Optimizing Operating Room Planning in Bariatric Clinics Through Data-Driven Insights

An Exploratory Case Study of Planning Processes at Vitalys Bariatric Clinic

BM51035: BME MSc Thesis

Catrien Stolle

Delft University of Technology

TU Delft

Optimizing Operating Room Planning in Bariatric Clinics Through Data-Driven Insights

An Exploratory Case Study of Planning Processes at Vitalys Bariatric Clinic

by

Catrien Stolle

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday 18th of December, 2024 at 09.30 AM.

Student number:	5805279	
Project duration:	May 1, 2023 – December 18, 2024	
Thesis committee:	Prof. dr. ir. J. van den Dobbelsteen,	TU Delft, supervisor
	Prof. dr. B. H. W. Hendriks	TU Delft
	E. Frassini	TU Delft
	D. Huisman	Johnson and Johnson, daily (external) supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

This thesis concludes my master's studies at TU Delft and represents a journey of discovery into the complexities of bariatric clinic operations and surgical planning. Collaborating with Vitalys Clinic and Johnson & Johnson Medtech provided me with the unique opportunity to witness the dedication of professionals striving to improve patient care and operational efficiency in dutch hospitals. Their passion and drive have been a constant source of motivation for me throughout this project. It is my hope that this thesis contributes to meaningful improvements in the patient planning and workflow, supporting Vitalys Clinic in advancing their processes.

This work would not have been possible without the guidance, encouragement, and support of many individuals. I would like to express my gratitude to Kristy Leenders from Vitalys Clinic for her invaluable insights and unwavering encouragement, as well as to Daitlin Huisman and the Johnson & Johnson Obesity Team for their expertise and thoughtful guidance throughout the research process. At TU Delft, I am thankful to John van den Dobbelsteen, whose trust in my abilities and constructive feedback allowed me the freedom to explore my ideas while refining my work.

Finally, I am profoundly grateful to my family and friends for their steadfast support and encouragement throughout this journey. Their belief in me, especially during challenging moments, has been an enduring source of strength. To everyone who contributed to this project in any capacity—thank you. Your support and involvement have been instrumental in making this work possible.

Catrien Stolle
Delft, November 2024

Summary

The transition from inpatient to outpatient care has revolutionized the healthcare, reducing hospital admissions and improving patient access through innovations such as laparoscopic surgery and advanced monitoring technologies (Richards, Seward, and Whaley 2022; Waghorn, McKee, and Thompson 1997). However, this shift has also introduced significant challenges to operational efficiency, particularly in clinics that manage chronic conditions like obesity (Aissaoui et al. 2024).

Obesity, identified as a global health crisis by the World Health Organization, presents significant public health challenges, particularly in the Netherlands, where over half of adults are classified as overweight or 14% as obese (World Health Organization 2021; National Institute for Public Health and the Environment (RIVM) 2021). Bariatric clinics like Vitalys aim to address these challenges by offering procedures such as gastric bypass and sleeve gastrectomy to improve health outcomes. However, inefficiencies in planning and scheduling processes at these clinics result in extended patient wait times, underutilized operating rooms, and administrative strain (Bielen and Demoulin 2007; Kern et al. 2021).

While traditional methods for optimizing planning focus on structured scheduling systems (Kam et al. 2021; Wojtys et al. 2009), techniques like neural networks, decision trees, and simulation models offer opportunities to predict patient flow and improve scheduling (Glowacka, Henry, and May 2009; Srinivas and Ravindran 2018). Studies have shown that leveraging AI and big data can optimize scheduling by accounting for patient preferences, no-shows, and cancellations, ultimately enhancing clinic performance (Golmohammadi 2021; Najmuddin, Ibrahim, Ismail, et al. 2010). Consequently, many AI-driven tools have been developed in recent years to improve planning and capacity in healthcare settings.

The Vitalys Bariatric Clinic, part of Rijnstate Hospital, highlights these challenges, struggling with inefficiencies in their planning and capacity that hinder operational efficiency and patient flow. Despite advancements in outpatient treatments and day procedures, the clinic currently schedules surgeries only two weeks in advance, resulting in extended wait times, underutilized resources, and increased administrative burdens. This study investigates the root causes of these inefficiencies through an exploratory analysis that integrates process mapping, stakeholder interviews, and data-driven insights to uncover bottlenecks in planning and scheduling. By addressing these issues, the research provides actionable recommendations for improving resource utilization and patient care while offering broader insights into workflow optimization for similar healthcare settings.

The research begins by analyzing operational workflows at Vitalys Clinic using the Functional Resonance Analysis Method (FRAM). A FRAM-as-imagined model was created to understand how the process is theoretically designed to function, while a FRAM-as-done model maps the actual execution of the process as it occurs in practice. These models helped identify sources of variability and inefficiencies in the planning process. Semi-structured interviews with stakeholders were conducted to validate these findings and uncover key drivers and barriers to scheduling efficiency. The results highlighted significant variability in the screening and operating room (OR) planning processes, primarily caused by administrative burdens and inefficiencies in the scheduling workflow. Additionally, a critical gap was identified: it remains unclear how many patients need to visit other departments during the process, which can introduce further delays and complicating the planning process.

The study also examines two years of patient and operational appointment data to identify trends, patterns, and bottlenecks in patient flow. A delay analysis highlights the most significant delays and their duration, revealing that bottlenecks often occur during pre-surgical phases where multi-department appointments introduce variability. The analysis found that only 25% of patients experience significant delays, defined as exceeding seven weeks after the start of the group trajectory. Further investigation into this delayed group revealed that patients who require visits to other departments are significantly more likely to encounter delays. Based on these findings, a new planning model is proposed. This model categorizes patients into two distinct groups: one group requiring additional departmental visits, which are scheduled later, and another group of patients who can be planned six weeks in advance.

This distinction aims to reduce delays and optimize the scheduling process by prioritizing patients based on their readiness for surgery.

The study further explores predictive modeling using AI techniques, including Random Forest and decision tree algorithms, to assess whether delays can be anticipated based on factors like appointment history and specialty-specific characteristics. While predictive models demonstrate potential in the literature, this study finds their effectiveness limited by the quality and availability of data. Importantly, the findings indicate that implementing AI is not strictly necessary to address the identified challenges. The proposed planning model, which leverages data-driven insights and minimal system adjustments, resolves many inefficiencies effectively. This suggests that traditional data-driven methods, when applied strategically, can sufficiently enhance planning processes without requiring advanced AI tools.

In conclusion, this thesis emphasizes that improving planning processes in bariatric clinics requires addressing several interconnected elements, including data analytics, stakeholder collaboration, and process analysis. The study shows that significant improvements can be achieved without the necessity of relying on Artificial Intelligence (AI). By utilizing data-driven insights and optimizing operational workflows, the proposed planning model at Vitalys Clinic has the potential to reduce delays, enhance operating room utilization, and improve patient care. The findings also suggest that while AI could further optimize planning processes in certain situations, it is not always crucial for achieving substantial improvements. Future research should focus on refining data collection methods, standardizing operational metrics, and exploring the scalability of these solutions to other healthcare environments, aiming to address a broader range of operational challenges.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
BMI	Body Mass Index
OR	Operating Room
FRAM	Functional Resonance Analysis Method
GP	General Practitioner
WHO	Wolrd Health Organization
WAI	Work-As-Imagind
WAD	Word-As-Done

Symbols

Term	Definition	Context
Vitalys Core Specialties	The primary departments within Vitalys Clinic that primarily manage patient care: Dietetics, Psychology, Nursing, Anesthesiology, and Surgery.	Bariatric Clinic
Bariatric Care	A specialized medical approach focused on weight loss treatment for obesity, often involving procedures such as gastric bypass or sleeve gastrectomy.	Healthcare
Screening Phase	The initial evaluation stage where patients are assessed for eligibility and readiness for bariatric surgery	Screenings Process
OR Phase	The Operating Room scheduling phase including the pretreatment group appointments	Surgical Planning
REDO Surgery	Follow-up bariatric surgeries for patients who previously underwent bariatric procedures	Bariatric Surgery
Predictive Modeling	Statistical and machine learning techniques applied to predict delays and streamline scheduling, enhancing workflow efficiency within the clinic.	Data Analysis
Lead Time	The total time from initial patient screening to surgery, first groups meeting to surgery used to assess scheduling efficiency.	Data Analysis
Postponed Patients	Patients whose surgical timelines have been delayed, often due to additional required appointments in departments outside of the Vitalys Core Specialties.	Scheduling

1 Introduction

One of the most serious global health issues today is the rise in chronic illnesses, particularly in obesity. The World Health Organization (WHO) has identified obesity as a major health concern due to its strong association with serious conditions such as cardiovascular disease, type 2 diabetes, and certain types of cancer (World Health Organization 2021). In the Netherlands, obesity rates are rising steadily, with over half of adults classified as overweight and over 14% as obese (National Institute for Public Health and the Environment (RIVM) 2021). This growing patient population underscores the importance of optimizing clinic operations and patient flow. For patients with chronic diseases, efficient patient flow is crucial for delivering coordinated care, reducing wait times, and improving outcomes.

Effectively managing patient flow and hospital operations has become increasingly challenging, particularly in clinics treating a growing number of patients with chronic conditions (Ordu et al. 2023). This challenge is further increased by the rising demand for healthcare services driven by population growth in the Netherlands, an aging demographic, and the introduction of more advanced medical treatments and procedures (Roy, Prasanna Venkatesan, and Goh 2021). At the same time, critical staffing shortages highlight the need for efficiency improvements to sustain healthcare systems. By optimizing processes, the same number of staff can deliver more care, keeping healthcare affordable and accessible even as demand continues to rise (Griffiths et al. 2020). Without these improvements, healthcare risks becoming financially unable to meet the needs of an expanding population.

Moreover, hospitals face increasing pressure to enhance patient care due to evolving healthcare standards, regulations, and advancements in medical treatments (Roy, Prasanna Venkatesan, and Goh 2021). Additionally, economic constraints and the growing complexity of healthcare systems have driven a significant shift from inpatient to outpatient settings to reduce costs and improve efficiency (Waghorn, McKee, and Thompson 1997). Over the past decade, this transition has reduced prolonged inpatient hospital admissions, demonstrating the effectiveness of new technologies and management strategies (Antony 2005), (Hensher et al. 1999). Innovations in monitoring technology, diagnostic tools, and treatment techniques, such as laparoscopic surgery, have enabled many procedures to be performed as outpatient treatments, including minor operations and same-day procedures (Richards, Seward, and Whaley 2022).

While these advancements have improved patient accessibility and reduced the need for hospitalization, they have also introduced new operational challenges (Aissaoui et al. 2024). Optimizing operational efficiency and patient flow has therefore become crucial to ensure equitable access to care for all patients. One key factor in this optimization is improving planning and scheduling, which enhances both patient satisfaction and operational efficiency. Previous research has demonstrated the effectiveness of lean methodologies and other strategies to enhance scheduling systems (Kam et al. 2021), (Wojtys et al. 2009). Furthermore, data-driven and Artificial Intelligence (AI) solutions have emerged as powerful tools for addressing scheduling challenges. Techniques such as neural networks and decision trees can predict service times and patient no-shows, enabling clinics to reduce idle time and improve scheduling reliability (Glowacka, Henry, and May 2009), (Srinivas and Ravindran 2018). For example, Lenin et al. 2015, successfully reduced waiting times in obstetrics and gynecology clinics by optimizing appointment systems and classifying patients based on their likelihood of attending appointments.

Additionally, leveraging big data can improve appointment scheduling by optimizing time and resource utilization. Simulation models, when combined with optimization algorithms, have further refined scheduling by accounting for patient preferences, no-shows, and cancellations, ultimately enhancing clinic performance (Golmohammadi 2021). For instance, Najmuddin, Ibrahim, Ismail, et al. 2010 utilized discrete event simulation to streamline patient flow and reduce waiting times in outpatient obstetric settings. With the growing demand for bariatric care in the Netherlands, refining their planning and scheduling systems in these clinics is essential to ensure timely, high-quality care for patients. This goal is increasingly supported by advancements in data analytics and artificial intelligence (L. Li, Diouf, and Gorkhali 2022), (Rutherford et al. 2017).

Vitalys Clinic, a division of Rijnstate Hospital in the Netherlands, specializes in obesity treatment through bariatric surgeries and medical interventions. The clinic performs procedures such as gastric bypass and gastric sleeve surgeries to help patients manage obesity and improve related health outcomes (Vitalys 2024). As an outpatient clinic, Vitalys delivers surgical care without overnight stays. However, the clinic faces significant operational challenges, particularly in scheduling and planning. These inefficiencies result in extended waiting times and delays, which directly impact patient satisfaction, a key measure of healthcare quality (Bielen and Demoulin 2007), (Kern et al. 2021). Long wait times not only reduce care quality but also negatively affect the patient experience. Additionally, administrative burdens and time constraints on healthcare personnel contribute to burnout, further impacting the quality of care (Kleiner and Wallace 2017).

Recognizing the potential of data analytics and AI, Vitalys Clinic aims to explore how these technologies can enhance operational efficiency and improve patient outcomes. AI and data analytics present significant opportunities to optimize scheduling, predict patient flow, and streamline daily operations, potentially reducing bottlenecks and improving service delivery. However, it is also crucial to evaluate whether AI-driven solutions are feasible and suitable for addressing the clinic's unique challenges. This exploratory case study focuses on the operational challenges faced by Vitalys Clinic in scheduling, planning, and capacity management. By leveraging data analytics and exploring the potential of artificial intelligence (AI), the study aims to optimize patient flow and improve scheduling processes to reduce waiting times. It will examine the clinic's planning and capacity workflows to identify inefficiencies and develop a more dynamic, patient-centered planning model. Additionally, the study will assess the applicability of AI in enhancing these processes, ensuring that the proposed solutions align with the clinic's operational goals. Ultimately, this approach seeks to align the clinic's operations with its objectives of improving efficiency, reducing delays, and enhancing patient outcomes.

1.1. Problem Statement

The Vitalys Bariatric Clinic, part of Rijnstate Hospital, faces ongoing challenges in its planning and capacity processes, which negatively impact operational efficiency and patient flow. Despite advancements in outpatient treatments and day procedures, the clinic continues to experience inefficiencies in surgery scheduling and can only schedule patients two weeks in advance. These limitations result in extended patient wait times, underutilized operating rooms, and increased administrative burdens, all of which undermine patient satisfaction and overall care quality. Through this case study of Vitalys Bariatric Clinic, this research aims to explore planning issues and operational efficiency challenges common to bariatric clinics. By addressing these problems within this clinic, the study will also contribute to broader insights on improving operational workflows and resource management in similar healthcare settings.

1.2. Research Objective and Scope

To address these challenges, this research will conduct an exploratory study to investigate the root causes of inefficiencies in the planning and capacity processes at the Vitalys Bariatric Clinic. The study will use methodologies including process analysis, data collection and analysis, and stakeholder interviews to identify operational bottlenecks and provide actionable recommendations for improving clinic performance and patient care efficiency. The research will focus on analyzing patient flow, workflow, and scheduling challenges, aiming to identify specific areas for improvement in planning and capacity management.

By leveraging data-driven insights, the study seeks to develop solutions that enhance clinic and patient flow operations and optimize resourche utilization. To achieve this, the research will pursue the following objectives:

- Investigate the current workflow at the clinic to identify bottlenecks and sources of variability in the scheduling process.
- Analyze stakeholder insights through interviews to better understand operational challenges.
- Use hospital data to detect trends, patterns, and inefficiencies in patient flow, resource utilization, and scheduling.
- Propose data-driven solutions and develop a new planning model to optimize operations and enhance patient flow.
- Investigate whether predictive models could enhance the planning process within Vitalys Clinic.

Additionally, the research will assess the trade-offs between AI-driven solutions and traditional methods for scheduling and capacity management. This includes evaluating whether AI implementation is necessary and appropriate, or if optimization through traditional data-driven methods could sufficiently address the clinic's challenges.

1.2.1. Research Scope

This study focuses on identifying areas for improvement and using data-driven approaches to address inefficiencies in patient flow and scheduling at the Vitalys Bariatric Clinic. Factors such as staff availability and material capacity are excluded, as they are assumed to remain stable within the scope of this research. The primary aim is to identify operational inefficiencies and explore how data, and potentially AI, can be leveraged to improve planning and capacity management for patients.

By doing so, the research seeks to enhance both patient and workflow efficiency in bariatric clinics. Furthermore, the findings may have broader implications for other outpatient services across healthcare centers in the Netherlands, providing data-driven solutions that could be generalized beyond the bariatric context.

1.3. Research Questions

This study seeks to answer the following main research question:

> "*What are the current challenges in planning and capacity at the Bariatric Clinic, and how can data analytics and Artificial Intelligence be leveraged to optimize these processes, thereby enhancing patient care and workflow efficiency?*"

The following subquestions will be explored in different chapters using various methods to identify current challenges and uncover data that can enhance the clinic's planning efficiency:

- What are the key steps in the current workflow at the Vitalys Bariatric Clinic and which contribute most to variability and inefficiencies in the workflow?
- How does variability affect overall workflow and scheduling efficiency, according to stakeholder feedback?
- What drivers and barriers are identified by stakeholders?
- What data can be used to analyze the plannings process, and what patterns or trends affecting patient flow can be identified?
- How can data-driven insights be applied to improve the efficiency of planning processes and be integrated into a planning model to improve the planning processes at Vitalys Clinic?
- Are there predictive possibilities to enhance the planning process at Vitalys Clinic and will it enhance the traditional scheduling process?

2

Background Information

2.1. Vitalys: Clinic Against Obesity

As obesity is recognized by the World Health Organization (WHO) (World Health Organization 2021) as one of the most critical public health challenges of the 21st century, many clinics have been established to combat this chronic disease. Obesity contributes to a wide range of health risks, including cardiovascular diseases (Poirier et al. 2006), type 2 diabetes, and certain cancers (Pati et al. 2023). Targeted therapies against obesity are essential to reduce health risks and improve patients' quality of life. Vitalys Clinic, part of Rijnstate Hospital, offers bariatric (surgical) and medical treatments to help individuals manage their weight effectively. As an outpatient clinic, Vitalys provides treatment that is more accessible and less disruptive for patients (Vitalys 2024)

2.1.1. Bariatric Care

The two main bariatric procedures carried out at Vitalys Clinic are the gastric bypass and the gastric sleeve, both of which are very successful in stimulating notable weight loss and enhancing related medical conditions (Grönroos et al. 2020), (Vitalys 2024). In the gastric bypass (Roux-en-Y), a small stomach pouch is created and directly connected to the small intestine, bypassing most of the stomach and a section of the small intestine. This reduces the quantity of food that can be consumed and the number of calories and nutrients that the body can absorb by avoiding the majority of the stomach and a section of the small intestine (Cui et al. 2021). Additionally, by decreasing the production of the hunger hormone ghrelin, the gastric bypass not only restricts food intake but also lessens appetite, resulting in greater long-lasting weight loss (Colquitt et al. 2014).

The gastric sleeve removes approximately 80% of the stomach, creating a structure known as the "sleeve." Like the gastric bypass, the sleeve limits food intake and significantly decreases hunger hormone production, aiding in effective weight loss (Karamanakos et al. 2008). Due to its ease of use and efficacy, this operation has gained popularity; many patients report better results for ailments like sleep apnea, high blood pressure, and type 2 diabetes following the treatment (Grönroos et al. 2020). REDO surgeries are performed for patients requiring further intervention due to complications or insufficient weight loss following initial bariatric procedures. Although less common, this surgery is sometimes essential for ensuring long-term success (Quezada et al. 2016).

At Vitalys Clinic, both surgeries are carried out laparoscopically, which guarantees less scarring, a quicker recovery, and less postoperative pain (Raj et al. 2018). People with a body mass index (BMI) of 40 or higher, or those with a BMI of 35 or higher who also have associated medical issues such diabetes or hypertension, are generally advised to have these operations (Kehagias et al. 2011). As mentioned, Vitalys uses a day-care model for bariatric surgery, allowing patients to arrive in the morning and go home after the procedure, which promotes efficiency and convenience for the patients. Both the gastric bypass and the sleeve gastrectomy result in substantial and long-lasting weight loss, coupled with improvements in metabolic conditions and low rates of complications (Cui et al. 2021); type 2 diabetes, hypertension, and sleep apnea have all been shown to be effectively improved by gastric bypass surgery, with results frequently continuing for up to ten years after the procedure (Syn et al. 2021).

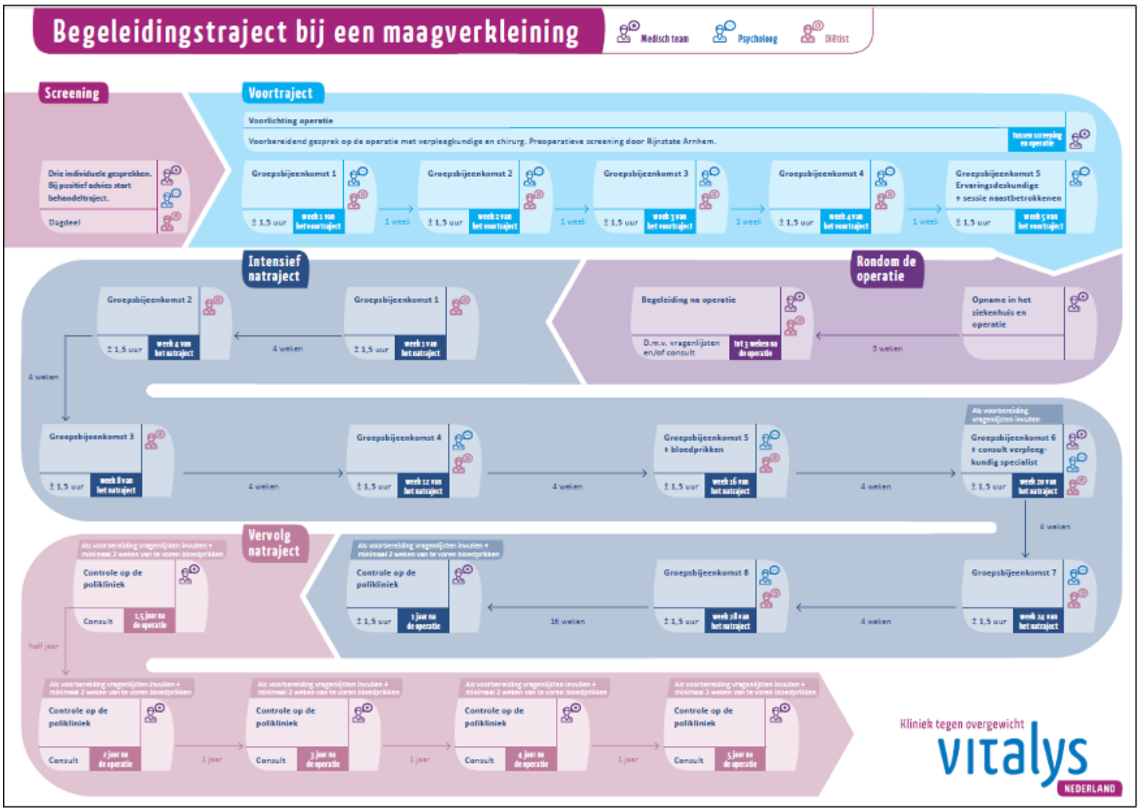

Figure 2.1: Treatment Plan for Bariatric Surgery by Vitalys (Vitalys 2024)

Patients at Vitalys Clinic undergo a thorough screening and preparation process before surgery to ensure they are physically and psychologically ready for the major lifestyle changes that follow these life-changing procedures. The initial screening is crucial, as it assesses whether patients are prepared for the postoperative recovery process and committed to the necessary lifestyle adjustments (Lauti et al. 2016). At Vitalys, these surgeries are part of a comprehensive treatment strategy, as shown in Figure 2.1. For most patients, the *"Begeleidingstraject bij een maagverkleining" (Guidance Pathway for Gastric Reduction)* (Vitalys 2024) includes five group appointments over six weeks, though some patients follow an individual pathway with four appointments during the same period.

In order to give patients full understanding regarding the procedure and its effects, the group visits include sessions with nurses, psychiatrists, and dietitians. These sessions focus on the important lifestyle adjustments that must be made following the surgery, especially with regard to eating habits and long-term behavioral adjustments that are essential for success. The sessions guarantee that patients are informed and assisted at every stage of their journey, assisting them in overcoming the difficulties related to a new lifestyle following surgery.

2.1.2. Medical care

Pharmacotherapy, next to surgical procedures, helps Vitalys Clinic patients lose weight. These drugs are typically recommended when diet and exercise prove insufficient, targeting patients with a BMI of 30 or higher or those with obesity-related conditions such as type 2 diabetes, hypertension, or cardiovascular disease (Aaseth et al. 2021)

One of the most frequent methods that is used is appetite suppression, where medications impact the central nervous system or imitate satiety chemicals. GLP-1 receptor agonists like liraglutide and semaglutide are frequently used for these treatments. These drugs promote weight loss by reducing calorie intake and creating longer-lasting fullness (Aaseth et al. 2021). Different drugs restrict fat absorption in the digestive tract, preventing the body from absorbing a portion of dietary fat (Pilitsi et al. 2019). Combining pharmacotherapy with lifestyle changes, including as diet modification, physical activity, and behavioral counseling, is most successful (Kheniser, 2021). It is part of a therapy plan to permanently lose weight and reduce obesity-related health risks. These medications form a component of an entire treatment; to enhance overall health and well-being, not as a solo solution (Saunders et al. 2018).

Vitalys Clinic's weight reduction treatment begins with a referral and initial consultation with an internist, nurse specialist, or physician assistant, followed by medical history, blood tests, and weight measures. Patients receive in-person and phone consultations over two years to track progress, modify medications, and promote weight loss. Vitalys Clinic's combination of surgery, drugs, and lifestyle adjustments is vital to combating the Netherlands' obesity crisis. Since obesity care is voluntary and may be scheduled ahead of time, scheduling and resource management can be improved. However, it is crucial to discover variability in outpatient clinic planning before surgery. Understand these sources of variability will help the clinic improve workflow, reduce wait times, and optimize resource use across the patient care journey.

2.2. Types of Variability impacting the Outpatient Clinic Efficiency

Vitalys Clinic and other outpatient clinics face numerous types of variability that disrupt patient flow and operational effectiveness, ultimately impacting planning and capacity. Based on insights from prior research (Stolle 2024), these challenges have been categorized, along with examples of how similar challenges have been addressed in existing studies.

Administrative Inefficiencies

Administrative inefficiencies significantly contribute to variability, often leading to prolonged patient wait times due to delays in processes such as registration, patient filing, and medical billing. For example, Mustapha et al. 2016 found that implementing electronic health record systems streamlined patient information management, alleviating bottlenecks caused by inefficient filing and sluggish registration processes. Similarly, Tayne et al. 2018 highlighted how automating patient intake workflows improved clinic operations and reduced inefficiencies. Both studies leveraged lean methodologies to address these issues.

At Vitalys, delays in the registration and intake process may hinder preoperative assessments, leading to scheduling disruptions. Incomplete patient records can require rescheduling or postponing appointments, further compounding inefficiencies and creating bottlenecks in the surgical scheduling process.

Resource Constraints

Limited availability of medical personnel, clinic space, and supplies is another significant challenge. Research has demonstrated that resource optimization strategies can mitigate these issues. For instance, Bernatchou et al. 2017 proposed predictive models to forecast resource requirements, while Kubala et al. 2021 investigated dynamic room assignment adjustments based on patient demand. Additionally, Liang et al. 2015 explored improved staff shift planning to alleviate resource shortages and enhance operational efficiency.

At Vitalys, where bariatric surgeries are in high demand, careful planning of surgical teams and effective room assignments are critical to maintaining consistent patient flow. Resource bottlenecks, such as insufficient staff or space, can disrupt the patient journey, resulting in prolonged wait times and underutilized resources.

Inefficient Scheduling

Suboptimal scheduling practices also contribute to variability. Clinics that schedule appointments too far in advance or fail to account for patient no-shows often experience delays and inefficiencies. Recent studies have demonstrated the effectiveness of data-driven models and optimization algorithms in addressing these challenges (Song, Bai, and Wen 2018; Srinivas and Ravindran 2018). For example, Kern et al. 2021 showed how machine learning could predict patient no-shows, enabling clinics to strategically overbook and minimize idle time. Similarly, Hua et al. 2023 demonstrated that data-driven scheduling models more accurately align appointment times with resource and staff availability.

At Vitalys, ineffective scheduling practices—such as failing to account for cancellations, only planning 2 weeks in advance or patient no-shows—can lead to underutilized operating rooms or overburdened staff. Adopting a more dynamic, data-driven scheduling approach could balance resource utilization and significantly improve operational efficiency.

Dynamic Factors

Dynamic factors related to human behavior, such as inconsistent patient arrival times and variable consultation lengths, further complicate clinic operations (Lee et al. 2022). For instance, Zou, Wang, and Cheng 2022 used simulation models to predict service time variability, optimizing clinic workflows despite these fluctuations. Likewise, Famiglietti et al. 2017 demonstrated how real-time data could help manage consultation time variability, ensuring a more consistent workflow. Additionally, Mesko et al. 2022 analyzed patient arrival patterns to design dynamic resource allocation systems.

At Vitalys, irregular consultation times or changing patient circumstances can complicate scheduling and the operational efficiency. These dynamic conditions often delay preoperative evaluations, underscoring the need for a robust, adaptable planning model capable of accommodating variability.

Understanding these sources of variability—whether stemming from administrative inefficiencies, resource limitations, ineffective scheduling, or dynamic human factors—is critical for improving Vitalys Clinic's planning processes. By addressing these inefficiencies and incorporating solutions such as data-driven strategies and adaptable planning models, this study aims to enhance workflow efficiency, optimize resource utilization, and reduce delays within the clinic.

2.3. Methods for Optimizing Planning and Capacity

Data-driven techniques and structured frameworks areused in this study to adress Vitalys' outpatient clinic planning inefficiencies. A widely referenced framework by Hans, Van Houdenhoven, and Hulshof 2011, is used to assess scheduling at strategic, tactical, and operational levels in healthcare. This framework deconstructs the scheduling process into multiple levels and facets, providing a comprehensive understanding of planning. This study focuses specifically on the offline operational level of the framework, where scheduling is pre-planned to enhance efficiency.

Additionally, the Functional Resonance Analysis Method (FRAM) (Hollnagel 2017) is applied to evaluate the impact of daily operational variability on clinic performance, encompassing preoperative evaluations and surgery scheduling. By breaking tasks into functional units, FRAM identifies interdependencies and critical areas where enhancements could reduce delays and optimize patient flow. During the first phase of this research, this structured approach is instrumental in pinpointing variability and deciding which areas of the planning process warranted focused improvement.

Data analytics complements these frameworks by analyzing appointment patterns over a two-year period to identify scheduling inefficiencies. The study investigates areas with significant delays and specialties most affected by them. To explore whether Artificial Intelligence (AI) could further enhance planning, predictive modeling techniques, such as decision trees, are evaluated to identify factors contributing to delays. By combining these frameworks and analytics, the study develops optimized strategies aimed at reducing patient wait times, improving planning efficiency, and enhancing overall clinic operations. The methodologies used in this research are detailed below.

2.3.1. A Framework for Healthcare Planning and Control

Due to the clinic's difficulties with forward planning, it is essential that its current procedures be reviewed. A structured method for controlling healthcare processes at the strategic, tactical, and operational levels is the Hierarchical Planning and Control Framework, which is shown in Figure 2.2 (Hans, Van Houdenhoven, and Hulshof 2011). Vitalys operates in a Low Variability, High Dependency (LH) environment, where surgeries like gastric bypass and sleeve gastrectomy are predictable but rely heavily on synchronized processes such as preoperative assessments, operating room availability, and postoperative care (Hans, Van Houdenhoven, and Hulshof 2011). Delays in these areas can cause scheduling bottlenecks and reduce efficiency.

At the strategic level, long-term decisions involve resource planning, such as determining the number of operating rooms required and investing in advanced equipment to meet patient volume projections.

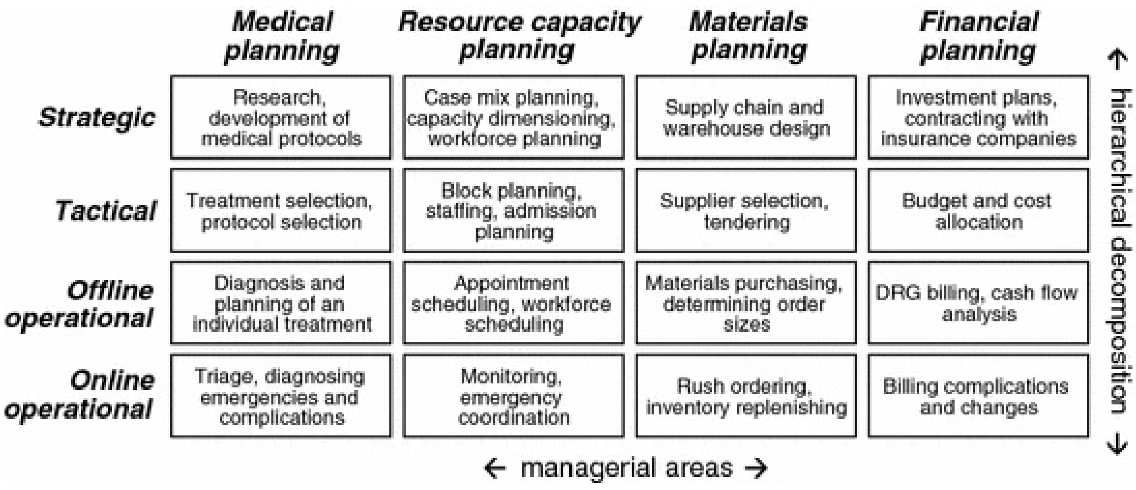

Figure 2.2: Hierarchical Framework by (Hans, Van Houdenhoven, and Hulshof 2011)

Tactical planning focuses on ensuring resource availability in the medium term, like surgical teams and medical supplies, to balance patient loads.

The operational level divides into offline and online planning, each managing distinct aspects of daily operations. Offline planning involves advance scheduling of surgeries, appointments, and related procedures based on forecasted needs, aiming to create structured schedules that optimize OR slots, equipment, and staffing. This approach reduces inefficiencies by organizing resources in advance to maximize productivity. On the other hand, online planning manages real-time disruptions, such as cancellations, delays, or emergency cases that require immediate OR access. It demands flexibility and quick decision-making to adjust schedules and reallocate resources promptly. This dynamic management minimizes disruptions, ensuring patient care continuity and efficient workflow.

This research focuses on offline operational planning, where structured advance scheduling can significantly improve Vitalys Clinic's ability to manage variability and optimize resources. In LH environments like Vitalys, where predictable procedures depend on interconnected processes, effective advance planning is essential to reduce bottlenecks and ensure efficient resource allocation.

2.3.2. Functional Resonance Analaysis Method (FRAM)

FRAM (Hollnagel 2017) is used to analyze daily operational variability in Vitalys' workflows. Unlike linear models, FRAM examines interdependencies between tasks to capture the complexity of real-world healthcare operations (Raben et al. 2018). It has been widely used in healthcare for analyzing sepsis management and emergency care (Sujan et al. 2023; Safi et al. 2023). The strategy has also been effectively applied to improve referral processes in outpatient clinics by mapping important activities and determining factors through first protocols and then staff interviews (Safi et al. 2023). This iterative process ensures that the model appropriately reflects real-world operations (work-as-done), rather than theoretical direction (work-as-imagined).

The FRAM approach splits jobs into hexagon-shaped functions, as seen in Figure 2.3. The six viewpoints of input (I), output (O), precondition (P), resources (R), time (T), and control (C) are used to analyze each function. This analysis shows how changes in one component, such as resource availability or time constraints, affect the system's overall performance (Hollnagel 2017; Clay-Williams, Hounsgaard, and Hollnagel 2015).

The FRAM-as-Imagined model represents the idealized, theoretical workflow as it is outlined in protocols, guidelines, or management expectations. This version provides a structured view of how the system is supposed to function under optimal conditions, without accounting for real-world complexities. In contrast, the FRAM-as-Done model focuses on how processes are carried out in practice. It incorporates the actual variability, workarounds, and interdependencies that arise due to real-world factors such as resource limitations, staff availability, or unexpected delays (Hollnagel 2017; Clay-Williams, Hounsgaard, and Hollnagel 2015).

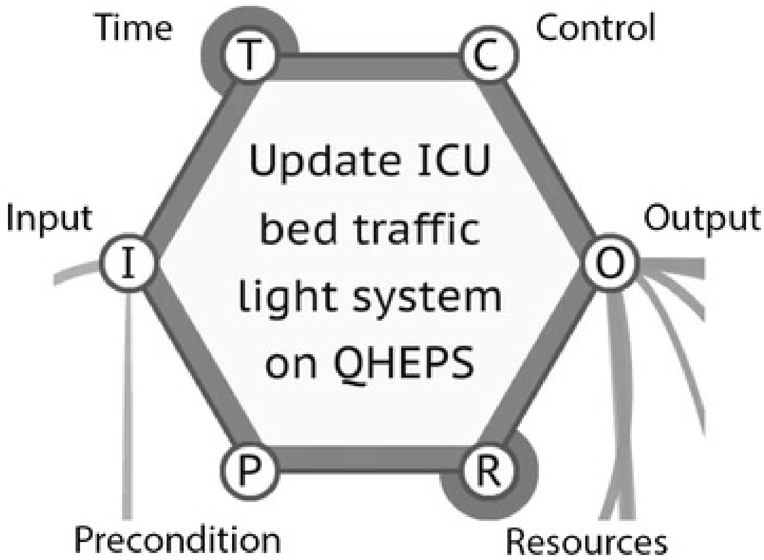

Figure 2.3: FRAM Function (Clay-Williams, Hounsgaard, and Hollnagel 2015)

The difference between these models often reveals critical gaps or inefficiencies. For instance, in healthcare, discrepancies may arise from miscommunication, incomplete guidelines, or variability in patient readiness. By comparing these two models, it becomes possible to identify areas where processes deviate from their idealized version and where interventions are most needed.

By first examining the FRAM-as-Imagined and then further analyzing the FRAM-as-Done, this study used FRAM at Vitalys Clinic to pinpoint areas of unpredictability in the clinic's planning and scheduling processes. Finally, by identifying these critical locations, the research will use data to explore and develop a flexible planning model that adapts to fluctuation while maintaining operational efficiency. By understanding how normal activities interact and pinpointing areas for improvement, research using FRAM has shown its effectiveness in boosting system resilience (Hollnagel 2017). By first examining the FRAM-as-Imagined through protocols and then further analyzing the FRAM-as-Done through interviews, this study used FRAM at Vitalys Clinic to pinpoint areas of unpredictability and variability in the clinic's planning and scheduling processes.

2.3.3. Data and Artificial Intelligence Methods for Planning Efficiency

Data analytics and AI are increasingly applied in healthcare to improve scheduling and patient flow management. This study uses techniques including time series analysis, median delay analysis, grouping-based planning models, and artificial intelligence (AI) for predictive modeling— some of which have been successfully used in healthcare settings in previous research.

Time series analysis is frequently used to identify trends and variations in patient visits, operations, and utilization of resources across time. Its efficacy in anticipating patient demand and streamlining appointment scheduling has been shown in earlier studies. For example, time series analysis was utilized to anticipate short-term patient flow in emergency departments by (Kadri et al. 2014) and to predict obstetric patient flow by (H. Li et al. 2021). Trends in appointments visits and surgeries at Vitalys Clinic during the two-year period from 2022 to 2024 are investigated in this study using time series analysis.

As we have mentioned before, inefficient healthcare systems can lead to longer wait times and reduced patient satisfaction due to delays between appointments and surgeries. This study uses median delay analysis to examine the interval of appointments between initial screening and surgery, identifying bottlenecks and specialties most affected by delays to enable targeted improvements in patient flow. For instance, models like those developed by (Naderi et al. 2021) and (Ogulata and Erol 2003) have optimized operating room scheduling through logic-based frameworks and hierarchical programming, balancing workloads and reducing wait times. The delay analysis provides insights into where and for which patients delays are most prevalent, enabling proactive planning and intervention.

Artificial intelligence, particularly predictive modeling, has become a valuable tool in healthcare for optimizing scheduling and managing resources. Techniques like machine learning, decision trees, and neural networks have been used to predict surgery delays, patient demand, and length of stay, improving preoperative planning and resource utiliziation (Lopez et al. 2022; Ramkumar et al. 2019). At Vitalys Clinic, predictive models using appointment data could identify factors contributing to delays, such as patient demographics and appointment history (Sapir-Pichhadze and Kaplan 2020). By anticipating potential delays, planners can make real-time adjustments to schedules, reducing bottlenecks and ensuring more efficient use of resources. This study evaluates the predictive capabilities of these AI models to determine their potential value in improving scheduling and patient flow management.

3 Thesis Structure

This thesis is structured into five parts, each addressing specific subquestions aimed at optimizing planning processes at the Vitalys Bariatric Clinic.

Part 1: Protocol Analysis The first part examines the current workflow at Vitalys Clinic by developing a FRAM-as-imagined model. This section maps the planning processes, visualizes staff roles, and analyzes activity timelines to understand the planning process and identify sources of variability that may impact efficiency.

Subquestion: What are the key steps in the current workflow at the Vitalys Bariatric Clinic and which contribute most to variability and inefficiencies in the workflow?

Part 2: Stakeholder Analysis Using semi-structured interviews with staff, this section constructs a FRAM-as-done model to capture real-world insights into variability points. It also identifies key drivers and barriers that influence the planning process.

Subquestions: How does variability affect overall workflow and scheduling efficiency, according to stakeholder feedback? What drivers and barriers are identified by stakeholders?

Part 3: Screening Phase Data Analysis This part focuses on analyzing preoperative data to uncover patterns and trends contributing to patient flow variability. The objective is to leverage these insights to optimize the pre-surgical screening process, minimize delays, and improve scheduling efficiency.

Subquestion: What data can be used to analyze the plannings process, and what patterns or trends affecting patient flow can be identified?

Part 4: OR Data Analysis and Planning Model Building on insights from the previous sections, this part looks at the OR Phase to ultimately develop a data-driven planning model. The model aims to optimize scheduling six weeks in advance, enhance OR utilization, and reduce delays, enabling better decision-making and capacity management.

Subquestion: How can data-driven insights be applied to improve the efficiency of planning processes and be integrated into a planning model to improve the planning processes at Vitalys Clinic?

Part 5: Predictive Analysis The final part evaluates the potential of predictive modeling to refine scheduling processes. This section examines the potential application of machine learning techniques to predict delays and optimize scheduling efficiency, assessing how predictive analytics can complement traditional scheduling methods.

Subquestion: Are there predictive possibilities within Vitalys Clinic that can be used to enhance the planning process?

4

Part 1: Protocol Analysis

This chapter will analyze workflows using information and protocols provided by Rijnstate Hospital and Vitalys Clinic to completely understand the planning process. The goal is to create a ”work-as-imagined” model of both the surgical and medical pathway, that accurately reflects the clinic’s expected procedures, as explained in chapter 2. At this moment, we have not discovered where or what the primary causes of variability are, nor have we recognized the specific operational issues. This section also includes a review of the workflow to learn about the planning process and uncover inefficiencies and unpredictability based on the FRAM. This will help us better understand and verify them when creating the FRAM-as-Done during interviews.

4.1. Objectives & Subquestions

The primary objective of this section is to conduct an analysis of the current workflow at Vitalys Clinic using the Functional Resonance Analysis Method (FRAM) to identify critical points of variability and inefficiencies within the planning processes of both the Surgical and Medical Pathways. The processes at Vitalys Clinic operate as complex systems subject to frequent changes and adaptations—characteristics often observed in healthcare settings (Cahill et al. 2010). In such environments, a significant gap often exists between ’work-as-imagined’ (WAI)—the protocols and guidelines established by planners—and ’work-as-done’ (WAD), or the actual practices of frontline workers such as surgeons, nurses, and administrative staff (Hollnagel 2017; Clay-Williams, Hounsgaard, and Hollnagel 2015). The FRAM-as-imagined model will help map and visualize Vitalys Clinic’s current planning workflows. By analyzing this model, we aim to understand how the workflow operates as imagined and identify initial points of variability. To address the objective, the following subquestion will be answered during this part:

- What are the key steps in the current workflow, and which contribute most to inefficiencies?

4.2. Method

In this section, we first analyzed the available documents and protocols at Vitalys Clinic to build a foundational understanding of the current workflow. The subsequent step involved developing the FRAM-as-Imagined model to investigate the planning process of the bariatric and medical pathways.

4.2.1. Document Analysis

The document analysis entailed reviewing key protocols provided by Vitalys Clinic to map out the workflow and treatment paths within their outpatient clinic. Both the bariatric surgical and medical pathways were examined through the lens of the Hierarchical Planning and Control Framework, which breaks down the planning process into distinct parts, as outlined in Chapter 2. This framework helped us focus on offline operational planning—the day-to-day activities related to the planning of both pathways

As shown in Figure 2.2 of the chapter, the analysis highlighted the need for planning documents related to individual treatments, particularly those involving appointment scheduling and workforce allocation

(Hans, Van Houdenhoven, and Hulshof 2011). The FRAM model was therefore based on three central protocols provided by Rijnstate and further informed by initial meetings with Vitalys Clinic management. The protocols, accessed from the hospital database, include:

1. Screening of bariatric patients
2. Pre- and post-operative group planning for bariatric patients
3. OR planning protocol for Vitalys
4. Medicine Traject

4.2.2. FRAM Model

As previously discussed, Vitalys Clinic provides two primary treatment pathways for bariatric patients: medical and surgical. To capture the intricacies of these workflows, we developed a FRAM-as-Imagined (Functional Resonance Analysis Method) model for each pathway using protocols provided by the clinic. The FRAM-as-Imagined model was created using a specialized tool designed for structured mapping of system functions (Zerprize 2024). With this tool, we were able to build out the functions within the FRAM models. For clarity and ease of interpretation, we structured the models with roles involved in the planning process on the y-axis and the timeline of activities, organized by phases, on the x-axis.

The y-axis represents the diverse stakeholders participating in the treatment process, which includes surgeons, nurses, the administrative office, OR planners, and group planners. Each role plays a crucial part at various stages of the patient's journey, contributing to the system's overall functionality. Meanwhile, the x-axis reflects the timeline of activities, divided into key phases that signify different stages of the patient's treatment, spanning from pre-operative assessments to surgery and postoperative care. In the surgical pathway, for example, these phases encompass pre-treatment planning, OR scheduling, and follow-up medical checks.

Through this model, we mapped out interactions among stakeholders and examined the flow of tasks, enabling us to understand the plannings process and identify critical points where variability may arise. The models serve as a foundational tool for identifying areas of variability and gaining insight into the operational dynamics affecting workflow efficiency.

4.3. Results

This section presents findings from a detailed document analysis that examines Vitalys Clinic's documented processes. Using protocols provided by Rijnstate Hospital, we created work-as-imagined diagrams to highlight areas of variability within the clinic's workflow. These protocols outline the key roles and responsibilities across both the bariatric surgical and medical pathways, which are summarized in Figure 4.1.

In this figure, each role is represented by an icon, illustrating its specific responsibilities within the planning processes. The figure helps clarify the distribution of tasks, showing how each role—whether administrative, clinical, or support-focused—interacts within the FRAM model. These interactions highlight the interdependent nature of tasks and roles within Vitalys Clinic's workflow, which are essential to achieving coordinated patient care and operational efficiency.

4.3.1. Bariatric Pathway

The Functional Resonance Analysis Method (FRAM) model for the bariatric surgery pathway at Vitalys Clinic includes five key stages in the planning process. Figure 4.2 illustrates the stages in the bariatric surgery pathway at Vitalys Clinic, covering screening, pre-treatment planning, surgery, and post-treatment and medical follow-up phases. Below, Table 4.1 offers a detailed breakdown of these stages, which are also organized by stakeholder responsibilities in the FRAM-as-Imagined model, as shown along the Y-axis in Figure 4.1.

The planning workflow is closely integrated with the patient journey at Vitalys Clinic. It starts with registration by the receptionists, followed by a screening that is sent to the administrative office. Here, patients are initially scheduled for OR pre-planning. If they are approved, patients proceed to the group planners, who organize either a group or individual treatment program. Group pathways typically consist of five sessions, while individual programs involve four appointments, all aligned with the same

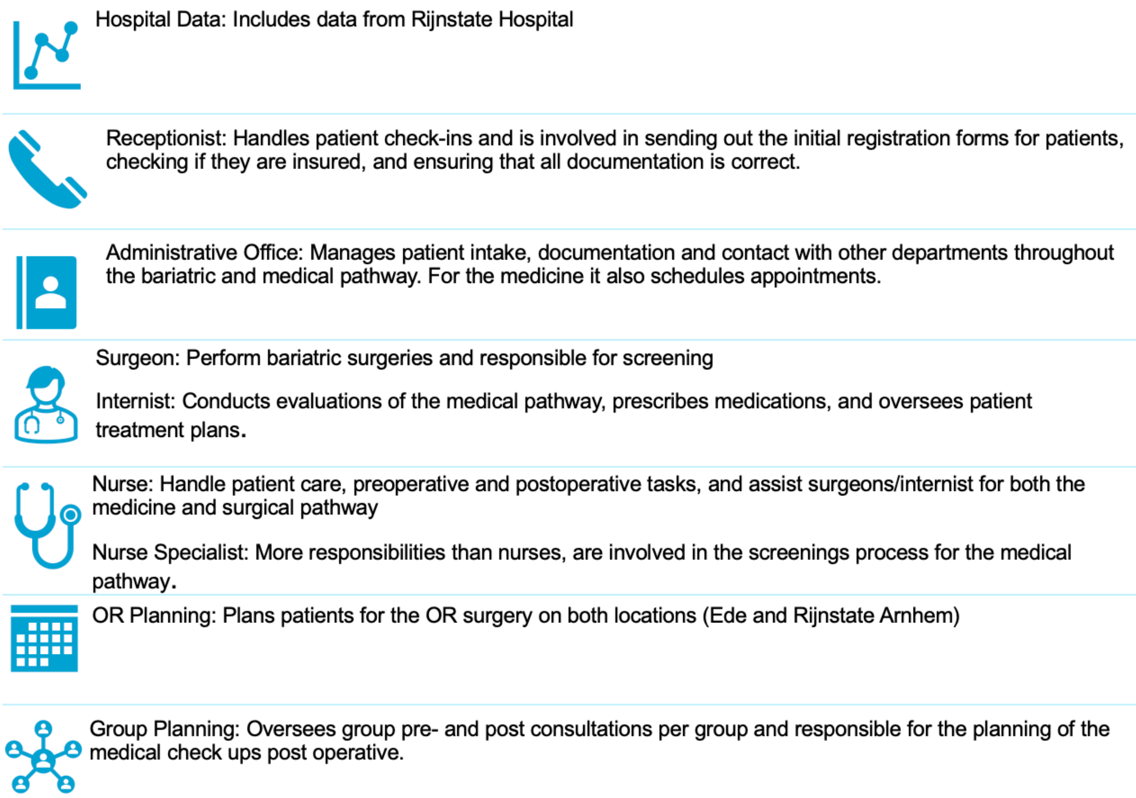

Figure 4.1: Roles and Responsibilities in the Bariatric and Medical Path at Vitalys Clinic

surgery timeline. Additionally, the administrative office oversees the appointments for preoperative screening (POS) or Internal Medicine consultations and can refer patients to other departments if further evaluations are needed.

The pre-treatment phase at Vitalys Clinic begins six weeks before surgery, featuring structured group sessions focused on preparing patients for sustainable weight loss through behavioral therapy and nutrition education. Once patients complete the group sessions and the POS screening, they are transferred to the OR planning team, where their surgeries are scheduled two weeks in advance. This step is conditional on approval during the initial screening and preoperative assessment.

Following surgery, patients enter a follow-up program that supports their lifestyle adjustments, with regular group meetings and one-on-one consultations to monitor their progress. Over time, these check-ups become less frequent, but patients continue to receive support for several years. Medical check-ups, the final phase, may extend up to five years after surgery, ensuring long-term care. Once these steps are completed, the patient's treatment journey at Vitalys Clinic concludes.

In analyzing the FRAM-as-Imagined model, it becomes evident that most activities converge at the OR planning phase. This stage involves high variability due to its dependence on preoperative assessments, staff availability, and patient readiness—all factors subject to changes. OR planning, according to the current protocol, is performed two weeks in advance and requires extensive coordination among surgeons, administrative staff, and nursing teams. Delays in OR planning can lead to extended patient wait times, affecting patient satisfaction and potentially impacting their health. Moreover, inconsistent scheduling practices can result in underutilized operating rooms at times and overburdened schedules at others. This variability also influences the availability of surgical teams, further complicating resource management. Additionally, any delays in the surgery phase impact the scheduling of follow-up appointments, which can postpone essential post-surgery monitoring and elevate the risk of complications for bariatric patients who require close observation.

Another issue with the FRAM-as-Imagined model is the lack of clear documentation regarding the

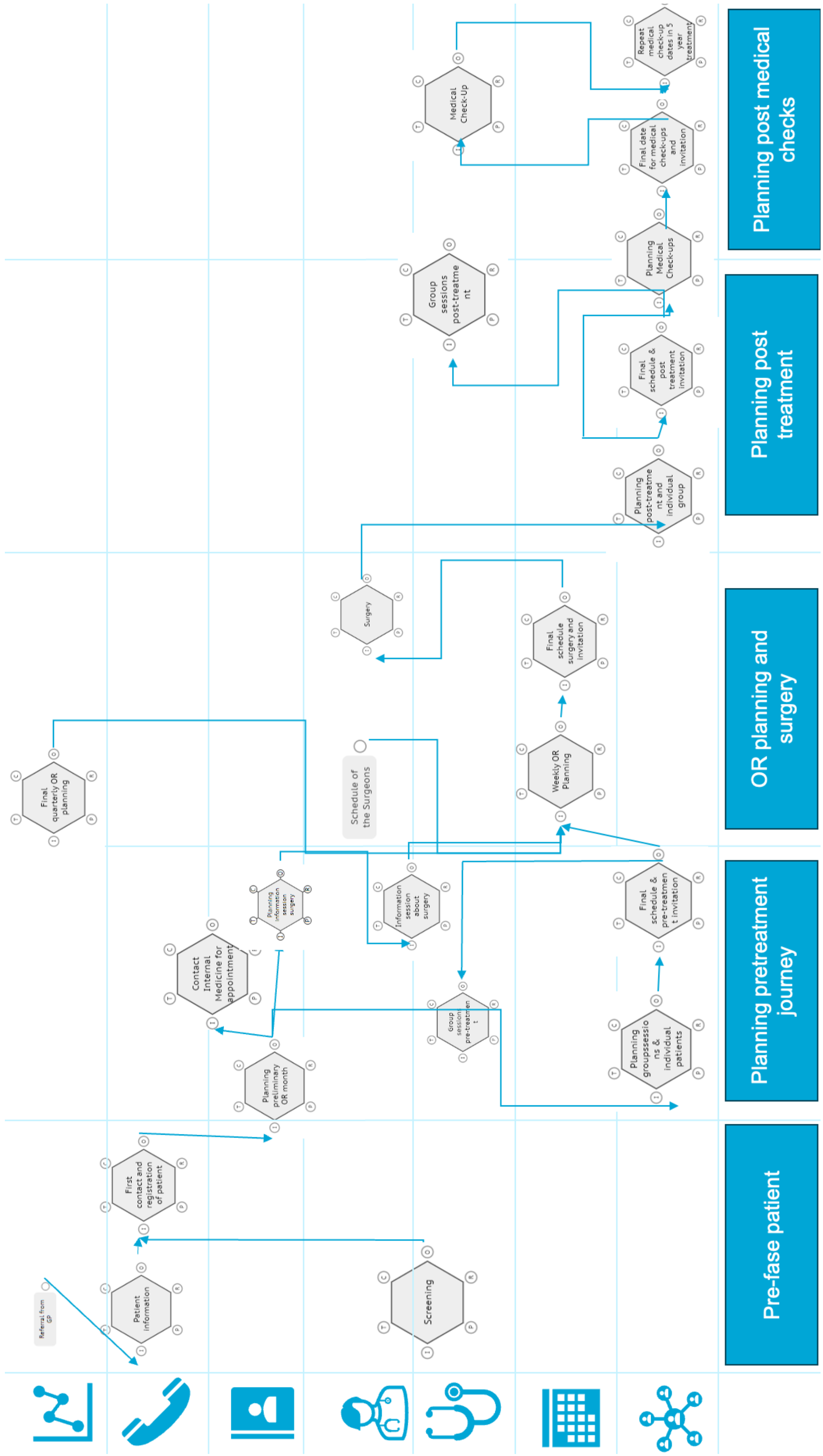

Figure 4.2: FRAM-as-Imagined for Bariatric Pathway

Table 4.1: Phases of the Bariatric Surgical Path at Vitalys Clinic

Phase	Description
Pre-phase Patient	Initial steps involving pre-surgical assessment, patient intake, and screening.
Planning Pre-treatment Journey	Coordination of pre-surgery group or individual treatments and consultations to ensure patient readiness.
OR Planning and Surgery	Scheduling the operating room, preparing the patient, and conducting the surgery.
Planning Post-treatment	Post-surgery follow-up treatments and recovery planning.
Planning Post Medical Checks	Post-operative check-ups to monitor recovery and long-term progress.

management of patients at the screening that might be excluded from the treatment, or who require additional tests or further evaluation during the screening process. The model does not clearly specify what happens during these phases, particularly in terms of how different patient needs are addressed. Ideally, patients who require more extensive testing or follow-up assessments during the screening phase would be identified early, but the protocols do not provide explicit guidance on this. This lack of clarity raises concerns that patients who need further tests may proceed through the workflow without proper handling, leading to inefficiencies in scheduling.

Additionally, a large number of stakeholders, including receptionists, administrative staff, nurses, surgeons, and group plan- ners, contribute to the process. The involvement of multiple roles adds complexity and variability, as each stakeholder has distinct responsibilities, and any delay or miscommunication between them can disrupt the process. In summary, the variability in the pre-surgery phases—such as patient screening, and OR planning—has a significant impact on the clinic's ability to manage its workflow effectively. Variability in the OR planning phase, in particular, affects the accuracy of surgery scheduling and overall resource management, potentially disrupting the continuity of patient care. Addressing these areas of variability, especially within OR planning, would help Vitalys Clinic minimize waiting times, optimize resource allocation, and improve the overall flow of patient care.

4.3.2. Medical Pathway

Starting when a patient is referred by their general practitioner (GP), the medical pathway at Vitalys Clinic follows a set series of steps to guarantee complete, ongoing treatment. The process begins with an initial assessment covering required blood tests and medication choice. After this evaluation, the patient has appointments for follow-up visits set for 4–6 weeks, 16 weeks, 9 months, and 1 year. After 1.5 years, the approach centers on steadily lowering drug levels, leading to a last visit at the 2-year point. Table 4.2 outlines the two phases in this medical pathway. The roles and responsibilities within this pathway are also presented in Figure 4.3.

Table 4.2: Phases of the Medical Path at Vitalys Clinic

Phase	Description
Pre-phase Patient	Initial steps involving patient referral, intake, and preliminary assessments.
Medicine Path	Structured consultations, medical checks, and periodic follow-ups to monitor medication-based treatment.

Figure 4.3 provides a visual representation of this structured and standardized pathway, reflecting how

the process is designed to function in its optimal state. The FRAM model divides the medical pathway into two main sections: the Pre-phase Patient and the Medicine Path. The pre-phase initiates with a referral from the patient's general practitioner (GP), followed by patient reception and registration. This stage includes key steps like arranging information sessions and collecting essential patient data. Stakeholders such as receptionists and administrative staff play a significant role here, ensuring smooth communication and proper patient registration. This structure creates minimal variability and a highly organized approach to patient intake.

Central to this process is the administrative office, which manages the scheduling of appointments throughout the patient's entire journey. This includes coordinating initial consultations, follow-up medical checks, and the timing of essential assessments like blood tests. By overseeing these scheduling tasks, the administrative office ensures seamless patient flow and efficient appointment coordination across all phases.

Within the Medicine Path section, where consultations, tests, and follow-ups occur, the FRAM model shows a well-streamlined process. The visual representation in the FRAM diagram illustrates the alignment of functions, with each step clearly connected to the next. There is minimal variability, and notably, no feedback loops appear, indicating that the pathway operates smoothly without constant adjustments. The absence of feedback loops supports the stability of this stage, allowing each step to proceed logically without significant disruptions.

Overall, the FRAM-as-Imagined model for the medical pathway at Vitalys Clinic demonstrates a well-organized and stable process with minimal variability. The structured pre-phase enables patients to seamlessly enter the more complex stages of the medical pathway. Once in the Medicine Path, the absence of feedback loops and the clear alignment of steps reflect an efficient and stable process. The administrative office's central role in managing appointments further reinforces the organization and consistency of patient care, helping Vitalys Clinic maintain effective patient flow and deliver reliable care outcomes.

4.4. Preliminary Conclusions

The FRAM-as-Imagined model provides a comprehensive overview of the bariatric surgery and medical pathways at Vitalys Clinic, highlighting both strengths and areas for improvement within these processes. The analysis identifies well-structured components of the pathways, particularly in the pre-treatment and post-treatment stages, where standardized protocols and minimal variability contribute to efficient workflows. However, critical points of variability are evident, especially in the OR planning phase.

The OR planning phase is a significant source of inefficiency due to the convergence of multiple pieces of information and the involvement of various stakeholders. This complexity often results in delays, making this phase a critical target for optimization. Enhancing coordination—particularly by implementing integrated digital tools—could reduce wait times and improve resource management. Additionally, clearly defining screening protocols and efficiently handling patients who require further testing would help prevent unnecessary delays and streamline the workflow.

A related issue is the lack of clear processes for managing patients requiring additional testing or evaluation during the screening phase. The responsibilities of multiple stakeholders overlap during this phase, complicating the maintenance of an organized workflow and creating opportunities for miscommunication and inefficiency. Standardizing processes and communication during screening could mitigate these challenges and support better coordination.

In contrast, the medical pathway exhibits a more linear and predictable sequence of activities. It begins with initial assessments and consultations, followed by well-structured follow-up appointments. The administrative office plays a central role in scheduling, ensuring stability and consistency within this pathway. While some variability arises from factors such as patient responses to treatment and the availability of healthcare professionals for follow-ups, these issues are less pronounced than those in the surgical pathway. Nevertheless, further improvements in resource allocation and coordination could enhance patient care efficiency.

In response to the question, "What are the key steps in the current workflow, and which contribute

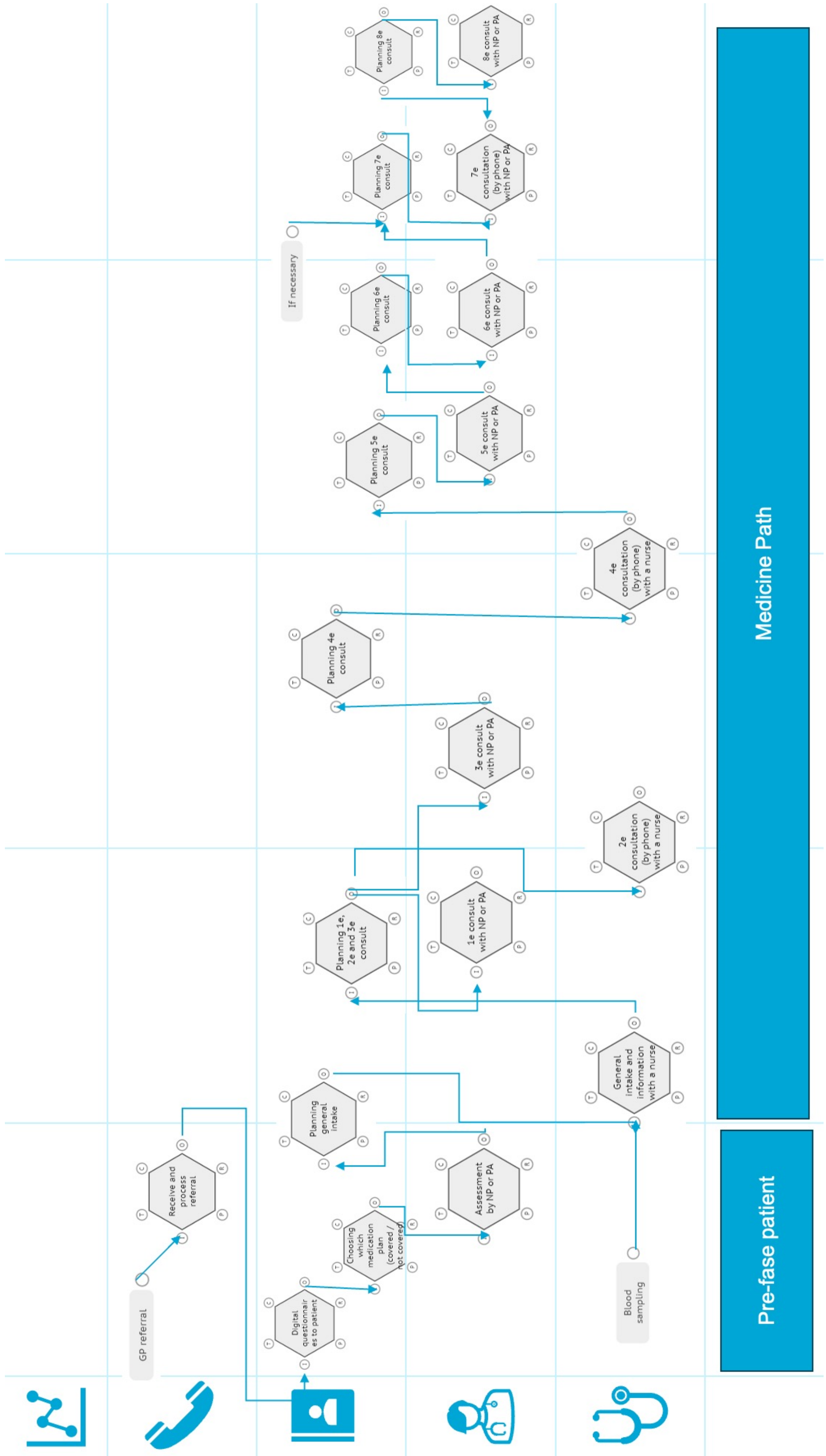

Figure 4.3: FRAM-as-Imagined for Medical Pathway

most to inefficiencies?", the FRAM-as-Imagined model identifies patient screening, registration, and OR planning as the primary contributors to inefficiencies. Streamlining protocols and adopting integrated tools in these areas could significantly improve workflow efficiency. However, these findings must be validated through interviews to create the FRAM-as-Done model.

Moving forward, this research will gather insights from key stakeholders at Vitalys Clinic through interviews. These interviews will be essential for validating the functions within the workflow and assessing the extent of variability experienced in practice. By comparing the FRAM-as-Imagined model with the "work-as-done" practices, the analysis will reveal discrepancies between theoretical protocols and real-world operations. This validation process will provide a deeper understanding of how variability affects each stage and enable the development of more precise recommendations for improving operational efficiency.

5

Part 2: Stakeholder Analysis

This chapter presents the findings from semi-structured interviews conducted to capture stakeholders' insights and experiences with the planning and capacity workflow processes at Vitalys Clinic. These interviews provided firsthand perspectives on the challenges faced by staff within the surgical and medical pathways. By developing the FRAM-as-Done model and analyzing interview data based on factors contributing to variability, drivers and barriers, we identified the most inefficient processes within the planning workflow at Vitalys Clinic. This analysis offers a clear understanding of critical areas that require improvement to enhance operational efficiency and patient flow.

5.1. Objectives & Subquestions

The objective of this section is to develop a "FRAM-as-Done" model that represents the actual workflow at Vitalys Clinic, allowing for a comparison with the "FRAM-as-Imagined" model created earlier. This comparison will help identify critical differences between the intended and real-world processes, highlighting sources of variability and inefficiencies in the planning process. Moreover, through analyzing interview transcriptions and applying coding to identify themes, this section also seeks to uncover key drivers and barriers impacting current planning processes. Ultimately, this approach will reveal how data-driven solutions could address these challenges and support workflow optimization.

This chapter will address the following subquestions:

- How does variability in different phases of the planning process affect overall workflow and scheduling efficiency, according to stakeholder feedback?
- What drivers and barriers are identified by stakeholders?

By answering these questions, we aim to provide a detailed understanding of stakeholder perspectives on the planning process and identify actionable insights for enhancing workflow efficiency.

5.2. Method

This part of the research involved conducting semi-structured interviews with various stakeholders to better understand the Work-as-Done (WAD) model at Vitalys Clinic. This approach follows a methodology similar to that used by Clay-Williams, Hounsgaard, and Hollnagel (2015). Using the same Functional Resonance Analysis Method (FRAM) and FMZ Visualiser, the goal was to create a WAD model that reflects actual practices. This was then compared with the "Work-as-Imagined" (WAI), representing the protocols and intended workflows, to identify discrepancies, potential inefficiencies, and areas of improvement.

In addition to mapping workflow variations, the study focused on analyzing sources of variability, drivers, and barriers within the planning processes at the clinic. Interview responses were coded and categorized to identify recurring themes and patterns, specifically focusing on quotes that highlighted varaibility elements, but also responses based on drivers and barriers. In the following sections, the coding and categorization process is explained in further detail, along with how these elements were incorporated into the WAD model to highlight actionable insights

5.2.1. Semi-Structured Interviews

Semi-structured interviews were conducted with key stakeholders directly involved in the planning processes at Vitalys Clinic. Given the focus on offline operational planning, the study included planners and healthcare providers who engage directly with the scheduling system, with no involvement from higher management. These interviews aimed to examine functions and roles within the bariatric and medical pathways, informed by the FRAM model outlined in Chapter 4.

The interview questions were adapted from previous research that combined FRAM analysis with qualitative interviews, notably the approach used by Clay-Williams, Hounsgaard, and Hollnagel (2015). These questions targeted key functions within the planning system and explored how data could potentially be leveraged to optimize processes. Interviewees were also encouraged to share their perspectives on areas for improvement. A guiding list of interview questions is provided in Appendix A.

In total, nine interviews were conducted with stakeholders involved in the surgical (bariatric) pathway, covering six distinct roles. An additional five interviews were held within the medical pathway, involving four different roles. Interviewees included group planners, OR planners, surgeons, nurse specialists, and medical secretaries. Each interview was categorized by the participant's role within either the surgical or medical pathway to ensure a targeted analysis of inefficiencies and variability specific to each area. A summary of the interviewee roles is presented in Table 5.1.

Table 5.1: Overview of Interviewees for Interview Analysis

Function	Number of Interviewees	Pathway
Group Planners	2	Bariatric
OR Planners	2	Bariatric
Receptionist	1	Bariatric
Medical Secretaries	1	Bariatric
Surgeons	1	Bariatric
Nurse Specialists	2	Bariatric
Internist	1	Medical
Medical Secretaries	2	Medical
Nurse	1	Medical
Nurse Specialists	1	Medical

5.2.2. Labeling Section

To systematically analyze the interview data, a structured approach was used to identify drivers, barriers, and sources of variability within the planning processes. The qualitative data analysis software Atlas.ti (*ATLAS.ti* 2024) was used the labeling and categorization of quotes from the interviews, allowing for an in-depth exploration of the workflows and pinpointing areas where improvements could be implemented.

First of all, the labeling process followed a phase-by-phase breakdown of the planning process as outlined in Chapter 4, with distinct analyses for both the medical and bariatric pathways. In each phase, sources of variability were then identified and grouped into factors, including administrative inefficiencies, resource constraints, scheduling practices, and dynamic human factors like patient behavior and staff availability, as explained in Chapter 2. This categorization provided insights into where significant variability occurred and how it impacted overall workflow at the clinic.

The interview data were then categorized into drivers and barriers specific to each phase of the planning processes. Drivers were defined as factors that support efficient planning and improve operational workflow, while Barriers were categorized as obstacles or inefficiencies that disrupt or hinder the planning process. Following the qualitative analysis, histograms were created to display the frequency and distribution of these drivers and barriers, providing a visual summary of the primary factors impacting the planning phases. A list of labeled quotations related to variabilities, drivers, and barriers is available in Appendix B.

In summary, the structured qualitative analysis conducted in this study, combined with visual and quantitative methods, enabled a thorough evaluation of the planning processes at Vitalys Clinic. These insights are critical for developing strategies aimed at optimizing workflow efficiency, reducing variability, and improving overall patient care.

5.3. Results

The results reveal that the highest variability is found in the bariatric surgical pathway. The ”As-Done” FRAM diagram (Figure 5.2) highlights notable deviations from the ideal process, particularly in the screening and OR planning phases. These stages involve multiple stakeholders—receptionists, administrative staff, and healthcare providers—leading to complex interactions, feedback loops, and, consequently, high variability.

Insights from interviews further emphasize the inefficiencies within these two phases. Figure 5.1 illustrates the variability observed across both the surgical and medical pathways, with the highest variability concentrated in Phase 1 (Screening) and Phase 3 (OR Planning) of the surgical pathway. By contrast, the medical pathway shows significantly less variability. The following sections analyze the factors contributing to high variability in these surgical phases, based on the quatations of the stakeholders.

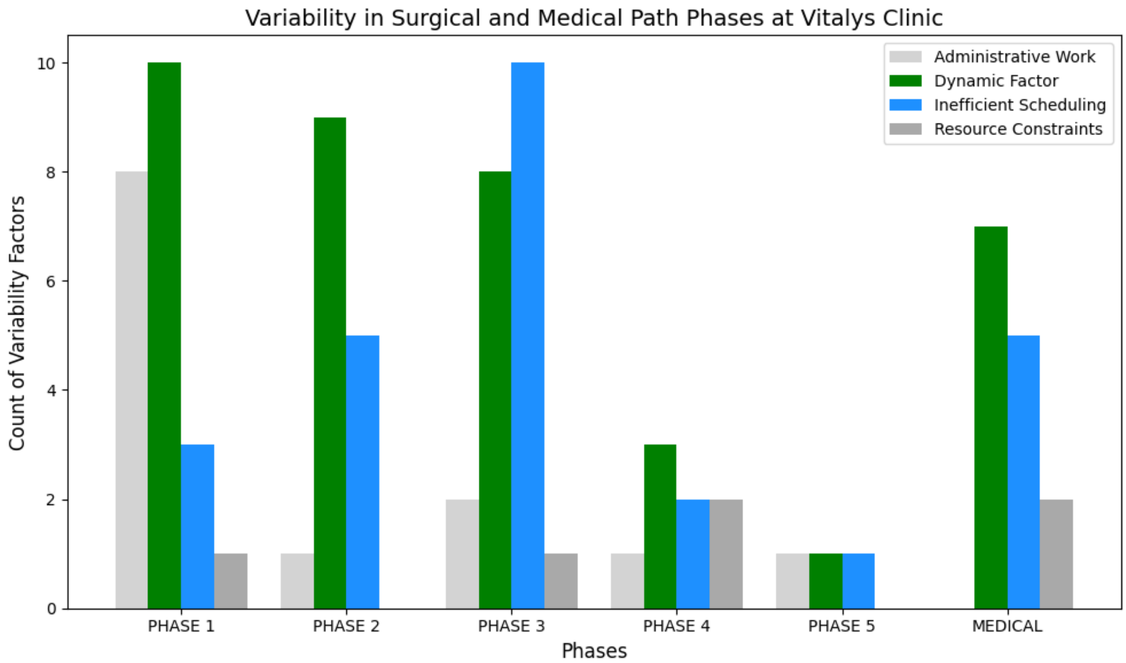

Figure 5.1: Variability of the Surgical and Medical Phases

5.3.1. Bariatric Pathway

The complete bariatric pathway shown in Figure 5.2 highlights the complexity in the first phase, screening. This phase involves several consultations and interactions with patients. The process starts with a referral from the general practitioner, leading to scheduling a screening at Vitalys. During screening, patients are categorized into three groups: red (not approved), orange (temporarily not approved), or green (approved). Patients marked as green move forward to the group planning phase. Those labeled orange need extra assessments, like consultations with a dietician or psychologist, before continuing. Patients who receive a red mark must return at a later time to start the process again. This setup is more complicated than what the protocols initially outlined in the “as-imagined” model. Further details on the screening process will be discussed later.

Following the screening, patients enter the group planning phase, where they are scheduled for pre-

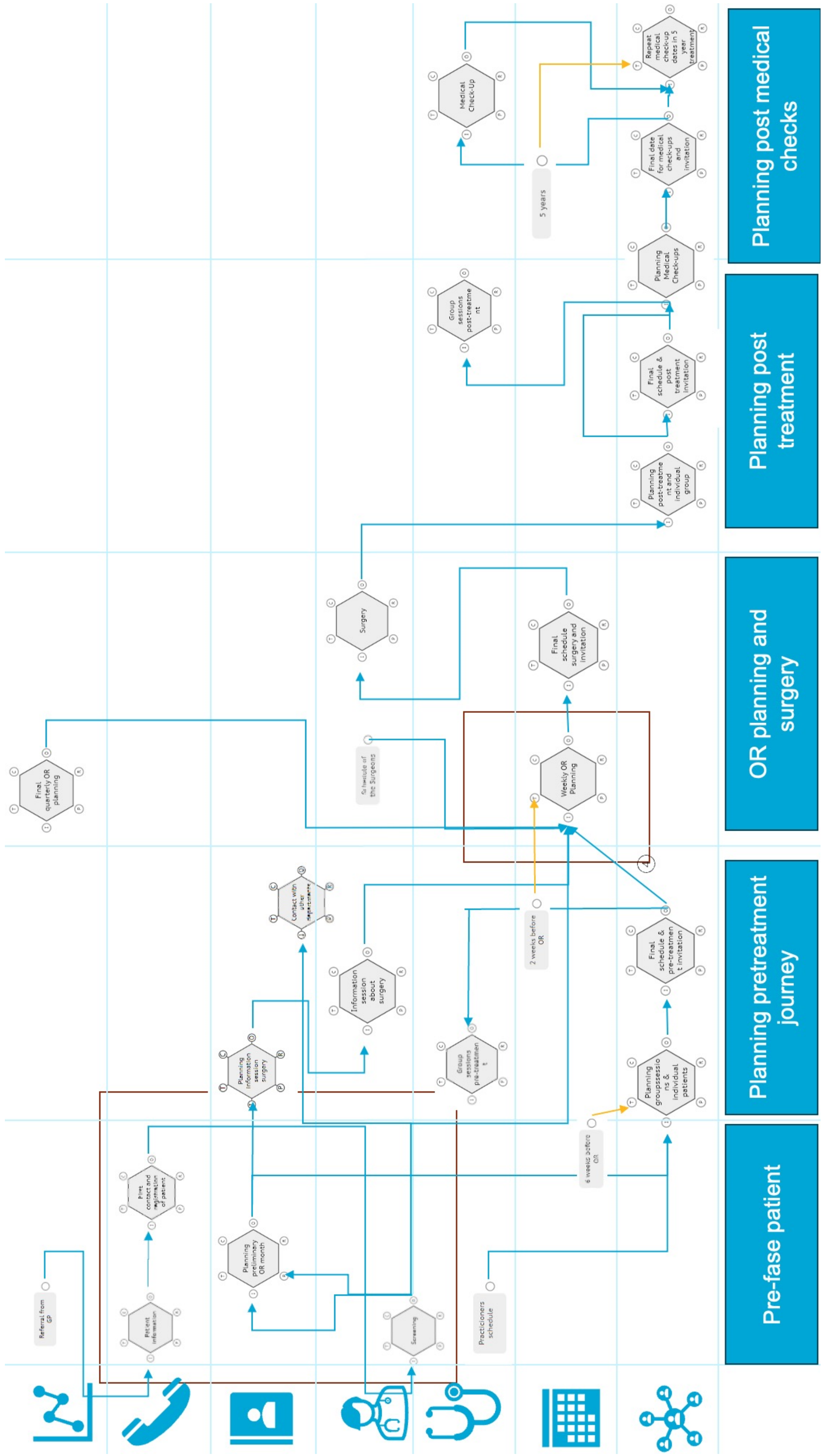

Figure 5.2: FRAM-as-Done for the Bariatric Pathway

operative assessments. The preoperative screening (POS) may happen during these group sessions, which adds complexity because it isn't always completed beforehand. Additionally, some patients might be referred to other departments for further assessments after the POS screening, which can lead to delays in the OR planning process. The most variability occurs in Phase 3 (OR Planning). Here, coordination between surgeons, nurses, and administrative staff is critical to ensure that resources are available and that preparations are completed on time. However, frequent delays and resource shortages disrupt the process, creating challenges in scheduling and managing resources. We will look more closely at the OR planning phase later.

In comparison, the phases related to pre-treatment planning, post-treatment planning, and post-medical check-ups show much less variability. These stages mainly consist of consultations and patient recovery and follow more structured routines with fewer changes. Overall, the ”As-Done” FRAM diagram highlights the complex challenges Vitalys Clinic faces in the bariatric surgical pathway. Phases such as patient preparation, OR planning, and pre-treatment show significant variability, indicating that targeted improvements in these areas could help increase both efficiency and patient care outcomes.

Phase 1: Screening

Phase 1, the screening phase, presents significantly more complexity than initially expected, as shown in Figure 5.3. This phase involves multiple stakeholders—receptionists, administrative staff, surgeons, and nurses—all contributing to the observed variability. The process begins with a referral from the general practitioner, followed by scheduling the screening at Vitalys. During the screening, patients are categorized into three main indicators: red (rejected), orange (temporarily rejected), or green (approved). Patients flagged green can proceed to the group planning phase, while those flagged orange require further assessments with specialists, such as a dietician or psychologist. Red-flagged patients must return for another screening at a later date.

In addition to these categories, an unofficial “yellow” category exists, comprising patients who are technically approved but still need one additional test before proceeding. This added step for yellow-flagged patients can create delays, as they start their treatment but must complete outstanding assessments.

Some green-flagged patients, despite being approved, may also require evaluations from other departments, introducing further delays in the process. One planner highlighted the challenge of this scenario:

> *“They go to another group. And people are not happy with that, right? They finally took the step to join the process, but then they have a 4-month delay because they need to see the internist. Or the MDL with 12 weeks, 11, 12 weeks waiting time. Yes, it is not normal what the waiting times are, right?”*

A surgeon emphasized the importance of having a systematic approach to tracking patients to prevent delays:

> *“Patients with orange flags are ideally monitored closely until they’re ready for surgery, but the current process lacks a systematic approach to managing these flags, which causes delays.”*

The need for accountability is further underscored in cases of flagged patients, where ownership of follow-up actions is essential to ensure patients don’t get lost in the system. As the surgeon explained:

> *“In the screening process, it is crucial that when a red flag arises, someone must take ownership of that problem and ensure it is followed up on.”*

This highlights that systematic tracking of patients is essential during the screening process, especially as flagged cases, such as those in the yellow category, tend to cause delays.

Interviews also revealed that administrative inefficiencies contribute to this variability, as we also see in Figure 5.1. One administrative staff member pointed out the manual documentation process as a particular challenge:

> *“We always print the receptionist’s order and check extra things like the patient’s screening date, insurance, height, and diabetes status.”*

These manual documentation practices are often inconsistent and inefficient, indicating a need for standardized procedures and better coordination between departments. Following the screening, a preliminary OR date is typically scheduled six weeks after the group session, and the patient's information is added to an Excel file for tracking. However, Vitalys Clinic's current system does not yet reliably provide a preliminary OR week or date, highlighting a gap in tracking efficiency.

In summary, Phase 1's variability stems from both administrative and human factors, underscoring the need for systematic tracking and documentation to manage flagged patients efficiently and reduce delays.

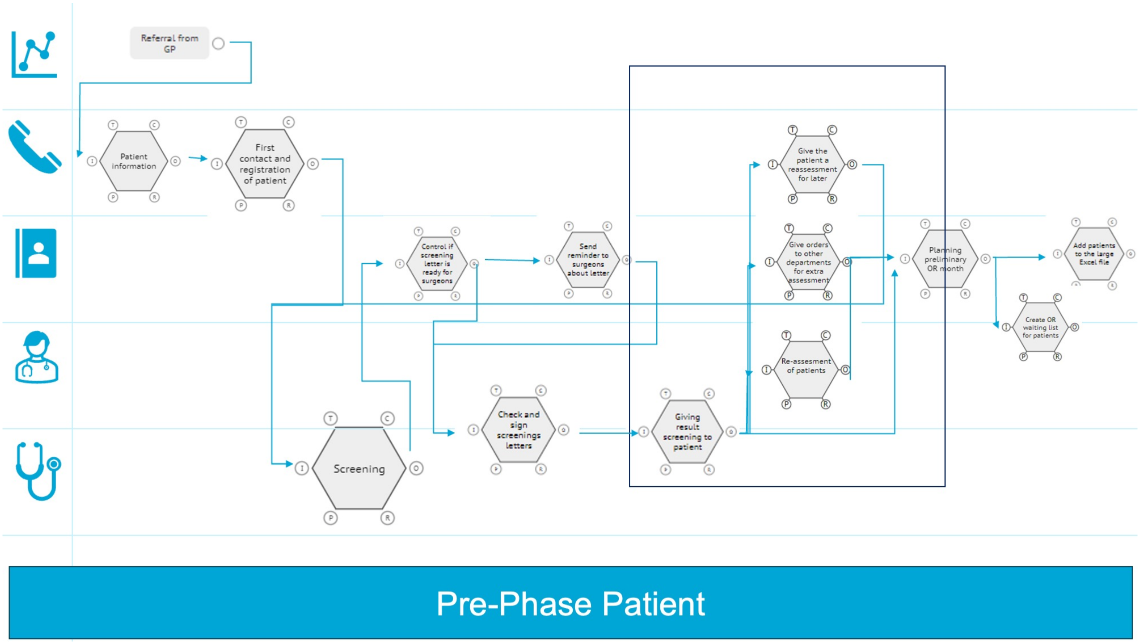

Figure 5.3: Zoom in: FRAM-as-Done for Screening Process Phase

Phase 3: OR Planning

Phase 3, OR Planning, represents the most complex and variable stage in the planning process at Vitalys. As shown in the "As-Done" diagram (Figure 5.4), this phase involves numerous interconnected feedback loops and pathways, all of which contribute to the challenge of finalizing the OR schedule. Effective coordination between surgeons, nurses, administrative staff, and external departments is essential, making this phase highly dynamic and susceptible to disruptions.

As shown in the figure, eight key inputs are needed to complete the OR planning process:

- Final quarterly OR Planning from Rijnstate
- Final schedule of the pre-treatment groups
- Information from nurses regarding screening
- Preoperative Screening (POS) information
- Schedule of surgeons
- Feedback loop: Additional information collection on the patient
- Feedback loop: Last-minute cancellations from patients regarding surgery dates
- Feedback loop: Control by surgeons

These inputs underscore the reliance on five primary information sources, all critical for generating the OR schedule. However, frequent disruptions from feedback loops—such as last-minute cancellations by patients or unexpected schedule adjustments by surgeons—challenge the process. As the interviews revealed, the need for additional information often arises late in the process, contributing to a scheduling horizon limited to just two weeks in advance.

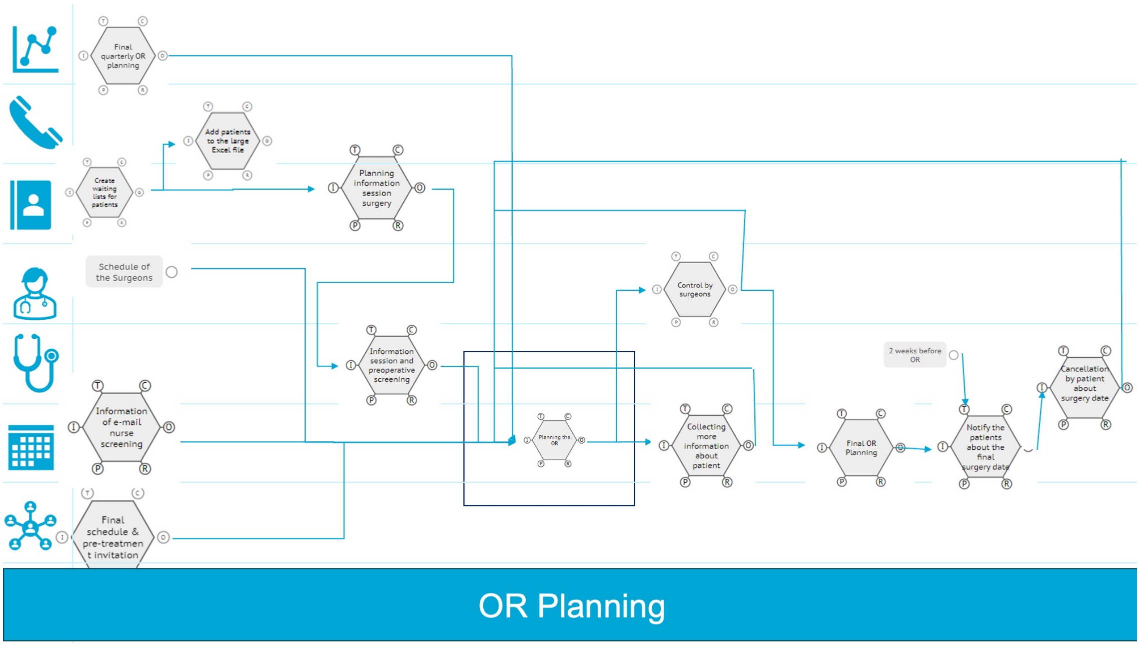

Figure 5.4: Zoom in: FRAM-as-Done for OR Planning Process Phase

Additionally, variability, as analyzed in Figure 5.1,is heavily influenced by dynamic factors and inefficiencies in scheduling practices. An OR planner explained the role of information collection in the planning process, detailing the impact of documentation requirements:

> *"That is actually collecting documentation, what you just said. Yes, yes. Planning. Planning, obviously. Contact with patients, surgeons, management, and scheduling. You need to know who operates on which day and track it carefully."*

Currently, the OR schedule is finalized two weeks in advance. While this provides some flexibility, it leaves little room for accommodating last-minute changes. The planner further explained:

> *"Yes, but basically Tuesday is the day, every Tuesday, that the list is sent for then two weeks later."*

Although this short-term planning period helps manage unanticipated changes—like surgeon availability or patient cancellations—it does not fully address the variability introduced by manual processes. The reliance on traditional methods, such as using Excel files and paperwork, further complicates efficient scheduling. As one planner described:

> *"We use a large Excel file with a list or the OR waiting list per group. We print these lists and manually plan each patient. It's quite old-fashioned."*

In addition to documentation challenges, delays frequently occur when patients must wait for evaluations in other departments, such as an appointment with an internist. This issue, particularly relevant for patients flagged as "yellow," was a recurring topic in interviews. One interviewee noted that, even after the preoperative screening (often referred to as the preoperative information session), some patients may still require additional assessments from other departments, leading to further delays in the process One planner provided insight into the ideal state of this process:

> *"Ideally, patients shouldn't come to the preoperative information session unless they've been fully cleared for surgery. This way, group planning becomes easier, and the process is smoother."*

In an ideal workflow, patients would be fully cleared for surgery before attending the preoperative screening, meaning they are entirely ready to proceed with surgery following this session. This would prevent any last-minute changes or additional appointments necessary after the preoperative screening, ensuring that group planning and OR scheduling phases operate without disruption.

To assess the full impact of these additional steps, it's important to examine the number of patients who still require follow-up visits to other departments after being initially cleared and enrolled in group sessions. Analyzing these cases can provide insights into how these extra steps affect the overall surgical planning and preparation timeline. By reducing these last-minute changes, Vitalys Clinic could streamline the process, enabling more predictable and efficient scheduling for the OR.

5.3.2. Medical Pathway

In contrast to the surgical pathway, the medical pathway exhibits significantly less variability and remains relatively stable. During the interviews, multiple respondents noted that the FRAM-as-Done closely aligns with the FRAM-as-Imagined. This suggests that the process is well-documented in protocols and operates efficiently. However, challenges were observed in managing their capacity, as the program is still in its early stages. Without keepign track of a current waitlist, it has been difficult to estimate the number of patients the clinic can accommodate. One administrative staff member commented:

> *"Planning is the most challenging part, especially managing capacity and avoiding long wait times."*

Some interviewees also pointed out that the manual scheduling of patients could be inefficient. Currently, the eight appointments required for the medical treatment plan (see Figure 4.3 of chapter 4) are scheduled manually. However, this manual system seems to be effective in managing patient flow. One interviewee emphasized:

> *"We still rely on manual planning for flexibility."*

The relatively low variability in the medical pathway, compared to the surgical pathway, indicates that improvements in capacity by starting to keep track of a waiting list, rather than a complete overhaul of the process, may provide the greatest benefit. The stability of this pathway is further evidenced by the minimal need for rescheduling, as one healthcare provider explained:

> *"How often do you adjust or reschedule appointments?" "Almost never. It's quite stable."*

Overall, while the medical phase runs smoothly, the management of increasing patient volumes and wait times remains an area for potential improvement.

5.3.3. Drivers and Barriers of Phase 1 and Phase 3

The interviews not only revealed significant variability within the bariatric surgical pathway but also shed light on key drivers and barriers in both Phase 1 (Screening) and Phase 3 (OR Planning). Understanding these factors is essential to assessing the perspectives of Vitalys Clinic's key stakeholders and uncovering opportunities for process improvement. As shown in Figure 5.5, the findings highlight major challenges and pinpoint areas within current processes where optimization could yield substantial benefits. Selected quotations illustrating these challenges are included in the following sections, with a list of identified drivers and barriers available in Appendix B.

Phase 1 is primarily administrative, and the most significant barrier identified in this phase is the burden of administrative work. This issue recurred frequently in the interviews, reflecting inefficiencies in how patient data is managed and processed. As one surgeon highlighted:

> *"As surgeons, we are involved in the preoperative process by reviewing the screening letters. We assess the screening letters and decide if we agree with the screening. I have to say that it is quite difficult to make a judgment without having seen the patient personally."*

The reliance on manual documentation and the use of Excel files to manage patient records was consistently noted as a barrier to efficiency. An administrative staff member expressed frustration with the current system, emphasizing:

> *"I find the total file in Excel often causes many problems."*

Other barriers in Phase 1 include additional consultations, lack of documentation, and delays in the process, though these are less frequent than the administrative challenges. Inefficiencies in handling patient information and scheduling create delays, complicating the coordination between departments and adding unnecessary layers of complexity to the screening process. Communication between colleagues was also mentioned as a barrier, suggesting that improving internal coordination could alleviate some of these burdens. This is also in line with what we have noticed before, that many stakeholders are involved in the plannings process.

Phase 3, OR planning, deals with more dynamic operational challenges, with inefficient scheduling emerging as the most prominent barrier. The complexity of coordinating surgery schedules, managing last-minute changes, and dealing with resource shortages creates substantial variability and disruptions in this phase. One OR planner noted:

> *"I think we generally plan well ahead, but it would help if an operation date is known earlier so we can prescribe medication earlier for admission."*

The challenge of scheduling operations, often just two weeks in advance, contributes to late operating dates and waiting lists. These barriers affect the entire workflow, making it difficult for staff to prepare adequately and for patients to plan their treatments around their personal lives. A patient highlighted the inconvenience caused by last-minute scheduling:

> *"You get called 14 days in advance on a Tuesday for the surgery date two weeks later. Yes, I find that so... At some point, it is inconvenient, and then work or with children."*

Additionally, communication between colleagues reappeared as a barrier in Phase 3, indicating that better coordination among surgeons, nurses, and administrative staff could help streamline the OR planning process.

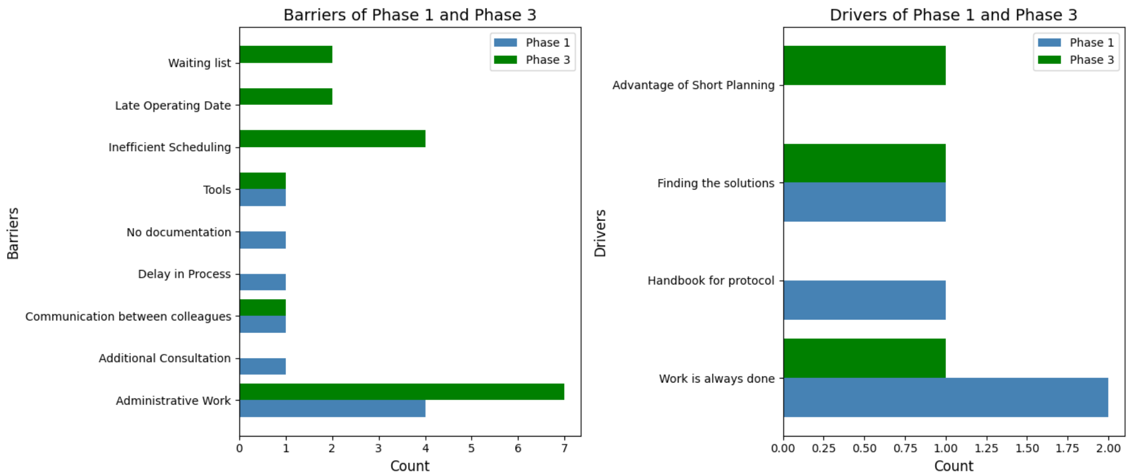

Figure 5.5: Drivers and Barriers in Phases 1 and 3 of the Surgical Pathway

In Phase 1, a notable driver is that work is always done, indicating a strong sense of reliability within the administrative staff despite the barriers faced. The *Handbook for protocol* also plays a role in maintaining structure and ensuring tasks are completed according to standards, although it does not eliminate the inefficiencies in the process. As one respondent noted:

> *"It runs smoothly, it seems everyone is clear on the steps."*

In Phase 3, the drivers are more operational. *Finding the solutions* and the *advantage of short planning* stand out as positive factors in this phase. This highlights the resilience of staff in addressing scheduling complexities and reacting quickly to changes, even though this phase is prone to inefficiencies like inefficient scheduling. Despite the barriers, the ability to adapt to last-minute changes helps keep the process moving forward. One planner mentioned the benefits of short planning:

> *"And how often do you change or adjust the planning? So actually if people drop out? Yes, in principle, never because it is 2 weeks in advance. In principle, it just stays like that."*

Comparatively, Phase 1 is more dependent on established protocols and administrative reliability, whereas Phase 3 is driven by problem-solving and adaptability in managing complex scheduling dynamics. What becomes evident from these findings is that stakeholders at Vitalys Clinic are open to change. Both the administrative and plannings staff express a willingness to adopt more efficient, data-driven tools to enhance scheduling and reduce variability. This suggests that while current processes face significant challenges, there is a strong foundation for improvement through targeted interventions.

5.4. Preliminary Conclusions

The semi-structured interviews and analysis conducted at Vitalys Clinic have uncovered significant inefficiencies and variability, revealing that the FRAM-as-Done, particularly in the screening and OR planning phases, differs notably from the FRAM-as-Imagined. Addressing these issues will require a multifaceted approach focused on improving documentation practices and integrating data analytics to streamline workflows. These efforts aim to enhance the clinic's operational efficiency and patient care outcomes.

One of the primary issues identified is the lack of detailed and standardized documentation during the preoperative screening phase. This problem, initially identified in the previous chapter, was further confirmed during interviews. The documentation gap disrupts patient intake flow and complicates subsequent steps in the care pathway. Interviewees highlighted variability in how patients are flagged during screening, particularly for those categorized as "yellow" (patients approved but requiring additional tests). Inconsistent handling of these "yellow" cases leads to delays and extended waiting times, especially for patients needing further assessments or specialist consultations. The analysis underscores the need for a standardized and automated system to manage patient flags and follow-ups, which would reduce inefficiencies in this critical early phase.

In the OR planning phase, scheduling surgeries well in advance is hindered by the volume of necessary information and reliance on manual processes. Staff members expressed a strong desire to schedule surgeries earlier, but the current system—relying heavily on manual Excel files and frequent last-minute adjustments—limits their ability to do so. These outdated methods contribute to delays in scheduling and postoperative follow-ups. Additionally, delays often arise when 'yellow' patients are referred to other departments, such as MDL, which frequently have long waiting periods. This complicates scheduling and further extends patient wait times.

To address these challenges, improvements in documentation and record-keeping during the screening phase are essential. Capturing detailed patient information, follow-up actions, and specific appointment details would streamline the planning process and support the implementation of a more efficient, data-driven scheduling system. Enhanced documentation would also enable the clinic to predict and plan surgeries more accurately, reducing variability and improving overall operational efficiency.

The data analysis will focus on understanding the screening-to-surgery process and its impact on patient flow, particularly examining the journey of patients flagged as 'yellow' to identify where delays occur and what factors contribute to these inefficiencies. Additionally, the OR planning process will be analyzed to develop a more robust, data-driven scheduling model that minimizes variability and enables better resource allocation. This approach will allow Vitalys Clinic to significantly reduce patient waiting times and enhance surgical planning efficiency.

6
Part 3: Screening Phase Data Analysis

In the initial phase of data analysis, gathering and examining relevant data is critical to understanding and improving the planning process at Vitalys Clinic. This analysis aims to uncover how existing data can be leveraged to streamline workflows and enhance overall efficiency. Insights gathered from stakeholder interviews have highlighted key challenges in managing patient flow along the care pathway, particularly in scheduling surgeries after the initial screening and during OR planning.

One significant issue involves patients who are not immediately approved for surgery or require additional consultations in other departments or outpatient clinics. These cases often lead to delays in the surgical schedule, underscoring the need for better tracking and reintegration processes. Understanding how these patients re-enter the scheduling system at a later stage is vital for minimizing disruptions and inefficiencies. This phase of the analysis will therefore focus on collecting and evaluating appointment data from the screening phase through to surgery. The goal is to identify sources of variability, analyze their impact on patient flow, and develop insights that can inform more effective scheduling and operational strategies at the clinic.

6.1. Objectives & Subquestions

The primary objective of this chapter is to analyze the existing data to create a comprehensive database that allows for evaluating patient flow within this outpatient clinic. This analysis will assess the completeness of the available data, identifying any gaps that may require further data collection. Additionally, the appointments will be reviewed to uncover patterns and trends that may influence the efficiency of the overall patient journey.

This chapter also aims to explore how screening appointments impact scheduling efficiency throughout a patient's surgical journey. By identifying the main factors contributing to delays, the analysis seeks to highlight bottlenecks and assess their effects on OR planning and scheduling. The insights gained from this analysis will serve as the basis for developing actionable solutions to enhance patient management and optimize OR planning.

This analysis is guided by the following subquestion:

- What data can be used to analyze the plannings process, and what patterns or trends affecting patient flow can be identified?

The findings from this chapter will provide a clearer understanding of how the screening phase affects patient flow and scheduling.

6.2. Method

This section details the data collection, preparation, and analysis processes applied to patient appointment and surgery data for the bariatric surgery pathway at Vitalys Clinic, focusing on identifying operational bottlenecks and enhancing patient flow for 2022 and 2023.

6.2.1. Data Collection & Preparation

Data was retrieved from Rijnstate Hospital's HiX system, focusing on patients registered in 2022 or 2023. For 2023, data was tracked through September 11, 2024, with the most recent surgery on September 6, 2024. Data extraction utilized CTCue (CTCue 2024), a tool provided by Vitalys Hospital and Chipsoft, to access details on surgeries, appointments, and demographic information within the HiX system (ChipSoft 2024). Key datasets included the appointments dataset, with information on appointment types (e.g., screenings, consultations) and dates, and the operations dataset, containing surgery details and patient demographics like BMI, age, and gender. In 2023, surgeries were performed at 2 locatiosn, Ede and Arnhem, where the these surgeries were added manually to the data set. Data summaries are presented in Table 6.1, and example of the data set is in Appendix C.

Table 6.1: Data Collection Summary for 2022 and 2023

Year	Number of Patients (Aanmeldformulier)	Number of Appointments
2023	1,342	33,679
2022	1,262	32,951

To prepare the data for analysis, several steps were taken. First, appointments were clustered by *clustering_id*, which allowed the grouping of similar appointment types for clearer analysis. Missing data was handled by manually updating records where possible, for example when information was found in the 'naslag' of patients or excluding them from the analysis if necessary. Very rare appointments were removed to focus on frequently occurring appointments, and duplicates in the database were identified and removed based on *pseudo_id*, *clustering_id*, and appointment date. The datasets were then merged to provide a comprehensive view of each patient's journey from screening to surgery. Post-surgery appointments were excluded as the analysis focused on pre-surgical activities. Patients were categorized based on their screening participation, trajectory participation, and surgery outcomes, with REDO surgeries analyzed separately due to their unique trajectory. Figure 6.1 outlines the data filtering process for 2022 and 2023. This filtering process was being performed in Python (version 3.9.6) and can be found in the Appendix D.

6.2.2. Data Analysis

The analysis focused on identifying patterns and bottlenecks in patient flow during the screening phase. To begin, a time series analysis was conducted for each patient to assess trends and variations in appointments and surgeries from 2022 to 2024. Appointments were grouped by specialty, forming two categories: "Vitalys Core Specialties" (including Psychology, Dietetics, Surgery, Anesthesiology, and Nurse appointments) and "Non-Vitalys Specialties." Monthly counts of appointments and surgeries were compiled to create time series plots, which allowed for comparison across specialties and time, offering insights into fluctuations in patient activity.

Patients were categorized into three groups based on their lead times: Standard (lead time under 49 days), Delayed (lead time over 49 days), and REDO. The 49-day threshold was set according to Vitalys Clinic's clinical guidelines, which anticipate a 6-week timeline for surgery scheduling, with an added buffer week to account for process variability. The delay period was measured from the patient's first group meeting, marking the official "start" of their pathway. Therefore, the screening phase occurs before this threshold. Delays between appointments were calculated using Python's `diff()` function to capture the time difference between consecutive appointments for each patient. Lead times from the initial screening to surgery were also calculated, with patients labeled as "Outliers" if their lead time exceeded 7 weeks.

Further analysis focused on delay patterns. Median lead times were calculated for screening time and group-to-surgery time for each specialty. The total number of patients and the delays between appointments were also analyzed. Each delay was attributed to the specialty waiting on an appointment. Given its robustness against outliers, the median was selected to represent lead times and delays accurately. This analysis was conducted separately for 2022 and 2023 to identify potential trends or changes over time. The analysis broke down results by appointment type, revealing which phases of the patient journey were most susceptible to delays. Bar charts illustrated median lead times from

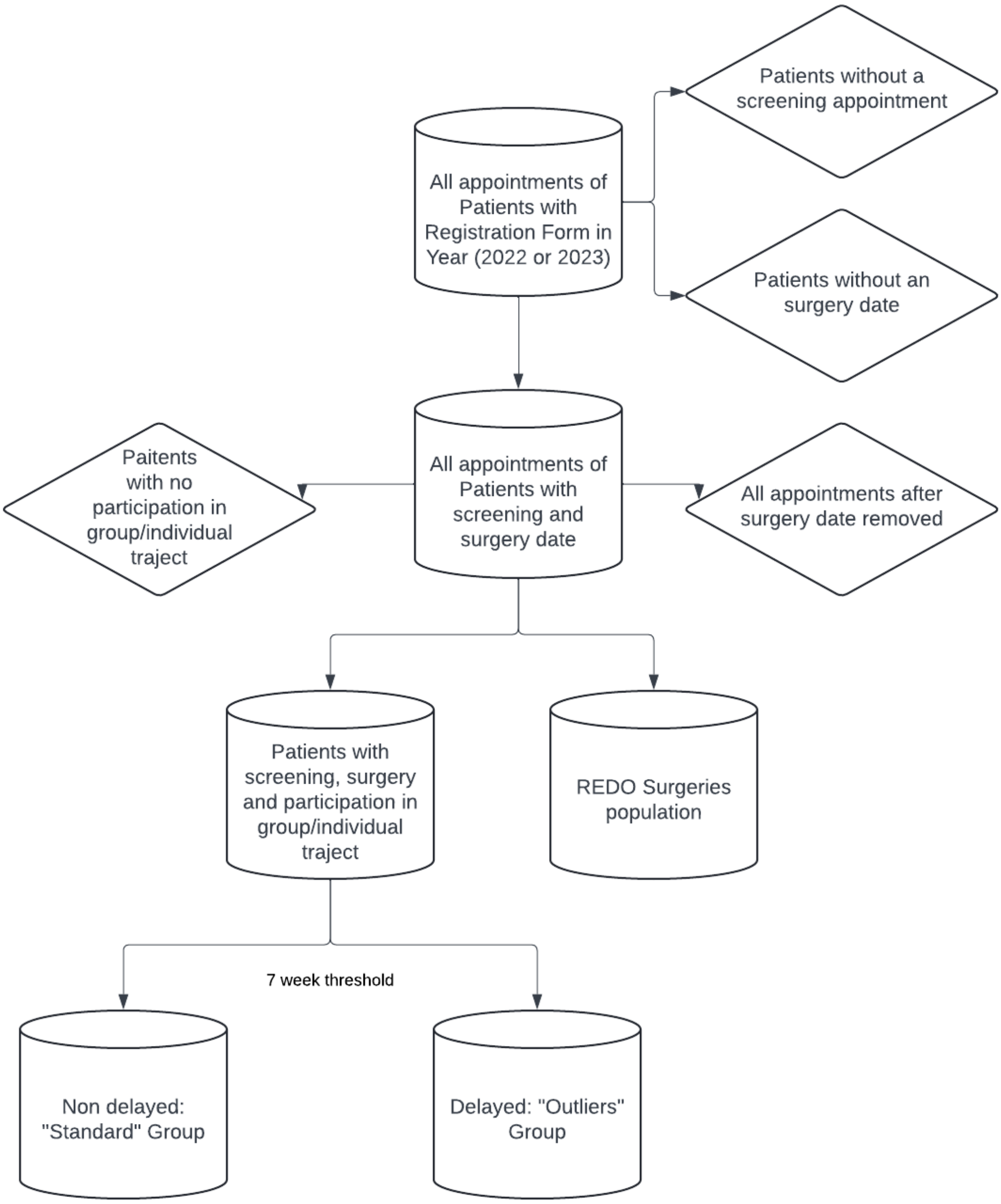

Figure 6.1: Data Filtering Process for 2022 and 2023

screening to surgery and from group meetings to surgery. Additionally, a secondary axis on these charts displayed the total number of patients per specialty, providing a comprehensive perspective on both lead times and patient volume.

Finally, the relationship between delays in appointments and lead times to surgery was examined by specialty. Specialties were grouped, with median lead times to surgery and median delays between appointments calculated for each. The top ten specialties with the longest median delays were then analyzed further. A scatter plot was generated to illustrate the relationship between median lead times and median delays for these top specialties, highlighting which areas most contributed to patient flow variability and where targeted improvements might yield significant results.

All data preparation and analysis were performed using Python (version 3.9.6) with Pandas, Matplotlib, and Seaborn libraries for data manipulation and visualization. The complete Python code for this analysis is included in the appendix. This structured analysis provided critical insights into the relationship between appointment delays and surgical lead times, identifying potential areas for improving patient flow and scheduling efficiency at Vitalys Clinic during the screening process.

6.3. Results

In this section, we present the findings from our data analysis conducted on patient records from 2022 and 2023 at Vitalys Clinic. The analysis focused on preoperative appointments, surgery scheduling, and patient demographic information. During the data collection process, several challenges emerged, particularly related to the completeness and accuracy of surgery data for certain procedures, which required manual adjustments.

6.3.1. Descriptive Statistics

Several issues arose during the data collection procedure for 2022 and 2023, particularly with regard to incomplete or inconsistent surgery data. For example, surgery dates for sleeve procedures were absent, requiring the manual inclusion of records for 482 patients by 2023. Some patient demographic information, such as age and weight, was also missing, possibly due to errors in data entry within the HiX system. We manually inserted this information into HiX after glancing at the written documentation. These challenges underscore the need for enhanced data management in order to ensure future datasets are comprehensive and accurate.

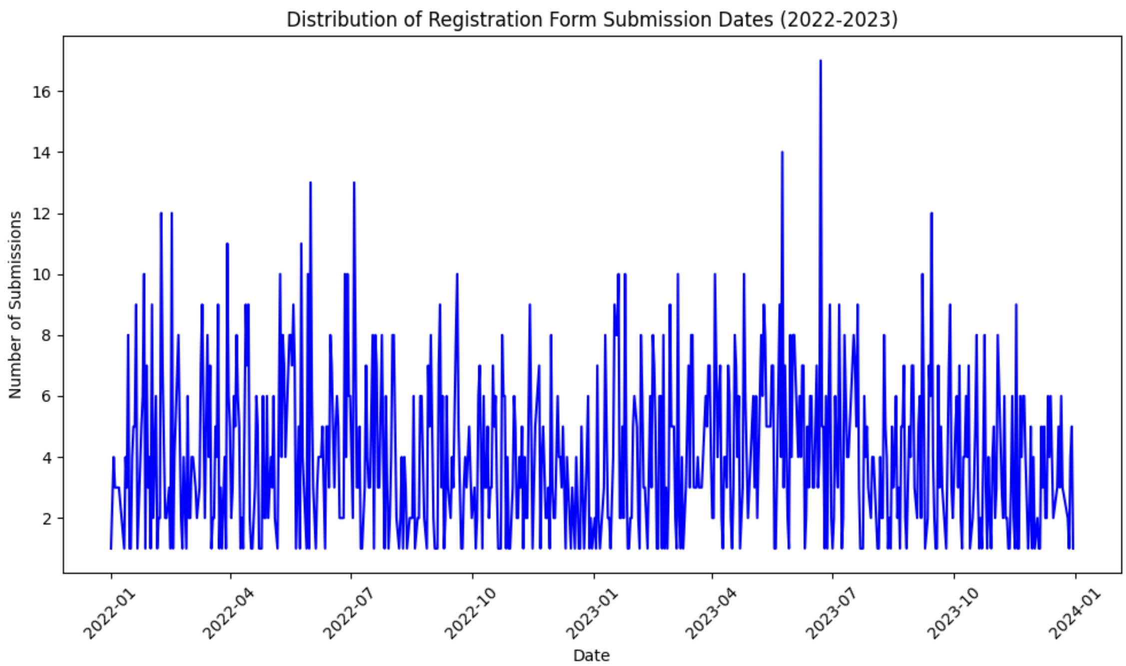

Figure 6.2: Submission dates

The initial 2023 dataset contained 1,342 patient records. Similarly, in 2022, 1,262 patients completed their submission dates. In Figure 6.2, it shows that the submission dates is relatively the same for all the months in 2022 and 2023, with a peak in the summer period of 2023. After filtering for relevant preoperative appointments, 12,071 appointments remained for analysis. Similarly, in 2022, 1,262 patients completed their screening, and after filtering, 9,718 preoperative appointments were retained. Table 6.2 summarizes the descriptive statistics for patients who underwent both screening and surgery.

Table 6.2: Descriptive Statistics for Patients with Screening and Operations

Statistic	2022	2023	Total
Total Patients (%)	1,262 (100%)	1,342 (100%)	2,603 (100%)
Patients with Screening and Operation (%)	868 (68.78%)	986 (73.53%)	1,854 (71.23%)
Patients with REDO (%)	77 (6.10%)	60 (4.47%)	137 (5.26%)
Patients with Screening, Group, and Surgery (%)	603 (47.79%)	922 (68.76%)	1,515 (58.60%)
Patients in the Standard Group (%)	445 (73.80%)	709 (76.90%)	1,147 (75.51%)
Patients in the Outliers Group (%)	158 (26.20%)	213 (23.10%)	368 (24.49%)

Table 6.2 shows that approximately 71% of patients across both years completed their screening and surgery without significant issues, but roughly 60% also completed a group traject. A larger proportion of patients in 2023 completed both the group sessions and surgery (68.76%) compared to 2022 (47.79%), reflecting improvements in workflow efficiency or more complete data in 2023. The exclusion of patients who did not attend group sessions likely contributed to the lower numbers for 2022. Notably, the total number of patients (1,515) is lower than the individual patient counts for 2022 and 2023. This discrepancy is likely due to patient overlap across both years.

Roughly 75% of these patients falling into the ’Standard’ group (those without major delays). Around 5.26% of patients underwent redo surgeries, which are important to consider for understanding overall patient flow. This means that approximately 24.49% of patients across both years experienced delays and were classified in the ’Outliers’ group. While most patients completed their journey within the anticipated time, the presence of outliers suggests areas of inefficiency in the scheduling process, warranting further investigation. Notably, the number of surgeries in 2022 was slightly lower, which may be attributed to disruptions caused by the COVID-19 pandemic. This impact likely affected both patient throughput and the scheduling of surgeries, leading to a slight drop in numbers.

Despite these challenges, the data reveals that Vitalys Clinic achieved a relatively high success rate in scheduling surgeries within the intended time frame, with the majority of patients experiencing smooth transitions from screening to surgery. Further analysis is required to understand the specific factors contributing to the delays observed in the remaining 25% of patients, which will be addressed in subsequent sections, because it still means that for 25% of the patients, there is a delay, and where is this delay then?

6.3.2. Time Series Analysis of Appointments and Surgeries (2022 and 2023)

A time series analysis was conducted to assess the distribution of appointments and surgeries over the period of 2022 to 2024. This analysis aims to highlight key patterns in both Vitalys and Non-Vitalys specialties, as well as the scheduling of surgical procedures. Figure 6.3 illustrates the trends observed in monthly counts for both appointment types and surgeries.

VVitalys appointments fluctuated significantly, peaking in late 2022 and early 2023, then steadily declining until mid-2024. This most likely reflects a spike in patient demand during certain times. Non-Vitalys appointments, on the other hand, were significantly lower and remained quite stable, which is a good indicator because it indicates that patients who required extra diagnostics or referrals outside of the core bariatric specialty remain consistent.

Surgical volumes remained mostly stable overall, with some spikes coinciding with Vitalys appointment peaks. This stability in surgeries, despite swings in appointments, suggests that the number of surgeries each week is well-calibrated to fulfill demand, especially during peak seasons. As these patients move from the preoperative to the surgical stages, it is projected that the surgery count will peak at the same time as Vitalys appointments. It is important to highlight that the lower numbers at the beginning

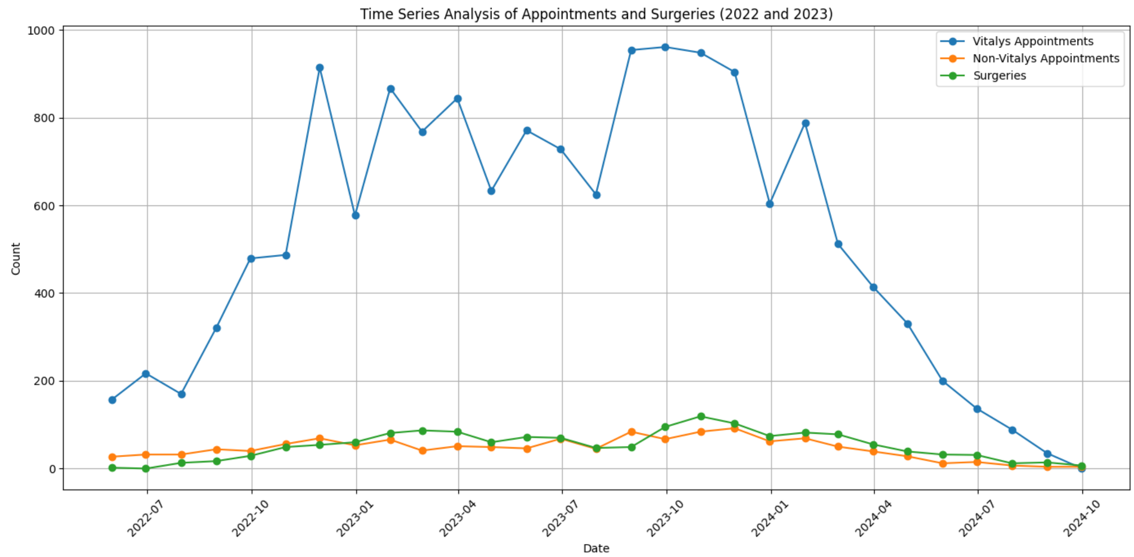

Figure 6.3: Time Series of Appointment and Surgeries for Vitalys Clinic

and end of the timeline are the result of the research focused just on patients examined in 2022 and 2023, ignoring new appointments in 2024.

6.3.3. Lead Time and Appointment count Analysis

Figures 6.4 and 6.5 present the lead times between the initial screening and surgery for patients undergoing bariatric surgery in 2022 and 2023. Patients were categorized into three groups based on their lead times. The first group, REDO patients (in red), consists of individuals who underwent a second surgery following a previous bariatric procedure and generally exhibited the longest lead times. The second group, Delayed patients (in yellow), includes those whose surgery occurred more than 7 weeks (49 days) after their first group appointment, showing lead times often comparable to the REDO group. Lastly, the Standard group (in green), whose surgeries were completed within 7 weeks after their first group appointment, demonstrated the shortest lead times.

While most patients in the Standard group did not experience significant delays, there are outliers in this group with lead times exceeding 200 days from screening to surgery. This indicates that, although the majority of patients progress through the process efficiently, a portion still encounters substantial delays despite initially falling within the standard timeline. A potential reason for these outliers is variability within the screening process itself. Some patients may need additional tests, consultations with other departments, or further medical evaluations before being cleared for surgery, which can extend the waiting time between their initial screening and first group appointment. This variability can push the lead time beyond beyond the 200 days.

We cannot definitively conclude that patients in the Standard group always experience shorter screening phases, there is evidence to suggest that delayed patients often have longer screening periods. It is likely that these extended screening phases contribute to the overall delays experienced by the delayed group. To better understand how the screening phase impacts lead times, further analysis of the appointments during this phase would be valuable. Gaining deeper insights into the screening process could reveal where delays occur and lead to actionable recommendations for optimizing the flow from screening to surgery. Such improvements would help reduce unnecessary delays and ensure more patients undergo surgery within the expected timeframe.

Appointment Count

Based on the graphs in Figure 6.6 and Figure 6.7, we see that the number of appointments does not directly correlate with experiencing delays. Although patients in the Standard group generally have a lower number of appointments compared to the Delayed and REDO groups, the number of appointments alone does not show a decisive pattern. For instance, in both 2022 and 2023, most patients

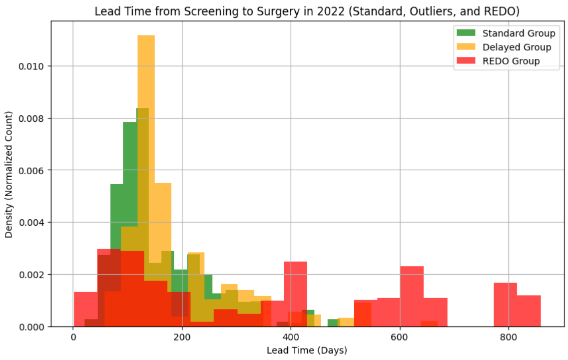

Figure 6.4: Lead time Screening to Surgery 2022

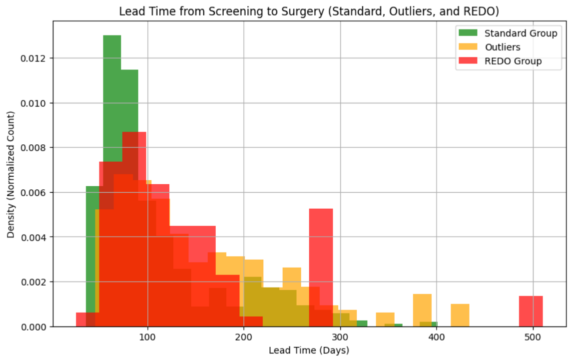

Figure 6.5: Lead time Screening to Surgery 2023

in the Standard group have between 10 and 20 appointments, yet there are patients within this group who still experienced longer lead times. This indicates that a higher number of appointments does not necessarily mean a patient will face delays, and conversely, patients with fewer appointments may still encounter delays.

In the Delayed and REDO groups, we see more variation and a greater number of outliers, suggesting that these patients generally require more care and follow-up. However, even within these groups, the number of appointments cannot definitively indicate whether a delay occurred, as some patients with many appointments still undergo surgery within the expected timeframe. Overall, while the number of appointments in the Standard group tends to be slightly lower, the appointment count alone cannot reliably predict delays in the care pathway.

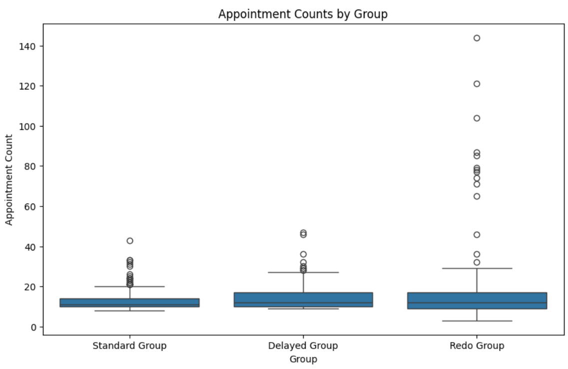

Figure 6.6: Appointment Counts Distribution 2022

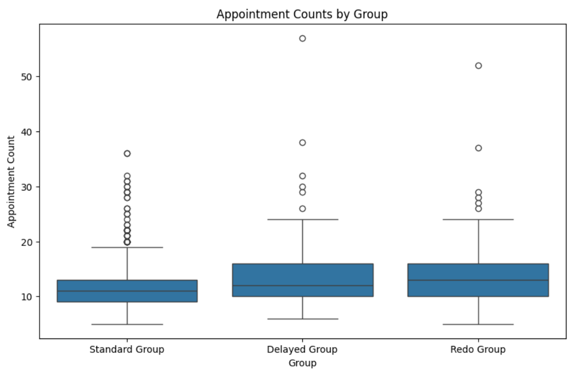

Figure 6.7: Appointment Counts Distribution 2023

6.3.4. Delay Analysis: Screening to Surgery

The patient journey from initial screening to surgery consists of several stages, with two critical components being the screening-to-surgery lead time and the group-to-surgery lead time. These phases are crucial for understanding how efficiently patients progress through the system and for identifying potential bottlenecks. The following analysis, illustrated in Figure 6.8 and Figure 6.5, presents data from 2022, highlighting the median delays across various specialties for the screening-to-surgery and group-to-surgery, the total number of patients per specialty, and the time between appointments. By examining these phases, we aim to identify areas where delays are most prominent and suggest improvements, particularly in specialties where planning and surgical capacities are most affected.

Key Findings from 2022 Data

The figure presented (Figure 6.8) illustrates the lead time from screening to surgery, the total number of patients per specialty, and the median delay between appointments across various departments. This analysis focuses on identifying specialties where patients experience significant delays in the screening process and assessing how this impacts the planning of surgical capacities. Each appointment is tagged with a corresponding specialty.

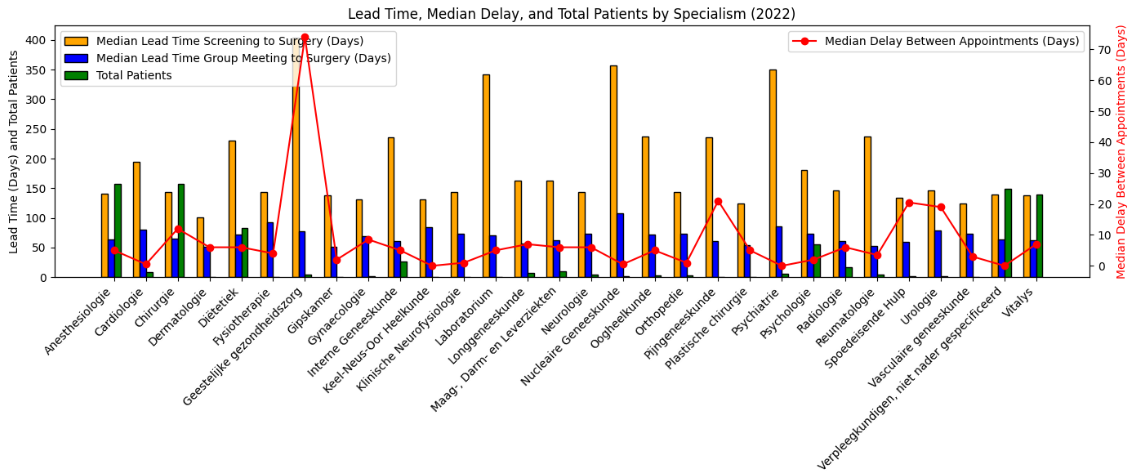

Figure 6.8: Delay Analysis: Screening to Surgery Patients Group 2022

The core specialties with higher patient volumes—such as Anesthesiology (*Anesthesiologie*), Surgery (*Chirurgie*), Dietetics (*Diëtetiek*), Psychology (*Psychologie*), Nursing (*Verpleegkundigen*), and Vitalys—are expected to have these volumes, as they represent mandatory steps in both individual and group treatment pathways. Despite this high volume, the median delays between appointments (represented by the red line in the graph) for these specialties remain relatively low, as does the overall lead time from screening to surgery. This suggests that the core stages of the process are well managed and do not significantly contribute to delays in the overall patient flow. Surgery (*Chirurgie*), however, shows a slightly longer lead time, which might be due to limited availability of consultations with the surgeons themselves.

However, specialties like Dietetics (*Diëtetiek*) and Psychology (*Psychologie*) do show longer screening-to-surgery lead times. This can be attributed to re-evaluation processes where patients may receive an 'orange flag' status, temporarily halting their progression as further assessments are needed. The average screening-to-surgery lead time for Vitalys is approximately 150 days, which we consider a standard benchmark for what is considered a 'normal' timeline.

Among the specialties, one notable outlier is Mental Health Care (*Geestelijke Gezondheidszorg*), which shows the longest median delay between appointments, approximately 70 days. Additionally, the median lead time from screening to surgery extends to nearly 400 days, significantly contributing to overall delays in patient progression. Despite these delays, once patients complete the screening phase, the time from group meetings to surgery is relatively short. A similar trend is observed in Psychiatry (*Psychiatrie*), where the delay between appointments is shorter, likely because Vitalys pre-reserves appointment slots for these patients, ensuring quicker access to care. Implementing similar pre-reservation strategies for Mental Health Care could potentially reduce delays and improve overall patient flow in this specialty.

Internal Medicine (*Interne Geneeskunde*) also shows a longer screening period and relatively higher patient volumes, although still low compared to the core specialties of Vitalys. This extended screening time can be explained by the fact that patients with conditions such as diabetes are referred to this specialty, resulting in a longer, more complex treatment trajectory. Radiologie heeft ook realtieve hogere patienten volume, maar zorgt niet perse voor vertraging zoals we zien in de grafiek.

Specialties such as Laboratory (*Laboratorium*), Nuclear Medicine (*Nucleaire Geneeskunde*), Pain Medicine (*Pijngeneeskunde*), Rheumatology (*Reumatologie*), and Ophthalmology (*Oogheelkunde*) show longer screening-to-surgery times. This may indicate that the diagnostic or evaluation phases within these specialties take longer. However, it is important to note that the actual number of patients in these specialties is relatively low, making these extended lead times less impactful on the overall patient flow. These departments can probably be considered exceptions.

Specialties like Cardiology (*Cardiologie*), Pulmonology (*Longgeneeskunde*), Gastroenterology (*Maag-*

Darm-Leverziekten), and Neurology (*Neurologie*) show moderate screening-to-surgery times with relatively few patients. Thankfully, these specialties do not significantly extend the overall lead times for the majority of patients. However, specialties such as Pain Medicine (*Pijngeneeskunde*), Urology (*Urologie*), and Emergency Care (*Spoedeisende Hulp*) exhibit higher median delays between appointments, indicating that delays within these departments may contribute to extended lead times for patients.

Other departments, such as Dermatology (*Dermatologie*), show average outcomes in terms of lead times and appointment counts. However, it is likely that many of these appointments are not directly related to the bariatric treatment pathway or are scheduled for patients with unrelated health issues. These appointments are unlikely to significantly affect the overall patient flow for bariatric surgery.

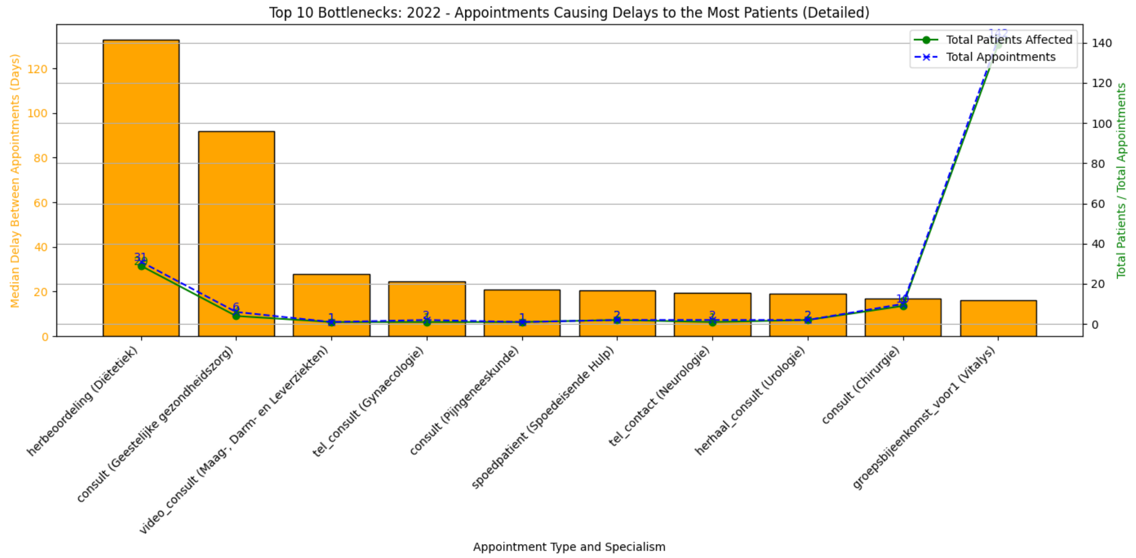

Figure 6.9: Bottlenecks Screening to Surgery 2022

Following the previous analysis of the 2022 screening-to-surgery lead times, Figure 6.9 provides a detailed breakdown of the top 10 bottlenecks—appointments that caused significant delays for the most patients. The median delay between appointments (represented by the orange bars) shows that specific appointment types and specialties contribute to extended lead times, while the green line tracks the number of affected patients. This visualization helps identify which steps in the patient journey require the most attention to reduce delays.

The first major bottleneck identified is Dietetics (*Diëtetiek*), specifically in the re-evaluation (*herbeoordling* process. With a median delay of approximately 140 days between appointments, this step impacts a large number of patients (143), representing a significant bottleneck in the patient flow. The re-evaluation process often leads to 'orange flag' patients, who require additional assessments before they can proceed, thus contributing to the long delay times observed. Another substantial delay occurs in Mental Health Care (*Geestelijke Gezondheidszorg*), where the median delay is around 90 days. This specialty affects fewer patients than Dietetics but still represents a notable barrier, as these consultations often involve complex assessments that can extend the patient's waiting time before surgery.

Specialties such as Gastroenterology (*Maag-, Darm-, en Leverziekten*), Gynecology (*Gynaecologie*), Pain Medicine (*Pijngeneeskunde*), Urology (*Urologie*), and Emergency Care (*Spoedeisende Hulp*) all exhibit median delays ranging up to 20 days for these specific appointment. While these specialties involve essential diagnostic or treatment evaluations, they affect a relatively small number of bariatric patients, suggesting these delays are more likely exceptions rather than widespread issues. Despite the smaller patient volumes, these extended waiting times can still contribute significantly to overall lead times in the patient journey. Even a few patients experiencing delays in these departments can create bottlenecks, prolonging the process from screening to surgery for those affected.

Interestingly, Surgery (*Chirurgie*) consultations show a noticeable median delay, aligning with the earlier observation of longer median delay times between appointments for this specialty. Additionally, group appointments (group sessions 1, *groepsbijeenkomst_voor1*) appear as a bottleneck in the analysis. However, this is understandable, as patients typically experience a short waiting period before their first group session, especially if they still need to complete other appointments beforehand.

Key findings from 2023 data

For the 2023 data (shown in Figure 6.10, the results for Vitalys' core specialties are similar to those of 2022, but with generally shorter lead times to surgery. This improvement suggests that the process from screening to surgery may have become more efficient within Vitalys, or that patients are being placed into groups for surgery more quickly. However, Dietetics (*Diëtetiek*) and Psychology (*Psychologie*) still show slightly elevated screening-to-surgery times, likely due to the reassessment processes involved for some patients. Unlike the previously mentioned specialties, Gynecology (*Gynaecologie*) and Surgery (*Chirurgie*) exhibit more moderate lead times.

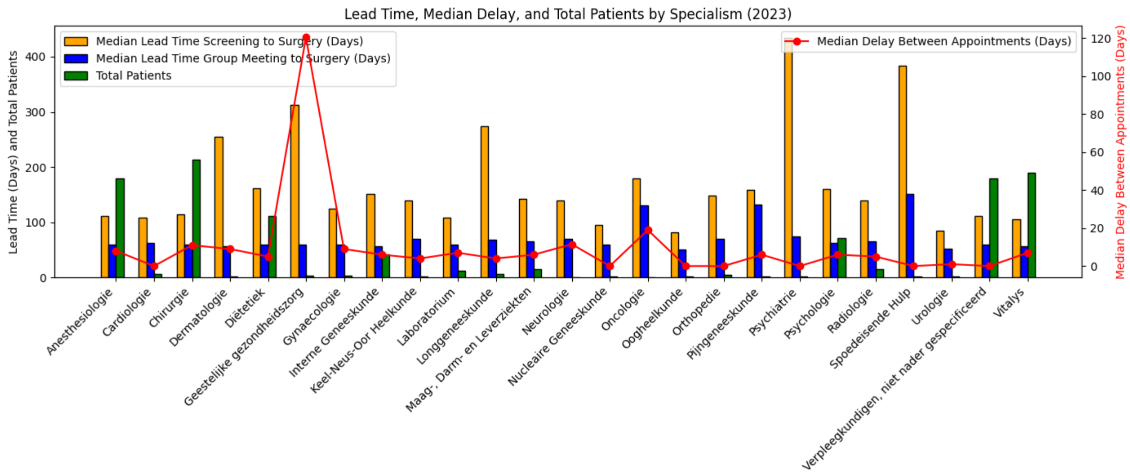

Figure 6.10: Delay Analysis: Screening to Surgery Patients Group 2023

Mental Health Care (*Geestelijke Gezondheidszorg*) again exhibits notably long delays between appointments, reinforcing the issue identified in 2022. This could be attributed to understaffing within the department, which is a known issue in these types of specialties. Psychiatry (*Psychiatrie*) also has longer lead times, but the delay between appointments is shorter than in Mental Health Care, likely due to the pre-reserved slots that Vitalys has for Psychiatry patients, allowing for faster access to appointments.

Similar to 2022, Emergency Care (*Spoedeisende Hulp*) shows high screening-to-surgery times. It is likely that when patients are referred to Emergency Care, there is already a serious underlying issue that will inherently delay surgery. This is further reflected in the longer group appointment-to-surgery times for these patients.

An interesting observation in 2023 is the high lead time for Dermatology (*Dermatologie*), though the number of affected patients is very low, suggesting that these cases are not directly related to the bariatric treatment pathway. Ear, Nose, and Throat (*Keel-Neus-Oorheelkunde*) also shows a slight increase in lead time, which is noteworthy but likely reflects isolated cases.

Internal Medicine (*Interne Geneeskunde*) had more patients in 2023 compared to 2022, likely due to an increase in diabetic patients referred to this department. In 2023, Pulmonology (*Longgeneeskunde*) also showed higher screening times, likely related to the need for additional tests before surgery. Gastroenterology (*Maag-, Darm-, en Leverziekten*) showed slightly elevated patient numbers due to required diagnostic tests, which is expected. While it was anticipated that Gastroenterology would show longer waiting times, the actual delays are relatively moderate. In contrast, Pulmonology shows notably longer delays, indicating that this specialty may require additional focus to address these issues.

Radiology (*Radiologie*) also saw more patients in 2023 but did not contribute significantly to overall delays, showing results similar to those of the Vitalys core group. Oncology (*Oncologie*), however, showed an increase in group-to-surgery time. This is likely an exception, as it could indicate a previously undiagnosed issue discovered during the bariatric pathway. Pain Medicine (*Pijngeneeskunde*) once again appears as a source of delays, with longer group-to-surgery times, signaling an area that requires attention.

Neurology (*Neurologie*) showed delays, but the number of affected patients remains very low, making it less impactful. Orthopedics (*Orthopedie*) also showed long lead times, but these likely represent exceptions rather than a systemic issue. It could be that the patients visiting these departments were for other reasons that for the bariatric surgery.

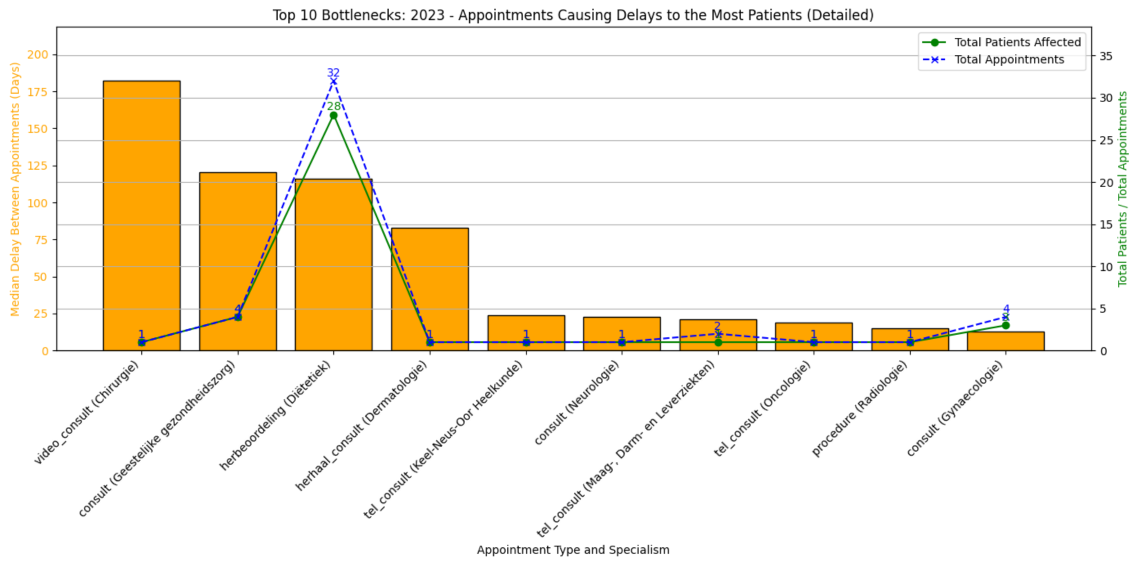

Figure 6.11: Bottlenecks Screening to Surgery 2023

The bottleneck analysis for 2023, as shown in Figure 6.11, highlights key delays across various appointment types and specialties. One of the most notable bottlenecks is the video consult for Surgery (*Chirurgie*), with a median delay of nearly 180 days. This significant delay could be due to limited availability of consultations with surgeons, which seems to be a recurring issue in both 2022 and 2023. The bottleneck here points to a potential capacity limitation within the surgery department that needs to be addressed to improve patient flow. However, compared to Figure 6.10, the median lead time to surgery stays comparable to Vitalys, so this means this appointment was probably exception.

Mental Health Care (*Geestelijke Gezondheidszorg*) again shows considerable delays, with a median delay of around 130 days between appointments. Despite the relatively low number of patients (4), these extended waiting times indicate that there are persistent issues with capacity or scheduling within this department, as also noted in the 2022 analysis. Similarly, Dietetics (*Diëtetiek*), particularly in the reassessment (*herbeoordeling*) process, shows a median delay of over 100 days. This reassessment often involves more complex evaluations, which contribute to the delay.

An interesting finding for 2023 is the high delay for Dermatology (*Dermatologie*), which was not as prominent in 2022. Although the patient count is very low (1), the median delay is still significant, suggesting that this may be an exceptional case rather than a systemic issue. The same can be said for other specialties like ENT (*Keel-Neus-Oorheelkunde*), Neurology (*Neurologie*), and Oncology (*Oncologie*), where delays are seen but with very few patients impacted.

The median delays for specialties like Gastroenterology (*Maag-, Darm-, en Leverziekten*) and Gynecology (*Gynaecologie*) are moderate compared to other specialties. In summary, while some bottlenecks remain consistent with the findings from 2022, such as delays in Surgery, Mental Health Care, and Dietetics, new specialties like Dermatology and ENT have appeared in the 2023 analysis. Although these specialties have a smaller impact due to the low number of affected patients, their presence in the bottleneck analysis highlights the need for continued attention to outlier cases that could still affect the overall patient flow.

6.3.5. Relationship Analysis: Lead Time with Delay Between Appointments

The scatter plot (Figure 6.12) compares the median delay between appointments (Y-axis) and the median lead time from screening to surgery (X-axis) for different specialisms. The purple dots represent the outliers group, while the orange dots represent the total group (both delayed and non-delayed patients) for the years 2022 and 2023. This comparison allows us to observe which specialisms tend to experience more delays and longer lead times in the patient journey. In Table 6.3, the corresponding descriptive for the outliers group are displayed, and Table 6.4 provides the same data for the total group, allowing us to compare these groups.

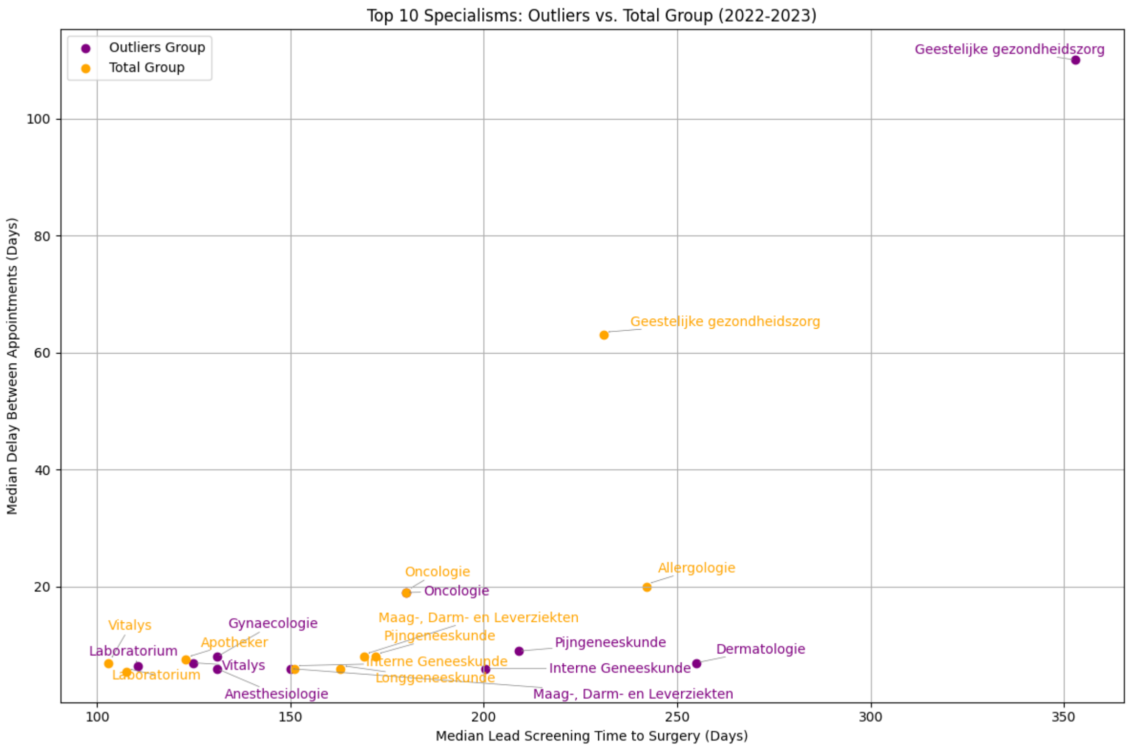

Figure 6.12: Top 10 Specialisms: Median Lead Screening Time to Surgery vs. Delay Between Appointments

In this section, we analyze the delays between appointments and lead times per specialism for both the outliers group and the total group of patients, highlighting the key differences and systemic issues that impact patient progression from screening to surgery. The following figures and tables present the top 10 specialisms that contribute to significant delays in the patient journey, allowing us to identify areas for improvement.

With median delays of 110 and 63 days, respectively, and lead times to surgery of 353 days for outliers and 231 days for the overall group, Mental Healthcare (*Geestelijke gezondheidszorg*) has the longest delays in both groups. Despite the modest sample sizes (8 outliers and 25 patients overall), our results show that mental healthcare is a major barrier, most likely as a result of intricate evaluations and drawn-out follow-ups. Reserving pre-scheduled spaces could be one method to address this and lessen the waits in this approach.

Although oncology (*Oncologie*) also exhibits significant lead times in both groups, only one patient in each group is impacted, indicating that the delays are most likely the result of unusual circumstances rather than systemic problems.

The median lead time for Pain Medicine (*Pijngeneeskunde*) is 172 days for the entire group and 209 days for outliers. The small study size (3 outliers and 9 total patients) suggests a limited impact on average patient timelines, despite the considerable median delays (9 days for outliers and 8 days overall). Nonetheless, pre-reserved appointment scheduling may help to reduce delays for this specialty.

The department with the highest patient load, Vitalys, exhibits comparatively effective scheduling, keeping both groups' median lead times at 125 days and delays at 7 days. However, the significant lead time standard deviation indicates that some patients encounter protracted delays, frequently as a result of necessary departmental visits. In some situations, prolonged delays raise the overall lead time variability throughout the patient group because all patients must complete this phase.

The median lead time for gastroenterology (*Maag-, Darm- en Leverziekten*) is 169 days for the entire group and 150 days for outliers, with 6–8 day waits between visits. Even while these delays are minor, the comparatively long lead times (25 outliers and 80 patients in total) suggest that more pre-operative planning would be necessary, which could lengthen the time frame if extra testing is required. To lessen these delays, reserved slots might be taken into consideration. Delays are less likely, nevertheless, if pre-operative evaluations are finished before the primary surgical trajectory. It could be worthwhile to investigate whether there are any delays for the 25 patients in the outliers group as a result of unfinished assessments during the trajectory phase.

With a median of 200.5 days for outliers and 151 days for the entire group, Internal Medicine (*Interne Geneeskunde*) likewise exhibits long lead periods. Variability in this area could have a substantial impact on patient timetables, perhaps as a result of complex patient illnesses like diabetes, given the greater sample size of 234 patients and 46 outliers. Keeping a closer eye on this group could help cut down on delays caused by complicated patients.

Table 6.3: Top 10 Specialisms by Median Delay, Standard Deviation, Lead Time to Surgery, and Unique Patients

No.	Specialism	Median Delay (Days)	Std Dev Delay (Days)	Median Lead Time (Days)	Std Dev Lead Time (Days)	Total Unique Patients
1	*Geestelijke gezondheidszorg*	110.0	54.78	353.0	129.71	8
2	*Oncologie*	19.0	-	180.0	-	1
3	*Pijngeneeskunde*	9.0	9.17	209.0	67.47	3
4	*Gynaecologie*	8.0	24.06	131.0	42.67	6
5	*Vitalys*	7.0	12.16	125.0	77.86	327
6	*Dermatologie*	7.0	34.66	255.0	93.17	3
7	*Laboratorium*	6.5	12.77	110.5	76.76	13
8	*Interne Geneeskunde*	6.0	16.45	200.5	120.39	70
9	*Maag-, Darm- en Leverziekten*	6.0	19.01	150.0	106.99	25
10	*Anesthesiologie*	6.0	27.34	131.0	90.99	334

In summary, the top specialisms identified reflect significant variability in lead times and delays across departments, with Mental Healthcare, Pain Medicine, Internal Medicine, and Gastroenterology contributing to observed delays. This suggests that proactive scheduling strategies, such as reserved slots for these departments, may help reduce delays and enhance patient flow from screening to surgery. However, given the relatively low patient numbers in certain specialisms, adjusting the system may need careful consideration to avoid introducing new inefficiencies.

6.4. Preliminary Conclusions

The analysis of the screening phase at Vitalys Clinic reveals an overall effective patient flow, with most patients progressing through the process without significant delays. Approximately 75% of patients fall within the Standard group, indicating that the majority are scheduled and proceed to surgery within expected timeframes. However, certain specialties do show notable delays, particularly within a subset of outlier patients, which represents around 25% of the total group. These patients face considerable delays, primarily stemming from the screening phase rather than from surgery scheduling itself. This underscores the importance of optimizing the screening process, especially for patients on more com-

Table 6.4: Top 10 Specialisms by Median Delay, Standard Deviation, Lead Time to Surgery, and Unique Patients (Total Group)

No.	Specialism	Median Delay (Days)	Std Dev Delay (Days)	Median Lead Time (Days)	Std Dev Lead Time (Days)	Total Unique Patients
1	*Geestelijke gezondheidszorg*	63.0	51.81	231.0	135.52	25
2	*Allergologie*	20.0	21.28	242.0	88.00	2
3	*Oncologie*	19.0	-	180.0	-	1
4	*Pijngeneeskunde*	8.0	11.65	172.0	48.98	9
5	*Maag-, Darm- en Leverziekten*	8.0	28.03	169.0	106.34	80
6	*Apotheker*	7.5	7.07	123.0	20.87	3
7	*Vitalys*	7.0	12.16	125.0	77.86	327
8	*Interne Geneeskunde*	6.0	16.45	151.0	103.96	234
9	*Longgeneeskunde*	6.0	11.99	163.0	107.84	31
10	*Laboratorium*	6.5	12.77	110.5	76.76	13

plex or non-standard pathways, to improve system-wide efficiency.

Among the specialties analyzed, Mental Healthcare, Gastroenterology, Internal Medicine, and Pain Medicine displayed the most significant delays in both the outliers and total patient groups. However, it is important to note that these cases are largely exceptions rather than indications of systemic scheduling issues within these departments. The relatively low sample sizes in some of these specialties and the sporadic nature of delays suggest that these occurrences are likely specific cases, rather than consistent bottlenecks impacting overall patient flow.

For instance, Mental Healthcare patients often require complex assessments and follow-up steps before surgery, contributing to extended lead times. Although this specialty consistently ranks among the longest delays, the limited number of patients affected suggests that these delays are not representative of systemic scheduling inefficiencies within the department. Similarly, in Pain Medicine and Gastroenterology, patients may require additional tests or interdisciplinary consultations, which extend timelines but are likely patient-specific rather than department-wide issues.

The data indicates that, while Vitalys Clinic's scheduling system is generally effective, targeted adjustments could improve patient flow further. For example, Vitalys, which handles the largest patient volumes, demonstrates efficient scheduling with a median delay of 7 days across both patient groups and a median lead time to surgery of 125 days. However, the high standard deviation in lead times suggests that some patients experience significant delays due to additional steps required by visits to other departments. Because all patients must pass through this phase, these extended delays in certain cases contribute to variability in overall lead times.

In Gastroenterology, the outliers group exhibits a median lead time of 150 days, while the total group averages 169 days. Although delays between appointments are moderate, the long lead times indicate that some patients face extended pre-surgical preparation, particularly if additional tests are necessary. In cases like these, where specific specialties consistently show longer lead times for some patients, pre-scheduling slots could help reduce bottlenecks. However, it is important to ensure that patients complete any required tests before the main surgical trajectory phase to prevent unnecessary delays.

The analysis also shows that other specialties, like Gynecology and Radiology, frequently appear in the outliers group but not in the total group. This suggests that patients in these specialties may experience delays due to the need for additional consultations or testing after the initial preoperative screening phase. It would be beneficial to investigate whether these patients are fully "approved" to proceed to surgery but still require further evaluations, which may be contributing to scheduling bottlenecks.

The relationship between appointment delays and overall lead times in this analysis emphasizes the importance of further optimizing the screening phase, particularly for Mental Healthcare, Gastroenterology, and Radiology. By implementing pre-scheduled appointments across more specialties and refining the timing of these appointments, Vitalys could achieve a reduction in delays and improve the patient journey from screening to surgery. Introducing dedicated slots for high-impact specialties, such as Mental Healthcare, could reduce bottlenecks and address the complex needs of patients within this department. Such a proactive approach in scheduling may be especially beneficial in minimizing the impact of individual specialty delays on the overall patient flow.

In conclusion, while the majority of the screening process functions efficiently, focusing on specialties with the longest lead times could further improve patient flow. Additionally, evaluating appointments scheduled after the initial group meeting and identifying those contributing significantly to delays could streamline transitions to surgery. Ensuring that all essential preoperative evaluations are completed before patients begin their main surgical trajectory is crucial to avoiding bottlenecks. Furthermore, it is essential to investigate what occurs during the main surgical trajectory phase, as ideally, all necessary screenings should be completed beforehand. This investigation will be conducted in the following section, where we analyze the impact of appointments within the trajectory phase on overall lead times. A thorough review of the screening and approval process could prevent unnecessary delays and enhance patient flow efficiency. Addressing these targeted areas will allow Vitalys to continue refining its scheduling system and improving timely care for all patients.

7

Part 4: OR Data Analysis and Planning Model

In the previous analyses (Parts 1 and 2), we identified significant variability during both the screening phase and the Operating Room (OR) planning phase at Vitalys. In Chapter 6, we explored the variability in the screening phase and found that while there is variability between patients, the number of patients needing to go to other departments remains relatively low, which is promising. However, it is crucial to investigate what happens after the first group meeting, as concluded in Chapter 6. The current planning process at Vitalys occurs only two weeks before surgery, which adds complexity to workflow efficiency and patient satisfaction. As revealed by the FRAM analysis, eight key variables converge during the OR planning phase, contributing to significant variability. These include patient information, surgeon schedules, hospital availability, and cancellations. In this chapter, we will analyze how the OR planning process can be optimized based on these data insights and develop a new planningsmodel.

7.1. Objectives and Subquestions

One of the main goals for Vitalys is to use data more effectively to improve the accuracy and timeliness of surgery planning. Currently, surgeries are planned only two weeks in advance, leading to inefficiencies and frequent rescheduling. This short planning window results from the need for detailed patient information and the dynamic nature of the treatment pathway, which makes it challenging to plan surgeries earlier. However, there is a strong desire to shift towards scheduling surgeries much earlier, which could streamline operations and improve resource utilization.

Another question that arises is whether the planning process could be improved by analyzing patient flow from the first group meeting to surgery. As discussed in earlier chapters, Vitalys plans surgeries only after patients complete the screening phase. This chapter will focus on investigating whether earlier planning, potentially after the fifth week of the treatment trajectory, could be feasible. Can the planning window be extended, and if so, how? As part of this objective, a new planning model has been developed to help address the challenges identified during this analysis. This model aims to optimize surgery scheduling by providing an earlier schedule date and reduce bottlenecks by leveraging data-driven insights from patient flows. In this section, the aim is to answer the subquestion:

- How can data-driven insights be applied to improve the efficiency of planning processes and be integrated into a planning model to improve the planning processes at Vitalys Clinic?

7.2. Method

The methodology for this section consists of two primary components: the analysis of operating room (OR) planning data and the development of a new theoretical planning model that can be integrated and tested at the Bariatric Clinic.

7.2.1. Data Collection and Analysis

The method for this section follows a systematic approach similar to that of the previous chapter. Data was collected from Rijnstate Hospital's HiX system for patients registered in 2022 and 2023. This dataset includes demographic information and lead times from the first group meeting to surgery.

The data underwent a cleaning process, focusing on separating delayed patients (lead time greater than 49 days) from non-delayed patients. For the analysis, appointments were filtered specifically for the timeframe between the first group meeting and surgery. After that, the same delay analysis was conducted as in Chapter 6. This examination was designed to identify bottlenecks and assess where delays occurred within this critical period. The focus was on determining which departments and appointment types contributed to delays, with separate analyses conducted for the years 2022 and 2023. Additionally, the same relationship analysis was performed to compare delayed patients with non-delayed patients, allowing for exploration of patterns in patient flow. The code used for the OR planning analysis can be found in Appendix D.

7.2.2. New Planning Model

Based on insights derived from the delay and relationship analyses, a structured planning strategy was designed to enhance the scheduling process at Vitalys Clinic. This model specifically examines when delays are most likely to occur and evaluates the impact of visits to specialties outside the Vitalys Core Specialties on these delays.

The total capacity for surgeries was determined to be 30.25 patients per week, given the availability of 5.5 slots per week for 5.5 patients. To enhance the planning process, a 6-week lookahead strategy was implemented. This proactive approach aimed to allocate slots for regular patients. The following key strategies were implemented to optimize the scheduling of OR slots based on patient flow and historical data:

- Slot Allocation: A total of 30 OR slots per week were allocated for surgeries across Vitalys' two operating locations. Out of these, 24 slots were reserved for standard patients requiring only appointments within the Vitalys Core Specialties, 4 slots were allocated for postponed patients, and 2 slots were designated for REDO surgeries.
- Postponed Patients: Patients flagged as needing visits to external departments (e.g., Radiology, Urology) were categorized as "postponed." These patients were scheduled for surgery two weeks after their final external appointment. If the slots for postponed patients were fully booked for the week, their surgery was postponed to the next available week, with a maximum of 4 slots reserved for these patients.
- REDO Surgeries: REDO surgeries were scheduled using the same principle as postponed surgeries, with a dedicated reserve of 2 slots per week. These surgeries were scheduled two weeks in advance and would only be rescheduled if no slots were available in the current week.
- Slot Overflow: If a week's OR slots were fully booked, the scheduling system automatically shifted the surgery to the following week. This ensured that no more than 30 total surgeries were booked per week across Vitalys' two locations.

By implementing these methods, this theoretical planning model seeks to reduce delays, optimize the utilization of OR slots, and streamline patient flow across Vitalys Clinic. This structured approach, incorporating a 6-week lookahead and dedicated slot allocations for postponed and REDO patients, allows for more effective scheduling management, balancing routine patient care with the additional demands posed by external and complex cases.

The code supporting these scheduling strategies and calculations is provided in Appendix D.

7.3. Results

In this section, we present the findings from our analysis of patient appointment data, focusing on the timeline from the first group meeting to surgery. These results are crucial for understanding the delays that can impact planning six weeks in advance and for identifying areas of improvement within the scheduling process at Vitalys Clinic.

We begin with an analysis of the delayed group to uncover insights regarding median lead times, delays between appointments, and the impact of various specialties on patient flow. The findings are segmented by year, allowing for a comparison of trends and patterns across 2022 and 2023. Following this, we explore the relationship between delays and appointment types, with the goal of informing future planning strategies and optimizing patient scheduling. After examining these aspects, we proposed a new planning strategy and model based on the data analyzed in previous chapters.

7.3.1. Delayed Group Analysis: 1st Group session to Surgery

In this section, we focus on appointments from the 1st group session to the surgery. This analysis is crucial because when a patient scheduled for a group meeting should be eligible for surgery, indicating that their treatment trajectory is progressing and should not be delayed. Therefore, understanding where delays occur and which appointments contribute to these delays is essential for optimizing patient care.

Key Findings: 2022

Figure 7.1 highlights several key insights concerning the Median Lead Time to Surgery, Median Delay Between Appointments, and the Total Number of Patients across various specializations in 2022.

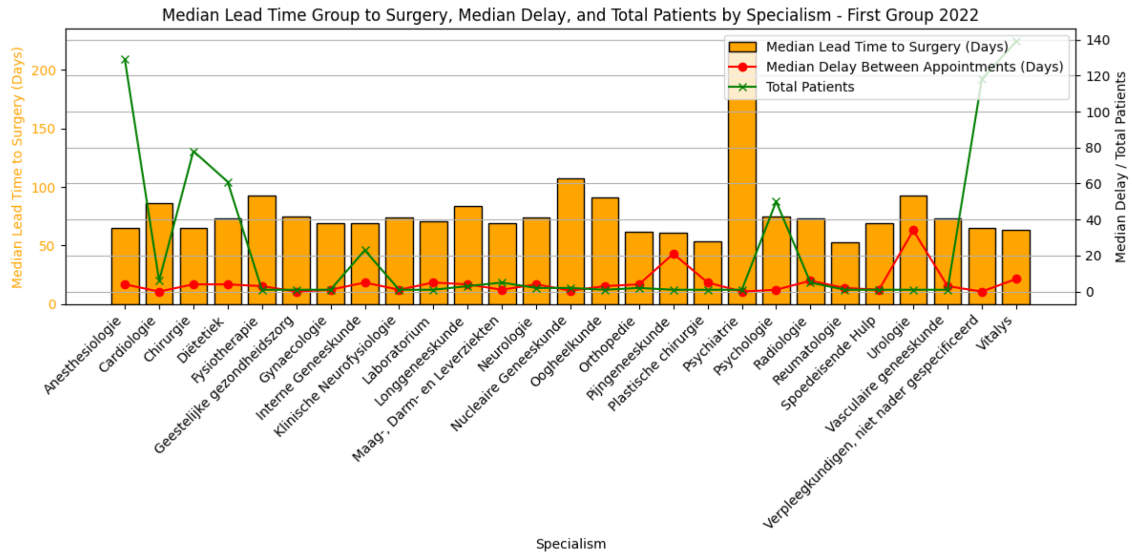

Figure 7.1: Median Lead Time Group to Surgery 2022

As expected, Vitalys has the highest number of patients, followed by Anesthesiology (*Anestesiologie)*. This suggests that nearly all preoperative screenings (POS) for 2022 patients occurred after the first group meeting, which, while not ideal, is important to check whether these POS screenings have caused follow-ups. As discussed in earlier chapters, this may introduce delays that could have been avoided with earlier scheduling of the POS.

Looking at patient numbers across other specialties, they remain relatively low. There is a slight increase in the number of patients seen by Internal Medicine (*Interne Geneeskunde*) (for diabetic patients), Dietetics (*Dietitiek*), and Psychology (*Psychologie*), reflecting individual appointments. Some surgical consultations also occurred. Contrary to previous concerns, the preoperative screening does not result in a significant number of additional appointments after the first group meeting. Slight increases are observed for Gastroenterology (*Maag-, Darm en Lever*) and Radiology (*Radiologie*), but the numbers remain below ten in both cases.

In terms of median delays between appointments, most specialties show low values except for Pain Medicine (*Pijngeneeskunde*) and Urology *Urologie*), where delays are notably longer. This pattern was also seen during the screening phase and is interesting to note. Both specialties show higher lead times, indicating bottlenecks, likely due to the complexity of cases or availability of specialists. Diagnostic procedures, follow-ups, or multidisciplinary care may contribute to these delays.

Psychiatry (*Psychiatrie*) has a notably high Median Lead Time to Surgery, although the number of patients affected is very small. This indicates that these patients are likely still undergoing approval processes or treatment, leading to significant delays before surgery can be scheduled. This suggests that the psychiatric assessment process should ideally be completed before the group meetings to prevent such delays.

Overall, most other specialties do not show significant delays in 2022. Vitalys itself shows lower median delays, while specialties like Cardiology, Physiotherapy, Pulmonology, Nuclear Medicine, Ophthalmology, and Urology exhibit slightly higher median lead times, though these affect very few patients and appointments.

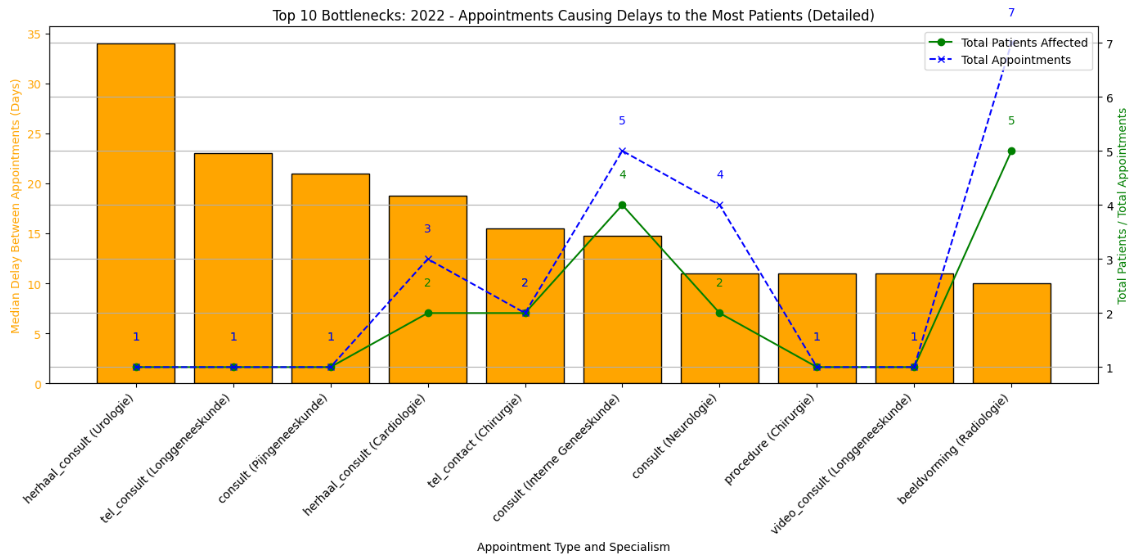

Figure 7.2: Bottlenecks Group to Surgery 2022

In Figure 7.2, the top 10 bottlenecks for 2022 are presented, highlighting appointments contributing most to patient delays. The orange bars represent the median delay between appointments (in days), the green line indicates the total number of patients affected, and the blue dashed line shows the total number of appointments.

The "herhaal_consult (Urologie)" category shows the highest median delay, exceeding 30 days. This long wait time significantly impacts the patient's overall treatment timeline, but given that only one patient was affected, this could be an isolated case rather than a systemic issue. Similar patterns are seen with Pulmonology and Pain Medicine, where low patient counts may skew the overall impact of these delays.

Internal Medicine consultations also show a higher median delay, likely due to the need for follow-up with diabetic patients. Given the importance of managing these cases, it would be prudent to reserve capacity in advance to prevent further delays.

Although the delay for imaging appointments is around 10 days, which is relatively modest, it still represents a bottleneck. This is likely due to follow-up checks after the preoperative screening (POS), as these delays were not observed during the screening analysis in Chapter 6. This suggests that imaging appointments are scheduled after the first group meeting, potentially contributing to delays.

The "tel_contact (Chirurgie)" category shows a median delay of around 15 days, affecting two patients. While this delay is not excessive, it is important to monitor surgical consultations closely, as any additional wait time could delay the surgery timeline. This is likely due to limited availability of surgical consultations, which may be contributing to these delays.

Overall, the number of patients affected by bottlenecks is relatively low, meaning that after the preoperative screening (POS) by Anesthesiology, there is no significant spike in appointments for other specialties. This indicates that the majority of necessary follow-up appointments are being handled efficiently, except for radiology (5 patients so not that many).

Key Findings: 2023
In the provided Figure 7.3 below, we can observe several key insights regarding the Median Lead Time to Surgery, Median Delay Between Appointments, and the Total Patients across various specializations of 2023:

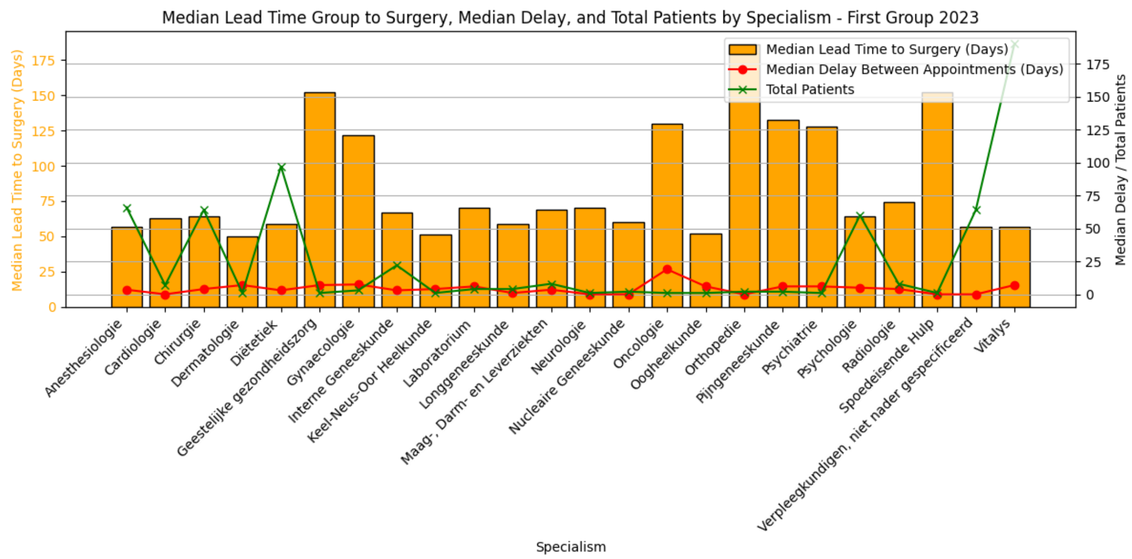

Figure 7.3: Median Lead Time Group to Surgery 2023

As expected, Vitalys has the highest number of patients, followed by Internal Medicine (*Interne Geneeskunde*), due to the increase in diabetic patients. However, the number of patients for Anesthesiology (*Anestesiologie*) is notably lower than in 2022, with just over 50 patients having their preoperative screenings (POS) after the first group meeting. This is a positive sign, as it indicates that the majority of POS screenings are now taking place before the group meetings, reducing the risk of extra tests and subsequent delays.

When looking at the patient numbers across other specializations, there is a slight increase, particularly in Internal Medicine (*Interne Geneeskunde*), as mentioned. However, many other specializations, including Dietetics (*Dietitiek*), Psychiatry (*Psychiatrie*), and Radiology (*Radiologie*), show relatively low patient counts. This suggests that most appointments needed for these specializations are already completed before the first group meeting, minimizing delays further down the line.

In terms of the median delay between appointments, represented by the red line, most specializations show a lower delay in 2023 compared to 2022, which is encouraging. However, there are some notable spikes in the Median Lead Time to Surgery. These spikes are observed in Mental Health Care (*Geestelijke Gezondheidszorg*), Gynecology (*Gynaecologie*), Oncology (*Oncologie*), Orthopedics (*Orthopedie*), Pain Medicine (*Pijngeneeskunde*), Psychiatry (*Psychiatrie*), and Emergency Care (*Spoedeisende Hulp*). Although these spikes are concerning, the low number of affected patients suggests that these may be isolated cases still in treatment. Nonetheless, it indicates that these patients tend to experience longer delays to surgery if they still have outstanding appointments.

In Figure 7.4, we present the top 10 bottlenecks for 2023, highlighting the appointments that cause the most significant delays. The orange bars show the median delay between appointments (in days), while the green line shows the total number of patients affected, and the blue dashed line represents the total number of appointments.

The ”consult (Gynecology)” category shows the highest median delay, exceeding 40 days, although it only affects two patients. Similarly, ”tel_consult (Oncology)” and ”onderzoek (Gastroenterology)” (*Maag-, Darm- en Leverziekten*) also show higher delays, but these delays affect a minimal number of

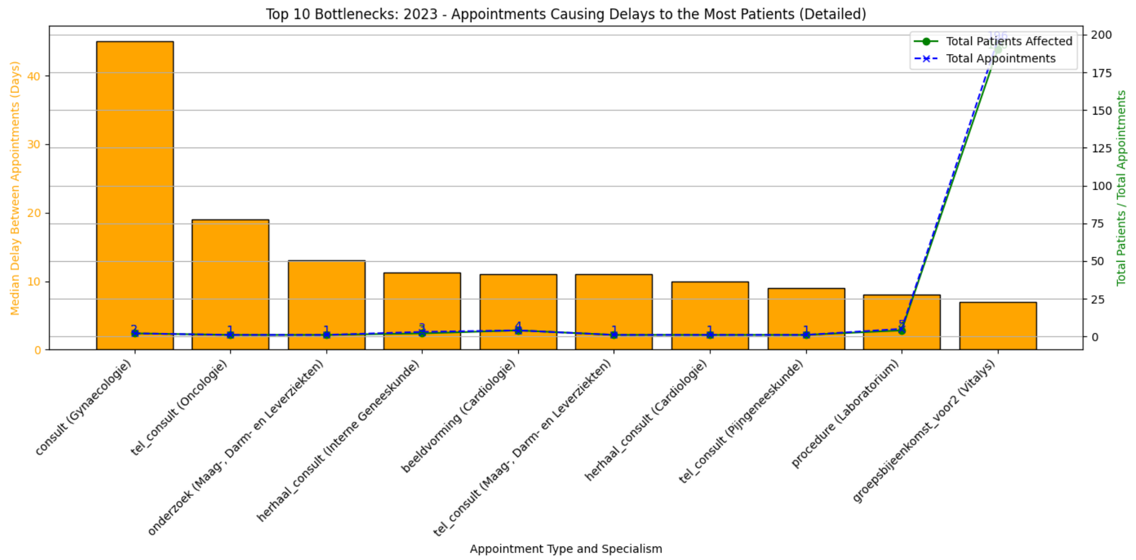

Figure 7.4: Bottlenecks Group to Surgery 2023

patients (one or two in each case), indicating that these are likely isolated cases rather than widespread bottlenecks.

Internal Medicine (*Interne Geneeskunde*) shows a slightly higher median delay, similar to 2022, due to follow-ups with diabetic patients. Reserving extra capacity for these patients may help reduce the delay. Imaging appointments (*Radiology*) also show a median delay of around 10 days, but like in 2022, this only affects a small number of patients, suggesting these delays occur after the first group meeting.

Interestingly, the "herhaal_consult (Cardiology)" and "tel_consult (Pain Medicine)" categories show a median delay, but again these appointments affect very few patients. However, given the complexities of these cases, it would be beneficial to keep an eye on these specializations to prevent future bottlenecks.

Overall, the 2023 analysis reveals that most bottlenecks affect only a few patients, and the majority of appointments are managed efficiently. The improvements in scheduling and processing are reflected in the reduced median delays between appointments across most specializations, although isolated cases still experience significant delays.

7.3.2. Relationship Analysis: Group with Delay between appointments

As in Chapter 6, we conducted a detailed analysis to understand the relationship between the delayed group and the total patient group, helping to identify areas of focus for planning optimization. This analysis is particularly valuable, as the findings can guide planners in future scheduling decisions. Figure 7.5 illustrates the relationship between median delays between appointments and overall lead times to surgery, with the total group shown in yellow and the outlier group in purple.

The data analysis examines how delays between appointments vary across specialisms and their influence on total surgery lead time. Table 7.1 and Table 7.2 present the top 10 specialisms, ranked by metrics including median delay, standard deviation, lead time to surgery, and the number of unique patients per specialty. Among the specialties, Urology (*Urologie*) stands out with the highest median delay of 34 days, signaling a significant bottleneck, although this affects only one unique patient in the outliers group. When comparing Urology to the total group, the delays are less severe, with a median lead time of 48 days. However, the standard deviation of 32.04 days indicates variability, suggesting that the potential for delays still exists. Similarly, Oncology (*Oncologie*) shows a substantial median delay of 19 days, but again, this only impacts one patient, which suggests that the delays might be more related to individual case complexities rather than systemic inefficiencies within the specialty.

Pain Medicine (*Pijngeneeskunde*), with a median delay of 9 days and 3 unique patients in both the

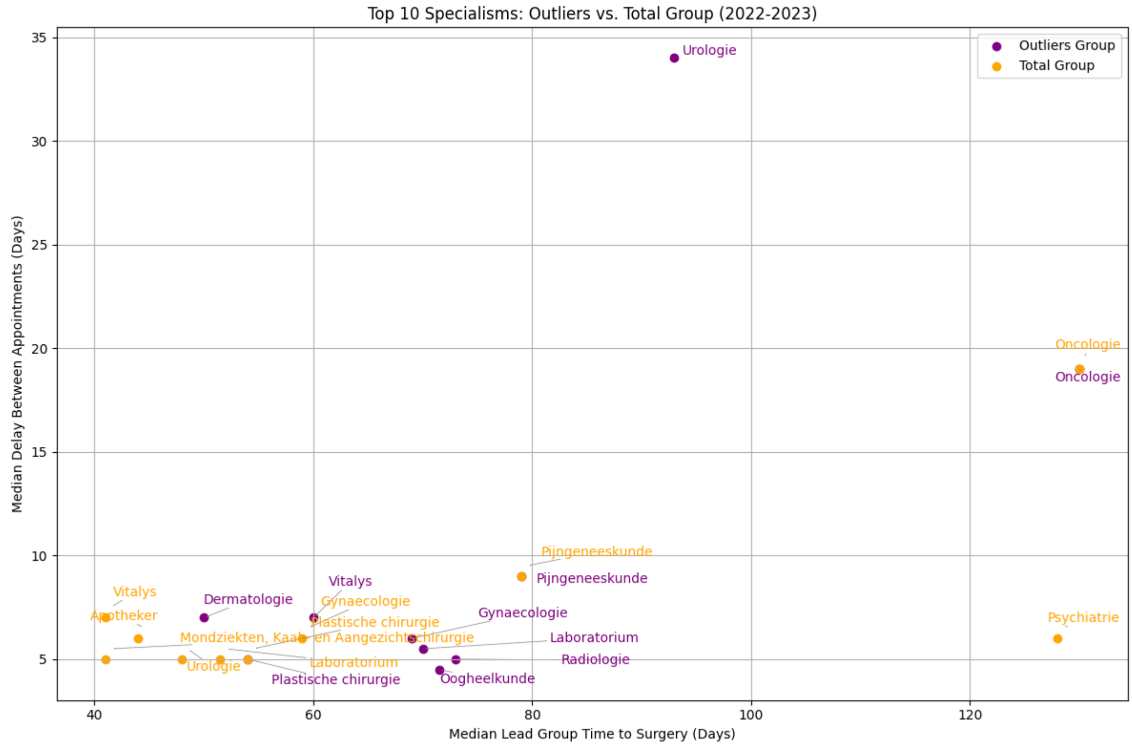

Figure 7.5: Top 10 Specialisms: Median Lead Group to Surgery vs. Delay between Appointments

outliers and total group, demonstrates the potential for significant delays, particularly for more complex cases. Gynecology (*Gynaecologie*) also appears in both groups, with 4 patients in the outliers group and 8 in the total group. This connection suggests that Gynecology contributes to delays, as it was absent from the screenings analysis (see Table 6.4).

A similar pattern is observed with Radiology (*Radiologie*), where 13 patients in the outliers group experienced delays due to additional appointments; however, these delays do not appear in the total group analysis. This suggests that visits to Radiology, although not frequent for the overall patient population, significantly impact timelines for the subset of patients requiring these additional consultations.

In other specialties, such as Dermatology (*Dietitiek*) and Maxillofacial Surgery (MKA) (*Mondziekten, Kaak- en Aangezichtschirurgie*), we observe isolated cases where delays likely resulted from one-off appointments outside the standard treatment trajectory. These individual cases caused delays for specific patients but are not part of broader systemic issues. The Laboratory department shows delays for 6 patients, likely due to the need for additional testing after the preoperative screening phase. Radiology, with 13 delayed patients over the past two years, follows a similar pattern. While most preoperative screenings in 2023 occurred before the first group meeting, suggesting improvements over 2022, these delays still impact patient timelines, particularly for those requiring additional tests.

Psychiatry (*Psychiatrie*) also shows up in the total group, with 4 patients experiencing delays. However, these psychiatric appointments did not seem to contribute directly to surgery delays, suggesting they might have been supplementary for patients needing extra support. That said, the high standard deviation in Psychiatry indicates that the timing of these appointments remains highly variable.

Overall, the findings indicate a pattern rather than a clear systemic error within specific departments. Patients who require visits to other departments seem more likely to experience delays, suggesting that additional appointments outside the core treatment pathway may contribute to these delays. However, further investigation is needed to confirm whether this correlation holds consistently across cases. If confirmed, identifying patients who may need external appointments early in the scheduling process could allow planners to anticipate and mitigate potential delays, resulting in a smoother workflow and reduced wait times.

Table 7.1: Top 10 Specialisms by Median Delay, Standard Deviation, Lead Group Time to Surgery, and Unique Patients (Outliers Group)

No.	Specialism	Median Delay (Days)	Std Dev Delay (Days)	Median Lead Time (Days)	Std Dev Lead Time (Days)	Total Unique Patients
1	Urologie	34.0	-	93.0	-	1
2	Oncologie	19.0	-	130.0	-	1
3	Pijngeneeskunde	9.0	9.17	79.0	67.57	3
4	Vitalys	7.0	12.18	60.0	36.55	327
5	Dermatologie	7.0	-	50.0	-	1
6	Gynaecologie	6.0	28.81	69.0	58.96	4
7	Laboratorium	5.5	7.08	70.0	26.59	5
8	Radiologie	5.0	22.57	73.0	67.29	13
9	Plastische chirurgie	5.0	7.21	54.0	-	1
10	Chirurgie	4.0	6.64	65.0	41.71	141

In summary, while there are no clear systemic inefficiencies across departments, the data suggests that patients needing appointments outside of the core specialties may be more prone to delays. By exploring this relationship further, Vitalys Clinic could refine its planning approach to minimize delays, particularly for patients requiring cross-department coordination, and improve the efficiency of the overall scheduling process.

Table 7.2: Top 10 Specialisms by Median Delay, Standard Deviation, Lead Time to Surgery, and Unique Patients (Total Group)

No.	Specialism	Median Delay (Days)	Std Dev Delay (Days)	Median Lead Time (Days)	Std Dev Lead Time (Days)	Total Unique Patients
1	Oncologie	19.0	-	130.0	-	1
2	Pijngeneeskunde	9.0	9.17	79.0	67.57	3
3	Vitalys	7.0	11.63	41.0	23.78	1368
4	Psychiatrie	6.0	5.39	128.0	91.88	4
5	Gynaecologie	6.0	23.22	59.0	54.80	8
6	Apotheker	6.0	-	44.0	-	1
7	Mondziekten, Kaak- en Aangezichtschirurgie	5.0	-	41.0	-	1
8	Urologie	5.0	18.36	48.0	32.04	3
9	Plastische chirurgie	5.0	7.21	54.0	-	1
10	Laboratorium	5.0	6.25	51.5	30.29	9

7.3.3. New Plannings Model

In our previous data analyses, we observed that the number of patients requiring visits to departments outside of the Vitalys Core Specialties is relatively low after the group meetings (see Figure 7.3 and Figure 7.1). By grouping Vitalys with departments such as Dietetics (*Dietitiek*), Psychology (*Psychologie*), Nurses (*Verpleegkundigen*), and Anesthesiology (*Anesthesiologie*), we found that a significant portion of patients only need to interact with these core departments. However, for the subset of patients who must visit external departments, the likelihood of experiencing delays increases substantially.

Figure 7.6 shows a stacked area chart that visualizes the number of delayed patients per specialty across varying weeks of delay, beginning with 7 weeks. Each color in the figure represents a particular specialty, demonstrating how the total number of delayed patients changes as the delay duration grows. This figure efficiently shows the distribution of delayed patients by specialty and duration of their delays,

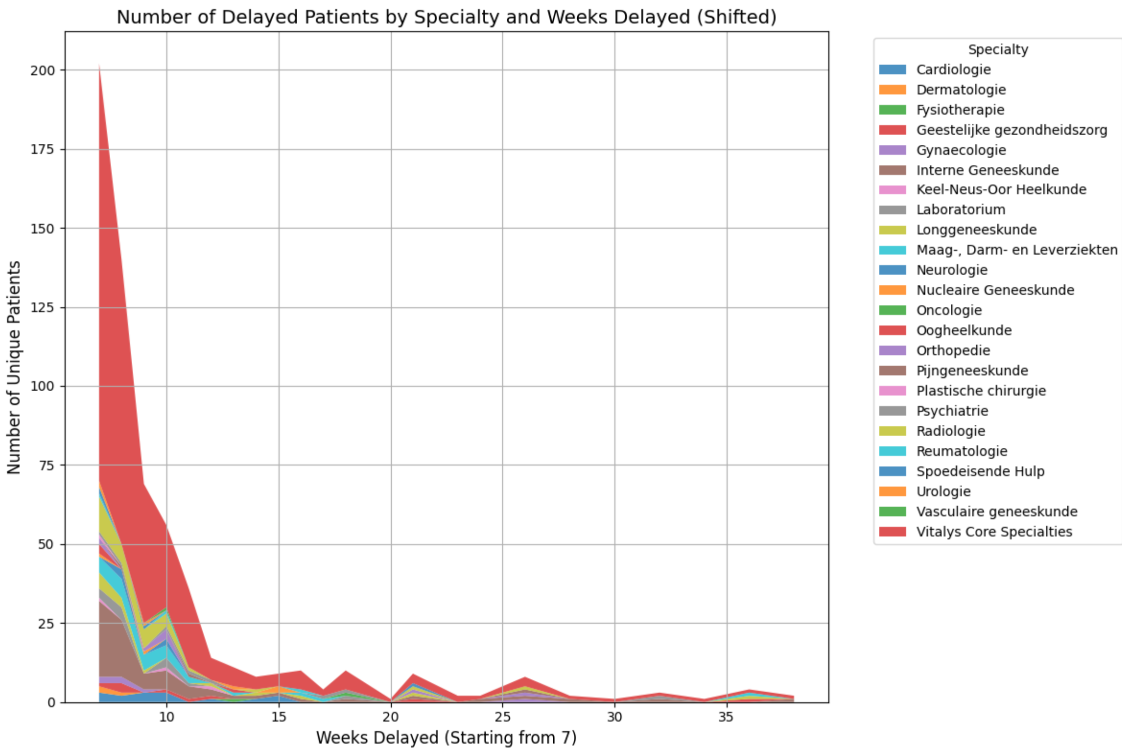

Figure 7.6: Stacked Diagram Specialisms across Delays

revealing which specializations cause the most delays over time.

At the start of the delay phase (7 weeks), there are more than 200 patients who are suffering delays. The red region of the figure represents the ”Vitalys Core Specialties,” which account for the majority of these delays. This suggests that a large proportion of patients who are initially delayed come from these primary departments. However, as the delay period progresses, the number of delayed patients at Vitalys Core Specialties decreases considerably.

The majority of delays occur within the first 7 to 10 weeks and include patients from Vitalys Core Specialties, indicating that early delays are most likely related to Vitalys' internal planning inefficiencies. After 10 weeks, the proportion of delayed patients from Vitalys Core Specialties reduces, and delays are more likely to involve patients who must visit other departments. This trend indicates that patients receiving care from many specialties, particularly those outside Vitalys, would experience longer delays due to the greater coordination required between these departments. These scenarios, which are generally more complicated and involve numerous departments, may necessitate specialist scheduling procedures because they fall outside of standard planning strategies.

When examining delays more closely in Figure 7.7, it becomes clear that patients who need to visit non-Vitalys departments experience a significant increase in delays. The red dashed line, indicating the percentage of delayed patients requiring appointments outside of Vitalys, shows a steep rise after 49 days, with delays intensifying beyond the threshold. This trend suggests that delays are more likely when patients need additional cross-departmental coordination with external specialties such as Radiology, Gynecology, and Pain Medicine, which fall outside Vitalys Core Specialties.

In the initial delay period, within the first 49 days, the majority of delayed patients—about 14% according to the chart—only require appointments within Vitalys Core Specialties. However, as delays extend, the percentage of patients needing visits to external departments steadily increases. This trend suggests that, over time, more patients face prolonged delays as they await appointments outside the Vitalys Core Specialties. This shift in required appointments highlights that delays often become more pronounced for patients needing coordination with non-Vitalys departments, particularly beyond the 7-week threshold.

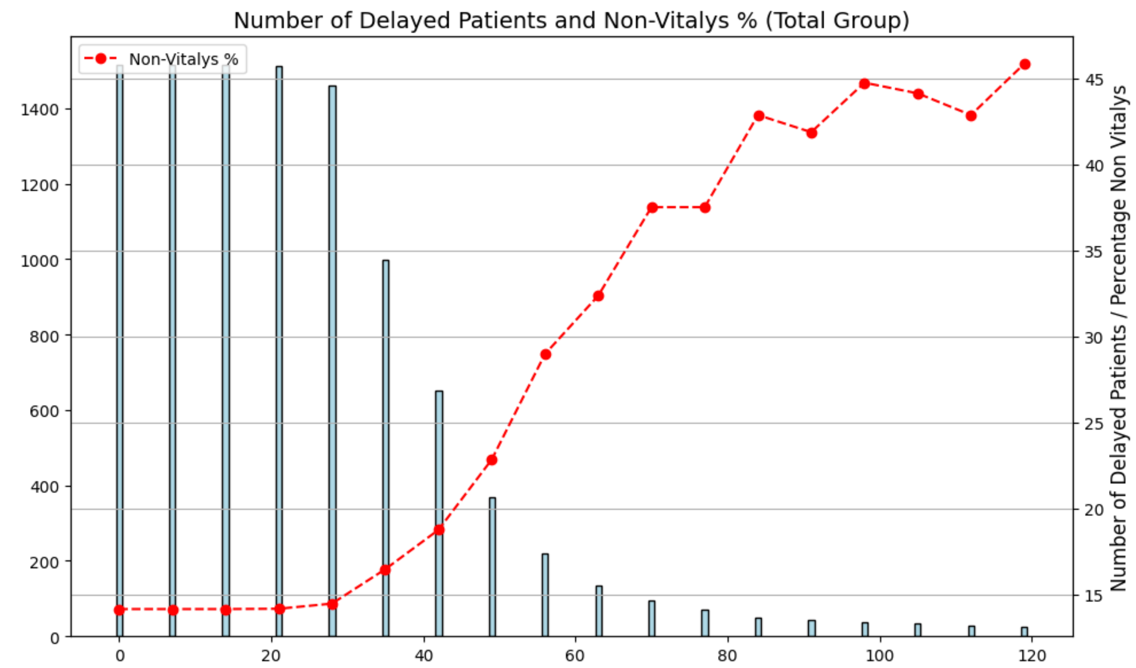

Figure 7.7: Delay Patients across percentage of Non Vitalys Specialisms

In summary, these findings underscore a strong link between the involvement of external departments and increased delays. Given this correlation, it may be effective to categorize patients needing external department visits as exceptions and refrain from scheduling their surgeries until all required evaluations are complete. By delaying the scheduling for these cases, planners can better manage these patients' unique timelines and reduce the likelihood of rescheduling. Improved coordination with non-Vitalys departments is also crucial to anticipate and mitigate these additional delays. This approach could help optimize patient flow, reduce overall wait times, and streamline surgical scheduling at Vitalys

Risk of Early Planning 6 Weeks Ahead: Operating Rooms

In examining the risk of early planning, particularly with a 6-week lead time for operating rooms, the "Comparison of Delayed Patients from First to Last Group Meeting" graph (Figure 7.8) highlights how delays vary across different stages of group meetings. This analysis included 1,515 patients, offering insight into how scheduling timing impacts delays.

The percentages of delayed patients at each meeting stage are as follows:

Group Meeting	Delayed Percentage	Number of Delayed Patients
First Group Meeting (voor1)	24.29%	371
Second Group Meeting (voor2)	23.03%	351
Third Group Meeting (voor3)	20.93%	319
Fourth Group Meeting (voor4)	17.76%	271
Fifth Group Meeting (voor5)	16.87%	257

Table 7.3: Delayed Patients Based on Group Meetings

This figure clearly illustrates the gradual decrease in the percentage of delayed patients from the first group meeting ("voor1") to the fifth group meeting ("voor5"). In table 7.3, we can find the percentage delayed at the time of the groupsmeeting. At the first meeting, 24.29% of patients experience delays, which progressively declines to 16.87% by the time of the final group meeting. This downward trend indicates that delays tend to decrease as patients move further along in the treatment process, however it is worth noting that the changes in delay percentages between meetings are relatively small. For example, the drop in delayed patients between the second and third group meetings is just over 2%.

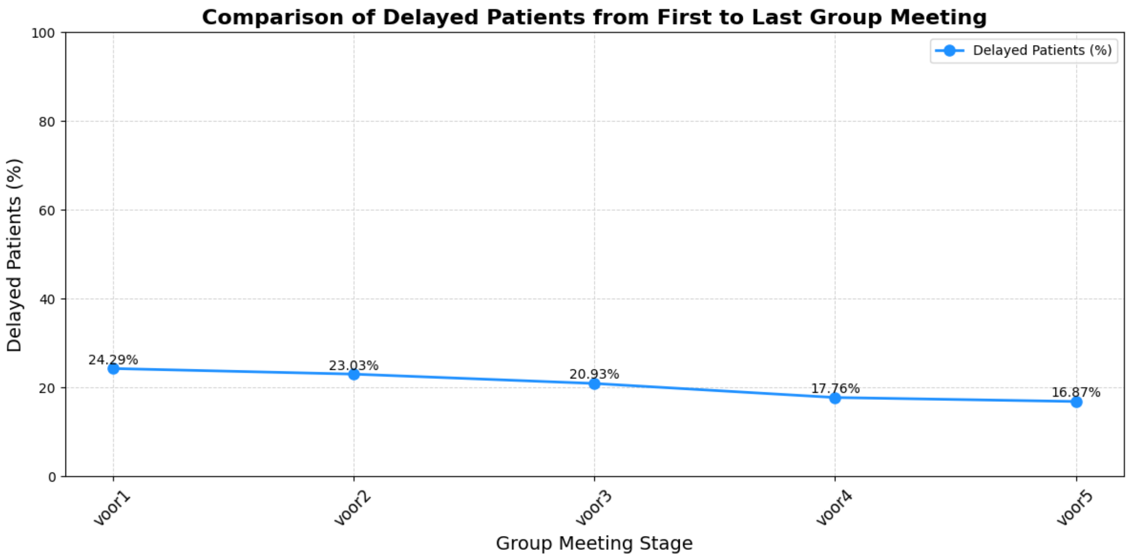

Figure 7.8: Comparison of Delayed Patients from First to Last Group Meeting

This suggests that while delays decrease, they do not dramatically reduce after the first meeting.

This says that even if you wouldn't change the schedule right now, the chance that you have to reschedule the patients 6 weeks before compared to 2 weeks before, is only 114 over 2 years. That is, calculated to per month, between 2-3 patients per month that have to be rescheduled due to delay. Analysis of the data alsoreveals that 65.26% of delayed patients required only appointments within Vitalys Core Specialties, including Psychology, Dietetics, Nursing, Anesthesiology, and Surgery. The remaining 34.74% of delayed patients required visits to departments outside these core specialties. However, as observed in earlier analyses, patients who need to visit other departments face a significantly higher likelihood of delays. This finding aligns with previous observations indicating that delays are often linked to the involvement of appointments outside Vitalys' core specialties.

Based on these insights, a more targeted scheduling strategy can be implemented. For patients who only require care within Vitalys Core Specialties, a 6-week planning lead time could be introduced. Given that this group has a lower likelihood of delays, their surgeries could be scheduled well in advance, reducing disruptions. By focusing on these patients, who are less prone to delays, Vitalys can manage its operating room capacity more effectively and reduce overall wait times. Furthermore, if patients are scheduled six weeks in advance, they are less likely to cancel, as they can better organize their personal schedules and mentally prepare for the surgery.

For the smaller group of patients who need appointments at external departments, it may be necessary to treat them as ”postponed” patients. These cases would require additional time for scheduling, and planners could account for potential delays by adjusting timelines or applying alternative scheduling processes. This approach would ensure that complex cases, involving multiple departments, are managed more effectively without affecting the majority of patients who can be scheduled well in advance.

In conclusion, the analysis suggests that early planning for surgeries, particularly for patients who remain within Vitalys Core Specialties, can be successfully implemented with a 6-week lead time. This would reduce delays and improve overall scheduling efficiency. Patients who require appointments outside of Vitalys can be treated as exceptions, with additional time or resources allocated to manage their delays. By adopting this approach, Vitalys can optimize its planning process, ensuring more reliable scheduling and minimizing delays for the majority of patients.

Planning Strategy for Vitalys: Optimizing Operating Room Slots

To develop a more effective planning strategy for Vitalys, we focus on optimizing the use of the available 5.5 operating slots per week, which are shared between the Ede and Rijnstate hospitals. This setup accommodates around 5 to 6 patients per slot, translating to approximately 30 patients per week. It is important to note that when the planning does not reach full capacity at Vitalys, the unused OR rooms are returned to the scheduling system for other purposes.

The data (see Figure 7.7) indicates that 30-35% of patients visiting departments outside of Vitalys experience delays beyond 49 days. Considering that roughly 25% of all patients face some form of delay, we can estimate the number of slots required for postponed patients. The calculation shows that approximately 3 slots per week should be allocated for postponed patients, calculated as follows:

$$\text{Number of slots needed for postponed patients} = 0.35 \times 0.25 \times 30 = 2.625 \text{ slots per week}$$

Therefore, we reserve 4 slots per week—1 extra patient slot is included for unforeseen cases, such as patients who, during the preoperative screening, receive news that additional tests are required, which may delay the surgery. These patients will be considered postponed patients for the scheduling model. In de resultaten zagen we echter dat dit heel weinig het geval is.

In addition to postponed patients, we must reserve slots for REDO surgeries, which require dedicated space in the operating schedule. As shown in Chapter 6, descriptive statistics indicate that approximately 60-80 patients per year undergo REDO operations. Given that there are 42 weeks in the year for Vitalys to operate, we reserve 2 slots per week for these procedures. REDO surgeries are scheduled 2 weeks in advance, with a limit of 2 slots per week. The scheduling logic ensures that REDO surgeries do not exceed the weekly capacity, and if a week is fully booked, the surgeries are rescheduled to the following week.

For both postponed and REDO surgeries, if a week becomes fully booked, the surgeries are shifted to the following week. This ensures that these patients experience a maximum delay of only 1 week. The scheduling rule is as follows: once the last appointment in a non-Vitalys department is completed, the patient is scheduled for surgery 2 weeks later. This approach helps ensure efficient scheduling based on each patient's appointment history outside of Vitalys. Per week, one slot (6 patients, postponed and REDO) is reserved for these "postponed" patients.

The next step was to write the code to implement the new scheduling system. We defined the available operating room slots per week, which includes a total of 30 slots, with 4 reserved for postponed patients and 2 reserved for REDO surgeries. The remaining 24 slots are allocated for standard patients—those who only need appointments within the Vitalys Core Specialties.

A new column `has_other_department` was created to flag patients who have appointments in departments outside the Vitalys Core Specialties. This step was essential to identify patients whose surgeries may need to be postponed due to visits to external departments, such as Radiology or Urology, which are likely to cause delays. Patients who had appointments in external departments were scheduled based on the latest appointment date in those departments, with a 2-week lead time applied. If the available slots for postponed patients in a given week were fully booked, the surgery was pushed to the following week. For patients who only visited the Vitalys Core Specialties, the surgery was scheduled 6 weeks after their first group meeting. Similarly, if the slots for scheduled patients in a given week were filled, the surgery was postponed to the next available week.

We handled REDO surgeries separately, ensuring that they were scheduled 2 weeks in advance of the operation date, just like the postponed patients. The appointments of the REDO patients can be different, and therefore should also be considered as separate cases. The scheduling system checked whether the allocated slots for REDO surgeries were available, and if not, the surgery was postponed to the following week. For each week, we tracked the number of scheduled, postponed, and REDO patients using the `patient_count_per_week` dictionary. This dictionary ensured that no more than 30 total slots were used in any given week, and surgeries were automatically postponed if the weekly capacity was exceeded.

Furthermore, when we calculate the number of patients planned according to the three scheduling categories, we obtain the results found in table 7.4. This means that over the last 2 years, only 215 patients who followed the "regular" trajectory were scheduled later than expected. The new theoretical planning model can be found in Figure 7.9 and the code can be found in appendix D.

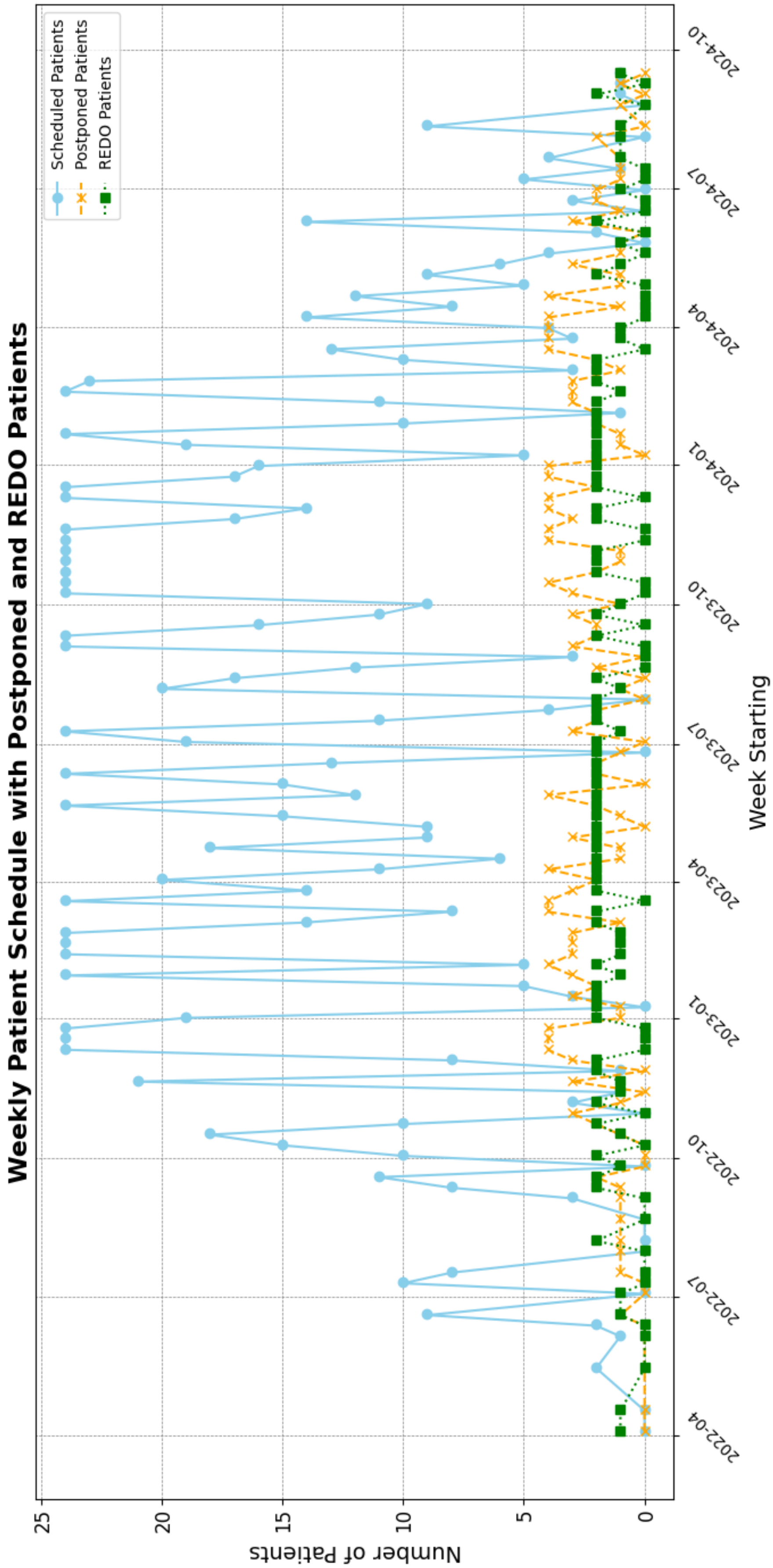

Figure 7.9: Theoretical Plannings Model

Table 7.4: Counts and Percentages of Scheduled, Postponed, and REDO Patients

Patient Category	**Count**	**Percentage (%)**
Standard Patients	1300	78.74%
Postponed Patients	215	12.96%
REDO Patients	137	8.30%
Total Patients	1652	100.00%

7.4. Preliminary Conclusions

The analysis of Vitalys' Operating Room (OR) planning process provides essential data-driven insights that offer clear pathways to enhance patient scheduling efficiency through improved planning strategies. Key findings highlight that specialties such as Pain Medicine, Radiology, and Gynaecology are associated with higher delays, particularly for patients requiring follow-up appointments with these departments after their group meetings. Although the number of patients affected in these areas is relatively small, the internal bottlenecks identified suggest an opportunity to streamline coordination to reduce lead times and improve overall patient flow. This improvement would require minimal structural changes and represents a straightforward area for optimization.

Data-driven insights underscore that delays are particularly pronounced for patients requiring visits to external departments, especially when these delays exceed the 49-day mark. This complexity, largely due to coordinating multiple appointments across departments, contributes significantly to extended lead times. Conversely, patients seen solely within the Vitalys Core Specialties experience fewer delays, indicating that internal scheduling inefficiencies are more manageable and responsive to planning adjustments. This finding emphasizes the value of cross-departmental coordination as a key area to address for patients who need external appointments, which will reduce their risk of delays.

Approximately 25% of the total patient group encounters delays, with 30-35% of patients who require external appointments experiencing delays beyond 49 days. To mitigate this, the data suggests reserving specific OR slots weekly for postponed patients, with 4 out of 30 OR slots allocated for this purpose. This targeted strategy helps address scheduling challenges associated with external departments and can be implemented in the existing system without substantial operational changes.

Similarly, REDO surgeries, though a smaller part of the scheduling load, benefit from reserved capacity. Allocating 2 slots weekly for REDO procedures ensures these patients are scheduled efficiently without competing with standard or postponed cases, further enhancing overall planning structure and reducing rescheduling needs.For standard patients—those needing only appointments within Vitalys Core Specialties—a proactive planning model with a 6-week lead time is feasible. By scheduling these patients in advance, resource utilization in the OR can be optimized, reducing the likelihood of last-minute changes. This pre-planned approach enables more stable scheduling and prevents last-minute cancellations, effectively using the 6-week forecast window to minimize disruptions.

In summary, data-driven insights can be seamlessly integrated into a planning model, enabling Vitalys to improve scheduling with minimal disruption to current systems. By establishing dedicated slots for postponed and REDO patients and scheduling standard patients 6 weeks in advance, Vitalys can anticipate a significant reduction in overall delays. This structured method for resource allocation fosters more effective use of OR capacity, enhancing patient flow and care efficiency.

In the following section, we explore predictive possibilities for further optimizing the planning model. Beyond identifying the need for external appointments, we will investigate other predictive factors contributing to delays. This includes the feasibility of building a predictive model for delay-prone patients and evaluating whether AI or traditional planning adjustments best support a more dynamic, data-responsive scheduling approach.

8

Part 5: Predictive Analysis

In the previous sections, we examined Vitalys Clinic bariatric patient flow delays and bottlenecks. These findings identified patient problems, particularly those requiring external department appointments, which significantly affected surgical scheduling. Chapter 7 proposes a planning model where patients with additional appointments at their initial group meeting are 'postponed,' meaning they are scheduled only for the last two weeks of the planning window.

In this section and given the previous observations, we turn to predictive modeling capabilities, a tool that may improve patient delay understanding and scheduling efficiency. This chapter evaluates the effectiveness of machine learning classifier models—namely Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP)—in predicting delays within the bariatric patient flow at Vitalys Clinic. The objective is to determine whether these models can enhance scheduling processes by accurately predicting the delays. With a predictive model in place, Vitalys Clinic could leverage real-time data to forecast potential delays and reduce the incidence of delayed surgeries, optimizing overall patient flow by flagging high-risk cases for further attention earlier in the scheduling process.

8.1. Research Objective

The primary objective of this chapter is to evaluate the predictive capabilities of machine learning models—specifically Decision Trees, Random Forests, and Multi-Layer Perceptron (MLP)—in forecasting scheduling delays within the bariatric patient flow at Vitalys Clinic. By assessing the effectiveness of these models, this chapter aims to determine whether predictive analytics can feasibly enhance scheduling efficiency. Additionally, this chapter explores the most significant factors contributing to delays, providing insights to address root causes and improve overall patient flow. For instance, it examines whether patients requiring appointments outside of Vitalys Core Specialties are more prone to delays compared to those managed entirely within core departments, as suggested in previous chapters. Understanding these distinctions is essential for developing targeted strategies that optimize scheduling processes and patient outcomes. This chapter seeks to answer the following question:

- Are there predictive possibilities to enhance the planning process at Vitalys Clinic and will it enhance the traditional scheduling process?

8.2. Method

This study utilized machine learning algorithms—including Random Forest, XGBoost, Decision Trees and Multi-Layer Perceptron (MLP)—to predict delays in bariatric surgery scheduling at Vitalys Clinic. The objective was to identify the model with the highest predictive accuracy and interpret the factors contributing to scheduling delays.

8.2.1. Data Preparation

The dataset consisted of the same 1,515 patients, as used in the previous analyses, including patient demographics used in previous analyses as well as additional information such as age, gender, health indicators (e.g., BMI, smoking habits), and medical history, including comorbidities and previous treatments. The target variable, `delayed_surgery`, was derived from the `lead_time_group_to_surgery` column, with a threshold set at 49 days. Patients with lead times exceeding 49 days were classified as delayed (1), while those with lead times of 49 days or fewer were classified as not delayed (0).

To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to oversample the minority class (delayed surgeries). This balancing process allowed the models to learn from both classes more effectively, mitigating biases and improving predictive performance for the minority class (Elreedy and Atiya 2019).

Statistical analysis was conducted to examine differences between delayed and non-delayed groups. For continuous variables such as age and BMI, independent t-tests were performed to test for significant differences. For categorical variables, chi-square tests were utilized to assess associations with the delay status, providing a deeper understanding of patient characteristics that might influence surgery delays.

8.2.2. Predictive Modeling

As mentioned, This study implemented and evaluated several machine learning models—specifically Random Forest, XGBoost and Multi-Layer Perceptron (MLP)—to predict surgery delays within the bariatric patient flow at Vitalys Clinic. Each model underwent systematic hyperparameter tuning to optimize its predictive capabilities. Hyperparameter tuning, a crucial step in machine learning, involves adjusting parameters that control a model's behavior during training but are not learned directly from the data itself. This process is essential to improve model accuracy and prevent overfitting (Xu, Coen-Pirani, and Jiang 2023). Additionally, we evaluated the influence of various variables that were statistically significant as well as those that were not statistically significant in prior analyses, as these could still impact the model's overall performance.

With this hyperparmeter tuning, key parameters for the Decision Trees, Random Forest, and Gradient Boosting models were carefully fine-tuned using GridSearchCV and RandomizedSearchCV. These parameters included tree depth, the number of estimators, and learning rates. The goal of this process was to get the models as close to perfect as possible while also avoiding overfitting. This made sure that each model could work well with new data. For the Multi-Layer Perceptron (MLP) model, things like the regularization factor (alpha), learning rate, and hidden layer configuration were changed to get the best results and help the model work well with data it had not seen before.

Learning curves were generated for each model—Random Forest, Gradient Boosting, and MLP—to examine generalization across training set sizes and detect any overfitting or underfitting. High training accuracy coupled with lower testing accuracy indicates overfitting. Specifically, when the training accuracy approaches 100% while testing accuracy remains significantly lower (e.g., around 60-70%), this suggests the model has learned the training data well but struggles to generalize to new data. Such a gap between training and testing accuracy is a clear marker of overfitting, indicating the model's reliance on specific patterns within the training data rather than capturing broader, generalizable trends.

Each model's performance is assessed using metrics such as accuracy, precision, recall, and F1-score. Special attention is given to the delayed surgery class to ensure that these models are effectively capturing delay predictions. Confusion matrices and classification reports provide insights into each model's strengths across classes. Furthermore, learning curves allow for visualization of the model's generalization over varying training sizes, which highlights any trends toward overfitting or underfitting.

Random Forest Model

The Random Forest model was selected for its robustness and ability to handle complex datasets with minimal preprocessing. Random Forests are especially fit for datasets where interactions between variables may not be clear, as they excel in capturing non-linear correlations between features (Wright and König 2019). After applying SMOTE for balancing the dataset, 10-fold cross-validation was used to ensure the model's robustness.

10-fold cross-validation involves dividing the dataset into 10 equal parts, or "folds." The model is then trained and evaluated 10 times, with each fold being used as the test set once, while the remaining 9 folds are used for training. This process allows the model to be trained and tested on different subsets of the data, providing a more reliable estimate of its performance and helping to minimize overfitting. After completing the 10 iterations, the performance metrics are averaged to give an overall evaluation of the model. Additionally, training was completed using the best hyperparameters identified during the cross-validation process.

XGBoost Model

The XGBoost model was selected for its efficiency and ability to handle large-scale datasets where capturing nuanced interactions and patterns is critical. Unlike Random Forest, which builds an ensemble of fully independent decision trees, XGBoost builds sequential, interconnected trees that iteratively improve on previous predictions. This approach, known as gradient boosting, allows XGBoost to capture subtle non-linear relationships and feature interactions effectively (Bentéjac, Csörgo, and Martínez-Muñoz 2019). The added regularization in XGBoost also reduces the likelihood of overfitting, making it advantageous for datasets where fine-tuning predictive accuracy is essential.

After applying SMOTE to balance the classes, hyperparameters such as learning rate, maximum depth, and number of estimators were fine-tuned using GridSearchCV to enhance model performance. To ensure robustness, a 5-fold cross-validation was performed, with training completed on the best hyperparameters identified.

Multi-Layer Perceptron (MLP) Model

The Multi-Layer Perceptron (MLP) model was used to evaluate neural networks' prediction skills, particularly in capturing complicated, non-linear patterns that typical ensemble models such as Random Forest or XGBoost may not detect as easily (Boateng and Yang 2023). Unlike tree-based models, that divide data based on feature values, MLP is a feedforward neural network with interconnected layers of neurons. Each neuron in the hidden layers conducts a weighted sum of inputs, applies a non-linear activation function, and sends the result to the next layer, allowing the model to learn more abstract representations of data via deep feature modification. MLP's layered nature makes it especially useful for modeling complex linkages and interactions between variables that may not be explicitly described in the dataset.

The MLP model in this study consisted of multiple hidden layers, with hyperparameters such as the number of neurons per layer, learning rate, and activation function rigorously optimized to maximize performance. After balancing the dataset with SMOTE, training was conducted with early stopping to prevent overfitting, making MLP a viable alternative for capturing complex patterns in the data that might be overlooked by more conventional models.

This analysis used several Python libraries essential for data handling, modeling, and evaluation. Pandas and NumPy were used for data manipulation, providing efficient data structures and statistical operations. For visualizations, Matplotlib and Seaborn offered clear, informative plots for analyzing trends and patterns. To build and assess machine learning models, scikit-learn provided tools for model selection (train_test_split, GridSearchCV), evaluation metrics (accuracy, confusion matrix), and scaling (StandardScaler, MinMaxScaler). Machine learning models like RandomForestClassifier, GradientBoostingClassifier, and MLPClassifier supported our classification tasks. The code can be found in Appendix D.

8.3. Results

This section evaluates each model based on its predictive performance and descriptive insights.

8.3.1. Descriptive Results

Table 8.1 presents a summary of patient characteristics, focusing on comorbidities and health behaviors within the dataset. The dataset comprises 1515 patients, with an average age of 43.44 years and a BMI of 42.46. The standard deviations of 12.22 for age and 5.42 for BMI indicate considerable variability within the population. For each categorical variable, "Category 0" indicates the absence of the condition or behavior, while "Category 1" indicates its presence. Notably, a significant portion of patients report conditions such as hypertension (461 patients with hypertension and 1054 without) and psychological support needs (478 patients with support and 1037 without). Conversely, the prevalence of diabetes is relatively low, with only 176 patients reporting this condition.

Table 8.1: Comorbidities, Health Behaviors, and Descriptive Statistics

Variable	**Category 0**	**Category 1**
Hypertension	1054	461
Psychological Help	1037	478
Diabetes	1339	176
Smoking	1263	252
Family History of Obesity	465	1050
Family History of Diabetes	727	788
Vitalys Only (Other Departments)	474	1041
Gender Dummy	1193	322
Delayed Surgery	1147	368
Numerical Variables	**Age**	**BMI**
Count	1515	1515
Mean	43.44	42.46
Std Dev	12.22	5.42
Min	18	31.53
25%	33	38.87
Median (50%)	43	41.40
75%	53	44.99
Max	71	90.43

The data also reveals a high incidence of family history of obesity, reported by 1050 patients. This insight underscores the potential influence of hereditary factors on obesity within this population. Additionally, many patients (1041) were treated exclusively within Vitalys, reflecting the clinic's ability to manage a substantial portion of cases in-house. However, 368 patients experienced delays in their surgeries, highlighting the need to focus on scheduling efficiency—particularly for patients who may need appointments in external departments.

Given the imbalanced nature of the delayed surgery outcomes, where the majority of patients do not experience delays, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) were used, as mentiond in the methods section. By using this technique, we hope to enhance the model's ability to accurately predict which patients are likely to experience delays, ultimately improving scheduling efficiency and patient flow at Vitalys Clinic.

Statistical Analysis

To inform the development of our predictive model, we evaluated various patient characteristics using statistical tests to assess their potential influence on surgery delays. Key variables tested included hypertension, diabetes, age, BMI, psychological support needs, family history of obesity and diabetes, the need for appointments in other departments, and gender.

Table 8.2 shows that several variables, such as diabetes, age, and the requirement for appointments in other departments, gave statistically significant p-values, indicating probable significance to surgical delays. Although several variables, such as hypertension and BMI, did not achieve stringent statistical significance, they are nonetheless clinically relevant and may play a role when combined with other variables in predicting surgery delays.

Hypertension and diabetes, for example, are well-known risk factors that might have an impact on a patient's overall health, potentially leading to delays in surgical routes. Thus, even if statistical tests show a weaker correlation, these variables may still have predictive value, particularly when combined with additional characteristics in a machine learning model.

To capture every aspect of these variables, we will use an iterative modeling approach. Selected vari-

Table 8.2: Statistical Test Results for Predictors of Surgery Delays

Variable	**Test**	**Statistic**	**p-value**
Hypertension	Chi-Squared	3.58	**0.059**
Psychological Help	Chi-Squared	0.91	0.34
Diabetes	Chi-Squared	4.83	**0.028**
Smoking	Chi-Squared	0.36	0.55
Family History of Obesity	Chi-Squared	0.04	0.85
Family History of Diabetes	Chi-Squared	0.00012	0.99
Other Departments	Chi-Squared	7.62	**0.006**
Gender Dummy	Chi-Squared	2.27	0.13
Age	T-test	2.44	**0.015**
BMI	T-test	1.37	0.17

ables, including statistically significant and therapeutically relevant parameters, will be used in Random Forest, MLP (Multi-Layer Perceptron), and XGBoost models. By iteratively testing and refining these variables in each model, we hope to determine which factors have the most impact on the delay of surgeries.

8.3.2. Results of Predictive Modeling

In this study, three predictive models—Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP)—were evaluated on the given dataset to assess the feasibility of using AI for reliable predictions. The results indicate that the dataset does not yield accurate predictions with these models, suggesting that significant data improvement would be required before AI can be implemented effectively. In the section below, every model is evaluated.

Random Forest Model

Desipte the efforts to balance the data using SMOTE to address class imbalance, the model was further refined through hyperparameter tuning, including parameters like the number of trees, maximum depth, and minimum samples required for splits. A 10-fold cross-validation process was used to assess model stability and consistency across various data subsets, the Random Forest model struggled to generalize well, achieving a test accuracy of only 50%. The classification report reflects a limited predictive capability, particularly for the delayed surgery class (class 1). Although precision and recall for non-delayed cases (class 0) were relatively higher, the low recall (0.43) and precision (0.29) for the delayed class indicate that the model struggles to capture this minority class accurately, even after balancing.

- **Accuracy**: 0.50
- **Confusion Matrix**:

	Predicted 0	Predicted 1
Actual 0	122	109
Actual 1	41	31

- **Classification Report**:

Class	Precision	Recall	F1-score	Support
0	0.75	0.53	0.62	231
1	0.22	0.43	0.29	72
Accuracy	0.50			
Macro Avg	0.48	0.48	0.46	303
Weighted Avg	0.62	0.50	0.54	303

The feature importance plot (Figure 8.2) highlights the most influential variables in predicting delays. Age and BMI emerged as the leading features, suggesting a potential association with surgery delays. Other factors, such as gender, visits to other departments, and specific comorbidities like diabetes and hypertension, also contributed to the predictions but to a lesser extent. These insights may help target areas for further investigation, particularly regarding which patient characteristics are most likely to contribute to delays.

The Random Forest model's high training accuracy of approximately 82%, contrasted with its lower test accuracy around 50%, emphasizes the model's overfitting tendencies, as depicted in the learning curve (Figure 8.1). While the model effectively learns patterns from the training data, it fails to generalize to new cases, reflecting limitations in data quality and complexity that impact its predictive power. This suggests that additional data and further optimization are necessary to make the model feasible for real-time clinical decision-making, where reliable and accurate delay prediction is essential

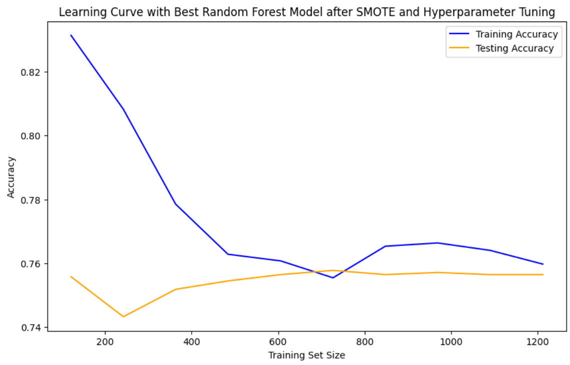

Figure 8.1: Learning Curve for the Random Forest Model after SMOTE and Hyperparameter Tuning

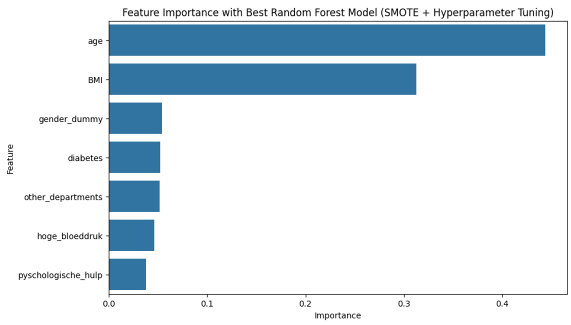

Figure 8.2: Feature Importance for the Random Forest Model (SMOTE + Hyperparameter Tuning)

Gradient Boosting Model

The Gradient Boosting model, known for its ability to combine weak learners into a strong predictive model, was implemented in this study to assess its performance in predicting surgery delays. Unlike Random Forests, which build independent decision trees and average their outputs, Gradient Boosting builds trees sequentially, each tree correcting the errors of the previous one. This approach enables Gradient Boosting to capture intricate patterns and is generally more effective for smaller datasets where complex relationships exist between features. However, despite these advantages, the Gradient Boosting model in this study yielded a modest test accuracy of only 50%, showing limited predictive capacity, similar to the Random Forest model.

- **Accuracy**: 0.50
- **Confusion Matrix**:

	Predicted 0	Predicted 1
Actual 0	117	114
Actual 1	42	30

- **Classification Report**:

Class	Precision	Recall	F1-score	Support
0	0.75	0.52	0.61	231
1	0.22	0.44	0.30	72
Accuracy		0.50		
Macro Avg	0.49	0.48	0.46	303
Weighted Avg	0.62	0.50	0.54	303

The classification report offers insight into the model's performance, especially for the delayed surgery class (class 1), where precision and recall are notably low. While non-delayed surgeries (class 0) reached a precision of 0.75, the recall of 0.52 indicates that the model correctly identifies just over half of true non-delayed cases. This lower recall, similar to the Random Forest model, shows that the Gradient Boosting model also faces challenges in identifying true cases accurately, especially when overfitting to certain patterns. For delayed cases, precision drops to 0.22 with a recall of 0.44, suggesting the model underperforms in reliably predicting delays, an essential goal in this context.

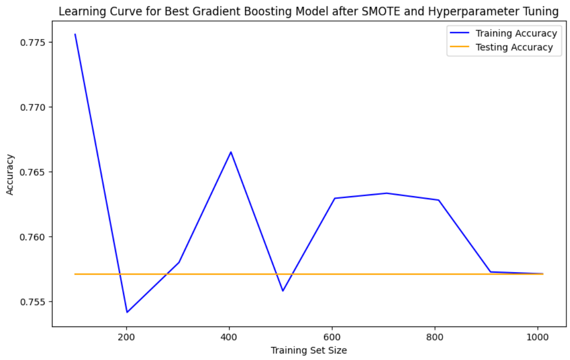

Figure 8.3: Learning Curve for the Gradient Boosting Model after SMOTE and Hyperparameter Tuning

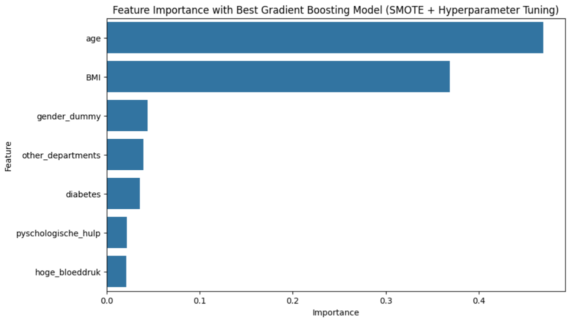

Figure 8.4: Feature Importance for the Gradient Boosting Model (SMOTE + Hyperparameter Tuning)

The learning curve (Figure 8.3) underscores the overfitting issue seen in the Random Forest model, with training accuracy nearing 1.0 while testing accuracy levels off around 50%. This gap illustrates that Gradient Boosting, while effective in memorizing training data, does not generalize well to new data. Similar to Random Forest, Gradient Boosting is susceptible to overfitting when patterns in the training data do not accurately represent the full data distribution. The lack of generalization limits the model's applicability in real-world clinical settings, where accurate predictions are crucial.

Feature importance analysis (Figure 8.4) shows that age and BMI are the primary predictors, aligning with the results from the Random Forest model, followed by factors like department visits and gender. Despite this ranking, the model's overall low predictive power suggests these features alone are insufficient for reliable delay prediction. While Gradient Boosting provides a sequentially built model that can theoretically capture deeper interactions than Random Forest, the results in this study indicate that both models face similar data limitations.

In summary, while Gradient Boosting offers theoretical advantages over Random Forest in handling feature interactions, it faces the same challenges with limited predictive capacity and overfitting in this dataset. Both models reveal age and BMI as potentially influential, but their overfitting tendencies and low predictive accuracy suggest that further data enhancements or other techniques are necessary for practical application in delay prediction.

Multi-Layer Perceptron (MLP) Model

The Multi-Layer Perceptron (MLP) model, a type of neural network, was implemented to capture complex, non-linear relationships in the dataset that may not be fully represented by tree-based models like Random Forest and Gradient Boosting. Unlike the previous two models, which rely on sequential or ensemble tree structures, the MLP model leverages multiple layers of connected neurons, allowing it to learn intricate patterns across features. The results from the tuned MLP model, however, show a test accuracy of 59%, highlighting some improvement over Random Forest and Gradient Boosting but still with limitations in predicting delays.

- **Accuracy**: 0.59
- **Confusion Matrix**:

	Predicted 0	Predicted 1
Actual 0	144	87
Actual 1	38	34

- **Classification Report**:

Class	Precision	Recall	F1-score	Support
0	0.79	0.62	0.70	231
1	0.28	0.47	0.35	72
Accuracy		0.59		
Macro Avg	0.54	0.55	0.52	303
Weighted Avg	0.67	0.59	0.62	303

The classification report details the model's performance across classes, showing that the MLP model performs better than Random Forest and Gradient Boosting in predicting non-delayed cases (class 0), achieving a precision of 0.79. This means that 79% of predicted non-delayed cases were accurate. The recall of 0.62 for this class indicates that the model identified 62% of true non-delayed cases, marking a balanced performance in terms of precision and recall with an F1-score of 0.70. However, similar to the previous models, the MLP's effectiveness diminishes when predicting delayed cases (class 1), achieving a precision of only 0.28 and a recall of 0.47. Although slightly better than the Random Forest and Gradient Boosting models, the MLP's F1-score of 0.35 for the delayed class shows it struggles to generalize and accurately predict delays.

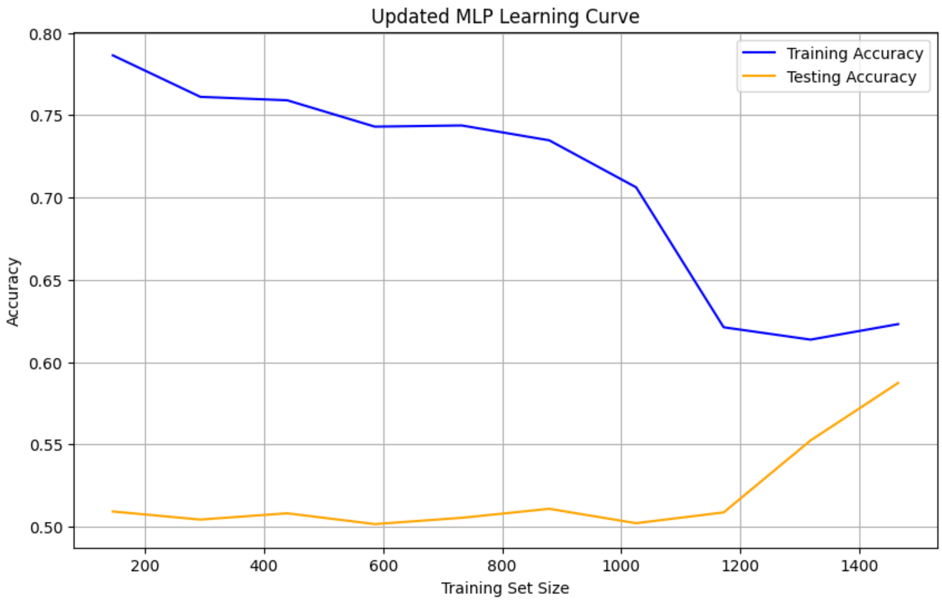

Figure 8.5: Learning Curve for the MLP Model after SMOTE and Hyperparameter Tuning

The learning curve in Figure 8.5 demonstrates slight overfitting, a common issue in neural networks. The training accuracy remains relatively high, close to 75%, while testing accuracy stabilizes between 55% and 60%, suggesting that the model learns the training data well but does not generalize effectively to unseen data. This gap between training and testing accuracy highlights overfitting, indicating that, similar to Random Forest and Gradient Boosting, the MLP model may be capturing patterns specific to the training data that do not translate to broader applicability.

When compared to Random Forest and Gradient Boosting, the MLP model shows moderate improvement in overall accuracy (59% compared to 50%), especially for the non-delayed class. This result suggests that MLP's neural architecture might offer an advantage in learning complex relationships that the tree-based models miss. However, MLP still struggles with the delayed surgery class, indicating that the current features and data quality may not be sufficient to capture the complexity required for accurate predictions of delay cases.

In summary, the MLP model demonstrates slightly better performance over Random Forest and Gradient Boosting in terms of accuracy and handling of non-delayed cases. Nevertheless, the model still faces limitations, particularly in predicting delays, as shown by low precision and recall in the delayed class and evidence of overfitting in the learning curve. To improve the MLP model's applicability in predicting surgery delays accurately, additional data or a more sophisticated neural architecture may be necessary. The comparison with Random Forest and Gradient Boosting highlights the need for refined strategies to overcome the limitations observed across all models for clinical decision-making applications.

8.4. Preliminary Conclusions

This chapter explored the potential of machine learning models—specifically Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP)—to predict scheduling delays within the bariatric patient flow at Vitalys Clinic. The aim was to assess whether predictive modeling could provide actionable insights for the scheduling process, with the ultimate goal of enhancing efficiency by identifying patients at higher risk of delays.

The results across all three models indicate limited predictive capacity in the current dataset. Each model achieved moderate accuracy but struggled particularly in identifying delayed surgery cases. Random Forest and Gradient Boosting, both tree-based models, encountered significant overfitting, as demonstrated by the learning curves where high training accuracy was met with lower, stabilized testing accuracy. This suggests that while the models capture patterns within the training data, they fail to generalize to new, unseen data, limiting their applicability in real-world clinical settings. The MLP model, while showing a slight improvement in test accuracy compared to the tree-based models, also exhibited overfitting and low precision and recall for the delayed cases.

These findings imply that the current dataset and features may not be sufficient for effective predictive modeling. The models showed that while certain patient characteristics, such as age, BMI, and the need for appointments in external departments, appear as influential features, they do not provide a strong enough basis for accurately predicting delays. The low precision and recall scores for delayed surgeries highlight the models' struggle to reliably identify this critical subset of patients.

The research question for this chapter was: 'Are there predictive possibilities to enhance the planning process at Vitalys Clinic, and will it enhance the traditional scheduling process?' Based on the findings, while predictive modeling holds potential as a tool for improving scheduling efficiency, the current dataset does not support reliable predictions of surgery delays. Thus, while there are possibilities, they are limited in the current context. For predictive analytics to enhance traditional scheduling processes at Vitalys Clinic, additional data points, refined feature engineering, or more sophisticated modeling approaches may be required. This could involve integrating more detailed patient information or exploring alternative variables that might contribute to delays.

When comparing these machine learning models to Vitalys Clinic's new proposed traditional scheduling approach, it becomes evident that the predictive models, in their current form, do not yet offer substantial advantages. The low predictive power and high overfitting suggest that, compared to existing scheduling processes, these AI-driven models would add limited value and could introduce additional complexity without a commensurate improvement in accuracy.

To conclude, although machine learning has the potential to improve patient flow predictions, the limitations in accuracy and overfitting indicate that current data and model configurations are insufficient for reliable delay prediction. To fully realize the benefits of predictive analytics in surgery scheduling, Vitalys Clinic would require a richer dataset and possibly alternative modeling approaches. For now, however, these AI models do not surpass traditional scheduling in terms of practicality and reliability, emphasizing that substantial enhancements in data and modeling are necessary for AI-driven delay prediction to become a viable tool in real-time clinical decision-making.

9 Discussion

The main objective of this study was to determine whether data-driven insights and artificial intelligence (AI) might be used to improve surgical scheduling and patient flow at Vitalys Clinic, specifically by identifying areas of variability and inefficiency in the bariatric care processes. The findings suggest that, while data-driven approaches, such as machine learning models, hold the potential to improve scheduling, a newly proposed system based on traditional methods but influenced by data insights has already proven effective—even without advanced AI capability.

Koushan, Wood, and Greatbanks 2021 conducted a systematic review of hospital-related factors that drive elective surgery cancellations, identifying primary causes as unavailable OR time, scheduling constraints, and insufficient resources, such as beds and adequate staffing. Our investigation of Vitalys Clinic indicated similar issues, particularly during the OR design phase. This is in line with the findings of Koushan et al., who claim that many hospital-driven scheduling inefficiencies can be reduced by better management techniques and long-term planning.

Palter et al. 2020 discovered that scheduling challenges were responsible for 61% of surgery cancellations, whereas bed shortages accounted for only 4.5%. They found that the majority of scheduling delays were caused by underestimated surgical times, unprepared patients, or complications from earlier cases. Hand et al. found that administrative considerations accounted for 60% of cancellations in a US teaching hospital, while bed constraints had less impact on elective procedures. The majority of cancellations were due to unsigned patient consent papers or last-minute medical developments, leaving insufficient time for discussions with family or a trusted doctor.

Interviews at Vitalys Clinic similarly identified patient-related difficulties, administrative bottlenecks, and scheduling inefficiencies as major factors to patient flow delays. These overlapping findings strongly indicate that established administrative rules and efficient scheduling techniques are critical for reducing delays and increasing bariatric treatment efficiency. The interviews also revealed a willingness among stakeholders to embrace improvements, particularly through the use of data-driven approaches or AI-based solutions.

To adress these challenges, the newly proposed scheduling system at Vitalys includes dedicated time slots for different patient groups, directly targeting critical bottlenecks identified during the study. This structured, data-driven model has shown significant gains in scheduling reliability and patient flow, making it a practical alternative to AI-based solutions. Importantly, the approach allows most patients to plan surgeries up to six weeks in advance, reducing last-minute changes, enhancing patient satisfaction, and improving overall operational efficiency.

Other research have demonstrated benefits by using data without using AI. Min and Yih 2010 suggest that incorporating patient priority into surgery scheduling decisions, such as picking patients from a waiting list based on urgency, can prevent suboptimal outcomes and enhance scheduling efficiency. Furthermore, other research show that discriminating between patient groups, such as first-time vs referral patients with varying time needs, can optimize scheduling and streamline operations (Safdar, Khan, and Shaukat 2022). This research highlights the possibility of structured, non-AI systems that use data to meet many of the requirements without the complex AI solutions.

Nonetheless, while traditional approaches have proven useful in this context, Artificial Intelligence has transformational potential, as evidenced by a large body of study. Alnsour et al. 2023 demonstrate how AI models may improve healthcare operations by enhancing resource allocation and capacity planning. They use an AI-based framework to reliably predict hospital length of stay (LOS). Lopez et al. 2022 show that AI/ML models can optimize preoperative patient selection and planning in elective surgery, while Ramkumar et al. 2019 show how neural networks predict inpatient charges and LOS, adapting efficiently in complex and variable scenarios.

One of the key trade-offs between using AI and traditional approaches is the amount of precision and predictability required for dependable decision-making. Traditional scheduling, as shown in the clinic's present planning model, offers an organized approach based on known patient paths and predictable workflows. This methodology has proven useful in decreasing delays and managing capacity in a controlled environment. However, it lacks the ability to respond flexibly to real-time changes such as unexpected cancellations and patient no-shows. This inflexibility can lead to bottlenecks, particularly during peak demand periods, which AI solutions try to alleviate by forecasting such situations.

Integrating predictive modeling at Vitalys could potentially yield useful insights into future delays. Sapir-Pichhadze and Kaplan 2020 found that decision trees and random forest models may forecast delays by assessing patient demographics, appointment histories, and specialty-specific information. Similarly, applying predictive modeling to Vitalys' appointment data could provide a mechanism to anticipate patient delays by dynamically revealing bottlenecks in the scheduling process.

Our own attempts to use machine learning models, such as Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP), revealed substantial limits, owing to overfitting and poor generalization. The limited descriptive factors available—such as patient age, BMI, and the need for trips to other departments—were insufficient for accurate delay forecasts. This finding emphasizes an important fact in healthcare predictive modeling: high accuracy frequently necessitates larger, more extensive datasets that represent the complexities of patient flow. The dataset's simplicity, with only 25% of cases experiencing delays, resulted in an imbalance that impacted model training and validation. This suggests that for AI to be a viable tool in this setting, a larger dataset with real-time information is required to provide successful prediction performance and support scheduling decisions.

Furthermore, because this study is Vitalys' initial exploration into machine learning for scheduling optimization in part 5, more research is needed. The models evaluated here—Random Forest, Gradient Boosting, and MLP—while complex, did not provide enough accuracy to support their use over a well-optimized classical scheduling strategy. The limited performance, caused by a tiny and imbalanced dataset, limited the models' capacity to generalize and make solid predictions. Given the basic structure of this initial model, adding more relevant features, such as real-time data points and thorough patient information, has the potential to dramatically improve prediction power. Future research might look at an enlarged model with more variables, which will certainly improve the usefulness and relevance of machine learning for scheduling in a hospital setting like Vitalys.

The study was also constrained by the scope and quality of the dataset, which primarily contained patient appointment data but lacked critical scheduling factors such as surgeon availability, staff schedules, and real-time updates on hospital resources such as bed availability or operating room occupancy. These features are critical for a completely optimized scheduling model because they have a direct impact on the ability to respond to last-minute changes, such as unexpected cancellations or urgent instances. Integrating these aspects would result in a more realistic, responsive scheduling system that could alter based on both patient demands and hospital capacity. Future study could benefit from creating a multi-layered dataset that incorporates both patient and resource data in order to improve machine learning models' predicted accuracy and the general practicality of a data-driven scheduling method.

Finally, the study's exclusive focus on elective bariatric procedures limited the applicability of findings to other departments or more complex medical situations, where variability and scheduling challenges may be greater. Bariatric surgery is an elective procedure, meaning that patients actively choose to undergo surgery rather than requiring immediate intervention. This planned nature of bariatric procedures typically results in fewer urgent changes compared to other surgical departments where emergency cases or complex conditions necessitate unpredictable adjustments. Consequently, the scheduling insights and improvements identified in this study may not fully extend to departments with higher urgency levels, where variability in patient needs and resource allocation is inherently greater.

To further elaborate on the comparison, while AI-based approaches have the potential to increase flexibility, machine learning models' current limitations—particularly their accuracy and susceptibility to overfitting—indicate that they may not be ready to reliably support Vitalys Clinic's scheduling needs. The observed overfitting and inaccuracy in predictions are due to the dataset's size and unpredictability, which, according to studies, are common issues in healthcare AI applications. These limitations highlight a tradeoff: AI solutions require solid datasets with enough real-time data to produce consistently trustworthy results, which are currently difficult to achieve in this environment. However, developing and maintaining such a dynamic dataset would necessitate a significant investment in data infrastructure, training, and continuing model validation, which may be incompatible with Vitalys Clinic's urgent demands and resources.

Despite the limitations of this study, artificial intelligence remains a promising technique for future schedule optimization in this, or other clinics. AI models may make more accurate forecasts and recommend alternate plans to reduce delays if more precise and real-time data were collected, including as staff availability, real-time surgical updates, and previous patient data. Future editions of the decision support system may include AI-based predictive models that dynamically change schedules depending on incoming data throughout the day. Furthermore, with better data, AI could help with decision-making by detecting trends and abnormalities that clinic staff may not see right away. Clinics that integrate predictive insights with traditional scheduling may benefit from a more dynamic approach to capacity management, reacting to real-time demands and avoiding interruptions caused by unpredictability in patient attendance and appointment durations.

In addition, broadening the research to include other medical departments, such as cardiology, could provide a greater knowledge of scheduling issues throughout the institution. For example, in cardiology, scheduling problems may occur as a result of a wider range of operations and patient needs, resulting in higher unpredictability in visits to different departments and trajectory durations. Thus, while AI may not be immediately required for Vitalys' bariatric scheduling, it has the potential to bring significant value in other sections of the hospital where situations are more unpredictable. Future study that includes other surgical specialties or patient demographics may provide a more complete picture of AI's function in healthcare scheduling. This expanded perspective would also disclose whether the findings from elective surgical scheduling can be applied to more dynamic and unexpected areas of hospital operations, potentially uncovering additional techniques to promote schedule optimization throughout the clinic.

Nonetheless, in research, it is essential to critically evaluate the necessity of AI implementation. For example, at Vitalys, the newly proposed "traditional" scheduling strategy already shows a strong likelihood of success. By focusing on practical enhancements, such as pre-scheduled slots for frequently delayed specialties and refining patient flow protocols, Vitalys could achieve significant efficiency gains without the added complexity of machine learning infrastructure. Why disrupt an already effective scheduling system with artificial intelligence if traditional methods can meet the clinic's needs? Given Vitalys' unique context, the newly proposed traditional scheduling strategy offers the most dependable solution for now, leveraging structured scheduling improvements that fit seamlessly within the clinic's existing framework. Should the clinic's scheduling requirements expand or become more complex, AI could be revisited as a valuable tool. With enhanced data collection and refined predictive capabilities, AI could then serve to further optimize patient flow and resource allocation.

10 Conclusion

The main objective of this study was to investigate how data analytics and artificial intelligence (AI) could improve surgical scheduling and patient flow within the bariatric treatment pathways at Vitalys Clinic. This research aimed to identify specific points of variability and inefficiency in both scheduling and treatment processes. We began by analyzing existing process data to identify bottlenecks and areas where scheduling inefficiencies commonly arise. To map out the planned processes, we developed a ”FRAM-as-imagined” model grounded in established clinical protocols. This model was further refined and tested by interviewing key stakeholders involved in patient care and scheduling, ultimately resulting in the “FRAM-as-Done” model. This combination of protocol analysis and stakeholder input offered a comprehensive view of the current processes and pinpointed specific areas where data-driven improvements could be effectively implemented.

The FRAM-as-Done model provided a detailed analysis of Vitalys Clinic’s bariatric surgery and medical pathways, highlighting areas where workflow improvements could yield substantial efficiency gains. Key points of variability, such as patient screening, registration, and OR planning, emerged as significant contributors to inefficiencies. These stages involve multiple stakeholders and often lack standardized interdepartmental communication, leading to bottlenecks. Through delay analysis, particularly examining delays between key appointments such as screening and surgery, we identified bottlenecks within specific departments and high-impact appointments, laying the foundation for the newly proposed scheduling system.

The analysis of the screening phase at Vitalys Clinic reveals an overall effective patient flow, with most patients progressing through the process without significant delays. Approximately 75% of patients fall within the Standard group, indicating that the majority are scheduled and proceed to surgery within expected timeframes. However, certain specialties do show notable delays, particularly within a subset of outlier patients, which represents around 25% of the total group. These patients face considerable delays, primarily stemming from the screening phase rather than from surgery scheduling itself. This underscores the importance of optimizing the screening process, especially for patients on more complex or non-standard pathways, to improve system-wide efficiency.

Among the specialties analyzed, Mental Healthcare, Gastroenterology, Internal Medicine, and Pain Medicine displayed the most significant delays in both the outliers and total patient groups. However, it is important to note that these cases are largely exceptions rather than indications of systemic scheduling issues within these departments. The relatively low sample sizes in some of these specialties and the sporadic nature of delays suggest that these occurrences are likely specific cases, rather than consistent bottlenecks impacting overall patient flow.

The delay analysis furthered revealed that patients requiring follow-up in external departments such as Gynecology, Psychiatry, Pain Medicine and other departments during group meetings faced prolonged wait times. To address these delays, the proposed model includes dedicated weekly slots: four for postponed patients to alleviate scheduling bottlenecks and two for necessary REDO procedures to avoid conflicts with standard cases. These patients are planned later to accommodate any additional requirements. Standard patients—those who only require core Vitalys services—are scheduled six weeks in advance, minimizing last-minute changes and optimizing resource utilization. This structured, data-informed model has already demonstrated significant improvements in scheduling reliability and patient flow, presenting a viable alternative to an AI-based approach while aligning well with the clinic's existing scheduling framework. Scheduling six weeks in advance reduces the likelihood of rescheduling or delays, creating a more predictable and manageable system for both patients and staff.

To explore the potential for predictive modeling, a Random Forest classifier was applied to assess the dataset's suitability for machine learning. While the model showed some ability to anticipate delays, limitations in performance underscored current constraints. This analysis suggests that the proposed data-enhanced traditional scheduling approach remains the most effective option under present conditions. While AI has transformative potential, this study shows that not every healthcare scheduling difficulty necessitates an AI-based solution. Vitalys Clinic's newly introduced data-driven scheduling strategy has already resulted in considerable improvements in patient flow and operational efficiency. However, AI is still a promising tool for the future. With more complete, real-time data, such as personnel availability, live operational updates, and specific patient pathway insights, AI models could improve delay prediction and support adaptive scheduling changes.

References

Aaseth, J. et al. (2021). “Diets and drugs for weight loss and health in obesity - An update.” In: *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 140, p. 111789. DOI: `10.1016/j.biopha.2021.111789`.

Aissaoui, Najla Omrane et al. (2024). “A BPMN-VSM based process analysis to improve the efficiency of multidisciplinary outpatient clinics”. In: *Production Planning & Control* 35.5, pp. 461–491.

Alnsour, Yazan et al. (2023). “Predicting Patient Length of Stay Using Artificial Intelligence to Assist Healthcare Professionals in Resource Planning and Scheduling Decisions”. In: *J. Glob. Inf. Manag.* 31, pp. 1–14. DOI: `10.4018/jgim.323059`.

Antony, Jiju (2005). “Assessing the Status of Six Sigma Implementation in the UK manufacturing Small and Medium Sized Enterprises (SMEs)”. In: *Second International Six Sigma Conference*.

ATLAS.ti (2024). Accessed: 2024-10-16. URL: `https://atlasti.com`.

Bentéjac, Candice, Anna Csörgo, and Gonzalo Martínez-Muñoz (2019). “A comparative analysis of gradient boosting algorithms”. In: *Artificial Intelligence Review* 54, pp. 1937–1967. DOI: `10.1007/s10462-020-09896-5`.

Bernatchou, Maryam et al. (2017). “Towards a simulation model of an outpatient chemotherapy unit”. In: *2017 International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*. IEEE, pp. 177–182.

Bielen, Frédéric and Nathalie Demoulin (2007). “Waiting time influence on the satisfaction-loyalty relationship in services”. In: *Managing Service Quality: An International Journal* 17.2, pp. 174–193.

Boateng, Veronica A. and Bo Yang (2023). “Ensemble Stacking with the Multi-Layer Perceptron Neural Network Meta-Learner for Passenger Train Delay Prediction”. In: *2023 IEEE Conference on Artificial Intelligence (CAI)*, pp. 21–22. DOI: `10.1109/cai54212.2023.00017`.

Cahill, Naomi E et al. (2010). “Understanding adherence to guidelines in the intensive care unit: development of a comprehensive framework”. In: *Journal of parenteral and Enteral Nutrition* 34.6, pp. 616–624.

ChipSoft (2024). *ChipSoft: HiX - Zorginformatieoplossingen*. Accessed: October 17, 2024. URL: `https://www.chipsoft.com/nl-NL/oplossingen/1`.

Clay-Williams, Robyn, Jeanette Hounsgaard, and Erik Hollnagel (2015). “Where the rubber meets the road: using FRAM to align work-as-imagined with work-as-done when implementing clinical guidelines”. In: *Implementation Science* 10, pp. 1–8.

Colquitt, J. et al. (2014). “Surgery for weight loss in adults.” In: *The Cochrane database of systematic reviews* 8, p. CD003641. DOI: `10.1002/14651858.CD003641.pub4`.

CTCue (2024). *CTCue: Find and use hospital data without programming*. Accessed: October 17, 2024. URL: `https://ctcue.com`.

Cui, Beibei et al. (2021). “Long-term outcomes of Roux-en-Y gastric bypass versus medical therapy for patients with type 2 diabetes: a meta-analysis of randomized controlled trials.” In: *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery*. DOI: 10.1016/j.soard.2021.03.001.

Elreedy, Dina and A. Atiya (2019). “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance”. In: *Inf. Sci.* 505, pp. 32–64. DOI: 10.1016/J.INS.2019.07.070.

Famiglietti, Robin M et al. (2017). “Using discrete-event simulation to promote quality improvement and efficiency in a radiation oncology treatment center”. In: *Quality Management in Healthcare* 26.4, pp. 184–189.

Glowacka, Karolina J, Raymond M Henry, and Jerrold H May (2009). “A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling”. In: *Journal of the Operational Research Society* 60.8, pp. 1056–1068.

Golmohammadi, Davood (2021). “A Decision-Making tool based on historical data for service time prediction in outpatient scheduling”. In: *International Journal of Medical Informatics* 156, p. 104591.

Griffiths, P. et al. (2020). “Beyond ratios - flexible and resilient nurse staffing options to deliver cost-effective hospital care and address staff shortages: A simulation and economic modelling study”. In: *International Journal of Nursing Studies* 117. DOI: 10.1101/2020.11.30.20240945.

Grönroos, S. et al. (2020). “Effect of Laparoscopic Sleeve Gastrectomy vs Roux-en-Y Gastric Bypass on Weight Loss and Quality of Life at 7 Years in Patients With Morbid Obesity: The SLEEVEPASS Randomized Clinical Trial.” In: *JAMA surgery*. DOI: 10.1001/jamasurg.2020.5666.

Hans, Erwin W, Mark Van Houdenhoven, and Peter JH Hulshof (2011). “A framework for healthcare planning and control”. In: *Handbook of healthcare system scheduling*. Springer, pp. 303–320.

Hensher, Martin et al. (1999). “Better out than in? Alternatives to acute hospital care”. In: *Bmj* 319.7217, pp. 1127–1130.

Hollnagel, Erik (2017). *FRAM: the functional resonance analysis method: modelling complex socio-technical systems*. Crc Press.

Hua, Li et al. (2023). “Research on outpatient capacity planning combining lean thinking and integer linear programming”. In: *BMC Medical Informatics and Decision Making* 23.1, p. 32.

Kadri, F. et al. (2014). “Time Series Modelling and Forecasting of Emergency Department Overcrowding”. In: *Journal of Medical Systems* 38, pp. 1–20. DOI: 10.1007/s10916-014-0107-0.

Kam, A. et al. (2021). “Using Lean Six Sigma techniques to improve efficiency in outpatient ophthalmology clinics”. In: *BMC Health Services Research* 21. DOI: 10.1186/s12913-020-06034-3.

Karamanakos, S. et al. (2008). “Weight Loss, Appetite Suppression, and Changes in Fasting and Postprandial Ghrelin and Peptide-YY Levels After Roux-en-Y Gastric Bypass and Sleeve Gastrectomy: A Prospective, Double Blind Study”. In: *Annals of Surgery* 247, pp. 401–407. DOI: 10.1097/SLA.0b013e318156f012.

Kehagias, I. et al. (2011). “Randomized Clinical Trial of Laparoscopic Roux-en-Y Gastric Bypass Versus Laparoscopic Sleeve Gastrectomy for the Management of Patients with BMI < 50 kg/m2”. In: *Obesity Surgery* 21, pp. 1650–1656. DOI: 10.1007/s11695-011-0479-x.

Kern, Christoph et al. (2021). “Big data simulations for capacity improvement in a general ophthalmology clinic”. In: *Graefe’s Archive for Clinical and Experimental Ophthalmology* 259, pp. 1289–1296.

Kleiner, Sibyl and Jean E Wallace (2017). “Oncologist burnout and compassion fatigue: investigating time pressure at work as a predictor and the mediating role of work-family conflict”. In: *BMC health services research* 17, pp. 1–8.

Koushan, Mona, Lincoln C Wood, and Richard Greatbanks (2021). “Evaluating factors associated with the cancellation and delay of elective surgical procedures: a systematic review”. In: *International Journal for Quality in Health Care* 33.2, mzab092.

Kubala, Michael et al. (2021). “Utilizing process improvement strategies to generate clinic templates and improve patient flow in pediatric otolaryngology”. In: *International Journal of Pediatric Otorhinolaryngology* 147, p. 110779.

Lauti, M. et al. (2016). “Weight Regain Following Sleeve Gastrectomy—a Systematic Review”. In: *Obesity Surgery* 26, pp. 1326–1334. DOI: 10.1007/s11695-016-2152-x.

Lee, Alex et al. (2022). “Exploring human factors in the operating room: scoping review of training offerings for healthcare professionals”. In: *BJS open* 6.2, zrac011.

Lenin, RB et al. (2015). “Optimizing appointment template and number of staff of an OB/GYN clinic–micro and macro simulation analyses”. In: *BMC health services research* 15, pp. 1–17.

Li, Hua et al. (2021). “Prediction of Obstetric Patient Flow and Horizontal Allocation of Medical Resources Based on Time Series Analysis”. In: *Frontiers in Public Health* 9. DOI: 10.3389/fpubh.2021.646157.

Li, Ling, Fatou Diouf, and Anjee Gorkhali (2022). “Managing outpatient flow via an artificial intelligence enabled solution”. In: *Systems Research and Behavioral Science* 39.3, pp. 415–427.

Liang, Bohui et al. (2015). “Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic”. In: *International Journal of Production Research* 53.24, pp. 7177–7190.

Lopez, C. et al. (2022). “Artificial Learning and Machine Learning Applications in Spine Surgery: A Systematic Review”. In: *Global Spine Journal* 12, pp. 1561–1572. DOI: 10.1177/21925682211049164.

Mesko, Shane et al. (2022). “Using patient flow analysis with real-time patient tracking to optimize radiation oncology consultation visits”. In: *BMC Health Services Research* 22.1, p. 1517.

Min, Daiki and Yuehwern Yih (2010). “An elective surgery scheduling problem considering patient priority”. In: *Computers & Operations Research* 37.6, pp. 1091–1099.

Mustapha, Adetoun et al. (2016). “Waiting Time is a Wasted Time: Obio Cottage Hospital Outpatient Transit Time Improvement”. In: *SPE African Health, Safety, Security, Environment, and Social Responsibility Conference and Exhibition*. SPE, SPE–183609.

Naderi, B. et al. (2021). “Increased Surgical Capacity without Additional Resources: Generalized Operating Room Planning and Scheduling”. In: *Production and Operations Management*. DOI: 10.1111/POMS.13397.

Najmuddin, AF, IM Ibrahim, SR Ismail, et al. (2010). “A simulation approach: improving patient waiting time for multiphase patient flow of obstetrics and gynecology department (O&G Department) in local specialist centre”. In: *WSEAS Trans Math* 9.10, pp. 778–790.

National Institute for Public Health and the Environment (RIVM) (2021). *Overweight and Obesity*. https://www.rivm.nl/en/overweight. Accessed: 2024-10-04.

Ogulata, S. N. and R. Erol (2003). “A Hierarchical Multiple Criteria Mathematical Programming Approach for Scheduling General Surgery Operations in Large Hospitals”. In: *Journal of Medical Systems* 27, pp. 259–270. DOI: 10.1023/A:1022575412017.

Ordu, Muhammed et al. (2023). “A comprehensive and integrated hospital decision support system for efficient and effective healthcare services delivery using discrete event simulation”. In: *Healthcare Analytics* 4, p. 100248.

Palter, Vanessa N et al. (2020). “Operating room utilization: a retrospective analysis of perioperative delays”. In: *Journal of Gynecologic Surgery* 36.3, pp. 109–114.

Pati, Sukanya et al. (2023). “Obesity and Cancer: A Current Overview of Epidemiology, Pathogenesis, Outcomes, and Management”. In: *Cancers* 15. DOI: 10.3390/cancers15020485.

Pilitsi, Eleni et al. (2019). “Pharmacotherapy of obesity: Available medications and drugs under investigation.” In: *Metabolism: clinical and experimental* 92, pp. 170–192. DOI: 10.1016/j.metabol.2018.10.010.

Poirier, P. et al. (2006). “Obesity and Cardiovascular Disease: Pathophysiology, Evaluation, and Effect of Weight Loss: An Update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease From the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism”. In: *Circulation* 113, pp. 898–918. DOI: 10.1161/CIRCULATIONAHA.106.171016.

Quezada, N. et al. (2016). “Laparoscopic sleeve gastrectomy conversion to Roux-en-Y gastric bypass: experience in 50 patients after 1 to 3 years of follow-up.” In: *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery* 12 8, pp. 1611–1615. DOI: 10.1016/j.soard.2016.05.025.

Raben, Ditte Caroline et al. (2018). “Application of a non-linear model to understand healthcare processes: using the functional resonance analysis method on a case study of the early detection of sepsis”. In: *Reliability Engineering & System Safety* 177, pp. 1–11.

Raj, P. P. et al. (2018). “Evolution and Standardisation of Techniques in Single-Incision Laparoscopic Bariatric Surgery”. In: *Obesity Surgery* 28, pp. 574–583. DOI: 10.1007/s11695-017-3036-4.

Ramkumar, P. et al. (2019). “Deep Learning Preoperatively Predicts Value Metrics for Primary Total Knee Arthroplasty: Development and Validation of an Artificial Neural Network Model.” In: *The Journal of arthroplasty*. DOI: 10.1016/j.arth.2019.05.034.

Richards, Michael R, Jonathan A Seward, and Christopher M Whaley (2022). “Treatment consolidation after vertical integration: evidence from outpatient procedure markets”. In: *Journal of health economics* 81, p. 102569.

Roy, Sumanta, Shanmugam Prasanna Venkatesan, and Mark Goh (2021). “Healthcare services: A systematic review of patient-centric logistics issues using simulation”. In: *Journal of the Operational Research Society* 72.10, pp. 2342–2364.

Rutherford, Patricia A et al. (2017). “Achieving hospital-wide patient flow”. In: *IHI White Paper. Cambridge: Institute for Healthcare Improvement*.

Safdar, Saria, S. Khan, and A. Shaukat (2022). “A Novel Data-Driven Adaptive Technique to Generate a Physician Visiting Schedule for Better Patient Experience”. In: *IEEE Access* 10, pp. 3935–3948. DOI: 10.1109/access.2021.3138460.

Safi, Mariam et al. (2023). "FRAM and LEAN as tools for describing and improving the referral process between outpatient clinics in a Danish Hospital: Complementary or conflicting?" In: *Safety Science* 166, p. 106230.

Sapir-Pichhadze, R. and B. Kaplan (2020). "Seeing the Forest for the Trees: Random Forest Models for Predicting Survival in Kidney Transplant Recipients." In: *Transplantation*. DOI: 10.1097/TP.0000000000002923.

Saunders, K. et al. (2018). "Obesity Pharmacotherapy." In: *The Medical clinics of North America* 102 1, pp. 135–148. DOI: 10.1016/j.mcna.2017.08.010.

Song, Jie, Yaqing Bai, and Jianpei Wen (2018). "Optimal appointment rule design in an outpatient department". In: *IEEE Transactions on Automation Science and Engineering* 16.1, pp. 100–114.

Srinivas, Sharan and A Ravi Ravindran (2018). "Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework". In: *Expert Systems with Applications* 102, pp. 245–261.

Stolle, Catrien (June 2024). *Optimizing Patient Flow in Outpatient Clinics: Evaluating the Impact of Different Methodologies*. Biomedical Engineering Master's Literature Review, TU Delft.

Sujan, M et al. (2023). "Operationalising FRAM in Healthcare: A critical reflection on practice". In: *Safety Science* 158, p. 105994.

Syn, N. et al. (2021). "Association of metabolic–bariatric surgery with long-term survival in adults with and without diabetes: a one-stage meta-analysis of matched cohort and prospective controlled studies with 174 772 participants". In: *The Lancet* 397, pp. 1830–1841. DOI: 10.1016/S0140-6736(21)00591-2.

Tayne, Samantha et al. (2018). "Maximizing operational efficiency using an in-house ambulatory surgery model at an academic medical center". In: *Journal of Healthcare Management* 63.2, pp. 118–129.

Vitalys (2024). *Vitalys Bariatric Clinic*. https://www.vitalys.nl. Accessed: 2024-10-04.

Waghorn, A, M McKee, and J Thompson (1997). "Surgical outpatients: challenges and responses". In: *British Journal of Surgery* 84.3, pp. 300–307.

Wojtys, E. et al. (2009). "Applying Lean Techniques to Improve the Patient Scheduling Process". In: *Journal For Healthcare Quality* 31, pp. 10–16. DOI: 10.1111/j.1945-1474.2009.00025.x.

World Health Organization (2021). *Obesity and Overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed: 2024-10-04.

Wright, Marvin N. and I. König (2019). "Splitting on categorical predictors in random forests". In: *PeerJ* 7. DOI: 10.7717/peerj.6339.

Xu, Chuhan, Pablo Coen-Pirani, and Xia Jiang (2023). "Empirical Study of Overfitting in Deep Learning for Predicting Breast Cancer Metastasis". In: *Cancers* 15. DOI: 10.3390/cancers15071969.

Zerprize (2024). *FMV*. Accessed: October 4, 2024. URL: https://fmv.zerprize.co.nz.

Zou, Chengye, Junwei Wang, and Yao Cheng (2022). "The impact of the variability of patient flow and service time on the efficiency of large-scale outpatient systems". In: *IEEE Transactions on Computational Social Systems* 10.3, pp. 1230–1240.

A

Interview Section

A.1. Interview Guide for Stakeholders at Vitalys Clinic

This interview guide was used for conducting semi-structured interviews with stakeholders at Vitalys Clinic. It is important to note that, as these interviews are semi-structured, the questions may deviate based on the interviewee's responses and the flow of the conversation. The primary goal of these interviews is to create a Work-as-Done diagram and to identify sources of variability, as well as the drivers and barriers within the clinic's operational processes.

A.1.1. General Introduction

Thank you for taking the time to speak with me today. My name is Catrien, and I am conducting research as part of my internship at Vitalys and Johnson & Johnson MedTech. The aim of my project is to explore how the planning workflow at Vitalys Clinic can be optimized and how data can potentially support this process. I would like to understand the processes you are involved in and where you believe there is room for improvement.

Before we begin, have you received and signed the informed consent form? If not, I can provide you with another copy. Also, would it be okay if I record this interview? The recording will be used for transcription purposes only, and all data will be anonymized before the recording is deleted.

Thank you for your time and cooperation.

A.1.2. Introduction (2 minutes)

In this brief section, I will introduce the purpose of the research and get to know the interviewee. This part includes understanding the interviewee's role within Vitalys Clinic.

- Can you tell me your name and your role at Vitalys Clinic?
- How long have you been working in this role?
- What is your primary responsibility within the clinic?

A.1.3. The Patient Journey (20 minutes)

The goal of this section is to understand the interviewee's role in the treatment process at Vitalys Clinic. We will use the Work-as-Imagined FRAM diagram to guide this discussion and clarify their tasks and responsibilities.

1. What are the main tasks you carry out in this stage of the process (the stage they pointed out in the diagram)?
2. What factors or events trigger the start of your work in this step? (e.g., patient data, communication from another department)

3. Are there any protocols or guidelines that dictate the actions you take during this step? If so, could you describe them? (Control)
4. Are there any prerequisites or conditions that need to be met before you can start working on this step? (Preconditions)
5. When do you consider this step completed, and what might delay the completion? (Output)
6. What resources (e.g., software, equipment, support) do you rely on to perform your tasks? How do you handle situations where these resources are unavailable? (Resources)
7. How do you manage unexpected events, such as last-minute changes, urgent tasks, or missing resources? (Variability)
8. How often do you have to make adjustments to the standard process? What factors typically cause these adjustments? (Variability)

A.1.4. Overall Process Evaluation (20 minutes)

This section will focus on identifying inefficiencies and potential improvements within the clinic's operations. The goal is to gather the interviewee's insights on how processes can be made more efficient and what data-driven solutions they envision.

1. Looking at the entire treatment journey, what do you see as the most challenging or inefficient part of the process?
2. How do these inefficiencies impact your work, the workflow of others, and patient care?
3. Have you encountered specific bottlenecks or delays during the planning or scheduling phases? If so, what are the causes?
4. What changes or improvements would you suggest to overcome these inefficiencies?
5. In your opinion, could data analytics or AI help address these issues? If yes, in what way?
6. Are there any processes or parts of the workflow that you believe work particularly well? What makes them effective?
7. Do you think there are aspects of the patient journey or operational processes that I haven't covered in this interview?
8. Do you have any suggestions for literature or topics I should explore to further understand these issues?

Thank the interviewee again for their time and valuable insights, and remind them that their feedback will be critical in improving operational processes at Vitalys Clinic.

B

Quotation Labels

This Appendix Chapter shows all the Quotations that are used to measure the Variability, but also the quotes that have been used for the Drivers and Barriers.

B.1. Variability between Phases

The quotations below are used to measure the variability between the different phases. Where is what the most?

Table B.1: Quotations and Drivers of Stakeholders at Vitalys Clinic

Who?	**Quotation**	**Phase**	**Type of Variability**
Administrative Office	Yes, we have a handbook that we follow. We only deviate from it when there are no available slots, causing delays.	Medical	Resource Constraint
Nurse Specialist	Sometimes there is doubt. It seems very black and white, all the regulations, but sometimes there is a gray area. A gray area takes more time and sometimes requires additional consultation.	Medical	Dynamic Factor
Nurse Specialist	Patients may not send the required reports, or they might not have completed a year of GLI. We wait until they complete it. Some patients register prematurely or plan their appointments too early, thinking they can start right away.	Medical	Dynamic Factor
Internist	Patients diverge if they have side effects or have questions in between. Also, sometimes patients want to switch to a different medication if they qualify for it.	Medical	Dynamic Factor
Internist	Sometimes people need to work on their stress or they don't lose enough weight. But if you lose 4% of your weight but your waist circumference decreases by 10 cm, you don't quite meet the criteria, but you should have another check-up after 16 weeks. I think the program works for about 70% of people, and for 30% it doesn't because people are different.	Medical	Dynamic Factor

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
Internist	The entire process can extend beyond two years due to delays, patient rescheduling, or changes in medication.	Medical	Resource Constraint, Dynamic Factor, Inefficient Scheduling
Administrative Office	Interviewer: Does every consultation need to be with an internist or just the first one? INTERVIEWEE: No, no, certain ones. Okay. So, there are some that can start directly with the internist or must start there for a reason. If there are no particular issues, it can generally be with the VS.	Medical	Dynamic Factor
Administrative Office	Just like the intake for non-insured patients, while insured patients' intake is done by the nurse specialist or the internist.	Medical	Inefficient Scheduling
Administrative Office	Yes, because, well, I think our appointments for the non-insured trajectory can be scheduled quickly, but these people still need another appointment with the nurse specialists. And that currently takes about 6 to 8 weeks or even 2 to 3 months before they can actually start the medication because of the high demand.	Medical	Inefficient Scheduling
Nurse Specialist	At Vitalys, it's very customer-centric, and any issue is quickly escalated to the doctor. Nurses sometimes doubt themselves unnecessarily, leading to more calls for the doctor. I think nurses need to feel more confident and recognize their importance. Some calls shouldn't go directly to the doctor. The planning and capacity are good; secretaries plan well.	Medical	Dynamic Factor
Administrative Office	Planning is the most challenging part, especially managing capacity and avoiding long wait times.	Medical	Inefficient Scheduling
Administrative Office	It's not just about scheduling information sessions and sending orders, but also thinking ahead for diabetes patients and planning their follow-ups.	Phase 1	Dynamic Factor, Administrative Work
Administrative Office	Yes, our work is constantly being adjusted. Yesterday we had adjustments in orders again, to make it as clear as possible for everyone.	Phase 1	Administrative Work
Administrative Office	You have to rely on what is written by the dietitians, the screening doctor, and the psychologists.	Phase 1	Dynamic Factor, Administrative Work
Administrative Office	An orange flag indicates that there are issues to work on, such as seeing a psychiatrist or psychologist or being discussed in a meeting. This is usually handled by the person who conducted the screening, often a nurse specialist or doctor.	Phase 1	Dynamic Factor
Administrative Office	With medication, the pathway is small and well-defined, but with the surgical clinic, it's broader.	Phase 1	Inefficient Scheduling

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
Administrative Office	If there are fewer screenings, we sometimes wait for patients. But it's picking up again. I saw an increase in referrals and registrations recently.	Phase 1	Dynamic Factor
Administrative Office	Look, if a screening is not approved, for whatever reason, 9 out of 10 times it's a deferred advice.	Phase 1	Dynamic Factor
Administrative Office	You'd be surprised how many orange flags we have, such as dietitian reassessments after three months. This accounts for about 20% of our screened patients.	Phase 1	Dynamic Factor
Receptionist	Yes, it's actually the data indeed that it is properly transferred. Yes, because if that's not correct... Does everything go by phone? Everything goes by phone. Oh yes. Quite difficult sometimes, but occasionally, to ensure that you understand correctly and so on... It is good to verify.	Phase 1	Administrative Work
Receptionist	We can handle some things ourselves, general questions. But we also regularly need to transfer calls.	Phase 1	Resource Constraint
Administrative Office	We always print the order and check everything—screening date, height, weight, insurance, diabetes status. We use HIX for planning and sending orders to anesthesia, and an Excel file for tracking. We have a separate file for the OR preliminary month planning.	Phase 1	Administrative Work
Surgeon	As surgeons, we are involved in the preoperative process by reviewing the screening letters. We assess the screening letters and decide if we agree with the screening. I have to say that it is quite difficult to make a judgment without having seen the patient personally.	Phase 1	Dynamic Factor, Administrative Work
Receptionist	Because sometimes we encounter things that are not correct. Or that the phone number has not been properly verified. Yes, super important information, of course.	Phase 1	Administrative Work
Administrative Office	What often delays is when we need to request medical records. If someone is deferred because the doctor wants to check first, what for example in the past surgeries, you have to ask for that, but you always have a consent form that you need to send.	Phase 1	Administrative Work
Surgeon	In the screening process, it is crucial that when a red flag arises, someone must take ownership of that problem and ensure it is followed up on. In this specific case, there were several alarm signs, but they were not documented or acted upon. I believe that during the screening, if there is a red flag, a problem owner should emerge to follow up on it.	Phase 1	Dynamic Factor
Group Planner	The administrative office often mentions when someone is on vacation. Then we make sure not to schedule an appointment during their vacation.	Phase 1, 2	Dynamic Factor

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
Surgeon	Patients who come to the information day should ideally have a green light without any issues. Then it's perfectly possible that they also have a preoperative trajectory planned. Ideally, the patient shouldn't come to the information session until everything is green-lighted. You would then have the same group with three groups coming that day, and you would have 24 to 30 patients, and then you have the individual patients in another session.	Phase 1, 2	Inefficient Scheduling
Nurse Specialist	And sometimes there's some confusion, and sometimes there are two questions in one, like both obesity analysis and medication trajectory, and we need to figure that out together. This process is still quite new.	Phase 1, Medical	Dynamic Factor
Group Planner	Then the patients have given their preferred time. And then they say, well, maybe I shouldn't have chosen that time. That happens sometimes, but that can happen to anyone. Then we start over again. So, you look again at where there is space.	Phase 2	Dynamic Factor
Group Planner	The biggest disruption can be if a practitioner is sick, and we need to switch them out. We try to keep the group sessions as scheduled to avoid rescheduling patients.	Phase 2	Dynamic Factor
Group Planner	If a patient withdraws from the pathway for any reason, we need to ensure they are assigned a new preliminary OR month with the administrative office. This is probably the biggest factor.	Phase 2	Dynamic Factor
Group Planner	We used to have a buffer, but we don't, so we plan people later, so they might join groups later. It's a bit more challenging now.	Phase 2	Dynamic Factor
Group Planner	We check if they want to be scheduled in that month. Summer can be tricky due to vacations, and patients need to decide if they want their procedure before or after their vacation.	Phase 2	Dynamic Factor
Group Planner	But we have an agreement that people can miss one appointment in the pre-operative trajectory. If they miss two, the operation does not go through and then it depends on whether we can still schedule one in between. It also depends on the motivation of course. We really have to try our best. But it is just nice if the operation is scheduled in that month so that it can continue.	Phase 2	Dynamic Factor
Group Planner	But the planning mainly affects dietitians and psychologists. We manage their vacation requests, so we can anticipate when replacements are needed.	Phase 2	Inefficient Scheduling
Group Planner	INTERVIEWER: And how often do you change or adjust the schedule? Does that happen often? INTERVIEWEE: Well, one week quite often, another week you think, oh well, it's actually running smoothly. Yes, it really varies per... Yes, exactly. But it still happens quite often. If a doctor is absent, then you also have to reschedule.	Phase 2	Dynamic Factor, Administrative Work

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
Surgeon	Another issue is that sometimes patients come to the information session with a fully planned trajectory, while others have nothing planned. It would be ideal if all patients attending the information session had their preoperative trajectory planned.	Phase 2, 3	Inefficient Scheduling
OR Planner	INTERVIEWER: So, for example, that a surgery needs to be done with two surgeons, or what kind of surgery? INTERVIEWEE: Yes, that comes from the information session. Sometimes from the screening, but usually from the information session. Then the surgeon assesses whether it is still a procedure with extra time. Or still with two surgeons or whatever.	Phase 3	Dynamic Factor
OR Planner	INTERVIEWER: Yes, do you see many patients who stay longer than a day? Actually, they are all day patients, right? INTERVIEWEE: Yes, that is preferred, but not everyone wants that. There are also many patients who are not allowed, cannot because of medication, traveling distance, or alimentation. And they also put the special notes of the patient.	Phase 3	Dynamic Factor
OR Planner	A lot always comes out of the information session that you need to process. For example, whether patients are admitted for day surgery. Or if they have to stay overnight. Or if further investigation needs to be done through a gastroscopy. Then you have to take action again.	Phase 3	Dynamic Factor, Administrative Work
OR Planner	INTERVIEWER: Do unexpected things happen with this planning? I think, yes. INTERVIEWEE: If an OR does not go through. Yes, and that can be because a surgeon is sick. Or an overlap from another. That something comes in between. Yes. INTERVIEWER: And then the patient is often admitted, of course. Yes. And then, what happens next? Do they get priority over other patients? INTERVIEWEE: Well, that is then the advantage of planning 14 days in advance.	Phase 3	Dynamic Factor
OR Planner	I am usually the check, check, goal, check person. There are a lot of checks in the OR planning process. Because not everyone can always think of the same thing, of course. No, true. And it's also annoying if a patient is scheduled for surgery and it can't go through. For whatever reason. Because the lab has not yet approved yet or something else.	Phase 3	Inefficient Scheduling
OR Planner	Then you don't say after a week... I don't know if that is an idea, I think it is, but then you also have to look if you can fill those weeks, right? That you plan all over the place. Then maybe you have gaps. In the OR week. Yes, because you want to have a day with six patients.	Phase 3	Inefficient Scheduling

Continued on next page

Table B.1 – continued from previous page

Who?	**Quotation**	**Phase**	**Type of Variability**
OR Planner	INTERVIEWER: Yes, but basically Tuesday is the day, every Tuesday, that the list is sent for then two weeks later. And that always happens in the hospital or not? INTERVIEWEE: Yes, that is a rule from Rijnstate. Some other departments are three weeks in advance. But usually, two weeks. I know from surgery. When I worked in admission planning, it was only one week.	Phase 3	Inefficient Scheduling
OR Planner	INTERVIEWER: And how often do you change or adjust the planning? So actually if people drop out? INTERVIEWEE: Yes, in principle, never. In principle, it just stays like that. Look at what I did in week 7. Week 7, yes. Let's see. Oh yes, well I had already scheduled this one, but I was going to call. And he said, I don't want this week at all; I want from July 18 or something. Yes. Yes, then you can put it back, back on the waiting list. Yes. And cancel again and change all lists again. A lot of work. Yes. And find new ones again.	Phase 3	Inefficient Scheduling
OR Planner	A lot of documentation needs to be collected, including details like whether the patient requires day surgery or overnight admission. This level of detail, while necessary, creates inefficiencies and delays, especially when combined with unexpected changes.	Phase 3	Administartive Work, Inefficient Scheduling
OR Planner	Sometimes they get called up, you are not yet approved by the others, but... All those negatives, that the surgery can't go ahead. Yes, that you have that as a kind of side note... Yes, that they also always know in the letter that it is well stated. These things should actually be arranged upfront. Yes, actually, you should only be scheduled after the group things if you need to go to the information session.	Phase 3	Inefficient Scheduling
OR Planner	But yes, so many changes still. And then, well, the nurse practitioner issues an order per January. Vacation starts on and with July 7. And then per July, they make the new OR waiting list again, and they put it back in the OR waiting list, vacation from July 1 to 9. Then we go again, so you get mistakes, right? Yes, because it can be so much leaner. Yes, and then there are just no mistakes made.	Phase 3	Inefficient Scheduling
OR Planner	Nowadays, we have a lot of teachers who want to be operated on during the holidays. In the Netherlands, the holiday is different in the north than in the south. It varies quite a bit. Yes, that is another thing I didn't expect to happen in such a scenario.	Phase 3	Resource Constraint
OR Planner	Not necessarily, but also, sometimes you need to see the internist. Yes, but you only find out you need to see the internist from the information session, of course. They have a waiting time of four months, I believe, the internist.	Phase 3	Dynamic Factor

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
OR Planner	INTERVIEWER: And what happens then? Does the patient still join the group phase, or do they go to another group? INTERVIEWEE: They go to another group. And people are not happy with that. They finally took the step to join the process, but then they have a 4-month delay because they need to see the internist. Or the MDL with 12 weeks, 11, 12 weeks waiting time. It is not normal what the waiting times are.	Phase 3	Dynamic Factor
OR Planner	Sometimes it's a redo and what kind of surgery they had before. It also depends on which waiting lists we have to deal with.	Phase 3	Dynamic Factor
OR Planning	INTERVIEWER: Yes, so you get that information from HICS or something? INTERVIEWEE: Surgeon head? No, that's from another program. We have to figure that out themselves.	Phase 3	Administrative Work
Nurse Specialist	INTERVIEWER: How often does a patient come back to the clinic with complaints? INTERVIEWEE: I think that varies too much to estimate.	Phase 4	Dynamic Factor
Nurse Specialist	Sometimes, orders are not placed, so someone is overlooked and not called.	Phase 4	Administrative Work
Nurse Specialist	INTERVIEWER: Yes, because it's not, yes, yes, no, everything stays anonymous, so you can be honest. But do the clinics not get adjusted despite having many long-term sick employees? INTERVIEWEE: Well, the clinics are closed because of this, but the influx of patients is still there. For example, surgical patients with issues, so you need to accommodate them. Our managers are strict about not overbooking, but it happens sometimes.	Phase 4	Resource Constraint
Nurse Specialist	Because the secretaries want to fit the patients in, and there are just not enough slots, or due to staff shortages, then they feel they have to fit them in somewhere, so it gets overbooked.	Phase 4	Resource Constraint
Group Planner	We depend on the medical team's schedule, which we don't control. There's a lot of shifting due to staff illnesses.	Phase 4, 5	Dynamic Factor
Group Planner	Yes, if someone reschedules, it requires a lot of adjustment. We're working on improving this with more experienced staff.	Phase 5	Inefficient Scheduling, Administrative Work
Nurse Specialist	Sometimes there's a planning issue with patients needing a swallow test, and they get scheduled for a consultation before the test results are back. So you end up rescheduling again.	Medical	Inefficient Scheduling
Nurse Specialist	I was thinking about my surgical clinics. I often have patients who need to do a blood test and haven't done it, so I can't do anything with that slot. Or if I ordered a stool test and it wasn't sent correctly by the secretariat, it delays everything. So we end up chasing the facts and have to reschedule everything.	Phase 2, 4	Inefficient Scheduling

Continued on next page

Table B.1 – continued from previous page

Who?	Quotation	Phase	Type of Variability
OR Planner	We use a large Excel file with a list per group. We print these lists and manually plan each patient. It's quite old-fashioned.	Phase 1, 2, 3, 4	Inefficient Scheduling
OR Planner	But there were times when I had planned the surgery, and then they still had to go to the information session. That is tricky. Yes, that is difficult.	Phase 3	Inefficient Scheduling
Group Planner	Occasionally, some get canceled if not needed. The biggest disruption can be if a practitioner is sick, and we need to switch them out.	Phase 2, 4	Dynamic Factor

B.2. Drivers

B.2.1. Drivers Phase 1

Table B.2: Quotations of Drivers at Phase 1 of Stakeholders at Vitalys Clinic

Who?	Quotation	Phase	Driver
Receptionist	The work is a continuous flow. Yes. Nothing is left for three days.	Phase 1	Work is always done
Administrative Office	We do it together now, and she has created a handbook for the entire secretariat. When there are updates, they are incorporated.	Phase 1	Handbook for the protocol
Receptionist	It runs smoothly, it seems everyone is clear on the steps.	Phase 1	Work is always done
Internist	The bariatric program has a nice digital setup with apps asking how far you've come with your 300 minutes, what's going well, what's not. I often have to suggest activities like swimming to patients, which they hadn't thought of.	Phase 1,2,3,4,5	Finding the solutions

B.2.2. Drivers Phase 3

Table B.3: Quotations of Drivers at Phase 3 of Stakeholders at Vitalys Clinic

Who?	Quotation	Phase	Driver
Internist	The bariatric program has a nice digital setup with apps asking how far you've come with your 300 minutes, what's going well, what's not. I often have to suggest activities like swimming to patients, which they hadn't thought of.	Phase 1,2,3,4,5	Finding the solutions
OR Planner	Interviewer: Yes, so the surgeon also shifts a bit with his own schedule. Sandra: Yes, or sometimes a fellow takes over an operation on that day, you know. Yes. Last time Bart Witteman was sick. I didn't even know, but everything continued. Oh wow. They just found a replacement themselves, actually.	Phase 3	Finding the solutions
OR Planner	Interviewer: Are these tasks always followed up properly? Yes, I think so. Yes, followed up... Yes, more that if someone else takes over when you are not there. Yes, so someone is responsible every day.	Phase 3	Work is always done

Continued on next page

Table B.3 – continued from previous page

Who?	Quotation	Phase	Driver
OR Planner	INTERVIEWER: And how often do you change or adjust the planning? So actually if people drop out? INTERVIEWEE: Yes, in principle, never because it is 2 weeks in advance. In principle, it just stays like that.	Phase 3	Advantage of Short Planning

B.3. Barriers

B.3.1. Barriers Phase 1

Table B.4: Quotations of Barriers at Phase 1 of Stakeholders at Vitalys Clinic

Who?	Quotation	Phase	Barrier
Receptionist	Yes, it's actually the data indeed that it is properly transferred. Yes, because if that's not correct... Does everything go by phone? Everything goes by phone. Oh yes. Quite difficult sometimes, but occasionally, to ensure that you understand correctly and so on... It is good to verify.	Phase 1	Administrative Work
Administrative Office	We always print the order and check everything—screening date, height, weight, insurance, diabetes status. We use HIX for planning and sending orders to anesthesia, and an Excel file for tracking. We have a separate file for the OR preliminary month planning.	Phase 1	Administrative Work
Surgeon	As surgeons, we are involved in the preoperative process by reviewing the screening letters. We assess the screening letters and decide if we agree with the screening. I have to say that it is quite difficult to make a judgment without having seen the patient personally.	Phase 1	Additional Consultation
Receptionist	Because sometimes we encounter things that are not correct. Or that the phone number has not been properly verified. Yes, super important information, of course.	Phase 1	Administrative Work
Receptionist	No, the issue is usually with the letter. It needs to be created by the specialist and supervised by the surgeon. We sometimes have to wait two to three weeks for this.	Phase 1	Communication between colleagues
Administrative Office	What often delays is when we need to request medical records. If someone is deferred because the doctor wants to check first, what for example in the past surgeries, you have to ask for that, but you always have a consent form that you need to send.	Phase 1	Delay in Process
Surgeon	In the screening process, it is crucial that when a red flag arises, someone must take ownership of that problem and ensure it is followed up on. In this specific case, there were several alarm signs, but they were not documented or acted upon. I believe that during the screening, if there is a red flag, a problem owner should emerge to follow up on it.	Phase 1	No Documentation
Group Planner	I find the total file in Excel often causes many problems.	Phase 1, 2, 3, 4	Tools

Continued on next page

Table B.4 – continued from previous page

Who?	Quotation	Phase	Barrier
Surgeon	There are only some examination rooms with a couch, and I find it very annoying that we don't have windows.	Phase 1, 2, 4, 5	Resource Constraints
OR Planner	We use a large Excel file with a list per group. We print these lists and manually plan each patient. It's quite old-fashioned.	Phase 1, 2, 3, 4	Administrative Work

B.3.2. Barriers Phase 3

Table B.5: Quotations of Barriers at Phase 3 of Stakeholders at Vitalys Clinic

Who?	Quotation	Phase	Barrier
Group Planner	I find the total file in Excel often causes many problems.	Phase 1, 2, 3, 4	Tools
Surgeon	Another issue is that sometimes patients come to the information session with a fully planned trajectory, while others have nothing planned. It would be ideal if all patients attending the information session had their preoperative trajectory planned.	Phase 2, 3	Inefficient Scheduling
OR Planner	Sometimes they get called up, you are not yet approved by the others, but... All those negatives, that the surgery can't go ahead. Yes, that you have that as a kind of side note... Yes, that they also always know in the letter that it is well stated. These things should actually be arranged upfront. Yes, actually, you should only be scheduled after the group things if you need to go to the information session.	Phase 3	Inefficient Scheduling
OR Planner	But yes, so many changes still. And then, well, the nurse practitioner issues an order per January. So I have from this Magdor in the English group. Vacation starts on and with July 7. Yes. And then per January, they make the OR waiting list again, and they put it back in the OR waiting list, vacation from July 1 to 9, yes, for example, and then we go again, so you get mistakes, right? Yes, because it can be so much leaner, do you understand? Yes, leaner, it can be so much leaner, it is really so. Yes, and then there are just no mistakes made.	Phase 3	Inefficient Scheduling
OR Planner	Not necessarily, but also, sometimes you need to see the internist. Yes, but you only find out you need to see the internist from the information session, of course. They have a waiting time of four months, I believe, the internist.	Phase 3	Waiting List
OR Planner	INTERVIEWER: And what happens then? Does the patient still join the group phase, or do they go to another group? INTERVIEWEE: They go to another group. And people are not happy with that, right? They finally took the step to join the process, but then they have a 4-month delay because they need to see the internist. Or the MDL with 12 weeks, 11, 12 weeks waiting time. Yes, it is not normal what the waiting times are, right?	Phase 3	Waiting List

Continued on next page

Table B.5 – continued from previous page

Who?	Quotation	Phase	Barrier
OR Planner	INTERVIEWER: Yes, so you get that information from HICS or something? INTERVIEWEE: Surgeon head? No, that's from another program. We have to figure that out themselves.	Phase 3	Administrative Work
OR Planner	INTERVIEWER: When do you start all these steps? So if they are scheduled in the group planning? Sandra: Yes, we get the target groups from the process planners. And then I look at the whole overview, what there is, I take out all the patients, which group they are in. I make an extraction, I put it in a folder. And then I completely work that out.	Phase 3	Administrative Work
OR Planner	INTERVIEWER: Yes, and where is all this written down? Sandra: In the OR waiting list. In the OR waiting list, so in that document, in that folder.	Phase 3	Administrative Work
OR Planner	It is difficult to say, difficult to speak. I would make a day free anyway. But you don't know what the reasons could be. Yes, but you don't know when it will take place. You get called 14 days in advance on a Tuesday for the surgery date two weeks later. Yes, I find that so... At some point, it is inconvenient, and then work or with children.	Phase 3	Late Operating Date
OR Planner	I think we generally plan well ahead, but it would help if an operation date is known earlier so we can prescribe medication earlier for admission.	Phase 3	Late Operating Date
OR Planner	INTERVIEWER: And you get that information passed on to you? Interviewee: Yes, we get a list from the nurse who was at the information session. Via email, and then we work that out in another document where we write it down.	Phase 3	Administrative Work
OR Planner	Yes, but now it is double work, right, because I put the changes in the document for the OR waiting list, and I put it in another document. Yes, that can maybe also be leaner, but I wouldn't know how.	Phase 3	Administrative Work
OR Planner	The amount of paperwork is the most inefficient.	Phase 3	Administrative Work
OR Planner	We use a large Excel file with a list per group. We print these lists and manually plan each patient. It's quite old-fashioned.	Phase 1, 2, 3, 4	Administrative Work
OR Planner	Yes. Also, sometimes in terms of communication. Yes. Yes. Because it starts, well, some say it already at the screening when they have vacation. Yes. Or that they can't or don't have. Yes. That they only want next year. Yes. So to speak. And then it comes to me, and no one knows it anymore.	Phase 3	Communication between Colleagues
OR Planner	But there were times when I had planned the surgery, and then they still had to go to the information session. That is tricky. Yes, that is difficult.	Phase 3	Inefficient Scheduling

C

Database Used for the Analysis

This appendix presents examples of the data tables used for the analysis. These include the combined surgeon and appointment data as well as demographic data for patients. The tables shown represent the initial structure of the data used in the delay analysis in Part 3 and Part 4 of the thesis.

C.1. Appointment and Surgeon Data

Table C.1: Example Database Combined Surgeon and Appointment Data Used for Delay Analysis (Part 3 and Part 4 of the Thesis)

ID	Clustering ID	Specialism	Location	Appointment Start Date	Operation Date	Type of Surgery
0..	Screening	Chirurgie	Hospital	2023-01-17 12:00:00	2023-08-21 00:00:00	Bypass
0..	Herbeoordeling	Diëtetiek	Hospital	2023-05-09 11:45:00	2023-08-21 00:00:00	Bypass
0..	Screening	Chirurgie	Hospital	2023-05-23 09:30:00	2023-08-21 00:00:00	Bypass
0..	Pos_screening	Anesthesie	Hospital	2023-07-05 09:40:00	2023-08-21 00:00:00	Bypass

C.2. Demographic Data of Patients

This section provides a sample of the patient demographic data used in the preoperative screening analysis. This data was essential for random forest analysis, to understanding the factors related to patient flow and delays in the process and see whether delays could be predicted.

Table C.2: Sample Patient Data from Preoperative Screening Analysis

ID	Specialisms	Gender	Age	Height (cm)	Weight (kg)	BMI	Lead Time (days)	HBP	Psychological Help	Obesity	Diabetes	Smoker	Other Departments
0..	Chirurgie, Diëtetiek, Anesthesiologie	M	54	187	138	39.46	41	1	0	1	0	0	0
0..	Chirurgie, Vitalys, Anesthesiologie	F	24	167	104	37.29	32	1	0	1	0	0	1
0..	Chirurgie, Vitalys	F	59	165	113	41.51	140	1	0	1	0	0	1
0..	Chirurgie, Anesthesiologie	F	61	173	128	42.77	45	0	0	1	0	0	1
0..	Chirurgie, Vitalys	F	33	164	119	44.24	37	0	0	1	0	0	1

D

Python Codes used for the research

The following codes were used to support the analyses in this thesis: filtering data, analyzing screening and OR delays, developing the planning model with sub-analyses, and applying a Random Forest model for predictive purposes.

D.1. Filtering Code

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#-----------------------------------------------------------
# Step 1: Load Data and Basic Descriptive Statistics of 2022 and 2023
#-----------------------------------------------------------

# Define file paths
file_path = '/Users/catrienstolle/Desktop/Data Set Thesis/merged_sheets_operations_new.xlsx'
appointments_file_path = '/Users/catrienstolle/Desktop/Data Set Thesis/merged_dataset_final_new.xlsx'

# Load datasets
overview_df = pd.read_excel(file_path, sheet_name='Merged sheet')
operations_df = pd.read_excel(file_path, sheet_name='Merged_Operations_Sheet')
data_df = pd.read_excel(appointments_file_path, sheet_name='Sheet1')

# Convert 'BMI' to numeric, handling invalid values
overview_df['BMI'] = pd.to_numeric(overview_df['BMI'], errors='coerce')

# Calculate basic statistics
average_bmi = overview_df['BMI'].mean()
average_age = overview_df['age'].mean()
gender_distribution = overview_df['gender'].value_counts()

# Count patients with and without screening
total_patients = len(overview_df)
patients_with_screening = overview_df[overview_df['screening_count'] > 0].shape[0]
patients_without_screening = total_patients - patients_with_screening

# Count patients with screening who had an operation vs. no operation
patients_with_screening_and_operation = overview_df[
    (overview_df['screening_count'] > 0) & (overview_df['afspraken_na_screening_preop'] > 0)
].shape[0]
patients_with_screening_no_operation = overview_df[
    (overview_df['screening_count'] > 0) & (overview_df['afspraken_na_screening_preop'] == 0)
].shape[0]

# Average number of appointments for patients with both screening and operation
average_appointments_screening_operation = overview_df[
    (overview_df['screening_count'] > 0) & (overview_df['operaties'] > 0)
```

```
]['afspraken_na_screening_preop'].mean()

# Display the descriptive statistics
text_summary = f"""
Descriptive Statistics Overview:

1. Average BMI: {average_bmi:.2f}
2. Average Age: {average_age:.2f} years
3. Gender Distribution:
   - Females: {gender_distribution.get('F', 0)}
   - Males: {gender_distribution.get('M', 0)}
4. Total Patients: {total_patients}
5. Patients with Screening: {patients_with_screening} ({(patients_with_screening/total_patients)*100:.2f}%)
6. Patients without Screening: {patients_without_screening} ({(patients_without_screening/total_patients)*100:.2f}%)
7. Patients with Screening and Operation: {patients_with_screening_and_operation} ({(patients_with_screening_and_operation/patients_with_screening)*100:.2f}%)
8. Patients with Screening but No Operation: {patients_with_screening_no_operation} ({(patients_with_screening_no_operation/patients_with_screening)*100:.2f}%)
9. Average Number of Appointments for Patients with Screening and Operation: {average_appointments_screening_operation:.2f} appointments
"""

print(text_summary)

#----------------------------------------------------------
# Step 2: Analyze Registration Form Submission Dates
#----------------------------------------------------------

# Convert 'aanmeldformulier_date' to datetime
overview_df['aanmeldformulier_date'] = pd.to_datetime(overview_df['aanmeldformulier_date'], errors='coerce')

# Plot registration form submission dates
plt.figure(figsize=(10, 6))
overview_df['aanmeldformulier_date'].dt.date.value_counts().sort_index().plot(kind='line', color='blue')
plt.title('Distribution of Registration Form Submission Dates')
plt.xlabel('Date')
plt.ylabel('Number of Submissions')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()


#----------------------------------------------------------
# Step 4: Data Cleaning and Duplicate Removal
#----------------------------------------------------------

# Remove duplicates based on 'pseudo_id', 'clustering_id', and 'afspraken_na_screening_alle_appointment_start_date_x'
df_cleaned = data_df.drop_duplicates(subset=['pseudo_id', 'clustering_id', 'afspraken_na_screening_alle_appointment_start_date_x'])

# Checking how many duplicates were removed
removed_duplicates = len(data_df) - len(df_cleaned)
print(f"Number of duplicates removed: {removed_duplicates}")

# Saving the cleaned dataframe to a new Excel file
output_cleaned_path = '/Users/catrienstolle/Desktop/data_cleaned_by_subset_doorloop_analyse.xlsx'
df_cleaned.to_excel(output_cleaned_path, index=False)

# Use the cleaned data for further analysis
data_df = df_cleaned

#----------------------------------------------------------
# Step 5: Count of Unique Patients
#----------------------------------------------------------

```

```python
unique_patient_count = data_df['pseudo_id'].nunique()
print(f"Total unique patients in the dataset: {unique_patient_count}")

#----------------------------------------------------------
# Step 9: Filter and Save Common and Rare Appointments
#----------------------------------------------------------

# Count the frequency of each 'clustering_id'
appointment_counts = data_df['clustering_id'].value_counts()

# Identify common and rare appointments
common_appointments = appointment_counts[appointment_counts > 1].index
rare_appointments = appointment_counts[appointment_counts == 1].index

# Filter data to keep only rows with common appointment descriptions
filtered_data_df = data_df[data_df['clustering_id'].isin(common_appointments)]
rare_appointments_df = data_df[data_df['clustering_id'].isin(rare_appointments)]

# Save both filtered and rare appointments to an Excel file
output_file_path = '/Users/catrienstolle/Desktop/filtered_data_doorloop_analyse_with_rare_appointments.xlsx'
with pd.ExcelWriter(output_file_path) as writer:
    filtered_data_df.to_excel(writer, sheet_name='Filtered_Appointments', index=False)
    rare_appointments_df.to_excel(writer, sheet_name='Rare_Appointments', index=False)

# Show a summary of the filtering process
print(f"Original number of appointments: {len(data_df)}")
print(f"Filtered number of appointments (common): {len(filtered_data_df)}")
print(f"Number of rare appointments (occur only once): {len(rare_appointments_df)}")
print(f"Filtered data and rare appointments saved to: {output_file_path}")

#---------- TIME SERIES ANALYSIS ------------------------------
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Assuming df is already loaded and processed from earlier steps

# Load the Excel file into a pandas DataFrame
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/merged_data_new.xlsx')

# Ensure date columns are in datetime format
df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'])
df['operation_date'] = pd.to_datetime(df['operation_date'])

# Define the specialty group
df['grouped_specialty'] = df['afspraken_na_screening_alle_appointment_specialism'].replace({
    'Psychologie': 'Vitalys Core Specialties',
    'Diëtetiek': 'Vitalys Core Specialties',
    'Chirurgie': 'Vitalys Core Specialties',
    'Vitalys': 'Vitalys Core Specialties',
    'Verpleegkundigen niet nader gespecificeerd': 'Vitalys Core Specialties',
    'Anesthesiologie': 'Vitalys Core Specialties',
})

# Filter out the first operation date per pseudo_id
first_operation_per_patient = df.groupby('pseudo_id')['operation_date'].min().reset_index()

# Merge this filtered operation date back to the original dataframe to preserve other columns
df = df.merge(first_operation_per_patient, on=['pseudo_id', 'operation_date'], how='inner')

# Filter Vitalys and non-Vitalys appointments based on the grouped_specialty
vitalys_appointments = df[df['grouped_specialty'] == 'Vitalys Core Specialties']
non_vitalys_appointments = df[df['grouped_specialty'] != 'Vitalys Core Specialties']

# Create time series data
vitalys_appointments_count = vitalys_appointments.resample('M', on='afspraken_na_screening_alle_appointment_start_date_x').size()
```

```python
non_vitalys_appointments_count = non_vitalys_appointments.resample('M', on='afspraken_na_screening_alle_appointment_start_date_x').size()
surgery_count = first_operation_per_patient.resample('M', on='operation_date').size()

# Create a DataFrame for plotting
time_series_data = pd.DataFrame({
    'Vitalys_Appointments': vitalys_appointments_count,
    'Non_Vitalys_Appointments': non_vitalys_appointments_count,
    'Surgeries': surgery_count
})

# Fill missing values using forward fill (you can use backward fill or other methods as needed)
time_series_data['Vitalys_Appointments'] = time_series_data['Vitalys_Appointments'].fillna(method='ffill')
time_series_data['Non_Vitalys_Appointments'] = time_series_data['Non_Vitalys_Appointments'].fillna(method='ffill')
time_series_data['Surgeries'] = time_series_data['Surgeries'].fillna(method='ffill')

# Drop rows with missing values
time_series_data = time_series_data.dropna()

# Plotting time series data
plt.figure(figsize=(14, 7))
plt.plot(time_series_data.index, time_series_data['Vitalys_Appointments'], label='Vitalys Appointments', marker='o')
plt.plot(time_series_data.index, time_series_data['Non_Vitalys_Appointments'], label='Non-Vitalys Appointments', marker='o')
plt.plot(time_series_data.index, time_series_data['Surgeries'], label='Surgeries', marker='o')

plt.title('Time Series Analysis of Appointments and Surgeries (2022 and 2023)')
plt.xlabel('Date')
plt.ylabel('Count')
plt.legend()
plt.grid()
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Decompose one of the series (Vitalys Appointments)
result = seasonal_decompose(time_series_data['Vitalys_Appointments'], model='additive', period=12)
result.plot()
plt.show()

# Decompose Non-Vitalys Appointments
result = seasonal_decompose(time_series_data['Non_Vitalys_Appointments'], model='additive', period=12)
result.plot()
plt.show()

# Decompose Surgeries
result = seasonal_decompose(time_series_data['Surgeries'], model='additive', period=12)
result.plot()
plt.show()

# Autocorrelation and Partial Autocorrelation Plots for Vitalys Appointments
plt.figure(figsize=(12, 6))
plot_acf(time_series_data['Vitalys_Appointments'], lags=20)
plt.title('Autocorrelation Function for Vitalys Appointments')
plt.show()

#--------------

import pandas as pd

# Stap 1: Dataset inladen
file_path = '/Users/catrienstolle/Desktop/Data Set Thesis/merged_dataset_final_new.xlsx'
data_df = pd.read_excel(file_path)

```

```python
# Stap 2: Datumkolommen omzetten naar datetime
data_df['aanmeldformulier_date'] = pd.to_datetime(data_df['aanmeldformulier_date'], errors='coerce')
data_df['afspraken_na_screening_alle_appointment_start_date_y'] = pd.to_datetime(data_df['afspraken_na_screening_alle_appointment_start_date_y'], errors='coerce')
data_df['operation_date'] = pd.to_datetime(data_df['operation_date'], errors='coerce')

# Stap 3: Bereken het aantal afspraken per patiënt
appointment_count_df = data_df['pseudo_id'].value_counts().reset_index()
appointment_count_df.columns = ['pseudo_id', 'appointment_count']
data_df = pd.merge(data_df, appointment_count_df, on='pseudo_id')

# Stap 4: Bepaal de modus van het aantal afspraken
mode_appointments = data_df['appointment_count'].mode()[0]
print(f"Meest voorkomend aantal afspraken (modus): {mode_appointments}")

# Stap 5: Bereken de doorlooptijd tussen screening en operatie
data_df['lead_time_screening_to_surgery'] = (data_df['operation_date'] - data_df['afspraken_na_screening_alle_appointment_start_date_y']).dt.days

# Stap 6: Bepaal de mediaan doorlooptijd en filter patiënten met acceptabele doorlooptijd
median_lead_time = data_df['lead_time_screening_to_surgery'].median()
upper_bound_time = median_lead_time
print(f"Mediaan doorlooptijd tussen screening en operatie: {median_lead_time} dagen")

# Stap 7: Bereken het aantal 'groepsbijeenkomsten' en 'screening' per patiënt
data_df['groepsbijeenkomst_count'] = data_df.groupby('pseudo_id')['clustering_id'].transform(
    lambda x: x.str.contains('groepsbijeenkomst_voor1|individueel_voor1|individueel_voorMOV', case=False, na=False).sum()
)
data_df['screening_count'] = data_df.groupby('pseudo_id')['clustering_id'].transform(lambda x: (x == 'screening').sum())

# Stap 8: Filter REDO patiënten
redo_data = data_df[data_df['clustering_surgeries_x'] == 'REDO']

# Stap 9: Filter patiënten met ten minste 1 groepsbijeenkomst en 1 screening
new_data = data_df.groupby('pseudo_id').filter(
    lambda group: group['groepsbijeenkomst_count'].max() >= 1 and
                  group['screening_count'].max() >= 1
)

# Stap 10: Verwijder REDO patiënten uit de gecombineerde dataset
total_patients_combined = new_data[~new_data['pseudo_id'].isin(redo_data['pseudo_id'])]
total_patients_combined = total_patients_combined.drop_duplicates(subset=['pseudo_id', 'clustering_id', 'afspraken_na_screening_alle_appointment_start_date_x'], keep='first')

# Aantal unieke patiënten in de gecombineerde dataset
total_patients_count = total_patients_combined['pseudo_id'].nunique()

# Stap 12: Verwijder duplicaten uit de standaardgroep
standard_group_combined = total_patients_combined.drop_duplicates(subset=['pseudo_id', 'clustering_id', 'afspraken_na_screening_alle_appointment_start_date_x'], keep='first')

# Stap 13: Filter de outliersgroep (patiënten die niet in de standaard- of REDO-groep zitten)
outliers = total_patients_combined[~total_patients_combined['pseudo_id'].isin(standard_group_combined['pseudo_id'])]

# Verwijder duplicaten uit de REDO- en outliersgroepen
redo_data = redo_data.drop_duplicates(keep='first')
outliers = outliers.drop_duplicates(keep='first')

# Stap 14: Print het aantal patiënten in de gecombineerde dataset
print(f"Totaal aantal patiënten in de gecombineerde dataset (Standaard + Outliers): {total_patients_count}")

# Stap 15: Sla de resultaten op in een Excel-bestand met meerdere tabbladen
output_file_path = '/Users/catrienstolle/Desktop/updated_patient_analysis_vitalys_v6_8oktober2022.xlsx'

with pd.ExcelWriter(output_file_path) as writer:
```

```python
    standard_group_combined.to_excel(writer, sheet_name='Standaardgroep', index=False)
    outliers.to_excel(writer, sheet_name='Outliersgroep', index=False)
    redo_data.to_excel(writer, sheet_name='REDO-groep', index=False)

print(f"Analyse succesvol opgeslagen in {output_file_path}")

# Stap 16: Print het aantal unieke patiënten in de REDO-groep
unique_count_redo = redo_data['pseudo_id'].nunique()
print(f"Aantal unieke patiënten in de REDO-groep: {unique_count_redo}")


import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

# Stap 1: Dataset voorbereiden en groepering maken -------------------------------------------
# Dataset inladen
vitalys_df = total_patients_combined

# Zet de datumkolommen om naar datetime
vitalys_df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(vitalys_df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')
vitalys_df['afspraken_na_screening_alle_appointment_start_date_y'] = pd.to_datetime(vitalys_df['afspraken_na_screening_alle_appointment_start_date_y'], errors='coerce')
vitalys_df['operation_date'] = pd.to_datetime(vitalys_df['operation_date'], errors='coerce')

# Sorteer de dataset op pseudo_id en afspraakdatum
vitalys_df = vitalys_df.sort_values(by=['pseudo_id', 'afspraken_na_screening_alle_appointment_start_date_x'])

# Verwijder irrelevante rijen zoals "Telefonisch contact"
irrelevant_filters = ['Telefonisch contact, geen consult', 'administrative', 'econsult', 'overleg_disciplines']
for term in irrelevant_filters:
    vitalys_df = vitalys_df[~vitalys_df['afspraken_na_screening_alle_appointment_description'].str.contains(term, case=False, na=False)]

# Bereken de vertraging tussen afspraken
vitalys_df['delay_between_appointments'] = vitalys_df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].diff().dt.days

# Bereken de doorlooptijd van screening tot operatie
vitalys_df['lead_time_screening_to_surgery'] = (vitalys_df['operation_date'] - vitalys_df['afspraken_na_screening_alle_appointment_start_date_y']).dt.days

# Opslaan van de delay analyse in een Excel-bestand
output_file = '/Users/catrienstolle/Desktop/doorloop_vitalys.xlsx'
with pd.ExcelWriter(output_file) as writer:
    vitalys_df.to_excel(writer, sheet_name='Standaard', index=False)
print(f"Delay analysis saved to {output_file}")

# Stap 2: Analyseer Herbeoordeling groep -------------------------------------------

# Totaal aantal unieke patiënten
total_patients_in_original = vitalys_df['pseudo_id'].nunique()


# Stap 3: Bereken doorlooptijd vanaf eerste groepsbijeenkomst tot operatie -------------------------------------------
# Vervang missing values (NaN) in 'clustering_id' door een lege string om problemen te voorkomen
vitalys_df['clustering_id'] = vitalys_df['clustering_id'].fillna('')

# Filter de data voor groepsbijeenkomsten
group_meetings = vitalys_df[vitalys_df['clustering_id'].str.contains('groepsbijeenkomst_voor1|individueel_voor|individueel_voorMOV', case=False, na=False)]

# Bereken de vroegste groepsbijeenkomst per patiënt
```

```python
group_meetings_earliest = group_meetings.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].min().reset_index()

# Geef de kolom met de eerste groepsbijeenkomst een duidelijkere naam
group_meetings_earliest.rename(columns={'afspraken_na_screening_alle_appointment_start_date_x': 'first_group_meeting_date'}, inplace=True)

# Voeg de eerste groepsbijeenkomst toe aan de dataset en bereken de doorlooptijd tot operatie
vitalys_df = pd.merge(vitalys_df, group_meetings_earliest, on='pseudo_id', how='left')
vitalys_df['lead_time_group_to_surgery'] = (vitalys_df['operation_date'] - vitalys_df['first_group_meeting_date']).dt.days

# Stap 4: Plot doorlooptijd distributie -------------------------------------------

# Plot histogram voor de doorlooptijd van groepsbijeenkomst tot operatie
plt.figure(figsize=(10, 6))
plt.hist(vitalys_df['lead_time_group_to_surgery'], bins=20, alpha=0.7, label='Lead Time to Surgery', color='green', density=True)
plt.axvline(x=49, color='red', linestyle='--', label='49 days threshold')
plt.title('Lead Time from Group Meeting to Surgery (with 49 days threshold)')
plt.xlabel('Lead Time (Days)')
plt.ylabel('Density (Normalized Count)')
plt.legend(loc='upper right')
plt.grid(True)
plt.tight_layout()
plt.show()

# Stap 5: Filter Standard en Outliers groepen -------------------------------------------

# Filteren op basis van een drempel van 49 dagen
standard_group_threshold = 49
standard_group_filtered = vitalys_df[vitalys_df['lead_time_group_to_surgery'] <= standard_group_threshold]
outliers_group_filtered = vitalys_df[vitalys_df['lead_time_group_to_surgery'] > standard_group_threshold]

# Aantal patiënten per groep berekenen
number_of_patients_standard = standard_group_filtered['pseudo_id'].nunique()
number_of_patients_outliers = outliers_group_filtered['pseudo_id'].nunique()

# Bereken percentages
percentage_standard = (number_of_patients_standard / total_patients_in_original) * 100
percentage_outliers = (number_of_patients_outliers / total_patients_in_original) * 100

# Resultaten printen
print(f"Totaal aantal patiënten in de originele dataset: {total_patients_in_original}")
print(f"Aantal patiënten in de Standard Group: {number_of_patients_standard} ({percentage_standard:.2f}%)")
print(f"Aantal patiënten in de Outliers Group: {number_of_patients_outliers} ({percentage_outliers:.2f}%)")

# Stap 6: Opslaan van de gefilterde groepen -------------------------------------------

output_file_standard = '/Users/catrienstolle/Desktop/standardgroep22022.xlsx'
output_file_outliers = '/Users/catrienstolle/Desktop/outliers2022.xlsx'
output_file_total = '/Users/catrienstolle/Desktop/biggroep_totaal_juist.xlsx'

# Opslaan van de totale dataset, standard groep en outliers in aparte Excel-bestanden
with pd.ExcelWriter(output_file_total) as writer:
    vitalys_df.to_excel(writer, sheet_name='Total', index=False)

with pd.ExcelWriter(output_file_standard) as writer:
    standard_group_filtered.to_excel(writer, sheet_name='Standard', index=False)

with pd.ExcelWriter(output_file_outliers) as writer:
    outliers_group_filtered.to_excel(writer, sheet_name='Outliers', index=False)

print(f"Filtered data saved to {output_file_total}, {output_file_standard}, and {output_file_outliers}")

```

```python
# Stap 7: Analyse van afspraak aantallen per groep --------------------------------------------

# Verwijder duplicaten om alleen unieke afspraak-aantallen per pseudo_id te behouden
standard_group_df_unique = standard_group_filtered.drop_duplicates(subset=['pseudo_id'])
outliers_group_df_unique = outliers_group_filtered.drop_duplicates(subset=['pseudo_id'])
redo_group_df_unique = redo_data.drop_duplicates(subset=['pseudo_id'])

# Voeg een kolom toe om de groep te identificeren
standard_group_df_unique['group'] = 'Standard Group'
outliers_group_df_unique['group'] = 'Delayed Group'
redo_group_df_unique['group'] = 'Redo Group'

# Combineer de data van alle drie de groepen in één dataframe
combined_df = pd.concat([standard_group_df_unique[['pseudo_id', 'appointment_count', 'group']],
                         outliers_group_df_unique[['pseudo_id', 'appointment_count', 'group']],
                         redo_group_df_unique[['pseudo_id', 'appointment_count', 'group']]])

# Maak een boxplot voor het aantal afspraken per groep
plt.figure(figsize=(10, 6))
sns.boxplot(data=combined_df, x='group', y='appointment_count')
plt.title('Appointment Counts by Group')
plt.ylabel('Appointment Count')
plt.xlabel('Group')
plt.show()

# Stap 8: Statistische analyses per groep uitvoeren --------------------------------------------

for group_name, group_data in combined_df.groupby('group'):
    print(f"\nStatistics for {group_name}:\n")

    # Beschrijvende statistieken
    mean_appointment = group_data['appointment_count'].mean()
    median_appointment = group_data['appointment_count'].median()
    std_appointment = group_data['appointment_count'].std()

    print(f"Mean Appointment Count: {mean_appointment}")
    print(f"Median Appointment Count: {median_appointment}")
    print(f"Standard Deviation: {std_appointment}")

    # Outliers detecteren met Z-score
    z_scores = stats.zscore(group_data['appointment_count'])
    outliers = group_data[z_scores.abs() > 3]
    print(f"Number of outliers: {len(outliers)}")

    # Normaliteitstest (D'Agostino en Pearson test)
    k2, p_value = stats.normaltest(group_data['appointment_count'])
    print(f"Normality test p-value: {p_value}")
    if p_value < 0.05:
        print("De afspraakverdeling is niet normaal.")
    else:
        print("De afspraakverdeling is normaal.")


# -----------------------------------------------------------
# Plot 1: Lead Time Distribution (Standard Group vs Outliers vs REDO)
plt.figure(figsize=(10, 6))

# Plot histograms for each group (Standard, Outliers, and REDO)
if 'lead_time_screening_to_surgery' in standard_group_filtered.columns:
    plt.hist(standard_group_filtered['lead_time_screening_to_surgery'], bins=20, alpha=0.7,
        label='Standard Group', color='green', density=True)
if 'lead_time_screening_to_surgery' in outliers_group_filtered.columns:
    plt.hist(outliers_group_filtered['lead_time_screening_to_surgery'], bins=20, alpha=0.7,
        label='Delayed Group', color='orange', density=True)
if 'lead_time_screening_to_surgery' in redo_data.columns:
    plt.hist(redo_data['lead_time_screening_to_surgery'], bins=20, alpha=0.7, label='REDO Group', color='red', density=True)
```

```python

plt.title('Lead Time from Screening to Surgery Total (Standard, Outliers, and REDO)')
plt.xlabel('Lead Time (Days)')
plt.ylabel('Density (Normalized Count)')
plt.legend(loc='upper right')
plt.grid(True)
plt.show()
```

D.2. Screenings Phase Analysis

```python
    import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Functie om dataset in te laden en voor te bereiden
def prepare_dataset(file_path, sheet_name):
    df = pd.read_excel(file_path, sheet_name=sheet_name)
    df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')
    df['afspraken_na_screening_alle_appointment_start_date_y'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_y'], errors='coerce')
    df['operation_date'] = pd.to_datetime(df['operation_date'], errors='coerce')

    # Sorteren en irrelevante rijen verwijderen
    df = df.sort_values(by=['pseudo_id', 'afspraken_na_screening_alle_appointment_start_date_x'])
    df = df[~df['afspraken_na_screening_alle_appointment_description'].str.contains('Telefonisch contact, geen consult', case=False, na=False)]
    df = df[~df['clustering_id'].str.contains('administrative|econsult|overleg|overleg_disciplines', case=False, na=False)]

    # Bereken vertragingen en doorlooptijd
    df['delay_between_appointments'] = df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].diff().dt.days
    df['delay_to_surgery'] = (df['operation_date'] - df['afspraken_na_screening_alle_appointment_start_date_x']).dt.days
    df['lead_time_screening_to_surgery'] = (df['operation_date'] - df['afspraken_na_screening_alle_appointment_start_date_y']).dt.days

    return df

# Functie om doorloopanalyse op te slaan
def save_delay_analysis(df, output_file):
    with pd.ExcelWriter(output_file) as writer:
        df.to_excel(writer, sheet_name='Standaard', index=False)
    print(f"Doorloop analysis saved to {output_file}")

# Functie om appointments na de eerste groepsbijeenkomst te filteren
def filter_after_first_group_meeting(df):
    df['first_group_meeting_date'] = df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].transform('min')
    filtered_df = df[df['afspraken_na_screening_alle_appointment_start_date_x'] >= df['first_group_meeting_date']]
    return filtered_df

# Functie om demografische informatie te visualiseren
def plot_demographics(df):
    patients_per_specialism = df.groupby('afspraken_na_screening_alle_appointment_specialism')['pseudo_id'].nunique().reset_index()
    patients_per_specialism.columns = ['Specialism', 'Number of Patients']
    patients_per_specialism['Percentage'] = (patients_per_specialism['Number of Patients'] / df['pseudo_id'].nunique()) * 100
    patients_per_specialism = patients_per_specialism.sort_values(by='Number of Patients', ascending=False)

    plt.figure(figsize=(10, 6))
    plt.barh(patients_per_specialism['Specialism'], patients_per_specialism['Number of Patients'], color='skyblue')
    plt.xlabel('Number of Patients')
    plt.title('Number of Patients per Specialism')
```

```python
    plt.tight_layout()
    plt.show()

# Functie om vertraging per specialisme te analyseren en te plotten, inclusief mediane doorlooptijd
def perform_delay_analysis(df, group_name, output_file):
    delay_analysis = df.groupby('afspraken_na_screening_alle_appointment_specialism').agg(
        median_screen_to_surgery=('lead_time_screening_to_surgery', 'median'),
        median_group_time_to_surgery=('lead_time_group_to_surgery', 'median'),  # Correcte kolom
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        total_patients=('pseudo_id', 'nunique')
    ).reset_index()

    delay_analysis.to_excel(output_file, index=False)
    print(f"Analysis saved to {output_file}")

    # Plot
    n = len(delay_analysis['afspraken_na_screening_alle_appointment_specialism'])
    indices = np.arange(n)
    bar_width = 0.2  # Smaller width to fit all bars

    fig, ax1 = plt.subplots(figsize=(14, 6))

    # Bar chart for Median Lead Time Screening to Surgery (orange)
    ax1.bar(indices - bar_width*1.5,
            delay_analysis['median_screen_to_surgery'],
            width=bar_width,
            color='orange',
            label='Median Lead Time Screening to Surgery (Days)',
            edgecolor='black')

    # Bar chart for Median Lead Time Group Meeting to Surgery (blue)
    ax1.bar(indices - bar_width/2,
            delay_analysis['median_group_time_to_surgery'],  # Nu correct weergegeven
            width=bar_width,
            color='blue',
            label='Median Lead Time Group Meeting to Surgery (Days)',
            edgecolor='black')

    # Bar chart for Total Patients (green)
    ax1.bar(indices + bar_width/2,
            delay_analysis['total_patients'],
            width=bar_width,
            color='green',
            label='Total Patients',
            edgecolor='black')

    # Set x-axis labels and ticks
    ax1.set_xticks(indices)
    ax1.set_xticklabels(delay_analysis['afspraken_na_screening_alle_appointment_specialism'], rotation=45, ha='right')

    # Set y-axis label for lead times and total patients
    ax1.set_ylabel('Lead Time (Days) and Total Patients')

    # Second axis for median delay
    ax2 = ax1.twinx()

    # Line plot for Median Delay Between Appointments (red)
    ax2.plot(indices,
             delay_analysis['median_delay_between_appointments'],
             color='red',
             marker='o',
             linestyle='-',
             label='Median Delay Between Appointments (Days)')

    # Set y-axis label for median delay
    ax2.set_ylabel('Median Delay Between Appointments (Days)', color='red')

    # Title and layout adjustments
```

```python
    plt.title(f'Lead␣Time,␣Median␣Delay,␣and␣Total␣Patients␣by␣Specialism␣({group_name})')

    # Combine legends from both axes
    ax1.legend(loc='upper␣left')
    ax2.legend(loc='upper␣right')

    # Tight layout for clarity
    plt.tight_layout()

    # Show the combined plot
    plt.show()

# Analyse uitvoeren voor 2022 en 2023
file_2022 = '/Users/catrienstolle/Desktop/Data␣Set␣Thesis/outliers2022correct.xlsx'
file_2023 = '/Users/catrienstolle/Desktop/Data␣Set␣Thesis/outliers.xlsx'

vitalys_df_2022 = prepare_dataset(file_2022, 'Outliers')
vitalys_df_2023 = prepare_dataset(file_2023, 'Outliers')

# Sla de analyses op
save_delay_analysis(vitalys_df_2022, '/Users/catrienstolle/Desktop/doorloop_2022.xlsx')
save_delay_analysis(vitalys_df_2023, '/Users/catrienstolle/Desktop/doorloop_2023.xlsx')

# Filteren na groepsbijeenkomst en opslaan
firstgroup_2022 = filter_after_first_group_meeting(vitalys_df_2022)
firstgroup_2023 = filter_after_first_group_meeting(vitalys_df_2023)
save_delay_analysis(firstgroup_2022, '/Users/catrienstolle/Desktop/firstgroup_2022.xlsx')
save_delay_analysis(firstgroup_2023, '/Users/catrienstolle/Desktop/firstgroup_2023.xlsx')

# Demografie plotten
plot_demographics(vitalys_df_2022)
plot_demographics(vitalys_df_2023)

# Vertraging per specialisme analyseren en plotten
perform_delay_analysis(vitalys_df_2022, '2022', '/Users/catrienstolle/Desktop/delay_analysis_2022.xlsx')
perform_delay_analysis(vitalys_df_2023, '2023', '/Users/catrienstolle/Desktop/delay_analysis_2023.xlsx')


# Function to perform a more detailed bottleneck analysis
def perform_detailed_bottleneck_analysis(df, year):
    """
    Performs a more detailed bottleneck analysis for the given dataset and year.
    Returns a plot of the top 10 bottlenecks with additional metrics and saves the analysis to an Excel file.

    Parameters:
    df (pd.DataFrame): The input dataset (filtered for the specific year).
    year (int): The year for labeling and file saving.
    """
    # Step 1: Calculate the delay between consecutive appointments
    df['delay_between_appointments'] = df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].diff().dt.days

    # Step 2: Fill NaN values in 'delay_between_appointments' with 0 if needed
    df['delay_between_appointments'] = df['delay_between_appointments'].fillna(0)

    # Step 3: Group by clustering_id, specialism, and pseudo_id to calculate statistics for unique patients
    grouped_df = df.groupby(['clustering_id', 'afspraken_na_screening_alle_appointment_specialism', 'pseudo_id']).agg(
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        max_delay=('delay_between_appointments', 'max'),  # Max delay per patient
        total_appointments=('clustering_id', 'size')  # Number of appointments per patient
    ).reset_index()

    # Step 4: Aggregate by clustering_id and specialism to calculate the median/max delay, total appointments, and total unique patients
    bottleneck_analysis = grouped_df.groupby(['clustering_id', 'afspraken_na_screening_alle_appointment_specialism']).agg(
```

```
        median_delay_between_appointments=('median_delay_between_appointments', 'median'),
        max_delay=('max_delay', 'max'),  # Max delay across patients
        avg_delay=('median_delay_between_appointments', 'mean'),  # Average delay across patients
        total_appointments=('total_appointments', 'sum'),  # Total appointments across patients
        total_patients=('pseudo_id', 'nunique')  # Unique patient count
    ).reset_index()

    # Step 5: Combine the appointment type and specialism for x-axis labels
    bottleneck_analysis['label'] = bottleneck_analysis['clustering_id'] + ' (' + bottleneck_analysis['afspraken_na_screening_alle_appointment_specialism'] + ')'

    # Step 6: Sort by the highest median delay and number of patients
    bottleneck_analysis = bottleneck_analysis.sort_values(by=['median_delay_between_appointments', 'total_patients'], ascending=False)

    # Step 7: Select the top 10 bottlenecks
    top_10_bottlenecks = bottleneck_analysis.head(10)

    # Step 8: Plot the top 10 bottlenecks by delay, patient count, and total appointments
    fig, ax1 = plt.subplots(figsize=(14, 7))

    # Bar chart for median delay
    bars = ax1.bar(top_10_bottlenecks['label'],
                   top_10_bottlenecks['median_delay_between_appointments'],
                   color='orange', label='Median Delay (Days)', edgecolor='black')

    ax1.set_xlabel('Appointment Type and Specialism')
    ax1.set_ylabel('Median Delay Between Appointments (Days)', color='orange')
    ax1.tick_params(axis='y', labelcolor='orange')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45, ha='right')

    # Second axis for total patients and total appointments
    ax2 = ax1.twinx()
    line1 = ax2.plot(top_10_bottlenecks['label'],
                     top_10_bottlenecks['total_patients'],
                     color='green', marker='o', label='Total Patients Affected')

    line2 = ax2.plot(top_10_bottlenecks['label'],
                     top_10_bottlenecks['total_appointments'],
                     color='blue', marker='x', linestyle='--', label='Total Appointments')

    # Adjust positioning of the patient numbers above the green line points
    for i, txt in enumerate(top_10_bottlenecks['total_patients']):
        ax2.text(i, top_10_bottlenecks['total_patients'].iloc[i] + 0.5, str(txt), color='green', ha='center', fontsize=10)

    # Adjust positioning of the appointment numbers above the blue line points
    for i, txt in enumerate(top_10_bottlenecks['total_appointments']):
        ax2.text(i, top_10_bottlenecks['total_appointments'].iloc[i] + 0.5, str(txt), color='blue', ha='center', fontsize=10)

    ax2.set_ylabel('Total Patients / Total Appointments', color='green')

    # Set limits for the y-axis to improve visualization
    ax1.set_ylim(0, top_10_bottlenecks['median_delay_between_appointments'].max() * 1.2)  # Adjust as needed
    ax2.set_ylim(0, top_10_bottlenecks[['total_patients', 'total_appointments']].max().max() * 1.2)  # Adjust as needed

    # Title and grid
    plt.title(f'Top 10 Bottlenecks: {year} - Appointments Causing Delays to the Most Patients (Detailed)')
    plt.grid(True)

    # Combine legends from both axes
    bars_list = list(bars)  # Convert bars container to list
    lines = bars_list + line1 + line2
```

```
    plt.legend(lines, [l.get_label() for l in lines], loc='upper right')

    # Show the plot
    plt.tight_layout()
    plt.show()

    # Step 9: Save the bottleneck analysis to an Excel file with additional columns
    output_file = f'/Users/catrienstolle/Desktop/bottleneck_analysis_detailed_{year}.xlsx'
    bottleneck_analysis.to_excel(output_file, index=False)
    print(f"Bottleneck analysis saved to {output_file}")

# Call the function for 2022 and 2023
perform_detailed_bottleneck_analysis(vitalys_df_2022, 2022)
perform_detailed_bottleneck_analysis(vitalys_df_2023, 2023)

import pandas as pd
import matplotlib.pyplot as plt

# Function to perform relationship analysis (median and standard deviation) for lead time and
     delays
def relationship_analysis(df):
    analysis = df.groupby('afspraken_na_screening_alle_appointment_specialism').agg(
        median_lead_time_to_surgery=('lead_time_screening_to_surgery', 'median'),
        std_lead_time_to_surgery=('lead_time_screening_to_surgery', 'std'),
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        std_delay_between_appointments=('delay_between_appointments', 'std')
    ).reset_index()

    # Calculate unique patient count per specialism
    unique_patient_count = df.groupby('afspraken_na_screening_alle_appointment_specialism')['
        pseudo_id'].nunique().reset_index()
    unique_patient_count.columns = ['afspraken_na_screening_alle_appointment_specialism', '
        total_unique_patients']

    # Merge the relationship analysis with the unique patient count
    final_analysis = pd.merge(analysis, unique_patient_count, on='
        afspraken_na_screening_alle_appointment_specialism')

    return final_analysis

# Function to plot top 10 specialisms by median delay
def plot_top_10_specialisms(df, group_name):
    # Sort by median delay and select the top 10 for plotting
    top_10_specialisms = df.sort_values(by='median_delay_between_appointments', ascending=
        False).head(10)

    # Assign unique numbers to specialisms for plotting
    top_10_specialisms['label_number'] = range(1, len(top_10_specialisms) + 1)

    # Create the scatter plot for the top 10 specialisms
    plt.figure(figsize=(10, 6))
    plt.scatter(top_10_specialisms['median_lead_time_to_surgery'],
                top_10_specialisms['median_delay_between_appointments'],
                color='purple')

    # Add numbered labels to each selected point
    for i, row in top_10_specialisms.iterrows():
        plt.text(row['median_lead_time_to_surgery'], row['median_delay_between_appointments'
            ],
                 str(row['label_number']), fontsize=9, ha='right')

    # Set axis labels and title
    plt.xlabel('Median Lead Screening Time to Surgery (Days)')
    plt.ylabel('Median Delay Between Appointments (Days)')
    plt.title(f'Top 10 Specialisms: {group_name}')

    # Add grid and layout adjustments
    plt.grid(True)
    plt.tight_layout()

    # Show the plot
```

```
    plt.show()

    # Print the legend mapping numbers to specialisms
    print(f"Legend␣for␣{group_name}:␣Mapping␣of␣numbers␣to␣top␣10␣specialisms")
    for i, row in top_10_specialisms.iterrows():
        print(f"{row['label_number']}:␣{row['afspraken_na_screening_alle_appointment_specialism']}")

# Function to save table data to CSV
def save_table(df, filename):
    output_file = f'/Users/catrienstolle/Desktop/{filename}.csv'
    df.to_csv(output_file, index=False)
    print(f"\nFull␣table␣saved␣to␣{output_file}")

# Load datasets
outliers_df = pd.read_excel('/Users/catrienstolle/Desktop/Data␣Set␣Thesis/totalmergedmapFINAL.xlsx', sheet_name='Blad1')
total_df = pd.read_excel('/Users/catrienstolle/Desktop/Data␣Set␣Thesis/merged_data_new.xlsx', sheet_name='Sheet1')

# Perform analysis for outliers group
outliers_analysis = relationship_analysis(outliers_df)
print("\nFull␣table␣of␣outliers␣group␣by␣Median␣Delay,␣Standard␣Deviation,␣Lead␣Time␣to␣Surgery,␣and␣Unique␣Patients:")
print(outliers_analysis)

# Perform analysis for total group (including non-delayed)
total_analysis = relationship_analysis(total_df)
print("\nFull␣table␣of␣total␣group␣by␣Median␣Delay,␣Standard␣Deviation,␣Lead␣Time␣to␣Surgery,␣and␣Unique␣Patients:")
print(total_analysis)

# Plot top 10 specialisms for both groups
plot_top_10_specialisms(outliers_analysis, 'Outliers␣Group␣(2022-2023)')
plot_top_10_specialisms(total_analysis, 'Total␣Group␣(2022-2023)')

# Save both tables to CSV
save_table(outliers_analysis, 'relationship_table_outliers_group')
save_table(total_analysis, 'relationship_table_total_group')

import matplotlib.pyplot as plt
from adjustText import adjust_text  # Import the adjustText library to avoid label overlap

# Function to plot top 10 specialisms for both groups on one graph with adjusted labels
def plot_combined_specialisms(outliers_df, total_df):
    # Sort and select the top 10 specialisms from both datasets
    top_10_outliers = outliers_df.sort_values(by='median_delay_between_appointments', ascending=False).head(10)
    top_10_total = total_df.sort_values(by='median_delay_between_appointments', ascending=False).head(10)

    # Create the combined scatter plot
    plt.figure(figsize=(12, 8))

    # Plot for outliers group (purple color)
    plt.scatter(top_10_outliers['median_lead_time_to_surgery'],
                top_10_outliers['median_delay_between_appointments'],
                color='purple', label='Outliers␣Group')

    # Plot for total group (orange color)
    plt.scatter(top_10_total['median_lead_time_to_surgery'],
                top_10_total['median_delay_between_appointments'],
                color='orange', label='Total␣Group')

    # Store text elements for adjustText to manage
    texts = []

    # Add specialism names to each point for outliers group
    for i, row in top_10_outliers.iterrows():
        texts.append(plt.text(row['median_lead_time_to_surgery'],
                              row['median_delay_between_appointments'],
```

```
                                row['afspraken_na_screening_alle_appointment_specialism'], fontsize=10, color='purple'))

    # Add specialism names to each point for total group
    for i, row in top_10_total.iterrows():
        texts.append(plt.text(row['median_lead_time_to_surgery'] + 0.5,  # Slight offset
                                row['median_delay_between_appointments'] + 0.5,  # Slight offset
                                row['afspraken_na_screening_alle_appointment_specialism'], fontsize=10, color='orange'))

    # Adjust text to avoid overlap
    adjust_text(texts, arrowprops=dict(arrowstyle='->', color='gray', lw=0.5))

    # Set axis labels and title
    plt.xlabel('Median␣Lead␣Screening␣Time␣to␣Surgery␣(Days)')
    plt.ylabel('Median␣Delay␣Between␣Appointments␣(Days)')
    plt.title('Top␣10␣Specialisms:␣Outliers␣vs.␣Total␣Group␣(2022-2023)')

    # Add legend, grid, and layout adjustments
    plt.legend()
    plt.grid(True)
    plt.tight_layout()

    # Show the combined plot
    plt.show()

# Ensure you have installed adjustText using pip: `pip install adjustText`

plot_combined_specialisms(outliers_analysis, total_analysis)
```

D.3. OR Phase Analysis

```
    import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Function to load and prepare dataset
def prepare_dataset(file_path, sheet_name):
    df = pd.read_excel(file_path, sheet_name=sheet_name)
    df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')
    df['afspraken_na_screening_alle_appointment_start_date_y'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_y'], errors='coerce')
    df['operation_date'] = pd.to_datetime(df['operation_date'], errors='coerce')

    # Sorting and filtering irrelevant rows
    df = df.sort_values(by=['pseudo_id', 'afspraken_na_screening_alle_appointment_start_date_x'])
    df = df[~df['afspraken_na_screening_alle_appointment_description'].str.contains('Telefonisch␣contact,␣geen␣consult', case=False, na=False)]
    df = df[~df['clustering_id'].str.contains('administrative|econsult|overleg|overleg_disciplines', case=False, na=False)]

    # Calculate delays and lead times
    df['delay_between_appointments'] = df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].diff().dt.days
    df['delay_to_surgery'] = (df['operation_date'] - df['afspraken_na_screening_alle_appointment_start_date_x']).dt.days
    df['lead_time_screening_to_surgery'] = (df['operation_date'] - df['afspraken_na_screening_alle_appointment_start_date_y']).dt.days

    return df

# Function to save delay analysis to Excel
def save_delay_analysis(df, output_file):
    with pd.ExcelWriter(output_file) as writer:
        df.to_excel(writer, sheet_name='Standaard', index=False)
    print(f"Doorloop␣analysis␣saved␣to␣{output_file}")

```

```
# Function to filter appointments after the first group meeting
# Step 1: Filter appointments after the first group meeting
def filter_after_first_group_meeting(group_df):
    """
    Filters the dataset to only include appointments that occur after and on the first group meeting,
    """
    # Ensure that all dates are in datetime format
    group_df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(
        group_df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')

    # Keep only appointments after the 1st group meeting
    filtered_df = group_df[group_df['afspraken_na_screening_alle_appointment_start_date_x'] >= group_df['first_group_meeting_date']]

    # Remove rows where the appointment type contains 'groepsbijeenkomst_voor'

    return filtered_df

# Function to perform delay analysis with median
def perform_delay_analysis_median(df, group_name, output_file):
    """
    Performs delay analysis grouped by specialism and appointment type using median.
    Saves the analysis to Excel and creates a combined bar/line plot.
    """
    # Step 1: Perform the delay analysis
    delay_analysis = df.groupby(
        ['afspraken_na_screening_alle_appointment_specialism']
    ).agg(
        median_screen_to_surgery=('lead_time_screening_to_surgery', 'median'),
        median_group_time_to_surgery=('lead_time_group_to_surgery', 'median'),  # Correct column name here
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        total_patients=('pseudo_id', 'nunique')
    ).reset_index()

    # Step 2: Save the results to Excel
    with pd.ExcelWriter(output_file) as writer:
        delay_analysis.to_excel(writer, sheet_name='Descriptive Stats', index=False)

    print(f"Analysis saved for {group_name} to: {output_file}")

    # Step 3: Create a bar plot for median lead time to surgery
    fig, ax1 = plt.subplots(figsize=(12, 6))

    # Bar chart for median lead time to surgery
    ax1.bar(delay_analysis['afspraken_na_screening_alle_appointment_specialism'],
            delay_analysis['median_group_time_to_surgery'], color='orange', label='Median Lead Time to Surgery (Days)', edgecolor='black')
    ax1.set_xlabel('Specialism')
    ax1.set_ylabel('Median Lead Time to Surgery (Days)', color='orange')
    ax1.tick_params(axis='y', labelcolor='orange')

    # Rotate x-axis labels for readability
    plt.xticks(rotation=45, ha='right')

    # Step 4: Create a secondary axis for median delay and total patients
    ax2 = ax1.twinx()

    # Line plot for median delay between appointments
    ax2.plot(delay_analysis['afspraken_na_screening_alle_appointment_specialism'],
             delay_analysis['median_delay_between_appointments'], color='red', marker='o', label='Median Delay Between Appointments (Days)')

    # Line plot for total patients
    ax2.plot(delay_analysis['afspraken_na_screening_alle_appointment_specialism'],
             delay_analysis['total_patients'], color='green', marker='x', label='Total Patients')

    ax2.set_ylabel('Median Delay / Total Patients', color='black')

```

```python
    # Title and grid
    plt.title(f'Median Lead Time Group to Surgery, Median Delay, and Total Patients by Specialism - {group_name}')

    # Combine both axes' legends
    fig.legend(loc='upper right', bbox_to_anchor=(1, 1), bbox_transform=ax1.transAxes)

    # Show the grid and the plot
    plt.grid(True)
    plt.tight_layout()
    plt.show()

# Load your datasets for 2022 and 2023
file_2022 = '/Users/catrienstolle/Desktop/Data Set Thesis/outliers2022correct.xlsx'
file_2023 = '/Users/catrienstolle/Desktop/Data Set Thesis/outliers.xlsx'

# Prepare the datasets
vitalys_df_2022 = prepare_dataset(file_2022, 'Outliers')
vitalys_df_2023 = prepare_dataset(file_2023, 'Outliers')

# Filter after the first group meeting
firstgroup_2022 = filter_after_first_group_meeting(vitalys_df_2022)
firstgroup_2023 = filter_after_first_group_meeting(vitalys_df_2023)

# Perform and save the delay analysis for 2022 and 2023
perform_delay_analysis_median(firstgroup_2022, 'First Group 2022', '/Users/catrienstolle/Desktop/firstgroup_2022_analysis.xlsx')
perform_delay_analysis_median(firstgroup_2023, 'First Group 2023', '/Users/catrienstolle/Desktop/firstgroup_2023_analysis.xlsx')

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Function to perform a more detailed bottleneck analysis
def perform_detailed_bottleneck_analysis(df, year):
    """
    Performs a more detailed bottleneck analysis for the given dataset and year.
    Returns a plot of the top 10 bottlenecks with additional metrics and saves the analysis
        to an Excel file.

    Parameters:
    df (pd.DataFrame): The input dataset (filtered for the specific year).
    year (int): The year for labeling and file saving.
    """
    # Step 1: Calculate the delay between consecutive appointments
    df['delay_between_appointments'] = df.groupby('pseudo_id')['afspraken_na_screening_alle_appointment_start_date_x'].diff().dt.days

    # Step 2: Fill NaN values in 'delay_between_appointments' with 0 if needed
    df['delay_between_appointments'] = df['delay_between_appointments'].fillna(0)

    # Step 3: Group by clustering_id, specialism, and pseudo_id to calculate statistics for
        unique patients
    grouped_df = df.groupby(['clustering_id', 'afspraken_na_screening_alle_appointment_specialism', 'pseudo_id']).agg(
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        max_delay=('delay_between_appointments', 'max'),  # Max delay per patient
        total_appointments=('clustering_id', 'size')  # Number of appointments per patient
    ).reset_index()

    # Step 4: Aggregate by clustering_id and specialism to calculate the median/max delay,
        total appointments, and total unique patients
    bottleneck_analysis = grouped_df.groupby(['clustering_id', 'afspraken_na_screening_alle_appointment_specialism']).agg(
        median_delay_between_appointments=('median_delay_between_appointments', 'median'),
        max_delay=('max_delay', 'max'),  # Max delay across patients
        avg_delay=('median_delay_between_appointments', 'mean'),  # Average delay across
            patients
        total_appointments=('total_appointments', 'sum'),  # Total appointments across
            patients
```

```python
        total_patients=('pseudo_id', 'nunique')  # Unique patient count
    ).reset_index()

    # Step 5: Combine the appointment type and specialism for x-axis labels
    bottleneck_analysis['label'] = bottleneck_analysis['clustering_id'] + ' (' +
        bottleneck_analysis['afspraken_na_screening_alle_appointment_specialism'] + ')'

    # Step 6: Sort by the highest median delay and number of patients
    bottleneck_analysis = bottleneck_analysis.sort_values(by=['
        median_delay_between_appointments', 'total_patients'], ascending=False)

    # Step 7: Select the top 10 bottlenecks
    top_10_bottlenecks = bottleneck_analysis.head(10)

    # Step 8: Plot the top 10 bottlenecks by delay, patient count, and total appointments
    fig, ax1 = plt.subplots(figsize=(14, 7))

    # Bar chart for median delay
    bars = ax1.bar(top_10_bottlenecks['label'],
                   top_10_bottlenecks['median_delay_between_appointments'],
                   color='orange', label='Median Delay (Days)', edgecolor='black')

    ax1.set_xlabel('Appointment Type and Specialism')
    ax1.set_ylabel('Median Delay Between Appointments (Days)', color='orange')
    ax1.tick_params(axis='y', labelcolor='orange')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45, ha='right')

    # Second axis for total patients and total appointments
    ax2 = ax1.twinx()
    line1 = ax2.plot(top_10_bottlenecks['label'],
                     top_10_bottlenecks['total_patients'],
                     color='green', marker='o', label='Total Patients Affected')

    line2 = ax2.plot(top_10_bottlenecks['label'],
                     top_10_bottlenecks['total_appointments'],
                     color='blue', marker='x', linestyle='--', label='Total Appointments')

    # Adjust positioning of the patient numbers above the green line points
    for i, txt in enumerate(top_10_bottlenecks['total_patients']):
        ax2.text(i, top_10_bottlenecks['total_patients'].iloc[i] + 0.5, str(txt), color='
            green', ha='center', fontsize=10)

    # Adjust positioning of the appointment numbers above the blue line points
    for i, txt in enumerate(top_10_bottlenecks['total_appointments']):
        ax2.text(i, top_10_bottlenecks['total_appointments'].iloc[i] + 0.5, str(txt), color='
            blue', ha='center', fontsize=10)

    ax2.set_ylabel('Total Patients / Total Appointments', color='green')

    # Title and grid
    plt.title(f'Top 10 Bottlenecks: {year} - Appointments Causing Delays to the Most Patients
         (Detailed)')
    plt.grid(True)

    # Combine legends from both axes
    bars_list = list(bars)  # Convert bars container to list
    lines = bars_list + line1 + line2
    labels = [l.get_label() for l in lines]
    plt.legend(lines, labels, loc='upper right')

    # Show the plot
    plt.tight_layout()
    plt.show()

    # Step 9: Save the bottleneck analysis to an Excel file with additional columns
    output_file = f'/Users/catrienstolle/Desktop/bottleneck_analysis_detailed_{year}.xlsx'
    bottleneck_analysis.to_excel(output_file, index=False)
    print(f"Bottleneck analysis saved to {output_file}")
```

```
# Call the function for 2022 and 2023
perform_detailed_bottleneck_analysis(firstgroup_2022, 2022)
perform_detailed_bottleneck_analysis(firstgroup_2023, 2023)

import pandas as pd
import matplotlib.pyplot as plt

# Function to perform relationship analysis (median and standard deviation) for lead time and delays
def relationship_analysis(df):
    analysis = df.groupby('afspraken_na_screening_alle_appointment_specialism').agg(
        median_lead_time_to_surgery=('lead_time_group_to_surgery', 'median'),
        std_lead_time_to_surgery=('lead_time_group_to_surgery', 'std'),
        median_delay_between_appointments=('delay_between_appointments', 'median'),
        std_delay_between_appointments=('delay_between_appointments', 'std')
    ).reset_index()

    # Calculate unique patient count per specialism
    unique_patient_count = df.groupby('afspraken_na_screening_alle_appointment_specialism')['pseudo_id'].nunique().reset_index()
    unique_patient_count.columns = ['afspraken_na_screening_alle_appointment_specialism', 'total_unique_patients']

    # Merge the relationship analysis with the unique patient count
    final_analysis = pd.merge(analysis, unique_patient_count, on='afspraken_na_screening_alle_appointment_specialism')

    return final_analysis

# Function to plot top 10 specialisms by median delay
def plot_top_10_specialisms(df, group_name):
    # Sort by median delay and select the top 10 for plotting
    top_10_specialisms = df.sort_values(by='median_delay_between_appointments', ascending=False).head(10)

    # Assign unique numbers to specialisms for plotting
    top_10_specialisms['label_number'] = range(1, len(top_10_specialisms) + 1)

    # Create the scatter plot for the top 10 specialisms
    plt.figure(figsize=(10, 6))
    plt.scatter(top_10_specialisms['median_lead_time_to_surgery'],
                top_10_specialisms['median_delay_between_appointments'],
                color='purple')

    # Add numbered labels to each selected point
    for i, row in top_10_specialisms.iterrows():
        plt.text(row['median_lead_time_to_surgery'], row['median_delay_between_appointments'],
                 str(row['label_number']), fontsize=9, ha='right')

    # Set axis labels and title
    plt.xlabel('Median␣Lead␣Screening␣Time␣to␣Surgery␣(Days)')
    plt.ylabel('Median␣Delay␣Between␣Appointments␣(Days)')
    plt.title(f'Top␣10␣Specialisms:␣{group_name}')

    # Add grid and layout adjustments
    plt.grid(True)
    plt.tight_layout()

    # Show the plot
    plt.show()

    # Print the legend mapping numbers to specialisms
    print(f"Legend␣for␣{group_name}:␣Mapping␣of␣numbers␣to␣top␣10␣specialisms")
    for i, row in top_10_specialisms.iterrows():
        print(f"{row['label_number']}:␣{row['afspraken_na_screening_alle_appointment_specialism']}")

# Function to save table data to CSV
def save_table(df, filename):
    output_file = f'/Users/catrienstolle/Desktop/{filename}.csv'
```

```
    df.to_csv(output_file, index=False)
    print(f"\nFull table saved to {output_file}")

# Load datasets
outliers_df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/totalmergedmapFINAL.xlsx', sheet_name='Blad1')
total_df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/merged_data_new.xlsx', sheet_name='Sheet1')

outliersafterfirst = filter_after_first_group_meeting(outliers_df)
totalafterfirst = filter_after_first_group_meeting(total_df)

# Perform analysis for outliers group
outliers_analysis = relationship_analysis(outliersafterfirst)
print("\nFull table of outliers group by Median Delay, Standard Deviation, Lead Time to Surgery, and Unique Patients:")
print(outliers_analysis)

# Perform analysis for total group (including non-delayed)
total_analysis = relationship_analysis(totalafterfirst)
print("\nFull table of total group by Median Delay, Standard Deviation, Lead Time to Surgery, and Unique Patients:")
print(total_analysis)

# Plot top 10 specialisms for both groups
plot_top_10_specialisms(outliers_analysis, 'Outliers Group (2022-2023)')
plot_top_10_specialisms(total_analysis, 'Total Group (2022-2023)')

# Save both tables to CSV
save_table(outliers_analysis, 'relationship_table_outliers_group')
save_table(total_analysis, 'relationship_table_total_group')

from adjustText import adjust_text  # Import the adjustText library to avoid label overlap

# Function to plot top 10 specialisms for both groups on one graph with adjusted labels
def plot_combined_specialisms(outliers_df, total_df):
    # Sort and select the top 10 specialisms from both datasets
    top_10_outliers = outliers_df.sort_values(by='median_delay_between_appointments', ascending=False).head(10)
    top_10_total = total_df.sort_values(by='median_delay_between_appointments', ascending=False).head(10)

    # Create the combined scatter plot
    plt.figure(figsize=(12, 8))

    # Plot for outliers group (purple color)
    plt.scatter(top_10_outliers['median_lead_time_to_surgery'],
                top_10_outliers['median_delay_between_appointments'],
                color='purple', label='Outliers Group')

    # Plot for total group (orange color)
    plt.scatter(top_10_total['median_lead_time_to_surgery'],
                top_10_total['median_delay_between_appointments'],
                color='orange', label='Total Group')

    # Store text elements for adjustText to manage
    texts = []

    # Add specialism names to each point for outliers group
    for i, row in top_10_outliers.iterrows():
        texts.append(plt.text(row['median_lead_time_to_surgery'],
                              row['median_delay_between_appointments'],
                              row['afspraken_na_screening_alle_appointment_specialism'],
                              fontsize=10, color='purple'))

    # Add specialism names to each point for total group
    for i, row in top_10_total.iterrows():
        texts.append(plt.text(row['median_lead_time_to_surgery'] + 0.5,  # Slight offset
                              row['median_delay_between_appointments'] + 0.5,  # Slight offset
                              row['afspraken_na_screening_alle_appointment_specialism'],
```

```
                                   fontsize=10, color='orange'))

    # Adjust text to avoid overlap
    adjust_text(texts, arrowprops=dict(arrowstyle='->', color='gray', lw=0.5))

    # Set axis labels and title
    plt.xlabel('Median Lead Screening Time to Surgery (Days)')
    plt.ylabel('Median Delay Between Appointments (Days)')
    plt.title('Top 10 Specialisms: Outliers vs. Total Group (2022-2023)')

    # Add legend, grid, and layout adjustments
    plt.legend()
    plt.grid(True)
    plt.tight_layout()

    # Show the combined plot
    plt.show()

# Ensure you have installed adjustText using pip: `pip install adjustText`

plot_combined_specialisms(outliers_analysis, total_analysis)
```

D.4. Plannings Model Analysis

```
import pandas as pd
import matplotlib.pyplot as plt

# Load your dataset (update with your actual file path)
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/merged_data_new.xlsx')

df = filter_after_first_group_meeting(df)

# Remove rows with missing or non-finite lead_time_group_to_surgery values
df_cleaned = df[df['lead_time_group_to_surgery'].notna()]

# Step 1: Calculate the number of weeks delayed (beyond 49 days)
df_cleaned['weeks_delayed'] = ((df_cleaned['lead_time_group_to_surgery'] - 49) // 7).astype(
    int)

# Step 2: Filter out only the patients who have a delay greater than 49 days
delayed_patients = df_cleaned[df_cleaned['lead_time_group_to_surgery'] > 49]

# Step 3: Group some specialties under the name 'Vitalys Core Specialties'
specialties_column = 'afspraken_na_screening_alle_appointment_specialism'
delayed_patients['grouped_specialty'] = delayed_patients[specialties_column].replace({
    'Psychologie': 'Vitalys Core Specialties',
    'Diëtetiek': 'Vitalys Core Specialties',
    'Chirurgie': 'Vitalys Core Specialties',
    'Vitalys': 'Vitalys Core Specialties',
    'Verpleegkundigen niet nader gespecificeerd': 'Vitalys Core Specialties',
    'Anesthesiologie': 'Vitalys Core Specialties',
})

# Step 4: Group by 'weeks_delayed' and 'grouped_specialty' to analyze delays per specialty
grouped_specialty_summary = delayed_patients.groupby(['weeks_delayed', 'grouped_specialty']).
    size().reset_index(name='Count')

# Save the results to an Excel file
output_dir = '/Users/catrienstolle/Desktop/'
grouped_specialty_summary.to_excel(f'{output_dir}grouped_specialty_summary.xlsx', index=False
    )

# Print out a message to indicate the results are saved
print(f"Grouped specialty summary (delayed patients only) has been saved to {output_dir}
    grouped_specialty_summary.xlsx.")

# Optionally, display the first few rows of the summary for review
print(grouped_specialty_summary.head())
```

```python
# Load your dataset (update with your actual file path)
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/merged_data_new.xlsx')

# Remove rows with missing or non-finite lead_time_group_to_surgery values
df_cleaned = df[df['lead_time_group_to_surgery'].notna()]

# Step 1: Calculate the number of weeks delayed (beyond 49 days)
df_cleaned['weeks_delayed'] = ((df_cleaned['lead_time_group_to_surgery'] - 49) // 7).astype(int)

# Step 2: Filter out only the patients who have a delay greater than 49 days
delayed_patients = df_cleaned[df_cleaned['lead_time_group_to_surgery'] > 49]

# Step 3: Group some specialties under the name 'Vitalys Core Specialties'
specialties_column = 'afspraken_na_screening_alle_appointment_specialism'
delayed_patients['grouped_specialty'] = delayed_patients[specialties_column].replace({
    'Psychologie': 'Vitalys Core Specialties',
    'Diëtetiek': 'Vitalys Core Specialties',
    'Chirurgie': 'Vitalys Core Specialties',
    'Vitalys': 'Vitalys Core Specialties',
    'Verpleegkundigen niet nader gespecificeerd': 'Vitalys Core Specialties',
    'Anesthesiologie': 'Vitalys Core Specialties',
})

# Step 4: Group by 'weeks_delayed', 'grouped_specialty', and 'pseudo_id' to count unique patients
grouped_specialty_summary = delayed_patients.groupby(['weeks_delayed', 'grouped_specialty'])['pseudo_id'].nunique().reset_index(name='Unique_Patient_Count')

# Save the results to an Excel file
output_dir = '/Users/catrienstolle/Desktop/'
grouped_specialty_summary.to_excel(f'{output_dir}grouped_specialty_summary_unique_patients.xlsx', index=False)

# Print out a message to indicate the results are saved
print(f"Grouped specialty summary (unique delayed patients) has been saved to {output_dir}grouped_specialty_summary_unique_patients.xlsx.")

# Optionally, display the first few rows of the summary for review
print(grouped_specialty_summary.head())


# Load your dataset (replace with your actual file path)
df = pd.read_excel('/Users/catrienstolle/Desktop/grouped_specialty_summary_unique_patients.xlsx')

# Step 1: Adjust the 'weeks_delayed' by adding 7 to each value (shift all weeks by 7)
df['weeks_delayed'] = df['weeks_delayed'] + 7

# Step 2: Pivot the data so that each department becomes its own column for counting unique patients
pivot_df = df.pivot_table(
    index='weeks_delayed',
    columns='grouped_specialty',
    values='Unique_Patient_Count',
    aggfunc='sum',
    fill_value=0
)

# Step 3: Plot the Stacked Area Chart for the number of unique patients (starting from week 7)
plt.figure(figsize=(12, 8))

# Plot a stacked area chart
plt.stackplot(pivot_df.index, pivot_df.T, labels=pivot_df.columns, alpha=0.8)

# Customize the plot
plt.title('Number of Delayed Patients by Specialty and Weeks Delayed', fontsize=14)
plt.xlabel('Weeks Delayed (Starting from 7)', fontsize=12)
plt.ylabel('Number of Unique Patients', fontsize=12)
plt.grid(True)
```

```python
plt.legend(title='Specialty', bbox_to_anchor=(1.05, 1), loc='upper left')

# Show the plot
plt.tight_layout()
plt.show()

# Save the adjusted summary to Excel (if needed)
output_dir = '/Users/catrienstolle/Desktop/'
pivot_df.to_excel(f'{output_dir}stacked_specialty_summary_shifted.xlsx', index=True)
print(f"Stacked summary with shifted weeks saved to {output_dir}stacked_specialty_summary_shifted.xlsx.")


#------- MAPPING
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/merged_data_new.xlsx')

# Convert relevant date columns to datetime
df['surgery_date'] = pd.to_datetime(df['operation_date'], errors='coerce')
df['appointment_start_date'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')

# Create a mapping for clustering_id to define the group meetings
meeting_mapping = {
    'voor1': {'groepsbijeenkomst': 'groepsbijeenkomst_voor1', 'individueel': ['individueel_voor1', 'individueel_voorMOV'], 'deadline': 49},
    'voor2': {'groepsbijeenkomst': 'groepsbijeenkomst_voor2', 'individueel': ['individueel_voor2'], 'deadline': 42},
    'voor3': {'groepsbijeenkomst': 'groepsbijeenkomst_voor3', 'individueel': ['individueel_voor3'], 'deadline': 35},
    'voor4': {'groepsbijeenkomst': 'groepsbijeenkomst_voor4', 'individueel': ['individueel_voor4'], 'deadline': 28},
    'voor5': {'groepsbijeenkomst': 'groepsbijeenkomst_voor5', 'individueel': ['individueel_voor5'], 'deadline': 21}
}

# Initialize an empty dictionary to store the percentage of delays and the number of patients
delayed_percentages = {}
patients_for_analysis = {}

# Loop through each group/individual meeting and calculate the percentage of delayed unique patients
for stage, details in meeting_mapping.items():
    meeting_patients = df[df['clustering_id'].isin([details['groepsbijeenkomst']] + details['individueel'])]
    meeting_patients = meeting_patients.drop_duplicates(subset='pseudo_id')
    meeting_patients['lead_time_to_surgery'] = (meeting_patients['surgery_date'] - meeting_patients['appointment_start_date']).dt.days

    delayed_patients = meeting_patients[meeting_patients['lead_time_to_surgery'] > details['deadline']]
    total_patients = len(meeting_patients)
    delayed_count = len(delayed_patients)

    delayed_percentage = (delayed_count / total_patients) * 100 if total_patients > 0 else 0
    delayed_percentages[stage] = {'total_patients': total_patients, 'delayed_percentage': delayed_percentage}

    # Store the delayed patients for later comparison
    patients_for_analysis[stage] = delayed_patients['pseudo_id'].unique()

# Prepare data for plotting
group_meetings = list(delayed_percentages.keys())
delayed_percentages_values = [data['delayed_percentage'] for data in delayed_percentages.values()]

# Create a figure and axis
plt.figure(figsize=(12, 6))

# Plot the delayed percentages with a line and markers
plt.plot(group_meetings, delayed_percentages_values, marker='o', linestyle='-', color='dodgerblue', markersize=8, linewidth=2, label='Delayed Patients (%)')
```

```

# Set the limits to 0-100% for the y-axis
plt.ylim(0, 100)

# Set the labels and title
plt.xlabel('Group Meeting Stage', fontsize=14)
plt.ylabel('Delayed Patients (%)', fontsize=14)
plt.title('Comparison of Delayed Patients from First to Last Group Meeting', fontsize=16, fontweight='bold')

# Display grid for readability
plt.grid(color='lightgrey', linestyle='--', linewidth=0.7)

# Annotate the percentages on the graph
for i, percentage in enumerate(delayed_percentages_values):
    plt.text(i, percentage + 1, f'{percentage:.2f}%', ha='center', fontsize=10, color='black')

# Add legend
plt.legend()

# Show the plot with better layout
plt.xticks(rotation=45, fontsize=12)  # Rotate x labels for better readability
plt.tight_layout()
plt.show()


import pandas as pd

# Load your dataset
df = total_df

# Convert relevant date columns to datetime
df['surgery_date'] = pd.to_datetime(df['operation_date'], errors='coerce')
df['appointment_start_date'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')

# Define the mapping for group meetings
meeting_mapping = {
    'voor1': {'groepsbijeenkomst': 'groepsbijeenkomst_voor1', 'individueel': ['individueel_voor1', 'individueel_voorMOV'], 'deadline': 49},
    'voor2': {'groepsbijeenkomst': 'groepsbijeenkomst_voor2', 'individueel': ['individueel_voor2'], 'deadline': 42},
    'voor3': {'groepsbijeenkomst': 'groepsbijeenkomst_voor3', 'individueel': ['individueel_voor3'], 'deadline': 35},
    'voor4': {'groepsbijeenkomst': 'groepsbijeenkomst_voor4', 'individueel': ['individueel_voor4'], 'deadline': 28},
    'voor5': {'groepsbijeenkomst': 'groepsbijeenkomst_voor5', 'individueel': ['individueel_voor5'], 'deadline': 21}
}

# Step 1: Find patients in voor1 but not in voor5
# Get patients delayed in voor1 and voor5
patients_vor1 = set(df[df['clustering_id'].isin([meeting_mapping['voor1']['groepsbijeenkomst']] + meeting_mapping['voor1']['individueel'])]['pseudo_id'])
patients_vor5 = set(df[df['clustering_id'].isin([meeting_mapping['voor5']['groepsbijeenkomst']] + meeting_mapping['voor5']['individueel'])]['pseudo_id'])

# Find remaining patients who are delayed in voor1 but not in voor5
remaining_patients_in_vor1 = patients_vor1 - patients_vor5

# Step 2: Get the specialties for the remaining patients
# Filter the DataFrame for the remaining patients and find their specialties
specialties_for_remaining_patients = df[df['pseudo_id'].isin(remaining_patients_in_vor1)]

# Assuming the column that indicates specialties is 'afspraken_na_screening_alle_appointment_specialism' (replace with your actual column)
specialties_column = 'afspraken_na_screening_alle_appointment_specialism'

# Step 3: Group Psychologie, Diëtetiek, and Chirurgie into a single category "Vitalys Core Specialties"
```

```
specialties_for_remaining_patients['grouped_specialty'] = specialties_for_remaining_patients[specialties_column].replace({
    'Psychologie': 'Vitalys Core Specialties',
    'Diëtetiek': 'Vitalys Core Specialties',
    'Chirurgie': 'Vitalys Core Specialties',
    'Vitalys': 'Vitalys Core Specialties',
    'Verpleegkundigen niet nader gespecificeerd': 'Vitalys Core Specialties',
    'Anesthesiologie': 'Vitalys Core Specialties',
})

# Step 4: Remove duplicate patient-specialty combinations (so each patient is counted only once per department)
specialties_for_remaining_patients_unique = specialties_for_remaining_patients.drop_duplicates(subset=['pseudo_id', 'grouped_specialty'])

# Step 5: Count the number of patients for each grouped specialty
specialty_counts = specialties_for_remaining_patients_unique['grouped_specialty'].value_counts()

# Step 6: Calculate percentages for each grouped specialty
total_remaining_patients = len(remaining_patients_in_vor1)  # Total unique patients in vor1 but not in vor5
specialty_percentages = (specialty_counts / total_remaining_patients) * 100

# Step 7: Save the specialty percentages to an Excel file
output_dir = '/Users/catrienstolle/Desktop/'  # Change to your actual output directory
specialty_percentages_df = pd.DataFrame({
    'Specialty': specialty_percentages.index,
    'Percentage': specialty_percentages.values
})
specialty_percentages_df.to_excel(f'{output_dir}specialty_percentages_for_remaining_patients_in_vor1_grouped.xlsx', index=False)

# Print out the specialty percentages for review
print(specialty_percentages_df)

# Display the percentage results in a more readable way
print(f"Specialty percentages for remaining patients in 'voor1' have been saved to {output_dir}specialty_percentages_for_remaining_patients_in_vor1_grouped.xlsx.")

# Step 4: Create a flag for patients who **only** visited Vitalys Core Specialties
# Step 4a: First, group by pseudo_id and check if all their appointments fall under "Vitalys Core Specialties"
patient_specialties_grouped = specialties_for_remaining_patients.groupby('pseudo_id')['grouped_specialty'].apply(lambda x: set(x))

# Step 4b: Create a new flag indicating whether a patient only visited Vitalys Core Specialties
def is_only_vitalys_core(specialty_set):
    return specialty_set == {'Vitalys Core Specialties'}

specialties_for_remaining_patients['only_vitalys_core'] = specialties_for_remaining_patients['pseudo_id'].map(patient_specialties_grouped).apply(is_only_vitalys_core)

# Step 5: Split the patients into two groups
only_vitalys_core_patients = specialties_for_remaining_patients[specialties_for_remaining_patients['only_vitalys_core'] == True]['pseudo_id'].nunique()
other_department_patients = specialties_for_remaining_patients[specialties_for_remaining_patients['only_vitalys_core'] == False]['pseudo_id'].nunique()

# Total patients for calculating percentages
total_remaining_patients = len(remaining_patients_in_vor1)

# Step 6: Calculate percentages
percentage_only_vitalys_core = (only_vitalys_core_patients / total_remaining_patients) * 100
percentage_other_departments = (other_department_patients / total_remaining_patients) * 100

# Step 7: Display the results
print(f"Percentage of patients who visited only Vitalys Core Specialties: {percentage_only_vitalys_core:.2f}%")
print(f"Percentage of patients who visited other departments: {percentage_other_departments
```

```
    :.2f}%")

# Save the results to an Excel file
output_dir = '/Users/catrienstolle/Desktop/'  # Change to your actual output directory
results_df = pd.DataFrame({
    'Category': ['Only␣Vitalys␣Core␣Specialties', 'Other␣Departments'],
    'Percentage': [percentage_only_vitalys_core, percentage_other_departments]
})
results_df.to_excel(f'{output_dir}comparison_vitalys_core_vs_other_departments.xlsx', index=
    False)

print(f"Comparison␣results␣saved␣to␣{output_dir}comparison_vitalys_core_vs_other_departments.
    xlsx.")

def analyze_patient_delays(df, group_name, output_file, max_weeks=10):
    """
    Analyzes delays between the first group meeting and surgery, plots the results,
    saves delayed patient appointments, and counts department visits.

    Parameters:
    df (pd.DataFrame): The input dataframe.
    group_name (str): A name for the group (used for titles and file names).
    output_file (str): The path to save the delayed patient appointments.
    max_weeks (int): The maximum number of weeks to analyze.
    """
    # Step 1: Define Vitalys Core departments
    df['grouped_specialty'] = df['afspraken_na_screening_alle_appointment_specialism'].
        replace({
        'Psychologie': 'Vitalys␣Core␣Specialties',
        'Diëtetiek': 'Vitalys␣Core␣Specialties',
        'Chirurgie': 'Vitalys␣Core␣Specialties',
        'Vitalys': 'Vitalys␣Core␣Specialties',
        'Verpleegkundigen␣niet␣nader␣gespecificeerd': 'Vitalys␣Core␣Specialties',
        'Anesthesiologie': 'Vitalys␣Core␣Specialties',
    })

    # Step 2: Calculate the delay between the first group appointment and surgery
    df['delay_from_first_group_to_surgery'] = (df['operation_date'] - df['
        first_group_meeting_date']).dt.days

    # Step 3: Calculate weeks delayed (starting from 49 days or 7 weeks)
    df['weeks_delayed'] = ((df['delay_from_first_group_to_surgery']) // 7).astype(int)

    # Step 4: Filter out patients with delay greater than 49 days
    delayed_patients = df[df['delay_from_first_group_to_surgery'] > 0]

    # Step 5: Analyze patient visits to identify if they visited non-core specialties
    patient_specialty_visits = delayed_patients.groupby('pseudo_id')['grouped_specialty'].
        apply(lambda x: set(x))

    # Step 6: Classify each patient based on whether they visited non-core departments
    patient_classification = patient_specialty_visits.apply(lambda x: 'Non-Vitalys' if '
        Vitalys␣Core␣Specialties' not in x or len(x) > 1 else 'Vitalys␣Core␣Specialties')

    # Step 7: Merge the classification back into the original dataframe
    delayed_patients = delayed_patients.merge(patient_classification.rename('
        visit_classification'), on='pseudo_id', how='left')

    # Step 8: Calculate unique patients who visited only Vitalys Core Specialties vs Non-
        Vitalys
    unique_patient_counts = delayed_patients.drop_duplicates(subset='pseudo_id').groupby(['
        weeks_delayed', 'visit_classification']).size().reset_index(name='
        Unique_Patient_Count')

    # Step 9: Prepare for plotting the number of unique patients over the weeks
    thresholds = range(0, (max_weeks + 1) * 7 + 49, 7)
    delayed_patient_counts = {}
    non_vitalys_percentage = {}

    # Create lists for plotting
    weeks = []
```

```python
    total_patients_per_week = []
    non_vitalys_percentages = []

    for threshold in thresholds:
        # Filter patients delayed beyond the threshold
        patients_beyond_threshold = delayed_patients[delayed_patients['delay_from_first_group_to_surgery'] > threshold]
        total_patients = patients_beyond_threshold['pseudo_id'].nunique()

        if total_patients > 0:
            # Calculate the number of patients visiting non-Vitalys departments
            non_vitalys_patients = patients_beyond_threshold[patients_beyond_threshold['visit_classification'] == 'Non-Vitalys']
            non_vitalys_count = non_vitalys_patients['pseudo_id'].nunique()

            non_vitalys_percentage[threshold] = (non_vitalys_count / total_patients) * 100

            # Store the count of delayed patients for this threshold
            delayed_patient_counts[threshold] = total_patients

            # Add data for plotting
            weeks.append(threshold)
            total_patients_per_week.append(total_patients)
            non_vitalys_percentages.append(non_vitalys_percentage[threshold])

            # Print the results for this threshold
            print(f'\nThreshold: {threshold} days')
            print(f'  Total patients delayed beyond {threshold} days: {total_patients}')
            print(f'  Non-Vitalys patients: {non_vitalys_count} ({non_vitalys_percentage[threshold]:.2f}%)')

        else:
            print(f'\nThreshold: {threshold} days')
            print('  No patients delayed beyond this threshold.')

    # Plot the counts of delayed patients for each threshold
    plt.figure(figsize=(10, 6))
    plt.bar(weeks, total_patients_per_week, color='lightblue', edgecolor='black', label='Total Delayed Patients')

    # Add a line plot for non-Vitalys percentage on the secondary y-axis
    plt.twinx()
    plt.plot(weeks, non_vitalys_percentages, color='red', marker='o', label='Non-Vitalys %', linestyle='--')
    plt.ylabel('Non-Vitalys Patients (%)', fontsize=12)

    # Set plot labels and title
    plt.title(f'Number of Delayed Patients and Non-Vitalys % ({group_name})', fontsize=14)
    plt.xlabel('Days Beyond First Group Appointment', fontsize=12)
    plt.ylabel('Number of Delayed Patients / Percentage Non Vitalys', fontsize=12)

    # Show grid and tight layout
    plt.grid(True)
    plt.legend(loc='upper left')
    plt.tight_layout()
    plt.show()

    # Return the unique patient counts for further use
    return unique_patient_counts, delayed_patient_counts, non_vitalys_percentage

# Example usage for your dataset
unique_patient_counts, delayed_patient_counts, non_vitalys_percentage = analyze_patient_delays(
    totalafterfirst,
    'Total Group',
    '/Users/catrienstolle/Desktop/delayed_outliers_appointments.xlsx',
    max_weeks=10
)

# Print out a summary of the results
print("\nDelayed Patient Counts and Non-Vitalys Percentages by Threshold:")
```

```python
for threshold in delayed_patient_counts.keys():
    print(f"Threshold:␣{threshold}␣days␣-␣Total␣Patients:␣{delayed_patient_counts[threshold]},␣Non-Vitalys␣Percentage:␣{non_vitalys_percentage[threshold]:.2f}%")

#- PLANNINGSMODEL

# Load your dataset
vitalys = pd.read_excel('/Users/catrienstolle/Desktop/Data␣Set␣Thesis/merged_data_new.xlsx', sheet_name='Sheet1')

def filter_after_first_group_meeting(group_df):
    """
    Filters the dataset to only include appointments that occur after and on the first group meeting,
    """
    # Ensure that all dates are in datetime format
    group_df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(
        group_df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')

    # Keep only appointments after the 1st group meeting
    filtered_df = group_df[group_df['afspraken_na_screening_alle_appointment_start_date_x'] >= group_df['first_group_meeting_date']]

    # Remove rows where the appointment type contains 'groepsbijeenkomst_voor'

    return filtered_df

df = filter_after_first_group_meeting(vitalys)

# Load the REDO surgeries dataset
redo_df = pd.read_excel('/Users/catrienstolle/Desktop/Data␣Set␣Thesis/redo_surgeries.xlsx', sheet_name='Blad1')  # Adjust the path as needed

# Ensure date columns are in datetime format
df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')
df['operation_date'] = pd.to_datetime(df['operation_date'], errors='coerce')
df['first_group_meeting_date'] = pd.to_datetime(df['first_group_meeting_date'], errors='coerce')
redo_df['operation_date'] = pd.to_datetime(redo_df['operation_date'], errors='coerce')  # Ensure the REDO date is in datetime format
redo_df['afspraken_na_screening_alle_appointment_start_date_x'] = pd.to_datetime(redo_df['afspraken_na_screening_alle_appointment_start_date_x'], errors='coerce')

# Set the number of slots available per week
slots_per_week = 28.0
reserved_slot_for_postponed = 4  # Reserve slots for postponed patients
reserved_slot_for_redo = 2 # Maximum REDO slots per week

# Define the number of weeks for lead time
lead_time_weeks = 6
postponed_weeks = 2
postponed_redo_weeks = 2
total_capacity_returned = 0

# List of excluded departments
excluded_departments = ['Psychologie', 'Diëtetiek', 'Vitalys', 'Verpleegkundigen␣niet␣nader␣gespecificeerd', 'Anesthesiologie', 'Chirurgie']

# Create a flag for patients who have appointments in other departments
df['has_other_department'] = df['afspraken_na_screening_alle_appointment_specialism'].apply(
    lambda x: 1 if x not in excluded_departments else 0
)

# Create a list to hold scheduled patients
scheduled_patients = []
patient_count_per_week = {}

# Get the latest appointment for each unique patient for scheduling
unique_patients = df.groupby('pseudo_id').agg({
    'first_group_meeting_date': 'min',
```

```python
    'afspraken_na_screening_alle_appointment_start_date_x': 'max',
    'has_other_department': 'sum'
}).reset_index()

# Iterate through each unique patient based on pseudo_id
for _, patient in unique_patients.iterrows():
    pseudo_id = patient['pseudo_id']
    last_appointment_date = patient['afspraken_na_screening_alle_appointment_start_date_x']
    first_group_meeting_date = patient['first_group_meeting_date']

    # Check if the patient has other department appointments
    if patient['has_other_department'] > 0:
        # If the patient has other department appointments, postpone scheduling
        appointment_date = last_appointment_date + pd.DateOffset(weeks=postponed_weeks)

        # Only append if appointment_date is valid
        if pd.notnull(appointment_date):
            # Check the week for scheduling
            while True:
                week_start = appointment_date.to_period('W').start_time
                if week_start not in patient_count_per_week:
                    patient_count_per_week[week_start] = {'Scheduled': 0, 'Postponed': 0, 'REDO': 0}

                # Check if slots are available for postponed patients
                if patient_count_per_week[week_start]['Postponed'] < reserved_slot_for_postponed:
                    # Schedule the patient
                    scheduled_patients.append({
                        'pseudo_id': pseudo_id,
                        'scheduled_date': appointment_date,
                        'status': 'Postponed due to other appointments'
                    })
                    patient_count_per_week[week_start]['Postponed'] += 1
                    break  # Exit the loop, patient is scheduled
                else:
                    # Move to the next week if over the limit
                    appointment_date += pd.DateOffset(weeks=1)  # Increment the appointment date by one week

    else:
        # Schedule based on the first group meeting date, 6 weeks ahead
        appointment_date = first_group_meeting_date + pd.DateOffset(weeks=lead_time_weeks)

        # Only append if appointment_date is valid
        if pd.notnull(appointment_date):
            # Check for slot availability and adjust if necessary
            while True:
                week_start = appointment_date.to_period('W').start_time
                if week_start not in patient_count_per_week:
                    patient_count_per_week[week_start] = {'Scheduled': 0, 'Postponed': 0, 'REDO': 0}

                # Check if slots are available for scheduled patients
                if patient_count_per_week[week_start]['Scheduled'] < (slots_per_week - reserved_slot_for_postponed):
                    # Schedule the patient
                    scheduled_patients.append({
                        'pseudo_id': pseudo_id,
                        'scheduled_date': appointment_date,
                        'status': 'Scheduled'
                    })
                    patient_count_per_week[week_start]['Scheduled'] += 1
                    break  # Exit the loop, patient is scheduled
                else:
                    # Move to the next week if over the limit
                    appointment_date += pd.DateOffset(weeks=1)  # Increment the appointment date by one week

# Schedule REDO surgeries, reserving slots two weeks in advance
unique_redos = redo_df.groupby('pseudo_id').agg({
```

```
    'operation_date': 'max'  # Get the latest operation date for each REDO patient
}).reset_index()

for _, redo in unique_redos.iterrows():
    appointment_date = redo['operation_date'] - pd.DateOffset(weeks=postponed_redo_weeks)  # Schedule 2 weeks in advance

    # Only append if appointment_date is valid
    while True:
        week_start = appointment_date.to_period('W').start_time
        if week_start not in patient_count_per_week:
            patient_count_per_week[week_start] = {'Scheduled': 0, 'Postponed': 0, 'REDO': 0}

        # Check if there's space for a REDO surgery considering the total capacity
        total_patients = (patient_count_per_week[week_start]['Scheduled'] +
                          patient_count_per_week[week_start]['Postponed'] +
                          patient_count_per_week[week_start]['REDO'])

        if total_patients < slots_per_week and patient_count_per_week[week_start]['REDO'] < reserved_slot_for_redo:
            # Schedule the REDO surgery
            scheduled_patients.append({
                'pseudo_id': redo['pseudo_id'],  # Assuming there's a pseudo_id for REDO patients
                'scheduled_date': appointment_date,
                'status': 'Scheduled REDO'
            })
            patient_count_per_week[week_start]['REDO'] += 1
            break  # Exit the loop, surgery is scheduled
        else:
            # Move to the next week if over the limit
            appointment_date += pd.DateOffset(weeks=1)  # Increment the appointment date by one week

# Convert scheduled patients to a DataFrame for further analysis
scheduled_patients_df = pd.DataFrame(scheduled_patients)

# Display scheduled patients
print(scheduled_patients_df)

for week, counts in patient_count_per_week.items():
    total_scheduled = counts['Scheduled'] + counts['Postponed'] + counts['REDO']
    available_slots = slots_per_week - total_scheduled
    if available_slots > 0:
        total_capacity_returned += available_slots

# Convert scheduled patients to a DataFrame for further analysis
scheduled_patients_df = pd.DataFrame(scheduled_patients)

# Display scheduled patients and total capacity returned
print(scheduled_patients_df)
print(f'Total capacity returned (slots available): {total_capacity_returned}')

# Convert the patient count dictionary to a DataFrame
patient_count_df = pd.DataFrame.from_dict(patient_count_per_week, orient='index').reset_index()
patient_count_df.columns = ['Week Starting', 'Scheduled Count', 'Postponed Count', 'REDO Count']

# Convert 'Week Starting' to datetime
patient_count_df['Week Starting'] = pd.to_datetime(patient_count_df['Week Starting'].astype(str))  # Ensure correct datetime format

patient_count_df = patient_count_df.sort_values(by='Week Starting')  # Sort by week date

# Save the weekly patient count to a CSV file
patient_count_df.to_csv('weekly_patient_schedule.csv', index=False)

# Create a plot
plt.figure(figsize=(14, 7))
plt.plot(patient_count_df['Week Starting'], patient_count_df['Scheduled Count'],
```

```python
         marker='o', linestyle='-', color='skyblue', label='Scheduled Patients')
plt.plot(patient_count_df['Week Starting'], patient_count_df['Postponed Count'],
         marker='x', linestyle='--', color='orange', label='Postponed Patients')
plt.plot(patient_count_df['Week Starting'], patient_count_df['REDO Count'],
         marker='s', linestyle=':', color='green', label='REDO Patients')

# Add titles and labels
plt.title('Weekly Patient Schedule with Postponed and REDO Patients', fontsize=16, fontweight='bold')
plt.xlabel('Week Starting', fontsize=14)
plt.ylabel('Number of Patients', fontsize=14)

# Customize the ticks
plt.xticks(rotation=45)
plt.yticks(fontsize=12)

# Add a grid for better readability
plt.grid(color='grey', linestyle='--', linewidth=0.5)

# Add a legend
plt.legend()

# Display the plot
plt.tight_layout()
plt.show()


# Convert scheduled patients to a DataFrame for further analysis
scheduled_patients_df = pd.DataFrame(scheduled_patients)

# Count the number of standard, postponed, and REDO patients
standard_count = len(scheduled_patients_df[scheduled_patients_df['status'] == 'Scheduled'])
postponed_count = len(scheduled_patients_df[scheduled_patients_df['status'] == 'Postponed due to other appointments'])
redo_count = len(scheduled_patients_df[scheduled_patients_df['status'] == 'Scheduled REDO'])

# Display the counts
print(f'Number of standard patients: {standard_count}')
print(f'Number of postponed patients: {postponed_count}')
print(f'Number of REDO patients: {redo_count}')

# Also print the total number of patients
total_patients = len(scheduled_patients_df)
print(f'Total number of patients: {total_patients}')

# Calculate the percentage for each category
standard_percentage = (standard_count / total_patients) * 100
postponed_percentage = (postponed_count / total_patients) * 100
redo_percentage = (redo_count / total_patients) * 100

print(f'Percentage of standard patients: {standard_percentage:.2f}%')
print(f'Percentage of postponed patients: {postponed_percentage:.2f}%')
print(f'Percentage of REDO patients: {redo_percentage:.2f}%')


# Initialize total capacity remaining (after accounting for all patients)
total_capacity_remaining = 0
total_slots_per_week = 30  # Define the total number of slots available per week

# Iterate through each week and calculate the remaining capacity after scheduling all patients
for week, counts in patient_count_per_week.items():
    total_scheduled = counts['Scheduled'] + counts['Postponed'] + counts['REDO']

    # Calculate remaining slots for this week
    remaining_slots = total_slots_per_week - total_scheduled

    # Only count remaining slots if they are available (positive value)
    if remaining_slots > 0:
        total_capacity_remaining += remaining_slots

```

```
# Print out the total capacity that could be returned after scheduling all patients
print(f'Total capacity remaining (after scheduling all patients): {total_capacity_remaining}')
```

D.5. Predictive Model Analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, GridSearchCV, StratifiedKFold, learning_curve, cross_val_score, RandomizedSearchCV
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from imblearn.over_sampling import SMOTE
from imblearn.under_sampling import RandomUnderSampler
from sklearn.preprocessing import StandardScaler, PolynomialFeatures, MinMaxScaler
from imblearn.over_sampling import SMOTE
from imblearn.under_sampling import RandomUnderSampler
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.neural_network import MLPClassifier


# Load the data
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/predictivemodel.xlsx', sheet_name='Blad1')

# Ensure numeric columns are properly formatted
df['age'] = pd.to_numeric(df['age'], errors='coerce')
df['BMI'] = pd.to_numeric(df['BMI'], errors='coerce')

# Scale 'age' and 'BMI' using StandardScaler
scaler = MinMaxScaler()
df[['age', 'BMI']] = scaler.fit_transform(df[['age', 'BMI']])

# Define a threshold to classify 'lead_time_group_to_surgery' and create the target variable
threshold = 49
df['delayed_surgery'] = df['lead_time_group_to_surgery'].apply(lambda x: 1 if x > threshold else 0)

# Handle missing values
df.dropna(subset=['age', 'BMI'], inplace=True)

# Create dummy variable for 'gender'
df['gender_dummy'] = df['gender'].map({'F': 0, 'M': 1})

# Optional: Create polynomial features for age and BMI
#poly = PolynomialFeatures(degree=2, include_bias=False)
#poly_features = poly.fit_transform(df[['age', 'BMI']])
#poly_feature_names = poly.get_feature_names_out(['age', 'BMI'])
#poly_df = pd.DataFrame(poly_features, columns=poly_feature_names)
#df = pd.concat([df, poly_df], axis=1)

# Define independent variables (X) excluding irrelevant columns
columns_to_use = [col for col in df.columns if col not in [
    'roken', 'pseudo_id', 'delayed_surgery','obesitas_familie', 'diabetes_familie', 'num_other_departments','gender', 'afspraken_na_screening_alle_appointment_specialism',
    'lead_time_group_to_surgery', 'lengte', 'gewicht', 'appointment_count', 'morbide_obese'
]]
X = df[columns_to_use]
y = df['delayed_surgery']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 1. SMOTE (Oversampling the Minority Class) only on training data
smote = SMOTE(random_state=42)
X_resampled_smote, y_resampled_smote = smote.fit_resample(X_train, y_train)

# Train a Random Forest classifier with SMOTE data
```

```python
rf_classifier_smote = RandomForestClassifier(random_state=42)
rf_classifier_smote.fit(X_resampled_smote, y_resampled_smote)
y_pred_smote = rf_classifier_smote.predict(X_test)

# Evaluate the model with SMOTE data
print("\n=== Random Forest with SMOTE ===")
print("Accuracy after SMOTE:", accuracy_score(y_test, y_pred_smote))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_smote))
print("Classification Report:\n", classification_report(y_test, y_pred_smote))

# 2. Undersampling the Majority Class only on training data
rus = RandomUnderSampler(random_state=42)
X_resampled_rus, y_resampled_rus = rus.fit_resample(X_train, y_train)

# Train a Random Forest classifier with undersampled data
rf_classifier_rus = RandomForestClassifier(random_state=42)
rf_classifier_rus.fit(X_resampled_rus, y_resampled_rus)
y_pred_rus = rf_classifier_rus.predict(X_test)

# Evaluate the model with undersampled data
print("\n=== Random Forest with Undersampling ===")
print("Accuracy after undersampling:", accuracy_score(y_test, y_pred_rus))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_rus))
print("Classification Report:\n", classification_report(y_test, y_pred_rus))

# 3. Hyperparameter Tuning for Random Forest with SMOTE
param_grid_rf = {
    'n_estimators': [100, 200],
    'max_depth': [3, 5],
    'min_samples_split': [2, 5],
    'min_samples_leaf': [1, 2],
    'bootstrap': [True, False]
}

# Create the Random Forest classifier
rf_classifier = RandomForestClassifier(random_state=42)

# Use GridSearchCV for hyperparameter tuning with 5-fold cross-validation
grid_search_rf = GridSearchCV(estimator=rf_classifier, param_grid=param_grid_rf, cv=5,
    scoring='accuracy', n_jobs=-1)
grid_search_rf.fit(X_resampled_smote, y_resampled_smote)

# Best parameters found by GridSearchCV
best_params_rf = grid_search_rf.best_params_
print(f"Best Hyperparameters (Random Forest): {best_params_rf}")

# Use the best Random Forest model after hyperparameter tuning
best_rf_classifier = grid_search_rf.best_estimator_

# Cross-validation with the best Random Forest model
cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
cv_scores = cross_val_score(best_rf_classifier, X_resampled_smote, y_resampled_smote, cv=cv,
    scoring='accuracy')
print(f'Average CV Accuracy: {np.mean(cv_scores):.4f}')

# Fit the best model on the full resampled dataset and make predictions on the test set
best_rf_classifier.fit(X_resampled_smote, y_resampled_smote)
y_pred_best_rf = best_rf_classifier.predict(X_test)

# Evaluate the best Random Forest model after SMOTE and hyperparameter tuning
print("\n=== Best Random Forest Model after SMOTE and Hyperparameter Tuning ===")
print("Accuracy:", accuracy_score(y_test, y_pred_best_rf))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_best_rf))
print("Classification Report:\n", classification_report(y_test, y_pred_best_rf))

# Learning Curve for Best Random Forest Model
train_sizes_rf, train_scores_rf, test_scores_rf = learning_curve(
    best_rf_classifier, X, y, cv=5, scoring='accuracy', train_sizes=np.linspace(0.1, 1.0, 10)
)
train_scores_mean_rf = np.mean(train_scores_rf, axis=1)
test_scores_mean_rf = np.mean(test_scores_rf, axis=1)
```

```python

plt.figure(figsize=(10, 6))
plt.plot(train_sizes_rf, train_scores_mean_rf, label='Training Accuracy', color='blue')
plt.plot(train_sizes_rf, test_scores_mean_rf, label='Testing Accuracy', color='orange')
plt.xlabel('Training Set Size')
plt.ylabel('Accuracy')
plt.title('Learning Curve with Best Random Forest Model after SMOTE and Hyperparameter Tuning
    ')
plt.legend()
plt.show()

# Feature Importance for Best Random Forest Model
importances_rf = best_rf_classifier.feature_importances_
feature_names_rf = X.columns
feature_importance_df_rf = pd.DataFrame({'Feature': feature_names_rf, 'Importance':
    importances_rf}).sort_values(by='Importance', ascending=False)

plt.figure(figsize=(10, 6))
sns.barplot(x='Importance', y='Feature', data=feature_importance_df_rf)
plt.title('Feature Importance with Best Random Forest Model (SMOTE + Hyperparameter Tuning)')
plt.show()

# Apply 10-Fold Cross-Validation
cv_scores = cross_val_score(best_rf_classifier, X_resampled_smote, y_resampled_smote, cv=10,
    scoring='accuracy')

# Print the average accuracy score from cross-validation
print(f'Average CV Accuracy: {np.mean(cv_scores):.4f}')

# ----------------------------------
# GRADIENT BOOSTING MODEL
# ----------------------------------
# Updated hyperparameter grid with stronger regularization
param_grid_gb = {
    'n_estimators': [200, 500],                # Allow more estimators with a lower learning
        rate
    'max_depth': [2, 3],                       # Shallow trees to reduce complexity
    'learning_rate': [0.005, 0.01],            # Lower learning rate
    'min_samples_split': [10, 20],             # Higher minimum samples for split
    'min_samples_leaf': [5, 10],               # Higher minimum samples per leaf
    'subsample': [0.5, 0.6]                    # Stronger subsampling to reduce overfitting
}

# Enable early stopping by setting a validation fraction
gb_classifier = GradientBoostingClassifier(random_state=42, validation_fraction=0.1,
    n_iter_no_change=10)

# Perform RandomizedSearchCV
randomized_search_gb = RandomizedSearchCV(estimator=gb_classifier, param_distributions=
    param_grid_gb,
                                          n_iter=10, cv=3, scoring='accuracy', n_jobs=-1,
                                              random_state=42)
randomized_search_gb.fit(X_resampled_smote, y_resampled_smote)

# Evaluate and get the best model and parameters
best_gb_classifier = randomized_search_gb.best_estimator_
print(f"Best Hyperparameters (Gradient Boosting): {randomized_search_gb.best_params_}")

# Cross-validation with the best Gradient Boosting model
cv_scores_gb = cross_val_score(best_gb_classifier, X_resampled_smote, y_resampled_smote, cv
    =3, scoring='accuracy')
print(f'Average CV Accuracy (Gradient Boosting): {np.mean(cv_scores_gb):.4f}')

# Fit the best model on the full resampled dataset and make predictions on the test set
best_gb_classifier.fit(X_resampled_smote, y_resampled_smote)
y_pred_best_gb = best_gb_classifier.predict(X_test)

# Evaluate the best Gradient Boosting model
print("\n=== Best Gradient Boosting Model after SMOTE and Hyperparameter Tuning ===")
print("Accuracy:", accuracy_score(y_test, y_pred_best_gb))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_best_gb))
```

```python
print("Classification Report:\n", classification_report(y_test, y_pred_best_gb))

# Learning Curve for Best Gradient Boosting Model
train_sizes_gb, train_scores_gb, test_scores_gb = learning_curve(
    best_gb_classifier, X, y, cv=3, scoring='accuracy', train_sizes=np.linspace(0.1, 1.0, 10)
)
train_scores_mean_gb = np.mean(train_scores_gb, axis=1)
test_scores_mean_gb = np.mean(test_scores_gb, axis=1)

plt.figure(figsize=(10, 6))
plt.plot(train_sizes_gb, train_scores_mean_gb, label='Training Accuracy', color='blue')
plt.plot(train_sizes_gb, test_scores_mean_gb, label='Testing Accuracy', color='orange')
plt.xlabel('Training Set Size')
plt.ylabel('Accuracy')
plt.title('Learning Curve for Best Gradient Boosting Model after SMOTE and Hyperparameter Tuning')
plt.legend()
plt.show()

# Feature Importance for Best Gradient Boosting Model
importances_gb = best_gb_classifier.feature_importances_
feature_importance_df_gb = pd.DataFrame({'Feature': X.columns, 'Importance': importances_gb}).sort_values(by='Importance', ascending=False)

plt.figure(figsize=(10, 6))
sns.barplot(x='Importance', y='Feature', data=feature_importance_df_gb)
plt.title('Feature Importance with Best Gradient Boosting Model (SMOTE + Hyperparameter Tuning)')
plt.show()

# -----------------------------------
# MLP  MODEL
# -----------------------------------
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, GridSearchCV, learning_curve
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.neural_network import MLPClassifier
from imblearn.over_sampling import SMOTE
from sklearn.preprocessing import StandardScaler

# Load and preprocess the data
df = pd.read_excel('/Users/catrienstolle/Desktop/Data Set Thesis/predictivemodel.xlsx', sheet_name='Blad1')

# Ensure numeric columns are properly formatted
df['age'] = pd.to_numeric(df['age'], errors='coerce')
df['BMI'] = pd.to_numeric(df['BMI'], errors='coerce')

# Scale 'age' and 'BMI' using StandardScaler
scaler = StandardScaler()
df[['age', 'BMI']] = scaler.fit_transform(df[['age', 'BMI']])

# Define the target variable based on threshold for 'lead_time_group_to_surgery'
threshold = 49
df['delayed_surgery'] = df['lead_time_group_to_surgery'].apply(lambda x: 1 if x > threshold else 0)

# Drop missing values
df.dropna(subset=['age', 'BMI'], inplace=True)

# Create dummy variable for 'gender'
df['gender_dummy'] = df['gender'].map({'F': 0, 'M': 1})

# Define features and target
columns_to_use = [col for col in df.columns if col not in [
    'pseudo_id', 'delayed_surgery', 'num_other_departments', 'gender', 'afspraken_na_screening_alle_appointment_specialism',
    'lead_time_group_to_surgery', 'lengte', 'gewicht', 'appointment_count', 'morbide_obese'
```

```
]]
X = df[columns_to_use]
y = df['delayed_surgery']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# SMOTE to handle class imbalance on training data only
smote = SMOTE(random_state=42)
X_resampled_smote, y_resampled_smote = smote.fit_resample(X_train, y_train)

# Initialize the MLP model with some regularization (alpha)
mlp = MLPClassifier(alpha=0.01, random_state=42, max_iter=500)

# Train the MLP model on the resampled training data
mlp.fit(X_resampled_smote, y_resampled_smote)

# Initial evaluation of the model
y_pred_mlp = mlp.predict(X_test)
print("Initial MLP Accuracy:", accuracy_score(y_test, y_pred_mlp))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_mlp))
print("Classification Report:\n", classification_report(y_test, y_pred_mlp))

from sklearn.neural_network import MLPClassifier

# Updated MLP with increased regularization, early stopping, and reduced complexity
mlp = MLPClassifier(
    hidden_layer_sizes=(50, 50),  # Fewer layers and/or neurons
    activation='relu',
    solver='adam',
    alpha=0.1,  # Increase regularization strength
    learning_rate='adaptive',
    learning_rate_init=0.0005,  # Lower learning rate
    early_stopping=True,  # Enable early stopping
    n_iter_no_change=10,
    max_iter=1000,  # Increase max_iter to allow convergence with early stopping
    random_state=42
)

# Fit the updated MLP on the resampled training data
mlp.fit(X_resampled_smote, y_resampled_smote)

# Predict and evaluate
y_pred_mlp = mlp.predict(X_test)
print("Updated MLP Accuracy:", accuracy_score(y_test, y_pred_mlp))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_mlp))
print("Classification Report:\n", classification_report(y_test, y_pred_mlp))


# Plot the learning curve for the best MLP model
def plot_learning_curve(model, X, y, title):
    train_sizes, train_scores, test_scores = learning_curve(model, X, y, cv=5, n_jobs=-1,
                                                            train_sizes=np.linspace(0.1, 1.0, 10),
                                                            scoring='accuracy')
    train_scores_mean = np.mean(train_scores, axis=1)
    test_scores_mean = np.mean(test_scores, axis=1)

    plt.figure(figsize=(10, 6))
    plt.plot(train_sizes, train_scores_mean, label='Training Accuracy', color='blue')
    plt.plot(train_sizes, test_scores_mean, label='Testing Accuracy', color='orange')
    plt.title(title)
    plt.xlabel('Training Set Size')
    plt.ylabel('Accuracy')
    plt.legend()
    plt.grid()
    plt.show()

# Learning curve
plot_learning_curve(mlp, X_resampled_smote, y_resampled_smote, 'Updated MLP Learning Curve')
```