up less frequently than the guideline advises
   - It depends on the specific case (please elaborate) …….

If respondent answered no in question 3:

Question 4

Your guideline does <u>not</u> distinguish between high and low risk patients.

c. How do you usually apply the guideline for a patient you perceive to be high-risk?
   - I follow the guideline-recommended frequency
   - I schedule follow-up more frequently than the guideline advises

- I schedule follow-up less frequently than the guideline advises
- It depends on the specific case (please elaborate) …….

d. How do you usually apply the guideline for a patient you perceive to be low-risk?
   - I follow the guideline-recommended frequency
   - I schedule follow-up more frequently than the guideline advises
   - I schedule follow-up less frequently than the guideline advises
   - It depends on the specific case (please elaborate) ……..

Question 5
Do you believe that a higher frequency of follow-up visits leads to higher chances of survival? (Please elaborate)

- Yes …….
- No …….

Question 6
Do you believe that timely detection of recurrence leads to higher chances of survival? (Please elaborate)

- Yes …….
- No …….

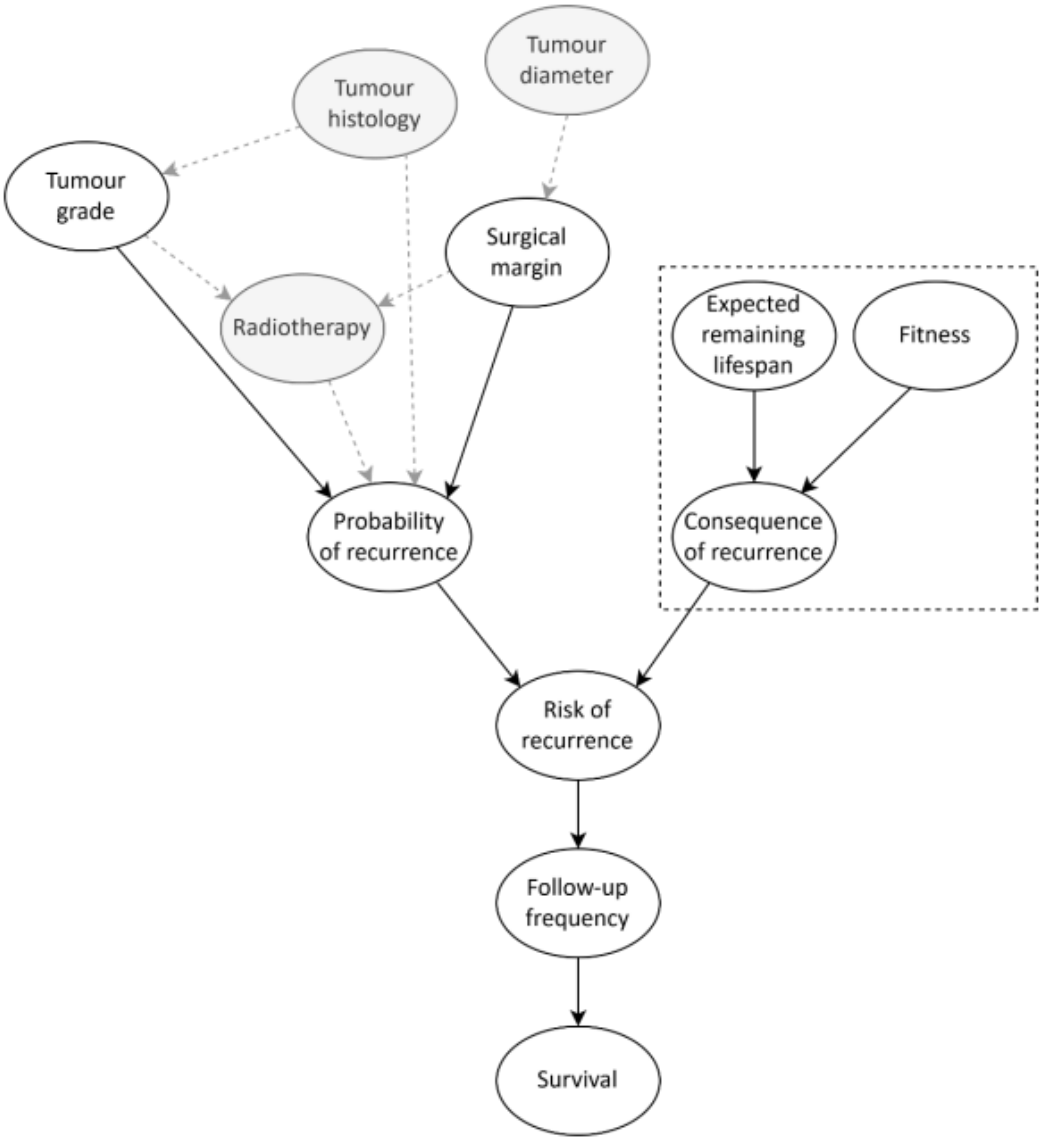
Section 2: Consequences for patients

As explained before, the model incorporates the consequence of a recurrence to reflect the acceptability of treatment harm for patients when a recurrence is detected (timely), which depends on:

- **Expected remaining lifespan**: For older patients with limited life expectancy, treatment harm will generally be less acceptable than for younger patients. So, consequences of timely detection will be lower for older patients.

- **Fitness**: Fitter patients will generally tolerate treatment better than unfit patients. So, consequences of timely detection will be higher for fitter patients. Fitness is based on a simplified WHO/ECOG Performance Status (see table below)

Simplified category	WHO/ECOG Performance Status score	Short description
Fit	0 and 1	Fully active or minor limitations
Unfit	2, 3 and 4	Limited self-care or bedridden part/all of the day



Question 7

Do you agree that the consequences of finding a recurrence depend on the patient's expected remaining lifespan and fitness level?

- Yes
- No (Please elaborate) …….

Question 8

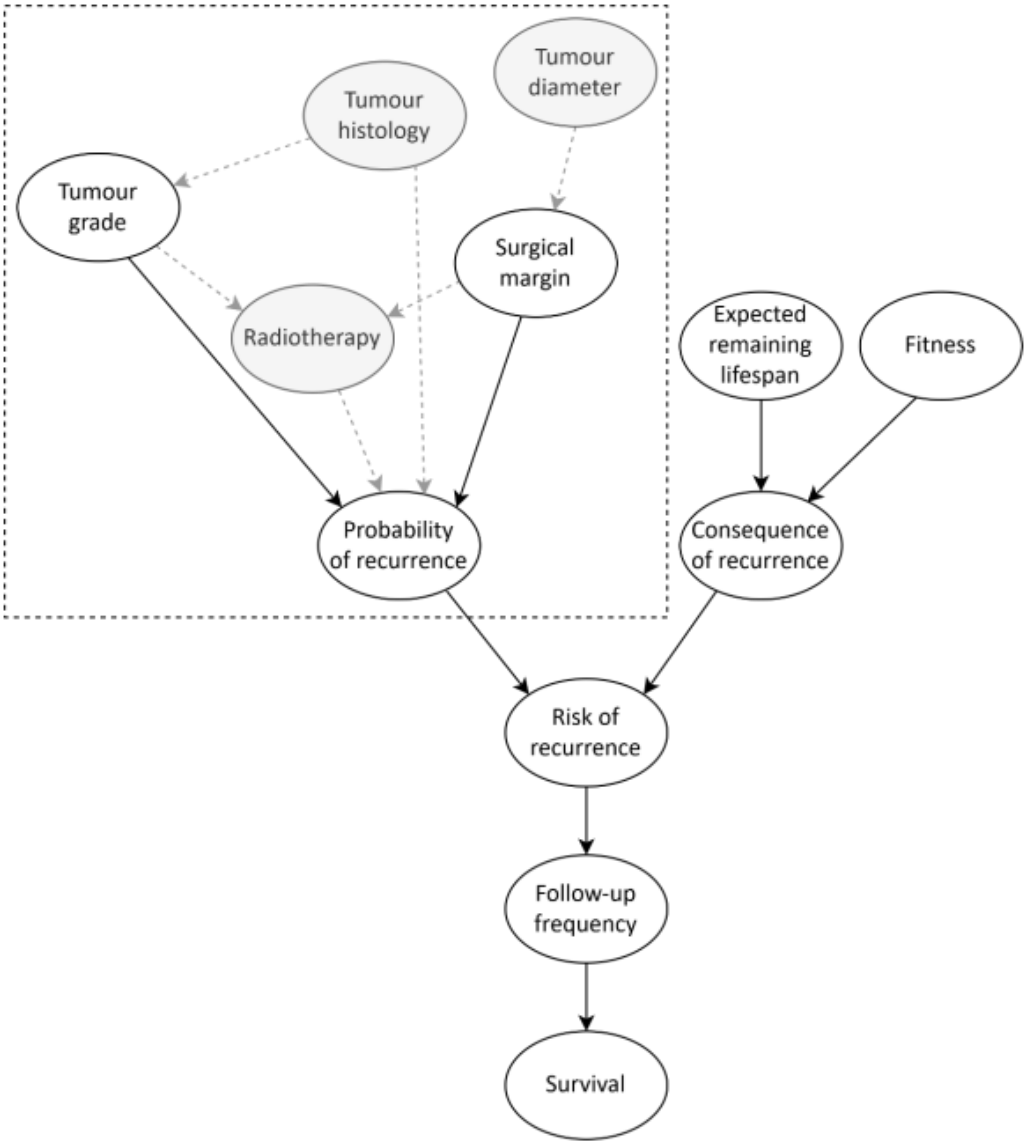
What level of consequence for finding a recurrence timely would you assign to the following patients?

	Low	Moderate	High
A fit patient with an expected remaining lifespan of 40 years or more?	☐	☐	☐
An unfit patient with an expected remaining lifespan of 40 years or more?	☐	☐	☐

A fit patient with an expected remaining lifespan between 20 and 40 years?	☐	☐	☐
An unfit patient with an expected remaining lifespan between 20 and 40 years?	☐	☐	☐
A fit patient with an expected remaining lifespan less than 20 years?	☐	☐	☐
An unfit patient with an expected remaining lifespan less than 20 years?	☐	☐	☐

Section 3: Factors influencing the likelihood of recurrence

The **probability of recurrence** depends on several tumour and treatment factors. Analyses of historical patient data have shown that these are interrelated, as shown in the figure below. These analyses showed that incorporating only tumour grade and surgical margin would be sufficient, as this would also incorporate the other factors, given the interrelations.



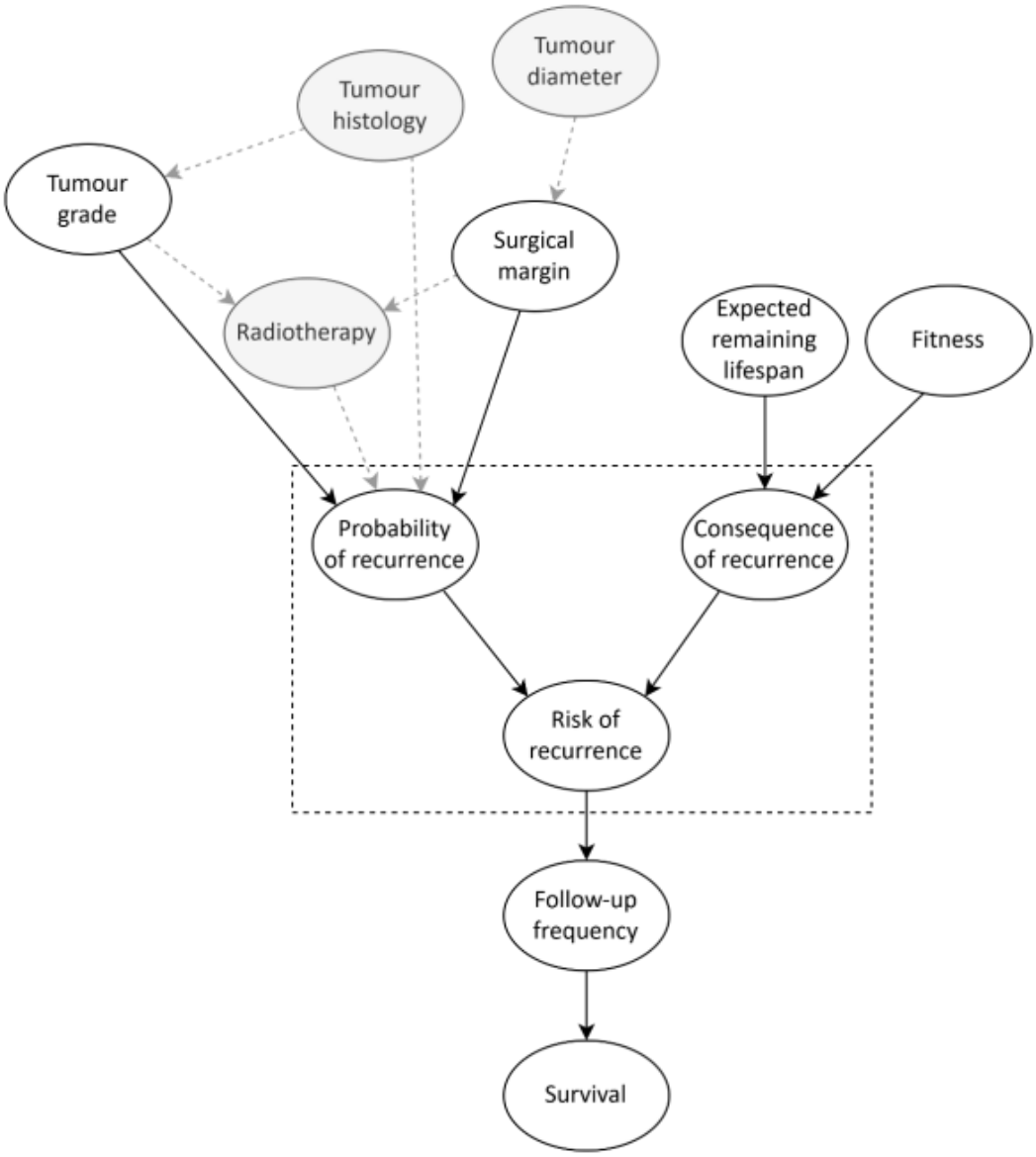
Question 9

Would you consider using a decision-support tool where the probability of recurrence is estimated based on tumour grade and surgical margin?

- Yes
- No (Please elaborate) …….

Section 4: Risk of recurrence

The **risk of recurrence** combines the probability of recurrence with the consequence of recurrence (reflecting acceptability of treatment harm in case of recurrence).



Question 10
What level of risk of recurrence would you assign to the following patients?

What risk level of recurrence do you assign to the following patients?	Low risk	High risk
Patient with a high probability of recurrence and high consequence of recurrence?	☐	☐
Patient with a high probability of recurrence and moderate consequence of recurrence?	☐	☐
Patient with a high probability of recurrence and low consequence of recurrence?	☐	☐
Patient with a low probability of recurrence and high consequence of recurrence?	☐	☐
Patient with a low probability of recurrence and moderate consequence of recurrence?	☐	☐
Patient with a low probability of recurrence and low consequence of recurrence?	☐	☐

Closing question
Do you have any additional comments or suggestions regarding the structure or content of the model?

A.3 Preliminary Influence Diagram Post-Treatment Context

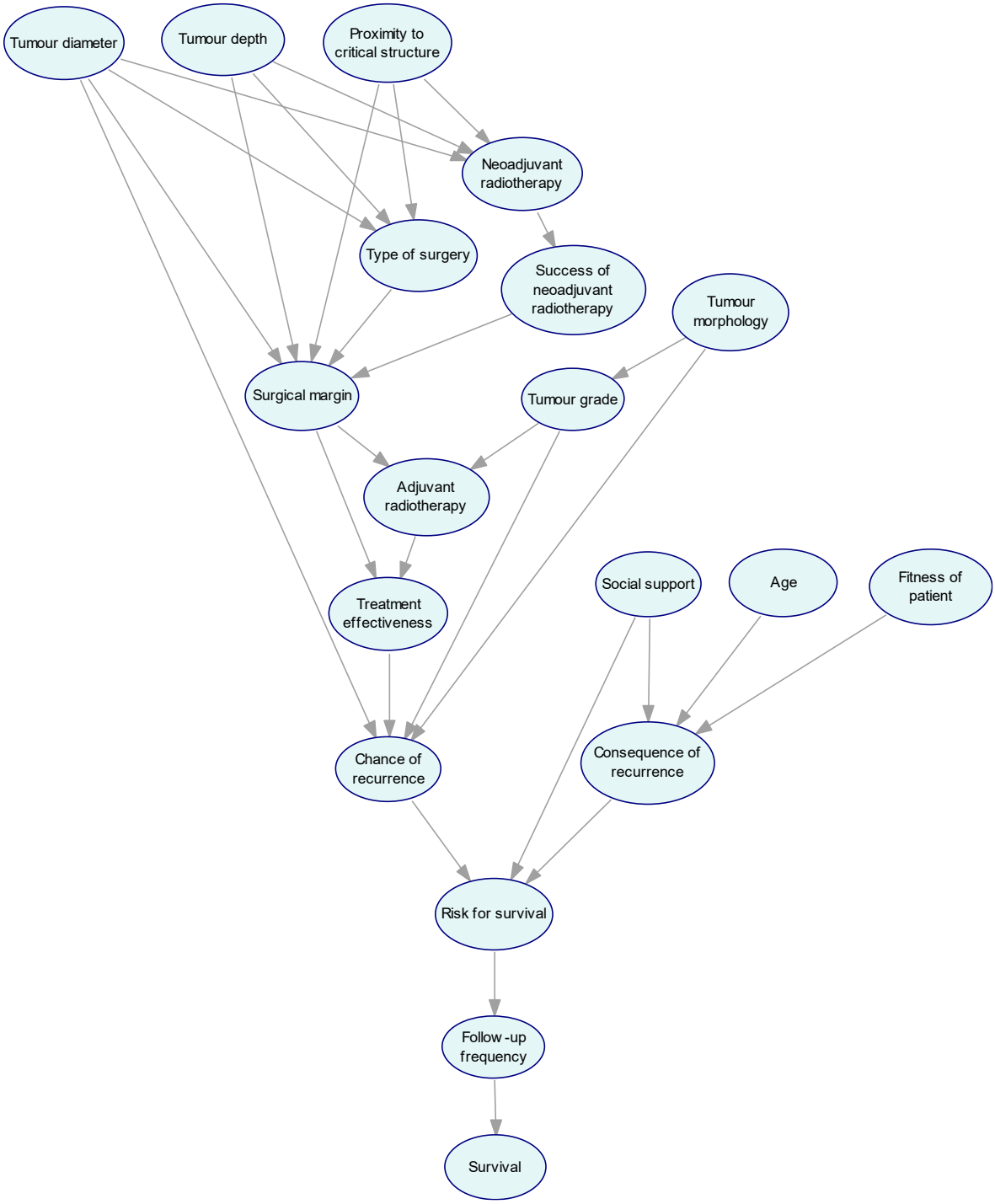


Figure 23: Preliminary influence diagram based on the exploratory discussion with an STS expert

A.4 Conditional Dependency between Surgical Margin and Radiotherapy

The total share of radiotherapy is almost identical for both surgical margin groups, namely 25.7% for R0 and 27.0% for R1/R2, as shown in Table 48. An interesting finding is that the percentage of neoadjuvant radiotherapy for the R0-group is nearly twice as high as for the R1/R2-group (Table 48). The fact that neoadjuvant radiotherapy is often used to shrink the tumour size before surgery, which would allow for easier R0-resection, could explain this finding.

Table 48: Margin R0 and R1/R2 cases and radiotherapy distribution, including share within margin groups and total population

Radiotherapy	Margin R0 cases (n, %)	Margin R1/R2 cases (n, %)	Total (n, %)
None	182 (74.3%)	89 (73.0%)	271 (73.8%)
Neoadjuvant	25 (10.2%)	7 (5.7%)	32 (8.7%)
Adjuvant	38 (15.5%)	26 (21.3%)	64 (17.4%)
Total	245 (100%)	122 (100%)	367 (100%)

Table 49: Margin R1/R2 percentage by radiotherapy

Radiotherapy	Margin R1/R2 percentage
None	32.8%
Neoadjuvant	21.9%
Adjuvant	40.6%
Total	65.4%

Table 50 shows that there is no clear relation between surgical margin, radiotherapy and recurrence. Tumour grade, as explained in the main text, is better at explaining the variation in recurrence percentages.

Table 50: Recurrence cases and percentages by surgical margin and radiotherapy

		Recurrence counts		Recurrence percentages	
Surgical margin	Radiotherapy	No	Yes	By surgical margin and radiotherapy	By surgical margin
R0	Adjuvant	26	12	31.6%	
	Neoadjuvant	15	10	40.0%	
	None	141	41	22.5%	25.7%
R1/R2	Adjuvant	13	13	50.0%	
	Neoadjuvant	5	2	28.6%	
	None	64	25	28.1%	32.8%

A.5 Survey Results

All questions in the survey were multiple choice questions and sometimes an elaboration was requested. For each possible answer choice, the number of times it was chosen is reported, together with the percentage compared to the total respondents.

Question 1

Is the risk of recurrence the main factor in deciding the follow-up frequency for your patients?

- Yes: 4
- No: 3
  Explanation:
  - It depends on a combination of tumor-related, treatment-related, and patient-related factors.
  - Risk of recurrence and risk of metastasis
  - We use a stratified protocol for follow-up, partly based on diagnosis (in some aspect this is recurrence) however for lesions with chance <1% for LM there is still a significant follow-up program with multiple follow-up moment per year, especially the first year. As this is protocol it is hard to do a less aggressive FU protocol
    Very little scientific evidence for CT scans vs radiographs.

Question 2

What guideline(s) do you follow when making follow-up decisions?

- ESMO-EURACAN: 5 (out of 7)
- NCCN: 2 (out of 7)
- Other: 4 (out of 7)
  - Our own
  - WeBot guideline
  - Our NOV leidraad
  - LUMC bone soft tissue follow-up scheme

Question 3A

Is there a distinction between high- and low-risk patients in your guideline?

- Yes: 7
- No: 0

Question 3B

Your guideline distinguishes between high and low risk patients, please tell us how it does that.
The guideline uses:

- Tumour grade: 6 (out of 7)
- Surgical margin: 1 (out of 7)
- Tumour histology: 7 (out of 7)
- Other: 2 (out of 7)
  - Histology includes type and grade
  - With or without ctx

Question 4A

Your guideline distinguishes between high and low risk patients.
How do you usually apply the guideline for a patient you perceive to be high-risk?

- I follow the guideline-recommended frequency: 6
- I schedule follow-up more frequently than the guideline advises: 0
- I schedule follow-up less frequently than the guideline advises: 0
- It depends on the specific case (please elaborate): 1
    - R1 resection in high grade I FU more frequent

Question 4B

Your guideline distinguishes between high and low risk patients.
How do you usually apply the guideline for a patient you perceive to be low-risk?

- I follow the guideline-recommended frequency: 4
- I schedule follow-up more frequently than the guideline advises: 0
- I schedule follow-up less frequently than the guideline advises: 1
- It depends on the specific case (please elaborate): 2
    - In chondrosarcoma, no difference in gr 2-3-dediff is made. in grade 2 lesions of hand and feet and after amputations, I schedule less frequent FU
    - If possible with advanced age then less frequent. But a protocol is a reflexion of our approach, we should adhere to the protocol or change the protocol not (randomly) do something else for some patients.

Question 5

Do you believe that a higher frequency of follow-up visits leads to higher chances of survival?
(Please elaborate)

- Yes: 1
    - For some specific indications with high risk of LM, marginal or r1 resections (LR)
- No: 6
    - Not in general. Maybe in specific cases when there is a high change of local recurrence
    - Not survival, maybe better local control
    - It will lead to an earlier detection of recurrence or metastases, but the current scheme is sufficient. Growth of the metastases/recurrence will not be so quick that a higher frequency is necessary

Question 6

Do you believe that timely detection of recurrence leads to higher chances of survival?
(Please elaborate)

- Yes: 4
    - metastatectomy is performed is possible to achieve better survival
    - When applicable for high grade lesions with singular LM
- No: 3
    - More control locally. survival is more dependent on metastasis yes/no

Question 7

Do you agree that the consequences of finding a recurrence depend on the patient's expected remaining lifespan and fitness level?

- Yes: 5
- No: 1
    - Fitness level; if unfit and recurrence is small, than less morbid/risk for treatment.

Question 8

Regardless of whether you agree with this part of the model structure, please answer the following questions as if it were accurate.
What level of consequence for finding a recurrence timely would you assign to the following patients?

	Low	Moderate	High
A fit patient with an expected remaining lifespan of 40 years or more?			6
An unfit patient with an expected remaining lifespan of 40 years or more?		2	4
A fit patient with an expected remaining lifespan between 10 and 40 years?			6
An unfit patient with an expected remaining lifespan between 10 and 40 years?	1	3	2
A fit patient with an expected remaining lifespan less than 10 years?		6	
An unfit patient with an expected remaining lifespan less than 10 years?	4	2	

Question 9

Would you consider using a decision-support tool where the probability of recurrence is estimated based on tumour grade and surgical margin?

- Yes: 6
- No: 0

Question 10

What level of risk of recurrence would you assign to the following patients?

What risk level of recurrence do you assign to the following patients?	Low risk	High risk
Patient with a high probability of recurrence and high consequence of recurrence?		4
Patient with a high probability of recurrence and moderate consequence of recurrence?		4
Patient with a high probability of recurrence and low consequence of recurrence?	4	2
Patient with a low probability of recurrence and high consequence of recurrence?	2	4
Patient with a low probability of recurrence and moderate consequence of recurrence?	4	
Patient with a low probability of recurrence and low consequence of recurrence?	4	

A.6 Prior Probability Tables for Root Nodes

Surgical margin

Table 51: Prior probability table for the variable Surgical margin (rounded to 3 decimals)

Surgical margin	Prior probability
R0	0.668
R1 or R2	0.332

Tumour grade

Table 52: Prior probability table for the variable Tumour grade (rounded to 3 decimals)

Tumour grade	Prior probability
Low	0.346
High	0.654

Expected remaining lifespan

Table 53: Prior probability table for the variable Expected remaining lifespan (rounded to 3 decimals)

Expected remaining lifespan	Prior probability
More than 40 years	0.166
Between 10 and 40 years	0.556
Less than 10 years	0.278

Fitness

Table 54: Prior probability table for the variable Fitness (rounded to 3 decimals)

Fitness	Prior probability
Fit	0.485
Unfit	0.515

A.7 Risk Percentage Threshold Determination

The high-risk percentages, as given by the model, for all possible combinations of input values are shown in Table 55. Based on the threshold as indicated in the columns (either 30%, 40%, 50% or 60%), the cells were coloured red if higher than the threshold, indicating high-risk, and green if lower than the threshold, indicating low-risk. Note that the simplified risk model does not distinguish between different surgical margins, as this variable is not included in this model version. The risk classifications (indicated by red and green) for each threshold and model type, are added to the historical patient dataset, based on their specific characteristics (or input values in this case). Fitness was not in the dataset, but everybody under 60 was considered fit, else unfit.

Table 55: High- and low risk classifications (red and green respectively), based on high-risk percentage outcomes for the model for all possible input combinations with different thresholds

Fitness	Expected remaining lifespan	Tumour grade	Surgical margin	Extensive risk model				Simplified risk model				Observable combinations
				30%	40%	50%	60%	30%	40%	50%	60%	
Fit	10-	High	R0	34	34	34	34	100	100	100	100	
			R1/R2	44	44	44	44					
		Low	R0	8	8	8	8	0	0	0	0	
			R1/R2	16	16	16	16					
	10 to 40	High	R0	100	100	100	100	100	100	100	100	X
			R1/R2	100	100	100	100					X
		Low	R0	100	100	100	100	100	100	100	100	X
			R1/R2	100	100	100	100					X
	40+	High	R0	100	100	100	100	100	100	100	100	X
			R1/R2	100	100	100	100					X
		Low	R0	100	100	100	100	100	100	100	100	X
			R1/R2	100	100	100	100					X
Unfit	10-	High	R0	11	11	11	11	33	33	33	33	X
			R1/R2	15	15	15	15					X
		Low	R0	3	3	3	3	0	0	0	0	X
			R1/R2	5	5	5	5					X
	10 to 40	High	R0	50	50	50	50	83	83	83	83	X
			R1/R2	56	56	56	56					X
		Low	R0	37	37	37	37	33	33	33	33	X
			R1/R2	41	41	41	41					X
	40+	High	R0	78	78	78	78	100	100	100	100	
			R1/R2	81	81	81	81					
		Low	R0	69	69	69	69	67	67	67	67	
			R1/R2	72	72	72	72					

After this, confusion matrices were made, putting the classifications made by the model next to the classification made by the STS expert. Since we had to estimate the fitness level for the patients in the dataset, only 16 out of 24 input combinations were possible to see following the extensive risk model, while this was 8 out of 12 for the simplified risk model. As can be seen in Table 55, the simplified risk model behaves similarly under the 40%, 50% and 60% thresholds. Therefore, only one confusion matrix for these thresholds is shown. The confusion matrices for the extensive risk model

are shown in Table 56, Table 57, Table 59 and Table 59, and for the simplified model in Table 60 and Table 61.

Table 56: Confusion matrix extensive risk model with a threshold of 30%

Risk classification from STS expert	**Risk classification Extensive risk model (30% threshold)** Low-risk	High-risk	Total
Low-risk	57	191	248
High-risk	45	74	119
Total	102	265	367

Table 57: Confusion matrix extensive risk model with a threshold of 40%

Risk classification from STS expert	**Risk classification Extensive risk model (40% threshold)** Low-risk	High-risk	Total
Low-risk	77	171	248
High-risk	45	74	119
Total	122	245	367

Table 58: Confusion matrix extensive risk model with a threshold of 50%

Risk classification from STS expert	**Risk classification Extensive risk model (50% threshold)** Low-risk	High-risk	Total
Low-risk	77	171	248
High-risk	63	56	119
Total	140	227	367

Table 59: Confusion matrix extensive risk model with a threshold of 60%

Risk classification from STS expert	**Risk classification Extensive risk model (60% threshold)** Low-risk	High-risk	Total
Low-risk	110	138	248
High-risk	79	40	119
Total	189	178	367

Table 60: Confusion matrix simplified risk model with a threshold of 30%

Risk classification from STS expert	Risk classification Simplified risk model (30% threshold) Low-risk	High-risk	Total
Low-risk	11	237	248
High-risk	13	106	119
Total	24	343	367

Table 61: Confusion matrix simplified risk model with a threshold of 40%, 50% and 60%

Risk classification from STS expert	Risk classification Simplified risk model (40%, 50% and 60% threshold) Low-risk	High-risk	Total
Low-risk	77	171	248
High-risk	63	56	119
Total	140	227	367

Based on Table 56 until Table 61, evaluation metrics can be calculated for each threshold, as shown in Table 62. It shows that the 40% threshold has the best accuracy and the (shared) lowest proportion that was classified too low by the extensive risk model. The 60% threshold scored better than 40% in the proportion of patients that were classified too high. While the 40% and 60% have a similar matching score, the 40% threshold is more careful as it underestimates the least. The 40% threshold scores the best in the extensive risk model, but also in the simplified risk model, since the accuracy is the highest. Therefore, the 40% threshold is chosen to be used for the rest of this study.

Table 62: Evaluation metrics for different thresholds in extensive risk model

Evaluation metrics	Extensive risk model				Simplified risk model	
	30%	40%	50%	60%	30%	40%-60%
Matching (accuracy)	35.7%	41.1%	36.2%	40.9%	31.9%	36.2%
Model classified higher than clinician	52.0%	46.6%	46.6%	37.6%	64.6%	46.6%
Model classified lower than clinician	12.3%	12.3%	17.2%	21.5%	3.5%	17.2%

A.8 CPTs for Nodes Following Risk of Recurrence

A.8.1 Extensive Risk Model

Table 63: CPT for LUMC follow-up guideline adherence year 1 and 2 by extensive model risk classification

Extensive model risk classification	LUMC follow-up guideline adherence Year 1 and 2		
	Below	On	Above
Low	0.656	0.311	0.033
High	0.559	0.371	0.069

Table 64: CPT for survival year 1 and 2 by extensive model risk classification

Extensive model risk classification	Survival Year 1 and 2	
	Survived	Died
Low	0.795	0.205
High	0.902	0.098

Table 65: CPT for LUMC follow-up guideline adherence year 3 to 5 by extensive model risk classification

Extensive model risk classification	LUMC follow-up guideline adherence Year 3 to 5		
	Below	On	Above
Low	0.812	0.174	0.014
High	0.733	0.221	0.047

Table 66: CPT for survival year 3 to 5 by extensive model risk classification

Extensive model risk classification	Survival Year 3 to 5 (Including patients with a recurrence in year 1 or 2)	
	Survived	Died
Low	0.856	0.144
High	0.891	0.109

Table 67: CPT for survival in total 5 years by extensive model risk classification

Extensive model risk classification	Survival total 5 years	
	Survived	Died
Low	0.680	0.320
High	0.804	0.196

Table 68: CPT for survival year 1 and 2 by extensive model risk classification and LUMC follow-up guideline adherence year 1 and 2

Extensive model risk classification	LUMC follow-up guideline adherence Year 1 and 2	Survival Year 1 and 2	
		Survived	Died
Low	Below	0.838	0.162
	On	0.737	0.263
	Above	0.500	0.500
High	Below	0.949	0.051
	On	0.857	0.143
	Above	0.765	0.235

Table 69: CPT for survival year 3 to 5 (if year 1 and 2 were recurrence-free) by extensive model risk classification and LUMC follow-up guideline adherence year 1 and 2, and 3 to 5

Extensive model risk classification	LUMC follow-up guideline adherence Year 1 and 2	LUMC follow-up guideline adherence Year 3 to 5	Survival Year 3 to 5 (if year 1 and 2 recurrence-free)	
			Survived	Died
Low	Below	Below	0.894	0.106
		On	1.000	0.000
		Above	NaN	NaN
	On	Below	0.889	0.111
		On	1.000	0.000
		Above	NaN	NaN
	Above	Below	NaN	NaN
		On	NaN	NaN
		Above	0.000	1.000
High	Below	Below	0.990	0.010
		On	0.917	0.083
		Above	NaN	NaN
	On	Below	1.000	0.000
		On	1.000	0.000
		Above	1.000	0.000
	Above	Below	NaN	NaN
		On	0.800	0.200
		Above	0.667	0.333

A.8.2 Simplified Risk Model

Table 70: CPT for LUMC follow-up guideline adherence year 1 and 2 by simplified model risk classification

Simplified model risk classification	LUMC follow-up guideline adherence Year 1 and 2		
	Below	On	Above
Low	0.686	0.286	0.029
High	0.533	0.392	0.075

Table 71: CPT for survival year 1 and 2 by simplified model risk classification

Simplified model risk classification	Survival Year 1 and 2	
	Survived	Died
Low	0.821	0.179
High	0.894	0.106

Table 72: CPT for LUMC follow-up guideline adherence year 3 to 5 by simplified model risk classification

Simplified model risk classification	LUMC follow-up guideline adherence Year 3 to 5		
	Below	On	Above
Low	0.843	0.145	0.012
High	0.709	0.241	0.051

Table 73: CPT for survival year 3 to 5 by simplified model risk classification

Simplified model risk classification	Survival Year 3 to 5 (Including patients with a recurrence in year 1 or 2)	
	Survived	Died
Low	0.878	0.122
High	0.882	0.118

Table 74: CPT for survival in total 5 years by simplified model risk classification

Simplified model risk classification	Survival total 5 years	
	Survived	Died
Low	0.721	0.279
High	0.789	0.211

Table 75: CPT for survival year 1 and 2 by simplified model risk classification and LUMC follow-up guideline adherence year 1 and 2

Simplified model risk classification	LUMC follow-up guideline adherence Year 1 and 2	Survival Year 1 and 2	
		Survived	Died
Low	Below	0.865	0.135
	On	0.750	0.250
	Above	0.500	0.500
High	Below	0.942	0.058
	On	0.854	0.146
	Above	0.765	0.235

Table 76: CPT for survival year 3 to 5 (if year 1 and 2 were recurrence-free) by simplified model risk classification and LUMC follow-up guideline adherence year 1 and 2, and 3 to 5

Simplified model risk classification	LUMC follow-up guideline adherence Year 1 and 2	LUMC follow-up guideline adherence Year 3 to 5	Survival Year 3 to 5 (if year 1 and 2 recurrence-free)	
			Survived	Died
Low	Below	Below	0.915	0.085
		On	1.000	0.000
		Above	NaN	NaN
	On	Below	0.909	0.091
		On	1.000	0.000
		Above	NaN	NaN
	Above	Below	NaN	NaN
		On	NaN	NaN
		Above	0.000	1.000
High	Below	Below	0.989	0.011
		On	0.917	0.083
		Above	NaN	NaN
	On	Below	1.000	0.000
		On	1.000	0.000
		Above	1.000	0.000
	Above	Below	NaN	NaN
		On	0.800	0.200
		Above	0.667	0.333

A.8.3 Clinician Model

Table 77: CPT for LUMC follow-up guideline adherence year 1 and 2 by clinician risk classification

Clinician risk classification	LUMC follow-up guideline adherence Year 1 and 2		
	Below	On	Above
Low	0.560	0.375	0.065
High	0.655	0.303	0.042

Table 78: CPT for survival year 1 and 2 by clinician risk classification

Clinician risk classification	Survival Year 1 and 2	
	Survived	Died
Low	0.875	0.125
High	0.849	0.151

Table 79: CPT for LUMC follow-up guideline adherence year 3 to 5 by clinician risk classification

Clinician risk classification	LUMC follow-up guideline adherence Year 3 to 5		
	Below	On	Above
Low	0.717	0.237	0.046
High	0.853	0.132	0.015

Table 80: CPT for survival year 3 to 5 by clinician risk classification

Clinician risk classification	Survival Year 3 to 5 (Including patients with a recurrence in year 1 or 2)	
	Survived	Died
Low	0.876	0.124
High	0.891	0.109

Table 81: CPT for survival in total 5 years by clinician risk classification

Clinician risk classification	Survival total 5 years	
	Survived	Died
Low	0.766	0.234
High	0.756	0.244

Table 82: CPT for survival year 1 and 2 by clinician risk classification and LUMC follow-up guideline adherence year 1 and 2

Clinician risk classification	LUMC follow-up guideline adherence Year 1 and 2	Survival Year 1 and 2	
		Survived	Died
Low	Below	0.906	0.094
	On	0.849	0.151
	Above	0.750	0.250
High	Below	0.910	0.090
	On	0.750	0.250
	Above	0.600	0.400

Table 83: CPT for survival year 3 to 5 (if year 1 and 2 were recurrence-free) by clinician risk classification and LUMC follow-up guideline adherence year 1 and 2, and 3 to 5

Clinician risk classification	LUMC follow-up guideline adherence Year 1 and 2	LUMC follow-up guideline adherence Year 3 to 5	Survival Year 3 to 5 (if year 1 and 2 recurrence-free)	
			Survived	Survived
Low	Below	Below	0.938	0.062
		On	0.933	0.067
		Above	NaN	NaN
	On	Below	0.963	0.037
		On	1.000	0.000
		Above	1.000	0.000
	Above	Below	NaN	NaN
		On	0.800	0.200
		Above	0.500	0.500
High	Below	Below	1.000	0.000
		On	1.000	0.000
		Above	NaN	NaN
	On	Below	1.000	0.000
		On	1.000	0.000
		Above	1.000	0.000
	Above	Below	NaN	NaN
		On	NaN	NaN
		Above	NaN	NaN

A.8.4 Grade Model

Table 84: CPT for LUMC follow-up guideline adherence year 1 and 2 by tumour grade

	LUMC follow-up guideline adherence Year 1 and 2		
Tumour grade	Below	On	Above
Low	0.850	0.142	0.008
High	0.454	0.463	0.083

Table 85: CPT for survival year 1 and 2 by tumour grade

	Survival Year 1 and 2	
Tumour grade	Survived	Died
Low	0.992	0.008
High	0.800	0.200

Table 86: CPT for ESMO follow-up guideline adherence year 3 to 5 by tumour grade

	ESMO follow-up guideline adherence Year 3 to 5		
Tumour grade	Below	On	Above
Low	0.874	0.117	0.009
High	0.654	0.285	0.062

Table 87: CPT for survival year 3 to 5 by tumour grade

	Survival Year 3 to 5 (Including patients with a recurrence in year 1 or 2)	
Tumour grade	**Survived**	**Died**
Low	0.944	0.056
High	0.839	0.161

Table 88: CPT for survival in total 5 years by tumour grade

	Survival total 5 years	
Tumour grade	**Survived**	**Died**
Low	0.937	0.063
High	0.671	0.329

Table 89: CPT for survival year 1 and 2 by tumour grade and ESMO follow-up guideline adherence year 1 and 2

		Survival Year 1 and 2	
Tumour grade	**ESMO follow-up guideline adherence Year 1 and 2**	**Survived**	**Died**
Low	Below	1.000	0.000
	On	0.944	0.056
	Above	1.000	0.000
High	Below	0.817	0.183
	On	0.802	0.198
	Above	0.700	0.300

Table 90: CPT for survival year 3 to 5 (if year 1 and 2 were recurrence-free) by tumour grade and ESMO follow-up guideline adherence year 1 and 2, and 3 to 5

Tumour grade	ESMO follow-up guideline adherence Year 1 and 2	ESMO follow-up guideline adherence Year 3 to 5	Survival Year 3 to 5 (if year 1 and 2 recurrence-free)	
			Survived	Survived
Low	Below	Below	0.978	0.043
		On	1.000	0.000
		Above	NaN	NaN
	On	Below	0.875	0.125
		On	1.000	0.000
		Above	NaN	NaN
	Above	Below	NaN	NaN
		On	NaN	NaN
		Above	0.000	1.000
High	Below	Below	0.932	0.068
		On	0.900	0.100
		Above	NaN	NaN
	On	Below	1.000	0.000
		On	1.000	0.000
		Above	1.000	0.000
	Above	Below	NaN	NaN
		On	0.800	0.200
		Above	0.667	0.333

A.9 Implementation of Clinician model and Tumour grade model

The implementations of the Clinician model, with risk of recurrence based on the STS expert's risk classification, and the Tumour grade model are shown in Figure 25 and Figure 26.

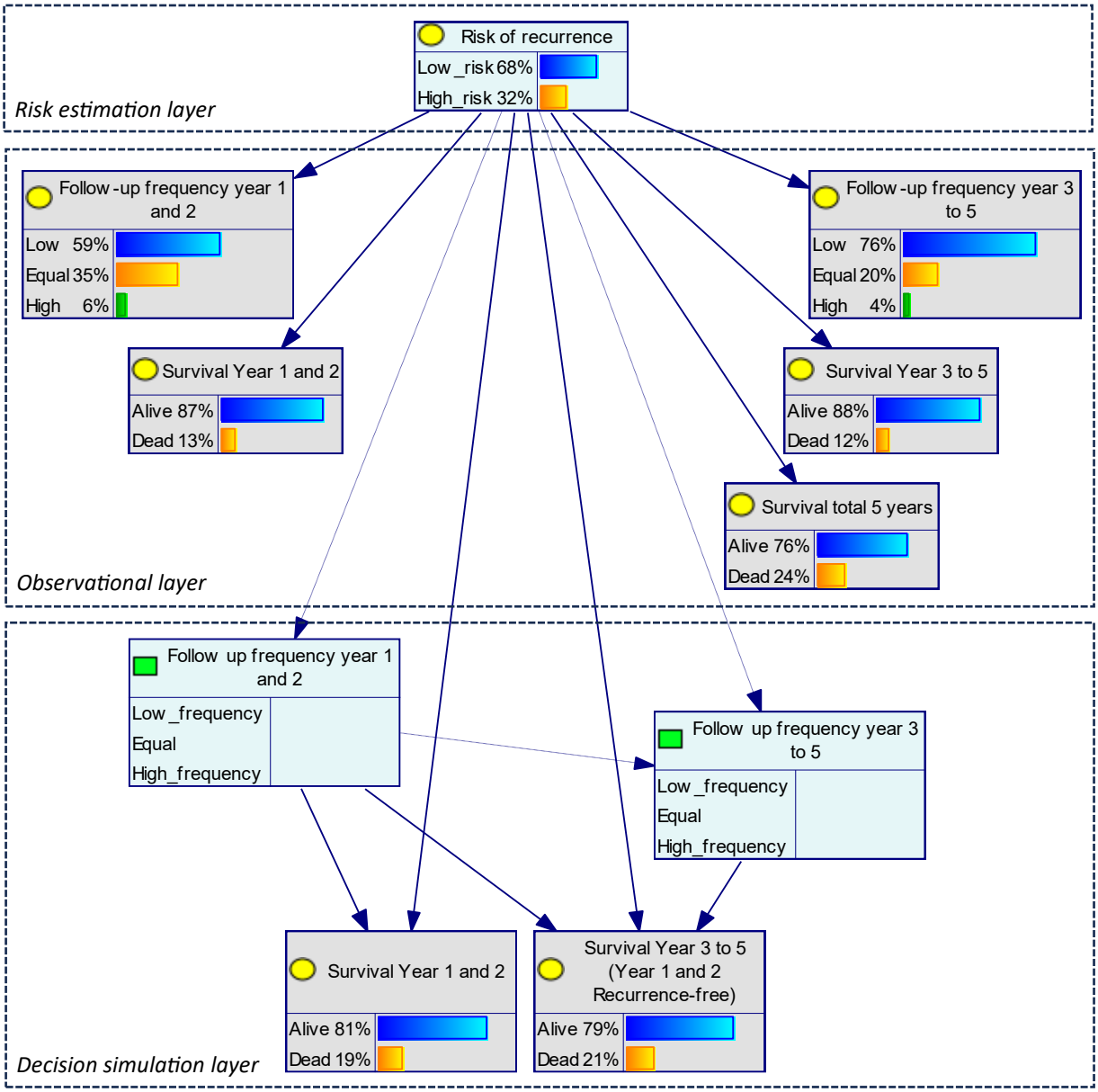


Figure 24: Implementation of Clinician Model as Bayesian network, indicating the three different layers, and nodes for which evidence should never be set are depicted in grey

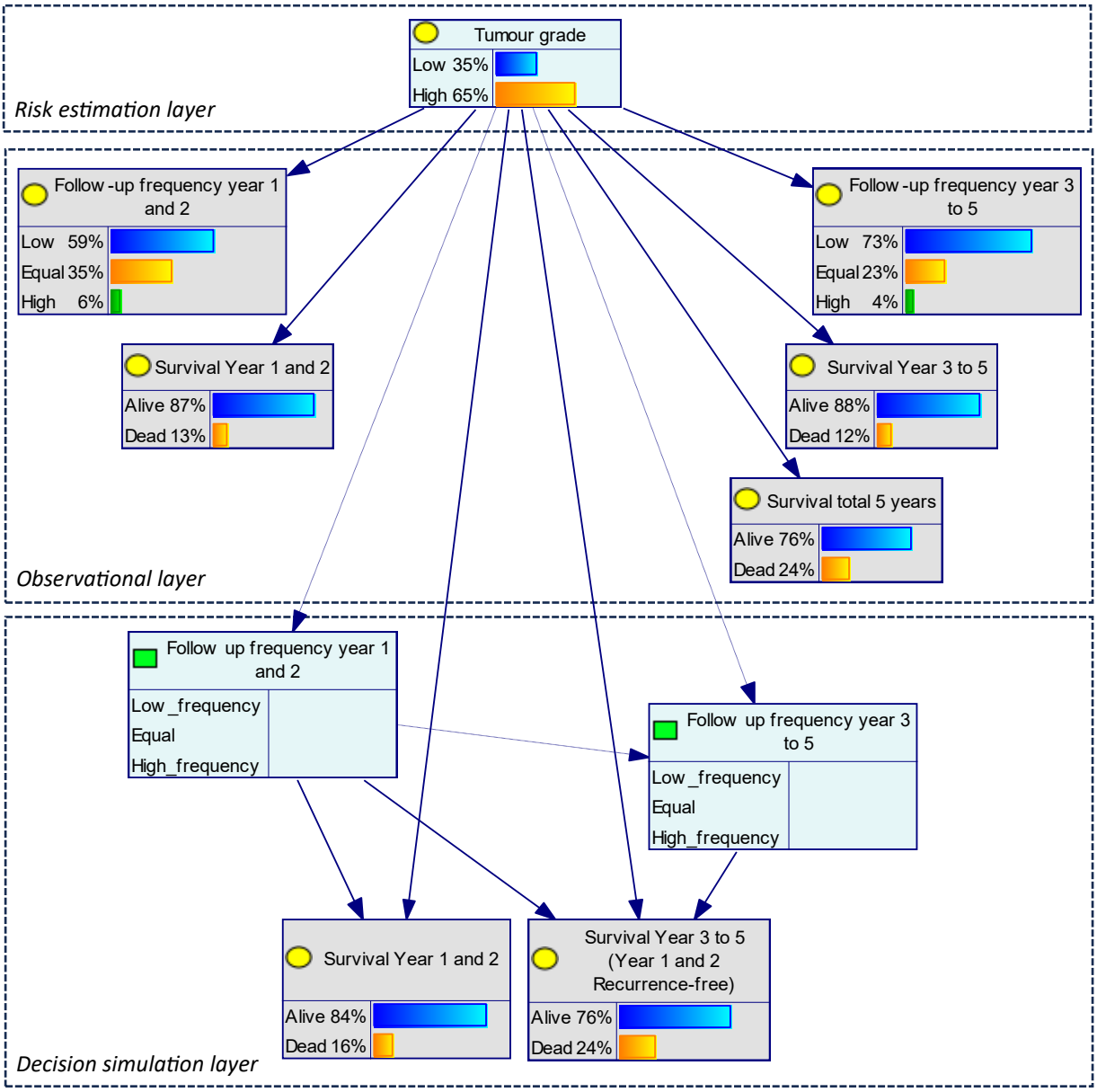


Figure 25: Implementation of Tumour grade model as Bayesian network, indicating the three different layers, and nodes for which evidence should never be set are depicted in grey