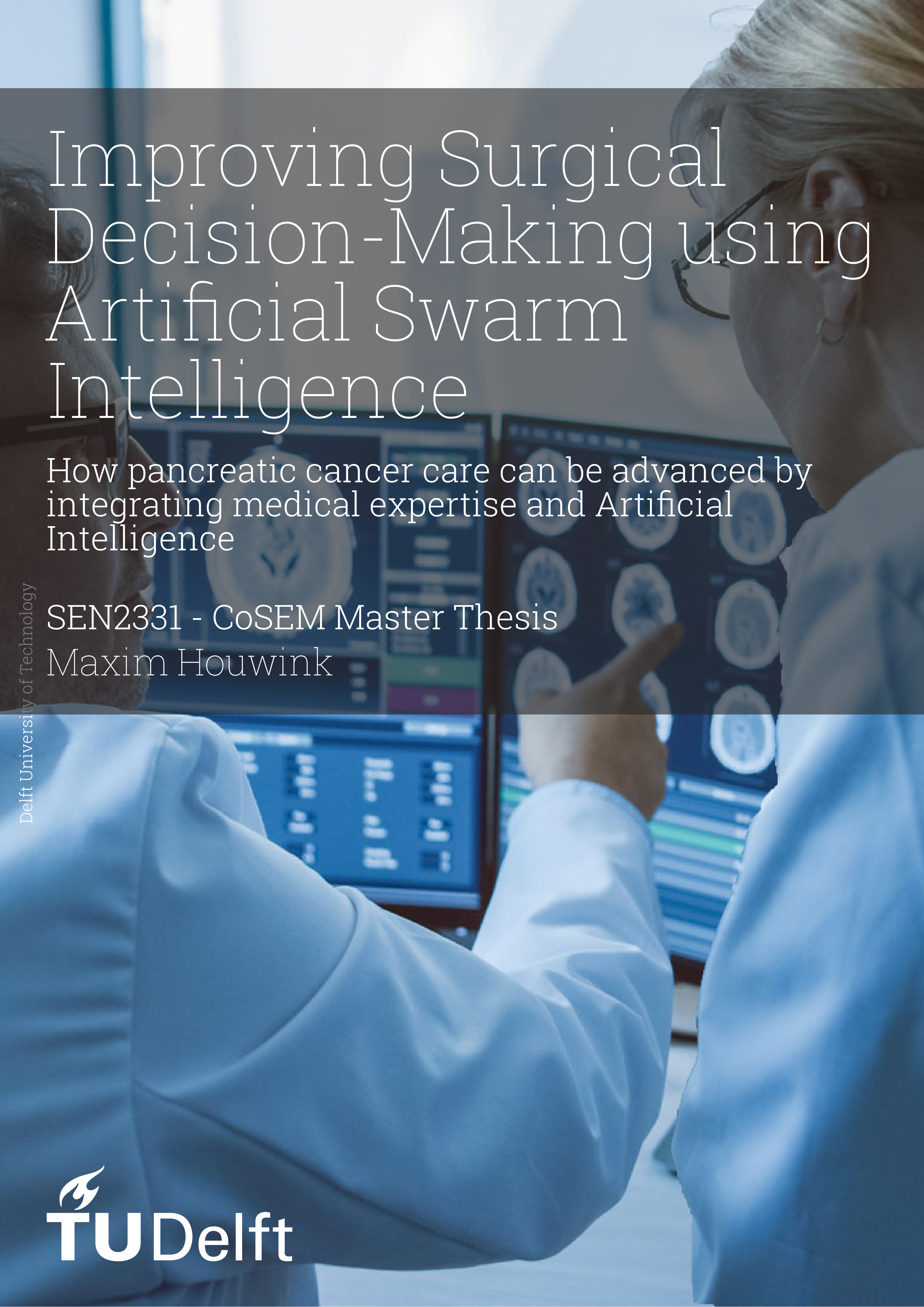


Improving Surgical Decision-Making using Artificial Swarm Intelligence

How pancreatic cancer care can be advanced by integrating medical expertise and Artificial Intelligence

SEN2331 - CoSEM Master Thesis

Maxim Houwink

Delft University of Technology

TU Delft

Improving Surgical Decision-Making using Artificial Swarm Intelligence

How pancreatic cancer care can be advanced by integrating medical expertise and Artificial Intelligence

by

Maxim Houwink

Chair:	Saba Hinrichs-Krapels
First Supervisor:	Kateřina Staňková
Second Supervisor:	Ashika Maharaj
Duration:	Q3-Q4 2024
Faculty:	Technology, Policy, and Management

TU Delft

Acknowledgement

With the completion of this master's thesis, I am beyond grateful to finish what has been one of the most professionally challenging yet rewarding journeys so far. Considering the uncertainties and obstacles in which this project started, apprehending the accomplishments achieved in terms of the experimental design, the execution, and the final product of this research is exceptionally remarkable. Particularly given the challenging personal circumstances in which this research was conducted, I cannot express the level of gratification with which I am reflecting on this journey.

First and foremost, I want to express my sincerest gratitude to my graduation committee: Saba Hinrichs-Krapels, Kateřina Staňková and Ashika Maharaj - each of you complemented each other to form an incredibly supportive yet critical supervisory team. Saba, your powerful and creative ideas continuously pushed me to think one step further in many aspects of the project. Kateřina, your critical and precise mindset forced me to challenge every step in the process. Lastly, Ashika: Your unwavering support contributed to more than just this research. Your willingness to help and positive guidance have made me a better student researcher and I am beyond grateful for your teachings throughout the entire journey of this research. Furthermore, I would also like to acknowledge the doctors at Leiden University Medical Center: Sven Mieog, Martijn van Dam, and Daphne Droogh. All three have contributed to an exceptional opportunity to set up the experiment in a real-world clinical setting. This research would not have been feasible without your help.

On a personal level, I would like to thank my family for their unwavering support. My mom, my dad, and my brother have been the foundation of my academic journey and their support lives on beyond this project. Similarly, I would also like to specifically recognize my friend Jos Groendijk. As a best friend, you have provided exceptional support, particularly during the tough times of this research as well as throughout my entire academic journey. Words are short for how much I appreciate your support.

Lastly, I would like to dedicate this project to my granddad, Eric Houwink. As a past professor and after graduating from Delft University of Technology as well, I hope to honor his work and professional achievements with this research. Although he is not able to witness my graduation in person anymore, I am convinced his life lessons and experiences still resonate with this research.

Closing off, the following quote serves as the foundation by which this thesis project as well as my academic journey has come to an end:

"The greater the obstacle, the greater the glory in overcoming it"

Maxim Houwink
Amsterdam, July 2024

Executive summary

Problem introduction

As one of the leading causes of cancer-related deaths, pancreatic cancer is currently one of the most challenging types of cancer for curative surgery. In treating patients with pancreatic cancer, multidisciplinary team (MDT) meetings have been established as the clinical standard for medical decision-making. Using MDT meetings, patient cases are discussed using combined expertise from various specializations including surgeons, oncologists, and radiologists. However, despite the combined expertise in surgical decision-making, research shows that 19%-33% of all pancreatic surgeries are observed to be prematurely abandoned due to locally advanced pancreatic cancer or metastatic disease. In these adverse cases, a patient is taken in for resection of the tumor, but the surgery is then abandoned to be informed that the patient has advanced cancer. Furthermore, MDT meetings typically feature an open discussion format that can be subject to social influence factors affecting the overall objectivity of individual expert opinions. Accordingly, a wide variety is observed in the format of discussion-based MDT meetings.

Research objective and main research question

To address these issues, this research explores Artificial Swarm Intelligence (ASI) as a potential technology to be used in MDT meetings. Based on swarm intelligence observed in natural species such as schools of fish, flocks of birds, or swarms of bees, this technology aims to bring these biological principles to human teams to improve decision-making processes. By amplifying collective intelligence using combined wisdom, expertise, and opinions, ASI has demonstrated to result in an enhanced performance when forecasting financial markets, an increased capability in correctly predicting sports games, and improved decision accuracy in medical diagnosis of pneumonia. In line with these findings, the objective of this research is to examine whether ASI can improve the surgical decision-making process, by assessing the impact of using ASI over discussion-based MDT meetings on the accuracy of tumor assessment and reduced social influence. To this end, the following main question is formulated:

In patients with potentially resectable PC, how does using ASI compared to discussion-based MDTs impact the tumor assessment accuracy and social influences in an MDT setting?

Following the main research question, the following subquestions are formulated:

- *RQ1: What are the reasons for adverse tumor assessment in PC?*
- *RQ2: To what extent does the accuracy of resectability assessment for PC cases change after the use of ASI in MDT meetings?*
- *RQ3: Can the use of ASI impact social influence during MDT meetings?*

Methods

This research deploys three methods. Most notably, this research deploys a novel approach by setting up the first experiment to test the use of ASI in a real-world setting. Whereas existing studies only compare the ASI performance against the average opinion of individual experts working in isolation, this experiment compares the tumor assessment accuracy using ASI against a simulated MDT meeting with experts in the field of pancreatic cancer. This experiment is set up at Leiden University Medical Center using de-identified CT images and case history to compare the test results against actual perioperative data. Second, to determine the setup and scope of the experiment, a preliminary literature review was conducted to identify the diagnostic classifications and underlying reasons for adverse tumor assessment. Third, after the experiment a survey was conducted to evaluate the impact of ASI on social influences in MDT meetings. To this end, two social influences are examined. On the one hand, the survey evaluates social loafing, which is defined as 'the decrease in individual effort when performing in groups as compared to the individual effort when performing alone'. On the other hand hand, social bias is examined to assess to what extent an individual's initial opinion is influenced by the behavior or opinions of other individuals in the group. By evaluating social loafing and social bias on evidenced factors that have been demonstrated to minimize these influences, the survey is used to indicate whether ASI could reduce social influence in MDT meetings.

Results

Based on the experiment results, the use of ASI shows equal assessment accuracy compared to assessing tumor resectability using discussion-based MDT meetings. The internal comparison showed a 100% concordance level, indicating that ASI did not affect the accuracy of tumor assessments in this experimental setup. Secondarily, in the ground truth comparison, a concordance level of 57.1% was observed, while the diagnosis comparison revealed a 62.5% concordance level. The survey results indicated that ASI has a modest impact on reducing social loafing and social bias. For social loafing, participants showed positive support for ASI's impact on task attractiveness and equalizing individual efforts. For social bias, the survey showed ASI's impact on decision anonymity and equal decision influence as important drivers to reduce social influence.

Discussion

The findings of this research suggest that ASI can match the decision accuracy of traditional MDT meetings while potentially reducing social biases. While the 100% concordance level suggests that ASI did not improve or reduce tumor assessment accuracy for this experiment, these results do not exclude that ASI could still contribute to reducing misdiagnosis in other forms of usage in MDT meetings. In addition, this research suggests that equal assessment accuracy could also highlight the role of other factors that play a role in the process of misdiagnosis. While the results of the social influence analysis show support for ASI's potential to reduce social loafing and bias, this research does not establish a direct causal relationship between ASI implementation and these social factors. The outcomes may also be influenced by the selected indicators, and including other relevant factors could significantly impact the analysis results.

The findings of this research with regard to diagnostic accuracy differ from existing studies which showed improved accuracy using ASI in medical diagnosis. This discrepancy may arise from the unique focus on group-based MDT approaches and the complex nature of tumor staging in the current study. Also, more detailed differences in the setup of the research may play a role. With regard to social influence, suggesting that ASI technology is designed to limit social bias effects. The research also supports the broader literature that standardized methods can reduce social influences in decision-making, as well as the importance of equal decision influence and anonymity is reinforced.

Future research

Following the findings of this research, the following implications for further research are suggested:

- **Targeting alternative reasons for adverse tumor assessment**
  The abandonment rate of pancreatic cancer surgeries is influenced by various factors beyond tumor assessment based on CT images. Future research could explore other causes of misdiagnoses and their roles in surgery abandonment, such as the impact of radiologist presentation and medical image quality
- **Testing alternative issues related to PC treatment**
  Future research could explore the use of ASI in addressing more complex and ambiguous medical decision-making issues. These issues could involve diverse interpretations from multiple disciplines and may not have a clear 'ground truth' but require consensus. Potential areas for this research include complex surgical interventions, multi-disciplinary treatment plans, and ethical considerations.
- **Demonstrating the effect of ASI and Social Influence** The last suggested avenue for further research is to investigate the causal relationship between ASI use and its impact on social loafing, social bias, and other behavioral factors. Understanding the effect of ASI on social influence contributes to a more objective decision-making process, even if it does not directly affect tumor assessment outcomes.

Contents

List of Figures

List of Tables

Glossary

Abbreviations

Abbreviation	Definition
ASI	Artificial Swarm Intelligence
AI	Artificial Intelligence
BRPC	Borderline Resectable Pancreatic Cancer
CDSS	Clinical Decision-Support Systems
CT	Computed Tomography
IT	Information Technology
LAPC	Locally Advanced Pancreatic Cancer
LUMC	Leiden University Medical Center
MDT	Multi-disciplinary teammeetings
MRI	Magnetic Resonance Imaging
ML	Machine Learning
PC	Pancreatic cancer
RQ	Research question

1 Introduction

Opening

This chapter elaborates on the core motivation of this research. By establishing the knowledge gap and explaining the relevant core concepts in the problem context, this chapter sets the foundation for introducing this research's objective and main question. This synthesis table provides an overview of relevant studies that identify relevant benefits and limits currently encountered in MDT meetings.
This chapter is structured as follows:

- Problem context
- Research objective and main research question
- Novelty of research
- Societal relevance
- Link to CoSEM program
- Thesis outline

1.1. Problem context

1.1.1. Pancreatic cancer and adverse surgery cases

With over 400.000 deaths in 2018, pancreatic cancer (PC) is among the leading causes of cancer-related deaths worldwide (Bray et al., 2018; Ferlay et al., 2019). While the current global 5-year survival rate for PC is less than 9% (Veldhuisen et al., 2018), current projections expect PC to become the second-leading cause of cancer death by 2030 in the United States (Khalaf et al., 2021; Rahib et al., 2014). Fundamental to the low survival rate are several key factors. To begin with, the symptoms associated with PC often present late which allows the tumor to have developed towards an advanced stage Kamisawa et al. (2016), Park et al. (2021), and Swords et al. (2015). This complicates the opportunity for survival since a delay in time from symptom onset to diagnosis has previously been shown to be a significant independent prognostic factor for survival (Swords et al., 2015). Furthermore, the number of symptoms displayed and the prominence of various symptoms for PC vary substantially (Hartwig et al., 2013). This, combined with the fact that many of these symptoms overlap with those of numerous benign abdominal conditions, has been suggested to be responsible for a delay in diagnosis in many patients with PC (Swords et al., 2015). Moreover, the aggressive nature of pancreatic cancer exacerbates the prognosis (Kamisawa et al., 2016). The rapid growth and spread of pancreatic tumors, often occurring before the disease is even detected, limit the window for effective intervention and treatment. This inherent aggressiveness, characterized by early metastasis and resistance to conventional therapies, further contributes to the poor survival outcomes for patients with PC.

As a result of these challenges, only 20% of the patients with PC are found eligible for curative resection at the time of diagnosis (Swords et al., 2015; Veldhuisen et al., 2018). However, within the population of patients having surgical resection, clinical registry data shows that in 19% to 33% of all cases the surgery is prematurely abandoned due to findings of locally advanced or metastatic disease (Section 4.2) (Allen et al., 2016; Hartwig et al., 2013; Hong et al., 2020; Strobel et al., 2019). In these adverse cases, the patient is taken into surgery for resection of the tumor, but the surgery is then prematurely abandoned as the tumor has developed to a stage where it can not be surgically removed due to unexpected disease spread or metastasis. Moreover, in the cases where the surgery operation is completed, 20% of all surgeries do not achieve a clear negative resection margin - implying that there are still malignant tumor cells present in the surrounding tissue of the patient even after the tumor has been surgically removed.

1.1.2. Multi-disciplinary team meetings

The early abandonment rate of PC surgeries imposes critical reflection on the decision-making process before surgery. To establish the optimal treatment strategy for an individual patient, multidisciplinary team (MDT) meetings are ideally organized by an expert panel - typically including radiologists, surgeons, oncologists, and nurses - as an important platform for surgical decision-making (Kee et al., 2004; H. K. Lim et al., 2016;

Maharaj et al., 2021). Commonly, an MDT meeting is conducted by collectively discussing and assessing patient portfolios, which typically include patient portfolios, health records, medical images, and other medical data (Fehervari et al., 2021; Kagan, 2005). By collecting experts with different specializations, in-depth team discussion enables the aggregation of all individual wisdom to arrive at the optimal patient treatment (Kagan, 2005). In this process, T. Gupta (2007), Maharaj et al. (2021) and Fehervari et al. (2021) also emphasize the excellent learning opportunity offered by MDT meetings across different disciplines, as well as for junior doctors by observing other specializations and senior colleagues.

However, despite MDT meetings having evolved to a standard of care (H. K. Lim et al., 2016), three MDT limitations can be identified in MDT evaluation literature. Firstly, although MDT meetings have been adopted by the majority of hospitals worldwide, the majority of papers acknowledge the difficulty of adopting a standardized protocol for structuring MDT discussions (T. G. Gupta, 2007; Lamb et al., 2012; Maharaj et al., 2021). Consequently, the selected specializations, as well as the procedure in which meetings are conducted, varies greatly across different healthcare settings and countries (Fehervari et al., 2021; Maharaj et al., 2021), also leading to varying performance and strongly impeding the ability to maintain complete focus on the patient (Lamb et al., 2012). Secondly, MDT meetings may commonly be sensitive to dominant individuals imposing their opinions based on their expertise or position within the hierarchy (Kagan, 2005; Kee et al., 2004). Following the aforementioned difficulty of structuring MDT meetings, T. G. Gupta (2007) states that no justified method exists for arriving at a collective decision. Resultingly, the presence of strong personalities may result in a biased outcome that differs from the optimal outcome when determined through an objective method or tool. Thirdly, since MDT meetings are typically managed by a chair, Fehervari et al. (2021), Kagan (2005) and Maharaj et al. (2021) observed that the quality of the discussions and management of divergent opinions may be largely dependent on the skills of the chair. Subsequently, Lamb et al. (2012) and Maharaj et al. (2021) observe that the input of some specializations - commonly palliative specialists and nurses - is regularly being overshadowed throughout MDT discussions.

1.1.3. Using Clinical Decision-Support Systems to improve MDT performance

To improve the performance of medical decision-making in MDT meetings, clinical decision-support systems (CDSSs) have been increasingly developed which assist the MDT meeting in arriving at substantiated, more accurate decisions in treatment strategies. Two notable applications include NAVIFY and Oncodoc (Hammer et al., 2020; Patkar et al., 2012; Séroussi et al., 2001). Although these applications show varying benefits on the efficiency at which MDT meetings are conducted, analysis of the literature concludes that these applications mainly assist MDT meetings by collecting, documenting, and presenting patient data before and during MDT meetings (Hammer et al., 2020; Patkar et al., 2012). Consequently, these applications do not directly impact discussion outcomes.

However, recent technological advancements enabled the further development of AI-driven evidence-based support systems such as MATE and Watson for Oncology (WFO) (Patkar et al., 2012; Somashekhar et al., 2018). This category of applications provides evidence-based therapy recommendations based on various sets of clinical data including published practice guidelines, scientific evidence, drug approvals, and treatment guidelines (Patkar et al., 2012; Somashekhar et al., 2018). To this end, WFO, which is the first prominent application based on Artificial Intelligence, continuously indexes literature, protocols, and patient charts (Kim et al., 2020). Across different cancer types and throughout different healthcare settings, several tests throughout the literature have shown a significant concordance (degree of agreement) of over 90% in treatment planning between the evidence-based CDSSs and MDT meetings (Kim et al., 2020; Patkar et al., 2012; Somashekhar et al., 2018). However, this category of applications still brings some notable limitations for reducing the abandonment rate of PC surgeries. Firstly, Kim et al. (2020) attributes the high concordance particularly to metastatic cases where the choice of therapy is relatively simple. Therefore, these CDSSs do not support addressing the main issue of PC where surgeries are abandoned due to unnoticed or unrecognized symptoms in the early tumor stages. As a result thereof, Kim et al. (2020) states that these CDSSs cannot deal with the complexity of all the patient-related factors (co-morbidity, insurance, socioeconomic state) associated with the treatment decisions at the early stage. Secondly, the effectiveness of its output is highly dependent on the data accuracy fed to the system (Somashekhar et al., 2018). Subsequently, Wang et al. (2023) and

Kim et al. (2020) have observed that such applications have difficulty in properly adjusting recommendations based on case history, patient factors, and anatomical variations.

1.1.4. Artificial Swarm Intelligence

Another potential technology to reduce the abandonment rate of PC surgeries is Artificial Swarm Intelligence (ASI). ASI is a state-of-the-art field of Artificial Intelligence that enables humans to collectively and efficiently arrive at a real-time decision through decentralized decision-making behavior and interaction (L. B. Rosenberg, 2015; Willcox et al., 2020). By combining the experience and knowledge of each individual, the technology amplifies the collective intelligence of a group by enabling its individuals to think and act as a collective mind. Correspondingly, this technology is observed to significantly improve overall decision accuracy in various situations (Metcalf et al., 2019; L. Rosenberg & Willcox, 2019b). Whilst ASI has proven to increase decision accuracy and improve consensus, the tool is also foreseen to eliminate decision bias due to decision hierarchy, social influence, and power imbalances (L. Rosenberg & Willcox, 2019b).

Examination of the early research on ASI in medical applications shows that the technology has significant potential. Most notably, L. Rosenberg et al. (2018) investigated the use of ASI to assess chest X-rays for the presence of pneumonia, accordingly observing a 33% increase in diagnostic accuracy when radiologists collaborated through ASI versus assessing the MRIs individually. Similarly, R. Shah et al. (2023) examined the elevated performance of ASI over majority voting in detecting meniscal lesions on knee MRI scans, thereby conceiving a 23% increase in accuracy in both individual and majority vote decisions. On top of this, both L. Rosenberg et al. (2018) and R. Shah et al. (2023) both carried out a comparison between two deep-learning systems (CheXNet and V-Net, respectively), concluding that ASI can outperform state-of-the-art AI models in terms of diagnostic accuracy.

However, although the aforementioned papers may show promising results, current studies on ASI in the medical field remain solely limited to controlled performance tests where the improved performance of ASI was assessed against the individual performance of each participant. Accordingly, this does not reflect a real-world medical setting where complex decision-making is conducted in MDT meetings by physically gathering experts to discuss a set of cases. Furthermore, the controlled environments in which ASI has currently been tested prevent any physical interaction among participants and thus exclude any process-related factors that may influence the actual outcome in a real-world MDT meeting.

1.1.5. Knowledge Gap

In treating patients diagnosed with PC, adverse tumor diagnosis can be considered a core problem for reducing the abandonment rate of PC surgeries. Currently, 10% to 15% of the surgeries lead to adverse outcomes: Whereas the patient is initially expected to have the pancreatic tumor surgically removed, the surgery is prematurely abandoned due to findings of advanced or metastatic disease. When examining the decision-making process leading to the adverse tumor diagnosis, two particular shortcomings of MDT meetings can be demonstrated in literature through their regular form. First, several papers acknowledge the lack of standardization in the way MDT decision-making is organized. Accordingly, the literature states that this can allow for social influence to emerge as dominant individuals tend to push through their own opinions. Furthermore, the lack of rationality in the decision process arising from the subjectivity of a chair can impact the performance of an MDT meeting.

While ASI can serve as a potential solution to these issues, research on the technology within medical decision-making currently remains insubstantial. In the present body of research, two studies are known to have tested the technology in medical decision-making. However, these studies have only tested ASI in controlled settings where its performance was compared against individual performances where each participant provided his or her opinion in isolation. Therefore, this may impose a gap between the tested, controlled environments and the use of ASI in a real-world MDT setting, since the controlled, isolated setting completely eliminates the fundamental nature of MDT decision-making by amplifying multi-disciplinary knowledge and combining diverse expert opinions through discussion. Therefore, it is currently unknown what the actual impact is on the decision-making process and accuracy when using ASI in a real-world MDT setting.

To substantiate the evidence presented in the problem context and knowledge gap, a synthesis table can be found in Appendix A that compares the current benefits and limitations of MDT meetings demonstrated in the literature. In terms of benefits, MDT meetings improve patient care by combining diverse medical expertise and providing a platform for professional development by allowing healthcare professionals to share knowledge. However, besides the aforementioned sensitivity to social influence, resource constraints and logistical issues also hinder effective MDT operations. Additionally, there is a lack of standardized guidelines for MDT composition and operation, leading to variability in implementation.

1.2. Research objective and main research question

In addressing the identified knowledge gap, this study conducts an exploratory research by applying ASI in the context of surgical oncology decision-making. The objective of this research is to examine whether ASI can improve the surgical decision-making process, by assessing the impact of using ASI over discussion-based MDT meetings on the accuracy of tumor assessment and reduced social influence. Using the PICO standard (Huang et al., 2006) for setting up clinical questions in the field of evidence-based medicine, the main research question has been formulated as follows:

In patients with potentially resectable PC, how does using ASI compared to discussion-based MDT meetings impact the tumor assessment accuracy and social influences in an MDT setting?

Following the aforementioned main research question, the sub-questions for this research have been formulated as follows:

1. *RQ1: What are the reasons for adverse tumor assessment in PC?*
   As the fundamental motivation of this research (Subsection 1.1.5), this section aims to identify the current evidence behind PC tumor diagnosis and the reasons for adverse tumor assessment. By identifying the factors that contribute to adverse tumor assessment, the results can be used as the foundation for targeting the setup and use of ASI in the following research questions.
2. *RQ2: To what extent does the accuracy of resectability assessment for PC cases change after the use of ASI in MDT meetings?*
   Based on the first benefit of ASI to medical decision-making identified in Subsection 2.3.5, this question aims to test the improved decision accuracy by collecting empirical data on the improved assessment accuracy of an MDT meeting after using ASI. To this end, the following hypothesis is formulated:
   H_1: In determining the resectability of patient X, the use of ASI in MDT meetings results in a reduced probability of adverse tumor assessment compared to discussion-based MDT meetings.
3. *RQ3: Can the use of ASI impact social influence during MDT meetings?*
   Identified as the second benefit of ASI to medical decision-making in Subsection 2.3.5, this question aims to evaluate the potential capability to reduce social influence in MDT meetings. This is evaluated by collecting empirical data on users based on their interaction with ASI in an MDT setting. For this evaluation, the following hypotheses are formulated:
   H_2: The use of ASI in MDT meetings can drive the reduction factors evaluation, attractiveness, and equal efforts to expedite reducing social loafing on decision outcome compared to discussion-based MDT meetings.
   H_3: The use of ASI in MDT meetings can drive the reduction factors of anonymity and equal decision influence to expedite reducing social bias on decision outcome compared to discussion-based MDT meetings.

1.3. Novelty of research

Profoundly, this research contributes to the existing body of scientific knowledge as the first study to apply ASI in a real-world MDT setting through a field test. To this end, an experiment is set up to simulate a real-world MDT meeting involving qualified medical experts in the field of PC (including surgeons, oncologists, doctors, and others) assessing past PC cases from patients previously treated at LUMC. By using ASI in a real-world MDT setting, this research advances on the existing studies that solely tested ASI in a controlled and isolated setting where participants do not cooperate but rather work in isolation (as previously identified in Subsection 1.1.5). Correspondingly, this unique approach intends to provide performance results that would resemble the real-world use of ASI in MDT meetings more closely, since the experiment compares the swarm performance against the performance of a simulated MDT meeting that collectively decides on the tumor resectability. In addition, using ASI in a practical MDT setting also enables empirical examination of its impact on the decision process and its social influences. This builds on the current body of knowledge, which is limited to a theoretical comparison of the decision outcomes and neglects other factors that indirectly impact the decision-making process. Altogether, the method employed in this research enables the collection of empirical data that is not yet gathered in existing research.

1.4. Societal relevance

By exploring the utilization of ASI in MDT meetings for PC diagnosis, this research holds three particular societal implications.

Tumor diagnosis for PC

The integration of ASI in MDT meetings can presumably enhance the accuracy of tumor assessments, leading to improved treatment planning and potentially higher survival rates for PC patients. On the other hand, accurate diagnoses can also reduce the abandonment rate of surgeries. From a patient perspective, this can attenuate the number of adverse cases whereby the patient sanguinely is being taken into surgery, but after waking up receives a palliative prognosis with a limited life expectancy of months (Veron Sanchez et al., 2023; Yadav et al., 2021). From a clinical perspective, this allows for enhanced quality of care by improved focus on curable patients - thus optimizing resource allocation, reducing healthcare operational costs, and thereby contributing to a more efficient healthcare system (Al Harbi et al., 2024).

Medical ethics

The use of ASI addresses critical ethical considerations in medical decision-making. By leveraging this technology, biases, and power imbalances that typically manifest in MDT meetings can be mitigated, thus ensuring a more equitable and fair decision-making process. ASI enables the inclusion of diverse expert opinions in a balanced manner, thereby preventing differences in personalities leading to varying decision influences in outcomes. Furthermore, the accurate and unbiased medical decisions facilitated by ASI uphold the ethical mandate of providing the best possible patient-centered care. Ensuring that every patient case is evaluated impartially and comprehensively aligns with the core ethical principles of justice, beneficence, and non-maleficence in medical practice (Della Croce, 2023).

Technology and innovation

This research highlights the transformative potential of AI technologies in complex decision-making scenarios in healthcare (Miller, 2024), particularly in the healthcare context. By demonstrating the efficacy of ASI in healthcare decision-making, this research may motivate further interdisciplinary research and development for technological advancements that extend beyond medical applications. Accordingly, the success of ASI in MDT meetings can drive similar innovations in other fields where collective decision-making is crucial, such as environmental management, education, and public policy. The wide range of cross-disciplinary influence highlights the broader societal relevance and potential impact of advancing AI technologies in complex decision-making processes.

1.5. Link to CoSEM program

The CoSEM program offered by TU Delft focuses on the establishment of complex, socio-technical systems based on a technical, institutional, and process perspective. As proposed in complex systems engineering literature (Baxter & Sommerville, 2011; Engelen, 2024; Gao et al., 2021; Lyytinen & Newman, 2008), the performance of such systems depends on the interaction and joint optimization of both social (institutions/structure and people/actors) and technical (technology and process) elements (Figure 1.1).

In investigating how ASI can benefit surgical oncology decision-making for patients diagnosed with PC, this research deploys a complex systems engineering perspective by testing the technology (ASI) using a holistic approach: Rather than focusing on the performance of the technology in isolation, it is also assessed how its use interacts with the involved actors, affects the decision-making process and its potential integration in organizational structures (Engelen, 2024; Gao et al., 2021). First, the technology in interaction with its users (medical experts) is tested to assess to what extent the interaction leads to improved overall performance. Second, how the technology affects existing (decision) processes is assessed by evaluating its effect on social influence and by identifying the potential enablers and barriers to the integration of the technology in MDT meetings. Third, the current structure of the MDT process is embraced by embedding ASI in the current protocol and structure of an MDT meeting.

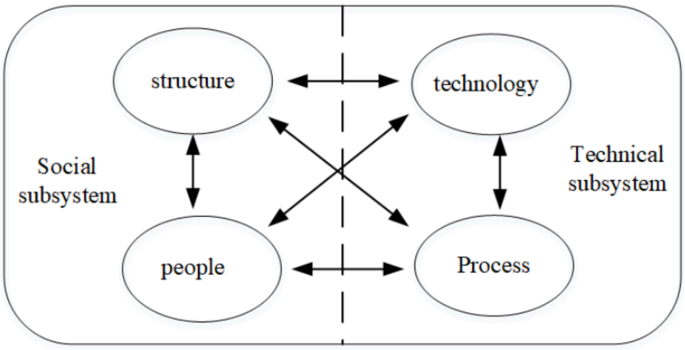


Figure 1.1: Diagram showing the interplay of the social and technological subsystems and their elements (taken from Gao et al. (2021))

1.6. Thesis outline

With the problem context, knowledge gap, main research question, and relevance of this research now documented in this introduction, the remainder of this research is structured as follows. To begin with, Chapter 2 introduces the necessary background to substantiate the topics of ASI and social influence. Thereafter, Chapter 3 presents the research design including the necessary datasets, the associated research methods, and ethical responsibility. Subsequently, Chapter 4 explicates the current evidence behind PC surgery and resectability assessment. Following this evidence, Chapter 5 and Chapter 6 present the experimental results of the performance analysis and social influence analysis, respectively. Based on these results, Chapter 7 provides an interpretation of the experimental findings, as well as the implications for further research. Lastly, Chapter 8 provides the findings of each research question to then conclude the main question of this research.

2 Literature Review

Opening

Recap

Previously in Chapter 1, the problem context, objective, and research question for this project was substantiated along with an outline showing the structure in which this research is presented. Furthermore, the problem context introduced several core concepts to this research along with several key takeaways within each concept as follows:

- **MDT meetings**
  - An important platform for surgical decision-making
  - Combines expertise and opinions from various backgrounds
  - Sensitivity to social influence (dominant individuals) resulting from lack of standardized protocols
  - Dependability on subjectivity and skills of chair
- **ASI**
  - An emerging field of AI that amplifies the human intelligence of a group by combining different expertise and knowledge
  - Shown to improve decision accuracy in various situations and improve consensus, while also eliminating decision biases and social influences
  - No existing literature evaluating the practical use or true performance when the ASI is applied in medical decision-making

Introduction

This literature review aims to substantiate on the aforementioned concepts to provide the necessary background for the remainder of this research. Due to the complexity of each concept, separate reviews have been conducted for a more profound background of each concept. An individual motivation for each review is introduced at the beginning of each section. This chapter first introduces an overall framework of all reviews, after which each review is separately presented in the same order as their knowledge is applied throughout the remainder of this research. Accordingly, the review is structured as follows:

- Framework of literature review
- Review of ASI fundamentals
- Review of social influence in MDT meetings
- Summary of literature reviews

2.1. Framework of literature review

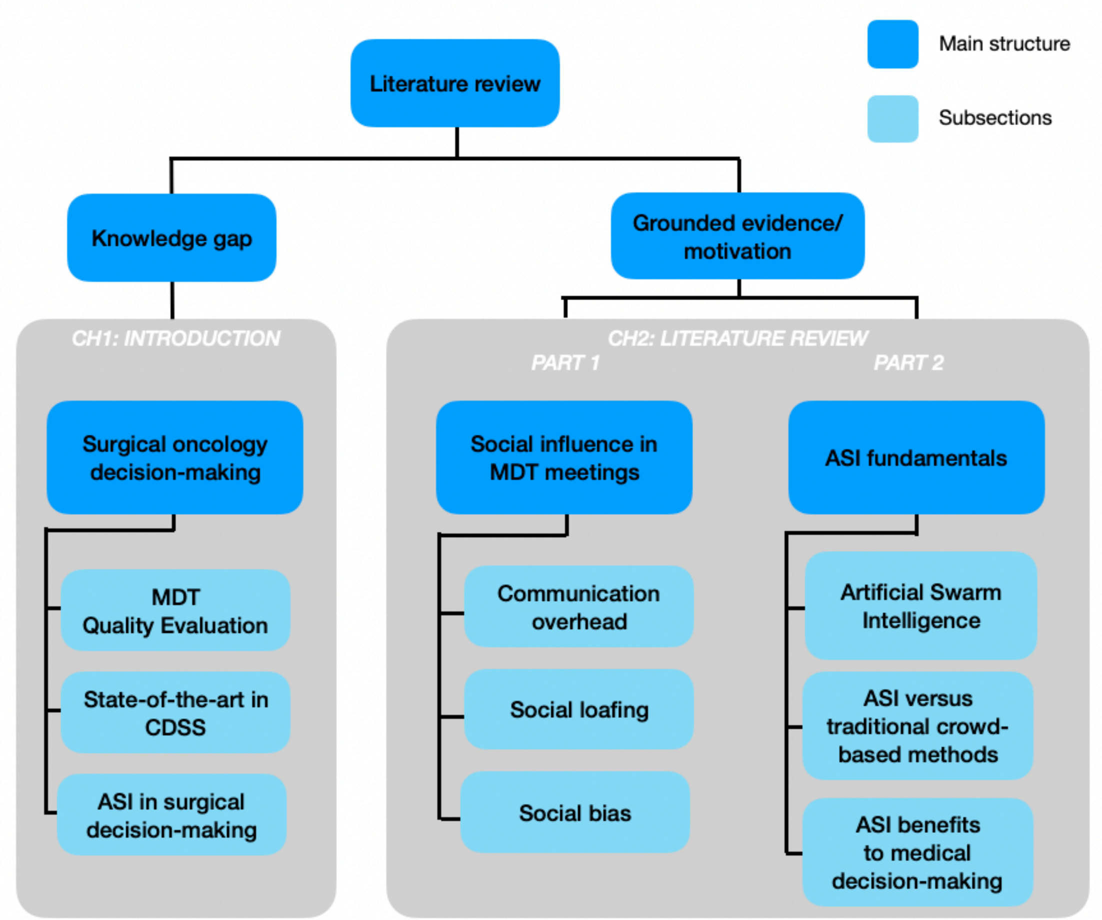


Figure 2.1: Literature review framework

This chapter establishes the core motivation for this research based on a comprehensive literature review. At its core, the review is centered around two components serving the following objectives (Figure 2.1):

- To identify the **knowledge gap** in the current state-of-the-art in MDT research
- To substantiate on the **grounded evidence** of the relevant concepts for this research

Each (sub)section of the two components is presented as follows:

1. Selection process
2. Review

Furthermore, the synthesis tables for the review on surgical oncology decision-making, social loafing, and social bias in MDT meetings can be found in Appendix A, Appendix B, and Appendix C, respectively. The synthesis appendix is therefore structured as follows:

- **Appendix A: Surgical Oncology decision-making**
  - A1: Synthesis table overview
  - A2: MDT Quality evaluation
  - A3: Existing solutions
  - A4: ASI in surgical decision-making
- **Appendix B: Synthesis table of social loafing**
- **Appendix C: Synthesis tables of social bias**

2.2. Review of social influence in MDT meetings

Opening

Recap

Previously in Chapter 1, MDT meetings were introduced as the main platform for surgical decision-making. However, the sensitivity of MDT meetings to social influence was suggested as one of the current limitations of MDT meetings. This was observed to lead to the following consequences:

- An important platform for surgical decision-making
- Combines expertise and opinions from various backgrounds
- Sensitivity to dominant individuals resulting from lack of standardized protocols
- Dependability on subjectivity and skills of chair

Motivation

As one of the fundamental motivations for setting up a field test for ASI in an MDT setting, this review aims to substantiate social influences that may significantly affect the decision-making process at an MDT meeting. The objective of this review thereby is to identify factors that can reduce these social influences in social settings. By selecting the factors that contribute to this reduction and understanding how they can be fostered, these factors can then be used as evaluation criteria after the experiment to assess the effect of ASI on each factor. In other words, the identified factors in this review serve as the basis for evaluating the impact of ASI on reducing social influence.

To substantiate the selected social influences, this review first begins with a definition and scope of the topic. Using this foundation, the remainder of this review is structured as follows:

- Definition and scope of social influence
- Selection process
- Communication overhead
- Social loafing
- Social bias

2.2.1. Definition and delineation of social influence

For this research, social influence is defined as follows:

”Any process whereby a person’s attitudes, opinions, beliefs, or behavior are altered or controlled by some form of social communication. It includes conformity, compliance, group polarization, minority social influence, obedience, persuasion, and the influence of social norms.” [1]

In this respect, it is important to note that existing literature predominantly refers to the same concept using the term *social bias*. For this research however, social influence is used as the overarching term whilst social bias is regarded as a subset of social influence as this is perceived to better fit the context of this research. For scoping relevant social influences for this research, preliminary research was conducted to identify social influences that are suggested to be particularly significant in MDT meetings using the following process.

1. Papers were collected using the following search string in Google Scholar: *”TITLE-ABS-KEY(“Social bias” AND “medical” AND “group decision-making”)”*
   As explained, despite the term ”social influence” used for this research, ”social bias” is used as the search string since this is observed to be the dominant term in academic literature
2. Papers were selected when based on the following criteria:
   - Paper is published after 01 January 2024
   - Paper examines two or more distinct social biases

[1] https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100515197

- Paper examines social biases that are potentially relevant to examine within the timeframe, scope, and resources of this research
- Paper examines social biases that are assumed to be directly resolved or impacted by ASI

3. The resulting selection included the following papers Choi et al. (2023, 2024) and Mannion and Thompson (2014)

Although the exact terminology rather varies across the papers, social bias, social loafing, and communication overhead are found as the overarching concepts based on the resulting selection. To adequately substantiate each of the three effects, this review explores each effect using a distinct search string formulated to cover each effect as comprehensively as possible. The explanation for each string is given hereafter in Subsection 2.2.2.

2.2.2. Selection process

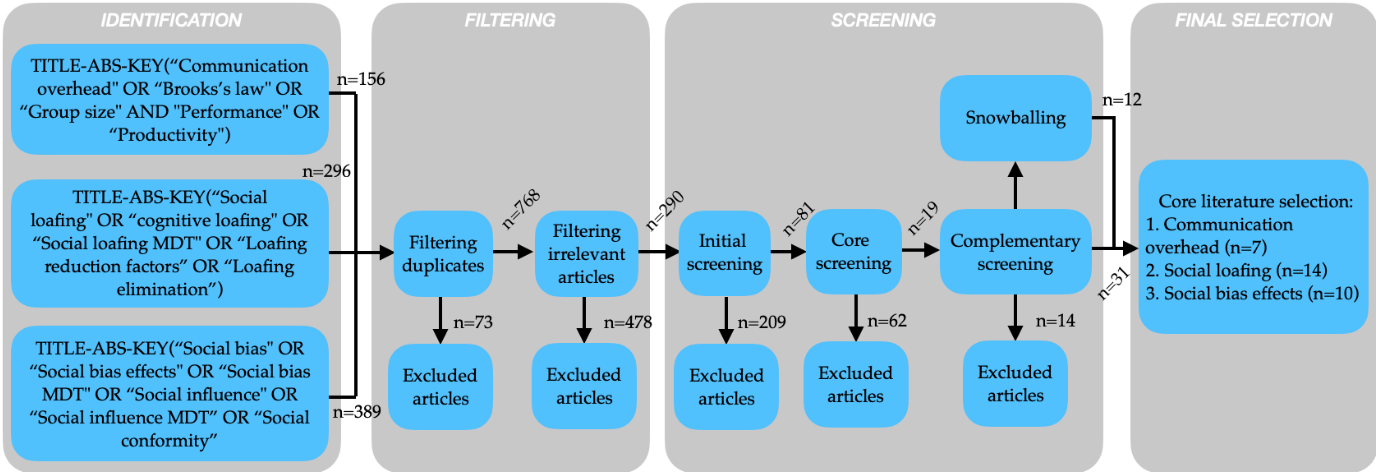


Figure 2.2: Overview of selection procedure

On 23 March 2024 (communication overhead), 24 April (social loafing), and 08 May 2024 (social bias effects), a systematic search was carried out using scientific databases Medline, Embase, Google Scholar, and IEEE. A schematic overview of the selection process is depicted in Figure 2.2 and, consistently across each subtopic, consists of the following phases:

1. **Identification**
   To determine the state-of-the-art across three topics, the initial collection of literature was composed by use of three search terms.
   (a) *"TITLE-ABS-KEY("Communication overhead" OR "Brooks's law" OR "Group size" AND "Performance" OR "Productivity")"*:
   This search string is intended to balance specificity and breadth by combining well-known theories and broader terms related to communication overhead. "Communication overhead" and "Brooks's law" are included since the preliminary review shows both terms are used interchangeably in literature for describing the same concept in group decision-making literature. The inclusion of "Group size" alongside "Performance" and "Productivity" ensures that the search addresses both the structural aspects of team composition and their direct effects on team output and efficiency. In this respect, productivity and performance are both included as each term is used interchangeably in literature as well.
   (b) *"TITLE-ABS-KEY("Social loafing" OR "cognitive loafing" OR "Social loafing MDT" OR "Loafing reduction factors" OR "Loafing elimination")"*:
   Since a preliminary review showed that social loafing predominantly has two forms (physical versus cognitive), "Social loafing" (as the overarching concept) and "cognitive loafing" (as a subset) are both included to combine a breadth search with a depth search specified to the relevant situation for MDT meetings. Moreover, based on the aim to also identify how social loafing can be

reduced, ”loafing reduction factors” and loafing elimination” are both included as they are used interchangeably in literature for this purpose.

(c) *”TITLE-ABS-KEY(“Social bias” OR “Social bias effects” OR “Social bias MDT” OR “Social influence” OR “Social influence MDT” OR “Social conformity”)”*:
Since a preliminary review reveals that social bias is a more generalized concept with a significantly larger body of research compared to social loafing, this search string also includes ”social bias MDT” and ”social influence MDT” to specify the results with respect to papers relevant to MDT settings. Also, as part of the larger body of research, the concept is more frequently referred to using alternative terms. Therefore, besides its overarching term ”social bias”, additional terms are selected that are commonly used interchangeably, which consist of ”social bias effects”, ”social influence” and ”social conformity”.

For simplicity reasons, the visualization of the selection process (Figure 2.2) provides a combined overview as each of the three searches progresses through mutual selection phases. However, each search was performed separately for each of the three subtopics. Accordingly, the search results from the identification phase toward the filtering phase can be interpreted as an aggregation formally expressed as combining the three search strings using an OR statement.

2. **Filtering**
   For filtering the initial collection, all papers that meet the following criteria were excluded:
   - For communication overhead and social bias effects: Papers published before 1 January 2003 (except for the original papers by Brooks (1995), Fried (1991), Galton (1907) and Janis (1982))
   - Paper is published in non-English and non-Dutch language
   - Paper is a conference paper
3. **Screening**
   The screening phase consisted of three components:
   - The initial screening was done by specifically filtering papers that were written as scientific papers or sections of books. The papers that were excluded consisted of conference papers
   - The core screening was done by scanning the resulting list of papers specifically addressing one of the three aforementioned three key topics based on their title, abstract, and conclusions.
   - The complementary screening was carried out on the remainder of the papers by specifically filtering the papers that (1) describe the group size paradox, (2) explicitly demonstrate social loafing and reduction factors for social loafing and (3) clarify social bias effects. In addition, the selection was completed through snowballing using in-paper references
4. **Final Selection**
   The final selection was completed by combining both the selected papers from the complementary screening and the snowballing process (Badampudi et al., 2015) to a cumulative of 31 papers

2.2.3. Synthesis

The synthesis table of the selected literature on social loafing and social bias can be found in Appendix B and Appendix C. To summarize the main findings of the selected literature on social loafing and social bias, an overview can be found below:

- **Synthesis table for social loafing**
  This table summarizes research on social loafing, where individuals exert less effort when working in a group compared to working alone. The key takeaways include the following:
  - **Cognitive loafing experiments**
    * (Petty et al., 1980): Demonstrated higher cognitive effort and better performance in individuals compared to groups.
    * (S. Harkins & Petty, 1982): Found that identifiable individual responses reduce loafing, especially on easy tasks.
    * (Weldon & Gargano, 1985, 1988): Showed that individuals produce more evaluations and

exhibit higher cognitive effort than groups.

  - **Loafing reduction factors**
    * Individual Identifiability: Making individual contributions identifiable reduces social loafing.
    * Evaluation of Individual Performance: Opportunities for performance comparison motivate individuals to contribute more.
    * Equal Perceived Group Efforts: High group cohesion and equal efforts enhance individual and team performance.
    * Task Attractiveness: Higher task attractiveness increases individual effort and reduces loafing.
    * Time-Restricted Tasks: Increased time pressure can enhance team workload and reduce loafing.

- **Synthesis table for social bias effects**
  This table shows the evidence on various social biases affecting group decision-making and ways to minimize these biases. The key takeaways include the following:
  - **Social bias effects**
    * Groupthink: Highly cohesive groups may prioritize unanimity over effective decision-making.
    * Social Dominance: Influence is attained through dominance (force) or prestige (skills).
    * Herding: Individuals conform to group opinions, especially when personal information is insufficient.
  - **Social bias minimization Factors**
    * Anonymity: Anonymity reduces peer pressure and stereotyping, encouraging more honest input.
    * Deliberation Engagement: Engaging in thorough discussions reduces the illusion of unanimity.
    * Equal Decision Influence: Ensuring equal influence among participants promotes fairness and reduces dominance-based biases.

2.2.4. Brooks's law: The group size paradox

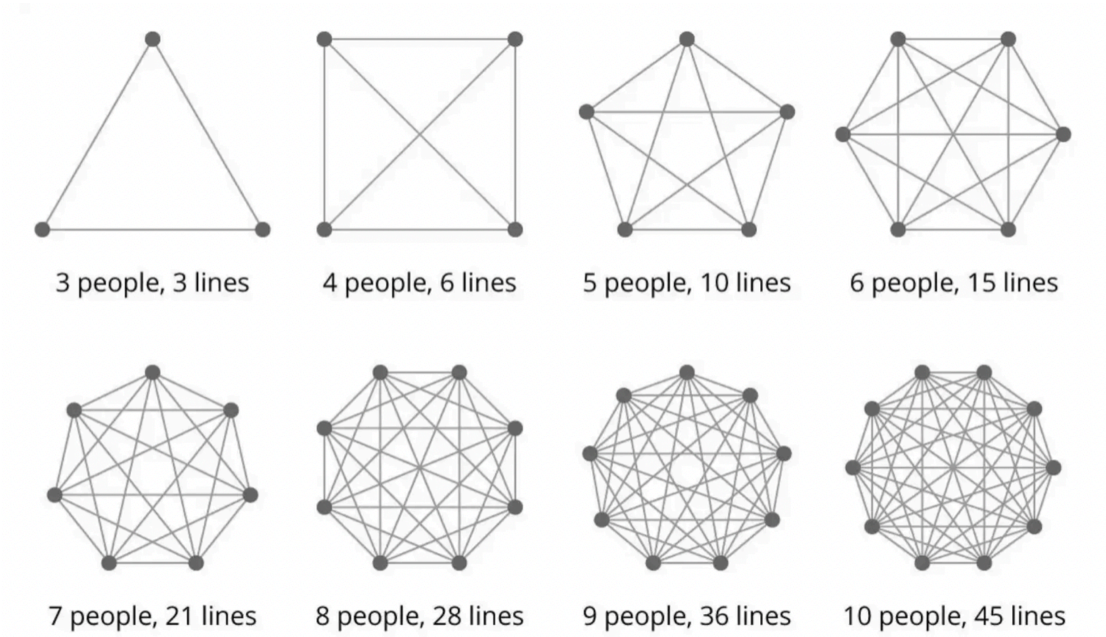


Figure 2.3: A schematic overview of the increasing number of communication lines for increasing group sizes (adapted from Brooks (1995)). The number of communication lines increases according to n(n-1)/2, where n represents the number of individuals. As can be deducted from the figure, a group size of 7 individuals leads to 21 communication lines (3x), while a size of 10 individuals leads to 45 communication lines (4.5x). This represents an exponential growth in communication lines, ultimately leading to communication overhead.

The lack of humans to be naturally capable of behaving on the principles of Swarm Intelligence can be mainly attributed to the fact that humans mainly convey opinions and knowledge through speech (Cristancho, 2021). Although speech enables humans to communicate complex affairs and ideas comprehensively to peers, maintaining all the necessary direct lines of communication in teams typically becomes difficult when group sizes increase (Staats et al., 2012). As originally shown by Brooks's law (Brooks, 1995), increasing group sizes leads to an exponentially increasing number of connections that need to be maintained between all individuals in a team (Figure 2.3). Resultingly, this can strongly restrict the opportunity for each individual to express his ideas or opinions to the group as a whole, leaving only a smaller part of the group able to communicate his ideas (Bechky, 2003). Given this constrained opportunity for only a selected few to speak out, this (time) restriction is at the foundation of the *law of diminishing returns* paradox (Fried, 1991): the point beyond which adding additional members yields progressively reduced output. Particularly from the perspective of MDT meetings, adding group members is typically believed to be associated with accumulated overall wisdom and expertise available in the group.

However, what is commonly overlooked is that, beyond some optimum, adding additional experts results in the increasing inability and restriction to verbally express their thoughts (Bechky, 2003; Staats et al., 2012). Resultingly, this leaves an augmented portion of the available expertise and opinion largely unnoticed, leading to redundancy and inactivity (Staats et al., 2012). When group size is further increased, this is what drives *communication overhead* (Kaufman, 2011): The proportion of time spent communicating with other members of the team at the cost of getting productive work done to achieve the overall goal. In the case of medical decision-making, conducting MDT meetings can arguably be regarded as what Brooks (1995) refers to as a *complex interrelationships task*: A process that requires many interpersonal dependencies. Subsequently, while enlarging the number of individuals in human teams is typically expected to accumulate the available knowledge and expertise within a group, collecting and aggregating the diverse opinions of all experts after some point contrarily leads to communication overhead and a counterproductive decision-making process (Figure 2.4) (Tornhill, 2019).

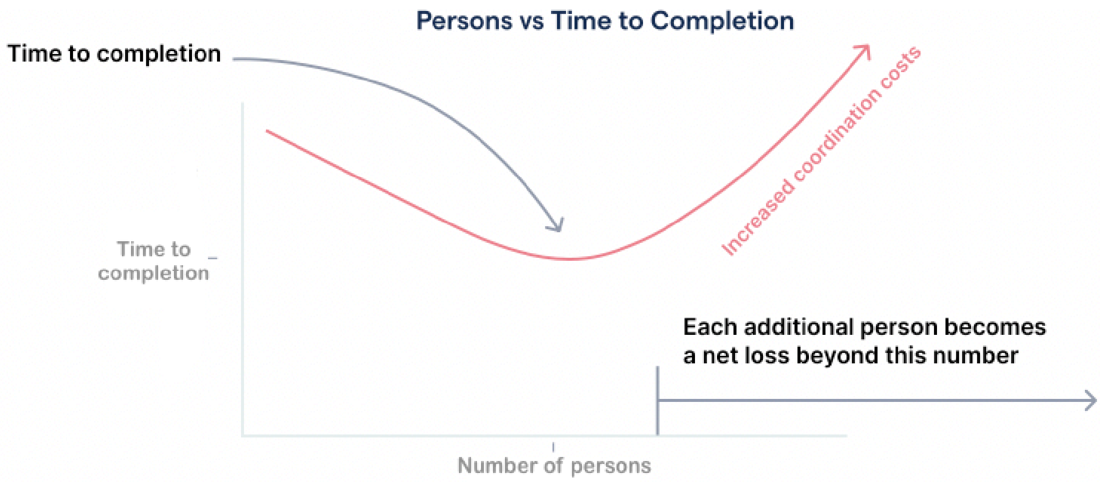


Figure 2.4: The overhead curve of complex interrelationships tasks (Tornhill, 2019). Initially, adding team members allows for spreading the workload of a task across an increasing number of individuals - increasing productivity (Total work per time unit). However, after some point, the portion of coordination and (mis)communication required among all individuals increases more significantly - thereby hampering overall efficiency (Total work over the amount of available resources (in this case: group members)).

2.2.5. Social Loafing

At the foundation of attenuated efficiency resulting from communication overhead is the phenomenon of *social loafing* (in literature also referred to as *free-riding*). Originally acknowledged as the Ringelmann-effect (Ingham et al., 1974; Liden et al., 2004), social loafing is referred to as 'the decrease in individual effort when performing in groups as compared to the individual effort when performing alone' (Kravitz & Lexington, 1986; Latane et al., 1979). Whereas the decreased effort was initially observed in a physical rope-pulling experiment with an increasing number of individuals by the French researcher Max Ringelmann (Kravitz & Lexington, 1986), Latane et al. (1979) later relabeled the Ringelmann-effect to social loafing after they observed a similar decline in individual motivation for team members performing cognitive tasks. Although this effect has been widely demonstrated in various tasks requiring physical efforts (including clapping and shouting (Figure 2.5) (Latane et al., 1979), pumping air (Kerr, 1981) and rope-pulling (Ingham et al., 1974), this review is focused on the efforts of cognitive tasks since this better aligns with the tasks performed during MDT meetings.

Social loafing in cognitive tasks, hereafter referred to as *cognitive loafing*, has been demonstrated in various experiments (Appendix C) after becoming a particularly popular field of research between 1980-1990. In their pioneering research, Petty et al. (1980) assigned a group of students to evaluate the performance of videotaped counseling psychologists, demonstrating that students believing their evaluation task was shared with many others produced fewer evaluative findings than those working alone. In their follow-up experiment, Petty et al. (1980) assigned a group of undergraduate students the task of peer-reviewing editorial essays written by other students, observing a significantly diminished thoroughness of arguments when the process was perceived as a shared responsibility. Afterward, S. Harkins and Petty (1982) first demonstrated similar results when comparing individual versus collective responsibility in brainstorming as many use cases as possible for an object. Secondly, S. Harkins and Petty (1982) tested the vigilance of individual and shared performance by reporting seldomly occurring visual signals on a TV screen, showing greater alertness when being solely responsible. Lastly, Weldon and Gargano (1985) carried out a multi-attribute judgment task, observing that individuals use more complex judgment strategies and collect more information to form a grounded judgment.

When considering the aforementioned conclusions in the context of MDT meetings, a considerable degree of resemblance can be observed with typical tasks and skills essential to MDT meetings. This includes (1) the comprehensive evaluation of medical images and other patient data, as well as thorough argumentation of experts medical opinions and convictions (as demonstrated by Petty et al. (1980)), (2) creativity in determining tailored treatment strategies for unique patient cases (similar to S. Harkins and Petty (1982)), (3) alertness in detected unusual, visible particularities (as shown by S. Harkins and Petty (1982)) and (4) (multi-)judgment tasks for interpreting unusualities in medical images (as investigated by Weldon and Gargano (1985). Based on these parallels, it can be argued that MDT meetings are social settings that are likely to be subject to cognitive loafing as well. This suggests the importance of identifying the underlying causes and resolutions to reduce its effects on the MDT outcome.

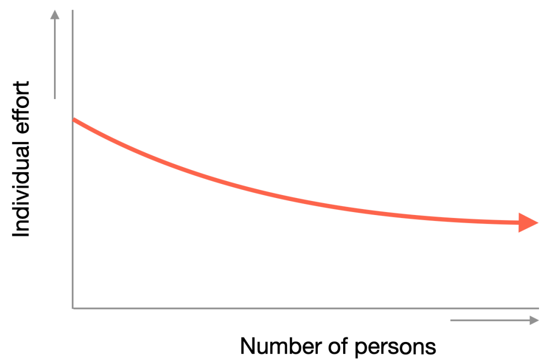


Figure 2.5: Reduced effort curve for cognitive and physical tasks. Whereas an individual typically exerts maximum effort when working alone, the individual effort is observed to reduce as group size increases and the individual can increasingly rely on his or her group members.

2.2.6. Reduction Factors to Cognitive Loafing

To reduce or eliminate social loafing, a wide range of reduction factors has been demonstrated in literature (S. Harkins & Petty, 1982; Jackson & Harkins, 1985; Karau & Wilhau, 2019; Liden et al., 2004; Mannion & Thompson, 2014; Petty et al., 1980; Zaccaro, 1984). This identifies the most significant resolutions proposed across the selected literature. From the perspective of MDT meetings and considering the potentially solvable resolutions by Swarm Intelligence, an overview of the relevant set of reduction factors is listed in Table 2.1.

Reduction factor	Description
Individual identifiability	When making individuals participate in a group identifiable, the efforts of an individual alone are observed to closely approach the efforts of the individual when working in a group (S. Harkins & Petty, 1982). According to Williams et al. (1981), identifiability eliminates the opportunity for individuals to "hide in the crowd".
Evaluation of individual performance	Another factor to reduce cognitive loafing is establishing a setup where the individual output can be evaluated against the output of the group (Karau & Wilhau, 2019; Xu et al., 2020). As stated by S. G. Harkins and Jackson (1985), the opportunity for comparison may lead the participants to believe that their performance can be evaluated, which motivates individual performance.
Task attractiveness	When individuals perceive the task as meaningful and attractive, individuals are less likely to loaf (Karau & Wilhau, 2019). In fact, according to (Zaccaro, 1984), the enhanced contribution arising from task attractiveness and meaningfulness commonly results in more peer pressure which is stated to increase as group size increases as well.
Time restricted tasks	To further reduce loafing behavior, another resolution is to impose time restrictions on the completion of the task (Cui et al., 2021; Luo et al., 2021; Sigauke, 2021). When tasks do not have a fixed deadline, time pressure is absent which allows for more opportunity for individuals to procrastinate (Luo et al., 2021; Sigauke, 2021).
Equal group efforts	When completing a task requires a collective effort, team cohesiveness has been demonstrated as an important factor in reducing loafing behavior (Høigaard et al., 2006; Karau & Williams, 1997). To this end, an important determinant is the opportunity for leveling (expectations of) efforts: Individuals show reduced loafing behavior when the expected or perceived efforts of their team members level with their individual perceived efforts (Jackson & Harkins, 1985).

Table 2.1: Factors reducing loafing behaviour

Although a large body of theories and frameworks have been established for identifying the effects of social loafing, the phenomenon remains by what Latane et al. (1979) refers to as a *social disease*: a negative effect (disease) for individuals and social institutions that is only observable by the presence or actions of other people (social). Hence, this provides a fundamental motive to establish empirical research to investigate social loafing in MDT settings.

2.2.7. Social bias effects

According to Mavrodiev et al. (2012), the *wisdom of the crowd* is defined as the phenomenon that the aggregate prediction or forecast of a group of individuals can be considerably more accurate than most individuals in the group, and typically – than any of the individuals comprising the group. Originally demonstrated by Galton (1907), he famously found at a farmer's fair that the independent estimations of the weight of a slaughtered ox from 784 farmers were more accurate than any of their individual estimations. Ever since, the effect has been widely studied across a wide range of social psychological research (Lorenz et al., 2011; Madirolas & de Polavieja, 2015; Mannes et al., 2014; Mavrodiev et al., 2012).

Despite its mathematical aggregation methodology based on independent and isolated opinions (Lorenz et al., 2011; Mannes et al., 2014; Mavrodiev et al., 2012), its rationality is demonstrated by a wide range of papers to suffer from various social bias effects observed when performing in groups (Kao et al., 2018; Madirolas & de Polavieja, 2015; Mannes et al., 2014; Mavrodiev et al., 2012). In this respect, a variety of cognitive and perceptual biases have been documented in which humans seemingly deviate from rational behavior (Kao et al., 2018), ultimately leading to a cumulative bias when aggregating all individual opinions to the collective outcome (Lorenz et al., 2011). Based on the literature selection (Appendix C), the following set of biases is selected with respect to their relevance to MDT meetings:

- **Group-think**
  According to Janis (1982), group thinking occurs when unanimity and harmony are preferred over the expression of an individual's true opinion. Hence, 'loyalty prevents each member to raise controversial issues' (Janis, 1982). In delivering high-quality and safe care, Mannion and Thompson (2014) and Lamb et al. (2013) state that group-think can militate against this when healthcare professionals feel uncomfortable in expressing dissenting views and are unwilling to speak openly about concerns. With respect to an MDT meeting, this is what typically leads to what Mannion and Thompson (2014) refer to as *illusion of unanimity*: when silence is interpreted as consent, making the group think it has made a 'unanimous' decision for instance in concluding cancer operability. Meanwhile, however, the group tends to conclude in 'false consensus' (Lamb et al., 2013).
- **Social dominance**
  In both cohesive and non-cohesive groups, a common observation in group discussions is what S. L. Lim and Bentley (2022) refer to as 'the loudest voice': more dominant personalities or extroverts overpower the modest individuals or introverts. In this respect, dominance can have multiple determinants including skills, expertise, and experience (Cheng, 2020; Ketterman & Maner, 2021). In the context of MDT meetings (Subsection 1.1.2), an example could be that more experienced experts (for instance surgeons and radiologists) having more influence on the decision outcome than other specializations (Kagan, 2005; Kee et al., 2004).
- **Herding**
  Herding, or social conformity, is the act of fitting within the group, manifested by the tendency of people to modify their opinions or judgments according to those of the group (Chen et al., 2022; Shamay-Tsoory et al., 2019; Zeigler-Hill et al., 2015). Whereas group-think inhibits individuals to express their opinion (Lamb et al., 2013), social conformity entails monitoring and copying the responses and actions of those we observe (Chen et al., 2022; Raafat et al., 2009). With respect to MDT meetings, herding may impose a significant effect on the outcome as discussion involves unequivocal exposure and consideration of other members: The open discussion setting (Subsection 1.1.2) of common MDT protocols allows for enhanced setting for individuals to adapt their opinion according to the group (Raafat et al., 2009).

2.2.8. Factors reducing social bias effects

Across the literature, a large variety of factors have been demonstrated to reduce the influence of social bias effects on collective performance. Considering the context of MDT meetings, the selection of factors deemed to be of particular interest to be served by Swarm Intelligence are listed in Table 2.2.

Reduction factor	Description
Anonymity	By deliberating anonymously toward group consensus, this property can eliminate peer pressure applied on or perceived by dissenters (Mannion & Thompson, 2014), causing them to inhibit their opposing opinions. Furthermore, anonymity can avoid 'stereotyping' (Mannion & Thompson, 2014): rejection of certain individuals based on their opposing views.
Deliberation engagement	To stimulate collective participation, specifically to those in the middle ground or less inclined to express their opinion by nature, engagement pushes every member to participate. This eliminates the aforementioned illusion of unanimity (Mannion & Thompson, 2014).
Equal decision influence	Establishing equal decision influence eliminates social bias effects arising from different personalities and traits (Cheng, 2020), which can neutralize dominant and submissive characters.

Table 2.2: Factors reducing social bias effects

2.2.9. Conclusion of social influence in MDT meetings review

Given the sensitivity of MDT meetings to social influence, this review substantiates the most important social influences as well as the most relevant reduction factors with respect to MDT settings. In conclusion, the literature on ASI reveals several key social influences that impact the effectiveness of MDT meetings. First, communication overhead highlights that increasing the number of group members can lead to a decline in overall productivity due to communication overhead, which complicates the maintenance of effective communication lines. Second, social loafing is defined as 'the decrease in individual effort when performing in groups as compared to the individual effort when performing alone'. This social influence can be mitigated by factors including individual identifiability, evaluation of individual performance, task attractiveness, time-restricted tasks, and ensuring equal group efforts. Lastly, social bias, which involves deviations from rational behavior due to cognitive and perceptual biases, can be reduced through anonymity, deliberation engagement, and ensuring equal decision influence.

Implications for this research

As stated in Subsection 1.1.2, MDT meetings are observed to be sensitive to social influence. To examine the impact of ASI on social influence in MDT meetings, this review substantiates the fundamentals of three particular social influences and the factors that reduce these influences. This review thereby constitutes the setup of this research as follows:

- Since these influences are only observed in social contexts, this serves as the fundamental basis for setting up empirical research in an actual MDT setting. Using an experimental setup, empirical data can be collected to determine whether ASI can contribute to the presented factors that reduce social influences.
- Besides substantiating the social influences themselves, identifying the reduction factors for social loafing and social bias is also key for setting up demonstrated criteria to evaluate the impact of ASI in reducing social influence. Therefore, these factors serve as the foundation for post-experimental assessment of the impact of ASI on social influence.
- While multiple social influences play a role in medical decision-making, this review elaborates on three influences in particular. From these three influences, the scope of this research is delineated to social loafing and social bias as the two social factors for further examination of this research. The third influence, communication overhead, is out of scope for this research due to its impact being particularly observable in larger group sizes (n>10).

2.3. Review of ASI fundamentals

Opening

Recap

Previously in Chapter 1, ASI was introduced as the core technology to be investigated in this research. The following takeaways were presented for testing this technology in improving decision-making in MDT settings.

- An emerging field of AI that amplifies the human intelligence of a group by combining different expertise and knowledge
- Shown to improve decision accuracy in various situations and improve consensus, while also eliminating decision biases and social influences
- No existing literature evaluating the practical use or true performance when the ASI is applied in medical decision-making

Motivation

As the fundamental technology around which this research is centered, this literature review examines the current evidence behind ASI to substantiate its fundamental features and benefits. Given the state-of-the-art status and the complexity of the technology, a dedicated review was conducted to provide a comprehensive background specifically focused on understanding of ASI and its strengths. This is deemed necessary for adequately understanding its potential usability for MDT meetings and the setup of this experiment. Altogether, the obtained findings of this literature review are presented as follows:

- Selection process
- Definition of Swarm Intelligence
- Artificial Swarm Intelligence
- ASI versus traditional crowd-based methods
- ASI benefits to medical decision-making

2.3.1. Selection process

To provide a comprehensive background to ASI, this review deployed a grey literature search. This method is conducted since there is currently strictly limited evidence available through traditional academic databases on ASI in the field of surgical oncology - particularly in the field of PC. Therefore, by broadening the research scope through a grey literature review, other sources such as reports, theses, and other (unpublished) documents can be included as well. On top of that, the inclusion of technical documents, best practice guidelines, or project reports is expected to also result in more practical insights into the technology.

Between 10 May 2024 and 21 May 2024, a grey literature search is carried out to collect research material

Figure 2.6: Selection process for ASI fundamentals review

for reviewing the current evidence behind ASI. In this process, the research material for this review is selected as follows (Figure 2.6):

1. **Scope of the review**
   The scope of this review is delineated according to the following criteria:
   - Focus of research material on the principles of ASI, its benefits over other methods, and its implications for medical decision-making
   - This review covers research articles, reviews, and relevant reports
   - The literature material for this review is published between 1 January 2005 and later
2. **Filtering**
   Literature material was included when meeting the following criteria:
   - The material is published in English or Dutch
   - The material was focused on the aforementioned themes including the foundational concepts of SI, the technological implications of ASI, and their comparative advantages over traditional crowd-based methods in decision-making
   - Articles solely focused on basic research, swarm optimization, and swarm in nature were excluded
3. **Screening and selection process**
   After filtering the relevant literature material, the titles, abstracts, and contents are reviewed to assess relevance, followed by a full-text review of the selected material. This process aims to ensure that the most relevant research material is included in the final review.

2.3.2. Definition of Swarm Intelligence

Figure 2.7: A visualization of Swarm Intelligence in nature: Individual agents interacting to create a flock of birds acting as a cohesive unity (Salmon, 2014)

According to L. Rosenberg and Willcox (2019b), Swarm Intelligence can be defined as a natural phenomenon enabling social species to quickly converge towards optimized group decisions by interacting as real-time, closed-loop systems. This process, which has been shown to amplify the collective intelligence of biological groups, has been widely studied in nature by analyzing the behavior of schools of fish, flocks of birds (Figure 2.7), and swarms of bees (H. Ahmed & Glasgow, 2012; Garg et al., 2009; Ginneken, 2009; Hassanien & Emary, 2018; Ioannou, 2017; Liu & Passino, 2000; Nayyar & Nguyen, 2018; Roy et al., 2014). Across all types of biological species, Ginneken (2009) identifies three overarching characteristics of a swarm:

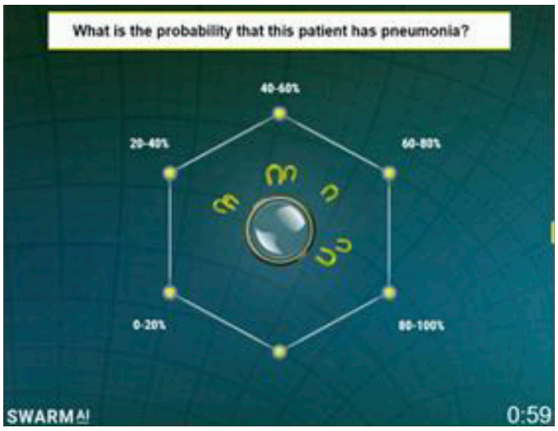
- A swarm contains a large number of freely moving agents, with each agent having equal decision influence.
- Every agent is capable of influencing and reacting to other agents, as well as adapting to external, environmental factors that can impact the swarm's behavior.
- All agents collectively form a complex but cohesive unity. In this sense, unity refers to the coordinated efforts of agents working together to achieve a common purpose.

2.3.3. Artificial Swarming Intelligence

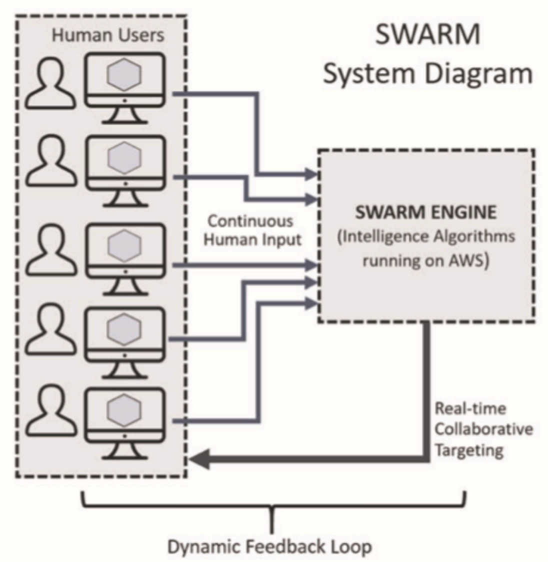
Artificial Swarm Intelligence (ASI) refers to the technology that brings the same benefits of Swarm Intelligence to decision-making in humans. Modeled on the same biological principles (Beni, 2005; Habboub et al., 2020; Hassanien & Emary, 2018), the technology uses the combined expertise, experience, wisdom and opinions of all team members to amplify the collective intelligence of the group for improved problem-solving and decision-making (Chakraborty & Kar, 2017; L. B. Rosenberg, 2015). In this respect, *artificial* thus refers to the technology that connects humans to interact and deliberate on complex issues in real time without the need for discussions through speech. Instead, the technology aims to create a networked human group resulting in what L. Rosenberg and Willcox (2019a) refer to as the 'brains-of-brains'-principle: a collective intelligence whose capacity exceeds that of any individual agent (Beni, 2005; Patel et al., 2019; L. Rosenberg, 2016).

In creating these 'hive minds', ASI is fundamentally based on the following core features:

- **Decentralisation**
  As opposed to traditional decision-making typically governed by centralized control or authority, ASI excels in amplifying the collective intelligence using decentralized decision-making (Garg et al., 2009; Hassanien & Emary, 2018). By fostering self-organization and emergent behavior, the technology eliminates power imbalances frequently observed in traditional decision-making hierarchies (Chakraborty & Kar, 2017; Hassanien & Emary, 2018).
- **Interactivity**
  ASI systems facilitate continuous interaction among participants, enabling the dynamic exchange of opinions and perspectives. To this end, participants act, react and deliberate according to the input of other participants, thereby creating a unified intelligence that converges to solutions based on combined wisdom and knowledge (Figure 2.8b) (Habboub et al., 2020; L. Rosenberg & Willcox, 2019a).
- **Synchronicity**
  In a synchronous decision process, all participants are present at the same time continuously influencing the decision (L. B. Rosenberg, 2015). With ASI operating in real-time, ensuring immediate response to changing conditions and enabling the swarm to tackle complex problems promptly and effectively.



(a) Interface of the Swarm.ai platform by Unanimous.ai (L. Rosenberg & Willcox, 2019b). The hexagon positions the provided outcomes of the question in space. The puck represents the collective decision, which is pulled by the individual participants reflected by the yellow magnets.



(b) Schematic overview of the closed-loop system using ASI (Patel et al., 2019). The participants provide input from their individual computers to the ASI engine. By aggregating the inputs from all individual inputs, the ASI engine determines the collective decision and feeds the result back to the individual users. This allows users to react to the collective opinion in real time.

Figure 2.8: Interface and system diagram of the ASI platform offered by Unanimous.ai

While several variations of ASI have been developed, this research uses the Swarm.ai platform by Unanimous.ai (California, USA) - also observed as the platform used for testing ASI in other medical studies (L. Rosenberg et al., 2018; A. Shah, 2019). This platform deploys a graphical interface in which the options are positioned along a hexagon (Figure 2.8a). Moreover, the collective decision is represented using a puck that is initially positioned in the center. In this interface, participants deliberate from their individual computers by moving the graphical puck toward the desired option. Participants influence the movement of the puck by adjusting graphical magnets to express their opinions on the collective decision. The input of each participant is recorded as a continuous stream of inputs, whereby the strength of each participant's conviction is indicated by the distance between their magnet and the puck. Altogether, the algorithm aggregates all individual inputs to determine the resulting movement of the puck as the collective decision. The resulting movement of the puck is then fed back to the individual's computer as input for the participant to react on (Figure 2.8b). After the time limit elapses (60 seconds by default) or the algorithm detects that the group decision has stabilized, an outcome is concluded and the process is terminated. Lastly, the algorithm intends to improve the outcome of the decision by considering various other factors, including the number of participants, initial preferences, consistency or changes in opinion of an individual, level of conviction, and type of answer choices.

2.3.4. ASI versus traditional crowd-based methods

Within crowd-based decision-making, ASI can be regarded as a specific type of method distinctive from traditional methods by creating a decision process where real-time interaction and deliberation among participants occurs. On the other hand, traditional methods for crowd-based decision-making for this research are considered those where respondents provide static, isolated inputs in that the individual opinions are aggregated without ongoing interaction among participants. As such, traditional crowd-based methods typically involve polls, votes, or surveys. As mentioned previously Section 2.2, the foundation of these methods lies in what is commonly referred to as the *wisdom of the crowd*: Capturing the average opinion across a population typically provides a more accurate view than most of the individual estimates (Lorenz et al., 2011; Mavrodiev et al., 2012). Over the years, many other more sophisticated and accurate methods have been proposed that have particular relevance to medical decision-making, including the Delphi method (an iterative process for gaining consensus through successive question rounds across a panel of experts) (Maharaj et al., 2019; E. Taylor, 2020) and the wisdom of select crowd strategy (where first the top-5 best performers or knowledgeable experts of a population are selected to then only combine their response while neglecting the rest of the crowd) (Mannes et al., 2014). However, all methods can be fundamentally recapitulated to the same underlying process: By collecting the individual inputs of a population and then statistically aggregating all results using various strategies, a consolidated outcome is determined that is supposed to represent the wisdom of the crowd.

Subsequently, several key differences arise between ASI and traditional crowd-based methods:

- **Input**
  When conducting traditional crowd-based methods, participants serve as 'respondents' who simply provide an isolated response which is then discretely captured before aggregation (R. Shah et al., 2023). From a statistical perspective, participants thus serve as *data-points* to be then combined with other responses for determining a collective outcome. On the other hand, ASI treats every participant as *data-processors*, each empowered to act, react, and interact to explore the decision-space and converge on solutions that optimize their combined knowledge, wisdom, insights, and opinions (L. Rosenberg & Willcox, 2019a).
- **Measurement**
  When capturing the input of each respondent using a single-point measurement, this obstructs one's ability to express any nuances. Examples of this include (1) when one person prefers option A, but is also perfectly comfortable choosing option B or (2) when a person is not sure about the preferred option, but is convinced that it is not option D. On the contrary, the continuous and interactive process enabled through ASI allows participants to explore and deliberate across the decision-space collectively. This allows for expressing a more nuanced opinion in the process before a decision is concluded.
- **Process**

To allow for more interaction, sequential input methods can capture inputs of respondents over time (including multi-measurements and up-/downvoting). However, although potentially making the decision process more interactive, such methods have been observed to be particularly sensitive to social bias effects (Section 2.2). For instance, Muchnik et al. (2013) and Weninger et al. (2015) carried out randomized experiments finding that a single up-vote, when inserted first on a particular post on Reddit, influenced the final decision of the group by 24% - 25%. Meanwhile, ASI is claimed to minimize these social bias effects by intelligently detecting the conviction class of each participant: Whereas highly convinced participants show less sensitivity to social bias effects, participants who less convinced are typically more sensitive to social bias effects. Consequently, while all participants have equal opportunity to influence the decision outcome, ASI intelligently provides more influence to those who show more conviction in their opinion. While this may not resolve an influenced opinion from the perspective of the individual participant, the technology does reduce the sensitivity from the perspective of the collective decision.

- **(Implied) decision outcome**
  When conducting traditional crowd-based methods, the results typically involve a dataset that exposes the dissonance across different outcomes. Especially when deciding according to a majority vote or other voting mechanisms, concluding the outcome requires analyzing the extent to which one option is preferred over another. Instead, ASI is based on concluding a decision through *unity*: rather than exposing the different opinions, the technology combines the opinions of different participants to deliberate as a uniform collective.

Elements	Traditional methods	Swarm Intelligence
Participants	'Data points'	'Data-processors'
Participation	Individual	Collective
Measurement	Single-point or sequential	Continuous
Deliberation	Isolated	Collaboration and interaction
Decision-making process	Statistical aggregation	Collaborative thinking
Driver	Polarisation	Convergence
(Implied) Outcome	Divergence	Unity

Table 2.3: Theoretical framework ASI vs. traditional crowd-based methods

2.3.5. ASI benefits to medical decision-making

Aside from the theoretical features, several practical benefits are identified in the grey literature when bringing the fundamentals of ASI to the setting of MDT meetings. A selection of potential benefits includes the following:

Improved decision accuracy or consensus
Questions can be either based on ground truth or non-truth (henceforth referred to as consensus) issues (Figure 2.9) (Camera et al., 2024). On the one hand, ASI is claimed to achieve noteworthy accuracy in truth- or fact-based issues across various applications. Previous ASI studies include a significantly enhanced effectiveness and usefulness when tasking teams of military pilots to collaboratively generate qualitative insights through ASI about the design of cockpits for Boeing (Befort et al., 2018), an enhanced accuracy of 25% for groups of financial traders when forecasting market key indicators (Oil, Gold, and S&P) through ASI (L. Rosenberg et al., 2017) and a 24% ROI after observing a group of average sports fans significantly outperforming the Vegas odds-market when tasked with predicting the outcome of 238 basketball games in the NBA through ASI (L. Rosenberg & Willcox, 2019a). On the other hand, ASI is perceived as a promising method for finding consensus in non-truth issues as well (L. B. Rosenberg, 2015): Issues where no ground truth can be established and therefore requires 'a generally accepted opinion or decision among a group of people' [2].
Reducing social influence
As introduced previously, MDT meetings can be subjected to various social influences (Section 2.2). In this

[2] https://dictionary.cambridge.org/dictionary/english/consensus

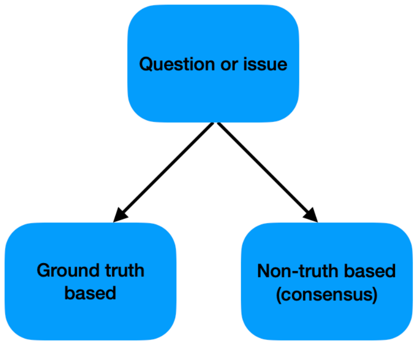


Figure 2.9: Two types of questions/issues (Camera et al., 2024)

respect, a typical phenomenon observed is the aforementioned 'illusion of unanimity' (Mannion & Thompson, 2014): when silence is interpreted as consent, leading the group to think it has made a unanimous decision while actually concluding in 'false consensus' (Lamb et al., 2013). To this end, ASI can serve as a powerful tool for the 'middle group' to speak out as its design facilitates a large portion of the reduction factors identified previously for social loafing (Table 2.1), including evaluation, attractiveness, and equal efforts. Likewise, ASI also preserves anonymity and equal decision influence. Especially when dealing with controversial cases, ASI thus can offer an accessible platform to reach quicker consensus among experts without the need for traditional crowd-based methods.

Intertwining man and machine

Over recent years, state-of-the-art AI- and ML technologies have been observed to be increasingly adopted into various applications of healthcare decision-making (Subsection 1.1.3). However, a frequent barrier currently encountered arising from ethical risks: Current AI- and ML-algorithms are typically designed to *substitute* humans while taking over medical decision-making, imposing significant resistance when it comes to decision responsibility and accountability (M. I. Ahmed et al., 2023). In the context of MDT meetings, this resistance and its implications for patient safety are what can result in considerable disruption against the adoption of current AI-powered CDSSs (Subsection 1.1.3). On the contrary, however, ASI is fundamentally designed to *co-operate* with humans for improved decision-making. In the context of MDT meetings, ASI thus amplifies the collective intelligence by 'connecting' groups of experts to work as a networked super-intelligence (L. Rosenberg & Willcox, 2019a). Correspondingly, the technology ensures humans remain in charge of the decision process at all times.

Remote meetings

Using ASI, medical experts do not need to be in the same room to participate in MDT meetings. Instead, experts can participate in the decision process from a remote device connected to the internet. Although the use of online has accelerated in recent years, particularly after COVID-19, group discussions through online meetings are still commonly regarded as less comfortable compared to physical meetings particularly when group sizes increase. To this end, ASI can serve as a useful platform since it eliminates the need for discussion through speech or the necessity to deliberate in the same physical room (Section 2.2). Instead, it enables experts to deliberate in real-time regardless of their physical location and group size. Hence, as part of the global transition toward a more remote, flexible working culture, ASI can potentially facilitate improved participation and organizational flexibility in MDT meetings without sacrificing the quality of the decision process.

Medical training and assessment

Whilst ASI aims to decentralize the decision process, the technology also enables the evaluation of the individual decision behavior of each participant. Aside from recording the individual decision outcome, ASI thereby also captures one's initial response along with the individual consideration process through which the participant eventually converges toward the outcome. Subsequently, using the analytics and visualizations of individual decision behavior, ASI could therefore serve as a comprehensive training tool for medical education and performance assessment. As such, evaluation of individual assessment performance can be used as feedback to both junior doctors and senior doctors.

2.3.6. Conclusion of ASI fundamentals review

This literature review examines the current evidence behind ASI, substantiating its fundamental features and concepts. The review highlights the significant advantages of ASI over other crowd-based methods, particularly in medical decision-making. The core findings are summarized as follows:

ASI is grounded in Swarm Intelligence, a natural phenomenon where species quickly converge towards optimized group decisions by interacting as a real-time, closed-loop system. ASI technology leverages these principles to enhance human decision-making through decentralization, interactivity, and synchronicity.

The literature identifies several key benefits of ASI compared to traditional crowd-based methods, including interactive participation, closed-loop feedback, minimized social bias, and collective convergence in decision outcomes. Furthermore, ASI demonstrates practical advantages in MDT settings, such as improved decision accuracy and consensus, reduced social influence, integration of human and machine intelligence, facilitation of remote meetings, and enhanced medical training and assessment.

Overall, the findings underscore the potential of ASI to transform decision-making processes by harnessing the collective intelligence of groups through advanced technological frameworks.

Implications for this research

As stated in Subsection 1.1.5, two core components of the knowledge gap in surgical decision-making were found to be the issue of adverse tumor assessment and the sensitivity of MDT meetings to dominant individuals. Considering this knowledge gap, the findings of this review contribute to this research as follows:

- Two particular ASI benefits that are relevant to MDT meetings are improved decision accuracy and reduced social influence. Given the knowledge gap, these two benefits are further examined as the pivotal variables in the setup of this research. This is further elaborated in Section 2.4.
- ASI shows demonstrated benefits in both ground-truth issues as well as consensus-based issues. Since the tumor assessment accuracy is assessed against a perioperative outcome considered as the clinical benchmark, this advocates for a ground-truth issue. This is further explained in Section 2.4.

2.4. Synthesis of literature reviews

To synthesize the results of all the reviews, an overview of all the identified research objects (Chapter 1, Section 2.3 and Section 2.2) along with the research scope is shown in Figure 2.10. When defining the scope of this research, the objective of this research serves as the foundation for examining two core research components: improved decision accuracy and reduced social influence. As part of the improved decision accuracy, a literature review is conducted to establish the evidence behind adverse tumor assessment. On the other hand, as part of the social influence analysis, two social factors are examined separately: social loafing and social bias. Altogether, the output of this research produces three deliverables: (1) An analysis of the evidence behind adverse tumor assessment, (2) the performance analysis, and (3) the social influence analysis.

With respect to the individual research objects, a more detailed explanation of the specific delineations can be found in clockwise order below:

- **Adverse case assessment for PC**
  In Chapter 1, the abandonment rate of PC surgeries is introduced as an important knowledge gap currently encountered in PC surgery. Therefore, the abandonment rate of PC surgeries serves as the core knowledge gap of this research, which is investigated using a holistic approach.

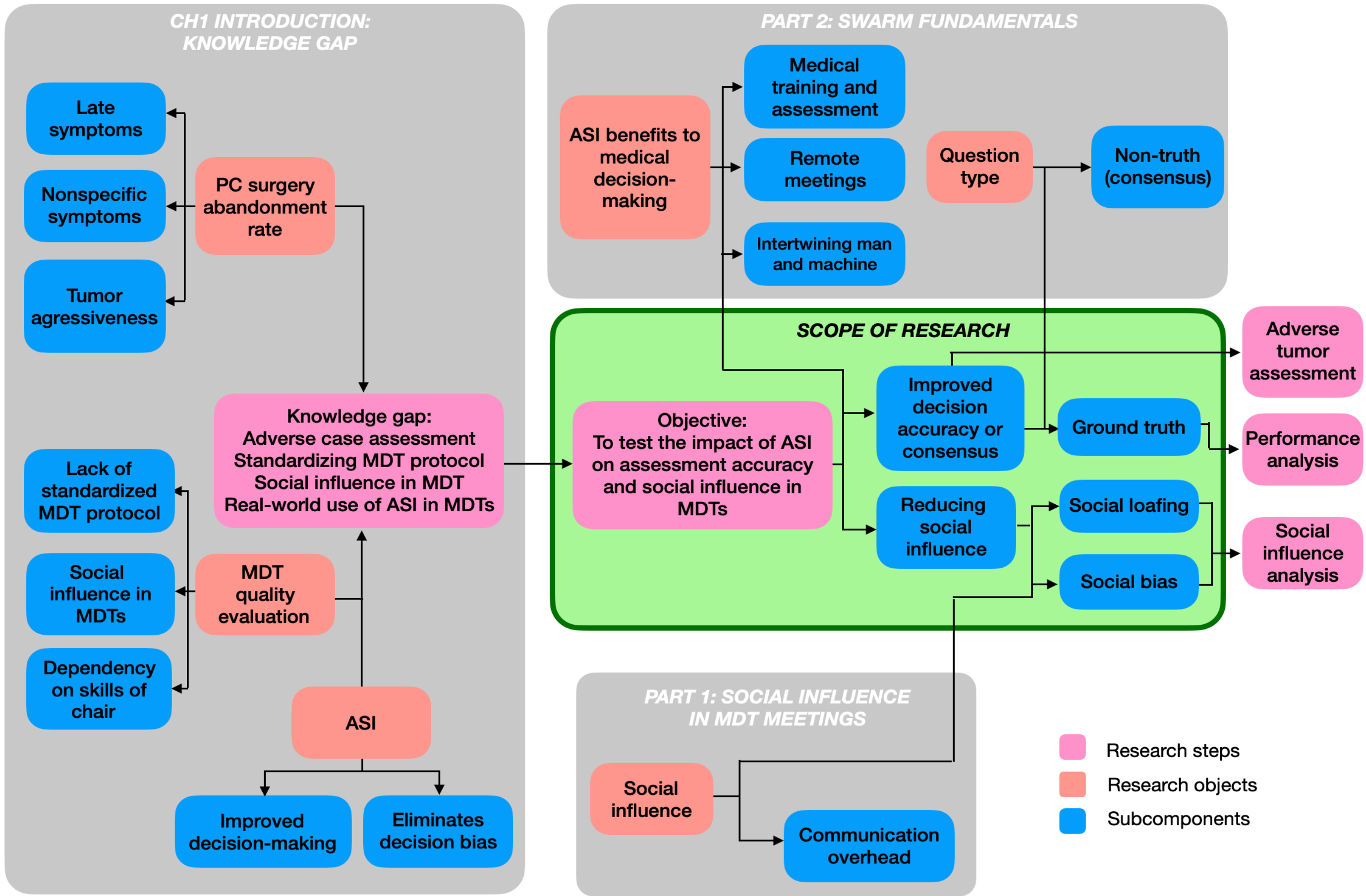


Figure 2.10: Research scope

- **ASI benefits to medical decision-making**
  From each of the five benefits identified for ASI in medical decision-making (Subsection 2.3.5), this research focuses on improved decision accuracy and reducing social influence as the core features utilized to improve tumor assessment accuracy and reducing social influence in MDT meetings. Given the present state of literature, these features are currently perceived as most relevant for in-depth research to MDT meetings.
- **Social influence**
  Based on the three components identified for social influence (Chapter 6), this research delineates social loafing and social bias as the conceptual latent variables to test the social impact of ASI. While communication overhead is also expected to have a direct impact on meeting efficiency, this factor is particularly relevant for larger group sizes (>10) which is deemed infeasible for the setup of this research.
- **MDT quality evaluation**
  The literature review (Subsection 1.1.2) revealed (1) lack of standardized MDT protocol, (2) social influence in MDT, and (3) dependence on the quality of the chair as three current limitations currently encountered in MDT meetings. Since the use of ASI is expected to affect each of the three limitations, all limitations are considered in the knowledge gap.

Implications for this research

Overall, this chapter presents a dedicated literature review on two core topics of this research:

- ASI fundamentals
- Social influence in MDT meetings

Furthermore, a third review of the evidence behind PC surgery assessment is presented in Chapter 4. With respect to the flow of this research, each review aims to substantiate *why* each of the topics is deemed relevant for the setup of this research. Afterward, by specifically delineating the explained research objects introduced in the literature reviews, the research scope (Section 2.4) defines *what* is included in the setup in this research. Subsequently, the next chapter on research design (Chapter 3) builds on this foundation by explaining *how* this study is set up using the gained background knowledge.

3 Research Design

Opening

Recap
Previously in Chapter 1, the main objective is formulated as follows:

Objective: This paper aims to explore whether the accuracy of tumor assessment for PC can be improved using ASI in MDT meetings. It thereby also examines whether ASI can be used to reduce social influences in the decision process of tumor assessment.

To this end, Chapter 1 introduces the abandonment rate of PC surgery as the underlying motivation for this research. Furthermore, Section 2.3 provided the background of this research. It introduces ASI as an AI technology, thereby substantiating its potential impact on improved decision performance. Furthermore, Section 2.2 also elaborates on social loafing and social bias as two notable types of social influence that can be potentially reduced by using ASI.

Introduction
This chapter presents the research design established for achieving the research objective. It thereby explains how each of the previously identified benefits (improved decision performance and reduced social influence) is addressed through this research. Furthermore, an overview of the research flow is given in Figure 3.1. To substantiate the design of this research, this chapter is structured as follows:

- Research subquestions
- Datasets
- Research methods
- Ethical approval and consent to participate

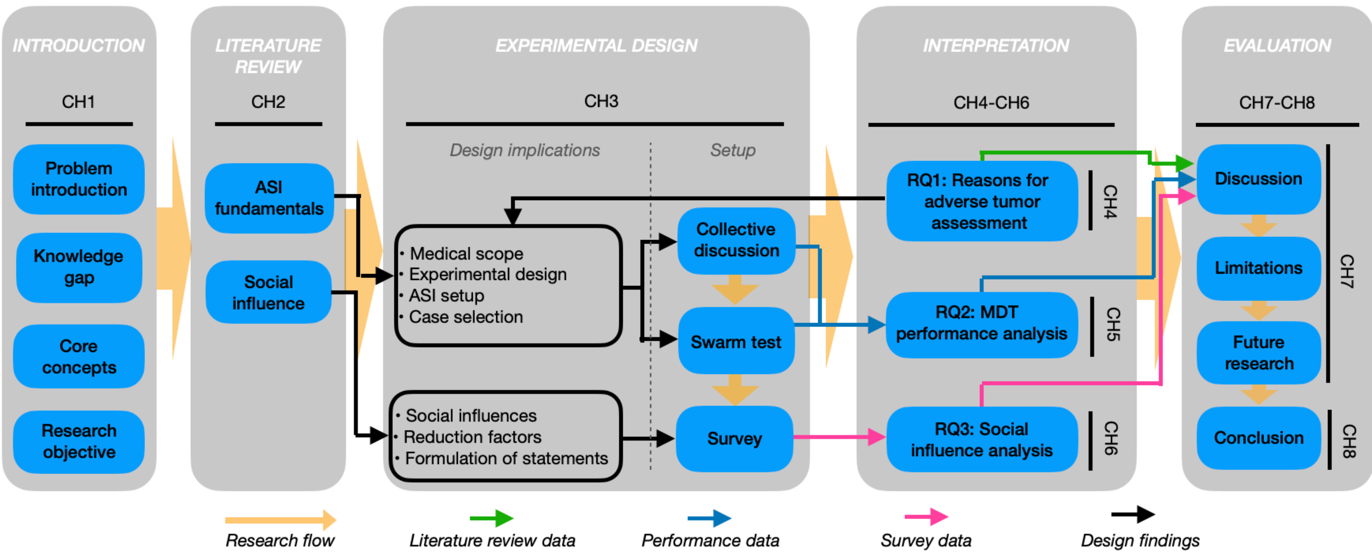


Figure 3.1: Research flow diagram

3.1. Datasets

Based on the proposed main research question and sub-questions, several datasets are to be collected. An overview of the use and interaction of all datasets is given in Figure 3.2. For each dataset, a substantiation for its collection is given below:

- **RQ1: Evidence on PC surgery**
  For identifying evidence on the abandonment rate of PC surgery, qualitative data is needed on the underlying causes and factors that contribute to the adverse case assessment. Firstly, this involves

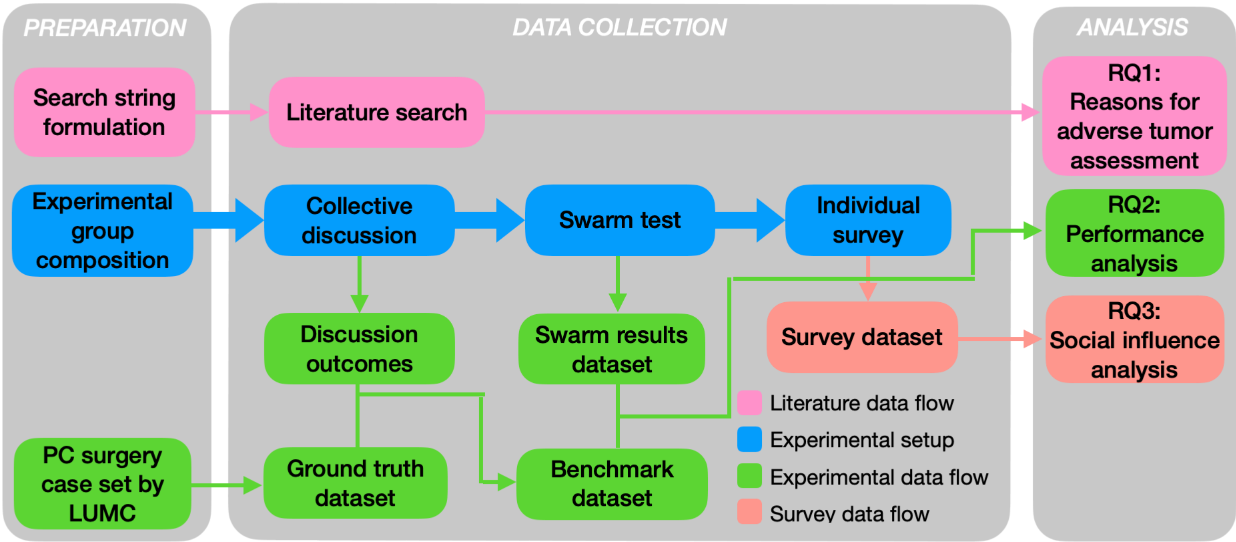


Figure 3.2: Data flow diagram

collecting findings on the current decision-making process in which the diagnosis of PC is established. Secondly, the data should reveal complications in the current process by which tumor resectability is assessed.

- **RQ2: Data to establish ground truth + Swarm performance**
  To demonstrate the impact of ASI on MDT performance, assessment data must be collected on the decision accuracy of the expert panel in determining the resectability of PC cases based on a presented medical image (CT scan). To this end, the software platform Swarm Analytics, an analysis platform offered supplementary to the swarm platform by Unanimous.ai, combined with Excel is used to collect and analyze the data. Herewith, three datasets need to be acquired:
  1. **Ground truth**: a dataset containing 10 past PC surgery cases, each containing patient CT scans and additional patient data along with the initial resectability assessment and perioperative outcome.
  2. **MDT performance**: a performance dataset based on the same 10 PC surgery cases, whereby the experimental group provides its initial resectability assessment of each of the 10 patient cases separately.
  3. **Swarm performance dataset**: a dataset based on the same 10 PC surgery cases, whereby all experiment participants collectively assess the resectability of each of the same 10 patient cases through ASI.
- **RQ3: Data on social influence**
  To determine the potential contribution that ASI can deliver for reducing social influence, evaluation data is to be gathered to determine whether ASI is perceived as a functional solution in driving reduction factors (Section 2.2). Since ASI is supposed to have a causal relationship with social influence, quantitative data is collected to indicate whether such a relationship seems plausible.

3.2. Research methods

For collecting the necessary datasets, two methods are employed for this research: An experiment and a survey (Table 3.1). For each subquestion, this section presents the deployed method and the corresponding setup.

3.2.1. RQ1: Literature review

Given the twofold types of data to be collected (Section 1.2), this literature review is structured through two subtopics: (1) PC tumor diagnosis and (2) reasons for adverse tumor assessment. On 28 June 2024 (PC tumor staging) and 29 June 2024 (reasons for adverse tumor assessment), a systematic search was carried

Method per Research Question (RQ)			
RQ	Method	Data Analysis Approach	Expected Output
1	Literature review	Research Approach	Reasons for adverse tumor assessment
2	Experiment	Quantitative	Performance analysis
3	Survey	Quantitative	Social influence analysis
Main	Convergence	Quantitative + Qualitative	Conclusion

Table 3.1: Research method per RQ

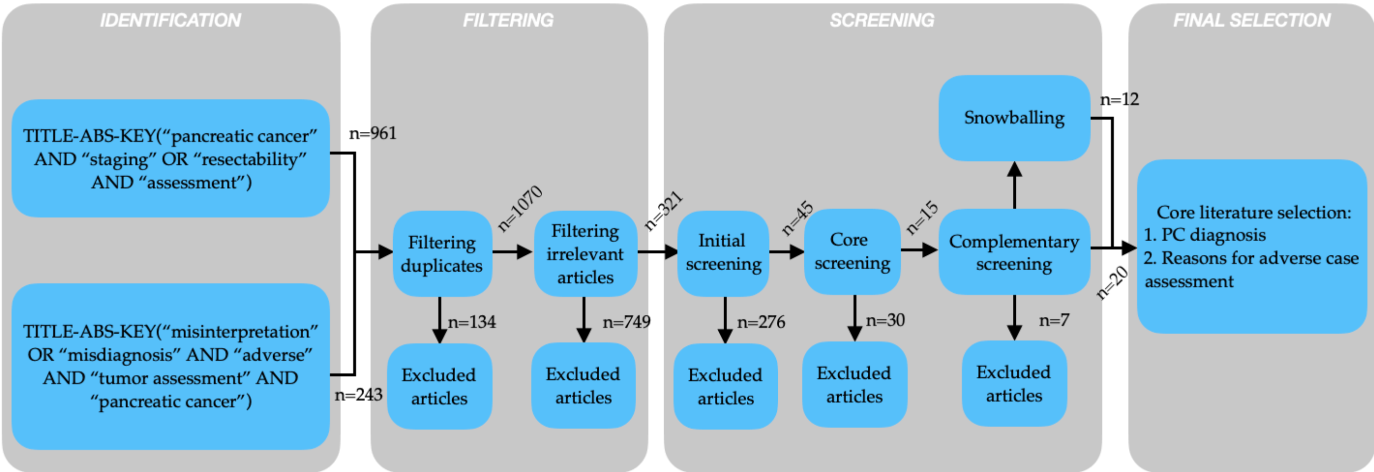


Figure 3.3: Selection procedure for PC resectability assessment review

out using scientific databases Google Scholar, Medline, Embase and IEEE. A schematic overview of the selection process is depicted in Figure 3.3. To profoundly substantiate both reviews, a dedicated search string was used for each of the two topics. Based on the explanation of the concepts of medical decision-making and adverse tumor assessment introduced previously (Chapter 1), a dedicated explanation for each string is given hereafter in step 1. Overall, the literature was selected according to the following process:

1. **Identification**
   To determine the state-of-the-art across two subtopics, the initial selection of literature was collected for both subtopics by use of the following search strings.
   (a) *"TITLE-ABS-KEY("pancreatic cancer" AND "staging" OR "resectability" AND "assessment")"*
   To begin with, this search term includes the terms "pancreatic cancer" and "assessment" for adequate scoping of the review to the topic of resectability assessment within the area of PC. Furthermore, while being distinct medical concepts, preliminary research on this topic shows that - within the area of tumor diagnosis - staging and resectability generally involve evaluating the extent and spread of the tumor using the same classification. Accordingly, both concepts are integrated in this term using the following term: ... "staging" OR "resectability"... .
   (b) *"TITLE-ABS-KEY("misinterpretation" OR "misdiagnosis" AND "adverse" AND "tumor assessment" AND "pancreatic cancer")"*
   Similarly to the previous string, the terms ..."adverse" AND "tumor assessment" AND "pancreatic cancer" ensure proper scoping of the review to the specific topic of adverse tumor assessment in the field of PC. Furthermore, for specifying papers where the adverse case assessments originate from misdiagnosis before the surgery, the search string uses "misinterpretation" and "misdiagnosis". Since preliminary research shows that both terms are used interchangeably, the terms are combined as "misinterpretation" OR "tumor assessment"....

For simplicity reasons, the visualization of the selection process (Figure 3.3) provides a combined overview as each of the two searches progresses through mutual selection phases. However, each search was performed separately for each of the two subtopics. Accordingly, the search results from the identification phase toward the filtering phase can be interpreted as an aggregation formally expressed as combining the two search strings using an OR statement.

2. **Filtering**
   For filtering the initial collection, all papers that meet the following criteria were excluded:
   - For both PC tumor staging and reasons for adverse tumor assessment, papers published before 1 January 2024 were excluded. An exception was made for (Armato et al., 2009), based on its time-independent relevance for radiology performance despite its publication date.
   - Paper is published in non-English and non-Dutch language
3. **Screening**
   The screening phase consisted of three components:
   - The initial screening was done by specifically filtering papers that were written as scientific papers or sections of books.
   - The core screening was done by scanning the resulting list of papers on specifically addressing one of the two aforementioned key topics based on their title, abstract, and conclusions. For adverse tumor assessment, the paper was screened specifically on containing reasons that contribute to adverse tumor assessment.
   - The complementary screening was carried out on the remainder of the papers by specifically filtering the papers that (1) describe the different stages and diagnostic process of PC or (2) provide evidence on factors that contribute to adverse tumor assessment. In addition, the selection was completed through snowballing using in-paper references.
4. **Final Selection**
   The final selection was completed by combining both the selected papers from the complementary screening and the snowballing process (Badampudi et al., 2015) to a cumulative of 20 papers

3.2.2. RQ2: Experimental set-up

To assess the performance of MDT meetings using ASI compared to discussion-based MDT meetings, the core of this research is centered around an experiment that uses a within-subject design (Charness et al., 2012). For this design, an experimental group is composed of medical experts whose performance is analyzed over two measurements: A collective discussion session ('untreated') and thereafter the collective swarm session ('treated'). An overview of the experimental flow using an RCT diagram is given in Figure 3.4.

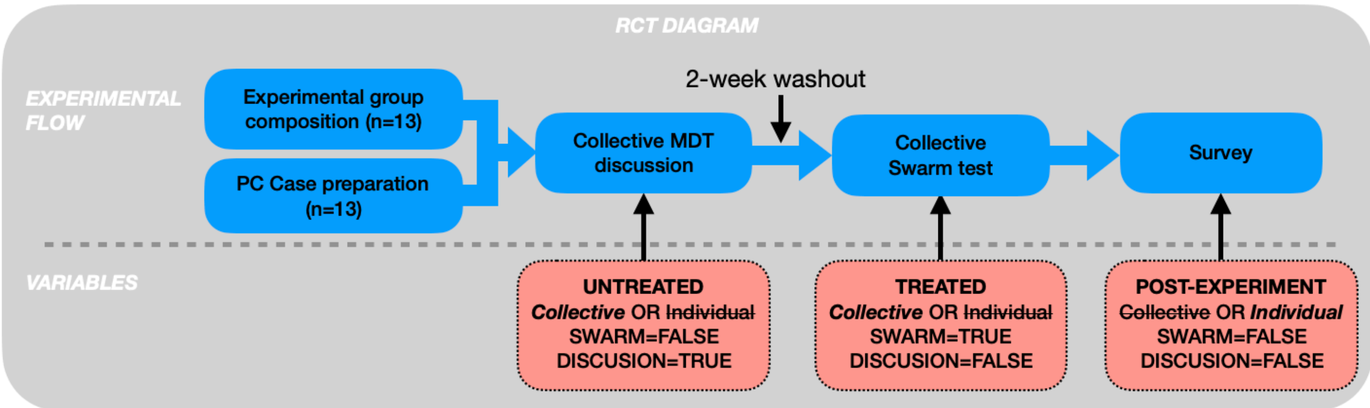


Figure 3.4: Theoretical RCT diagram. For this research, a population of thirteen participants (n=13) was invited and a set of thirteen (n=13) cases was prepared. Moreover, a two-week washout period was installed to reduce framing bias (middle). Lastly, experimental variables are given for (1) collective vs individual (how the test/survey is conducted: collectively vs individually), (2) swarm (whether the research step involves ASI: true/false), and (3) discussion (whether the research step involves discussion: true/false) (below)

Group composition

The invited experimental group is composed of thirteen medical experts (hereafter: participants) actively working or researching in the field of PC at Leiden University Medical Center (LUMC), The Netherlands.

Every participant in the selected group is invited based on their regular (weekly) presence at the LUMC MDT meetings for pancreatic surgery. To eliminate biased decision outcomes arising from different group compositions (Subsection 1.1.2), participants could only participate in the experiment if they could attend the collective discussion *and* swarm test. This ensures that the composition of the experimental group remains constant across both sessions. Equivalently, given the crucial role and dependency of the radiologist (as further elaborated in Section 4.4), the collective discussion and swarm test were both chaired by the same radiologist.

Case selection

For assessing the (difference in) performance of both sessions, a selection of thirteen PC cases was prepared from past patient cases at LUMC (Figure 3.4). Using the LUMC patient database, cases were selected as follows:

- Cases were selected from patients who have previously received surgical treatment at LUMC after 01 January 2020.
- Cases contained *anonymized* medical images, disease details, and patient history. From the CT scans, a pancreatic tumor was visible along with its surrounding tissue. For every case, the CT scans were used that were taken at the first presentation of the patient at LUMC.
- Among the case set, a selection of four *adverse* cases was included: Cases where the original MDT conclusion suspected a (borderline (BRPC)) resectable tumor, whereas the perioperative (or post-surgical) results revealed the tumor staging to be locally advanced (LAPC), metastatic or otherwise irresectable. This classification is based on the findings of and further explained in Chapter 4.
- Since the specific boundary criteria between different tumor stages can vary between (international) guidelines (Section 4.2), the guidelines of the Dutch Pancreatic Cancer Group (Veldhuisen et al., 2018) were followed for this experiment since these comprise the standard assessment criteria adopted by LUMC and thus commensurate with the original diagnostic outcomes. This is further explained in Chapter 4.

Collective discussion

During the collective discussion, the experimental group discusses the case set to collectively assess the resectability of each patient. Given the great variety in which MDT meetings are conducted across healthcare institutions and countries (Subsection 1.1.2), the collective discussion is designed to mimic the standard protocol by which MDT meetings are conducted at LUMC as closely as possible. Consequently, the session was thus structured as follows:

1. All cases are discussed separately and sequentially, reviewing as many of the thirteen cases as possible within 60 minutes.
2. The collective discussion is conducted using the same room and computer facilities as regular pancreatic MDT meetings at LUMC.
3. In line with the standard MDT meeting protocol LUMC, each case is prepared and introduced by the radiologist who presents her clinical findings based on her review.
4. After the introduction by the radiologist, the participants collectively discussed each case to converge toward consensus regarding the resectability of the tumor.
5. For objective comparison with the ASI test, the experimental group concludes each case using one of the discrete options in the option set given in Table 3.2. This option set is based on the findings of Chapter 4.

Washout-period

Since both sessions used the same set of cases, a 14-day washout period was installed between 8 May 2024 and 22 May 2024 to minimize the influence of carryover effects (Cnops et al., 2022). Based on other studies (Donald Harvey et al., 2021; R. Shah et al., 2023) and the availability constraints of the participants, two weeks is installed between the two sessions.

No.	Option
1	Resectable
2	Borderline resectable (BRPC)
3	Locally advanced (LAPC)
4	(Suspicion of) metastatic disease
5	Irresectable
6	No conclusion

Table 3.2: Standard option set for each case using which the participants systematically respond to the experimental question: What is the resectability of patient X?

Swarm test

For assessing the pre-selected case set using ASI, the same group of experts evaluates the case set simulta-

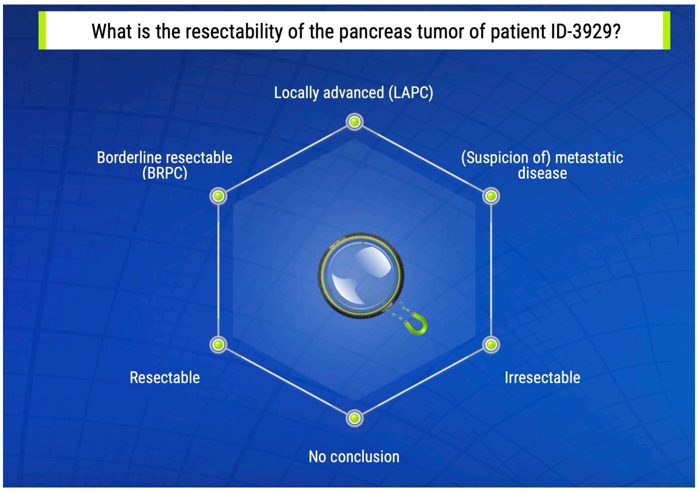


Figure 3.5: Swarm interface

Figure 3.6: Technical setup of swarm test

neously using the Swarm.ai platform by Unanimous.ai (California, USA). To this end, each of the participants collaborates in a swarm session to collectively decide on the resectability from their computer (Figure 3.6). Using this setup, the swarm test was conducted as follows:

1. All cases are discussed separately and sequentially.
2. To avoid unintentional framing or other bias caused by the presentation of different radiologists (Section 4.4), each case is prepared and introduced by the same radiologist as the collective discussion, presenting the same clinical findings.
3. To eliminate any social influence (Section 2.2), interpersonal discussion concerning the patient cases is prohibited.
4. After the introduction by the radiologist, all participants deliberate through the Swarm.ai platform to converge to one of the same options listed in Table 3.2. For each swarm, the maximum time was restricted to 60 seconds and chat functionality was disabled. In addition, each participant joined the swarm environment using an anonymized Swarm-ID to preserve full anonymity throughout the swarm test.

Comparison of experimental results

The output of the experiment produces two datasets: the collective discussion outcomes and the swarm test outcomes (hereafter referred to as 'experimental outcomes'). Using the experimental outcomes and case data (perioperative and original diagnosis outcomes), three comparisons are conducted (Figure 3.7):

1. Internal comparison of experimental results: collective discussion versus swarm test (Chapter 5)

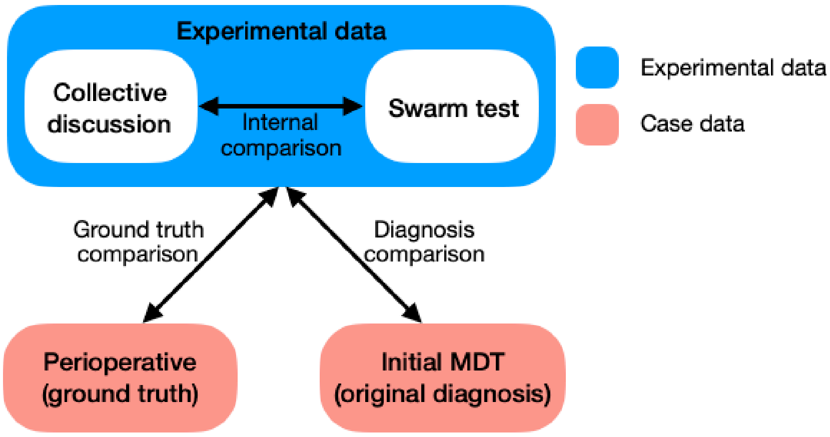


Figure 3.7: Comparison structure for the performance analysis. The comparison is composed of three components: (1) internal comparison, (2) ground truth comparison, and (3) diagnosis comparison

2. Ground truth comparison: experiment versus ground truth (Appendix F)
3. Diagnosis comparison: experiment versus initial MDT (Appendix F)

For each comparison, the level of concordance between two data sets is determined quantitatively. To this end, the concordance is defined as follows:

$$\text{concordance level} = \frac{\text{number of cases with identical case outcome}}{\text{total number of cases in case set}} \times 100\%$$

Accordingly, for each comparison, a higher concordance level corresponds to a relatively higher number of identical case outcomes. Furthermore, with regard to the assessment accuracy, this specifically refers to the comparison between the experimental outcomes and the perioperative (ground truth) outcomes. The assessment accuracy thus aims to reflect to what extent the participants during the experiment assess the resectability of the case in concordance with the 'true' tumor resectability originally established from perioperative outcomes.

3.2.3. RQ3: Survey

Based on the social loafing and social bias effects identified previously in Section 2.2, an anonymized survey is conducted among the participants to capture their attitude toward the potential capability of ASI to minimize social influences in MDT meetings. While the original survey is conducted in Dutch, a translated version of this survey to English can be found in Appendix E. Using a structured setup (closed-ended questions), the survey uses a quantitative approach to assess the effects of ASI on social loafing and social bias.

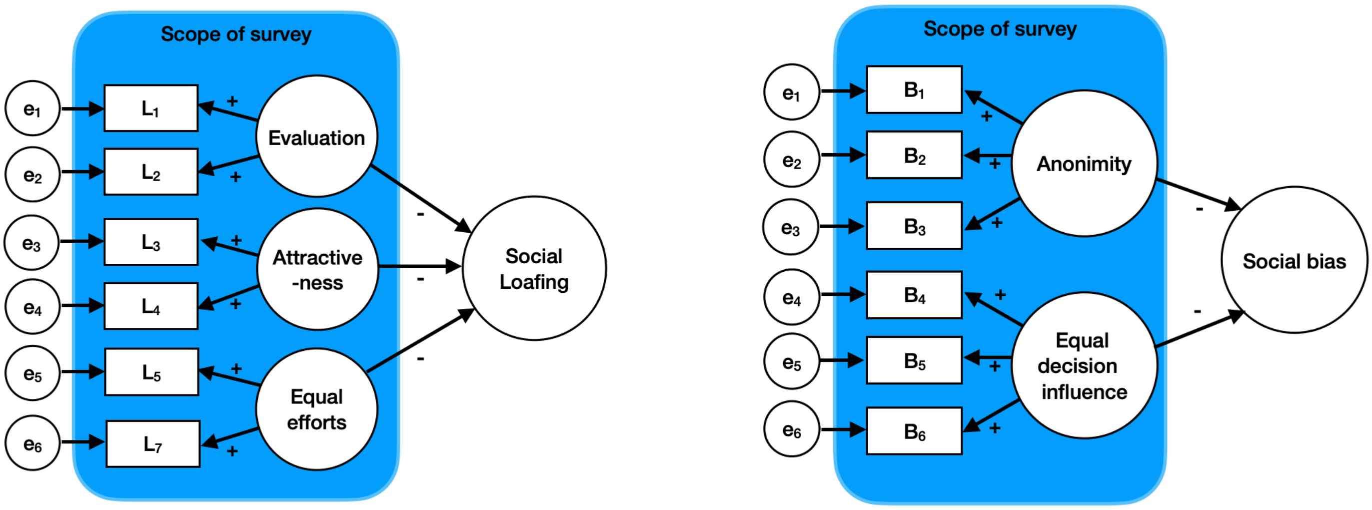
Conceptual measurement model

For each of the two social factors, a selection of reduction factors is created from the relevant ASI factors introduced previously in Section 2.2. By expressing social loafing and social bias effects each as latent variables (Weller et al., 2020), a conceptual measurement model is used to provide an intuition as to how the survey intends to capture both social factors (Figure 3.8). While the measurement model is not statistically tested for this research, it presents the conceptual association of each reduction factor to its social factor based on which the survey statements are formulated.

Survey setup

Based on the aforementioned conceptual measurement models, the survey is set up as follows:

- The survey is conducted a posteriori (after the experiment) as a single-point measurement among the experimental group
- Given the time constraints imposed by the availability of the participants, a selection of reduction factors (presented in Section 2.2) is created based on the evaluability of each reduction factor and its relevance



(a) Causal model for social loafing

(b) Causal model for social bias

Figure 3.8: Conceptual measurement models for social loafing and social bias

with respect to the experiment. For social loafing, the selection consists of *evaluation*, *attractiveness* and *equal efforts* {L_1, L_2, L_3, L_4, L_5, L_6}, whilst for social bias the reduction factors *anonimity* and *equal decision influence* {B_1, B_2, B_3, B_4, B_5, B_6} are selected (Figure 3.8).

- For reliability (defined as 'the accuracy or the extent to which a research instrument consistently provides the same results' (Heale & Twycross, 2015)), multiple statements for each factor have been formulated such that they are to some extent interchangeable and can therefore be regarded as repeated measures. Hence, the more their scores are similar, the higher the reliability (Molin, 2024). An overview of all statements is given in Table 3.3 and Table 3.4.
- Each statement is answered individually using Likert-scale answers between 1-5. On this scale, 1, 2, 3, 4, and 5 refer to strongly disagree, disagree, neutral, agree, and strongly agree, respectively. Additionally, participants can provide additional comments as open input regarding the statements.

Indicator statements for social loafing		
Indicator	**Statement**	**Reduction factor**
L_1	The opportunity offered by ASI to compare my opinion against the group pushes me to provide more input compared to regular MDTs	Evaluation
L_2	Being aware that my individual performance can be evaluated using ASI, this pushes me to provide more input compared to regular MDTs	Evaluation
L_3	The visualization of the collective decision process by ASI pushes me to provide more input compared to regular MDTs	Attractiveness
L_4	ASI provides a more attractive way to participate in the decision process, which pushes me to provide more input compared to regular MDTs	Attractiveness
L_5	Since ASI enables me to see others participate in the decision process, this pushes me to also better participate compared to regular MDTs	Equal efforts
L_6	Since ASI pushes everyone to provide equal input, this pushes me to provide more input compared to regular MDTs	Equal efforts

Table 3.3: Indicator statements for social loafing (translated from Dutch)

Data aggregation

Indicator statements for social loafing		
Indicator	**Statement**	**Reduction factor**
B_1	Due to the anonymity provided by ASI, I would be more inclined to provide my opinion when disagreeing with the collective opinion compared to regular MDTs	Anonimity
B_2	Due to the anonymity provided by ASI, I feel more courage to express my opinion when disagreeing with the collective opinion compared to regular MDTs	Anonimity
B_3	Due to the anonymity provided by ASI, I feel more motivated to continue expressing my opinion when disagreeing with the collective opinion compared to regular MDTs	Anonimity
B_4	ASI provides me the impression that I always have as much influence on the decision outcome as all other participants	Equal decision influence
B_5	ASI eliminates additional influence gained based on personality, years of experience, and specialisation	Equal decision influence
B_6	During the experiment, I have expressed my opinion across all cases more than I have during the collective discussion	Equal decision influence

Table 3.4: Indicator statements for social bias (translated from Dutch)

The survey results are interpreted based on a quantitative interpretation of the Likert-scale data. To this end, the mean (μ), median ($\tilde{x}$), and variance (σ^2) are calculated for each reduction factor as follows:

- **Mean (μ)**
  The mean (μ) represents the average score of the responses. It provides a measure of central tendency, indicating the typical response of the population. Including the mean in the interpretation helps identify the overall level of agreement or disagreement among respondents. It is calculated using the formula:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

  where N is the total number of responses and x_i is the value of the i-th response.
- **Median ($\tilde{x}$)**
  For an odd number of responses, the median is the value at position $\left(\frac{N+1}{2}\right)$. For an even number of responses, it is the average of the values at positions $\left(\frac{N}{2}\right)$ and $\left(\frac{N}{2}+1\right)$. The median is a robust measure of central tendency, less affected by outliers than the mean. Including the median in the interpretation ensures a comprehensive understanding of the distribution of responses, particularly in skewed data.
- **Variance (σ^2)**
  The variance (σ^2) measures the dispersion of the responses around the median. For this analysis, the sample variance is calculated using the median to eliminate outlier results. It is computed using the formula:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \tilde{x})^2$$

  where $\tilde{x}$ is the median of the responses and x_i is the value of the i-th response. For interpretation of the variance, the following thresholds are used:
  - $\sigma^2 < 0.5$: Low variability, indicating a strong consensus among respondents.
  - $0.5 \leq \sigma^2 < 1$: Moderate variability, indicating a moderate level of agreement among respondents.
  - $\sigma^2 \geq 1$: High variability, indicating a larger dispersion in opinions among respondents.

Using the converted Likert scale responses into these statistical measures, an interpretation of the data is given based on a quantitative analysis of the central tendency and variance of the population responses.

3.3. Ethical approval and consent to participate

Ethical approval

This research is approved by Delft University of Technology. This research is conducted following the ethical standards outlined by the Human Ethics Research Committee (HREC) of Delft University of Technology (HREC approval number: TUD/TBM-4275). The ethical approval confirms that the research adheres to all ethical guidelines and standards proposed by the Delft University of Technology for the protection of human participants. The experiment is conducted at Leiden University Medical Center (LUMC) using their facilities. To this end, a research appointment is provided to obtain research privilege as a TU Delft researcher at LUMC (Department: GEO, project number: 8204/70852). Accordingly, the setup of the experiment, the experiment run, and data storage were performed completely within the premises and using the facilities of LUMC.

Confidentiality and anonymity are strictly maintained throughout the research. Participants' data are anonymized and securely stored in compliance with the data protection policies of Delft University of Technology and applicable laws. Only the research team has access to the data, and no personally identifiable information was included in any publications or presentations resulting from this research.

Participation and consent

For this experiment, only participants were recruited who are actively working at LUMC. All participants were provided beforehand with a comprehensive information sheet detailing the purpose of the study, the procedures involved, potential risks, the data protection safeguards in use, and their rights as participants. Before participation, all participants provided written informed consent. The consent form explicitly stated that participation was voluntary, and participants could withdraw from the study at any time without penalty or loss of benefits to which they were otherwise entitled. If participants had any questions or concerns about the study, they were encouraged to contact the corresponding researcher, Maxim Houwink. Additionally, participants were informed that they could contact the responsible researcher, Saba Hinrichs-Krapels if they had any concerns about their rights or the ethical conduct of the study.

Medical images and swarm data

All medical data (medical images and anonymized patient data) remained within the LUMC data storage facilities throughout the entire research. For the experiment, the use of medical images was accordingly secured using, but not limited to, the following safeguards:

- Only images (with anonymized case data) were used that were available within the LUMC databases and were secured accordingly. Therefore, no transfer of medical images between external institutions or public platforms occurred.
- For use during the experiment, medical images were prepared in and accessed from the institutional Archiving and Communication System (PACS). All images were fully de-identified and were assigned anonymized case IDs.
- The medical images were only shown to participants who have seen the medical images in the past and, in line with LUMC guidelines, had the authority to access the selected images.

All swarm data produced during the experiment remained within the secured storage platform of Unanimous.ai. The swarm data was protected using the following safeguards:

- The swarm data was stored securely in a digital environment to which only the corresponding researcher had access. Hence, also Unanimous.ai did not have access to the data.
- No case data or media (medical images) were exchanged with the swarm platform. Hence, all sensitive data remained within the premises of LUMC.
- To link the question (outcome) to the corresponding case, the question only contained the de-identified case ID.

4 Evidence behind PC tumor diagnosis

Opening

Recap

Previously in Chapter 1, the topic of PC surgery was introduced along with a set of challenges currently encountered in this field. The three key takeaways from this topic were as follows:

- PC is among the leading causes of cancer-related deaths worldwide. Its low survival rate can be attributed to late and nonspecific symptoms.
- As a result of these factors, a high abandonment rate of PC surgeries is observed. This results from adverse case outcomes, where the patient is initially deemed eligible for resection but during the operation is found to have advanced or metastatic disease.
- In cases where the surgery is completed, there is still a significant portion of cases where no clear negative resection margin can be achieved.

Motivation

As the fundamental motivation of this research, this section aims to identify the current evidence behind PC tumor diagnosis and adverse surgery outcomes. To profoundly substantiate both reviews, a dedicated search string was used for each of the two topics. Since the core concepts of medical decision-making and adverse tumor assessment have been introduced previously (Chapter 1), an individual explanation for each string is given hereafter in Subsection 3.2.1. Accordingly, this literature review is structured through two subtopics: (1) PC tumor diagnosis and (2) reasons for adverse tumor assessment. Overall, this review is structured as follows.

- Selection process
- Introduction to tumor resectability for PC
- Diagnosis of tumor resectability
- Reasons for adverse tumor assessment

4.1. Synthesis

A synthesis table for reasons of adverse tumor assessment support by literature can be found in Appendix D. The synthesis table consolidates findings from the selected papers, identifying key factors that contribute to the challenges in accurately diagnosing and assessing tumors. These factors are categorized into three main domains: Organizational, MDT, and patient-related factors.

Organizational factors

Organizational factors refer to the broader structural elements in the clinical setting that indirectly influence tumor assessment. Affecting the patient more indirectly, organizational refers to the logistical and scheduling aspects. These include the selection of cases, availability of resources, and planning restrictions within healthcare institutions. Effective organizational structures ensure that the right cases are selected for surgical intervention (Hoogenboom et al., 2022; Lamb et al., 2014; Strobel et al., 2019), that adequate resources such as advanced imaging technologies are available (Fehervari et al., 2021; Frija et al., 2021; Hricak et al., 2021; Lamb et al., 2014), and that planning and coordination are streamlined to avoid delays in diagnosis and treatment (Halle-Smith et al., 2024; Hartwig et al., 2013; Taberna et al., 2020).

MDT factors

MDT factors pertain to issues that arise within the diagnostic interval: The calculated time between first presentation of the patient to the diagnosis of cancer (Swords et al., 2015). During this phase, the factors contributing to adverse tumor assessment can be classified across two sublevels. At the interpretations sublevel, these factors include operator dependence (T. M. Ahmed et al., 2024; Armato et al., 2009; Kang et al., 2021; LeBlanc et al., 2023), cognitive errors (Haj-Mirzaian et al., 2020; Hoogenboom et al., 2022), and cognitive bias (Busby et al., 2018; Kang et al., 2021; LeBlanc et al., 2023). At the technological sublevel, imaging techniques (Haj-Mirzaian et al., 2020; Hartwig et al., 2013; Hong et al., 2020; Kang et al., 2021), high-quality imaging quality (T. M. Ahmed et al., 2024; Budigi et al., 2022; Haj-Mirzaian et al., 2020; LeBlanc et al., 2023) and complete anatomical coverage (Haj-Mirzaian et al., 2020) are essential for accurate tumor assessment.

Patient-related factors

Patient-related factors are specific to the biological characteristics of the patient and the tumor. The specific biological behavior and growth patterns of the tumor, as well as how rapidly it progresses, contribute to the adverse tumor assessment since these factors are uncertain, patient-specific, and therefore difficult to predict. These include intrinsic tumor features (Budigi et al., 2022; Haj-Mirzaian et al., 2020; Kang et al., 2021; Strobel et al., 2019), the progression of the tumor (Halle-Smith et al., 2024; Hartwig et al., 2013; Swords et al., 2015; Wainberg et al., 2024), other patient-related factors (Haj-Mirzaian et al., 2020; Park et al., 2021; Strobel et al., 2019), and coexisting or underlying pathologies (T. M. Ahmed et al., 2024; Halle-Smith et al., 2024; Kang et al., 2021; LeBlanc et al., 2023).

4.2. Introduction to tumor resectability for PC

Despite continuous developments in cancer therapy, PC remains one of the most aggressive tumor entities and is one of the five most frequent causes of tumor-associated deaths in the European Union and the USA (Conroy et al., 2023; Park et al., 2021). According to Kamisawa et al. (2016), Park et al. (2021), and Veldhuisen et al. (2018), most patients with PC (in literature also referred to as Pancreatic ductal adenocarcinoma, or PDAC) present with nonspecific symptoms at an advanced stage, with disease that is not amenable to curative surgery. Accordingly, Park et al. (2021) and Veldhuisen et al. (2018) observe that the 5-year survival rate remains around 10% and is observed to have stagnated at this point despite advancements in PC treatment (Hartwig et al., 2013).

For classifying the resectability of PC, guidelines in the Netherlands (Dutch Pancreatic Cancer Group (Veldhuisen et al., 2018)) and worldwide (such as ESMO Clinical Practice Guideline, UICC TNM or NCCN (Conroy et al., 2023; Hartwig et al., 2013; Park et al., 2021)) distinguish four stages of tumor development. Across these guidelines, the generally accepted classification distinguishes four stages of PC tumor development: resectable, borderline resectable (BRPC), locally advanced (LAPC), and metastatic disease. Although the exact boundary criteria for each stage varies across guidelines (Conroy et al., 2023; Park et al., 2021; Veldhuisen et al., 2018), an overview of the tumor stages with concordant characteristics across guidelines is provided in Table 4.1.

4.3. Diagnosis of tumor resectability

At the time of diagnosis, Park et al. (2021), Kang et al. (2021) Veldhuisen et al. (2018) state that 30% of the cases are presented with LAPC while in 40%-60% the tumor has progressed to metastatic disease. For diagnosis of PC, Computed tomography (CT) is the main modality (Conroy et al., 2023; Park et al., 2021). Furthermore, for cases where the CT is deemed inconclusive or further differentiation of cystic is needed, an MRI is recommended for further assessment (Conroy et al., 2023; Park et al., 2021; Veldhuisen et al., 2018), although Hartwig et al. (2013) states it does not add information about resectability. Alternatively, a diagnostic laparotomy can be advised. This procedure, whereby a small telescope is inserted inside the abdomen using a large incision, is observed to have an increased probability of 29%-33% for identification of metastasis to other parts of the body (Takadate et al., 2021).

Clinical diagnosis of PC remains challenging. According to Allen et al. (2016), Hartwig et al. (2013), Hong et al. (2020), and Strobel et al. (2019), a demonstrated range of 19% to 41% of patients are misdiagnosed based on initial CT assessments. In this respect, misdiagnosis is caused by a diagnostic error: errors that result in incorrect, delayed, or missed diagnosis (Busby et al., 2018). Across the literature, several consequences can be identified for diagnostic errors in PC staging assessment.

- Initial misdiagnosis of PC is associated with a delayed diagnosis of PC and higher risk of locally advanced or metastatic disease at the time of presentation (Kang et al., 2021; Swords et al., 2015). In their research, when a patient with symptoms is misdiagnosed or underestimated at his first presenta-

Stage	Definition	Key characteristics
Resectable	Resectable PC tumors can be entirely removed through surgery due to the absence of tumor involvement with major blood vessels near the pancreas. Resectable PC coincides with the highest probability of complete tumor resection, also referred to as margin-negative or R_0 resection.	• Tumors are confined to the pancreas • No involvement of nearby major arteries or veins • No evidence of distant metastasis
Borderline resectable (BRPC)	Borderline resectable pancreatic cancer involves tumors that have limited involvement with nearby blood vessels, making surgical removal challenging but still possible with advanced surgical techniques and preoperative (neoadjuvant) therapy.	• Tumors may be in contact with surrounding veins or arteries but not encasing them. • Often treated with neoadjuvant therapy (chemotherapy and/or radiation) to shrink the tumor before actual resection.
Locally advanced (LAPC)	Locally advanced pancreatic cancer refers to tumors that have extended beyond the pancreas to involve nearby structures, such as major blood vessels, making surgical removal highly unlikely without significant risk.	• Tumors encase nearby arteries (such as the celiac axis or superior mesenteric artery) or veins. • No evidence of distant metastasis. • Typically not resectable at diagnosis; treatment focuses on controlling the disease
Metastatic disease	Metastatic pancreatic cancer has spread to distant parts of the body, such as the liver, lungs, or peritoneum. This stage is considered advanced and not curable with surgery.	• Presence of distant metastases regardless of the size or extent of the primary tumor. • Treatment is aimed at controlling symptoms and prolonging life

Table 4.1: Tumor staging classifications for PC

tion to a healthcare provider, Kang et al. (2021) and Swords et al. (2015) demonstrated that this results in a longer imaging misdiagnosis interval (calculated time between first imaging examination to the diagnosis of PC) - leading to a higher risk of tumor unnoticed progression toward a more severe stage.

- Another consequence of misdiagnosis is that a substantial number of patients undergo unnecessary major operations with literature suggesting a range between 19% to 33% (Hartwig et al., 2013; Hong et al., 2020; Strobel et al., 2019). For these 'adverse' cases, the patient undergoes major surgery for resection of a tumor that is initially presumed to be resectable, but the surgery is then abandoned due to findings of locally advanced or metastatic disease (Allen et al., 2016).
- As another consequence of misdiagnosis, Conroy et al. (2023) and Kang et al. (2021) state that another 10%-15% of patients undergo unnecessary laparotomy (opening the abdomen using a large incision) with lack of curative resectability identified only during the laparotomy. While these operations may not be as invasive as a complete surgical resection, these techniques can still result in physical and emotional stress for patients, along with increased healthcare costs and recovery time (Hartwig et al., 2013; Hong et al., 2020).

4.4. Reasons for adverse tumor assessment

When evaluating the reasons for the adverse tumor assessment, a wide range of factors is suggested by literature across varying domains (Appendix D). For this review, this section aims to provide an overview

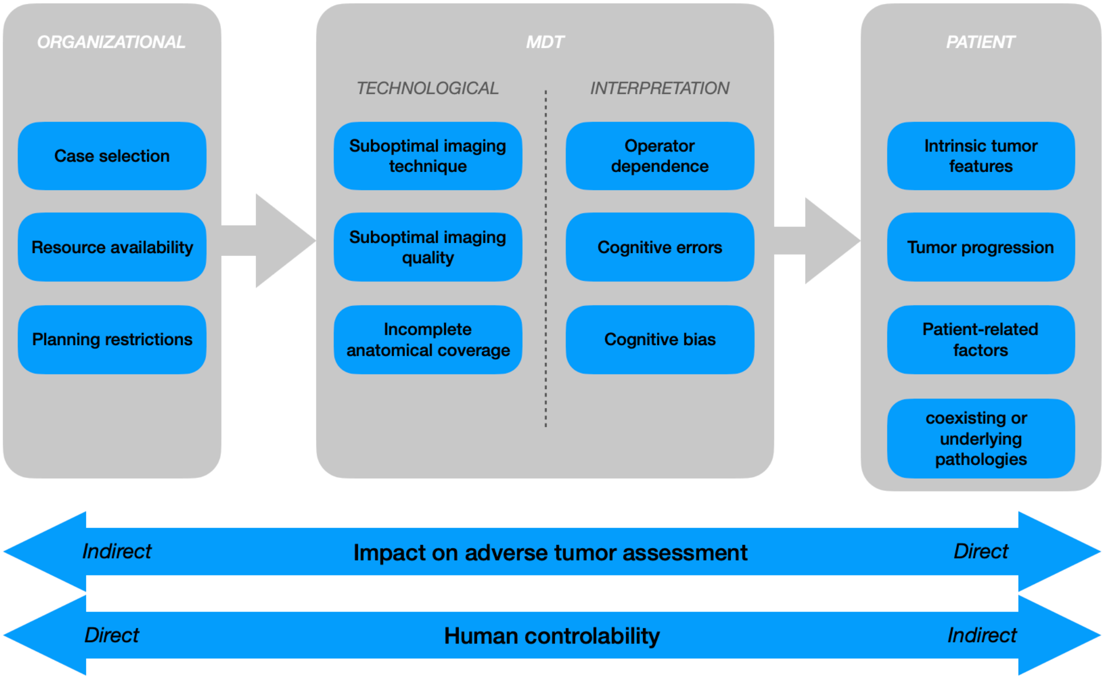


Figure 4.1: Factors of adverse case assessment based on literature

(breadth) of contributing factors observed across the literature. Correspondingly, all factors have been categorized across the three domains: organizational, MDT, and patient (Figure 4.1).

Organisational

- **Case selection**: Given that 24% (H. K. Lim et al., 2016; Pillay et al., 2016) to 45% (Maharaj et al., 2021) of the cases reviewed during an MDT leads to treatment modifications, non-selection of PC cases in pancreatic cancer can miss the benefit of comprehensive expert review for complex cases and challenge initial (individual) findings (Strobel et al., 2019). Reasons for not bringing a case to the MDT include unintentional factors (missed intrinsic tumor features, secondary findings or incidental findings (Haj-Mirzaian et al., 2020; Hoogenboom et al., 2022; Strobel et al., 2019)) or intentional factors (case deemed irrelevant for MDT discussion, imposed prioritization of cases due to time constraints (Lamb et al., 2014).
- **Resource availability**: Resource availability plays a critical role in the adverse assessment of PC. Limited access to advanced imaging equipment, such as multiphasic CT scanners and high-resolution MRI can result in suboptimal imaging quality and inaccurate tumor staging (Frija et al., 2021; Hricak et al., 2021). Additionally, the lack of availability of medical experts (Fehervari et al., 2021), particularly radiologists given their key role in tumor identification (Armato et al., 2009), can further contribute to misinterpretation of imaging studies.
- **Planning restrictions**: Planning restrictions can lead to misalignments in treatment planning. Restricted access to healthcare providers or logistical delays in patient diagnosis or treatment can lead to tumor progression before actual surgery Halle-Smith et al. (2024) and Hartwig et al. (2013)/. Alternatively, Taberna et al. (2020) states that other scheduling issues can arise due to infrequent or irregular MDT planning or improper coordination of preoperative planning.

MDT

- **Technological**
  - **Suboptimal imaging technique**: Suboptimal imaging techniques can significantly impact the inaccurate assessment of pancreatic tumors, often leading to adverse outcomes (Hartwig et al., 2013). While widely used due to their speed and availability, CT scans may not provide sufficient soft tis-

sue contrast to distinguish between pancreatic tumors and adjacent structures (Haj-Mirzaian et al., 2020). Conversely, MRI provides superior soft tissue contrast (Hong et al., 2020) but can be limited by motion artifacts and lower availability, which may lead to inappropriate selection between the two methods (Haj-Mirzaian et al., 2020). This misjudgment can result in either underestimation or overestimation of the tumor extent, thereby affecting treatment decisions and outcomes.

- **Suboptimal imaging quality**: The diagnosis of PC is routinely based on the presence of hypoenhancing pancreatic mass (as the classic primary finding), as well as several secondary imaging features (T. M. Ahmed et al., 2024; Budigi et al., 2022). However, lacking soft tissue contrast (particularly with CT), suboptimal slice thickness, and suboptimal contrast dosing can obscure the distinction between tumor tissues and adjacent structures, resulting in inaccurate delineation of tumor boundaries and involvement with surrounding veins (LeBlanc et al., 2023). In addition, particularly in MRI, motion artifacts during scans - arising from patient movement, respiratory movements, bowel movements, and cardiac and vascular pulsations - further degrade image quality (Haj-Mirzaian et al., 2020).
- **Incomplete anatomical coverage**: When imaging does not encompass the entire pancreas or its surrounding structures, crucial details about the tumor's extent and involvement with adjacent organs and vessels may be missed (Haj-Mirzaian et al., 2020). This incomplete anatomical coverage can be caused by variability in patients' breath-hold and can lead to underestimating the tumor size or missing metastatic spread to nearby lymph nodes or distant organs.

- **Interpretation**
  - **Operator dependence**: For initial diagnosis of PC, T. M. Ahmed et al. (2024) and Kang et al. (2021) state that radiologists are often the first physicians to diagnose or raise the possibility. Accordingly, given the variability in skill and experience of radiologists observed within cancer departments (Armato et al., 2009), performing and interpreting the imaging studies play a crucial role in the (in)accurate assessment (T. M. Ahmed et al., 2024; LeBlanc et al., 2023).
  - **Cognitive errors**: Cognitive error refers to misinterpretation due to physical difficulties. To this end, Haj-Mirzaian et al. (2020) identifies physical distractions, including reader fatigue, interruptions to look at other cases, and phone calls. Moreover, Haj-Mirzaian et al. (2020) names the occurrence of incidental findings, where pancreatic masses are found unintentionally after initially focusing on other clinical indications due to lack of systematic review.
  - **Cognitive bias**: Cognitive biases influence radiologist decision-making and can result in medical errors or adverse patient outcomes (Busby et al., 2018). Examples include framing bias (when the same information leads to different diagnostic conclusions depending on how clinical information is presented), anchoring bias (when one remains fixed on his original diagnostic finding despite being presented with subsequent, contrary information), and confirmation bias (the tendency to search for data supporting one's initial opinion rather than looking for data that competes the opinion) (Busby et al., 2018).

Patient

- **Intrinsic tumor features**: Intrinsic tumor features such as size-related, isoattenuating, and non-contour deforming characteristics significantly impact the assessment of pancreatic cancer (Budigi et al., 2022). Small tumors (<2cm) can be difficult to accurately measure due to obscuration by surrounding tissues (Budigi et al., 2022; Kang et al., 2021). Similarly, isoattenuating tumors (Figure 4.2), which blend in with normal pancreatic tissue, and non-contour deforming tumors, which do not alter the pancreas's shape, can be easily overlooked on standard imaging (Haj-Mirzaian et al., 2020; Kang et al., 2021).
- **Tumor progression**: Tumor progression in pancreatic cancer significantly complicates accurate assessment by introducing rapid changes in tumor size, infiltration into surrounding structures, and metastatic spread during the diagnostic interval (Swords et al., 2015; Wainberg et al., 2024). Although the maximum diagnostic interval is therefore by clinical guidelines (Veldhuisen et al., 2018), the progression of the disease can still lead to discrepancies between initial imaging and the actual extent of the disease (Halle-Smith et al., 2024).

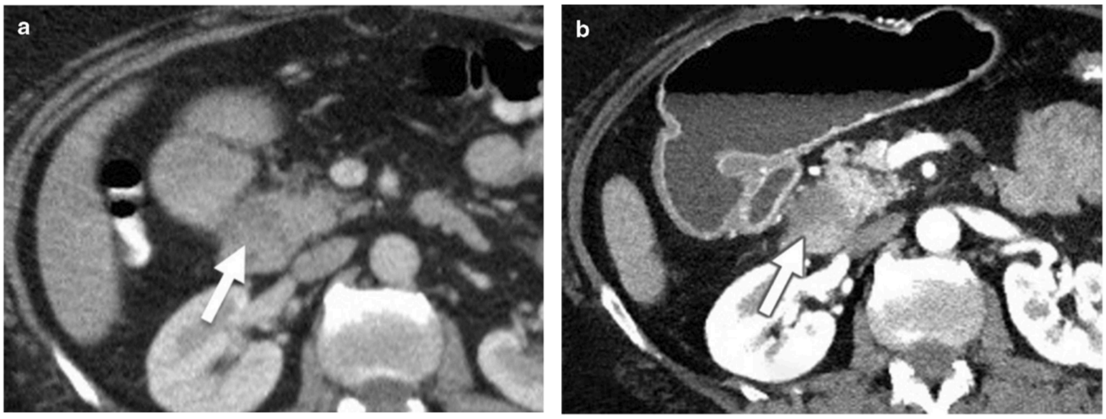


Figure 4.2: An example of how the complexity of image interpretation can result in missed diagnosis using two CT images of a 47-year old female (taken from (Kang et al., 2021)). (a) Axial contrast-enhanced CT image (portal venous phase) showing a hypoattenuated mass. The ill-defined region was initially missed and, despite multiple emergency trips to the emergency department, no cause for the patient's non-specific symptoms was identified. (b) Axial contrast-enhanced CT image in the portal venous phase 6 weeks later of the same patient after non-specific symptoms remained. Only after administration of contrast-enhancing drugs, a pancreatic head mass was observed with invasion to adjacent structures. The patient received chemotherapy and died 2 years after her initial exam

- **Patient-related factors**: Patient-related factors can impose difficulties in assessing tumor extent and involvement with adjacent organs and vessels. For instance, for patients with obesity, excessive adipose tissue can obscure the pancreas and surrounding structures on imaging studies (Strobel et al., 2019). Similarly, age-related degenerative changes and or the presence of extensive fibrosis and inflammation around the pancreas can obscure tumor visualization and impact the quality of imaging (Haj-Mirzaian et al., 2020)
- **Concurrent pancreatic pathology**: Concurrent pancreatic pathologies such as chronic pancreatitis, cystic lesions, and benign tumors significantly complicate the assessment of pancreatic cancer (Kang et al., 2021). Chronic pancreatitis can create inflammatory masses and fibrosis that mimic or obscure malignant tumors on imaging, leading to diagnostic confusion (LeBlanc et al., 2023). Similarly, benign tumors can be mistakenly interpreted for malignant ones, while cystic lesions can obscure visualization of underlying cancers (T. M. Ahmed et al., 2024; Halle-Smith et al., 2024)

4.5. Conclusion of PC resectability assessment review

Most notably, this review shows the complexity of addressing the issue of misdiagnosis for PC. The findings of this review show that the abandonment rate of PC surgeries can be attributed to misdiagnosis, where the patient is initially deemed eligible for resection but during the operation is found to have advanced or metastatic disease. As the most important output of these findings, this review identifies critical reasons for misdiagnosis across the organizational, MDT, and patient levels.

Firstly, the assessment of tumor resectability in PC distinguishes four tumor stages: resectable, borderline resectable (BRPC), locally advanced (LAPC), and metastatic disease. Misdiagnosis rates based on initial CT assessments range from 19% to 41%, leading to severe consequences such as delayed diagnoses, unnecessary major operations, and unwarranted laparotomies. Furthermore, the reasons for adverse tumor assessment can be categorized into three hierarchical layers: organizational, MDT, and patient. Organizational factors include case selection processes, resource availability, and planning constraints. Within the MDT domain, technological limitations such as suboptimal imaging techniques, poor image quality, and incomplete anatomical coverage significantly impact diagnostic accuracy. Additionally, interpretation issues including operator dependence, cognitive errors, and biases further contribute to misdiagnoses. Lastly, patient-related factors, including intrinsic tumor characteristics, tumor progression, patient-specific factors, and concurrent pancreatic pathologies, complicate the accurate assessment of tumor resectability.

Implications for this research

As stated in Subsection 1.1.5, adverse tumor assessment is what fundamentally drives the significant abandonment rate of PC surgeries. After identifying the current evidence behind the diagnosis of PC surgery and the underlying causes of adverse tumor assessment, these findings contribute to this research as follows.

- To reduce the amount of misdiagnosis, the findings of Section 4.2 suggest that the use of ASI should be targeted at improving the accuracy of the correct tumor staging. Hence, the design of the research and the collected data should be aimed at correctly assessing the tumor resectability through the identified classifications: resectable, borderline resectable (BRPC), locally advanced (LAPC), and metastatic disease.
- By identifying the landscape of factors contributing to adverse tumor assessment, a research scope can be defined by delineating the factors that can be resolved using ASI. To this end, ASI is expected to target the *MDT interpretation level*. Accordingly, the factors in this level serve as important conditions that need to be reflected in the experiment (hereafter substantiated in Chapter 5).
- By taking a systems engineering approach to the interpretation of the reasons for adverse tumor assessment and considering their interplay, the abandonment rate can be considered the output of a complex process that is currently operating suboptimally. In this process, the use of ASI is considered to specifically target the cognitive bias, cognitive error, and operator dependency factors at the MDT-interpretation level. Subsequently, distinguishing the different levels (organizational, MDT, patient-related) and underlying factors allows the delineation of the scope and identifies the potential impact of ASI in reducing the abandonment rate of PC surgeries more specifically.

5 Performance results

Opening

Recap Previously, the foundational principles of Swarm Intelligence (SI) and Artificial Swarm Intelligence (ASI) were examined. A comparative analysis between ASI and traditional crowd-based methods was conducted, as well as an analysis of potential benefits for medical decision-making. With respect to medical decision-making, a selection of the potential benefits of MDT meetings was proposed, among which the potential improved decision performance in MDT meetings

Introduction This chapter addresses the potential benefit of improved decision performance in MDT meetings. Based on the experimental setup introduced previously (Subsection 3.2.2), this chapter presents the outcomes of the experiment in assessing patient resectability, as well as an interpretation of the results. Central to this chapter is RQ2:

To what extent does the accuracy of resectability assessment for PC cases change after the use of ASI in MDT meetings?

For addressing RQ2, the following hypothesis is tested:

H_1: In determining the resectability of patient X, the use of ASI in MDT meetings results in a reduced probability of adverse tumor assessment compared to discussion-based MDT meetings

To this end, this chapter is structured as follows:
- Practical experimental setup
- Results

5.1. Practical experimental setup

5.1.1. Practical RCT diagram

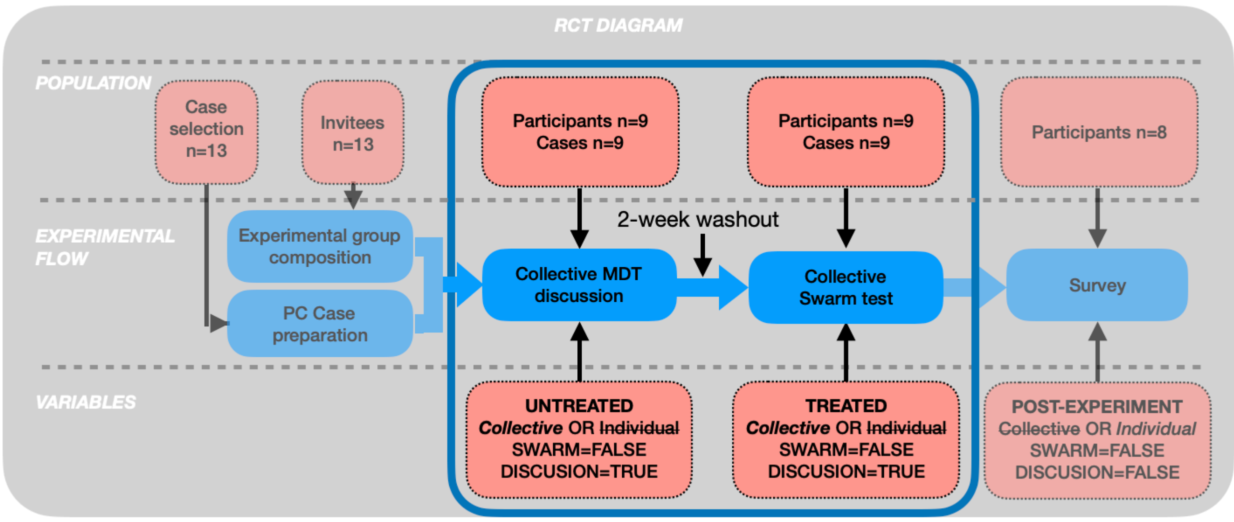


Figure 5.1: Practical RCT diagram with focus on experiment (blue scope)

On 8 May 2024 and 22 May 2024, the experiment was conducted based on the practical experimental setup at LUMC (Figure 5.1). Compared to the theoretical setup (Subsection 3.2.2), the experiment was conducted subject to the following adjustments:

- Due to the time constraints of both sessions (60 minutes), nine cases were discussed across both sessions. A rank in cases was prepared in advance based on the expected ambiguity of each PC case.

- Due to availability constraints, nine participants took part in both experiments. The experimental group remained nearly consistent across both sessions. During the second session, one participant missed out on one case which was then replaced by a stand-by expert.
- Due to availability reasons, the wash-out period between both sessions was extended to fourteen days.
- Due to facility reasons, whilst the collective discussion was conducted at the regular MDT meeting room at LUMC, the swarm experiment was conducted at a LUMC computer room other than the MDT meeting room. This ensured the availability of computers for every participant.

5.1.2. Final experimental group composition

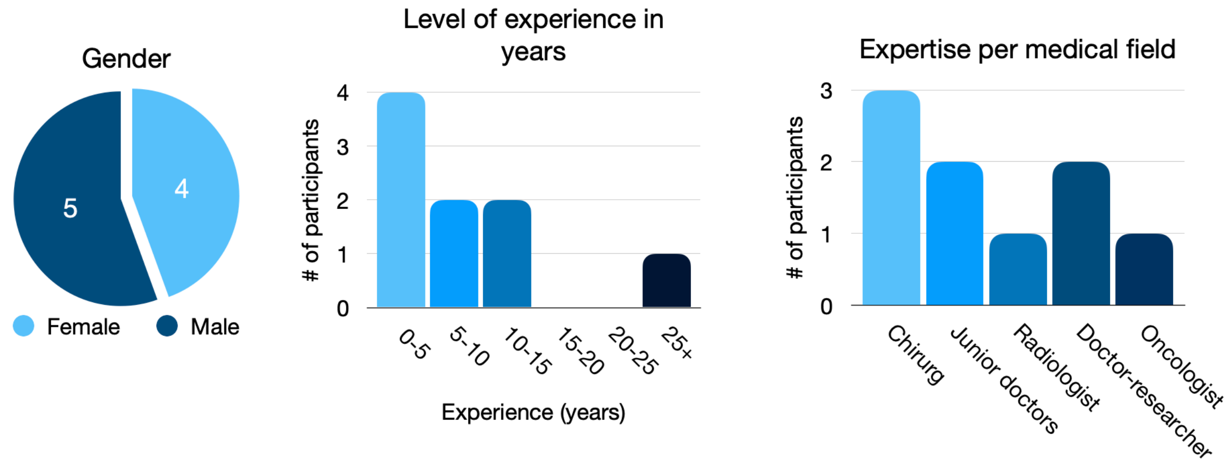


Figure 5.2: Sociodemographic data of experimental group

The final experimental group is composed of nine ($n = 9$) participants actively working or researching in the field of PC at Leiden University Medical Center (LUMC), The Netherlands (Figure 5.2). Distributions of the experimental group include a man:women ratio of $55.6\% : 44.4\%$ with mean experience level (in years) $\mu = 8.61$ and median experience level (in years) $\tilde{x} = 5.0$. Moreover, the population consists of a medical expertise range of $n = 21$ participants for specializations the surgeon ($n = 3$), junior doctors ($n = 2$), radiologist ($n = 1$), doctor-researchers ($n = 3$) and oncologists ($n = 1$).

5.2. Results

5.2.1. Comparison structure

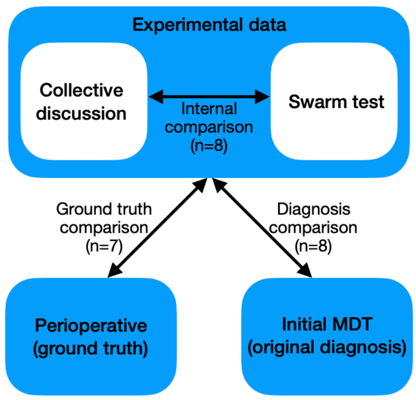


Figure 5.3: Overview of comparison structure

As introduced previously (Section 3.2), the following three comparisons are conducted:

1. Internal comparison of experimental results: collective discussion versus swarm test
2. Ground truth comparison: experiment versus ground truth
3. Diagnosis comparison: experiment versus initial MDT

For valid comparison results, the case set is modified resulting in an adapted case selection for the internal comparison ($n = 8$), ground truth comparison ($n = 7$), and diagnosis comparison ($n = 8$) due to the following reasons:

- For case SWARM-2359 (*), the patient was treated with induction therapy (chemotherapy, ChT) between the initial MDT and perioperative monitoring. This information was revealed after conducting the experiment. While the perioperative outcome reflects the outcome after the patient receives induction therapy, the patient data presented during the experiment consisted of pre-operative CT-scans data before receiving induction therapy. Subsequently, this case is included in the internal comparison and diagnosis comparison but excluded from the ground truth comparison.
- For case SWARM-1268 (**), a crucial consideration is that the outcome of this case is subject to human error: the radiologist presenting the cases indicated to have accidentally provided inconsistent information between the collective discussion and the swarm test. At the swarm test, information was provided about 'vascular contact with SMA', whereas this information was omitted during the collective discussion. Therefore, given the potential bias that the outcome of case SWARM-1268 may have been subjected to this information consistency, this case is excluded for the internal comparison, the ground truth comparison, and the diagnosis comparison

5.2.2. Results table

An overview of the experimental results versus the case data (ground truth) is given in Table 5.1.

Ground truth versus experimental outcomes				
	Experimental data		Case data	
Case-ID	Collective discussion	Swarm test	Initial MDT	Perioperative (Ground truth)
SWARM-8945	LAPC	LAPC	Resectable	Irresectable
SWARM-4156	LAPC	LAPC	LAPC	Irresectable
SWARM-6324	Resectable	Resectable	Resectable	Metastasis
SWARM-4694	BRPC	BRPC	BRPC	BRPC
SWARM-3167	BRPC	BRPC	BRPC	BRPC
SWARM-9641	LAPC	LAPC	BRPC	LAPC
SWARM-4498	Resectable	Resectable	Resectable	Resectable
SWARM-2359*	Metastasis	Metastasis	LAPC	BRPC
SWARM-1268**	No conclusion	BRPC	Resectable	Metastasis

Table 5.1: Outcomes of experimental data versus case data

5.2.3. Internal comparison (collective discussion versus swarm test

The duration of the collective discussion for all cases ($n = 9$) was $t = 43m30s$ (average time per case: $t = 4m49s$), while the duration of the swarm test was $t = 38m10s$ (average time per case: $t = 4m14s$). When analyzing the experimental results, eight cases from the total case set ($n = 8/8$) resulted in identical outcomes between the collective discussion versus the swarm test:

Case set internal comparison		
Case-ID	Collective discussion	Swarm test
SWARM-8945	LAPC	LAPC
SWARM-4156	LAPC	LAPC
SWARM-6324	Resectable	Resectable
SWARM-4694	BRPC	BRPC
SWARM-3167	BRPC	BRPC
SWARM-9641	LAPC	LAPC
SWARM-4498	Resectable	Resectable
SWARM-2359*	Metastasis	Metastasis

Table 5.2: Outcomes for internal comparison

Concordant (100%)

- SWARM-8945: LAPC
- SWARM-4156: LAPC
- SWARM-6324: Resectable
- SWARM-4694: BRPC
- SWARM-3167: BRPC
- SWARM-9641: LAPC
- SWARM-4498: Resectable
- SWARM-2359: Metastasis

When analyzing the severity of tumor assessment, the 100% concordance level implies an equally severe assessment between collective discussion and swarm test. Both sessions resulted in cases SWARM-6324 and SWARM-4498 to be assessed resectable, SWARM-4694 and SWARM-3167 to be assessed BRPC, SWARM-8945, SWARM-4156, and SWARM-9641 to be assessed LAPC and SWARM-2359 to be assessed metastasis. Accordingly, this is observed in the confusion matrix (Figure 5.4) where all assessment outcomes are situated on the diagonal (G_1, E_1=1, G_2, E_2=2, G_3, E_3=3, G_4, E_4=1) and no cases are observed off the diagonal.

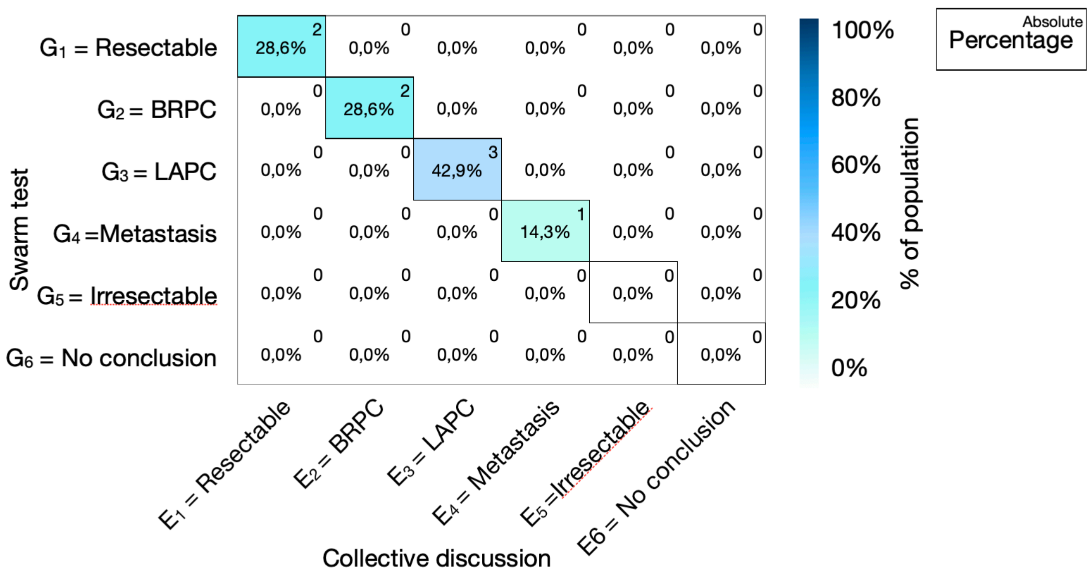


Figure 5.4: Confusion matrix internal comparison

5.2.4. Ground truth comparison: Experimental outcomes versus ground truth

When analyzing the experimental results against the perioperative outcomes, a ground truth concordance of four out of nine cases ($n = 4/7$) is observed:

Concordant (57.1%)

- SWARM-4694: BRPC
- SWARM-3167: BRPC
- SWARM-9641: LAPC
- SWARM-4498: Resectable

Discordant (42.9%) (Experiment vs perioperative)

- SWARM-8945: LAPC vs Irresectable
- SWARM-4156: LAPC vs Irresectable
- SWARM-6324: Resectable vs Metastasis

Case set ground truth comparison		
Case-ID	Experimental outcomes	Perioperative outcomes
SWARM-8945	LAPC	Irresectable
SWARM-4156	LAPC	Irresectable
SWARM-6324	Resectable	Metastasis
SWARM-4694	BRPC	BRPC
SWARM-3167	BRPC	BRPC
SWARM-9641	LAPC	LAPC
SWARM-4498	Resectable	Resectable

Table 5.3: Outcomes for ground truth comparison

When analyzing the discordant cases, the perioperative outcomes of all cases prove to be more severe compared to the experimental outcomes. Firstly, SWARM-8945 and SWARM-4156 (G_5, E_3=2) were assessed LAPC during the experiment, while the perioperative outcome revealed the tumor staging to be irresectable. Secondly, SWARM-6324 (G_4, E_1=1) was assessed resectable during the experiment, while the perioperative outcome showed the tumor to be metastatic. As observed correspondingly in the CM (Figure 5.5), all discordant case outcomes SWARM-6324 (G_4, E_1), SWARM-8945 (G_5, E_3) and SWARM-4156 (G_5, E_3) are situated below the diagonal.

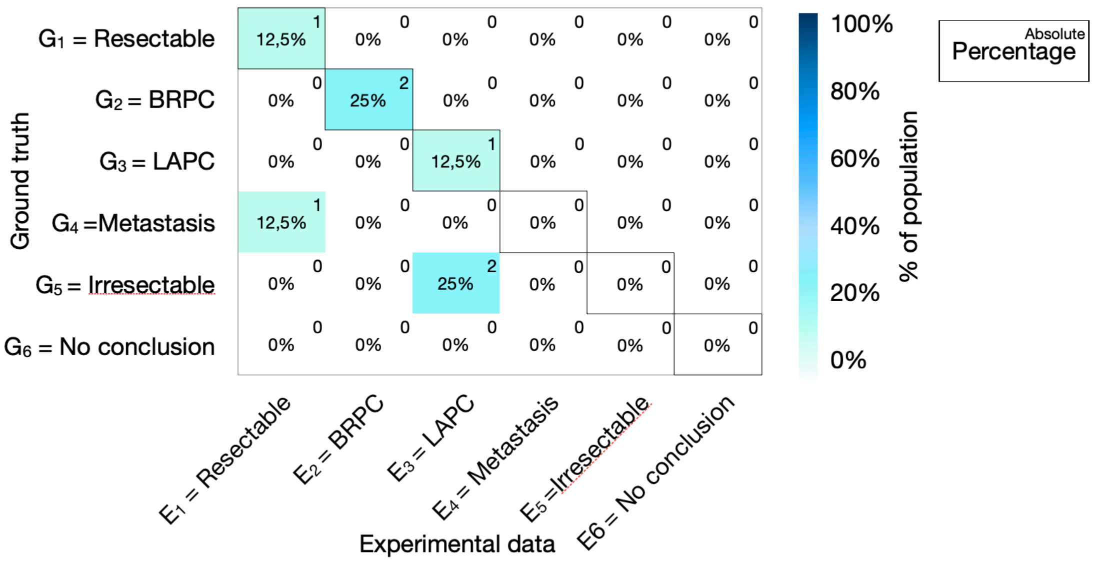
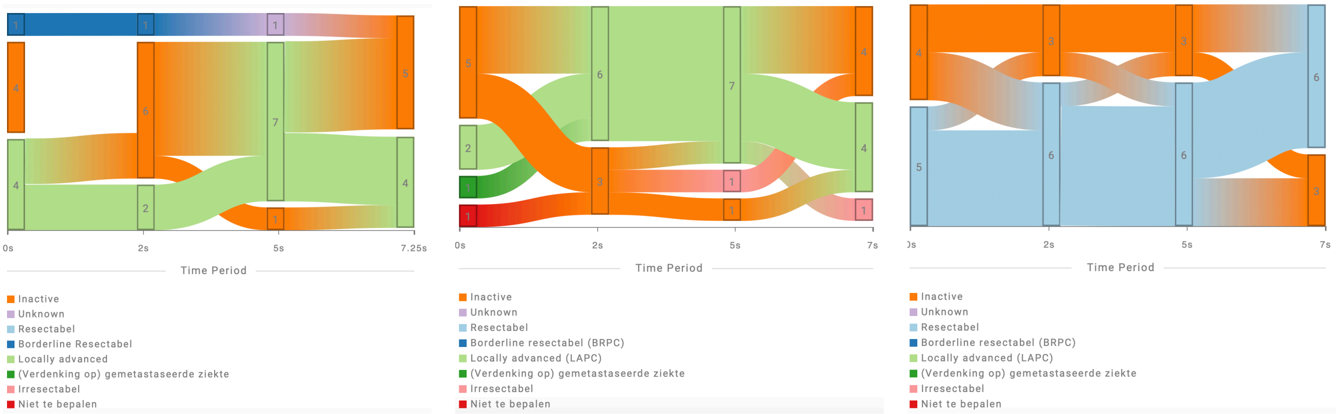


Figure 5.5: Confusion matrix experimental data versus ground truth



(a) SWARM-8945 **(b)** SWARM-4156 **(c)** SWARM-6324

Figure 5.6: Faction change graphs of discordant cases

For the discordant cases of the ground truth comparison, additional data is collected in the form of faction change graphs through the Swarm.ai platform (Figure 5.6) and treatment history (through the LUMC database). For functional comparison of the faction change graphs, the portions of 'inactive users' in the faction change graphs (indicated in orange) are neglected. When comparing the faction change graphs, the most significant level of deliberation is observed for SWARM-4156. Firstly, the graph of SWARM-4156 shows to be the only case that concludes in a (slightly) dissident outcome between LAPC (green) and irresectable (pink). Second, the graph of SWARM-4156 shows notable deliberation (decision heterogeneity), implying that the participants have changed their input to a notable extent throughout the swarm session. On the contrary, SWARM-8945 and SWARM-6324 show a strictly uniform outcome when neglecting the inactive users. Moreover, less deliberation (decision heterogeneity) is observed for both cases, where particularly the outcome of SWARM-6324 shows to have developed unanimously towards resectable (blue) when neglecting the inactive users.

Furthermore, when analyzing the treatment history of the discordant cases, the following observations are made:

- **SWARM-8945**
  A more detailed look at the LUMC treatment history data concludes that this case involved a considerable rate of tumor progression. Whereas the CT scan of the initial MDT (which was also used during the experiment) showed no vascular involvement, a second scan was taken within 30 days in which a notable degree of vascular involvement (±180°) was observed. Thereafter, surgery was undertaken less than 30 days after the second scan, during which an advanced degree of tissue involvement, vascular involvement, and encasement of the pancreatic head was revealed- therefore leading to an irresectable tumor.
- **SWARM-4156**
  An in-depth analysis of the case history showed that this was a boundary case between BRPC and LAPC, where the MDT could not establish a definitive conclusion based on the available medical imaging at that time. Resultingly, exploratory research was conducted more than six weeks after the date of initial CT scans, thus allowing the opportunity for excessive tumor progression. During the exploratory research, the tumor was then deemed to be overly metastasized by the LUMC surgical team to achieve a reasonable chance for R_0 resection. The case was therefore deemed irresectable.
- **SWARM-6324**
  When analyzing the treatment history, it is recorded that the metastatic disease was only revealed during exploratory research as no evidence for advanced tumor staging could be discerned using the initial medical imaging presented at the original MDT meeting. However, these initial medical images which did not provide any evidence for advanced tumor staging were also the images presented for this case during the experiment.

5.2.5. Diagnosis comparison: Experimental outcomes versus initial MDT outcomes

When comparing the experimental outcomes against the initial MDT outcomes, a diagnostic concordance of five out of nine cases ($n = 5/8$) is observed:

Concordant (62.5%)

- SWARM-4156: LAPC
- SWARM-6324: Resectable
- SWARM-4694: BRPC
- SWARM-3167: BRPC
- SWARM-4498: Resectable

Discordant (37.5%) (Experiment vs initial MDT)

- SWARM-8945: LAPC vs Resectable
- SWARM-9641: LAPC vs BRPC
- SWARM-2359: Metastasis vs LAPC

Case set diagnosis comparison		
Case-ID	Experimental outcomes	Initial MDT
SWARM-8945	LAPC	Resectable
SWARM-4156	LAPC	LAPC
SWARM-6324	Resectable	Resectable
SWARM-4694	BRPC	BRPC
SWARM-3167	BRPC	BRPC
SWARM-9641	LAPC	BRPC
SWARM-4498	Resectable	Resectable
SWARM-2359*	Metastasis	LAPC

Table 5.4: Outcomes for diagnosis comparison

When analyzing the discordant cases, tumor resectability for all discordant cases is assessed to a more advanced stage during the experiment compared to the initial MDT. This is reflected in the confusion matrix (Figure 5.7) as all discordant cases are situated above the diagonal (G_1,E_3=1, G_2,E_3=1 and G_3,E_4=1).

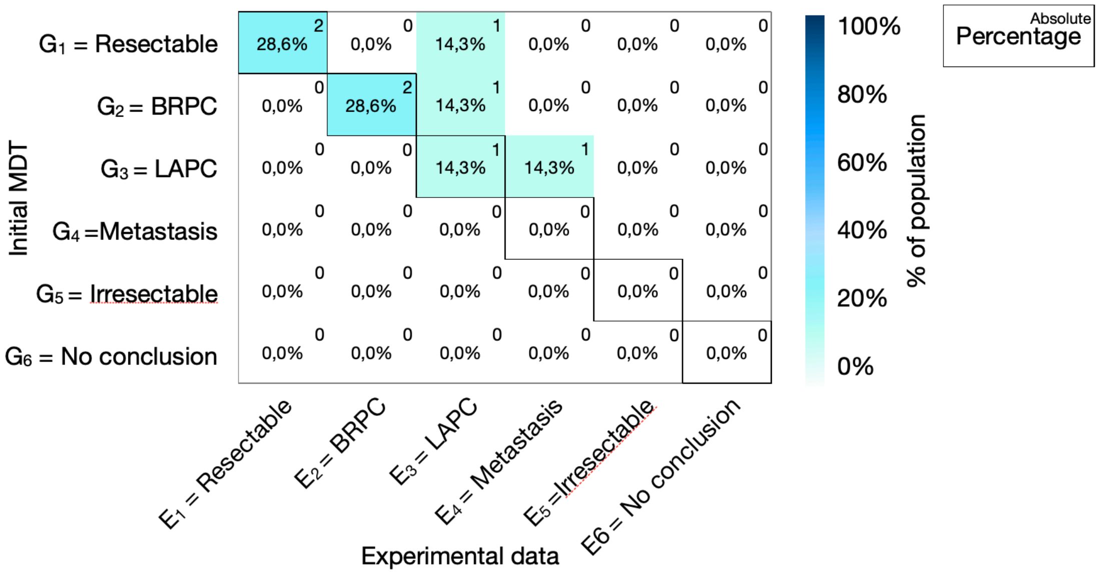


Figure 5.7: Confusion matrix experimental data versus initial MDT

6 Social influence results

Opening

Recap Previously, several benefits of ASI were identified that could potentially benefit surgical decision-making in MDT meetings. From these benefits, the previous chapter investigated the extent to which the implementation of ASI could improve resectability assessment compared to regular MDT meetings. In terms of experimental data (collective discussion versus swarm test), no significant change in assessment accuracy was observed.

Introduction This chapter addresses another potential benefit of ASI in MDT meetings: the potential capability of ASI to minimize social influence in MDT meetings. Based on the experimental setup introduced previously (Subsection 3.2.2), this chapter presents an analysis of the a posteriori survey results alongside its interpretation. Central to this chapter is RQ3:

To what extent can the use of ASI drive reduction factors that reduce social influence during MDT meetings?

To address RQ3, the following hypotheses are tested:

H_2: The use of ASI in MDT meetings can drive the reduction factors evaluation, attractiveness, and equal efforts to expedite reducing social loafing on decision outcome compared to discussion-based MDT meetings
H_3: The use of ASI in MDT meetings can drive the reduction factors anonymity and equal decision influence to expedite reducing social bias on decision outcome compared to discussion-based MDT meetings

To this end, this chapter is structured as follows:
- Practical survey group composition
- Results
- Interpretation

6.1. Practical survey group composition

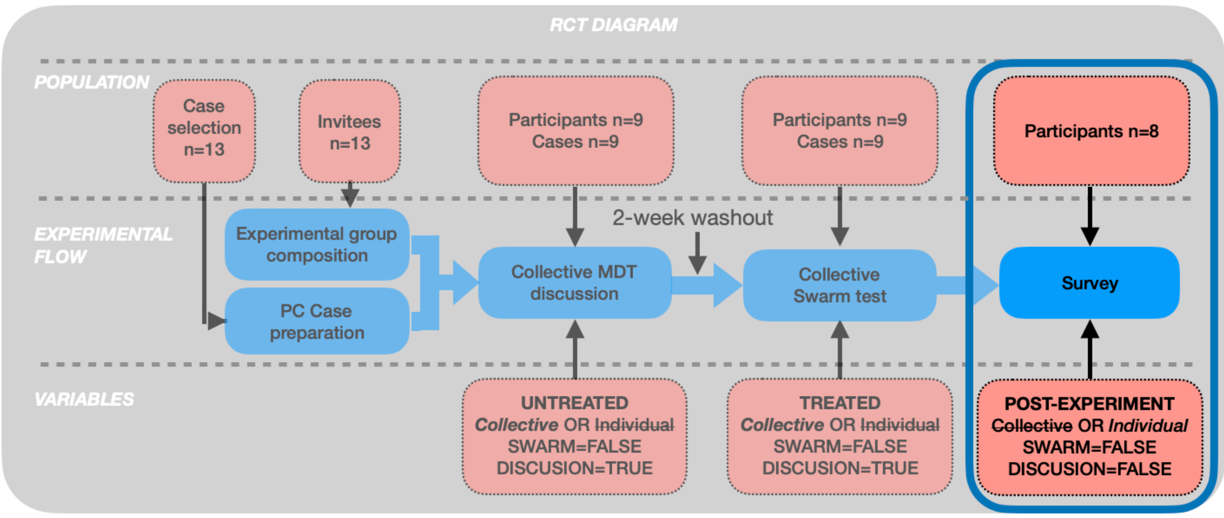


Figure 6.1: Practical RCT diagram with focus on the survey (blue scope)

Between 24 May 2024 and 31 May 2024, the survey was completed by eight LUMC respondents ($n = 8$) (Figure 6.2) - each of which participated in both sessions of the experiment. Therefore, the number of survey respondents is equivalent to 88.89% of the experimental participants (Figure 5.2). Distributions of the survey group include a man:women ratio of $50\% : 50\%$ with mean experience level (in years) $\mu = 6.25$ years and median experience level (in years) $\tilde{x} = 2.5$. Furthermore, the survey population consists of a medical expertise range of $n = 1, 2$ for surgeons ($n = 2$), junior doctors ($n = 2$), radiologists ($n = 1$), doctor-researchers ($n = 2$) and oncologist ($n = 1$).

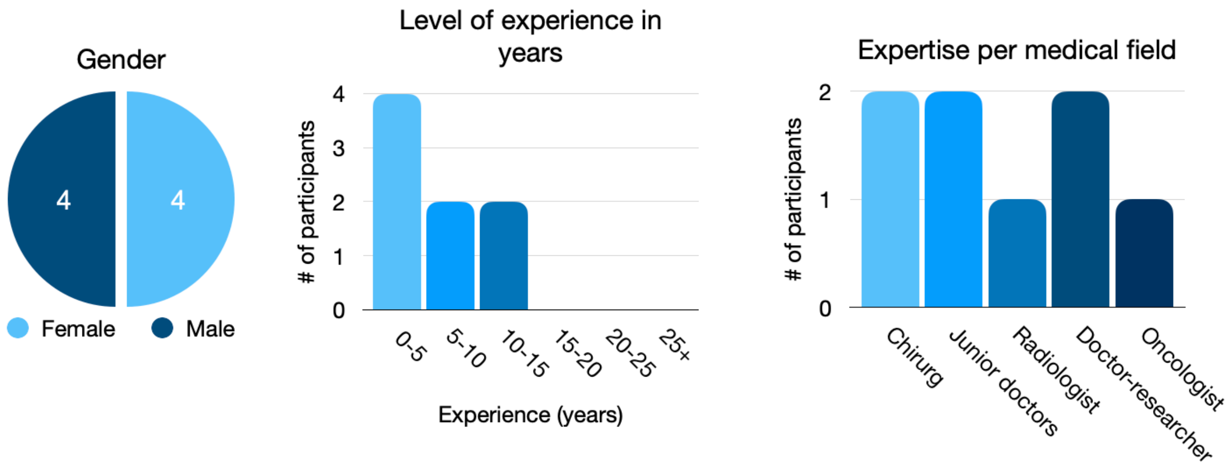


Figure 6.2: Sociodemographic data of survey group

6.2. Results

6.2.1. Social loafing

		Per indicator			Average per reduction factor		
		Mean (μ)	Median ($\tilde{x}$)	Variance (σ^2)	Mean (μ)	Median ($\tilde{x}$)	Variance (σ^2)
Evaluation	L_1	2.75	2	1.714	2.68	2.25	1.38
	L_2	2.625	2.5	1.036			
Attractiveness	L_3	3.625	4	0.714	3.63	4.00	0.57
	L_4	3.625	4	0.429			
Equal efforts	L_5	2.875	3	0.714	3.19	3.50	0.79
	L_6	3.5	4	0.857			
Average		3.17	3.25	0.91			

Table 6.1: Descriptive statistics for social loafing

Based on the social loafing results, a results table (Table 6.1), frequency histogram (Figure 6.3), and heatmap (Figure 6.4) are presented. As introduced in Section 3.2, the results only include the mean, median and variance due to the population size ($n = 8$).

Based on the results table, the following observations can be made:

- **Evaluation (L_1 and L_2)**
  - The mean scores for the evaluation indicators (L_1 and L_2) are $\mu = 2.75$ and $\mu = 2.625$, respectively, indicating a slightly below average agreement among respondents that evaluation by ASI encourages more input compared to regular MDT meetings.
  - The median values for these indicators are $\tilde{x} = 2$ and $\tilde{x} = 2.5$, which aligns closely with the mean values, confirming that the central tendency of responses is around neutral to disagree.
  - The variances are $\sigma^2 = 1.714$ and $\sigma^2 = 1.036$, respectively. Variance for both L_1 and L_2 indicates high variability, suggesting diverse opinions on individual performance evaluation.
- **Attractiveness (L_3 and L_4)**

- The mean scores for attractiveness (L_3 and L_4) are $\mu = 3.625$, indicating that respondents generally agree that the attractiveness of ASI contributes positively to their input.
- The median for these indicators is $\tilde{x} = 4$, suggesting that the majority of respondents agreed with these statements.
- The variances are $\sigma^2 = 0.714$ and $\sigma^2 = 0.429$, respectively, indicating low variability in responses, suggesting a strong consensus on the attractiveness of ASI.

- **Equal Efforts (L_5 and L_6)**
  - The mean scores for equal efforts (L_5 and L_6) are $\mu = 2.875$ and $\mu = 3.5$, indicating a positive trend towards agreement that ASI promotes equal participation.
  - The median values are $\tilde{x} = 3$ and $\tilde{x} = 4$, respectively, indicating that responses range from neutral to agree.
  - The variances are $\sigma^2 = 0.714$ and $\sigma^2 = 0.857$, respectively, indicating low to moderate variability, suggesting a consensus among participants that ASI promotes equal input.
- **Overall Trends**
  - The average mean score across all indicators is $\mu = 3.17$, suggesting a general agreement among respondents that ASI contributes to reducing social loafing.
  - The average median is $\tilde{x} = 3.25$, which supports the average mean, indicating a central tendency towards agreement.
  - The overall variance is $\sigma^2 = 0.91$, indicating low to moderate variability in responses, suggesting that while there is general agreement, individual opinions do vary. However, considering a variance of $\sigma^2 = 1$ is commonly accepted as the threshold between low versus high variance for Likert scale data (Savage, 1957), the overall variance is still found to be within tolerable limits.

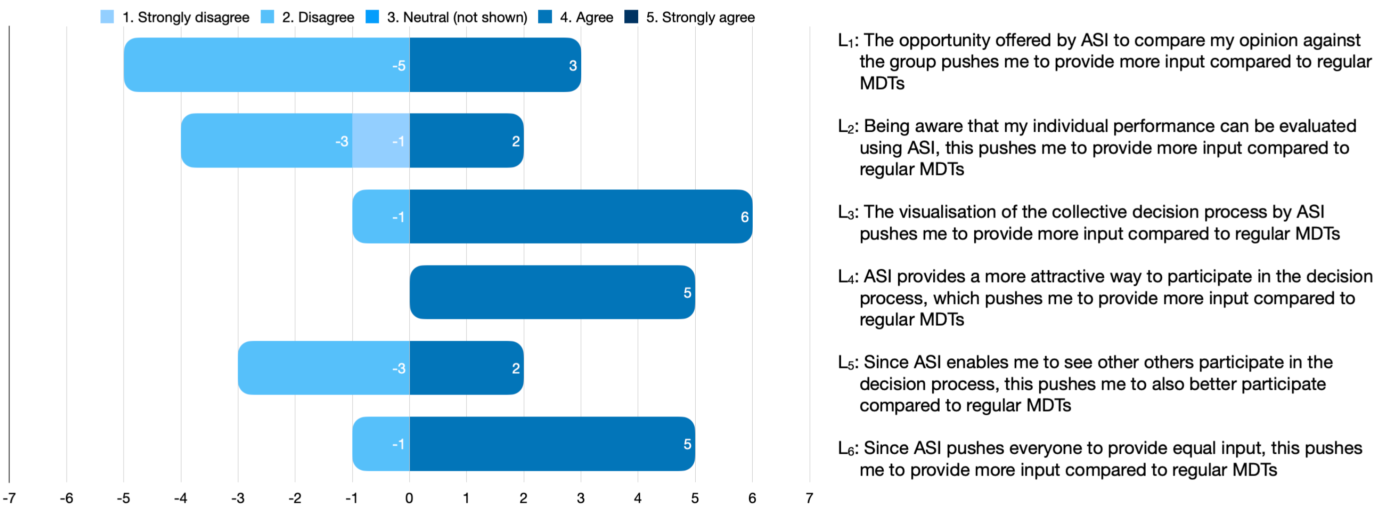


Figure 6.3: Histogram of social loafing results

6.2.2. Social bias

Based on the results table, the following observations can be made:

- **Anonymity (B_1, B_2, and B_3)**
  - The mean scores for the anonymity indicators (B_1, B_2, and B_3) are $\mu = 3.375$, $\mu = 3.25$, and $\mu = 2.875$, respectively. These scores indicate a general agreement among respondents that anonymity provided by ASI encourages them to provide input, with B_1 showing the highest level of agreement.
  - The median values for these indicators are $\tilde{x} = 4$, $\tilde{x} = 3.5$, and $\tilde{x} = 3$, respectively, which align closely with the mean values, confirming a central tendency towards agreement.
  - The variances are $\sigma^2 = 0.836$, $\sigma^2 = 0.786$, and $\sigma^2 = 0.696$, respectively. These values indicate low

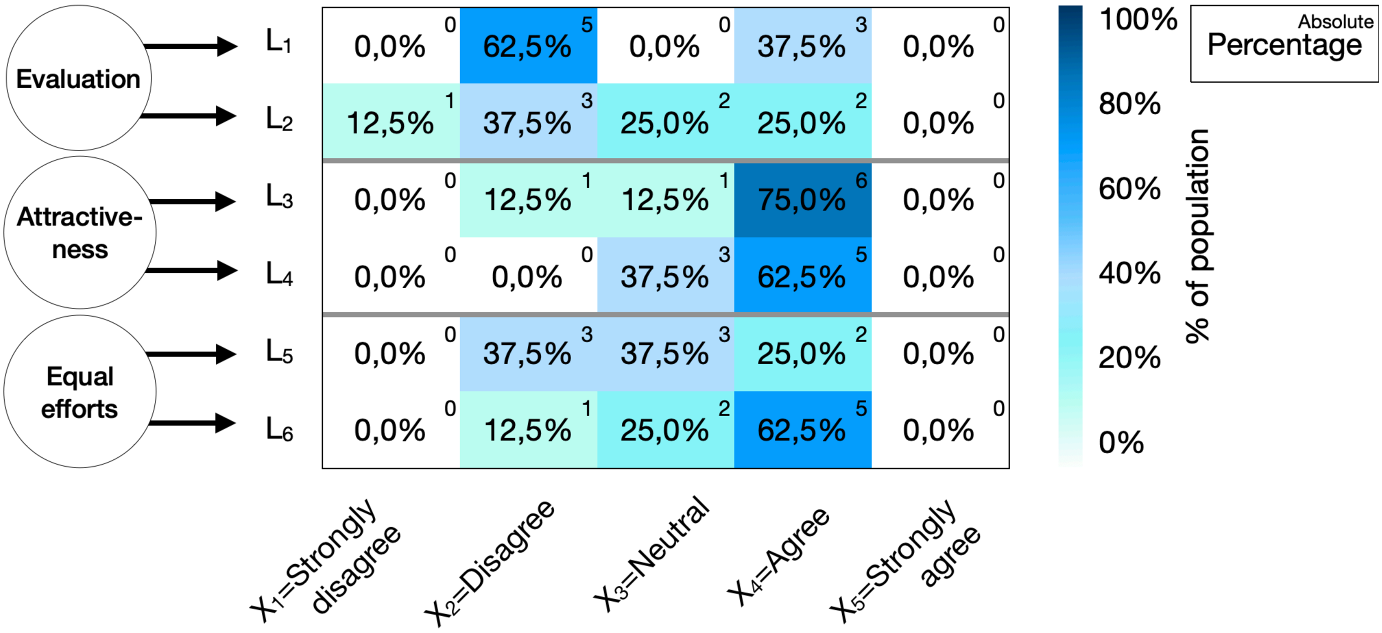


Figure 6.4: Heatmap of social loafing results

		Per indicator			Average per reduction factor		
		Mean (μ)	Median ($\tilde{x}$)	Variance (σ^2)	Mean (μ)	Median ($\tilde{x}$)	Variance (σ^2)
Anonymity	B_1	3.375	4	0.836	3.17	3.50	0.77
	B_2	3.25	3.5	0.786			
	B_3	2.875	3	0.696			
Equal decision influence	B_4	3.125	3.5	1.255	3.46	3.50	0.98
	B_5	4.125	4	0.984			
	B_6	3.125	3	0.696			
Average		3.31	3.50	0.88			

Table 6.2: Descriptive statistics for social bias

to moderate variability in responses, suggesting some diversity in opinions regarding the influence of anonymity on providing input.

ruby Copy code

- **Equal Decision Influence (B_4, B_5, and B_6)**
  - The mean scores for equal decision influence indicators (B_4, B_5, and B_6) are $\mu = 3.125$, $\mu = 4.125$, and $\mu = 3.125$, respectively. These scores indicate a positive trend towards agreement that ASI promotes equal decision influence, with B_5 showing the highest level of agreement.
  - The median values are $\tilde{x} = 3.5$, $\tilde{x} = 4$, and $\tilde{x} = 3$, respectively, indicating that responses range from neutral to agree.
  - The variances are $\sigma^2 = 1.255$, $\sigma^2 = 0.984$, and $\sigma^2 = 0.696$, respectively, indicating moderate to high variability. This suggests a wider range of opinions regarding the equal decision influence provided by ASI.
- **Overall Trends**
  - The average mean score across all indicators is $\mu = 3.31$, suggesting a general agreement among respondents that ASI contributes to reducing bias in decision-making.
  - The overall median is $\tilde{x} = 3.50$, which supports the average mean, indicating a central tendency towards agreement.
  - The overall variance is $\sigma^2 = 0.88$, indicating low to moderate variability in responses. This suggests that while there is general agreement, individual opinions vary. However, considering a variance

of $\sigma^2 = 1$ is commonly accepted as the threshold between low versus high variance for Likert scale data, these results indicate that the variability is within an acceptable range.

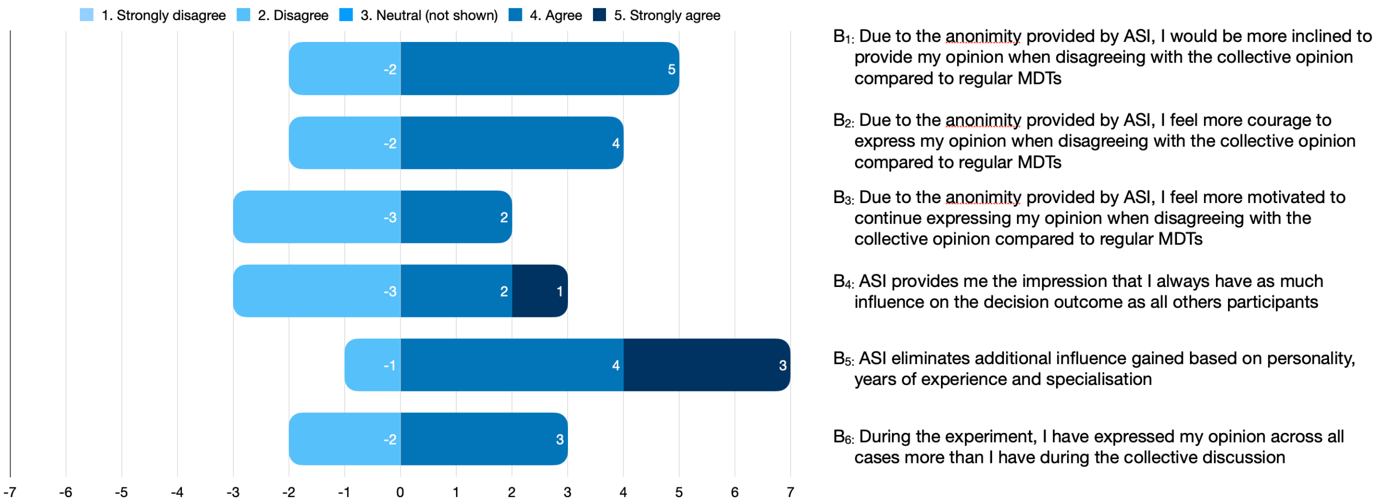


Figure 6.5: Histogram of social bias results

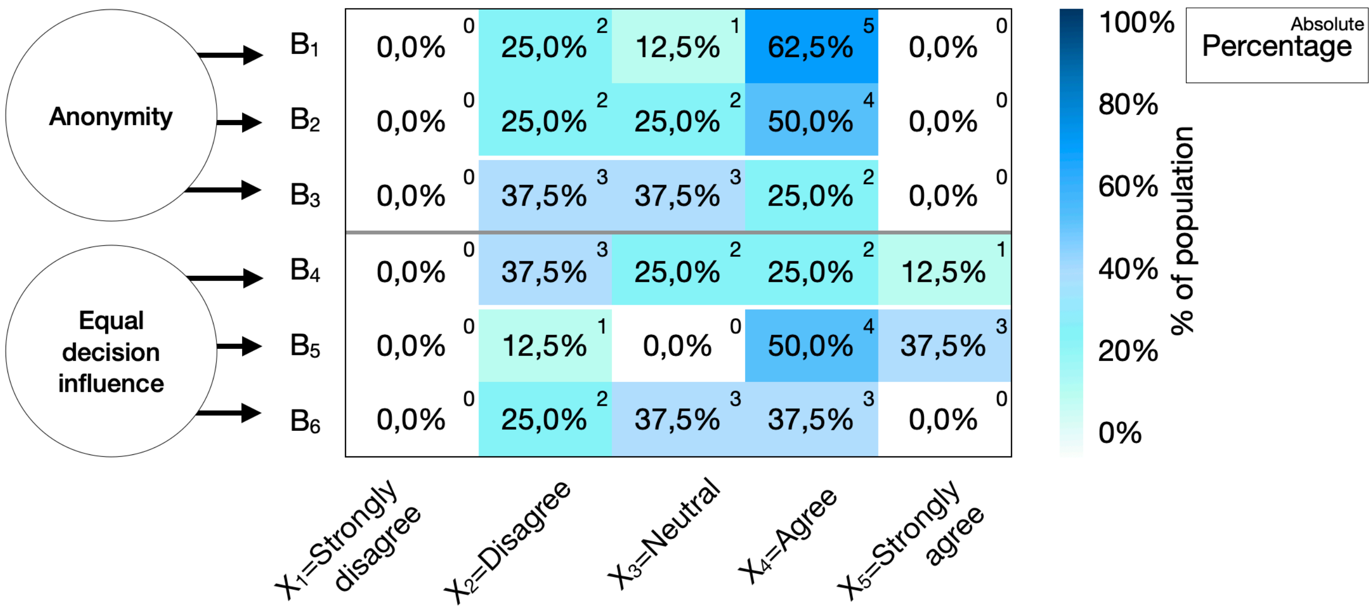


Figure 6.6: Heatmap of social bias results

Closing

This chapter addresses RQ3 by testing the following hypotheses:

Could the use of ASI impact social influence during MDT meetings?

H_2: The use of ASI in MDT meetings can drive the reduction factors evaluation, attractiveness, and equal efforts to expedite reducing social loafing on decision outcome compared to discussion-based MDT meetings.
H_3: The use of ASI in MDT meetings can drive the reduction factors of anonymity and equal decision influence to expedite reducing social bias on decision outcome compared to discussion-based MDT meetings.

Based on the survey results conducted across a population nearly consistent with the experimental group ($n = 8$), it is observed that the overall population shows a modest positive attitude toward the capability of ASI to reduce social influence. For social loafing, the central tendency ($\mu = 3.17$, $\tilde{x} = 3.25$) is in line with H_2, suggesting modest support for the notion that the implementation of ASI can help reduce social loafing. For social bias, the central tendency of its results ($\mu = 3.31$, $\tilde{x} = 3.50$) also aligns with H_3, albeit with a lower degree of dispersion in the results compared to social loafing. Nonetheless, the slight positive tendency of both social factors suggests a positive influence of ASI on reducing social influence.

7 Discussion

Opening

Recap
In Section 1.2, the research objective was introduced as follows:

The objective of this research is to examine whether ASI can improve the surgical decision-making process, by assessing the impact of using ASI over discussion-based MDT meetings on the accuracy of tumor assessment and reduced social influence

Accordingly, Chapter 4 identified the foundational classification of tumor diagnosis, as well as a selection of reasons that contribute to adverse tumor assessment. Based on these findings, Chapter 5 and Chapter 6 presented the results of the experiment and survey for the performance and social influence analyses, respectively. The performance analysis showed a 100% concordance level for the internal comparison, a 57.1% concordance for the ground truth comparison, and a 62.6% concordance level for the diagnosis comparison. In turn, the social influence analysis showed nearly identical central tendency results for social loafing ($\mu = 3.17$, $\tilde{x} = 3.25$) and social bias ($\mu = 3.31$, $\tilde{x} = 3.50$).

Introduction
This chapter discusses the results of the three analyses. It thereby examines the results that specifically contribute to the research objective. For the performance analysis, the internal comparison thereby serves as the primary findings to be discussed for determining the impact of ASI compared to discussion-based MDT meetings. On the contrary, the ground truth comparison and diagnosis comparison are considered secondary findings for which a more detailed discussion can therefore be found in Appendix F.
Moreover, this chapter also aims to provide a broader perspective on the results of this research. This is done by positioning the findings in the existing landscape of academic literature and by reflecting on how this research can contribute to further research in this field. Therefore, this chapter builds on the findings of the discussion using the following section:

- Limitations
- Implications for future research

By setting up an experiment using a simulated MDT meeting, this research deployed a novel approach to test the use of ASI in a real-world setting for PC. To this end, this research first conducted a literature review to establish the current evidence behind PC diagnosis and adverse tumor assessment. Based on these findings, the experiment was then set up at LUMC by forming an experimental group of nine medical experts to test the impact of using ASI in a simulated MDT meeting on determining tumor assessment accuracy. To this end, the experimental group first assessed a set of nine past patient cases using a discussion-based MDT, followed by the assessment of the same case set using ASI. Using this approach, the use of ASI was tested in a practical setting that would resemble the real-world use of the technology in MDT meetings more closely compared to existing studies testing ASI in medical decision-making. Lastly, testing the technology in the field also allowed for collecting data on the impact of the technology on social influence. On this account, a survey was conducted among the participants to evaluate their perceived impact of using ASI on reducing social loafing and social bias.

When identifying the current evidence behind the tumor staging and reasons for adverse tumor assessment in the field of PC, the abandonment rate (as the fundamental motivation of this research) can be interpreted as the outcome of a complex process (Section 1.5) in which numerous other factors are involved. By taking a systems perspective on this diagnostic process and its embeddedness in its socio-technical setting, the abandonment rate can be regarded as an indicator of a socio-technical system performing sub-optimally as a result of the interplay between all of its interrelated factors. Consequently, by distinguishing the different levels (organizational, MDT, patient-related) and the total landscape of factors contributing to adverse tumor assessment, this allowed to identification and scope of the specific level (MDT interpretation) and factors (operator dependency, cognitive error, and cognitive bias) to be targeted by ASI more accurately.

In determining the impact of ASI on assessment accuracy, the internal concordance of n=8/8 (100%) suggests that the use of ASI in this experiment neither enhances nor reduces the assessment accuracy. Therefore, the results of this experiment thus do not support H_1 in the notion that the use of ASI yields a reduced probability for adverse tumor assessment. However, given the setup of this experiment and the potential impact of operator dependency (Section 4.4), these results do not exclude that the use of ASI could still improve the assessment accuracy of an MDT when the role of operator dependency is eliminated. Furthermore, the results may also highlight the role of other causes for adverse tumor assessment including operator dependency and technological reasons (Section 4.4). Altogether, the performance results suggest careful consideration of the issue, scope, and stage within the decision-making process when using ASI to reduce tumor misdiagnosis.

When positioning the findings of this research to existing studies, the equal assessment accuracy between ASI and discussion-based MDT meetings opposes the results of existing studies examining the accuracy of ASI in medical decision-making. In contrast to L. Rosenberg et al. (2018) and A. Shah (2019) showing improved accuracy when using ASI for the diagnosis of pneumonia and meniscal lesions, respectively, the performance results of this research do not provide support for improved assessment accuracy in the diagnosis of PC tumors. In general, the fundamentally distinct method of this research may have contributed to opposing results compared to existing studies: Whereas L. Rosenberg et al. (2018) and A. Shah (2019) both compared the ASI performance to the individual performance where experts worked in isolation, this research measured the ASI performance against the group performance of a simulated, discussion-based MDT. Furthermore, when comparing the setup of this research to existing studies more profoundly, the study of L. Rosenberg et al. (2018) is essentially based on determining *whether* medical images show the presence of pneumonia, by assessing the probability that a CT-scan shows the presence of pneumonia (provided on a 1%-100% scale) and transforming the responses into a binary classification (yes versus no). On the contrary, this research deals with tumor *resectability* or *staging* (Section 3.2), therefore addressing an issue in which more factors need to be considered for correct assessment. Furthermore, R. Shah et al. (2023) used the interpretation of an experienced or the most experienced radiologist as ground truth for assessing the performance of the individual performance and swarm performance. On the other hand, this research uses the perioperative (post-surgery) case data owned by LUMC to eliminate the possibility of human error in establishing the ground truth.

With respect to social influence, the results suggest ASI has some impact in reducing social influence in MDT meetings - which aligns with H_2 and H_3. Based on the central tendency of social loafing, the respondents positively support the impact of ASI on loafing reduction factors - particularly favoring task attractiveness and equal efforts when using ASI in MDT meetings. For social bias, its central tendency supports the impact of ASI on bias reduction factors anonymity, and equal decision influence - but with a lower level of dispersion compared to social loafing. Within the field of ASI, the findings of this research are in line with the review of L. Rosenberg and Willcox (2019a), in which they explain how the technology is supposed to limit social bias effects. However, data or evidence on the true impact of ASI on social influence currently remains unavailable. Across a wider field of studies, the findings of this research are in line with the notion that a standardized method can reduce the exposure to social influences in decision-making (Cheng, 2020; C. Taylor et al., 2012; E. Taylor, 2020). While varying forms of aggregation methods exist, implementing structured methods including ASI can thus increase objectivity and fairness throughout the process. Moreover, the significance of equal decision influence and anonymity is also in support of other studies highlighting the role of these aspects in reducing social influence (Dietrich & Spiekermann, 2022; Javadi Khasraghi & Hirschheim, 2022; E. Taylor, 2020).

Limitations

While the methods deployed in this research gained new empirical findings on the use of ASI in MDT meetings, this research is subject to several limitations. This section presents the most significant limitations of each method separately.

Limitations to the performance analysis

Operator dependency

The setup of this research entails that the radiologist first presents the findings to the participants, after which the experimental group collectively discusses the findings. As noted previously (Section 4.4), this imposes significant dependency on the skills, interpretation, and performance of the radiologist when concluding the assessment outcome. Moreover, the fact that the presenting radiologist is included in the experimental setup ascertains to what extent the performance analysis outcomes can be fully attributed to the collective performance of the experimental group versus to what extent the outcomes have been influenced by the radiologist.

Lack of ambiguity

By assessing tumor resectability, the issue addressed in this research may have been too apparent for the participants to conclude in various interpretations. Alternatively, the issue of tumor resectability did not require a sufficient level of deliberation across multiple disciplines. Both underlying reasons follow from the low level of faction changes (Appendix F) and the observed uniformity during the decision-making process. Evaluation of the experiment concludes that the lack of ambiguity can be caused by concluding the tumor assessment after the presentation of the radiologist and the quality or tumor visibility of available medical images.

Sample size of case set

Due to the time constraints of both experimental sessions and the availability of the LUMC experts, the small sample of PC-cases (practical case set, n=9) provides increased exposure to potential sampling bias (Panzeri et al., 2008), which may originate from selecting a case set that does not appropriately reflect the typical case set. This may lead to skewed outcomes and may restrain the generalizability of the outcomes, restricting their applicability to broader clinical settings.

Tumor progression

For the ground truth comparison (Subsection 5.2.4), an important limitation is that the CT scans shown during the experiment do not necessarily reflect the *true* resectability of the tumor associated with the perioperative outcome (ground truth). In fact, between the date of the CT scan and registration of perioperative outcome (date of surgery), a period of several weeks or months can elapse where the case can be subject to tumor progression. In such case, the resectability on a CT scan may have been for instance 'correctly' classified as BRPC, but the observed perioperative resectability may have progressed toward LAPC (Section 4.4).

Patient history

Due to privacy and ethics restrictions, the experiment only used past patient cases that have been previously discussed and mostly treated at LUMC from 01 January 2020 onwards. Since some participants may thus have been familiar with the cases, this likely results in biased outcomes as these participants are then also aware of the perioperative outcomes of the cases.

Washout-period

Despite a practical washout period of two weeks (Chapter 5), the sequential research setup is still vulnerable to carryover effects (Cnops et al., 2022): the influence that the outcomes or experiences of a prior experimental condition have on the responses in subsequent conditions. Accordingly, participants may have been biased by the outcomes of the collective discussion.

Limitations to the social influence analysis

Population size

Given the constrained availability of the LUMC experts, a survey was used to gather data for the evaluation of social influence. However, the small population size strictly limits the significance of the social influence results and generalization to a wider population. For this reason, it is explicitly mentioned that the survey results only provide an *indication* as to whether ASI could minimize social influence.

Single-point measurement (post-experiment)
For assessing the impact of ASI on social loafing and social bias, no baseline measurement could be taken for establishing an individual benchmark due to the availability of the LUMC experts. Correspondingly, the single-point measurement only allows for providing an indication *whether* the use of ASI has a potential impact on social loafing and social bias effects, thus does not suffice in assessing *the magnitude* of the impact which would require longitudinal data (Hsieh et al., 2020; A. Shah, 2019).

Selection of reduction factors
For the evaluation of social loafing and social bias, the outcome is largely affected by the included indicators as including other factors may have significantly impacted the outcomes of the social influence analysis. While the selection of reduction factors for this survey was based on their relevance to MDT meetings, the final selection did not involve objective assessment and may have therefore been subjected to (selection) biases.
Causal relationship
While the results show overall support that ASI can drive reduction factors demonstrated in the literature (Section 2.2), this research does not formally establish whether there is a direct, *causal* relationship between the use of ASI as independent variable and social factors (social loafing or social bias) as dependent variables. This limits the significance of the social influence findings.

Statement formulation
The statements have been formulated as a cause-effect statement (e.g. 'ASI provides a more attractive way to participate in the decision process, which pushes me to provide more input compared to regular MDTs'). Whilst the cause-effect formulations are intended to more directly assess the impact of the reduction factor on social bias and social loafing, this formulation may have introduced a biased result since participants may agree with the cause (e.g. a more attractive way), but do not necessarily agree with the effect (e.g. pushes them to provide more input).

Implications for future research
This section presents how the findings of this research can contribute to future research. Following the limitations identified previously (Chapter 7) as well as the analysis of enablers and barriers (Appendix G), an overview of potential avenues for further research is given below:

Targeting alternative reasons for adverse tumor assessment
As the fundamental motivation of this research, the abandonment rate of PC surgeries is considered the adverse outcome of a complex process that involves multiple factors than the decision of tumor assessment based on CT-images alone (as identified in Section 4.4). For this reason, as the scope of this research is targeted at eliminating cognitive bias and cognitive error Section 2.4, this may suggest further research into other causes of misdiagnoses and reducing role in the abandonment rate of PC surgeries. Following the limitation of including the radiologist's presentation in this experiment (Chapter 7), investigating the impact of operator dependency would be most closely related to this research. Furthermore, the obscured tumor visibility (relating to the image quality of the CT scan) of two cases (Appendix F) suggests further research at the technological layer as well.

Testing alternative issues related to PC treatment
Following the limitation that the issue of tumor resectability may appear to be too unambiguous (Chapter 7, future research could advance on the use of ASI by addressing other, more ambiguous issues in medical decision-making that involve more diverse interpretations from multiple disciplines. Subsequently, this could potentially extend toward issues where no 'ground truth' can be established but rather requires some form of consensus (Appendix G and Section 2.4). Further avenues in this respect can include issues affecting complex surgical interventions, multi-disciplinary treatment plans, or complicated ethical considerations.

Exploring the effect of ASI and social influence
Following the limitations of the survey of this research (Chapter 7), further research is suggested to investigate

whether a causal interference relationship exists between the use of ASI and the effect on social loafing behaviour, social bias, or other behavioral factors. Further research can advance on this topic by increasing the population size of surveys in this area, as well as by conducting a longitudinal study that evaluates social influences over a sequence of measurements. Although this may not directly affect the outcomes of tumor assessment, understanding the effect of ASI on social influence does contribute to a more objective decision-making process.

Decision influence based on experience and expertise

Based on the findings of Appendix G, further research could assess the accuracy of decision-making through ASI while assigning different weights to individual opinions based on the experience and expertise of each medical expert. While equal decision influence is one of the core features of ASI in achieving improved decision-making (Section 2.3), experience and expertise are critical factors for the reliability of individual assessments by medical experts - based on the input given in Appendix G. Therefore, this approach can help balance the benefits of democratized input with the practical reality that more experienced or specialized experts may provide more reliable insights.

Usage protocols

Following the findings of Appendix G, another avenue for further research involves specifying the practical role of ASI in medical decision-making by refining guidelines on the use of the technology in MDT meetings. Suggestions for such protocol guidelines include (1) limiting the use of ASI to only those cases considered controversial for avoiding unnecessary delayed meetings, (2) instructions for regulating the type of issues that can be served best through ASI, and (3) a hybrid meeting structure that combines decision-making through conventional discussion (for the ability to exchange viewpoints between experts) with ASI for issues where discussion does not conclude a decisive answer. Future research could test these protocol guidelines in real-world settings, assessing their impact on meeting efficiency, time to complete, decision quality, and participant satisfaction.

8 Conclusion

To improve the accuracy of resectability assessments in patients with PC, this research investigates the potential benefits of implementing ASI within MDT meetings. Central to this research is the following main question:

In patients with potentially resectable PC, how does using ASI compared to discussion-based MDTs impact the tumor assessment accuracy and social influences in an MDT setting?

To conclude this research, a consolidated overview of the conclusion for each subquestion is provided before concluding the main research question below:

RQ1: What are the reasons for adverse tumor assessment in PC?
For RQ1, a literature review is conducted to establish the evidence behind the causes of adverse tumor assessment for PC. Firstly, the results show that the adverse tumor assessment fundamentally comes down to misclassification of tumor resectability across the following stages: resectability, borderline resectability (BRPC), locally advanced (LAPC), and metastatic disease. When identifying the reasons for this misclassification, a landscape of factors was identified across (1) the organizational level, (2) the MDT level, and (3) the patient level. Acknowledging the abandonment rate of PC surgeries as the result of misclassification in the diagnostic stage served as the foundation for setting up the experiment to address tumor resectability. Furthermore, the identification of all factors across these levels enabled to delineation of the scope of impact for ASI and to determine the position of the technology within the socio-technical system.

RQ2: To what extent does the accuracy of resectability assessment for PC cases change after the use of ASI in MDT meetings?
For RQ2 (Chapter 5), a performance analysis is conducted to test the extent to which the use of ASI leads to improved assessment accuracy in tumor resectability using an experiment. Most importantly, the internal comparison reveals complete concordance (100%) between the collective discussion and swarm test outcomes across eight cases, suggesting that the ASI use does not significantly improve or reduce the accuracy of resectability assessments compared to the simulated MDT meeting. Therefore, ASI appears to perform equivalently to the discussion-based MDT meeting. However, the ground truth comparison (57.1% concordance) and diagnosis comparison (62.5% concordance) show more significant differences in assessment. While the ground truth comparison shows a slight under-assessment of tumor severity during the experiment, the diagnosis comparison shows a more severe tumor assessment during the experimental outcomes compared to initial MDT outcomes.

RQ3: Could the use of ASI impact social influence during MDT meetings?
For RQ3 (Chapter 6), a social influence analysis is conducted among the experiment participants using a survey. First, the analysis concludes with moderate support for the notion that the use of ASI can reduce social loafing in MDT meetings. Second, the analysis concludes with a slight support of the notion that the use of ASI contributes to a reduced level of social bias in MDT meetings as well, although a larger level of dispersion is observed for this notion. However, based on the single-point measurement of the survey, the results of the social influence analysis only suggest *whether* the use of ASI could impact social influence but does not suffice in establishing *the extent* to which the technology can reduce social loafing or social bias.

Main RQ

In concluding the main research question, the findings of this research show that ASI in this experiment does not significantly enhance or reduce the accuracy of tumor resectability assessments compared to conventional MDT methods. However, it must be considered that this experiment simulated an MDT meeting where tumor assessment occurred after and based on the presentation of the radiologist. Correspondingly, given the significant impact of operator dependency on MDT interpretation of the tumor, these results do not exclude the possibility that ASI could still improve assessment accuracy when used in other forms or setups, potentially without the presence of a radiologist. Furthermore, the abandonment rate of PC surgeries is regarded as the outcome of a complex process influenced by many alternative interrelated factors beyond MDT meetings only. Therefore, the unchanged assessment accuracy despite using ASI may highlight a more significant role of other factors, such as operator dependency and technological limitations, suggesting further research in these areas.

Regarding social influence, the results indicate that ASI has a modest impact on reducing social loafing and bias in MDT meetings. Participants reported a positive impact of ASI on factors such as task attractiveness, equal efforts, anonymity, and equal decision influence. However, the study did not establish a direct causal relationship between the use of ASI and social influences due to the compact survey form and small population size. The selection of survey factors, based on their relevance to MDT meetings, may also have influenced the outcomes since including additional relevant factors could significantly affect the social influence analysis results. Altogether, the positive results of the social influence analysis suggest further research on the impact of ASI on social influence.

Following the findings of this research, several avenues for future research are suggested in this field. First, given the complexity of the factors leading to the abandonment rate of PC surgeries, other reasons for misdiagnoses and their roles in the treatment process are suggested. Specifically, exploring the impact of operator dependency and the quality of CT scan images serve as two notable directions based on the experimental findings of this research. Additionally, future studies could advance this research by studying the use of ASI for more complex issues in medical decision-making, such as complex surgical interventions, multi-disciplinary treatment plans, and ethical considerations. Following the setup of this research, other studies could thereby extend to consensus-based questions rather than ground truth questions. Furthermore, future research could also focus on establishing a causal relationship between ASI use and social influences. While this may not directly lead to improved results, understanding the effect of ASI on social influence can contribute to a more objective decision-making process. Lastly, based on the enablers and barriers analysis, other avenues for future research include (1) exploring decision influence based on experience and expertise and (2) varying usage protocols.

References

Ahmed, H., & Glasgow, J. (2012). *Swarm Intelligence: Concepts, Models and Applications* (tech. rep.).

Ahmed, M. I., Spooner, B., Isherwood, J., Lane, M., Orrock, E., & Dennison, A. (2023). A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus*. https://doi.org/10.7759/cureus.46454

Ahmed, T. M., Chu, L. C., Javed, A. A., Yasrab, M., Blanco, A., Hruban, R. H., Fishman, E. K., & Kawamoto, S. (2024). Hidden in plain sight: commonly missed early signs of pancreatic cancer on CT. https://doi.org/10.1007/s00261-024-04334-4

Al Harbi, S., Aljohani, B., Elmasry, L., Baldovino, F. L., Raviz, K. B., Altowairqi, L., & Alshlowi, S. (2024). Streamlining patient flow and enhancing operational efficiency through case management implementation. *BMJ Open Quality*, *13*(1). https://doi.org/10.1136/bmjoq-2023-002484

Allen, V. B., Gurusamy, K. S., Takwoingi, Y., Kalia, A., & Davidson, B. R. (2016, July). Diagnostic accuracy of laparoscopy following computed tomography (CT) scanning for assessing the resectability with curative intent in pancreatic and periampullary cancer. https://doi.org/10.1002/14651858.CD009323.pub3

Armato, S. G., Roberts, R. Y., Kocherginsky, M., Aberle, D. R., Kazerooni, E. A., MacMahon, H., van Beek, E. J., Yankelevitz, D., McLennan, G., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Caligiuri, P., Quint, L. E., Sundaram, B., Croft, B. Y., & Clarke, L. P. (2009). Assessment of Radiologist Performance in the Detection of Lung Nodules. Dependence on the Definition of "Truth". *Academic Radiology*, *16*(1), 28–38. https://doi.org/10.1016/j.acra.2008.05.022

Badampudi, D., Wohlin, C., & Petersen, K. (2015). Experiences from using snowballing and database searches in systematic literature studies. *ACM International Conference Proceeding Series*, *27-29-April-2015*. https://doi.org/10.1145/2745802.2745818

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. https://doi.org/10.1287/orsc.14.3.312.15162

Befort, K., Baltaxe, D., Itt, C., & Durbin, D. (2018). Artificial swarm intelligence technology enables better subjective rating judgment in pilots compared to traditional data collection methods. *Proceedings of the Human Factors and Ergonomics Society*, *3*, 2033–2036. https://doi.org/10.1177/1541931218621459

Beni, G. (2005). From swarm intelligence to swarm robotics. *Lecture Notes in Computer Science*, *3342*, 1–9. https://doi.org/10.1007/978-3-540-30552-1{_}1

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Brooks, F. P. (1995, August). *Mythical Man-Month, The: Essays on Software Engineering* (tech. rep.). https://tudelft.on.worldcat.org/oclc/31166617

Budigi, B., Oliphant, M., & Itri, J. (2022, July). Pancreatic Adenocarcinoma: Diagnostic Errors, Contributing Factors and Solutions. https://doi.org/10.1016/j.acra.2021.10.025

Busby, L. P., Courtier, J. L., & Glastonbury, C. M. (2018). Bias in radiology: The how and why of misses and misinterpretations. *Radiographics*, *38*(1), 236–247. https://doi.org/10.1148/rg.2018170107

Camera, G., Garratt, R., Monnet, C., & Garratt, R. (2024). *Truth by Consensus: A Theoretical and Empirical Investigation Truth by Consensus: A Theoretical and Empirical Investigation †* (tech. rep.).

Chakraborty, A., & Kar, A. K. (2017). Swarm intelligence: A review of algorithms. In *Modeling and optimization in science and technologies* (pp. 475–494, Vol. 10). Springer Verlag. https://doi.org/10.1007/978-3-319-50920-4{_}19

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Chen, X., Li, S., Zhang, Y., Zhai, Y., Zhang, Z., & Feng, C. (2022). Different drives of herding: An exploratory study of motivations underlying social conformity. *PsyCh Journal*, *11*(2), 247–258. https://doi.org/10.1002/pchj.515

Cheng, J. T. (2020, June). Dominance, prestige, and the role of leveling in human social hierarchy and equality. https://doi.org/10.1016/j.copsyc.2019.10.004

Choi, J. J., Gribben, J., Lin, M., Abramson, E. L., & Aizer, J. (2023). Using an experiential learning model to teach clinical reasoning theory and cognitive bias: an evaluation of a first-year medical student curriculum. *Medical Education Online*, *28*(1). https://doi.org/10.1080/10872981.2022.2153782

Choi, J. J., Mhaimeed, N., Al-Mohanadi, D., & Mahmoud, M. A. (2024). Medical residents' perceptions of group biases in medical decision making: a qualitative study. *BMC Medical Education*, *24*(1). https://doi.org/10.1186/s12909-024-05643-4

Cnops, V., Iyer, V. R., Parathy, N., Wong, P., & Dawe, G. S. (2022, April). Test, rinse, repeat: A review of carryover effects in rodent behavioral assays. https://doi.org/10.1016/j.neubiorev.2022.104560

Conroy, T., Pfeiffer, P., Vilgrain, V., Lamarca, A., Seufferlein, T., O'Reilly, E. M., Hackert, T., Golan, T., Prager, G., Haustermans, K., Vogel, A., & Ducreux, M. (2023). Pancreatic cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up□. *Annals of Oncology*, *34*(11), 987–1002. https://doi.org/10.1016/j.annonc.2023.08.009

Cristancho, S. M. (2021). On collective self-healing and traces: How can swarm intelligence help us think differently about team adaptation? *Medical Education*, *55*(4), 441–447. https://doi.org/10.1111/medu.14358

Cui, X., Zhang, Y., Zhou, Y., Huang, T., & Li, Z. (2021). Measurements of team workload: A time pressure and scenario complexity study for maritime operation tasks. *International Journal of Industrial Ergonomics*, *83*. https://doi.org/10.1016/j.ergon.2021.103110

Della Croce, Y. (2023). Epistemic Injustice and Nonmaleficence. *Journal of Bioethical Inquiry*, *20*(3), 447–456. https://doi.org/10.1007/s11673-023-10273-4

Dietrich, F., & Spiekermann, K. (2022). *Deliberation and the Wisdom of Crowds 1* (tech. rep.). https://ssrn.com/abstract=4145405

Donald Harvey, R., Mileham, K. F., Bhatnagar, V., Brewer, J. R., Rahman, A., Moravek, C., Kennedy, A. S., Ness, E. A., Claire Dees, E., Percy Ivy, S., Ebbinghaus, S. W., Schenkel, C., & Uldrick, T. S. (2021, May). Modernizing clinical trial eligibility criteria: recommendations of the ASCO-friends of cancer research washout period and concomitant medication work group. https://doi.org/10.1158/1078-0432.CCR-20-3855

Engelen, Y. (2024). *Explaining socio-technical systems* (tech. rep.). https://flatland.agency/en/blog/

Fehervari, M., Hamrang-Yousefi, S., Fadel, M. G., Mills, S. C., Warren, O. J., Tekkis, P. P., & Kontovounisios, C. (2021, May). A systematic review of colorectal multidisciplinary team meetings: An international comparison. https://doi.org/10.1093/bjsopen/zrab044

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019, April). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. https://doi.org/10.1002/ijc.31937

Fried, L. (1991). Team size and productivity in systems development: Bigger does not always mean better. *Journal of Information Systems Management*, *8*(3), 27–35. https://doi.org/10.1080/07399019108964994

Frija, G., Blažić, I., Frush, D. P., Hierath, M., Kawooya, M., Donoso-Bach, L., & Brkljačić, B. (2021, August). How to improve access to medical imaging in low- and middle-income countries ? https://doi.org/10.1016/j.eclinm.2021.101034

Galton, F. (1907). Vox Populi. *Nature*, *75*, 509–510.

Gao, T. P., Su, H., & Yu, T. (2021). The Connotation and Logical Construction of Government Digital Transformation - Based on the Analysis of Sociotechnical System Theory. *E3S Web of Conferences*, *251*. https://doi.org/10.1051/e3sconf/202125103069

Garg, A., Gill, P., Rathi, P., & Garg, K. K. (2009). *An Insight into Swarm Intelligence* (tech. rep. No. 8).

Ginneken, J. v. (2009). *De kracht van de zwerm*. Business Contact.

Gupta, T. (2007). Multidisciplinary care in the community setting. *Journal of Oncology Practice*, *3*(2), 71–72. https://doi.org/10.1200/JOP.0722504

Gupta, T. G. (2007). Multidisciplinary Clinics in Oncology: The Hidden Pitfalls. *Journal of Oncology Practice*, *3*(2), 71–72. https://doi.org/10.1200/JOP.0722504

Habboub, G., Tish, S., Chakravarthy, V., Mroz, T., Ghogawala, Z., & Benzel, E. (2020). Interpretable Artificial Swarm Intelligence in Complex Neurosurgical Decision Making: Part 1. *World Neurosurgery*, *141*, 518–520. https://doi.org/10.1016/j.wneu.2020.07.005

Haj-Mirzaian, A., Kawamoto, S., Zaheer, A., Hruban, R. H., Fishman, E. K., & Chu, L. C. (2020, February). Pitfalls in the MDCT of pancreatic cancer: strategies for minimizing errors. https://doi.org/10.1007/s00261-019-02390-9

Halle-Smith, J. M., Bartlett, D., Chatzizacharias, N., Dasari, B. V., Marudanayagam, R., Sutcliffe, R. P., Brown, R. M., & Roberts, K. J. (2024). Investigating misdiagnosis of suspected cancer among patients undergoing pancreatoduodenectomy: three decades of experience. *HPB*, *26*(3), 344–351. https://doi.org/10.1016/j.hpb.2023.11.011

Hammer, R. D., Fowler, D., Sheets, L. R., Siadimas, A., Guo, C., & Prime, M. S. (2020). Digital Tumor Board Solutions Have Significant Impact on Case Preparation. *JCO Clin Cancer Inform*, *4*, 757–768. https://doi.org/10.1200/CCI.20

Harkins, S., & Petty, R. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology*, *43*(6), 1214–1229.

Harkins, S. G., & Jackson, J. M. (1985). The Role of Evaluation in Eliminating Social Loafing. *Personality and Social Psychology Bulletin*, *11*(4), 457–465. https://doi.org/10.1177/0146167285114011

Hartwig, W., Werner, J., Jäger, D., Debus, J., & Büchler, M. W. (2013). *Review Improvement of surgical results for pancreatic cancer* (tech. rep.). www.thelancet.com/oncology

Hassanien, A. E., & Emary, E. (2018). *Swarm Intelligence: principles, advances, and applications*.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. https://doi.org/10.1136/eb-2015-102129

Høigaard, R., Tofteland, I., & Ommundsen, Y. (2006). *The Effect of Team Cohesion on Social Loafing in Relay Teams* (tech. rep. No. 1).

Hong, S. B., Choi, S. H., Kim, K. w., Kim, S. Y., Kim, J. H., Kim, S., & Lee, N. K. (2020). Meta-analysis of MRI for the diagnosis of liver metastasis in patients with pancreatic adenocarcinoma. *Journal of Magnetic Resonance Imaging*, *51*(6), 1737–1744. https://doi.org/10.1002/jmri.26969

Hoogenboom, S. A., Engels, M. M., Chuprin, A. V., van Hooft, J. E., LeGout, J. D., Wallace, M. B., & Bolan, C. W. (2022). Prevalence, features, and explanations of missed and misinterpreted pancreatic cancer on imaging: a matched case–control study. *Abdominal Radiology*, *47*(12), 4160–4172. https://doi.org/10.1007/s00261-022-03671-6

Hricak, H., Abdel-Wahab, M., Atun, R., Lette, M. M., Paez, D., Brink, J. A., Donoso-Bach, L., Frija, G., Hierath, M., Holmberg, O., Khong, P. L., Lewis, J. S., McGinty, G., Oyen, W. J., Shulman, L. N., Ward, Z. J., & Scott, A. M. (2021, April). Medical imaging and nuclear medicine: a Lancet Oncology Commission. https://doi.org/10.1016/S1470-2045(20)30751-8

Hsieh, C. J., Fifić, M., & Yang, C. T. (2020). A new measure of group decision-making efficiency. *Cognitive Research: Principles and Implications*, *5*(1). https://doi.org/10.1186/s41235-020-00244-3

Huang, X., Lin, J., & Demner-Fushman, D. (2006). *Evaluation of PICO as a Knowledge Representation for Clinical Questions* (tech. rep.). University of Maryland, College Park. http://www.fpin.org/

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). *The Ringelmann Effect: Studies of Group Size and Group Performance1* (tech. rep.).

Ioannou, C. C. (2017, August). Swarm intelligence in fish? The difficulty in demonstrating distributed and self-organised collective intelligence in (some) animal groups. https://doi.org/10.1016/j.beproc.2016.10.005

Jackson, J. M., & Harkins, S. G. (1985). Equity in Effort: An Explanation of the Social Loafing Effect. *Journal of Personality and Social Psychology*, *49*(5), 1199–1206.

Janis, I. L. (L. (1982). *Groupthink : psychological studies of policy decisions and fiascoes*. Houghton Mifflin.

Javadi Khasraghi, H., & Hirschheim, R. (2022). Collaboration in crowdsourcing contests: how different levels of collaboration affect team performance. *Behaviour and Information Technology*, *41*(7), 1566–1582. https://doi.org/10.1080/0144929X.2021.1887354

Kagan, A. R. (2005, March). The multidisciplinary clinic. https://doi.org/10.1016/j.ijrobp.2004.10.040

Kamisawa, T., Wood, L. D., Itoi, T., & Takaori, K. (2016, July). Pancreatic cancer. https://doi.org/10.1016/S0140-6736(16)00141-0

Kang, J. D., Clarke, S. E., & Costa, A. F. (2021). Factors associated with missed and misinterpreted cases of pancreatic ductal adenocarcinoma. *European Radiology*, *31*. https://doi.org/10.1007/s00330-020-07307-5/Published

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Loannou, C. C., Giam, X., & Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society Interface*, *15*(141). https://doi.org/10.1098/rsif.2018.0130

Karau, S. J., & Wilhau, A. J. (2019, October). Social loafing and motivation gains in groups: An integrative review. In *Individual motivation within groups: Social loafing and motivation gains in work, academic, and sports teams* (pp. 3–51). Elsevier Inc. https://doi.org/10.1016/B978-0-12-849867-5.00001-X

Karau, S. J., & Williams, K. D. (1997). *The Effects of Group Cohesiveness on Social Loafing and Social Compensation* (tech. rep. No. 2).

Kaufman, J. (2011). Communication Overhead. In *The personal mba : A world-class business education in a single volume*. https://www.google.nl/books/edition/The_Personal_MBA/QMEIP4DtifcC?hl=en&gbpv=1&dq=8%20Symptoms%20of%20Bureaucratic%20Breakdown&pg=PT246&printsec=frontcover

Kee, F., Owen, T., & Leathem, R. (2004). Decision making in a multidisciplinary cancer team: Does team discussion result in better quality decisions? *Medical Decision Making*, *24*(6), 602–613. https://doi.org/10.1177/0272989X04271047

Kerr, N. (1981). Ringlemann revisited: Alternative explanations for the social loafing effect. *Personality and Sociology Bulletin*, *7*(2).

Ketterman, A. B., & Maner, J. K. (2021). Complaisant or coercive? The role of dominance and prestige in social influence. *Personality and Individual Differences*, *177*, 110814. https://doi.org/10.1016/j.paid.2021.110814

Khalaf, N., El-Serag, H. B., Abrams, H. R., & Thrift, A. P. (2021, May). Burden of Pancreatic Cancer: From Epidemiology to Practice. https://doi.org/10.1016/j.cgh.2020.02.054

Kim, M. S., Park, H. Y., Kho, B. G., Park, C. K., Oh, I. J., Kim, Y. C., Kim, S., Yun, J. S., Song, S. Y., Na, K. J., Jeong, J. U., Yoon, M. S., Ahn, S. J., Yoo, S. W., Kang, S. R., Kwon, S. Y., Bom, H. S., Jang, W. Y., Kim, I. Y., … Choi, Y. D. (2020). Artificial intelligence and lung cancer treatment decision: Agreement with recommendation of multidisciplinary tumor board. *Translational Lung Cancer Research*, *9*(3), 507–514. https://doi.org/10.21037/tlcr.2020.04.11

Kravitz, D. A., & Lexington, B. M. (1986). *Ringelmann Rediscovered: The Original Article* (tech. rep. No. 5).

Lamb, B. W., Sevdalis, N., Vincent, C., & Green, J. S. (2012, June). Development and evaluation of a checklist to support decision making in cancer multidisciplinary team meetings: MDT-QuIC. https://doi.org/10.1245/s10434-011-2187-0

Lamb, B. W., Taylor, C., Lamb, J. N., Strickland, S. L., Vincent, C., Green, J. S., & Sevdalis, N. (2013). Facilitators and barriers to teamworking and patient centeredness in multidisciplinary cancer teams: Findings of a national study. *Annals of Surgical Oncology*, *20*(5), 1408–1416. https://doi.org/10.1245/s10434-012-2676-9

Lamb, B. W., Jalil, R. T., Sevdalis, N., Vincent, C., & Green, J. S. (2014). Strategies to improve the efficiency and utility of multidisciplinary team meetings in urology cancer care: A survey study. *BMC Health Services Research*, *14*(1). https://doi.org/10.1186/1472-6963-14-377

Latane, B., Williams, K., & Harkins, S. (1979). *Many Hands Make Light the Work: The Causes and Consequences of Social Loafing* (tech. rep. No. 6).

LeBlanc, M., Kang, J., & Costa, A. F. (2023). Can we rely on contrast-enhanced CT to identify pancreatic ductal adenocarcinoma? A population-based study in sensitivity and factors associated with false negatives. *European Radiology*, *33*(11), 7656–7664. https://doi.org/10.1007/s00330-023-09758-y

Liden, R. C., Wayne, S. J., Jaworski, R. A., & Bennett, N. (2004). Social loafing: A field investigation. *Journal of Management*, *30*(2), 285–304. https://doi.org/10.1016/j.jm.2003.02.002

Lim, H. K., Stiven, P. N., & Aly, A. (2016). Reinterpretation of radiological findings in oesophago-gastric multidisciplinary meetings. *ANZ Journal of Surgery*, *86*(5), 377–380. https://doi.org/10.1111/ans.12537

Lim, S. L., & Bentley, P. J. (2022). Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-22856-z

Liu, Y., & Passino, K. M. (2000). *Swarm Intelligence: Literature Overview* (tech. rep.).

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

Luo, Z., Marnburg, E., Øgaard, T., & Okumus, F. (2021). Exploring antecedents of social loafing in students' group work: A mixed-methods approach. *Journal of Hospitality, Leisure, Sport and Tourism Education*, *28*. https://doi.org/10.1016/j.jhlste.2021.100314

Lyytinen, K., & Newman, M. (2008). Explaining information systems change: A punctuated socio-technical change model. *European Journal of Information Systems*, *17*(6), 589–613. https://doi.org/10.1057/ejis.2008.50

Madirolas, G., & de Polavieja, G. G. (2015). Improving Collective Estimations Using Resistance to Social Influence. *PLoS Computational Biology*, *11*(11). https://doi.org/10.1371/journal.pcbi.1004594

Maharaj, A. D., Evans, S. M., Zalcberg, J. R., Ioannou, L. J., Graco, M., Croagh, D., Pilgrim, C. H., Dodson, T., Goldstein, D., Philip, J., Kench, J. G., Merrett, N. D., Neale, R. E., White, K., Evans, P., & Leong, T. (2021). Barriers and enablers to the implementation of multidisciplinary team meetings: A qualitative study using the theoretical domains framework. *BMJ Quality and Safety*, *30*(10), 792–803. https://doi.org/10.1136/bmjqs-2020-011793

Maharaj, A. D., Ioannou, L., Croagh, D., Zalcberg, J., Neale, R. E., Goldstein, D., Merrett, N., Kench, J. G., White, K., Pilgrim, C. H., Chantrill, L., Cosman, P., Kneebone, A., Lipton, L., Nikfarjam, M., Philip, J., Sandroussi, C., Tagkalidis, P., Chye, R., … Evans, S. M. (2019). Monitoring quality of care for patients with pancreatic cancer: a modified Delphi consensus. *HPB*, *21*(4), 444–455. https://doi.org/10.1016/j.hpb.2018.08.016

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276–299. https://doi.org/10.1037/a0036677

Mannion, R., & Thompson, C. (2014). Systematic biases in group decision-making: Implications for patient safety. *International Journal for Quality in Health Care*, *26*(6), 606–612. https://doi.org/10.1093/intqhc/mzu083

Mavrodiev, P., Tessone, C. J., & Schweitzer, F. (2012). Effects of Social Influence on the Wisdom of Crowds. http://arxiv.org/abs/1204.3463

Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review*, *61*(4), 84–109. https://doi.org/10.1177/0008125619862256

Miller, G. (2024). *Streamlining Hospital Operations, Optimizing Resource Allocation, and Improving Efciency with AI Predictive Analytics and Machine Learning Algorithms* (tech. rep.). https://www.healthcareittoday.com/2024/06/18/streamlining-hospital-operations-optimizing-resource-allocation-and-improving-efficiency-with-ai-predictive-ana

Molin, E. (2024). Structural Equation Models SEN1721.

Muchnik, L., Aral, S., & Taylor, S. (2013). Social Influence Bias: A Randomised Experiment. *Science*, *341*(6146), 643–647. https://doi.org/10.1126/science.1235053

Nayyar, A., & Nguyen, N. G. (2018). *Introduction to Swarm Intelligence* (Vol. 1).

Panzeri, S., Magri, C., & Carraro, L. (2008). Sampling bias. *Scholarpedia*, *3*(9), 4258. https://doi.org/10.4249/scholarpedia.4258

Park, W., Chawla, A., & O'Reilly, E. M. (2021, September). Pancreatic Cancer: A Review. https://doi.org/10.1001/jama.2021.13027

Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., Rajpurkar, P., Amrhein, T., Gupta, R., Halabi, S., Langlotz, C., Lo, E., Mammarappallil, J., Mariano, A. J., Riley, G., Seekins, J., Shen, L., Zucker, E., & Lungren, M. (2019). Human–machine partnership with artificial intelligence for chest radiograph diagnosis. *npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0189-7

Patkar, V., Acosta, D., Davidson, T., Jones, A., Fox, J., & Keshtgar, M. (2012). Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance. *BMJ Open*, *2*(3). https://doi.org/10.1136/bmjopen-2011-000439

Petty, R., Harkins, S., & Williams, K. (1980). The effects of group diffusion of cognitive effort on attitudes: An information-processing view. *Journal of Personality and Social Psychology*, *38*(1).

Pillay, B., Wootten, A. C., Crowe, H., Corcoran, N., Tran, B., Bowden, P., Crowe, J., & Costello, A. J. (2016, January). The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: A systematic review of the literature. https://doi.org/10.1016/j.ctrv.2015.11.007

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. https://doi.org/10.1016/j.tics.2009.08.002

Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., & Matrisian, L. M. (2014, June). Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states. https://doi.org/10.1158/0008-5472.CAN-14-0155

Rosenberg, L., Pescetelli, N., & Willcox, G. (2017). Artificial Swarm Intelligence Amplifies Accuracy when Predicting Financial Markets. *IEEE 8th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*.

Rosenberg, L. (2016). *Artificial Swarm Intelligence, a Human-in-the-Loop Approach to A.I.* (tech. rep.). www.aaai.org

Rosenberg, L., Lungren, M., Halabi, S., Willcox, G., Baltaxe, D., & Lyons, M. (2018). Artificial Swarm Intelligence employed to Amplify Diagnostic Accuracy in Radiology. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, 1186–1191. https://doi.org/10.1109/IEMCON.2018.8614883

Rosenberg, L., & Willcox, G. (2019a). *Artificial Swarm Intelligence* (tech. rep.). https://www.researchgate.net/publication/334898388

Rosenberg, L., & Willcox, G. (2019b). *Artificial Swarm Intelligence Artificial Swarm Intelligence View project Virtual Reality and Augmented Reality View project* (tech. rep.). Unanimous AI. https://www.researchgate.net/publication/334544553

Rosenberg, L. B. (2015, July). Human Swarms, a real-time method for collective intelligence. https://doi.org/10.7551/978-0-262-33027-5-ch117

Roy, S., Biswas, S., & Sinha Chaudhuri, S. (2014). Nature-Inspired Swarm Intelligence and Its Applications. *International Journal of Modern Education and Computer Science*, *6*(12), 55–65. https://doi.org/10.5815/ijmecs.2014.12.08

Salmon, P. (2014). Nonlinear pattern formation in bone growth and architecture. *Frontiers in Endocrinology*, *5*(DEC). https://doi.org/10.3389/fendo.2014.00239

Savage, I. R. (1957). Nonparametric Statistics. *Journal of the American Statistical Association*, *52*(279), 331–344. https://doi.org/10.1080/01621459.1957.10501392

Sedgwick, P., & Greenwood, N. (2015, September). Understanding the hawthorne effect. https://doi.org/10.1136/bmj.h4672

Séroussi, B., Bouaud, J., & Antoine, E.-C. (2001). *ONCODOC: a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer* (tech. rep.). University of Paris.

Shah, A. (2019). Using data for improvement. *BMJ (Online)*, *364*. https://doi.org/10.1136/bmj.l189

Shah, R., Astuto Arouche Nunes, B., Gleason, T., Fletcher, W., Banaga, J., Sweetwood, K., Ye, A., Patel, R., McGill, K., Link, T., Crane, J., Pedoia, V., & Majumdar, S. (2023). Utilizing a Digital Swarm Intelligence Platform to Improve Consensus Among Radiologists and Exploring Its Applications. *Journal of Digital Imaging*, *36*(2), 401–413. https://doi.org/10.1007/s10278-022-00662-3

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019, March). Herding Brains: A Core Neural Mechanism for Social Alignment. https://doi.org/10.1016/j.tics.2019.01.002

Sigauke, Y. (2021). *The relationship between workplace boredom, procrastination, work engagement and turnover intention within the South African context* (tech. rep.).

Somashekhar, S. P., Sepúlveda, M. J., Puglielli, S., Norden, A. D., Shortliffe, E. H., Rohit Kumar, C., Rauthan, A., Arun Kumar, N., Patil, P., Rhee, K., & Ramya, Y. (2018). Watson for Oncology and breast cancer treatment recommendations: Agreement with an expert multidisciplinary tumor board. *Annals of Oncology*, *29*(2), 418–423. https://doi.org/10.1093/annonc/mdx781

Staats, B. R., Milkman, K. L., & Fox, C. R. (2012). The team scaling fallacy: Underestimating the declining efficiency of larger teams. *Organizational Behavior and Human Decision Processes*, *118*(2), 132–142. https://doi.org/10.1016/j.obhdp.2012.03.002

Strobel, O., Neoptolemos, J., Jäger, D., & Büchler, M. W. (2019, January). Optimizing the outcomes of pancreatic cancer surgery. https://doi.org/10.1038/s41571-018-0112-1

Swords, D. S., Mone, M. C., Zhang, C., Presson, A. P., Mulvihill, S. J., & Scaife, C. L. (2015). Initial Misdiagnosis of Proximal Pancreatic Adenocarcinoma Is Associated with Delay in Diagnosis and Advanced Stage at Presentation. *Journal of Gastrointestinal Surgery*, *19*(10), 1813–1821. https://doi.org/10.1007/s11605-015-2923-z

Taberna, M., Gil Moncayo, F., Jané-Salas, E., Antonio, M., Arribas, L., Vilajosana, E., Peralvez Torres, E., & Mesía, R. (2020, March). The Multidisciplinary Team (MDT) Approach and Quality of Care. https://doi.org/10.3389/fonc.2020.00085

Takadate, T., Morikawa, T., Ishida, M., Aoki, S., Hata, T., Iseki, M., Miura, T., Ariake, K., Maeda, S., Kawaguchi, K., Masuda, K., Ohtsuka, H., Mizuma, M., Hayashi, H., Nakagawa, K., Motoi, F., Kamei, T., Naitoh, T., & Unno, M. (2021). Staging laparoscopy is mandatory for the treatment of pancreatic cancer to avoid missing radiologically negative metastases. *Surgery Today*, *51*(5), 686–694. https://doi.org/10.1007/s00595-020-02121-4

Taylor, C., Brown, K., Lamb, B., Harris, J., Sevdalis, N., & Green, J. S. (2012, December). Developing and testing TEAM (Team Evaluation and Assessment Measure), a self-assessment tool to improve cancer multidisciplinary teamwork. https://doi.org/10.1245/s10434-012-2493-1

Taylor, E. (2020). We Agree, Don't We? The Delphi Method for Health Environments Research. *Health Environments Research and Design Journal*, *13*(1), 11–23. https://doi.org/10.1177/1937586719887709

Tornhill, A. (2019). Brooks's Law.

Veldhuisen, E., Brada, d. L., Walma, d. M., dr Molenaar, p. I., van Lienden, d. K., Wilmink, d. J., dr Besselink, p. M., & dr Busch, p. O. (2018). *Resection of locally advanced pancreatic cancer after induction FOLFIRINOX chemotherapy: new developments* (tech. rep.). www.dpcg.nl

Veron Sanchez, A., Bennouna, I., Coquelet, N., Cabo Bolado, J., Pinilla Fernandez, I., Mullor Delgado, L. A., Pezzullo, M., Liberale, G., Gomez Galdon, M., & Bali, M. A. (2023, July). Unravelling Peritoneal Carcinomatosis Using Cross-Sectional Imaging Modalities. https://doi.org/10.3390/diagnostics13132253

Wainberg, Z. A., Kang, Y. K., Lee, K. W., Qin, S., Yamaguchi, K., Kim, I. H., Saeed, A., Oh, S. C., Li, J., Turk, H. M., Teixeira, A., Hitre, E., Udrea, A. A., Cardellino, G. G., Sanchez, R. G., Zahlten-Kümeli, A., Taylor, K., & Enzinger, P. C. (2024). Bemarituzumab as first-line treatment for locally advanced or metastatic gastric/gastroesophageal junction adenocarcinoma: final analysis of the randomized phase 2 FIGHT trial. *Gastric Cancer*, *27*(3), 558–570. https://doi.org/10.1007/s10120-024-01466-w

Wang, L., Chen, X., Zhang, L., Li, L., Huang, Y., Sun, Y., & Yuan, X. (2023). Artificial intelligence in clinical decision support systems for oncology. https://doi.org/10.7150/ijms.77205

Weldon, E., & Gargano, G. (1985). *Cognitive Effort in Additive Task Groups: The Effects of Shared Responsibility on the Quality of Multiattribute Judgments* (tech. rep.).

Weldon, E., & Gargano, G. (1988). Cognitive loafing: The effects of accountability and shared responsibility on cognitive effort. *Journal of Personality and Social Psychology*, *14*(1), 159–171.

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *46*(4), 287–311. https://doi.org/10.1177/0095798420930932

Weninger, T., Johnston, T. J., & Glenski, M. (2015). Random voting effects in social-digital spaces: A case study of reddit post submissions. *HT 2015 - Proceedings of the 26th ACM Conference on Hypertext and Social Media*, 293–297. https://doi.org/10.1145/2700171.2791054

Willcox, G., Rosenberg, L., Askay, D., Metcalf, L., Harris, E., & Domnauer, C. (2020). Artificial Swarming Shown to Amplify Accuracy of Group Decisions in Subjective Judgment Tasks. https://doi.org/10.1007/978-3-030-12385-7{_}29

Williams, K., Harkins, S., & Latane, B. (1981). *Identifiability as a Deterrent to Social Loafing: Two Cheering Experiments* (tech. rep. No. 2).

Xu, X., Wang, L., & Wang, L. (2020). Social facilitation or social loafing – threshold effect of group size on customer's complaint intention. *Journal of Contemporary Marketing Science*, *3*(2), 243–263. https://doi.org/10.1108/jcmars-12-2019-0049

Yadav, T., Kurdia, K., & Irrinki, S. (2021, January). Intraperitoneal paclitaxel for pancreatic cancer with peritoneal metastasis. https://doi.org/10.1093/bjs/znaa032

Zaccaro, S. J. (1984). Social Loafing: The role of Task Attractiveness. *Personality and Social Psychology Bulletin*, *10*(1), 99–106.

Zeigler-Hill, V., Welling, L. L., & Shackelford, T. K. (2015). *Evolutionary Psychology Series Editors* (tech. rep.). http://www.springer.com/series/10583

A Synthesis table MDT in Surgical Oncology Decision-making

A.1. Synthesis table overview

	Author	MDT Benefits	MDT limitations
MDT quality evaluation A2	Kagan (2004)	• Opinion-based discussion from more than one type of oncologist • Physicians from multidisciplinary clinics learn from each other • In-depth discussion of the patient's clinical factors	• Censoring of different opinions (if ran by a dominating physician), leading to decreased decision quality • Chair skills highly impact the tone of the clinic • Arguments may be superfluous, wasting time, harmful, verbal abuse, or tangential to the target • Physicians may sometimes focus on self-conceit instead of the patient's needs
	Maharaj et al. (2021)	• Higher accuracy and complete preoperative staging and neoadjuvant or palliative treatment • Engagement of sub-specialists relevant to their specific circumstances • Optimize patient selection for specific treatments • An excellent opportunity to learn and interact across disciplines	• The impact of MDT meetings on survival outcomes in cancer has not been definitely proven • Lacking capacity to adequately cater to the volume of patients to be discussed and scheduling difficulties • Lack of palliative care input in MDT meetings (majority of PC patients are diagnosed at an advanced stage) • Difficulty for some specialists (palliative) to participate in discussions due to a lack of confidence or respect • MDT meetings are dependent on the skills of the chair in ensuring all participants are included in the discussions
	Gupta (2007)	• MDT improves decision-making by closing the knowledge gaps among the members • MDC can be a vital tool to enhance members' skills in a short period • MDC can lead to integrated teaching • MDT can lead to the creation of databases to review outcome data	• MDT decisions are regularly based on lack of response, autocratic choice or deference to expert opinion • Regularly, egoist individual dominates proceedings and dictates terms in an MDC • Collective decision-making may falsely reduce the sense of individual responsibility for decisions and patient care • Collective decision-making may make the members believe their decisions are more ethical because they are based on "consensus" opinion
	Fehervari et al. (2021)	• MDT coordinator has been added to the list of core members in order to fulfil all administrative tasks in a meeting, ensuring all relevant imaging and results are available for the meeting • The benefit of MDT discussion is most marked in patients with advanced disease and clinicians are not always aware of what their colleagues in other specialties might be able to offer • MDTs are a valuable training opportunity for junior doctors	• Most common issue internationally is lack of staffing and resources • There are no existing guidelines on who specifically should lead and attend CR MDTs • The list of core members varies greatly per country
	Lamb et al. (2012)	• Can reduce communication error • Improve safety in surgery • Evidence-based	• May be difficult to enforce • May be perceived to undermine autonomy, skill or knowledge • In organisations, the enormous number of checklist may cause 'checklist fatigue'
Existing solutions A3	Taylor (2021)	• anonymity(reducing dominant personalities by using questionnaires/surveys) • controlled feedback (summary results reported to respondents in each round) • statistical group response (an aggregate of individual opinions to reduce pressure for conformity).	• The method is significantly time-consuming • Difficult to keep the population of respondents consistent throughout the process
	Burmeister et al. (2016)	-	• Highly variability in the way MDT are incorporated into patient care • Some specialisations (most notably gastroenterology) commonly not represented in MDT • Divergent views on whether *all* patients vs. Only resectable patients should be presented in MDT • MDT meetings mostly led by surgeons who are found to undervalue presence of other health specialisations (notably palliative care workers)
	Patkar et al. (2012)	• MATE: Structurized patient data capturing, presentation and audit for use in preparation to or during MDT meetings • Advanced evidence-based decision-support • Provides automatic recommendation on treatment plans based on clinical guidelines • High proven rate of concordance (93.2%) between MDT decision and MATE recommendation	• Incompatibility with existing data capture systems, resulting in double data entry • Costs and resources • Clincical buy-in • Scalability • Voor in het verslag: Data has to be prepared in advance to the MDT (time constraint) • Designed as complementary in the MDT decision (voor in het verslag: cannot compensate for inherent weaknesses in team composition, Organization or operation)
	Hammer et al. (2020)	• NAVIFY Timer Board Solution (NTB): Patient data aggregation for preparing, presenting and documentation during Tumer Boards (TB) • Significantly decreases preparation time on timer board preparations in different cancer categories • Decreases variation in preparation times • Standardization of the preparation process	• Focuses on the preparation of the MDT meeting • Does not structure or speed up the decision-process itself • Does not improve the inclusiveness of the MDT meeting
	Wang et al. (2023)	• CDSS can guide inexperienced doctors and interns • WFO provides evidence to support recommended treatments • WFO can provide rapid and accurate suggestions for most cancer patients	• The effectiveness of CDSS is highly dependent on the data set fed to the system • The effectiveness of WFO limited to a selection of cancers, restricting its usage in co-morbidities • WFO fails to adapt to genetic mutations originating from devious demographic properties • Test reporting of CDSS performance currently is not representative, as most studies do not mention unsupported cases •
	Somashekhar et al. (2018)	• Clinical decision-support systems help oncologists in personalizing care through rapidly changing scientific evidence, drug approvals, and treatment guidelines • Watson for Oncology (WFO), an AI for CDSS, provides 90+ concordance in treatment between the CDF and multidisciplinary tumor board • WFO provides evidence for its decision, enabling oncologists to examinee the data which the WFO's recommendation is based on	• Effectiveness of CDSS is highly dependent on the data accuracy fed to the system • CDSS may lead to over-reliance on technology, subsequently resulting in more errors • Although CDSSs are intended to be support-based, they function as a tool that could 'take over' the final therapy decision
	Kim et al. (2020)	• CDSS provides +90% concordance with oncology expertise at metastatic stage • CDSS continuously updates its knowledge base • CDSS models intelligent processes in order to lower human intervention in treatment decision-making	• WFO is still less suitable when patient is in non-metastatic stage, as the decision complexity at this stage is still high due to many patient-related factors (co-mobordity, insurance, socioeconomic state) • CDSS commonly trained through standardised data, WFO could be unable to properly adjust its recommendations based on local demographic properties • Cancer patients worry about receiving treatment recommendations from WFO
ASI in surgical decision-making A4	Rosenberg et al. (2018)	• Testing the implementation of ASI in the diagnosis of Pneumonia using 50 test cases • Achieving a 82% diagnostic accuracy through ASI • Significantly improved performance compared to individual accuracy (73%) and Machine-Learning assisted diagnosis (60%)	• Focuses on the diagnosis (Pneumonia) rather than decision-making • Accuracy compared to a 'ground truth' dataset that is acknowledged to be error-prone
	Shah et al. (2023)	• Testing the use of ASI in diagnosis of meniscal lesions • Achieving an increased performance of between 23% and 30% across different cohorts • ASI would provide a useful tool for collaborating and evaluating complex or ambiguous cases	• Focuses on the diagnosis (meniscal lesions) rather than decision-making • Test performance may have been influenced by time and COVID-constraints
	Habboub et al. (2020)	• Conceptual framework describing two-component model: Swarm periphery and Swarm Center • Improved model trust through Machine-Learning interpretability	• Does not describe operationalisation • Lacking verifiability and testability

A.2. MDT quality evaluation

	Author	MDT Benefits
MDT quality evaluation	Kagan (2004)	• Opinion-based discussion from more than one type of oncologist • Physicians from multidisciplinary clinics learn from each other • In-depth discussion of the patient's clinical factors
	Maharaj et al. (2021)	• Higher accuracy and complete preoperative staging and neoadjuvant or palliative treatment • Engagement of sub-specialists relevant to their specific circumstances • Optimize patient selection for specific treatments • An excellent opportunity to learn and interact across disciplines
	Gupta (2007)	• MDT improves decision-making by closing the knowledge gaps among the members • MDC can be a vital tool to enhance members' skills in a short period • MDC can lead to integrated teaching • MDT can lead to the creation of databases to review outcome data
	Fehervari et al. (2021)	• MDT coordinator has been added to the list of core members in order to fulfil all administrative tasks in a meeting, ensuring all relevant imaging and results are available for the meeting • The benefit of MDT discussion is most marked in patients with advanced disease and clinicians are not always aware of what their colleagues in other specialties might be able to offer • MDTs are a valuable training opportunity for junior doctors
	Lamb et al. (2012)	• Can reduce communication error • Improve safety in surgery • Evidence-based

	Author	MDT limitations
MDT quality evaluation	Kagan (2004)	• Censoring of different opinions (if ran by a dominating physician), leading to decreased decision quality • Chair skills highly impact the tone of the clinic • Arguments may be superfluous, wasting time, harmful, verbal abuse, or tangential to the target • Physicians may sometimes focus on self-conceit instead of the patient's needs
	Maharaj et al. (2021)	• The impact of MDT meetings on survival outcomes in cancer has not been definitely proven • Lacking capacity to adequately cater to the volume of patients to be discussed and scheduling difficulties • Lack of palliative care input in MDT meetings (majority of PC patients are diagnosed at an advanced stage) • Difficulty for some specialists (palliative) to participate in discussions due to a lack of confidence or respect • MDT meetings are dependent on the skills of the chair in ensuring all participants are included in the discussions
	Gupta (2007)	• MDT decisions are regularly based on lack of response, autocratic choice or deference to expert opinion • Regularly, egoist individual dominates proceedings and dictates terms in an MDC • Collective decision-making may falsely reduce the sense of individual responsibility for decisions and patient care • Collective decision-making may make the members believe their decisions are more ethical because they are based on "consensus" opinion
	Fehervari et al. (2021)	• Most common issue internationally is lack of staffing and resources • There are no existing guidelines on who specifically should lead and attend CR MDTs • The list of core members varies greatly per country
	Lamb et al. (2012)	• May be difficult to enforce • May be perceived to undermine autonomy, skill or knowledge • In organisations, the enormous number of checklist may cause 'checklist fatigue'

A.3. Existing solutions

	Author	MDT Benefits
Existing solutions	Taylor (2021)	• anonymity(reducing dominant personalities by using questionnaires/surveys) • controlled feedback (summary results reported to respondents in each round) • statistical group response (an aggregate of individual opinions to reduce pressure for conformity).
	Burmeister et al. (2016)	-
	Patkar et al. (2012)	• MATE: Structurized patient data capturing, presentation and audit for use in preparation to or during MDT meetings • Advanced evidence-based decision-support • Provides automatic recommendation on treatment plans based on clinical guidelines • High proven rate of concordance (93.2%) between MDT decision and MATE recommendation
	Hammer et al. (2020)	• NAVIFY Timer Board Solution (NTB): Patient data aggregation for preparing, presenting and documentation during Turner Boards (TB) • Significantly decreases preparation time on timer board preparations in different cancer categories • Decreases variation in preparation times • Standardization of the preparation process
	Wang et al. (2023)	• CDSS can guide inexperienced doctors and interns • WFO provides evidence to support recommended treatments • WFO can provide rapid and accurate suggestions for most cancer patients
	Somashekhar et al. (2018)	• Clinical decision-support systems help oncologists in personalizing care through rapidly changing scientific evidence, drug approvals, and treatment guidelines • Watson for Oncology (WFO), an AI for CDSS, provides 90+ concordance in treatment between the CDF and multidisciplinary tumor board • WFO provides evidence for its decision, enabling oncologists to examinee the data which the WFO's recommendation is based on
	Kim et al. (2020)	• CDSS provides +90% concordance with oncology expertise at metastatic stage • CDSS continuously updates its knowledge base • CDSS models intelligent processes in order to lower human intervention in treatment decision-making

	Author	MDT limitations
Existing solutions	Taylor (2021)	• The method is significantly time-consuming • Difficult to keep the population of respondents consistent throughout the process
	Burmeister et al. (2016)	• Highly variability in the way MDT are incorporated into patient care • Some specialisations (most notably gastroenterology) commonly not represented in MDT • Divergent views on whether *all* patients vs. Only resectable patients should be presented in MDT • MDT meetings mostly led by surgeons who are found to undervalue presence of other health specialisations (notably palliative care workers)
	Patkar et al. (2012)	• Incompatibility with existing data capture systems, resulting in double data entry • Costs and resources • Clincical buy-in • Scalability • Voor in het verslag: Data has to be prepared in advance to the MDT (time constraint) • Designed as complementary in the MDT decision (voor in het verslag: cannot compensate for inherent weaknesses in team composition, Organization or operation)
	Hammer et al. (2020)	• Focuses on the preparation of the MDT meeting • Does not structure or speed up the decision-process itself • Does not improve the inclusiveness of the MDT meeting
	Wang et al. (2023)	• The effectiveness of CDSS is highly dependent on the data set fed to the system • The effectiveness of WFO limited to a selection of cancers, restricting its usage in co-morbidities • WFO fails to adapt to genetic mutations originating from devious demographic properties • Test reporting of CDSS performance currently is not representative, as most studies do not mention unsupported cases •
	Somashekhar et al. (2018)	• Effectiveness of CDSS is highly dependent on the data accuracy fed to the system • CDSS may lead to over-reliance on technology, subsequently resulting in more errors • Although CDSSs are intended to be support-based, they function as a tool that could 'take over' the final therapy decision
	Kim et al. (2020)	• WFO is still less suitable when patient is in non-metastatic stage, as the decision complexity at this stage is still high due to many patient-related factors (co-mobordity, insurance, socioeconomic state) • CDSS commonly trained through standardised data, WFO could be unable to properly adjust its recommendations based on local demographic properties • Cancer patients worry about receiving treatment recommendations from WFO

A.4. ASI in surgical decision-making

	Author	MDT Benefits
ASI in surgical decision-making	Rosenberg et al. (2018)	• Testing the implementation of ASI in the diagnosis of Pneumonia using 50 test cases • Achieving a 82% diagnostic accuracy through ASI • Significantly improved performance compared to individual accuracy (73%) and Machine-Learning assisted diagnosis (60%)
	Shah et al. (2023)	• Testing the use of ASI in diagnosis of meniscal lesions • Achieving an increased performance of between 23% and 30% across different cohorts • ASI would provide a useful tool for collaborating and evaluating complex or ambiguous cases
	Habboub et al. (2020)	• Conceptual framework describing two-component model: Swarm periphery and Swarm Center • Improved model trust through Machine-Learning interpretability

	Author	MDT limitations
ASI in surgical decision-making	Rosenberg et al. (2018)	• Focuses on the diagnosis (Pneumonia) rather than decision-making • Accuracy compared to a 'ground truth' dataset that is acknowledged to be error-prone
	Shah et al. (2023)	• Focuses on the diagnosis (meniscal lesions) rather than decision-making • Test performance may have been influenced by time and COVID-constraints
	Habboub et al. (2020)	• Does not describe operationalisation • Lacking verifiability and testability

B Synthesis table of social loafing

	Author	Method	Independent variables	Dependent variables	Outcome
Cognitive loafing experiments	Petty & Harkins. (1980)	1. Evaluating the performance of a video-taped counseling psychologist 2. Reviewing editorial essays of other students	1. Group size 1. Good vs. Bad therapist manipulation 2. Group size 2. Quality of editorial manipulation	1. Evaluation measures 1. Cognitive effort measures 2. Evaluation measures 2. Cognitive effort measures	1. Individuals showed higher cognitive efforts 1. Individuals provided higher ratings to good therapist manipulation 1. Individuals provided more positive arguments on likable stimulus and negative arguments on dislikable stimulus 2. Individuals showed higher cognitive involvement 2. Individuals provided more thorough arguments
	Harkins & Petty (1982)	1. Brainstorming and generating uses for an object 2. Watching a TV screen and reporting seldomly occurring signals	1. Individual vs. Group (64) 1. Difficult object (detached doorknob) vs. easy object (knife) 2. Individual vs. Group (122) 2. High contrast (easy) vs. Low contrast (difficult) level 2. Identifiable response vs. Anonymous response	1. Cognitive effort measures 1. Number of uses 2. Number of correct detections 2. Identifiable vs. Anonymous 2. Cognitive effort	1. For easy object, individuals produce more cognitive effort and more uses 1. For difficult object, individuals work as hard as group members, even though their outputs are **not** identifiable 2. For easy level: Identifiable, individual responses outperformed group, anonymous responses 2. For difficult level: Anonymous response does not produce cognitive loafing
	Weldon & Gargano (1985)	1. Evaluating job descriptions 2. Check understanding and assess attitude toward experiment 1	1. Individual vs. group (16) 2. Individual vs. Group (16)	1. Number of evaluations 1. Indication of information use 2. (Non)linear information use 2. (Non)linear utility functions	1. Individuals produced more evaluations and stronger linear main effects 2. Individual judges showed less nonlinear information use 2. Individual evaluators showed more significant main effects
	Jackson & Harkins (1985)	Participation in shouting experiment based on perceived effort of others (partner)	Individual vs. Pair (2) Expected low- vs. High-effort condition of partner	Cognitive effort measures Decibel scores	Individuals provide less effort when knowing their partner report to put less effort too
	Weldon & Gargano (1988)	Evaluating job descriptions based on perceived accountability	Individual vs. Group (15-20) Accountable vs. Unaccountable	Level of information processing Cognitive effort measures	Individuals with shared responsibility that were unaccountable showed less effortful judgement strategies compared to working alone Accountability levelled differences between individual and group efforts

	Reduction factor	Literature evidence			Citations
Loafing reduction factors	Individual identifiability	Williams et al. (1981)	Harkins & Petty (1982)		- "Present data suggests that when the subjects' performances cannot be evaluated, they loaf… (Harkins & Petty, 1982)
	Evaluation of individual performance	Harkins (1985)	Xu et al. (2020)	(Karau & Wilhau, 2019)	- "Individual output must be evaluatable against outputs of others" (Harkins, 1985) - "The opportunity for comparison may lead the participants to believe that their performances can be evaluated, which are seen to motivate performance" (Harkins, 1985) - "Numerous studies…have shown that social loafing can be reduced or eliminated in many situations when individuals' contributions to the collective task can be identified … for evaluation" (Karau & Wilhau, 2019)
	Equal perceived group efforts	Jackson & Harkins (1985)	Høigaard et al. (2006)	Karau & Williams (1997)	- "When the partner suggested that she would always work hard, the participants worked hard, whether working alone or with their partner, and when the partner suggested that she would hardly work, the participants followed suit and reduced their alone performances to their group levels." (Jackson & Harkins, 1985) - "Apparently, one function of team cohesiveness may be that a sense of social attraction and integration to the team counteracts social loafing, and thereby enhances individual effort and team performance. Put differently, high group cohesion may increase team members' group orientation and enhance the motivation to work for the team. " (Høigaard et al., 2006)
	Task attractiveness	Zaccaro (1984)			- "While performance in the low task attractiveness conditions declined across increases in group size, performance in high task attractiveness conditions increased as group size increased" (Zaccaro, 1984) - "The enhancement of performance in high task attractive groups suggests that these pressures reverse the diffusion of effort that produces loafing" (Zaccaro, 1984)
	Time restricted tasks	Luo et al. (2021)	Cui et al. (2021)	Sigauke (2021)	- "The results lend support to Hypothesis 1: Subjective team workload increases with time pressure" (Cui, 2021) - "Procrastination was associated with workplace boredom (Metin et al., 2016) because it is a condition characterised by under-stimulation (Reijseger et al., 2013). Those employees with lessened physical-cognitive energy are not committed to their work." (Sigauke, 2021)

C Synthesis tables of social bias

	Effect	Literature evidence				Citations
Social bias effects	Group-think	Janis (1982)	Mannion & Thompson (2014)	Lamb et al. (2013)		- "Highly cohesive groups with strongly connected members may inhibit the expression of (true) opinion; in such cases, group harmony and unanimity may be privileged over effective decision-making (Mannion & Thompson, 2014) - "Clearly, groupthink militates against the delivery of high quality and safe care where health care professionals should feel comfortable in expressing dissenting views and are willing to speak openly about concerns they may have about the quality of care." (Mannion & Thompson, 2014)
	Social dominance	Lim & Bently (2022)	Cheng (2020)	Knetterman & Maner (2021)		"Research in evolutionary psychology has identified two general strategies – dominance and prestige - used to attain influence and high social rank within groups. Whereas dominance is defined by the use of force to gain social rank, prestige is defined by the display of valued skills and abilities." (Knetterman & Maner, 2021)
	Herding	Shamay-Tsoory et al. (2019)	Zeigler-Hill et al. (2015)	Chen et al. (2022)	Rabat et al. (2009)	- "Social conformity is the act of fitting in with the group, manifested in the tendency of people to modify their opinions or judgments according to those of the group" (Shamay-Tsoory et al, 2019) - "Most of our daily decisions are made in a social context, in which people often have only limited information when making decisions. Therefore, integrating social information might help to exploit the informational content and lead to make the optimal decision when private information is insufficient" (Chen et al., 2022)

	Minimisation factor	Literature evidence	Citations
Social bias minimisation factors	Anonymity	Mannion & Thompson (2014)	- "'Direct pressure on dissenters': opposing opinions or questioning a decision's rationale results in the group applying peer pressure to coerce the dissenter into conformity." (Mannion & Thompson, 2014) - "'Stereotyping': 'outsiders' are viewed as different or inferior. The group then uses this perceived inferiority to discredit those who oppose their views. An example in health care is when a senior clinician dismisses or fails to take seriously the concerns of other (non-clinical) staff or patients about poor quality care." (Mannion & Thompson, 2014)
	Deliberation engagement	Mannion & Thompson (2014)	"'Illusion of unanimity': silence is interpreted as consent; consequently, the group believes that a 'unanimous' decision has been made." (Mannion & Thompson, 2014)
	Equal decision influence	Cheng (2020)	"Across societies, humans seek out and defer — out of personal choice — to people who possess skills, attributes, and locally valued knowledge attributes that inspire respect. (…) with highly influential leaders exercising differential weight and influence over key decisions and opinions in the community" (Cheng, 2020)

D Synthesis table for adverse case assessment

D.1. Overview of synthesis table

D2

	Effect	Literature evidence				Citations
Organizational	Case selection	Hoogenboom et al. (2022)	Strobel et al. (2019)	Lamb et al. (2014)		- “Adequate patient selection is a key determining factor of the outcome of pancreatic cancer surgery (Fig. 1) and has to be based on assessments of the patient’s tumour stage and the risks associated with surgery.” (Strobel et al., 2019) - “Proposed solutions include treating simple cases by protocol and approving such treatment plans outside the MDT meeting; prioritisation of cases; and splitting larger, high- volume MDTs into smaller, more specialist and hence” (Lamb et al., 2014)
	Resource availability	Frija et al. (2021)	Hricak et al. (2021)	Fehervari et al. (2021)	Lamb et al. (2014)	- “Imaging is essential to ensure timely diagnosis and appro- priate treatment of diseases and it would be unethical not to transfer the benefits that state-of-the-art imaging provides in developed countries, to LMICs” (Frija et al., 2021) - “Our global assessment of imaging and nuclear medicine resources identified substantial shortages in equipment and workforce, particularly in low-income and middle-income countries” (Hricak et al., 2021) - The most common issues with MDTs described are a lack of staffing and resources. MDT meetings may not be part of a clinician’s job plan and often clash with other activities resulting in non-attendance. (Fehervari et al., 2021)
	Planning restrictions	Hartwig et al. (2013)	Taberna et al. (2020)	Halle-Smith et al. (2024)		- “Even at experienced, high-volume centres, these data show that patients with cystic lesions need a thorough interdisciplinary evaluation and careful selection for observation or surgery to supply adequate therapy. Importantly, unnecessarily long follow-up might enable progression of branch-duct IPMNs into ductal adenocarcinomas.” - “As an essential member of the MDT, the role of the specialized clinical nurse in this disease is to support patients during the whole diagnostic and treatment process, which will include not only performing nursing interventions (i.e., symptom, toxicity and/or wound management) but also operational case management such as treatment planning and coordination.” (Taberna et al., 2020)

D3

		Effect	Literature evidence				Citations
MDT	**Technological**	Suboptimal imaging technique	Haj-Mirzaian et al. (2020)	Hong et al. (2020)	Hartwig et al. (2013)	Kang et al. (2021)	- “Small tumour size is one of the most relevant positive predictive factors in pancreatic cancer, and high-quality imaging techniques play a crucial part in the diagnosis of pancreatic tumours” (Hartwig et al., 2013) - “MRI can improve the detection and char- acterization of small focal hepatic lesions in comparison with CT, because MRI can provide more information in the form of various image sequences including T2-weighted image, DWI, and HBP images of hepatocyte-specific contrast- enhanced MRI.” (Hong et al., 2020)
		Suboptimal imaging quality	Haj-Mirzaian et al. (2020)	Budigi et al., (2022)	Ahmed et al., (2024)	LeBlanc et al. (2023)	- “The main reason for not resecting PDAC lesions initially believed to be resectable based on preoperative CT was unsuspected liver metastases identified during surgery (26). Liver metasta- ses missed on CT are associated with lower contrast-to-noise ratio, subcapsular location, hepatic steatosis that masks lesions, and examination performed for indications other malignancy (27)” (Budigi et al., 2022) - “These images should be interpreted with caution because of the possibility of motion and other artifacts [39]. MRI is more susceptible to motion artifacts in comparison to MDCT; these motion artifacts can obscure small tumors and lead to misdiagnosis. Motion can result in image blurriness and artifact“ (Haj-Mirzaian et al., 2020)
		Incomplete anatomical coverage	Haj-Mirzaian et al. (2020)				- “Furthermore, the pancreatic head and uncinate process may be incompletely included in axial MRI due to variability in patients’ breath-hold
	Process	Operator dependence	Armato (2009)	Ahmed et al. (2024)	Kang et al. (2021)	LeBlanc et al. (2023)	- “Substantial variability exists across radiologists in the task of lung nodule identification in CT scans” (Armato, 2009) - “radiologists are often the first physicians to diagnose or raise the possibility” (Ahmed et al., 2024) - “”Understanding tumor spread patterns, identi- fication of errors and their contributing factors as well as knowledge of approaches to reduce diagnostic errors enable radiologists to accurately stage/restage PDAC” (Budigi et al., 2022)
		Cognitive error	Haj-Mirzaian et al. (2020)	Hoogenboom et al. (2022)			- "Regarding the possible radiological errors that may have hampered accu- rate detection at the time of imaging, underreading was the most commonly identified error in our study (47% of errors in MRI, 57% in CT)” (Hoogenboom et al., 2022)
		Cognitive bias	LeBlanc et al. (2023)	Kang et al. (2021)	Busby et al. (2018)		- This highlights the importance of referring physicians to provide a thorough history on the imaging requisition, and may also reflect the cognitive biases that radiologists are susceptible to. An example of a cognitive bias is framing bias, which is the tendency to be influenced by how a question is asked or how a problem is presented. (LeBlanc et al., 2023)

D4

	Effect	Literature evidence				Citations
Patient	Intrinsic tumor features	Haj-Mirzaian et al. (2020)	Strobel et al. (2019)	Budigi et al., (2022)	Kang et al. (2021)	- “A small but clinically relevant proportion of patients are initially understaged and undergo laparotomy for resection fol- lowing CT, followed by abandonment of resection owing either to encroachment of the tumour on local structures or, more commonly, because micrometastases are iden- tified intraoperatively in the peritoneum and/or liver.” (Strobel et al., 2019) - “Most subtle tumors missed in our experience were small (<2 cm) hypoenhancing lesions in the uncinate process, isoattenuating tumors, and tumors on unenhanced imaging that were subtle but visible in retrospect” (Budigi et al., 2022)
	Tumor progression	Halle-Smith et al. (2024)	Wainberg et al. (2024)	Swords et al. (2015)	Hartwig et al. (2013)	- “Due to the rapid disease progression associated with these cancers, in particular pancreatic ductal adenocarcinoma (PDAC) and cholangiocarcinoma, it is highly desirable to establish a rapid preoperative diagnosis and initiate treatment quickly.“ (Halle-Smith et al., 2024) - They concluded that once PDAC becomes clinically detectable, progression from low-stage to advanced-stage disease is rapid.” (Swords et al., 2015) - “Importantly, unnecessarily long follow-up might enable progression of branch-duct IPMNs into ductal adenocarcinomas” (“Hartwig et al., 2013)
	Patient-related factors	Strobel et al. (2019)	Park et al. (2021)	Haj-Mirzaian et al. (2020)		- “Well-defined risk factors for perioperative morbidity and mortality in patients undergoing pancreatic surgery can be classi- fied either as patient-related factors, such as age, obesity and comorbidities, or tumour-related factors, including the extent of local involvement of neighbouring organs and blood vessels” (Strobel et al., 2019) - “Focal fatty infiltration is a relatively common finding in the adult population, especially in old or obese subjects” (Haj-Mirzaian et al., 2020)
	coexisting or underlying pathologies	LeBlanc et al. (2023)	Kang et al. (2021)	Halle-Smith et al. (2024)	Ahmed et al. (2024)	- “Pancreatitis presents a challenge with respect to the imaging diagnosis and staging of PDAC for two reasons: first, PDAC can be misinterpreted as acute pancreatitis, and pancreatitis, particularly mass-forming pancreatitis, can be mistaken for malignancy. Second, inflammatory changes can mimic locally invasive disease, limiting assessment of resectability” (Kang et al., 2021) - “Given that a multitude of benign pathology, such as autoimmune pancreatitis or complex cystic lesions, can mimic malignant conditions, there is a challenge to providing surgical treatment within the limitations of the current diagnostic landscape” (Halle-Smith et al., 2024) - “There exists a significant overlap in the imaging features, clinical presentation, and risk factors of pancreatitis and PDAC” (Ahmed et al., (2024) - Imaging findings can also be misinterpreted for other pathology; for example, acute pancreatitis can mimic or mask PDAC” (LeBlanc et al., 2023)

D.2. Synthesis table for adverse case assessment: Organizational layer

	Effect	Literature evidence				Citations
Organizational	Case selection	Hoogenboom et al. (2022)	Strobel et al. (2019)	Lamb et al. (2014)		- "Adequate patient selection is a key determining factor of the outcome of pancreatic cancer surgery (Fig. 1) and has to be based on assessments of the patient's tumour stage and the risks associated with surgery." (Strobel et al., 2019) - "Proposed solutions include treating simple cases by protocol and approving such treatment plans outside the MDT meeting; prioritisation of cases; and splitting larger, high- volume MDTs into smaller, more specialist and hence" (Lamb et al., 2014)
	Resource availability	Frija et al. (2021)	Hricak et al. (2021)	Fehervari et al. (2021)	Lamb et al. (2014)	- "Imaging is essential to ensure timely diagnosis and appro- priate treatment of diseases and it would be unethical not to transfer the benefits that state-of-the-art imaging provides in developed countries, to LMICs" (Frija et al., 2021) - "Our global assessment of imaging and nuclear medicine resources identified substantial shortages in equipment and workforce, particularly in low-income and middle-income countries" (Hricak et al., 2021) - The most common issues with MDTs described are a lack of staffing and resources. MDT meetings may not be part of a clinician's job plan and often clash with other activities resulting in non-attendance. (Fehervari et al., 2021)
	Planning restrictions	Hartwig et al. (2013)	Taberna et al. (2020)	Halle-Smith et al. (2024)		- "Even at experienced, high-volume centres, these data show that patients with cystic lesions need a thorough interdisciplinary evaluation and careful selection for observation or surgery to supply adequate therapy. Importantly, unnecessarily long follow-up might enable progression of branch-duct IPMNs into ductal adenocarcinomas." - "As an essential member of the MDT, the role of the specialized clinical nurse in this disease is to support patients during the whole diagnostic and treatment process, which will include not only performing nursing interventions (i.e., symptom, toxicity and/or wound management) but also operational case management such as treatment planning and coordination." (Taberna et al., 2020)

D.3. Synthesis table for adverse case assessment: MDT layer

		Effect	Literature evidence				Citations
MDT	**Technological**	Suboptimal imaging technique	Haj-Mirzaian et al. (2020)	Hong et al. (2020)	Hartwig et al. (2013)	Kang et al. (2021)	- "Small tumour size is one of the most relevant positive predictive factors in pancreatic cancer, and high-quality imaging techniques play a crucial part in the diagnosis of pancreatic tumours" (Hartwig et al., 2013) - "MRI can improve the detection and characterization of small focal hepatic lesions in comparison with CT, because MRI can provide more information in the form of various image sequences including T2-weighted image, DWI, and HBP images of hepatocyte-specific contrast- enhanced MRI." (Hong et al., 2020)
		Suboptimal imaging quality	Haj-Mirzaian et al. (2020)	Budigi et al., (2022)	Ahmed et al., (2024)	LeBlanc et al. (2023)	- "The main reason for not resecting PDAC lesions initially believed to be resectable based on preoperative CT was unsuspected liver metastases identified during surgery (26). Liver metasta- ses missed on CT are associated with lower contrast-to-noise ratio, subcapsular location, hepatic steatosis that masks lesions, and examination performed for indications other malignancy (27)" (Budigi et al., 2022) - "These images should be interpreted with caution because of the possibility of motion and other artifacts [39]. MRI is more susceptible to motion artifacts in comparison to MDCT; these motion artifacts can obscure small tumors and lead to misdiagnosis. Motion can result in image blurriness and artifact" (Haj-Mirzaian et al., 2020)
		Incomplete anatomical coverage	Haj-Mirzaian et al. (2020)				- "Furthermore, the pancreatic head and uncinate process may be incompletely included in axial MRI due to variability in patients' breath-hold
	Process	Operator dependence	Armato (2009)	Ahmed et al. (2024)	Kang et al. (2021)	LeBlanc et al. (2023)	- "Substantial variability exists across radiologists in the task of lung nodule identification in CT scans" (Armato, 2009) - "radiologists are often the first physicians to diagnose or raise the possibility" (Ahmed et al., 2024) - ""Understanding tumor spread patterns, identification of errors and their contributing factors as well as knowledge of approaches to reduce diagnostic errors enable radiologists to accurately stage/restage PDAC" (Budigi et al., 2022)
		Cognitive error	Haj-Mirzaian et al. (2020)	Hoogenboom et al. (2022)			- "Regarding the possible radiological errors that may have hampered accu- rate detection at the time of imaging, underreading was the most commonly identified error in our study (47% of errors in MRI, 57% in CT)" (Hoogenboom et al., 2022)
		Cognitive bias	LeBlanc et al. (2023)	Kang et al. (2021)	Busby et al. (2018)		- This highlights the importance of referring physicians to provide a thorough history on the imaging requisition, and may also reflect the cognitive biases that radiologists are susceptible to. An example of a cognitive bias is framing bias, which is the tendency to be influenced by how a question is asked or how a problem is presented. (LeBlanc et al., 2023)

D.4. Synthesis table for adverse case assessment: Patient layer

	Effect	Literature evidence				Citations
Patient	Intrinsic tumor features	Haj-Mirzaian et al. (2020)	Strobel et al. (2019)	Budigi et al., (2022)	Kang et al. (2021)	- "A small but clinically relevant proportion of patients are initially understaged and undergo laparotomy for resection fol- lowing CT, followed by abandonment of resection owing either to encroachment of the tumour on local structures or, more commonly, because micrometastases are iden- tified intraoperatively in the peritoneum and/ or liver." (Strobel et al., 2019) - "Most subtle tumors missed in our experience were small (<2 cm) hypoenhancing lesions in the uncinate process, isoattenuating tumors, and tumors on unenhanced imaging that were subtle but visible in retrospect" (Budigi et al., 2022)
	Tumor progression	Halle-Smith et al. (2024)	Wainberg et al. (2024)	Swords et al. (2015)	Hartwig et al. (2013)	- "Due to the rapid disease progression associated with these cancers, in particular pancreatic ductal adenocarcinoma (PDAC) and cholangiocarcinoma, it is highly desirable to establish a rapid preoperative diagnosis and initiate treatment quickly." (Halle-Smith et al., 2024) - They concluded that once PDAC becomes clinically detectable, progression from low-stage to advanced-stage disease is rapid." (Swords et al., 2015) - "Importantly, unnecessarily long follow-up might enable progression of branch-duct IPMNs into ductal adenocarcinomas" ("Hartwig et al., 2013)
	Patient-related factors	Strobel et al. (2019)	Park et al. (2021)	Haj-Mirzaian et al. (2020)		- "Well-defined risk factors for perioperative morbidity and mortality in patients undergoing pancreatic surgery can be classi- fied either as patient-related factors, such as age, obesity and comorbidities, or tumour-related factors, including the extent of local involvement of neighbouring organs and blood vessels" (Strobel et al., 2019) - "Focal fatty infiltration is a relatively common finding in the adult population, especially in old or obese subjects" (Haj-Mirzaian et al., 2020)
	coexisting or underlying pathologies	LeBlanc et al. (2023)	Kang et al. (2021)	Halle-Smith et al. (2024)	Ahmed et al. (2024)	- "Pancreatitis presents a challenge with respect to the imaging diagnosis and staging of PDAC for two reasons: first, PDAC can be misinterpreted as acute pancreatitis, and pancreatitis, particularly mass-forming pancreatitis, can be mistaken for malignancy. Second, inflammatory changes can mimic locally invasive disease, limiting assessment of resectability" (Kang et al., 2021) - "Given that a multitude of benign pathology, such as autoimmune pancreatitis or complex cystic lesions, can mimic malignant conditions, there is a challenge to providing surgical treatment within the limitations of the current diagnostic landscape" (Halle-Smith et al., 2024) - "There exists a significant overlap in the imaging features, clinical presentation, and risk factors of pancreatitis and PDAC" (Ahmed et al., (2024) - Imaging findings can also be misinterpreted for other pathology; for example, acute pancreatitis can mimic or mask PDAC" (LeBlanc et al., 2023)

E Survey form

Part 1 of 2 *

This first part of the questionnaire evaluates your experience with ASI compared to the current setup used for pancreatic MDTs at LUMC. By 'the current setup', it is meant that the outcome of regular MDTs is achieved based on verbal discussion and/or deliberation.

Each statement thus follows the sentence:
"In comparison to the way in which our regular MDT's are conducted, ..."

	1. Strongly disagree	2. Disagree	3. Neutral	4. Agree	5. Strongly agree
... the opportunity to compare my opinion against the group pushes me to provide more input	○	○	○	○	○
... being aware that my individual performance can be evaluated by the tool, pushes me to participate more	○	○	○	○	○
... the visualisation of the collective deliberation process of the pushes me to participate more	○	○	○	○	○
... the tool provides a more attractive way to participate which pushes me to provide more input	○	○	○	○	○
... The tool makes me confident that every member equally participates, pushing me to participate equally as well	○	○	○	○	○
... The fact that the tool requires participation pushes me to participate more	○	○	○	○	○

Part 2 of 2 *

This second part of the questionnaire evaluates your experience with ASI compared to the current setup used for pancreatic MDTs at LUMC. By 'the current setup', it is meant that the outcome of regular MDTs is achieved based on verbal discussion and/or deliberation.

Each statement thus follows the sentence:
"In comparison to the way in which our regular MDTs are conducted, ..."

	1. Strongly disagree	2. Disagree	3. Neutral	4. Agree	5. Strongly agree
... the anonymity provided by ASI encourages me more to express my opinion even when disagreeing with the collective decision	○	○	○	○	○
... ASI provides me the feeling that I have as equal influence on the decision outcome as all others	○	○	○	○	○
... the anonymity provided by ASI makes me feel more comfortable expressing my opinion when disagreeing with the collective decision	○	○	○	○	○
... the anonymity provided by ASI makes me more motivated to continue participating in the decision-making process when disagreeing with the group	○	○	○	○	○
... ASI eliminates influence caused by different personalities, experience and seniority in the group (i.e. introvert vs extrovert, expertise, specialisation)	○	○	○	○	○

Enablers and Barriers

6. From your perspective, name 3 potential enablers for the implementation of ASI to MDT meetings *

   In this respect, 'enabler' refers to anything organisational, technological or implementational that promotes the potential use of ASI in MDT meetings

   Enter your answer

7. From your perspective, name 3 potential barriers for the implementation of ASI to MDT meetings *

   In this respect, 'barrier' refers to anything organisational, technological or implementational that hinders the potential use of ASI in MDT meetings

   Enter your answer

F Extension of performance analysis: Ground truth and diagnostic comparison

Opening

Recap

In Chapter 5, the results were presented from the experiment conducted to address the improved assessment accuracy. For interpretation of the results, three comparisons are presented in Section 3.2: (1) An internal comparison, (2) a ground truth comparison, and (3) a diagnostic comparison. Previously in Chapter 7, the results from the internal comparison are discussed.

Introduction

This section progresses on the internal comparison by substantiating on the results of the ground truth comparison and diagnostic comparison. Whereas the internal comparison showed a 100% concordance level with no discordant case outcomes, the ground truth comparison and diagnostic comparison each have multiple discordant cases. Therefore, this section provides a more detailed examination of the discordant case outcomes and aims to identify underlying reasons for the modest concordance levels.

F.1. Ground truth comparison: Experimental outcomes versus ground truth

Based on the case set for ground truth comparison (Table 5.3), a 57.1% concordance level shows a moderate performance in resectability assessment, as this implies that in 42.9% of the cases, the experimental group has concluded an adverse outcome compared to the ground truth. Therefore, these deviating results could be perceived as deficient in terms of assessment accuracy. Particularly the large deviation in staging of SWARM-6324 may be initially perceived as notably apprehensive.

However, it should be noted that a conservative interpretation is suggested for the ground truth comparison since direct contrasting of the resectability assessment is subject to the following confounding factors.

- **Image quality**
  In line with reliance on image quality introduced previously (Section 4.4), analysis of the LUMC case history shows that the visibility of the PC tumor resectability for SWARM-4156 and SWARM-6324 was limited on the available CT-scans for these cases. Although the LUMC dataset on patient history did not confirm the underlying reason for obscured visibility, this is likely caused by patient-related factors including tumor location, surrounding tissue or anatomical variations (as introduced previously in Section 4.4).
- **Tumor progression**
  Analysis of the LUMC case history also showed the likely impact of tumor progression between the time point of preoperative diagnosis and the perioperative outcome on the assessment outcomes, particularly for cases SWARM-8945 and SWARM-4156. The observation that the perioperative outcomes of both cases (irresectable) were found more severe than their experimental outcomes (LAPC) based on CT scans of the initial MDT further supports the role of tumor progression as a potential contributor to the adverse assessment. Consequently (Section 4.4), although these cases show a discordant outcome between the experiment and ground truth datasets, it must therefore be considered that the experimental assessment of this case may have been correct given the available CT-images during the experiment.

All in all, direct comparison of the experimental and ground truth outcomes may overlook crucial complexities (at the patient level (Section 4.4) caused by image quality and tumor progression. Therefore, a nuanced approach is suggested when interpreting the discrepant assessment outcomes between the experiment and the ground truth dataset.

When analyzing the decision behavior of the discordant cases, the faction change graphs (Figure 5.6) show consistent results to the aforementioned confounding factors as follows:

- Overall, a lower degree of faction change is observed for the concordant cases during the swarm compared to the discordant cases. This suggests a higher level of conviction for concordant cases as participants tend to remain more consistent with their initial choice compared to the discordant cases
- SWARM-8945: When neglecting the 'inactive' users, the swarm uniformly concludes in LAPC (green). This suggests that the information presented during the swarm test was perceived as unambiguous, which may indicate that the seemingly different outcome of the swarm test (LAPC) and initial MDT (Resectable) is caused by different information or medical imaging presented between the Swarm test and the initial MDT
- SWARM-4156: When neglecting 'inactive' users, the swarm shows a discordant conclusion between LAPC (green) and irresectable (pink). This observation could support the notion that SWARM-4156 is the only discordant case where obscured tumor visibility was present, potentially leading to a higher degree of ambiguity in medical image interpretation.
- SWARM-6324: When neglecting the 'inactive' users, the swarm shows a uniform conclusion to resectable (light blue). This could affirm that, while the true tumor ultimately revealed to be metastasized during surgery, its resectability remained indiscernible from resectable tumors when assessing the CT scans available at the initial MDT (and thus collective discussion).

F.2. Diagnosis comparison: Experimental outcomes versus initial MDT outcomes

The 62.5% concordance level observed in the diagnostic case set (Table 5.4) also shows a moderate performance in resectability assessment against the 37.5% discordance level. When analyzing the discordant cases, a noticeable pattern in the confusion matrix (Figure 5.7) suggests that the experimental group may have been more severe in assessing patient resectability compared to the initial MDT. However, contrary to the ground truth comparison, the CT images used during the experiment are (assumed to be) the same as the CT images used during the initial MDT. Accordingly, the influence caused by confounding factors on the patient level (Section 4.4) such as tumor progression and image quality are therefore deemed unlikely. Instead, other confounding factors at the MDT level (Section 4.4) that may contribute to the discordant case outcomes are the following:

- **Contextual Information**
  The initial MDT likely had access to a broader context, including patient-related factors, patient history data, physical examination findings, additional medical (MRI) images, and other clinical information that was not available to the experimental group. Although it is acknowledged that this does not directly improve the assessment accuracy (due to cognitive biases arising from additional information (Section 4.4), this additional information can influence resectability decisions, potentially making the initial assessments less severe.
- **Group consistency**
  Although the experimental group remained consistent, the cases ranged between 01 January 2020 and 12 March 2024 and have therefore been assessed by various group compositions during the initial MDT. Therefore, it is likely that the original diagnosis of the majority of the cases was established by a varying selection of participants.
- **Consistency in radiological presentation**
  Following the range in case presentation (01 January 2020 and 12 March 2024), LUMC case data shows that the initial MDT meetings - where all cases selected for this experiment were presented - had different radiologists presenting their initial findings. Correspondingly, as introduced previously (Section 4.4), this inherently imposes inconsistency at the interpretation level due to variability in operator dependence, cognitive errors, and cognitive bias.
- **The Hawthorne effect**

The Hawthorne effect refers to the phenomenon where individuals alter their behavior in response to being observed or knowing that they are part of an experiment (Sedgwick & Greenwood, 2015). For this experiment, the setup may have provided exposure to the Hawthorne effect contributing to a more severe resectability assessment of the average participant.

G Enablers and barriers analysis

G.1. Results

Opening

Recap As one of the fundamental motivations for this research, testing ASI in the field allows for collecting empirical data on the use of the technology in a practical MDT setting. Aside from collecting the necessary datasets for this research, the experiment therefore also enables to gathering of additional evaluations and opinions regarding the potential use in a real-world MDT setting.

Introduction Secondary to the data collected for evaluating the impact of ASI on social influence, the survey is also used to capture additional data regarding potential enablers and barriers to the use of ASI in MDT meetings. To this end, the respondents were requested to provide feedback on these domains based on their own experience with ASI. The results of this survey can serve as a foundation for future research. This section presents the results of the enablers and barriers survey.

G.2. Practical survey group composition

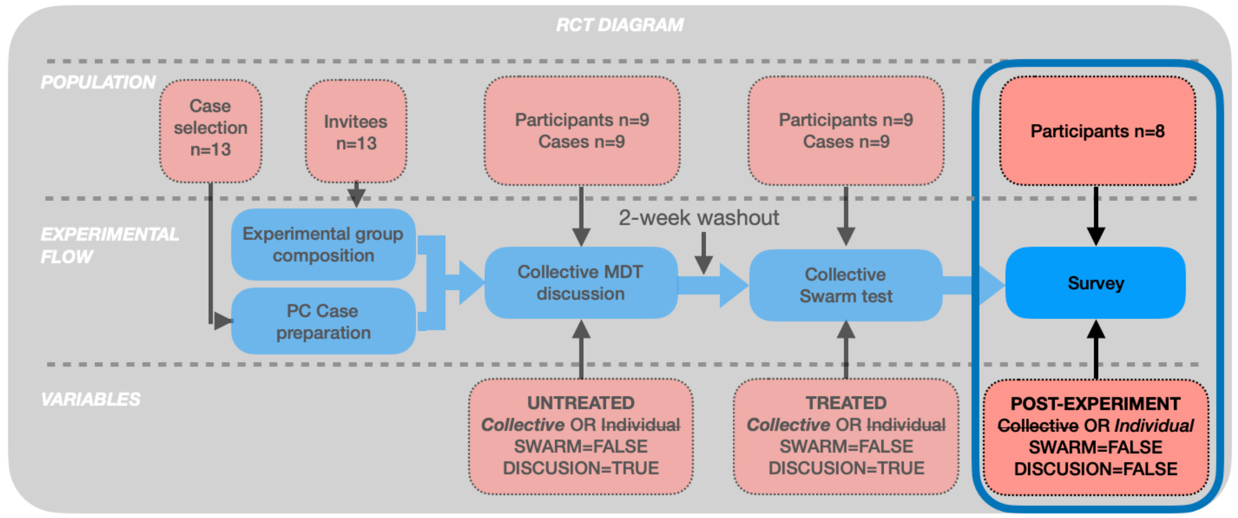


Figure G.1: Practical RCT diagram with focus on the survey (blue scope)

As presented previously (Section 3.2), the data for this analysis was collected as the second part of the survey described for the previous analysis (Chapter 6) (Figure G.1). Across both survey parts, the survey group composition remained consistent: Eight LUMC respondents (n=8) completed the survey between 24 May 2024 and 31 May 2024 based on their participation in the experiment. Distributions of the survey group thus remained consistent as well (Figure G.2): man-women=50%-50% with mean experience level=6.25 years and a medical expertise range of n=1-2 participants for specializations the *surgeon (n=1)*, *junior doctors (n=2)*, *radiologist (n=2)*, *doctor-researchers (n=3)* and *oncologists (n=1)*.

G.3. Results

G.3.1. Enablers

Qualitative analysis of the survey results shows a decent degree of concordant responses in terms of potential enablers for the use of ASI in MDT meetings. Based on the total set of responses, a set of eight themes is identified over 20 responses as follows:

- **Ease of use**
  Three (n=3) users reported unanimously that ASI provides an easy-to-use interface that is intuitive and uses a seamless workflow. The interface thereby contributes to an accessible method of participating,

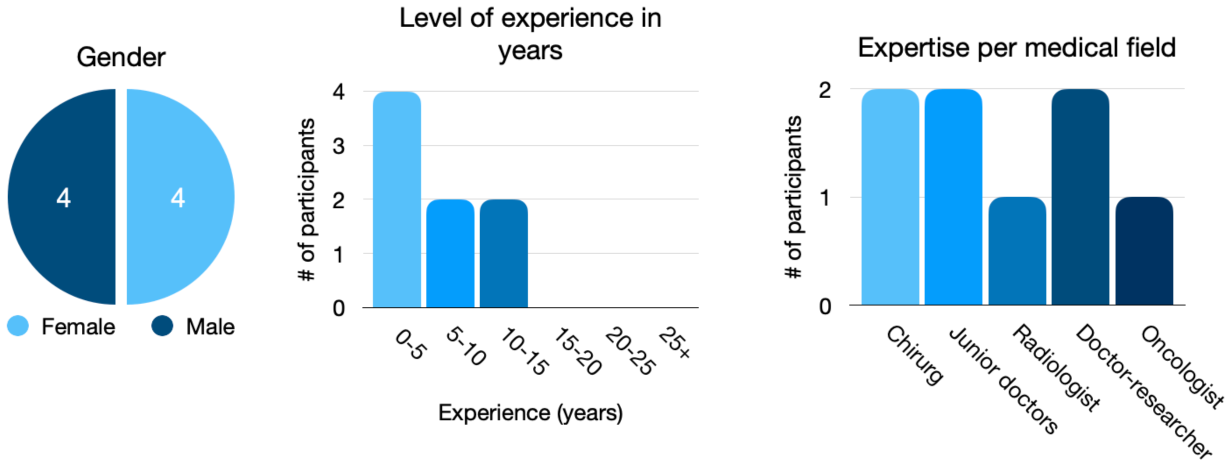


Figure G.2: Sociodemographic data of survey group

especially for users new to the technology.

- **Decision Quality**
  Another theme (n=3) entails the improved decision quality facilitated by ASI. Participants highlighted the demonstrated benefits of the technology, emphasizing enhanced decision-making processes and better quality decisions as enablers. Furthermore, the opportunity to involve only specialists ensures that the outcomes are based on (selected) expertise.
- **Engagement**
  Three participants (n=3) highlighted the role of ASI in fostering engagement among MDT participants. The inclusive nature of the platform pushes for active participation during an MDT, whilst the emphasis on teamwork takes optimal advantage of the available expertise.
- **Practical Implementation**
  Highlighted by three respondents (n=3), the implementation theme refers to other practical applications outside of the scope of this research. Respondents mentioned the importance of interface training to expand the usability of the technology to issues other than resectability assessment. In addition, the application of ASI to non-truth issues was mentioned, including patient treatment pathways. Instead of establishing the *truth*, ASI would then serve to create *consensus* for multi-perspective issues.
- **Equal Decision Influence**
  Two respondents (n=2) emphasized the equal decision influence afforded by ASI. The system ensures that each specialist has equal influence in the decision-making process, promoting fairness and inclusivity. This approach helps to balance the influence of different participants, regardless of their experience or status, leading to more democratic and equitable decision outcomes.
- **Visualisation**
  Two participants (n=2) noted the visualization capabilities of ASI as a key enabler. The platform's ability to provide clear and concise visual representations of complex data aids in the decision-making process, making it easier for participants to interpret and analyze information. The layout of the interface further enhances its usability, contributing to a more efficient and effective decision-making environment.
- **Easy Log-in**
  Two respondents (n=2) identified the easy log-in feature of ASI as an important enabler. The platform's design ensures that users can quickly and easily access the system, minimizing delays and disruptions. This feature is particularly beneficial in a medical setting, where timely access to information and decision-making tools is crucial.
- **Anonymity**
  Two participants (n=2) highlighted the anonymity feature of ASI as a significant enabler. By allowing users to participate anonymously, the platform encourages more honest and potentially unbiased (Chapter 6) contributions. This feature helps to mitigate the influence of hierarchical structures and

social pressures, ensuring that decisions are based solely on the merits of the information presented.

Theme	ID	Response
Ease of use	3	Ease of use
	5	Ease of use
	6	Ease of use
Decision quality	5	Demonstrated benefits
	7	Improved decision quality
	8	Only specialists
Engagement	4	Inclusivity for each MDT participant
	6	Teamwork
	7	Demonstrated more participation
Practical implementation	6	Interface training
	7	Application to non-truth issues
	6	Time benefits
Equal decision influence	4	Equal decision influence
	8	1 'vote' per specialist
Visualisation	1	Visualisation
	3	Lay-out
Easy log-in	2	Easy log-in
	8	Easy access (log-in)
Anonymity	3	Anonymity
	4	Anonymity

Table G.1: Enablers per theme

G.3.2. Barriers

For the aggregation of the barriers results, a set of five themes is identified over sixteen responses for the use of ASI in MDT meetings. The results have been aggregated as follows:

- **Lacking Exchange of Viewpoints**
  Four participants (n=4) identified several barriers within this theme. Most prominently, three respondents highlighted a lack ability of to argue different viewpoints between specializations, restricted opportunity to discuss resectability or treatment nuances and details, and no possibility to discuss treatment for patients. Particularly, one respondent emphasized that the considerations are sometimes more important than the outcome in patient care. These barriers thus limit the thorough exchange of ideas necessary for comprehensive decision-making.
- **Neglecting Distinction in Experience and Knowledge**
  Three participants (n=3) reported perceiving the equal decision influence as a key barrier. Especially since experience is reported as a significant factor in accurate assessment, this sociodemographic is noted as a significant restriction that could undermine the expertise-driven nature of decision-making in MDT meetings.
- **Endured Decision-Making Process**
  Contrary to time benefits as an enabler, three respondents (n=3) identified barriers associated with the enduring nature of the process. Especially for 'clear' cases, longer deliberation durations are found to (unnecessarily) delay decision-making. This can be considered a significant barrier in the typically time-pressured schedules of medical experts.
- **Technical Infrastructure**
  Four participants (n=4) pointed out barriers related to technical infrastructure, including issues with the availability of computer facilities, platform implementation, accessibility to computers, and the costs of computer facilities. These technical barriers could hinder the implementation of ASI into existing MDT workflows.
- **Biased Group Confirmation**

Two respondents highlighted barriers in the illusion of group confirmation. They reported a lack of influence when an individual opinion opinion diverges from the collective opinion. Also, another respondent questioned whether group thinking is truly minimized when individuals see the collective discussion.

Theme	ID	Response
Lacking exchange of viewpoints	4	Excessive focus on outcome instead of considerations
	6	Lacking argumentation of viewpoints across specializations
	5	Restricted discussion of viewpoint nuances and details
	7	No possibility to discuss treatment for patients
Neglecting distinction in experience and knowledge	6	Lacking decision influence based on experience
	7	Equal influence despite variability in experience/knowledge
	8	Neglecting difference in experience
Endured decision-making process	1	Enduring process
	5	Longer duration than discussion especially for ‘clear’ cases
	7	Delayed decision-making for ‘clear’ cases
Technical infrastructure	1	Availability of computer facilities
	3	Platform implementation
	5	Accessibility to computers
	6	Costs of computer facilities
Biased group confirmation	4	Lacking influence of divergent opinions
	4	Sensitivity to collective decision

Table G.2: Barriers per theme

G.4. Enablers and barriers analysis

G.4.1. Enablers

When comparing the number of themes over the number of responses, a greater dispersion is observed across the population for the themes of enablers in using ASI in MDT meetings compared to the barriers. Predominantly, the ease of use, decision quality, engagement, and practical implementation were observed as the most supported enablers of ASI - each supported by three respondents. For a more detailed interpretation of each theme or group of congruent themes, a description is given below

- **Ease of use, visualization, and easy log-in**
  These themes are perceived as key enablers for supporting further development and research of ASI in medical decision-making. Both enablers are deemed particularly important given the time availability of medical experts and the priority of making technological innovations accessible for them (Subsection 1.1.2).
- **Decision quality, engagement, equal decision influence, and anonymity**
  These enablers closely coincide with the core features of ASI Section 2.3. Consequently, this may support the notion that, despite its indifferent performance observed during the experiment Subsection 5.2.3, ASI offers features that are still perceived beneficial to medical decision-making - albeit in a broader range of questions or issues.
- **Practical implementation**
  Practical implementation also shows to be another potential enabler particularly significant when the technology is further developed. This includes the alignment of interface training with existing training, testing the technology to other types of questions or issues, and the potential time benefits of extensive discussions when optimizing the usage procedure of ASI.

G.4.2. Barriers

In contrast to the aggregation of responses for enablers, a more concentrated set of themes was derived for the barriers to using ASI in MDT meetings. Across all themes, lacking exchange of viewpoints and technical infrastructure were the two most strongly observed themes. For each theme, a description is given below:

- **Lacking exchange of viewpoints**
  As one of the most supported barriers, the lack of exchange of viewpoints largely is based on the inability ’to discuss treatment for patients’ - thereby leading to a ’restricted discussion of viewpoint nuances and details’. This barrier is justifiable from the experimental restriction that participants were disallowed to discuss during the swarm test, as well as the restricted interface option to discuss through chat during the swarm test. Nonetheless, to overcome these barriers, further avenues for research are suggested in Chapter 7.
- **Neglecting distinction in experience and knowledge**
  Contrary to situations where the knowledge and experience of individuals in a group are commensurate or interchangeable, the heterogeneous disciplines in MDT meetings necessitate the need for variability in decision influence. However, since a variability in decision influence contradicts one of the fundamental features of ASI, this advocates for testing the technology in a broader field of application. Both avenues for further research are discussed in Chapter 7
- **Endured decision-making process**
  Several respondents report a ’longer duration than discussion, especially for clear cases. This barrier can be justified by the experimental procedure where every case was assessed separately during the swarm test, regardless of the case complexity. However, this also entails a barrier that can be overcome through further research, as further discussed in Chapter 7.
- **Technical infrastructure**
  Another barrier reported for ASI is its reliance on the ’availability of computer facilities’ and the associated ’costs of computer facilities’. This is regarded as a barrier fundamental to the technology whenever deployed through large-scale implementation, implying potential implementation to also affect decision-making on an institutional or policy level.
- **Biased group confirmation**
  Despite a positive tendency to reduce social influence (Chapter 6), two respondents reported the ’sensitivity to collective decision’ in parallel with ’lacking the influence of divergent opinions’. Accordingly, this notion emphasizes further research on social bias, as further discussed in Chapter 7.