Improving the homogeneity of brachytherapy treatment plans generated by BRIGHT using a hotspot registration method based on connected component analysis

Joost Commandeur

Technische Universiteit Delft

TU Delft

Improving the homogeneity of brachytherapy treatment plans generated by BRIGHT using a hotspot registration method based on connected component analysis

by

Joost Commandeur

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on 05 November 2021, 09:30

Student number:	4381149	
Project duration:	November 16, 2020 – November 05, 2021	
Thesis committee:	Prof. dr. P.A.N. Bosman	CWI, TU Delft, supervisor
	Dr. T. Alderliesten	LUMC, supervisor
	Prof. dr. A. Panichella	TU Delft
	A. Bouter	CWI
	L.R.M. Dickhoff	LUMC

This thesis is confidential and cannot be made public until November 05, 2021.

TU Delft

Preface

For the last 12 months I have dived deep into the world of brachytherapy for prostate cancer. It has been an incredibly interesting adventure to say the least, with one of its highlights: a visit to Amsterdam Medical Center where I could see (and even helping during some of) the steps of a treatment procedure for a patient. This adventure has resulted in this master thesis which, after more than 7 years, is the closing act of my student life which I enjoyed immensely.

I could not have done this work by myself and therefore I would like to express my gratitude to everyone who has supported me during the process. In particular I would like to thank the following people in no particular order. Prof. Dr. Peter Bosman for his extensive feedback on the direction of my research and for the patience he had when I was again taking too long whilst giving one of my many 'update' presentations. Dr. Tanja Alderliesten for her keen eye in spotting required changes in my work and her subtle remarks on my 'too short through the corner' thoughts on clinical (data) procedures. MSc Leah Dickhoff for her critical questions on the assumptions I made and for her down-to-earth mentality when I was overthinking. Finally, I would like to thank MSc Anton Bouter for his expertise on everything for which I needed help (even when it was 02.00am, or 09.00pm on a Sunday) and his ability to always ask the right questions.

Finally I would like to thank everyone who has made my student life absolutely epic!

Joost Commandeur
Rotterdam, November 2021

Executive summary

High-dose rate brachytherapy for prostate cancer is a radiation treatment method that radiates the tumour from inside the body. It does so by making use of catheters, which are inserted in the prostate. A radioactive source is guided through these catheters and stopped at certain positions, called dwell positions, to give off radiation. The time that the radioactive source is stopped at a dwell position is called a dwell time. The longer the dwell time, the more radiation is given off at that dwell position. A treatment plan for this form of brachytherapy consists of a set of dwell times for the dwell positions. The goal of the treatment is to cover the prostate and the seminal vesicles with radiation as well as possible according to a prescribed dose, whilst sparing surrounding organs from receiving too much radiation. These goals are captured in the clinical protocol, which specify radiation limits and targets for all organs involved. The construction of treatment plans can be quite cumbersome and difficult due to the inherent trade-off between covering and sparing. For this reason, automation is applied.

On such automation is BRIGHT, **BR**achytherapy via Artificially **I**ntelligent **G**OMEA **H**euristic based **T**reatment planning. BRIGHT is a multi-objective real-valued evolutionary algorithm that finds a set of high quality treatment plans with different trade-offs between coverage and sparing in a short amount of time. Since March 2020, BRIGHT has been in use at the Amsterdam University Medical Centers, location AMC, and has proven its value. BRIGHT has outperformed the previous clinical practice in a clinical observer study.

Even-though BRIGHT outperformed the previous clinical practice, in practice manual adjustments to the treatment plans are still done to meet additional preferences. Particularly, the minimization of hotspots. Hotspots are contiguous volumes of high dose. Hotspots can result in negative side-effects for the patients and should therefore be prevented. The goal of this thesis is to explore means to reduce hotspots in BRIGHT, while minimally deteriorating the goals in the clinical protocol.

In the first part of this thesis the consequences and causes of hotspots are discussed. The reason why BRIGHT produces hotspots is the simple fact that it is allowed to do so. The clinical protocol that it optimizes does not account for hotspots, since it is build up of aggregate metrics which do not account for contiguousness of high dose. In literature no adequate metrics have been found that capture hotspots. A new hotspot metric is proposed in this thesis, called the Hotspot Size Index (HSI) which directly captures the hotspots. To use this metric in BRIGHT, a hotspot registration method is developed which makes use of the graph-based connected component algorithm Afforest.

In the second part of this thesis, different options for adapting BRIGHT to mitigate hotspots are explored. These options consist of both type of adaptation, i.e. adding a constraint, an objective or augmenting the current objectives, as well as what type of metric should be used to steer on. The addition of an extra objective is chosen for in this thesis. For the type of metric to steer on, four different options are explored. Two metrics are based on the dwell times, which are the Dwell Time Modulation Restriction (DTMR) and the Dwell Length Duration Modulation (DLDM). DTMR is a method used in clinics around the world and DLDM is a new proposed metric in this paper. Next to the two dwell time based metrics, there are two metrics that are based on the calculated received dose. These are the HSI and the sum of extra Volume-indices. Volume-indices are used in the clinical protocol to formulate goals. The extra added Volume indices are targeted towards higher dose, but do not account for contiguousness.

In the last part of this thesis, the different BRIGHT adaptations are tested. From the results it became clear that steering directly upon the HSI successfully addresses the goal of reducing hotspots, while minimally deteriorating the goals in the clinical protocol. For smaller hotspots it did have more difficulty. The sum of extra added Volume-indices also resulted in reduced hotspots and better performance for smaller hotspots, but suffered from more deterioration in clinical protocol goals. The DTMR also resulted in a reduction in hotspots for some of the patients, but was less consistent and also suffered from deterioration. Lastly, the DLDM was least suitable to reduce hotspots, although it did result in less hotspot for some patients.

The final take-away of this thesis is that we successfully adapted BRIGHT to mitigate hotspots, while minimally deteriorating the goals specified in the clinical protocol.

Contents

1 Introduction

Artificial Intelligence (AI) has seen a great increase in popularity and development in the last years. AI is now being applied in an increasing number of fields, ranging from autonomous cars to fraud detection at banks. Healthcare is another field where AI is gaining traction. Applying AI in healthcare can for instance replace tasks that are very tedious, time consuming or even impossible for humans. One of these AI healthcare advancements is BRIGHT, **BR**achytherapy via Artificially **I**ntelligent **G**OMEA **H**euristic based **T**reatment planning, which is applied in radiation oncology for prostate cancer. The subject of this thesis is the exploration of potential improvements of BRIGHT, based on preferences expressed by the clinic in which it is used.

In this chapter, the background of this master thesis will be discussed in the following order. Firstly, an introduction will be given on prostate cancer and one of its treatment modalities i.e. high-dose rate brachytherapy. Secondly, the clinical workflow at the Amsterdam University Medical Centers, location AMC (AMC), which makes use of BRIGHT, will be discussed. Thirdly, BRIGHT will be elaborated on, followed by the problem statement, research questions, and set-up of this research.

1.1. Brachytherapy for prostate cancer

Prostate cancer is the second most occurring type of cancer being diagnosed in men in the world (Ferlay et al., 2019). The number of patients receiving a prostate cancer diagnosis is expected to grow as the world population and its life expectancy grows (Schröder et al., 2012). There are several different methods of treatment in use today, ranging from surgical procedures to remove the prostate entirely, to irradiating the tumor to stop the growth and destoy the tumor cells. A common method of treating prostate cancer is the combination of External Beam Radiotherapy (EBRT) and brachytherapy. In recent years brachytherapy, a type of internal radiation cancer treatment, is being used more and more as a mono-therapy (Yoshioka et al., 2013). The goal of radiotherapy, such as EBRT and brachytherapy, is to damage the DNA of the cancerous cells such that the cells stop dividing and eventually die. Cells have some tolerance for radiation, therefore the cancerous cells must receive enough radiation to sufficiently damage their DNA. In this thesis the focus will be on brachytherapy.

1.1.1. High-Dose Rate Brachytherapy

Brachytherapy is a treatment method that delivers radiation to cancer tissue from inside the body, rather than from outside the body as is done in EBRT. This is done by placing the radioactive source inside the body. The advantage of this type of treatment is that it is able to more specifically target the tumour and thereby reduce the amount of radiation received by the surrounding tissue and organs, since the radioactive source is located in the target volume. High-dose rate brachytherapy (HDR-BT) is a type of brachytherapy that is characterized by the short (<30 min) but high dosage treatment. In HDR-BT a radioactive source is placed inside the body using catheters (hollow needles). Through these catheters, the radioactive source is guided, and halted at predetermined locations, called the dwell positions. The duration of halting the radioactive source at a dwell position is called the dwell time. The longer the dwell time, the more radiation is delivered to the surroundings of the dwell position. A visual representation of what this type of treatment looks like is presented in Figure 1.1.

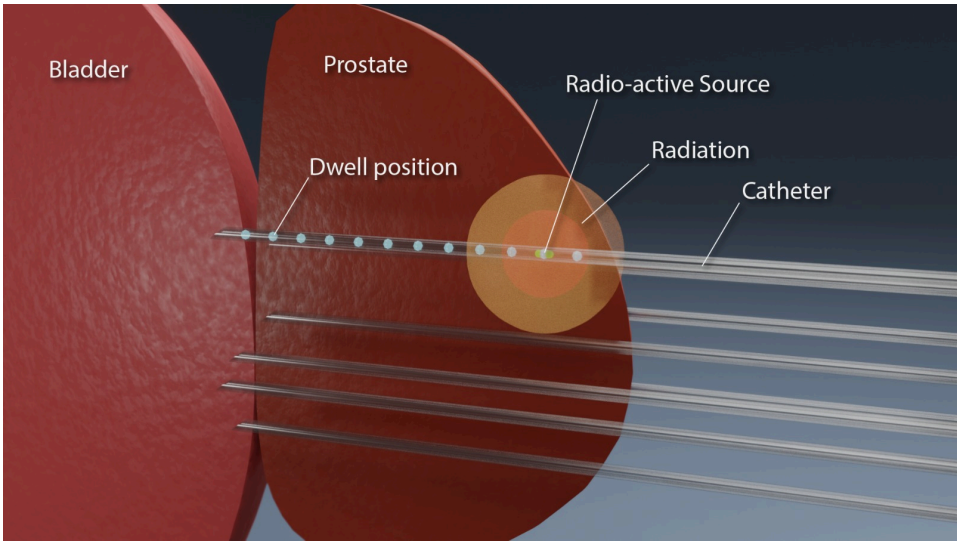

Figure 1.1: Visual representation of prostate HDR-BT using a single dwell position. Catheters (silver/grey cylinders) are inserted in the prostate which is neighboured by the bladder. Note that only one dwell position is used, for simplicity.

The goal of HDR-BT is to deliver the prescribed radiation dose, measured in Gy (Gray), to the target volumes whilst sparing surrounding healthy Organ At Risk (OARs) as well as possible. To achieve this, a treatment plan must be constructed. This plan consists of the dwell times of the radioactive source at the dwell positions, or in other words, how long the radioactive source stays at each of the dwell positions. The goal is to sufficiently treat the tumour, while sparing the OARs as best as possible. In these catheters there are several positions in which the radioactive source can reside. The longer the radioactive source stays at a dwell position, the more radiation is delivered in the proximity of that dwell position. The quality of a treatment plan is assessed using different quantitative metrics per organ and by visually inspecting the dose distribution. These quantitative metrics are described in the clinical protocol.

1.1.2. Clinical protocol

The plan to be constructed should at least adhere to the treatment planning criteria, captured in the clinical protocol, which specify sparing and coverage goals. These criteria can vary between different hospitals. Within the AMC, the criteria are based on dose-volume indices (DVIs) which are adapted from the GEC-ESTRO HDR Prostate Guidelines (Hoskin et al., 2013). These dose-volume indices are specified for the target volumes and the OARs. The target volumes in HDR-BT are the prostate and potentially the seminal vesicles. The OARs are all organs involved, namely the prostate, seminal vesicles, bladder, rectum and urethra. The criteria to adhere to, DVIs, are generally described using a prescription dose. A volume index (V) is a way of describing how much volume should receive less or more than a certain percentage of the prescribed dose. For example, $V_{100\%}^{Prostate} > 95\%$ specifies that at least 95% of the prostate volume should receive at least 100% of the prescribed dose. This can then be seen as a coverage aim. An example of a sparing aim would then be $V_{150\%}^{Prostate} < 50\%$, which specifies that no more than 50% of the target volume can receive 150% or more of the prescribed dose. Next to the volume indices there are dose indices (D), which define the dose received by the most irradiated sub-volume of an organ. For example, $D_{1cm3}^{Bladder} < 86\%$ states that the highest irradiated cumulative 1 cubic centimeter of bladder tissue should not receive more than 86% of the prescribed dose. This dose index can thus be seen as a sparing index. An example of a dose-coverage index is $D_{90\%}^{Prostate} > 100\%$ which states that the highest irradiated 90% sub-volume of the prostate should receive at least 100% of the prescribed dose. Table 1.1 shows the protocol as used by the AMC for the prostate with a prescribed dose of 15 Gy.

Table 1.1: Clinical protocol as used by the AMC for HDR-BT of prostate cancer with a prescribed dose of 15 Gy. Protocol is split out in columns per organ.

Prostate	Bladder	Rectum	Urethra	Seminal Vesicles
$D_{90\%} > 100\%$	$D_{1cm3} < 86.7\%$	$D_{1cm3} < 73.3\%$	$D_{0.1cm3} < 120\%$	$V_{73} > 95\%$
$V_{100} > 95\%$	$D_{2cm3} < 80\%$	$D_{2cm3} < 63.3\%$	$D_{30\%} < 110\%$	
$V_{150} < 40\%$				
$V_{200} < 15\%$				

1.2. Clinical workflow at the AMC

To better understand the treatment procedure and to give more background information on HDR-BT, the clinical workflow at the AMC for HDR-BT is described here.

The treatment starts with the placement of the catheters. Catheters are inserted into the body via the perineum, the area between the anus and the scrotum, using ultrasound imaging to give an indication of the placement. The number of catheters used at the AMC is typically around 20. Using more catheters gives the advantage of more degrees of freedom when constructing the treatment plan, but comes at the cost of increased risk of complications due to more insertions required. During insertion, the catheters are placed as optimally as possible such that the treatment plan can be constructed as well as possible for the patient at hand. Due to difficulties arising from the internal configuration of the patient and the inability of the catheters to bend beyond a certain angle, it is not always possible to create an implant (configuration of inserted catheters) which allows for a treatment plan that adheres to all clinical criteria. The best possible plan which can be achieved depends on the combination of implant configuration and the anatomy of the patient.

After the catheters have been placed, a Magnetic Resonance Imaging (MRI) scan is made. This is done to capture where the catheters have been placed in relation to the anatomy of the patient. These images are required for making the treatment plan and calculating how much radiation will be received, given a treatment plan, by the different organs involved. The closer an organ is located to a dwell position, the more radiation it will receive from it. To enable calculation of the dose received, organs and other volumes of interest are delineated/contoured and catheters are reconstructed in the MRI images.

The treatment plans will be constructed after the delineations are made. Since April 2020, the AMC makes use of BRIGHT, to automatically generate treatment plans. BRIGHT produces a set of high quality treatment plans with different trade-offs between coverage of the tumour and sparing of the OARs. From these treatment plans a selection, typically between 1 and 5 plans, is made based on the achieved DVI values with different trade-offs being made. This selection of plans is then exported, for dose distribution inspection and possible manual adjustment to Oncentra Brachy (Elekta AB, Stockholm, Sweden) (OB). OB is a software program that visualizes the delivered dose by showing iso-dose lines over the delineated MRI scans, and shows the achieved DVI values as recalculated by OB. OB is a CE certified BT planning system and therefore used for the final assessment. Adjustments of the treatment plan are made by the clinical expert if desired after a single plan has been selected. These adjustments are mainly motivated by patient-specific circumstances that motivate the clinical expert to deviate from the clinical protocol or to improve the homogeneity of the planned dose (Barten et al., 2021). An adjustment can for instance be to deliver less dose to the bladder region because of other clinical consideration specific to that patient.

After approval of a treatment plan by the team of clinical experts, the patient will undergo radiation treatment. After treatment, the catheters are removed. The time between inserting the implant and performing the radiation therapy should be minimized. This is due to several reasons. One of them is minimizing the discomfort for the patient as the catheters will only be removed after radiation has been performed. Another reason is to minimize the risk of deformations in either the implant configuration due to movement or deformations in the anatomy of the patient (i.e. bowel movements). This is why there is a motivation from the clinical experts to automate procedures to speed up the process, besides the main motivation of improving treatment plan quality.

1.3. BRIGHT

An automation that has been successfully put into clinical practice is BRIGHT. BRIGHT both improves the treatment plan quality and the required time to create a treatment plan. This is done by the automatic creation of high-quality treatment plans, and has been in use since March 2020. As of July 2020 it has been used for 12 patients.

Creating treatment plans manually, also called forward planning, is a tedious and complex task due to two main reasons. First of all, there is a large number of dwell positions in which the radioactive source can reside. For the prostate cancer patients in the AMC treated between April 2020 and July 2021, between 17 to 24 catheters are used. In each of these catheters the stepping distance between dwell positions is 2.0 mm (AMC default) These catheters have a size of 400 mm, which means 200 positions per catheter. However, not all positions are relevant. Positions that are far away from the target organs are disabled since they do not have to be used for delivering the prescribed dose. In practice this results in a median of 333 active positions (range 216–429). Having this number of variables to be set with a real-valued time can be a complex and time-consuming task for humans to perform. The second aspect that makes manually creation of treatment plans difficult is the inherent trade-off in plan quality. Both the coverage of the target volumes as well as the sparing of the OARs need to be taken into account. These goals, however, are conflicting since increasing a dwell time will result in more radiation delivered which is advantageous for covering, but potentially disadvantageous for sparing.

Due to these difficulties, optimization is used to automatically find the dwell times to achieve an acceptable treatment plan. Creating treatment plans in an automated way by finding a plan that satisfies the criteria is called inverse planning. Several methods for optimizing treatment plan(s) have been proposed in literature. In most cases the problem is formulated as an optimization problem. The formulation is done either with a single objective (Lessard and Pouliot, 2001; Lahanas et al., 2003) or in a bi-objective manner (Luong et al., 2017), but more objectives could be used.

1.3.1. Objective function formulation

For the single objective formulation the most common method is the linear penalty model (LPM), which is employed by HIPO (Karabis et al., 2005) and IPSA (Lessard and Pouliot, 2001). This model penalizes the deviation from the received and prescribed dose values by multiple points in the volume (dose points) based on the treatment planning criteria. This is an easy to solve model, but has been shown to weakly correlate with the planning criteria (Morén et al., 2018a). Instead of the single-objective approach, BRIGHT makes use of a bi-objective objective function formulation. This formulation directly captures all treatment planning criteria in a worst-case manner, and therefore correlates strongly with the criteria.

Based on the clinical protocol, a distinction is made between sparing targets and coverage targets. All criteria, as described in Table 1.1, with a 'greater than' sign are covering targets and all constraints with a less than sign are sparing targets. These targets are captured in two objectives, the least sparing index (LSI) and the least coverage index (LCI). This is done in the following way:

$$LCI(t) = \min\{\delta_c(V_{100}^{Prostate}), \delta_c(D_{90\%}^{Prostate}), \delta_c(V_{73}^{Vesicles})\} \tag{1.1}$$

$$\begin{aligned} LSI(t) = \min\{&\delta_s(V_{150}^{Prostate}), \delta_s(V_{200}^{Prostate}), \delta_s(D_{1cm3}^{Bladder}), \delta_s(D_{2cm3}^{Bladder}), \\ &\delta_s(D_{1cm3}^{Rectum}), \delta_s(D_{2cm3}^{Rectum}), \delta_s(D_{0.1cm3}^{Urethra}), \delta_s(D_{30\%}^{Urethra})\} \end{aligned} \tag{1.2}$$

Where:

$$\delta_c(V_d^o) = \frac{V_d^o - V_d^{o,min}}{V_d^{o,min}}$$

$$\delta_s(D_v^o) = \frac{D_v^{o,max} - D_v^o}{D_v^{o,max}}$$

$$\delta_s(V_d^o) = \frac{V_d^{o,max} - V_d^o}{V_d^{o,max}}$$

The LCI and LSI values are determined by the worst performing sub-function, where sub-functions are of the form $\delta_c(V_d^o), \delta_s(D_v^o)$, or $\delta_s(V_d^o)$. These sub-functions calculate the distance between the achieved value for a planning aim for a particular plan and the minimum or maximum allowed value. For example, when in a treatment plan 90% of the prostate volume receives at least 102% of the prescribed dose where the aim was at least 100%, then the distance for that criteria is 2%. When combining these distances in a worst-case manner, and if the value for LCI is 2%, then the other LCI sub-functions have achieved at least a positive distance of 2% with its minimally aspired value. These distances are normalized to their target value, to allow for a better comparison of criteria. The values of the criteria are calculated using dose calculation points (DCPs). DCPs are randomly sampled points in target and sparing volumes, for which the dose is calculated. Using these DCPs, the metrics of the LCI and LSI can be approximated.

The advantage of the used objective function formulation, besides the direct link with the criteria, is that it offers more insight into the trade-off between coverage and sparing. In general, a multi/bi-objective optimization problem does not have a single solution as the final answer. The optimization results in a Pareto approximation front, which is a set of non-dominated solutions. A non-dominated solution is a solution for which there is no other solution in the set that is at least equal in all objectives and better in at least one objective. An example of such a Pareto approximation set, which consists of non-dominated solutions is shown in Figure 1.2.

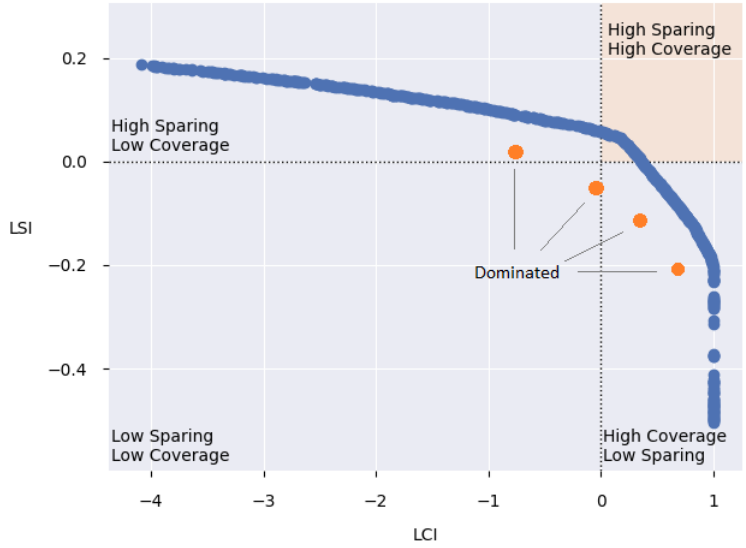

Figure 1.2: Example Pareto approximation front of BRIGHT for HDR-BT for prostate cancer. Each blue dot represents a single treatment plan of the Pareto approximation front with different values for the D- and V-indices. The Pareto approximation shows the trade-off between covering and sparing of organs. The orange region in the top right corner is called the 'golden corner'. In this region all treatment plans adhere to all treatment planning aims, and thus have a positive LCI and LSI. The orange dots indicate treatment plans that would not make part of the Pareto approximation front, since they would be dominated.

In the Pareto approximation front it can be seen that no treatment plan can be found for which there is another treatment plan that dominates it.

1.3.2. Optimization Procedure

The goal of BRIGHT is to focus on the trade-off between coverage and sparing, thereby finding a large number of high-quality treatment plans within a short amount of time. This allows the clinical experts to spend their time on selecting the preferred treatment plan for the patient at hand, rather than adjusting a single one to be as acceptable as possible in the limited amount of time. For this purpose, the underlying optimization procedure of BRIGHT is an Evolutionary Algorithm (EA) (Holland et al., 1992, De Jong, 1975). An EA is a meta search heuristic.

The general idea behind EAs for optimization is that it maintains a *population* of solutions. With every generation it will try to improve the *fitness* of each solution as expressed by the values for the

objectives. To do this the algorithm will apply *variation* to the population, thereby creating *offspring*. Variation is done by adjusting parts of the solution. This can be done in an intelligent way by for instance making use of the learned characteristics of the problem at hand and its potential solutions. After this variation, a from of selection is applied to promote the features of the best solutions given the current population. To determine the best solution, given 2 or more objectives, a general approach is to use Pareto dominance. Pareto dominance means that only if a solution strictly dominates another solution, meaning that one solution is at least equal in all objectives and superior in at least 1 objective, then it is better than the other.

BRIGHT is based on an EA called MO-RV-GOMEA (Multi-Objective, Real-Valued, Gene-Pool Optimal Mixing Evolutionary Algorithm). Specific elements that are of importance to this research will be discussed in more detail in Section 2.2. In theory the algorithm could run forever and still find new solutions that are ever so slightly better than the previous ones or that fill the gap between two existing solutions. Given the time requirements, this is however not feasible. To terminate the algorithm, a termination condition has to be set. There are multiple ways of setting a termination condition, such as maximum number of generations. Given the clinical time requirement, a maximum time termination condition is used in the clinic. At the AMC this is 3 minutes, as that has been shown to allow BRIGHT to reach Pareto approximation fronts in which no further meaningful improvements are found with more time (Bouter et al., 2019).

To improve the speed of the algorithm, parallelization on a Graphics Processing Unit (GPU) has been applied. The specifics of this can be found in Bouter et al., 2019. The main concept that GPU parallelization exploits is that solutions can be adjusted and evaluated independently from each other in a single generation.

1.4. Problem Statement

BRIGHT has been proven to be an effective automated treatment planning method for prostate HDR-BT (Maree et al., 2019). In a retrospective observer study based on 18 different patients, it was shown that the plans generated by BRIGHT were considered equal or superior to the used clinical plans from the previous treatment planning process. The observers also stated that they valued the comparison of multiple plans with the different trade-offs, as compared to the single plan from the previous used method. These multiple plans showed them novel insight into trade-offs for the patient. BRIGHT has been taken in use for the first patient in March, 2020 and 12 patients have been treated with the help of BRIGHT since July, 2021.

For plans in the golden corner, which adhere to all clinical criteria, clinical experts may use their expertise to judge plans based on criteria that are not included in the clinical protocol. When this is the case, other aspects than the ones stated in the clinical protocol are considered and adjusted for. In the evaluation study based on the patients treated with the help of BRIGHT it became apparent that the plans generated by BRIGHT did not always fully satisfy the preferences by the clinical experts (Barten et al., 2021) even though plans adhered to the clinical protocol. One of the aspects, that was targeted with manual adjustments after optimization, was the homogeneity of the treatment plans. In the current set-up of BRIGHT for HDR-BT, homogeneity is not explicitly accounted for. This raises the main research question and problem statement:

Research Question: Problem statement

How can BRIGHT best be improved for the homogeneity of the resulting treatment plans for high dose rate brachytherapy for prostate cancer?

1.5. Research Questions

The problem statement raises a number of questions to be answered. To start, it is not entirely clear what is meant exactly with homogeneity. Different clinical experts respond to questions regarding this topic with different answers. So the first question to be answered is:

Research Question: Why homogeneity?

What is homogeneity in HDR-BT treatment plans for prostate cancer and why is it desired?

When it is clear what the clinical experts aim for with regards to homogeneity it must be quantified. This quantification must be as good in line with the aim as possible. So this raises the question:

Research Question: Measuring homogeneity

How can and should homogeneity be quantified for HDR-BT treatment plans for prostate cancer?

Given that a quantification of homogeneity is established it is important to understand why non-homogeneous plans can be constructed by the algorithm. This will give insights in understanding the problem as well as possible ways of tackling the issue. The research question to be answered is:

Research Question: Heterogeneity causes

What causes BRIGHT to produce non-homogeneous i.e. heterogeneous treatment plans?

As a follow-up question it is then important to look for ways of improvement. The research question to be answered next is:

Research Question: BRIGHT adjustments

How can BRIGHT be enhanced to potentially produce more homogeneous treatment plans?

There will be multiple aspects that will play a role in evaluating the performance of the different adjustments, such as the required extra time and the ability to capture the preferences of the clinical experts in the resulting treatment plans. This raises two related questions:

Research Question: BRIGHT performance

What are the important performance aspects of BRIGHT for HDR-BT and how do different enhancements score on these aspects?

In order to answer these questions, a review of the literature will be conducted, as well as interviews with medical specialists and algorithmic experiments.

1.6. Outline of this Document

This thesis starts in Chapter 2 with a general explanation of evolutionary algorithm and BRIGHT in particular. Next, in Chapter 3 a description is given of homogeneity in HDR-BT for prostate cancer, its consequences, the causes of in-homogeneity, the current methods for mitigating homogeneity and the current methods of measuring homogeneity. This is followed by a description of a new way of measuring homogeneity and how that can be implemented in BRIGHT. In Chapter 4 different options for potentially improving the homogeneity of the treatment plans from BRIGHT are discussed as well as a method for potentially improving the run-time of these improvements. In chapter 5 the results of some preliminary experiments are discussed which are aimed at finding the right settings for the main

experiments of this research. The experiment set-up of the main research is discussed in Chapter 6. The results from the main experiments are described in chapter 7, which is followed by a discussion of the results in chapter 8 and conclusions together with future work in chapter 9. In the appendices the more detailed results are shown as well as a conference abstract based on this work.

2

Background: Evolutionary Algorithms

In this chapter, background information will be given on the used optimization methods in BRIGHT. This background information helps in understanding the potential improvements of BRIGHT. This chapter will start off in Section 2.1 with a general explanation of evolutionary algorithms (EAs), the algorithmic concept on which BRIGHT is built, and why they are a preferred option for multi-objective optimization will be explained in Section 2.1.1. In Section 2.2, MO-RV-GOMEA will be discussed. MO-RV-GOMEA is a specific type of model based EA for multi-objective real-valued optimization and forms the basis of BRIGHT.

2.1. Evolutionary Algorithms

BRIGHT is build on the concept of EAs. EAs are a class of optimization algorithms that are inspired by the process of evolution in nature. In mimicking nature, EAs maintain a population of individuals, i.e. solutions to the optimization problem. Individuals encode a solution in such a way that it can altered during optimization. An example of such an encoding is an array of real-valued variables, where each variable corresponds to a variable in the objective function. During optimization, new generations of the population are generated based on the previous generation. The goal of creating a new generation, also called offspring, is to improve upon the previous generation. To improve upon the previous generation, information about the fitness of the individuals in the previous generation is exploited. The fitness of an individual describes the quality of the solution, i.e. how well it performs on the objective function from the optimization problem. The calculation of the fitness value of an individual is called an evaluation. The exploitation of fitness information is generally done through selection and variation. In selection, all individuals are evaluated on the objective function and the best performing individuals are selected for variation, the other individuals are discarded. This selection is based on natural selection from nature, where the strongest/best adapted individuals survive and generate more offspring and thereby pass their genes onto the new generation. After selection, variation is applied on the the individuals. This variation will change the solutions, with the aim of improving their fitness values. In variation there are generally two methods, cross-over and mutation. Cross-over can be seen as a child receiving a selection of the genes from both parents. In EAs this means that an individual gets parts of the solution from different individuals in the previous generation. Mutation is applied on parts of the individual to generate new solution features, that were potentially not present in the population. This can also be seen in nature, where random mutations of the DNA of individuals can lead to beneficial trades that render the individual better adapted to its environment. Selection is applied again after variation. This selection step will pressure the optimization to better solutions. The iterations of selection and variation can continue indefinitely, therefore a termination condition is set. This termination condition can be based on multiple aspects of the algorithm, as for instance the maximum allowed run-time or when a certain fitness level has been reached. To start the algorithm, a population must be initialized. This initialization of solutions can be done in a random fashion or more intelligently to 'warm-start' the algorithm, although warm-starting the algorithm can have negative consequences for the course of the optimization procedure. The general outline of EAs is shown in Figure 2.1.

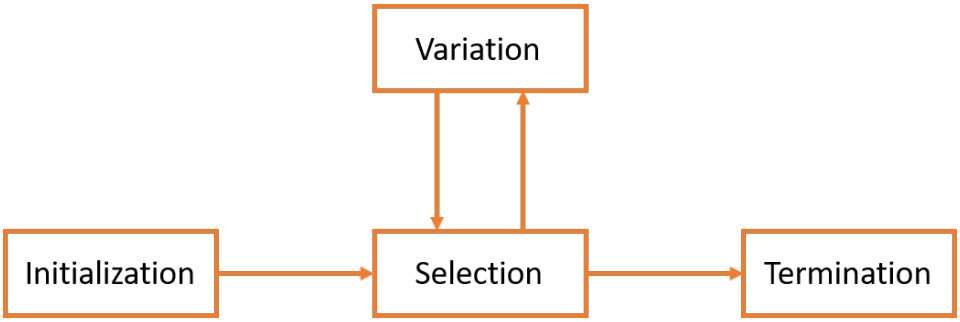

Figure 2.1: General outline of the steps performed in EAs. EAs generally start with the initialization of a set of solutions, called *individuals* which make up a *population*. From these *individuals* the best ones are selected to be used in the generation of *offspring*, i.e. the new population. This generation of *offspring* is done by applying *variation* on these selected best *individuals*. *Variation* generally consists of *cross-over* and *mutation*. After *variation*, the best solutions from the offspring are selected for another round of *variation*. This procedure of selection and variation can continue indefinitely, therefore a termination condition is set. If this condition is met, then the algorithm terminates and returns the best solutions.

2.1.1. Evolutionary algorithms for multi-objective optimization

Multi-objective optimization problems are a class of problems, for which the goal is to find the best solution(s) according to multiple objectives. As compared to a single-objective optimization problem, multi-objective optimization problems generally do not have a single best solution, but multiple solutions that show a trade-off between the objectives. These solutions are called Pareto-optimal solutions. Optimization techniques that are state-of-the-art for single-optimization problems do not generally work well on these set of problems, as they will only find a single solution. To make single optimization techniques, such as linear optimization algorithms, work for multi-objective optimization they would be required to be applied repetitively (Deb, 2014). Since EAs make use of and produce a population of solutions, rather than a single solution, they are better equipped to find Pareto-optimal solutions and therefore a popular method for multi-objective optimization (Deb, 2014). The problem of finding treatment plans is a multi-objective problem, as there is a trade-off between organ sparing and clinical target coverage. Therefore, an EA is used in BRIGHT.

2.2. MO-RV-GOMEA

The specific EA that BRIGHT is based on is called MO-RV-GOMEA, Multi-Objective Real-Valued Gene-pool Optimal Mixing Evolutionary Algorithm. MO-RV-GOMEA is an EA that is capable of optimizing real-valued multi-objective optimization problems. Treatment plans, the individuals to optimize, consist of a set of dwell times which are real-valued. MO-RV-GOMEA is a very powerful multi-objective EA, that has been shown to be the state-of-the-art EA for multi-objective optimization of high-dose-rate brachytherapy planning for prostate cancer treatment (Luong et al., 2018). MO-RV-GOMEA is effective in exploiting the dependencies between the decision variables and uses that to efficiently solve the planning problem. Furthermore, the variation method used in MO-RV-GOMEA allows for partial evaluations of the objective function which makes the calculation of fitness values of the offspring more efficient. Not all optimization problems allow for partial evaluations of the objective function. To enable this, the objective function must be understood, if this is the case it can be called Grey-Box optimization in contrast to Black-Box optimization where there is little to no information about the objective function.

A general overview of the implementation details of MO-RV-GOMEA and BRIGHT that are deemed important for this thesis will be discussed in the next subsections. This is structured according to the general concepts of variation, selection and individuals of EAs. The full overview of implementation details of MO-RV-GOMEA can be found in Bouter et al., 2017. The full overview of implementation details of BRIGHT can be found in Bouter et al., 2019, the used method in BRIGHT for partial evaluations can be found in Bouter et al., 2018.

2.2.1. Individuals, population and elitist archive

The individuals in BRIGHT are the treatment plans. These treatment plans consist of a set of dwell times for the set of dwell positions in the catheters. The representation of an individual in BRIGHT is an array of real valued variables. Each variable represents the set dwell time for a specific dwell position.

Next to the population of individuals on which the variation is applied, BRIGHT also stores a set

of the best found non-dominated individuals that have been found throughout the generations. This is called the elitist archive. A non-dominated individual means that there is no other individual that performs at least equal in all objectives and better in at least one. This dominance relation is also called Pareto-dominance. After every round of variations in individuals it is checked whether the newly created individuals should belong in the elitist archive by comparing them to every individual in the elitist archive. If an individual is added to the elitist archive, all individuals that are dominated by the new individual are deleted from the archive. The elitist archive has a maximum capacity, since it would otherwise potentially grow indefinitely. If the maximum capacity has been reached, then individuals that are more similar to each other, with regards to their fitness values, are removed.

2.2.2. Selection and variation

In MO-RV-GOMEA, the selection and variation procedures work as follows. Based on the best 35% of individuals in the population, q different equal sized clusters are identified in the population. A visual representation of this is shown in Figure 2.2. For these q different clusters, multi-variate Gaussian distributions are fitted on the dwell times of individuals in the cluster. These Gaussian distributions are fitted on the different FOS-elements (Family of Subsets). FOS elements are subsets of the total set of dwell positions which are considered more dependent on each other. In BRIGHT this means dwell times that are in close proximity to one another. These multi-variate Gaussian distributions will be used to generate the offspring.

Based on these multi-variate Gaussian distributions, new partial solutions are sampled based on the FOS-elements in random order. These partial solutions consists of new dwell times for a FOS-element. This partial solution is then inserted in every individual that belongs to the cluster for which the multi-variate Gaussian was fitted. After insertion, the individuals are evaluated. If the insertion, i.e. variation, has resulted in an improved fitness, then the variation is accepted. For evaluating the fitness of an individual it is not required to evaluate the full objective function, since only a subset of the variables has changed. A partial evaluation of the objective function can be done instead, the specifics of this can be found in Bouter et al., 2018 and Bouter et al., 2019. After all FOS-elements for all clusters have been sampled, and their insertion into the individuals has been evaluated, it is checked if the new individuals should be inserted in the elitist archive as described in Section 2.2.1. The changed population of individuals is now the new generation. Note that no individuals are discarded, as was described in the general EA procedure.

By using the clusters, the improvement of the population is guided in different directions as shown in 2.2. Every cluster has its own pressure direction on the Pareto approximation front, this means that some clusters will push to more organ sparing treatment plans and other clusters to more target coverage treatment plans.

This cycle of selection, clustering, Gaussian fitting and sampling is repeated until the termination condition has been reached. The termination condition for BRIGHT is a maximum allocated run-time, since in the clinic time is limited.

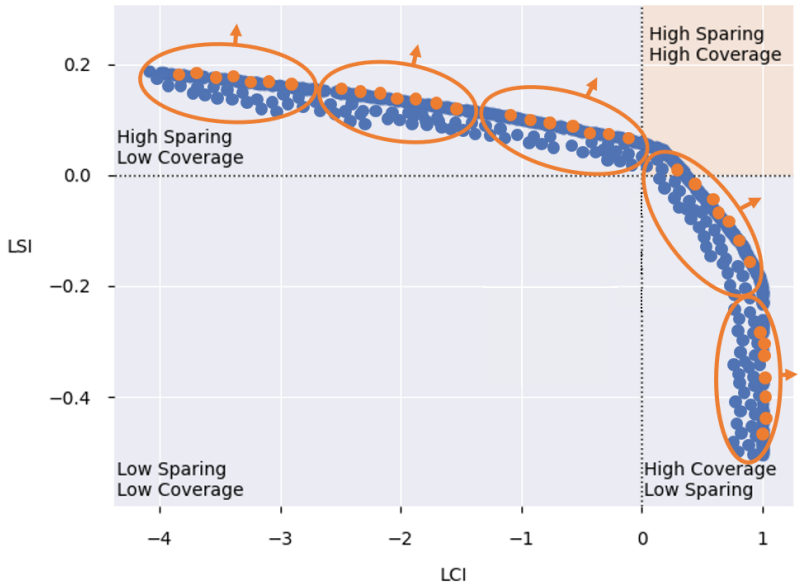

Figure 2.2: Population of individuals, **not only the Pareto approximation front**. Every individual is represented by a blue dot or an orange dot. The equal sized clusters are represented by the orange ovals. Note that every individual is part of a single cluster. The best 35% of individuals that determine a cluster is indicated by the orange dots. The orange arrows point to the direction in which the clusters apply pressure on the Pareto approximation front. The cluster in the top left corner applies more pressure towards the sparing of organs, while the cluster in the bottom right corner applies more pressure towards the coverage of treatment plans.

3

Homogeneity of Treatment Plans

This chapter is dedicated to describing the motivation for promoting homogeneity in high dose rate brachytherapy treatment plans as described by the clinical experts from the AMC. The following questions will be discussed in order of elicitation: What are the different consequences of non-homogeneous treatment plans? How is that caused by non-homogeneity? What mitigation measures and practices are currently in place to mitigate these consequences? What are the shortcomings of these current practices?

3.1. Consequences of non-homogeneous treatment plans

The term homogeneous has several definitions in the dictionary: *of the same or similar kind or nature / of uniform structure or composition throughout / ...*. In the case of treatment plan quality, homogeneity can best be described using the second definition. A homogeneous treatment plan is a treatment plan in which the distribution of radiation dose does not show steep slopes. There are two main downsides associated with non-homogeneous treatment plans which are related to high-dose volumes and to susceptibility to disturbances.

3.1.1. High-Dose Sub-Volume (Radio-Necrosis)

The first downside of non-homogeneous treatment plans is caused by so-called 'hotspots'. Hotspots are contiguous volumes of high dosage. In the evaluation of the use of BRIGHT in the clinic it became apparent that the plans that BRIGHT produces can still contain hotspots, which need to be minimized (Barten et al., 2021). The downside of having hotspots is the risk of radio-necrosis. If there is too much radiation, this can kill cells for which the body is not able to regenerate it in a normal way. This can have several negative side effects such as bladder dysfunction (incontinence, Zakariaee et al., 2017), and erectile dysfunction (impotence, DiBiase et al., 2000). Furthermore, necrosis is toxic for the body and should be avoided.

When speaking to the clinical experts at the AMC, no single minimum dosage value was immediately decided upon for the definition of hotspots. After discussion and evaluation of adjustments made in the clinic, the lower bound value of the hotspots was set to 300% of the prescribed dose (15 Gy) for the target volumes (prostate and seminal vesicles) and 200% of the prescribed dosage for the normal-tissue surrounding the target volumes. For the other OARs no hotspot lower bound was set as no hotspots were formed there. The lower bound volume for a hotspot has been set to 0.1 cm^3, since lower volumes are deemed to be insignificant.

Hotspots are formed in two ways. The most straightforward one is from a single dwell position where the radioactive source resides too long. The second one is the interplay of multiple dwell positions. A single point in, for example, the prostate receives its radiation from multiple dwell positions. The received dose is then the sum of doses received from each of these dwell positions. This makes it possible to form a hotspot without the need for a single very long dwell time in a single dwell position.

A plan could be considered homogeneous when it has a very high dose overall, but this would then lead to very large hotspots. Therefore, in this research these plans are also considered to not be homogeneous.

3.1.2. Susceptibility to Disturbances

The second downside to heterogeneous treatment plans is the higher susceptibility to disturbances and uncertainties. Disturbances and uncertainties can be and are introduced in several ways (Bel et al., 2020). During the clinical workflow it is possible that the catheters are shifted after the creation of the treatment plans due to patient or organ movement. This will result in a different outcome of the treatment plan than that was planned for. Another source of uncertainty is the delineation of the organs. The delineation of the organs is done by hand and will affect the calculated DVI values, as they are based on the volume of the organs which is dictated by the delineation. The reconstruction of the catheters in the MRIs also entails some uncertainties in the dwell positions.

Heterogeneous treatment plans are more susceptible to deviations in anticipated outcomes than homogeneous plans. These plans have more peaks in delivered radiation. A shift of a catheter will cause a change in the received radiation for a specific location in the patient. This becomes a problem if this peak shifts from a place where this peak is allowed to a place where it will cause negative side effects. Heterogeneity is therefore considered to be a bigger problem at the tips of the catheters, as they are closer to the organs at risk. A more homogeneous plan without high peaks suffers less from the problem of susceptibility to deviations and uncertainties.

This study will focus on the mitigation of hotspots rather than the susceptibility to disturbances. Including the latter would increase the scope of this project too much for a master thesis project. Nonetheless, since it is an important aspect of homogeneity, the smoothness of dwell time distribution will be reported on for the experiments. For a full assessment of susceptibility to disturbances it would be required to simulate these disturbances.

With these findings the first research question can be answered:

Research Question: Why homogeneity?

What is homogeneity in HDR-BT for prostate cancer and why is it desired?

Homogeneity of HDR-BT treatment plans for prostate cancer is how smooth the distribution of doses is over the target organs. It is desired both for the prevention of hotspots as well as promoting treatment plan robustness. The first one will be the focus of this thesis.

3.2. Causes of Non-Homogeneous Treatment Plans

The reason why BRIGHT produces plans which can suffer from hotspots is because the desire of the clinical experts regarding hotspots is not directly captured in the clinical protocol. The clinical protocol is based on DVIs. The DVIs are aggregate measures of the dose received by the different volumes of interest. Since they are aggregate measures, two different plans can be constructed which lead to the same DVI values but show different dose distributions (Morén et al., 2018b). A simplified example of what this looks like is shown in Figure 3.1. In the figure, two simplified plans are presented which both have an equal V_{200} index (the percentage of the volume of a region of interest that receives 200% of the prescribed dose) but show a different distribution. For BRIGHT, these plans are considered equal in their quality. This means that BRIGHT has a blind-spot for hotspots.

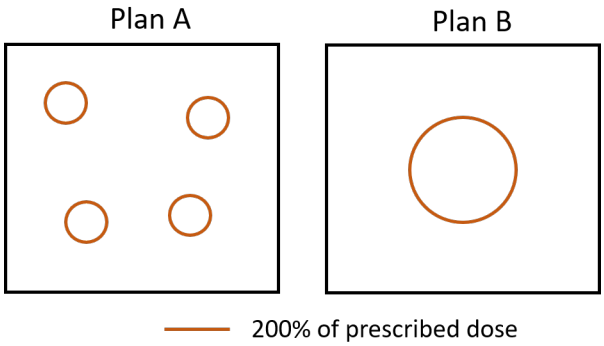

Figure 3.1: Dummy visualisation of iso-dose lines for two different treatment plans with equal DVIs but different distribution of radiation dose. Only the 200% iso-dose lines are shown. The treatment plan on the left has a more scattered distribution of 200% dose as compared to the plan on the right, which is more prone to having hotspots.

The formation of hotspots could potentially be further strengthened through the way BRIGHT is applying variation to treatment plans. This variation is applied to find improvements for the different solutions BRIGHT has in the population. For each treatment plan BRIGHT will simultaneously change the dwell times of dwell positions in FOS-elements. FOS-elements, Family Of Subsets, are subsets of dwell positions which are considered more dependent on each other. This means of exploitation could be described as local improvement, since only a subset of variables is changed at a time and checked for improved solutions. These local changes will contribute to the formation of hotspots. The neighbouring dwell times of the dwell positions of the changed dwell positions are not changed accordingly and therefore more heterogeneity can be introduced. This results in a higher chance of a single or a small number of dwell positions to take on longer dwell times, as it is less likely to change a multitude of dwell positions simultaneously. These longer dwell times can then lead to hotspots. However, these partial changes are the strong-suit of BRIGHT when it comes to run-time efficiency, as it allows for partial evaluations which make it suited for parallelization which greatly improves the run-time (Bouter et al., 2019). In the Linear-Penalty-Model optimization techniques, similar formation of hotspots by using a lower number of active dwell positions but with higher dwell time has been observed (Holm et al., 2012). The optimization techniques for the Linear-Penalty-Model did not make use of partial variations and evaluations.

With these findings the first research question can be answered:

Research Question: Heterogeneity causes

What causes BRIGHT to produce non-homogeneous, i.e. heterogeneous, treatment plans?

Non-homogeneous treatment plans are produced by BRIGHT since it is allowed by the clinical protocol which BRIGHT optimizes on, i.e. BRIGHT has a blind-spot for hotspots. Furthermore, BRIGHT exploits local improvements which can potentially result in more locally connected dose and therefore potentially more and/or larger hotspots. This hypothesis however, will not be further tested in this thesis as it is deemed out of scope.

3.3. Current practices to overcome and prevent non-homogeneous treatment plans

Some mitigation strategies for heterogeneity have been proposed in literature and applied in optimization methods used in practice, which were also used by AMC before BRIGHT. The most important one is Dwell Time Modulation (DTM). DTM has been implemented in different forms in HIPO and IPSA(Lessard and Pouliot, 2001, Lahanas et al., 2003). In HIPO it is called Dwell Time Gradient Restriction, and in IPSA it is called Dwell Time Deviation Constraint. DTM consists of controlling

the amount of deviation that neighbouring dwell positions can have in their dwell times. By reducing the amount of time deviation from one dwell position to another, homogeneity is promoted. The exact implementation of the DTM in HIPO and IPSA is not publicly available.

In the paper by Balvert et al., 2014, several implementations of DTM have been tested to evaluate their effect on the quality of the treatment plans. The impact of DTM was evaluated using both the linear penalty model (LPM) as well as a direct optimization on the DVHs. The DTM was implemented as a hard constraint (DTM restriction, DMTR). This constraint renders all solutions with an excessive value for the DTM metric infeasible. Three different implementations of DTM were evaluated. These three implementations were the relative, absolute and quadratic difference between dwell times of neighbouring dwell positions given a maximum distance. In their paper, the authors experimented with different cut-off values for the constraints. The limitation of this study is that they only had the data available for 3 different patients. Therefore they encouraged other institutions to quantitatively assess the influence of DTMR. The conclusion of the paper was that DTMR was able to reduce the largest contiguous volume of 150% of the prescribed dose, but that it came at the cost of reduced $D^{P}_{90\%}rostate$ which was often rendered unsatisfactory for the used clinical protocol.

What is also important to realize is that DTMR on itself does not guarantee the absence of hotspots, it merely provides a more homogeneous dose distribution. In Figure 3.2 this is illustrated. Plan A performs worse on DTM, since the dwell times show higher deviation, but better on hotspot size. DTM in combination with an adequate measure to reduce high dwell times could potentially result in homogeneous plans without hotspots.

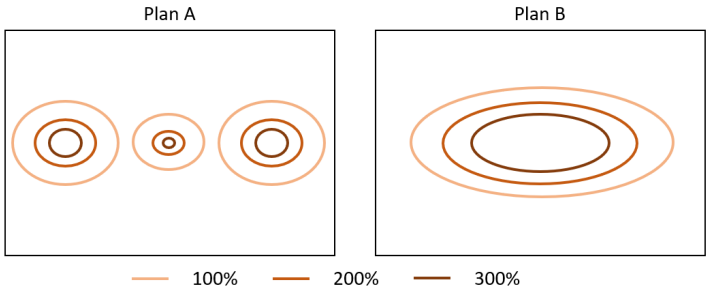

Figure 3.2: Dummy visualisation of iso-dose lines of two different treatment plans with 3 dwell positions. The shown iso-dose lines are 300%, 200% and 100% of the prescribed dose. The treatment plans have a quality difference in DTM values, and inverse quality difference in hotspot size. The treatment plan on the left has smaller hotspots, but more capricious dwell time distribution as compared to the plan on the right.

The DTMR registers the deviation of neighbouring dwell positions. In the paper by Balvert et al., 2014 the deviation of the 2 nearest neighbours was accounted for. Given the spacing of the catheters, i.e., 2 mm stepping distance in catheters between dwell positions and 5 mm spacing between catheters as dictated by the template used to insert the catheters (see Figure 3.3), the 2 nearest neighbours are within the same catheter. The DTMR will be further elaborated on in Section 4.2.1.

Besides the DTMR in the two different LPM algorithms, a common approach for mitigating hotspots (and thereby improving homogeneity) is to visually inspect the resulting iso-dose lines of the resulting treatment plans. By inspecting the iso-dose lines the hotspots are identified. Next, the dwell times that cause these hotspots are manually adjusted.

3.4. Scoring Homogeneity

It is necessary to utilize a method of scoring the treatment plans in order to be able to make quantitative judgments on the treatment play quality in terms of homogeneity. In this subsection, the scoring of homogeneity will be discussed.

3.4.1. Clinical requirements

Before discussing metrics to capture hotspots an overview will be given on what the clinical requirements are with regards to hotspots. As discussed before, homogeneity is an important aspect of HDR-BT. This statement however needs some more clarification, because hotspots are an inherent property of brachytherapy. Whenever a dwell position is activated and the dwell time is set, it is

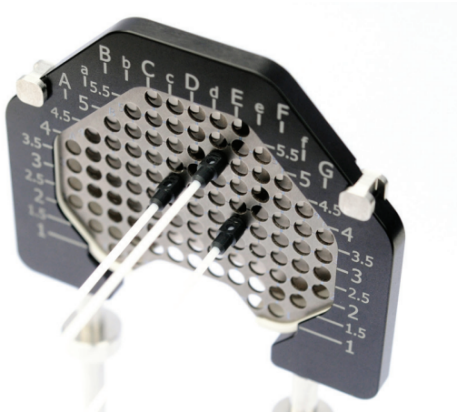

Figure 3.3: Image of the Martinez Brachytherapy Template (source: Elekta, 2021). This template is used to guide and hold the catheters in place during treatment. The distance between the circular cutouts is 5 mm. The template is place in front of the area between the scrotum and the anus, called the perineum.

inevitable that a small hotspot will be formed. This hotspot is a small region in and around the dwell position which will receive a high dose above the prescription dose. This can be seen in Figures 3.4 and 3.5, where all dwell positions have been set to the same small dwell time.

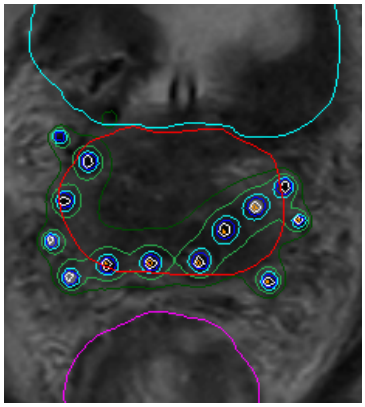

Figure 3.4: Visualisation of a treatment plan with small ”hotspots” showing all iso-dose lines (from inside out: 300%, 200%, 150%, 100% and 80%). The rectum is delineated by the purple line, the prostate by the red line and the bladder by the blue line.

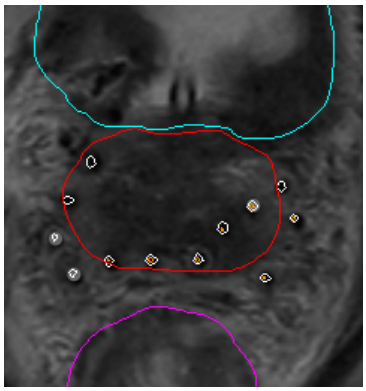

Figure 3.5: Visualisation of a treatment plan with small ”hotspots” showing only 200% (white) and 300% (orange) iso-dose lines. The rectum is delineated by the purple line, the prostate by the red line and the bladder by the blue line.

However, the doses will quickly decrease with the increase in distance to the dwell position, which is called the dose fall-off. The sharp dose fall-off is the strong suit of HDR-BT since it allows for high radiation values in the target volumes, while not radiating the surrounding tissues too much. In Figure 3.6, an example drawn to relative scale of the dose rate fall-off based on the TG-34 model (Rivard et al., 2004) is visualized. The TG-34 model is a widely used method for calculating dose values in brachytherapy.

Because of this contradictory situation of trying to limit hotspots, but ’hotspots’ are the strong suit of HDR-BT, it is required to sharply define what type of dose distribution is deemed undesirable and what type of dose distribution is desired. In general, how should homogeneity/hotspots be measured? The situation which the clinical experts have expressed that should be avoided is large contiguous sub-volumes of high received dose. A single 300% contiguous sub-volume of 2cc is worse than 2 smaller 300% contiguous sub-volumes that sum to 2cc. On top of that, there is also a distinction that should be made between hotspots in the prostate/seminal vesicles and hotspots in other tissue and organs. A contiguous volume of 200% received doses in the prostate is not perceived as a hotspot by the clinical experts, whereas the same high-dose sub-volume in the tissue surrounding the target volumes is

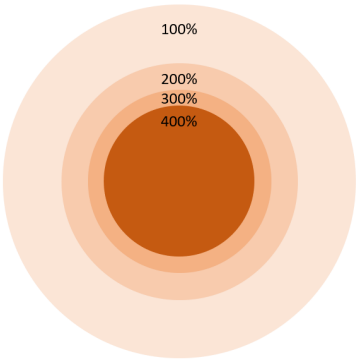

Figure 3.6: Scale visualization of dose-rate fall-off from a single point of radiation based on the TG-34 model for dose calculations Rivard et al., 2004. Different colours indicate different levels of delivered radiation (i.e. 100%, 200%, 300% and 400% of the prescribed dose). The non-linear relation between radius and delivered dose can be seen.

undesirable. In the organs at risk, a hotspot is viewed as a contiguous sub-volume of at least 100%. Furthermore, whenever the hotspot is small enough it is also not perceived as a problem. In the discussions, the clinical experts expressed that a hotspot with a volume lower than 0.1 cc is negligible. As was already shown, there will be small hotspots formed surrounding and in the active dwell positions. These are also not deemed undesirable. Simply setting a lower bound on hotspot volume would however not suffice. When measuring hotspots one would potentially automatically register long hotspots that are within the catheters that exceed the lower bound but are not undesirable. In Figures 3.7 and 3.8, this is visualized. On the right image in the left catheter there are two contiguous sub-volumes of 300%, but only the top one is perceived as a hotspot since it is wider.

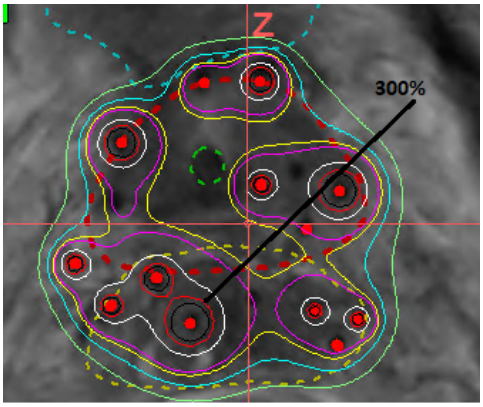

Figure 3.7: Visualisation of iso-dose lines viewed in the transverse direction (from feet to head, back-view for catheters). Black line is a pointer to indicate the red line which depicts the 300% iso-dose line, which is the minimum dose value for hotspots in the prostate and the seminal-vesicles. Red dots show dwell positions. The red dotted line delineates the prostate, the yellow dotted line the rectum, the green dotted line the urethra and the blue dotted line the bladder.

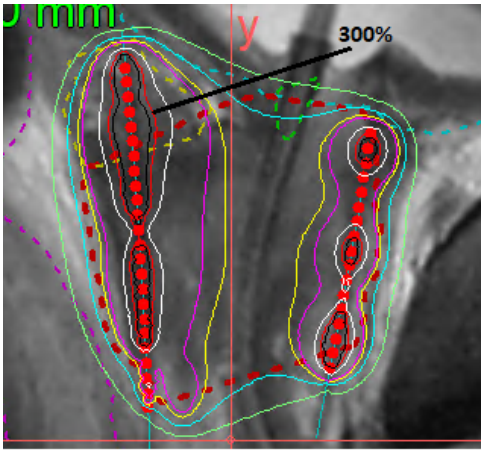

Figure 3.8: Visualisation of iso-dose lines viewed in the coronal direction (from toes to heel, side-view for catheters). Black line is a pointer to indicate the red line which depicts the 300% iso-dose line, which is the minimum dose value for hotspots in the prostate and the seminal-vesicles. Red dots show dwell positions. The red dotted line delineates the prostate, the yellow dotted line the rectum, the green dotted line the urethra and the blue dotted line the bladder.

Therefore, another addition is made to our definition of hotspots, which is that only regions outside 1 mm radius of the center of the dwell positions are considered when registering hotspots. The value of 1 mm is based on the fact that adjacent dwell positions are 2 mm apart. By setting the radius to 1 mm a single large hotspot within the catheters will not be registered. This brings us to the final definition of undesired hotspots, which is summarized in Table 3.1:

	Dose	Volume	Distance*
Target volume	≥300 %	≥0.1 cc	> 1 mm
Normal-Tissue	≥200 %	≥0.1 cc	> 1 mm
Organs at Risk	≥100 %	≥0.1 cc	> 1 mm

Table 3.1: Summary of hotspot definition defined per volume of interest, split out on volume with received dose and *minimum distance from the active dwell positions

3.4.2. Metrics from literature

In literature, several different metrics have been proposed to assess the homogeneity quality of treatment plans. There was and is a need for automatic assessment. Automated assessment allows for the comparison of larger numbers of treatment plans, since it requires less manual labor of either a clinical expert or a researcher as no dose distributions have to be visually inspected slice by slice.

The most common metrics for assessing homogeneity are the V_{150} and V_{200} indices for the prostate. However, as previously discussed, these measures are aggregates and will not differentiate between a single large 150% or 200% contiguous volume and multiple small ones. Based on the V-indices, another metric could be created which looks at the V_{300} or even the V_{400}. This will then also suffer from the lack of spatial accounting. Nonetheless, given dose distribution in hotspot due to the dose-fall off, as shown in Figure 3.6, large hotspots will have larger regions of 400% while they might be non-existent in smaller hotspots. The only downside then is that hotspots that are made up from multiple dwell positions, where the high-dose iso-dose lines overlap, are potentially not accounted for.

Another metric is the Homogeneity Index (HI). The HI has been formulated in various different ways (Kataria et al., 2012, Patel et al., 2020). These can be summarized as the relation between volume percentage of low, normal and high dose. One such formulation, which is used in the paper by Kataria et al., 2012 is shown in Equation 3.1:

$$HI = \frac{D_5\% - D_{90\%}}{D_p} \cdot 100\% \tag{3.1}$$

In this equation, $D_5\%$ and $D_{90\%}$ represent the minimum dose in respectively the strongest irradiated 5% and 90% of the volume of the prostate and D_p represents the prescription dose. Since the different HIs are built upon the same V and D indices as the clinical protocol, they will suffer from the same spatial awareness shortcoming.

The S-index is an improvement on HI. Instead of using only a few D- and V-indices and comparing them, the S-index focuses on the deviation in the Dose Volume Histogram (DVH) curve. The S-index is thus basically the standard deviation of received dosage in the volume over which it is defined. The definition of the S-index is given in Equation 3.2:

$$\text{S-index} = \sqrt{\sum (D_i - D_{mean})^2 \cdot \frac{v_i}{V}} \tag{3.2}$$

where D_i is the radiation received by the i^{th} sampled point in a volume, D_{mean} the mean received dose by all DC points, v_i the volume of the i^{th} DC point and *V* the total volume of all DC points. The S-index thus is standard deviation of the dose distribution normalized for volume. This formulation however also does not take the exact spatial distribution of dosage into account.

Kim et al., 2016 formulated the Bubble Index (BI) with the specific aim of preventing ”hotspot agglomeration”. They defined the BI as ”the maximum rate of change of iso-dose surface area-to-volume ratio (SA:V) with respect to dose”. Unfortunately they did not provide any more explanation or, more importantly, implementation details. In their paper it seems that the Bubble Index is automated, rather than calculated by hand. However, since no implementation details were given or extensive explanation of the results it could not be used in this research.

The last metric was used in Thomas et al., 2007 and Golshan et al., 2014. This metric is an extension upon the normal V index, named the contiguous V index (V^C) and denoted by for example V_{200}^1, where *C = 1* tells us the size of the largest contiguous volume receiving 200% of the prescribed dose and V_{200}^2 the second largest. Although this comes close to the requirements, it is not complete, since it does not account explicitly for the location. The largest hotspot can for instance be formed inside the catheters. Furthermore, they did not provide the implementation of measuring the V^C. It was only stated that it was calculated after optimization using VariSeed, a software program for brachytherapy treatment planning.

3.4.3. Creating a metric

Since the proposed metrics in literature do not match exactly with the requirements as set by the clinical experts, a new metric is proposed, called the Hotspot Size Index (HSI). The 'size' is added to prevent confusion with the already defined Homogeneity Index (HI). The new HSI definition is shown in Equation 3.3.

$$HSI = \sum_{i \in H(v_o, d_o, l_o)} (S_i)^f \tag{3.3}$$

Where S_i denotes the volume of hotspot i, f the exponent factor to allow for increased or decreased accounting of larger hotspots and $H(v_o, d_o, l_o)$ the set of hotspots with a minimum size/volume of v_o, dose d_o, and distance to dwell positions l_o where o represents the value per organ. This means that single hotspots can consist of different lower-bound values on received doses if the hotspot overlap multiple organs/tissues. The factor f is included to allow for more flexibility in measuring hotspots, however this research is limited to using only a factor of 1 resulting in the sum of hotspots which fit the requirements. The HSI value thus expresses what the total weighed volume is of hotspots in a treatment plan.

In this research, the size of all individual hotspots are recorded that are at least of size v and of dosage d_o.

With these findings the first research question can be answered:

Research Question: Measuring homogeneity

How can and should homogeneity be quantified for HDR-BT treatment plans for prostate cancer?

Homogeneity has been quantified in literature but these quantification have failed to directly address the occurrence of hotspots. The literature that did directly measure the size and number of hotspots failed to provide implementation details. In this research a new quantification is proposed in the form of the Hotspot Size Index (HSI), which is the weighted sum of hotspots that satisfy predefined hotspot criteria.

3.5. Hotspot Registration

In this section the automation of calculating the HSI value will be discussed. In order to discuss different potential implementations of measuring homogeneity through the registration of hotspots, it is important to know how the dose distributions of different treatment plans are calculated. In BRIGHT, as in many other automated treatment planning applications, the dose distribution is approximated using dose calculation points.

3.5.1. Dose Calculation Points

Dose calculation points (DCPs) are points that are sampled within the different tissues and organs for which one or more DVI is required to be calculated. Every DCP approximates a sub-volume of the organ or tissue that it is sampled in. Using the DCPs the different DVIs can now be calculated easily by sorting the DCPs per organ on their received doses. Using the sorted array of received dose per DCP, the V_x index can be calculated by determining what percentage of DCPs in an organ has received at least x% radiation. The volume that each DCP represents depends on the number of DCPs sampled and the volume of the organ or tissue that the DCPs are sampling in. Each DCP in a single organ or tissue represents the same volume. In Figure 3.9, a visualization of DCPs approximating a set of organs is presented, where the prostate is red, the bladder is blue, the rectum is orange and the seminal vesicles are light brown.

Figure 3.9: Visualisation of dose calculation points approximating organs. Every dot represents a dose calculation point. The blue dots represent the bladder, the red dots the prostate, the light brown dots the seminal vesicles and the orange dots the rectum.

The DCPs are sampled uniformly at random in the target volume rather than placed on a grid. The motivation for this lies in the ability of the DCPs to approximate the DVIs. In the paper by Niemierko and Goitein, 1990 it was shown that the number of DCPs required to accurately approximate the DVIs was at least 50 times smaller when sampled uniformly at random, compared to a regular grid. The random sampling in BRIGHT is done by sampling in a bounding box of the contours of each organ and accepting a sampled DCP if it falls within the contours and otherwise rejecting it.

3.5.2. Hotspot registration in literature

In the paper by Thomas et al., 2007 the largest hotspots were recorded as found in multiple treatment plans. This was done using the software of VariSeed after optimization. Unfortunately the VariSeed software is not open-source and therefore the implementation could not be inspected for use in this research.

Zakariaee et al., 2016 used another approach to determine hotspot sizes. A Fast-Random-Walker algorithm (Andrews et al., 2010) was adapted to detect the hotspots. Unfortunately no detailed implementation description was given. However, in the Fast-Random-Walker paper it became clear that this procedure required user-interaction and was probabilistic. In this research, the method for hotspot registration should be fully automatic as the available time of the clinical experts to contribute to this research is limited. Furthermore, ideally the method should also be deterministic to minimize the required number of runs for fair comparison of treatment plans.

The last hotspot registration method found in literature is given by Morén et al., 2018b. The premise on which their method is built is that within a small sub-volume, all DCPs are so close to each other that they can be perceived as connected. A high value for the V_{200}^{k}, for each small sub-volume k, then indicates a hotspot in that small sub-volume. Although the implementation of this method is straightforward, it does not satisfy the requirements as discussed earlier. Although the lack of spatial awareness of the DVI-based hotspot/homogeneity metrics is improved, it is not sufficient yet. The interconnection of multiple small sub-volumes is discarded, which results in this method being unable to completely register the full size of the hotspots, which again results in multiple hotspot configurations being possible under the same reported metric value. A different distribution is not necessarily worse, but if all these sub-volumes are clustered together, then a single large hotspot is formed which is what should be prevented.

3.5.3. Hotspot registration requirements

Since no open-source method of registering hotspots has been found, a new method is developed for this research. Here, we list a number of requirements that we perceive to be important for a hotspot registration method, to be used in BRIGHT, that matches clinical requirements, as discussed in Section 3.4.1.

The first requirement is that the method should be compatible for implementation on the GPU. This is because in BRIGHT, the calculated dose of the DCPs for all the different plans in the population are stored in GPU memory. Having a method which only performs well on CPU would then require copying the dose-values from GPU memory to CPU memory during optimization to assess the HSI. This would greatly impact the overall run-time of the algorithm.

This brings us to the second requirement, the required run-time. The method should be as fast as possible, since time is limited. Especially if the HSI needs to be calculated during optimization then this is of importance if adaptation in the clinic is envisioned since during optimization a large number of treatment plans need to be evaluated. That said, the execution speed is not the main priority in this research, but nonetheless the focus will be on state-of-the-art implementations to already give some insight in the required run-time.

The third and fourth requirements are that the method should be fully automatic, since there will be many evaluations of plans required (over 200.000 evaluations were used in the clinic per patient), and the method should be deterministic to reduce the required number of evaluations per plan.

The fifth and last requirement is that the method should be adaptable to using the DCPs, as the dose distribution is captured using DCPs in BRIGHT and developing another means of registering dose is deemed out of scope for this research.

This set of requirements led to two methods which were further inspected, the Afforrest-algorithm (Sutton et al., 2018) for graphs and the Block-Based Union-Find (BUF) algorithm (Allegretti et al., 2019) for a regular grid of voxels. Both have been shown to be state-of-the-art on detecting connected components (Hong et al., 2020, Lemaitre et al., 2021). The problem of hotspot detection can be reduced to the problem of detecting connected components as will be discussed in the next sections.

3.5.4. BUF algorithm

Connected component detection is finding subsets of either voxels or nodes that are connected. Voxels are points in a regular spaced grid, they can be seen as equally sized cubes. Nodes are connected by sharing an edge (graph) and voxels by sharing a side/corner. In connected component labeling, all nodes/voxels that are connected will receive the same label. Our hotspot detection problem can be reduced to the problem of labeling connected components, since it can be limited to labeling only the nodes/voxels (DCPs) that have received a dose of above a certain lower-bound.

To reduce the hotspot registration to the BUF algorithm, which works on voxels, these voxels need to be created first from the DCPs. The strong suit of the BUF algorithm is its exploitation of regular access pattern of the voxels. However, to reduce the DCPs to these voxels they have to be summarized into the voxels. Next, for every voxel it should be determined if it is a 'hotspot-voxel' or not to enable connected component analysis. This would then result in a loss in accuracy because an aggregate needs to be made of all DCPs that are in the voxel. Another potential issue is the absence of DCPs in a defined voxel. Then no dose is calculated for that voxel. This raises the question what value the voxel should take, since it can then either connect a hotspot or breakup a hotspot. This can happen since the DCPs are sampled in a random way. Furthermore, the resolution of the voxels will need to be the same in every organ/tissue, but in BRIGHT the number of DCPs is equal in every organ/tissue whereas their volumes are not. Given these shortcomings, the BUF algorithm was not chosen for further testing and implementation. In theory BUF could be implemented for hotspot detection, but a solution has to be found for all these shortcomings. Due to these shortcoming it was chosen to implement a graph-based method that does not suffer from these shortcomings.

3.5.5. Afforrest hotspot detection

The Afforrest algorithm is an adaptation of the Shiloach-Vishkin algorithm (Shiloach and Vishkin, 1980). It is a label propagation algorithm, which transfers the lowest node-label through the graph to the nodes it is connected to via edges. However, before diving into the details of the implementation, first the construction of the graph will be discussed.

Creating the graph

As discussed before, the calculation of the DVIs is done using dose calculation points (DCPs). Using these DCPs the graph will be constructed. The DCPs will be the nodes in the graph, the only thing that remains is the creation of edges. The goal is to be able to detect connected components

of DCPs which all have at least a certain received doses. These connected components then are the hotspots. A simplified visualization of this is given in Figure 3.10.

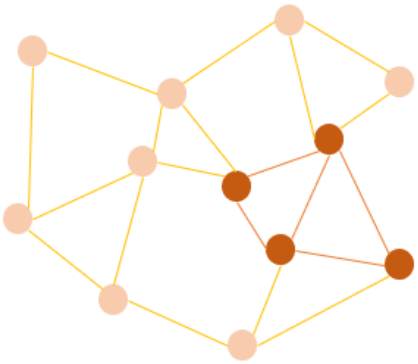

Figure 3.10: Dummy visualization of a graph based hotspot, each dot represents a dose calculation point. A light dot colour indicates a received dose under the hotspot lower bound dose. A dark dot colour indicates a received dose above the hotspot lower bound dose. A collection of connected dark dots represents a hotspot.

In Figure 3.10 the DCPs with the darker shade of orange make up the hotspot. The hotspot is thus detected since they are connected to each other.

To define edges between the DCPs, a maximum edge length is set. If the distance between two DCPs is less or equal to the maximum edge length, then the two DCPs will be connected by an edge. The maximum edge length will dictate when two hotspots will be seen as two separate hotspots or as a single larger hotspot. A simplified visualization of this is shown if Figure 3.11.

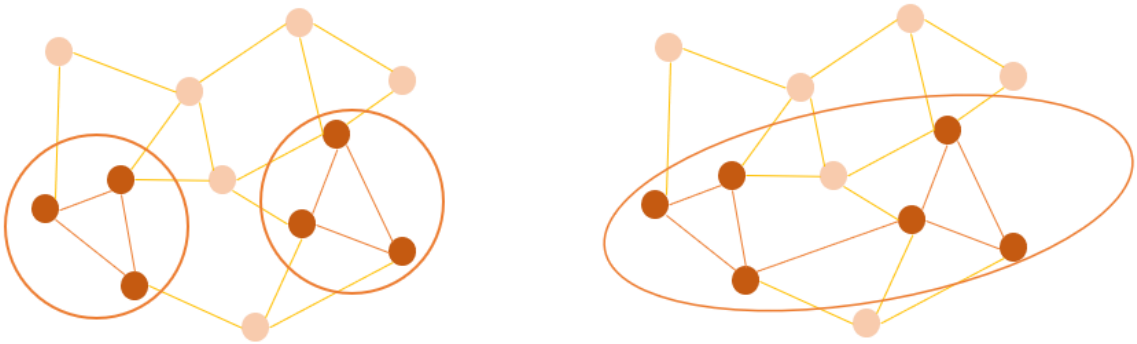

Figure 3.11: Visualisation of two disconnected hotspots on the left and a single connected hotspot on the right. Defined on the same set of dose calculations points, but a different set of edges. The hotspot on the right is a single connected hotspot because all 'dark' nodes are connected via a path without 'non-dark' nodes on the path.

On the left side, the maximum edge length is shorter, thereby disconnecting the two hotspots, whereas in the graph on the right, the maximum edge length is longer which makes the two smaller hotpots connected to each other. This shows that finding the right edge length is important, since too small a length will lead to underestimation of the hotspot size and too long a length will result in over-estimation of hotspot size. In discussion with the clinical experts, the maximum edge length has been set to 0.5 mm as the ground truth. This is equal to the resolution in which Oncentra Brachy visualizes the iso-dose lines on which the clinical experts form their opinions. By fixing this edge length, the only cause for over or under

The creation of the graph only has to be done twice, since the DCPs are also only sampled twice. Once for evaluation during the optimization procedure and once for the reevaluation of the optimization results. The number of dose calculation points used during optimization in the clinic is 20.000 points per organ. During reevaluation, this number is increased to 100.000 points. In Bouter et al., 2019 it

was shown that the reduced accuracy during optimization was acceptable given the reevaluation of the results based on the DVIs. However, the naive implementation of generating the graph using a maximum edge length would require a pair wise distance calculation of all DCPs. Given the number of DCPs this would greatly impact the run-time of BRIGHT. Therefore, a simple speed-up has been implemented. Given that the maximum edge length is known upfront, the total bounding box of all organs will be filled with voxels with a side-length equal to that of the maximum edge length. All DCPs will be sorted into their respective voxel. Now the only distance calculations that have to be made are between DPCs in a single voxel or between adjacent voxels, since the distance between any other pair of DCPs will be larger than the maximum edge length.

Connected component detection procedure

In this section, a very brief overview will be given on how the Afforrest algorithm works, a more detailed description is given in Sutton et al., 2018 and Hong et al., 2020. Note that in the explanation in this thesis the abbreviation DCPs is used instead of nodes, as is the case in the Afforrest literature. The adaptations which make the algorithm find hotspots rather than connected components will be discussed after the explanation of Afforrest.

In order to detect connected components, Afforrest makes use of Union-Find. One step of the procedure is shown in Figure 3.12. All DCPs have a parent pointing label (yellow arrow), i.e. a label which states what the index of its parent DCP is in the array of DCPs. First all DCPs are initialized with their own index as parent pointing label, thus creating a forest of trees consisting of only 1 DCP. Then for every edge (green connector) in the graph, the algorithm looks at its DCPs and for both DCPs it will find the root DCP (DCP that is pointing to itself) of the tree that the DCP is connected to (blue arrows show the procedure of finding the root DCP). This is found by traversing through the graph based on the DCPs that are pointing to one another. Then if both root DCPs are found for both DCPs, they will be compared and the root DCP with the higher index will now not point to itself but to the other root DCP.

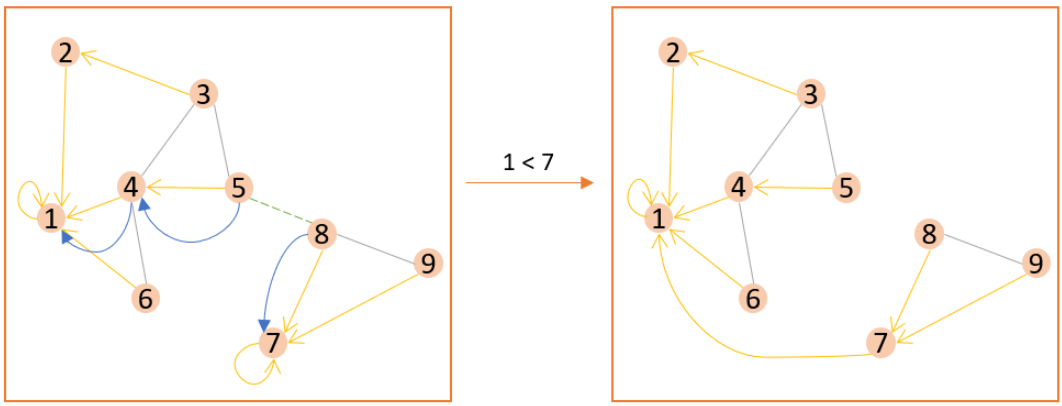

Figure 3.12: Visualisation of a single edge reduction step in the Afforrest connected component procedure. Every circle represents a DCP. Every DCP has its own index (black number) and parent DCP it is pointing to (yellow arrow). In this visualisation the reduction step is shown for the edge between DCPs 5 and 8. For each DCP the root DCP of the tree that it is pointing to is determined by traversing parents, shown with the blue arrows. After the root DCP have been found, their labels/indices are compared and the DCP with the highest index will be pointed to the DCP with the lowest index, in this case DCP 1.

The algorithm performs this procedure of reducing labels in parallel for all the edges of a sub-graph. This sub-graph consists of all DCPs but a subset of the edges. One edge per DCP is chosen for the sub-graph. This procedure is repeated until each edge has been reduced once. This means that the procedure is called upon k-times, where k is the maximum degree of the graph. During each procedure call, the parent pointing labels are reduced to the root label of the tree it is connected to. The convergence is guaranteed by making use of the atomic asynchronous compare_and _swap method from CUDA for writing to the memory of the parent pointing labels in parallel. This method prevents data race problems, i.e. two threads writing to the same memory location leading to unpredictable outcomes. CUDA is the GPU programming interface of Nvidia. By performing the procedure on sub-graphs, the total amount of work is reduced since the average depth of the created trees is reduced in

each step as compared to performing the procedure on the full graph in parallel.

The last part of the algorithm is large component skipping. In the first executions of the procedure on sub-graphs, the largest (or a very large) component should already become apparent as it has the most DCPs pointing towards it. The largest component is estimated by randomly sampling parent pointing labels and finding the most frequent root label. The large component skipping trick exploits the fact that some edges can be skipped, while still reaching convergence. The edges that will be skipped are the edges of DCPs that are already pointing towards the largest component. This results in a reduced amount of edges to be reduced. This trick only works if edges are stored twice as two directed edges, since then there will be edges pointing towards the largest component which will be reduced. The largest component skipping procedure is visualized in Figure 3.13.

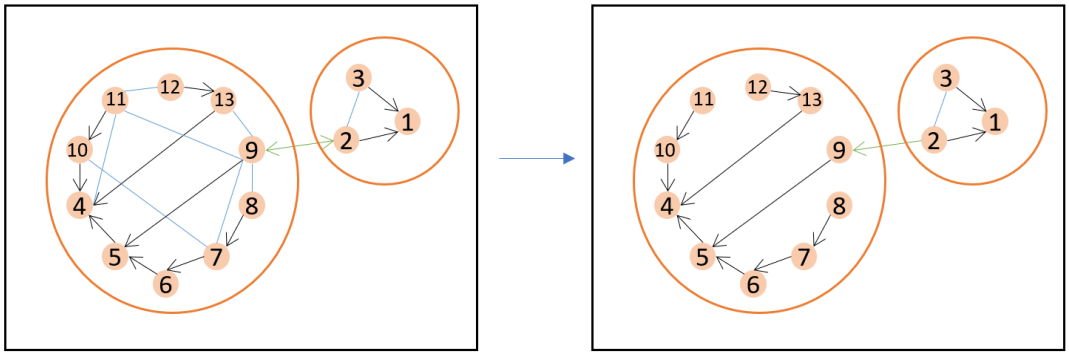

Figure 3.13: Visualisation of the largest component skipping procedure of the Afforrest algorithm. In the images two clusters/hotspots/components are identified, indicated by the two orange circles. In the left image all edges are still active and can be used in later steps. The black arrows indicate edges that have already been reduced, the blue lines are edges that have not yet been reduced. After determining the largest component, in this case the component with root DCP 4 (left component), all outgoing edges are discarded and will not be reduced in later steps. This results in the graph on the right. Only these resulting edges are used for further steps. Since the edge from DCP 2 to DCP 9, shown in green and not yet reduced, will not be skipped, the procedure calls that will follow will eventually reduce the root label of the component on the left to root DCP 1.

In Figure 3.13 the largest component is identified as the left component with root DCP 4. If the algorithm performs the reducing procedure on an edge, but the starting DCP is already pointing to the largest component we can skip it. This can be seen in the edges of node 9. The edges going from DCP 9 to DCP 11, 7, 8, 13 and 2 will be skipped, since DCP 9 is already pointing towards the 'largest' component. Since every edge is saved twice in memory, once for both DCP it connects, it will still guarantee that the lowest root DCP will be found. In Figure 3.13 above, it wrongly identified 2 hotspots (root DCP 4 and 1) but since DCP 9 and 2 are still connected with an edge going from 2 to 9 these two hotspots will eventually be merged as this edge will still be sampled.

Adapting connected component for hotspot detection

A typical result from Afforrest is shown in Figure 3.14. Here all DCP indices have been replaced by the root DCP index.

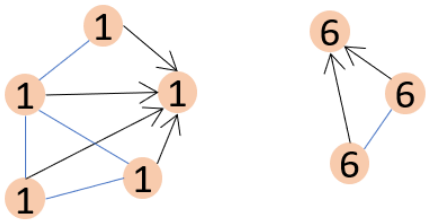

Figure 3.14: Visualisation of the result of Afforrest, which is a directed forest i.e. a set of trees/edges pointing to a single root DCP.

However, in the case of hotspot detection, we only want to connect DCPs which have received dose of above the lower bound given the organ they are in. Therefore the initialization procedure and

the linking/reduction procedure of Afforrest have been adapted. In the initialization procedure where every DCP gets a unique parent pointing label, DCPs are only assigned a unique parent pointing label if the received dose value is high enough, otherwise they get a placeholder value of -1. Then in the linking/reduction procedure, whenever the algorithm encounters a DCP with a parent pointing label of value -1, it is simply ignored. This procedure is visualized in Figure 3.15.

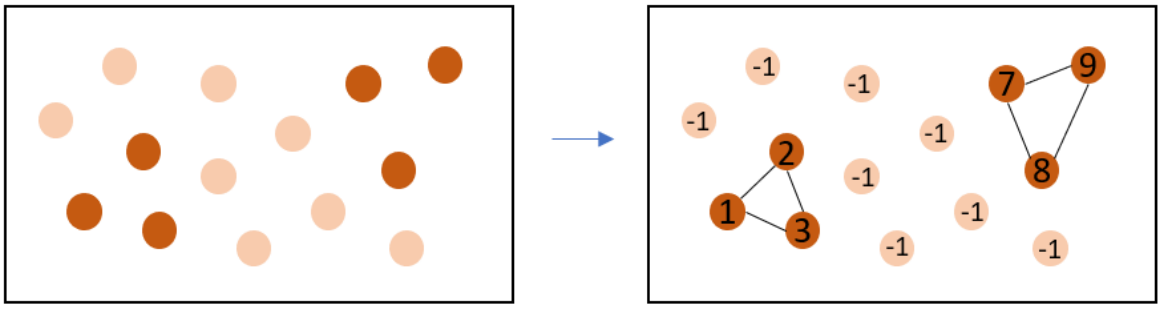

Figure 3.15: Visualisation step of the initialization of hotspot detection. Every dot represents a DCP. Dark dots represent DCPs for which the received radiation is more than the lower bound of hotspot dose, lighter dots received less than the lower bound of hotspot dose.

In section 3.4.1 it was discussed that hotspots are only of importance whenever they exceed a certain distance from the catheters, since it is no problem if a large hotspot is formed within the catheters. To adhere to this requirement we only want to measure hotspots outside of the 1.1 mm radius of the active dwell positions, to prevent registering a long hotspot within the catheters. Therefore, besides giving the all non-hotspot-dose DCPs a initial parent pointing label of -1, all DCPs that fall within a 1.1 mm radius of the active dwell positions are also given the initial parent pointing label of -1.

Now that the algorithm has been equipped to register hotspots, the end result of the procedure should be transformed to the Hotspot Size Index metric, which was the sum of hotspot sizes of hotspots with a size of at least 0.1 cc. In order to calculate this a weighted sum is calculated for all DCPs within the same connected component. The sum is taken over the respective volume that each DCP represents. So if there are 10 DCPs pointing to root DCP 1, and they all have a volume of 0.01 cc, the resulting sum would be 0.1 cc for the hotspot of DCP 1. By then comparing the size of the hotspot to the hotspot lower-bound volume of 0.1 cc we can sum the hotspots and thereby calculate the HSI value of a treatment plan. Furthermore, it is also possible to save all hotspot sizes for later inspection when required.

Psuedocode and computational complexity of hotspot detection methods

The theoretical worst-case time complexity of this procedure is $\mathcal{O}(D^2)$, where D is the set of dose calculation points. This time complexity is determined by the pair-wise distance calculations that need to be made between the DPCs. However, by pre-processing the DCPs into buckets, the number of pair-wise comparisons is reduced. This greatly sped-up the process of graph generation, as not for all pairs of dose calculation points the distance had to be calculated and the cost of pre-processing only costs $\mathcal{O}(D)$.

When there are different maximum edge lengths per organ/volume, then the maximum of those edge lengths is used for bucket creation. When comparing pair-wise distances to the maximum edge length, then the maximum of the maximum edge lengths from the two organs/volumes is used for the comparison.

Algorithm 1 Graph Generation

```
1:  procedure Graph Generation(D, E_max)          // D = set of DCPs
2:                                                // E_max = max edge length
3:
4:      E ← {}                                    // Set of edges, stored in CSR matrix format
5:      B_X,Y,Z ← {}                              // Set of buckets, each dimension divided into
6:                                                // segments equal to the maximum edge length
7:      for all d ∈ D do                          // Fill the buckets with the DCPs
8:          x ← ⌊d_x / E_max⌋
9:          y ← ⌊d_y / E_max⌋
10:         z ← ⌊d_z / E_max⌋
11:         b_x,y,z ← d
12:
13:     for all b_x,y,z ∈ B_X,Y,Z do
14:
15:         for all d_1 ∈ b_x,y,z do
16:
17:             for all d_2 ∈ b_x,y,z do           // Compare distances for DCPs in same bucket
18:                 if d_1 ≠ d_2 then
19:                     if distance(d_1, d_2) ≤ E_max then // If distance smaller than max distance
20:                         E ← e_d1,d2
21:
22:             for all d_2 ∈ N(b_x,y,z) do        // Compare distances for neighbour bucket DCPs
23:                 if distance(d_1, d_2) ≤ E_max then
24:                     E ← e_d1,d2
25:
26:     return E                                  // Returns the created edge list
```

Algorithm 2 Hotspot Detection

```
1:
2:  procedure LabelInitialization(D, R_{lb,o})        // D = Set of DCPs per treatment plan
3:                                                    // R_{lb,o} = hotspot radiation lower bound per organ o
4:      L ← []                                        // Array of DCP labels per solution
5:      for all D_x ∈ D  in parallel do
6:          for all d ∈ D  in parallel do
7:              if R(d) ≥ R_{lb,o} then               // check against radiation lower bound of the organ
8:                                                    // that the DCP is sampled in
9:                  L_x[d_i] ← d_i                    // d_i = index of DCP d
10:             else
11:                 L_x[d_i] ← −1
12:     return L
13:
14: procedure Afforest(L, E, #rounds)                 // L = labels of DCPs per treatment plan
15:                                                   // E = set of edges between DCPs
16:     for i ← 1, #rounds do
17:         for all L_x ∈ L  in parallel do           // L_x = labels for treatment plan x
18:             for all l ∈ L_x : i ≤ ||N(l)||  in parallel do
19:                 union_async(l, N(l)_i, L_x)       // N(l) = neighbours of DCP l
20:             for all l ∈ L_x  in parallel do
21:                 FindAndCompress(l, L_x)
22:     c_x ← most_most_frequent_component(L_x)
23:     for all L_x ∈ L  in parallel do
24:         for all l ∈ L_x : l ≠ c_x  in parallel do
25:             for i ← #rounds, ||N(l)|| in parallel do
26:                 union_async(l, N(v)_i, L_x)
27:             for l ∈ L_x  in parallel do
28:                 FindAndCompress(l, L_x)
29: procedure union_async(u, v, L_x)                  // p_u = parent label of DCP u
30:     p_u ← FindAndCompress(u, L_x), p_v ← FindAndCompress(v, L_x)
31:     while p_u ≠ p_v do
32:         if p_u = L_x[p_u] and CAS(&P[u], p_u, p_v) then
33:             return
34:         p_u ← FindAndCompres(u, L_x), p_v ← FindAndCompress(v, L_x)
35: procedure FindAndCompress(u, L_x)
36:     r ← u
37:     if L_x[r] = r then
38:         return r
39:     while r ≠ L_x[r] do
40:         r ← L_x[r]
41:     while j ← L_x[u] > r do
42:         L_x[u] ← r, u ← j
```

Algorithm 3 HSI

```
1:  procedure HotspotVolumeConversion(L, V)      // L = set of labels for DPC per solution
2:                                               // V = array of approximated volume per DCP
3:      H ← {}                                   // H = set of hotspots per solution
4:      for all L_x ∈ L in parallel do
5:          for all l ∈ L_x in parallel do
6:              if l ≠ −1 then
7:                  AtomicAdd(H_x[l], V_l)       // AtomicAdd to prevent Data Race
8:      return H
9:
10: procedure HSI(H, V_lb)                       // H = set of hotspot per solution
11:                                              // V_lb = volume lower bound on hotspots
12:     HSI ← []
13:     for H_x ∈ H in parallel do
14:         for h ∈ H_x do
15:             if h ≥ V_lb then
16:                 HSI[x] ← HSI[x] + h
```

The run-time complexity of this algorithm is determined by the FindAndCompress method, which is not done in parallel. This method has a run-time complexity, defined for parallel computing, of $\mathcal{O}(|D|)$ *depth* complexity and $\mathcal{O}(|D|^2)$ *work* complexity. What should be noted here is that this complexity is also linear in the maximum degree of the graph, as for every edge of a node the FindAndCompress step has to be performed.

4

Equipping BRIGHT to Improve Homogeneity

This chapter will be dedicated to describing how BRIGHT is adjusted such that it is potentially better equipped to generate more homogeneous treatment plans. These different means of adjustment will all be tested in the experiments.

4.1. Options for Implementation

This chapter will first start off with a more detailed explanation of the different ways in which BRIGHT is able to steer improvement based on the quality of the treatment plans. Then, the different ways of assessing the quality of the treatment plans will be discussed. There are three main routes for steering on treatment plan quality, which all have their advantages and disadvantages. These three different routes are the creation of an extra objective, a new constraint or the augmentation of the current objectives.

4.1.1. Creating an Extra Objective

As discussed in Section 1.3.1 BRIGHT tries to improve the treatment plans by looking at the Least Sparing Index (LSI) and the Least Coverage Index (LCI). It does so by comparing different treatment plans based on these two metrics. Using Pareto-dominance, in every generation the set of non-dominated solutions is determined based on the population and the offspring of that population. For a solution to not be dominated means that there is no other solution that is equal in all objectives and better in at least one. The set of non-dominated solutions is called the Pareto approximation front, as there is no guarantee that the front has reached optimality. The set of solutions forms a front in 2D space as visualized in Figure 4.1.

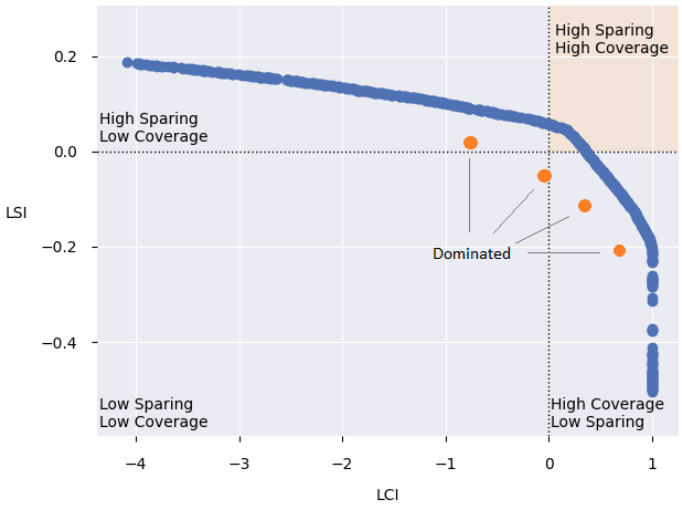

Figure 4.1: Example Pareto Approximation Front. Every dot represents a treatment plan. On the x-axis the LCI-score and on the y-axis the LSI-score of the treatments plans are shown. The orange region in the top right corner is the golden corner where all treatment plans have positive LCI and LSI.

When an extra objective is added, BRIGHT tries to improve that new objective as well as possible, thereby doing exactly what is intended if that objective relates to the HSI. Furthermore it even provides more insights in the possibilities for the patient as it will show the trade-off there is to make when it comes to homogeneity and the other two objectives, if any. Visualizing the trade-off between 3 objectives is less trivial than two objectives without using motion. However, motion is difficult to show in a text file, as can be seen in Figure 4.2, where the viewing angle is adjusted between different plots.

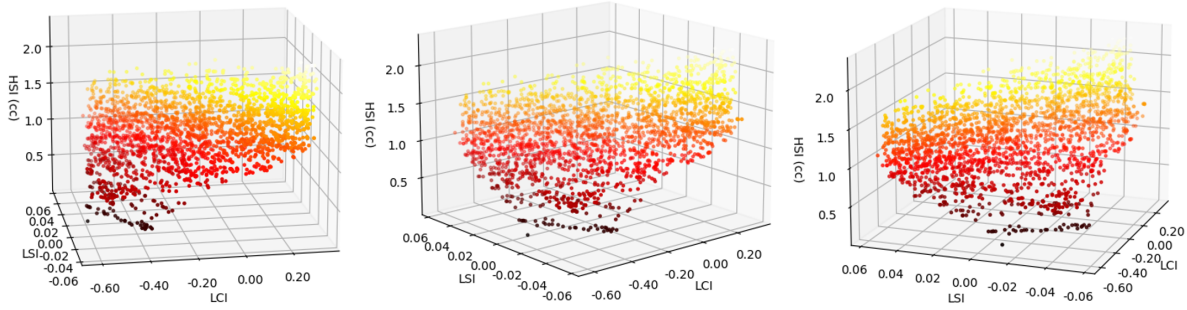

Figure 4.2: The same 3D Pareto approximation front with different viewing angles per plot. The axes show the LCI, LSI and HSI in this example. The colour coding of the treatment plans is linked to their HSI value, which will be further explained in Figure4.3.

Since using motion is not ideal in a text based document such as this one, slicing is used. In Figure 4.3, a Pareto approximation front is shown, which has three axes rather than two. This means the third objective is flattened, which can result in treatment plans placed on top of each other. This happens when two treatment plans have a combination of LSI and LCI values that are just slightly different, but have a different third objective value. In that case, because of the size of the marker used to indicate a treatment plan, one treatment plan will be placed on top of the other. Slicing through the Pareto approximation front using upper bounds on the third objective is then required to get a good view of the achieved third objective values. The different plots in Figure 4.3 have a different upper-bound (2.0, 1.0 and 0.5 cc) for the third objective (HSI).

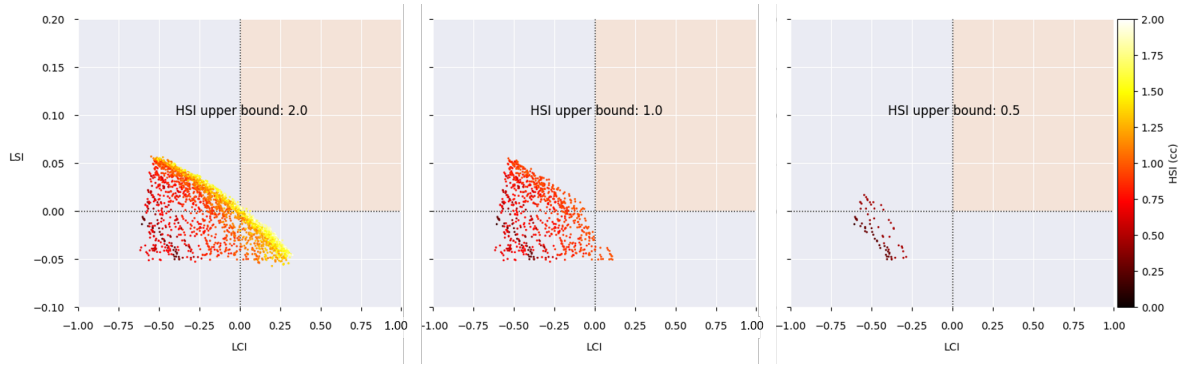

Figure 4.3: Pareto approximation fronts with different upper-bounds for achieved HSI values. On the right side of the graphs a colour bar is shown. The colour of a treatment plan as shown in the Pareto approximation front corresponds to the achieved HSI value of that plan. In every plot only the plans that adhere to the HSI upper bound are shown. The upper bounds are 2, 1 and 0.5 cc.

This extra dimension however comes at a cost. Adding an extra objective increases the computational complexity of the problem. Rather than a curved line for the approximation front, there now is a curved 3D surface. Having this third objective to account for in comparing treatment plans for dominance means there is more space for the solutions to fill without being dominated. This extra dimension results in a decrease in efficiency of population-based optimization methods that make use of domination and aim to cover the entire Pareto front (Deb, Saxena, et al., 2006), which BRIGHT is, since the chance increases that a treatment plan is non-dominated. This in turn reduces the pressure that the optimization has towards the Pareto approximation front, which results in an increase in convergence time. In the worst case the required population size to approximate the Pareto front increases exponentially with an increase in the number of objectives. The formula for determining the population size in that worst-case scenario would be $\mathcal{O}(N^M)$, where N is population size needed for a one-dimensional problem and M is the number of dimensions. In the bi-objective problem formulation the ideal population size was found to be 96 (Bouter et al., 2019). This would mean a population size of $(\sqrt{96})^3 \approx 941$ for the tri-objective problem. This worst-case behaviour is only true when there is an equal trade-off between all objectives. If two or more objectives in a multi-objective problem are non-conflicting and thus show now trade-off, then the dimensionality of the Pareto front reduces. This would then in turn also lead to a reduction in the required population size.

In this research the population size has been set to 288. This number has been found in preliminary experiments to be effective for three objectives. In BRIGHT, a new generation is generated using clusters which are fitted along the approximation front to promote the pressure towards the approximation front. The number of clusters found to be effective in the bi-objective formulation is five (Bouter et al., 2019). Due to the increase dimensionality also the number of clusters needs to increase. In this research the number of clusters has been set to twelve. This number has been found in preliminary experiments to be effective for three objectives.

An important advantage of using a third objective is that no aspiration value for the third objective has to be set upfront for the patient, which will be the case for the next two means of steering on treatment plan quality.

4.1.2. Setting a Constraint

Another option would be to simply add a constraint based on the homogeneity quality of the treatment plans. This constraint would render the treatment plans which do not adhere to the set cut-off value infeasible. In the generational check for Pareto dominance by BRIGHT, also some hard constraints are checked. If one of the solutions is violating a constraint, then automatically the other treatment plans dominate. This results in the infeasible treatment plan being discarded. This method of constraint handling is called 'constraint domination'.

The advantage of using a constraint over adding a third objective is that it will not result in added complexity other than the need to calculate the homogeneity quality/constraint value itself.

The difficulty then, however, is finding a value that satisfies the clinical requirements, but still allows for the algorithm to find good treatment plans for all patients. If the constraint is set too tight, this could result in the optimization not being able to properly explore the solution space. This could result in a

Pareto approximation front that is not able to reach the Golden Corner. If the Golden Corner is not reached but it could have been possible without the constraint, then that would implicitly mean that BRIGHT values satisfying the constraint more than reaching the Golden Corner. This would not be in line with the preferences of the clinical experts.

To see if there could be a constraint value that generalizes well to the whole patient set one could first use the treatment plan quality indicator as a third objective to see for each patient what is possible.

An option for improving the downside of not being able to find the right constraint could be to make a dynamic constraint. This constraint would be based on the already learned parts of the solution space. This however is left open for further research.

4.1.3. Augmenting Current Objectives

The last option for steering upon a homogeneity indicator is augmenting the current objectives. The current objectives, LSI and LCI, are both a combination of DVIs. As discussed in Section 1.3.1, the LSI and LCI are given the value of the worst performing DVI in the set of DVIs they represent. This performance is measured by the achieved value of the DVI and its aspiration value. The set of DVIs of both LSI and LCI can be extended with a new indicator for the homogeneity, with its own aspiration value. The logical option would be to extend the LSI, as this focuses on the sparing of organs and tissues which the homogeneity requirements are also focused on.

The advantage of adding the homogeneity indicator to the current objectives over adding it as its own objective is one of computational complexity. The problem will remain a bi-objective one, but will account for homogeneity. The advantage over adding it as a constraint is that it will not render part of the solution space infeasible, allowing BRIGHT to explore all of the solution space.

A downside of this approach is that the last statement is only partially true. To add the new indicator to the set of LSI, one would be required to set an aspiration value. The problem with this is that it is not known if a good aspiration value is achievable. If the aspiration value is set too ambitious for a patient, then this indicator will dominate the behaviour of the LSI. This means that BRIGHT will effectively disregard the other indicators in the LSI and only focus on the homogeneity indicator, which could lead to BRIGHT not reaching the Golden Corner, as was the case with adding a new constraint. If the aspiration value is set not ambitious enough and is achieved too easily, then the other indicators will dominate the behaviour of the LSI. This would result in unexplored parts of the solution space, as BRIGHT will disregard further exploring the minimization of the new homogeneity indicator, and will be unable to find plans with sufficiently small hotspots.

4.2. Options to measure homogeneity

When trying to improve the homogeneity of the treatment plans, the dwell times have to be adjusted. To do this, BRIGHT must be adapted to promote the right adjustments. Currently the metrics that BRIGHT can steer on are the LCI and the LSI which are a proxy for the DVIs. In formulating metrics to steer upon for improving homogeneity, one can either look directly at the distribution of dwell times or look at the resulting dose distribution. The other choice to make is to either directly measure homogeneity or indirectly.

The advantage of looking at the dwell times is that it is less complex from a computational point of view, as there are fewer dwell positions (200-400) than dose calculation points (>100.000). The advantage of looking indirectly at homogeneity is that the method can be created in a way that is computationally cheaper, as less detail is required. However, these indirect methods might be correlated strongly enough with the formation of hotspots so that they will do the job, without the expensive operation of registering hotspots. Therefore different strategies will be tested for both dwell time-oriented as well as dose distribution-oriented improvements, and direct registration as well as indirect. In the paragraphs below the different indicators which will be tested in this research will be discussed.

4.2.1. Dwell time oriented indicators

A treatment plan consists of a set of dwell times for a set of dwell positions. Different dwell times will result in different dose distributions. By controlling the distribution of dwell times one could in theory control the formation of hotspots. The question is how the dwell times should be controlled in order to reduce the formation of hotspots.

As discussed in Section 3.2 the main issue of hotspots is the connectedness of high-dose regions.

A single small high-dose region is not an issue, but a large high-dose region consisting of one long dwell time or multiple medium dwell times is the problem. Potentially, a combination of the dwell time as well as the distance between longer dwell times should be taken into account when looking for homogeneous treatment plans.

Dwell time modulation

In literature and practice, as discussed in Section 3.3, the main means of improving treatment plan homogeneity is the use of Dwell Time Modulation (DTM) which has been formulated in several ways. The general idea behind the DTM is to reduce/limit the amount of deviation that adjacent dwell positions have in relation to each other. The goal of the DTM is to result in a smooth distribution of dose over the target volumes. This can thus be seen as an indirect way of mitigating hotspots.

The DTM formulation that will be tested in this research is shown in Equation 4.1, which is the absolute DTMR from Balvert et al., 2014.

$$DTM = \sum_{i \in D} \sum_{j \in C_i} \max(0, |T_i - T_j| - \theta) \tag{4.1}$$

D is the set of dwell positions, C_i is the set of the k-nearest neighbours of dwell position *i*, T_i is the dwell time of dwell position *i* and θ is the control parameter for the allowed deviation.

The conclusion in the paper by Balvert et al., 2014, was that adding the DTMR did result in improved robustness but did not yield a reduction in hotspot sizes without simultaneously deteriorating the $D_{90\%}^{Prostate}$. However, in their paper they only had three patients and the model used was a single-objective one. By using a multi-objective optimization it might be possible to see if there is a correlation between hotspots and DTM value.

The computational complexity of calculating the DTM value is linear in the number of dwell positions to calculate deviation for, i.e. $\mathcal{O}(D)$. The location of the dwell positions will not change during the optimization and reevaluation. Thus, by calculating the k-nearest neighbours of each dwell positions once and storing it in memory, it can be reused. Then during optimization and reevaluation only k computations have to be made for each dwell position ($\mathcal{O}(kD)$). Given that the number of dwell positions is relatively low (200-400) this is not an expensive operation. If for instance k would be set to 6 (aiming for two neighbours in each direction x, y and z), this would result in approximately $400 \cdot 6 = 2400$ calculations per treatment plan.

The downside of DTM is that it only focuses on the deviation between dwell times, but not on the length of the dwell times themselves. In theory, a treatment plan with a uniform distribution of very long dwell times would receive a perfect score upon the DTM. However, this would result in a very large hotspot which covers all organs. But given that BRIGHT is not only steering upon one metric, it might result in more homogeneous treatment plans. In combination with the V200, which indirectly reduces the dwell time lengths, the DTM might be able to prevent hotspots.

DLDM

In the paper by Morén et al., 2018b, which focused on reducing hotspots, a novel approach to reduce hotspots was presented. The results showed that with their approach a potential reduction in hotspots was achieved. However, their means of measuring hotspots was indirect, by focusing on multiple sub-volumes and measuring the V_{200} in those volumes.

This used metric/objective focused on the dose distribution rather than the dwell time distribution, but it was an inspiration for a reformulation to dwell times which tries to directly minimize hotspots. Using their formulation both the minimization of maximum dwell times as well as maximizing the distance between longer dwell times can be established. In their paper they did not name the metric, so the liberty has been taken to name it Dwell Length and Distance Modulation (DLDM). The formulation of the metric is shown in Equation 4.2:

$$DLDM = \sum_{i,j \in D : i \neq j \wedge d(i,j) \leq l} \frac{g(T_i) \times g(T_j)}{d(i,j)} + \sum_{i \in D} g(T_i), \tag{4.2}$$

where *D* is the set of dwell positions, T_i is the dwell time for dwell position *i*, *d(i,j)* the Euclidean distance between dwell positions i and j, *l* the maximum distance between dwell positions and *g* is the function

described in Equation 4.3 where *B* is a lower bound on the dwell time.

$$g(D_i) = \max\{0, D_i - B\} \tag{4.3}$$

The value for *l*, the maximum length between dwell times for which the DLDM will be calculated, has been set to 10 mm. If no upper-bound value would be set, then BRIGHT would be pressured to push out the longer dwell times as far out from each other as possible. This could then result in more doses in the edges of the prostate, closer to the normal tissue. This is not necessarily the behaviour that is sought after. Therefore, by setting a maximum distance, only high dwell positions that are too close to each other are penalized. The value 10 mm has been chosen because the implant which guides the catheters will put adjacent catheters at least 5 mm apart from each other (Elekta, 2021). Based on the hotspots seen in practice only hotspots are formed between 2 adjacent dwell positions in transversal direction, therefore $2 \cdot 5 = 10$ mm has been chosen.

The value for *B*, the lower bound on the dwell time, is dependent on the source strength that is being used. Based on that, the time is calculated which will generate a x% iso-dose line that has a radius of 2.5 mm, given that the spacing between adjacent dwell positions is either 2 mm or 5 mm. In setting a lower bound on the dwell time for the calculation of DLDM, more focus is applied on the higher dwell times which cause hotspots. If all dwell times were taken into account then a small reduction in a low dwell time would also result in an improvement of the DLDM. This behaviour however is not what this metric is intended for. The value used in this research, which showed the best results in the experiments was 100%.

The computational complexity of this metric shares some similarities to that of the DTM. The distance matrix can be computed once and stored in memory. During optimization, the complexity of the evaluation of one treatment plan then is $\mathcal{O}(|D|^2)$. The number of calculations would then be, for a patient with 400 dwell positions, $\frac{n\cdot(n-1)}{2} = \frac{400\cdot399}{2} = 79.800$ calculations. This falls in range with the computation of the DCPs, therefore using dwell position calculation will in general not worsen the time complexity. Some optimizations can however be put in practice, such as sorting the distance to other dwell positions and stopping the DLDM calculation for that dwell position once the dwell positions are further away than 10 mm. This significantly reduces the number of calculations to be made.

4.2.2. Dose distribution oriented indicators

In the dose distribution oriented indicators there are multiple strategies to follow. As discussed in Section 3.4.2 about the metrics to measure homogeneity from literature, there are multiple ways to define homogeneity, from aggregate measures to the more exact measure described in Section 3.5.5.

HSI

The first dose distribution based metric is the HSI index as described in Sections 3.4.3 and 3.5.5. This metric is exactly what needs to be minimized to prevent hotspots from forming. The HSI is also the metric which will be used for evaluating the results of all other possible improvement of BRIGHT. However, there is a downside to this method, which is the computational and space complexity of the procedure.

To calculate the HSI in an exact way, which would be able to separate hotspots on a distance of 0.5 mm, the required number of dose calculation points is large. When assuming that a DCP represents a perfect cube in volume, then the required volume per DCP would be $(0.5)^3 = 0.125$ mm3. The volumes of organs in the patient data set used in this research are shown in Table 4.1. Next to that the volume per DCP if 20.000 points were used during optimization is shown. In the last column the required number of DCPs is shown to reach a volume of 0.125 mm^3 per DCP.

Volume (cc)	**Average**	**Minimum**	**Maximum**	**DCP(mm^3)**	**Required #DCPs**
Prostate	45.62	32.13	64.67	2.28	364,993
Normal-tissue Prostate	78.20	46.99	108.33	3.91	625,639
Vesicles	4.57	1.27	7.68	0.23	36,525
Normal-tissue Vesicles	28.38	14.94	39.80	1.42	227,051

Table 4.1: Overview of organ (mean, min and max) volumes in the used patient data set and the volume a single DCP approximates (V/DCP) given that organ volume in combination with the used number of DCPs. The required #DCPs, shows the number of DCPs that would be needed to allow for a separability of 0.5 mm between hotspots.

This shows that the required number of DCPs is even larger than what was used during reevaluation, which was 100,000 per organ. Only the seminal vesicles are small enough in volume. If the other organs were given 20,000 DCPs, then the total number of points required during optimization would come at a total of 2.5 million. Given that the received dose by a DCP is stored as a float of 4 bytes, then for a single treatment plan one would already need 10 MB. During the optimization procedure BRIGHT stores more than 1000 treatment plans on GPU memory, which results in at least 10 GB of required GPU memory. This is simply too much required space as this space is also required for other data structures and computations. Therefore using HSI in the exact way with 0.5 mm separability is infeasible even without looking at the computational complexity. For reference, the machines on which this research has been conducted had 12 GB of VRAM, i.e. GPU memory.

Nevertheless, the HSI could also be used in an approximate way by increasing the maximum edge length and keeping the 20.000 DCPs. This reduces the space requirement but comes at the potential cost of accuracy. The consequence of this could be overestimation of hotspot size when two hotspots are close to each other which should be separated based on the 0.5 mm requirement. Another consequence could be less accurate hotspot size measurement. There are simply less DCPs, meaning that the accuracy of hotspot volume also decreases. In the study done by Bouter et al., 2019 a similar problem arose with the number of DCPs for DVI calculations. They eventually came to the conclusion that the reduced accuracy of 20,000 DCPs compared to the 100,000 DCPs for the calculation of the DVIs was acceptable. A similar question has to be answered for the HSI. With the reduced number of DCPs it was chosen to adjust the maximum edge length for generating the graph. The edge length will now instead of 0.5 mm be determined based on the volume of the organ that the DCPs are in. So for every organ, a different maximum edge length will be determined based on the volume of a single DCP represents. This is simply done by taking the cube-root of the volume per DCP, again assuming a DCP represents a cube volume. The motivation for defining the maximum edge length in this way is that if the assumption of a perfect uniformly distributed set of DCPs is true, then the generated graph is connected as a lattice. Now for constructing the graph, if two DCPs from two different organs are a potential link, then the largest maximum length is taken to check if the connection should be there. If the assumption of cubic volume does not hold, which is highly likely, then the edge length should be dependent on direction to create a perfect lattice. However, hotspots are ignorant of direction when it comes to connectivity and therefore a single edge length for every direction should be chosen.

The computational complexity of the HSI can be found in Section 3.5.5. The computational complexity is worse than that of the previous metrics, which focused on the dwell times rather than the DCPs. This is a downside of this approach.

Extra V indices

The second dose distribution based metric is a simpler one in terms of computational complexity as compared to the HSI but still looks at the dose distribution itself. The metric is the sum of extra V indices as described in Equation 4.4. This is an indirect way of trying to reduce hotspots.

$$V_{indices}(cc) = V_{300}^{Prostate} + V_{300}^{SeminalVesicles} + V_{200}^{Normal-TissueProstate} + V_{200}^{Normal-TissueSeminalVesicles} \quad (4.4)$$

The idea behind the metric is that it might not be necessary to exactly know where the hotspots are, when they can be reduced them with more general means. The shortcoming of the $V_{200}^{Prostate}$, as discussed in Section 3.2, was that it did not discriminate between multiple small hotspots below the volume lower bound and a single (or a small number) of large ones. However, given the union-like geometry of radiation given off by active dwell positions (see Figure 3.6), the larger hotspots are more likely to have 'hotter' insides than smaller hotspots. By not focusing on the V_{200} but on the V_{300} it might be possible to reduce most of the hotspots that are formed.

The non-linear dose-rate fall-off means that there is a non-linear relationship between distance from the dwell position and the received dose. The lower the dosage the less strong the fall-off is over distance. Using the TG-43 dose rate model (Rivard et al., 2004), one can calculate the difference in 300% volume between a single and multiple points of radiation achieving the same 100% iso-dose volume. This is shown in Figure 4.4. There it can be seen that two dwell positions result in a 12% decrease of V_{300} iso-dose volume as compared to a single dwell position for the same 100% iso-dose volume. This effect combined with steering on V_{300} could result in more shorter dwell times as this is advantageous for the V_{300} and potentially also for the formation of hotspots.

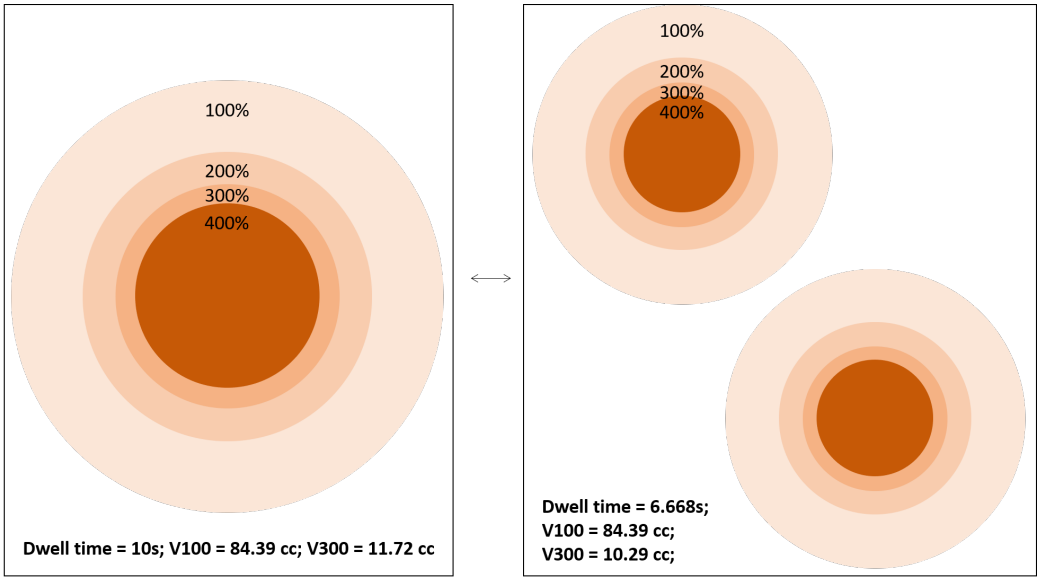

Figure 4.4: Visualisation of iso-dose lines of a single vs. double active dwell position(s) with equal 100% iso-dose volume. Every colour represents a different iso-dose, i.e. $400\%, 300\%, 200\%$ and 100%. The resulting iso-dose volumes for 100% and 300% are shown in the text below, as well as the used dwell time in seconds.

With these findings the following research question can be answered:

Research Question: BRIGHT adjustments

How can BRIGHT be enhanced to potentially produce more homogeneous treatment plans?

BRIGHT can be enhanced to produce more homogeneous treatment plans by adding an extra objective, a new constraint or by augmenting the current objectives. These additions need a metric to be defined. These metrics can be either directly or indirectly related with hotspots. The advantage of adding an extra objective over a new constraint or augmenting current objectives is that no prior aspiration value has to be set. The potential advantage of using a metric that is indirectly related to hotspots is a reduced computational complexity as compared directly related metrics.

4.3. Adaptive steering

The last adjustment made to BRIGHT in this research is not directly related to homogeneity but more to intelligently reducing the search space during the optimization. The motivation for this is two-fold. Firstly, the number of plans of interest can be increased. Secondly, the computational time to reach convergence can be decreased. These potential advantages are especially important when an additional objective is introduced, because the increased size of the objective space makes it more difficult to reach deep into the golden corner.

In the clinical results it was shown that the plans selected from the Pareto approximation front for further inspection and adjustment in OncentraBrachy were all very close to the Golden Corner. This can be seen in the figures in Section B.1 in the appendix. However, during optimization, BRIGHT also puts effort (allocating individuals) in exploring better plans in the tails of the Pareto approximation front. One could conclude that this is all done in vain, since these plans are not of interest to the clinical experts.

If during optimization, when learning what is achievable for a patient in terms of LSI and LCI, a method is applied to shrink the search space, then the computational power is allocated more efficiently towards the main goal of the optimization. Given that BRIGHT is a population-based optimization

method this is especially important, since now more of the individuals contribute to the goal of pushing the approximation front towards the golden corner and more individuals will be created in and around the golden corner for the final result.

The method used to restrict the search space based upon what is learned during optimization is adaptive steering, which was inspired by Alderliesten et al., 2015. Adaptive steering is steering the search effort of BRIGHT throughout the optimization phase based upon the achieved results. This is done by restricting the search space, i.e. tightening the bounds of the search space throughout the generations. A simplified example of the process is shown in Figures 4.6.

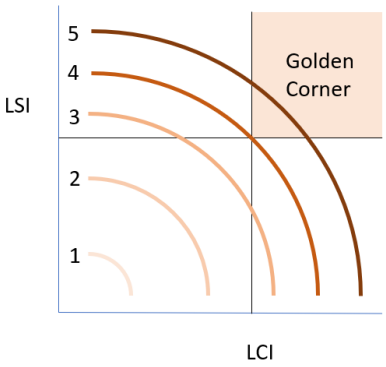

Figure 4.5: Generational progress of Pareto approximation fronts without adaptive steering. 5 generations of Pareto approximation fronts are shown, starting from 1 ending at 5.

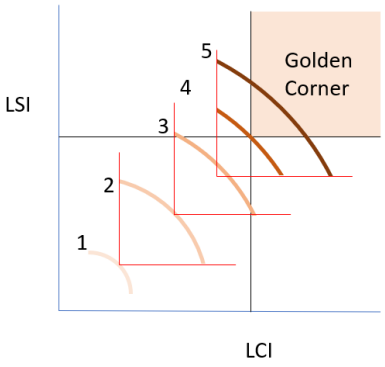

Figure 4.6: Generational progress of Pareto approximation fronts with adaptive steering. The short straight red lines are the enforced constraints per generation, starting at generation 2.

In each generation an assessment is made on the spread of the generation in both LCI and LSI. Based on this spread a new constraint is determined which cuts off the bottom x% on both the LSI and LCI. This constraint will render any treatment plan that has a worse LSI or LCI infeasible, thereby pushing BRIGHT in the direction of the Golden Corner. This procedure is only started if it can be reasonably assumed that well-enough spread has been established in the population. In the start of the optimization phase random plans will be generated. If the adaptive steering would be active right from the start then this might lead to a too restrictive search space, which will hinder the exploration of BRIGHT, and might lead to premature convergence. This problem is visualized in Figure 4.7, where if the adaptive steering would be active the trajectory towards the Golden Corner would be hindered.

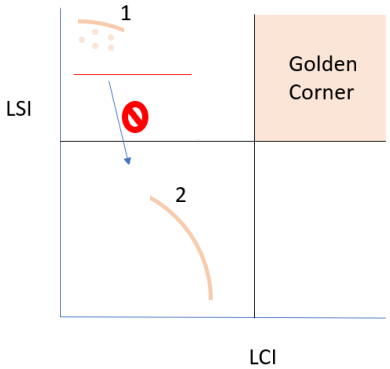

Figure 4.7: Visualization of the potential initialization problem with adaptive steering. The curved lines show the Pareto approximation fronts of 2 generations. The short straight red lines show the enforced constraints. The optimization is not able to reach generation 2, due to the enforced constraint after generation 1.

If a new constraint is enforced upon the population, then the elitist archive will also need to be evaluated to see if any of the individuals in it are now rendered infeasible and have to be discarded. When a new constraint is very 'aggressive' and renders a large portion of the elitist archive infeasible then it might not be wise to apply a new round adaptive steering on the next generation, as this might

lead to a reduced performance of the optimization procedure due to the inability to learn. Therefore the same check as for starting the adaptive steering is also applied in each generation.

If the new constraint, which has been calculated upon the spread in either the LCI or the LSI, does not improve more than 1% on the previous constraint for that objective, then it is not applied to prevent the need to reevaluate the elitist archive over and over, without meaningful improvement.

Lastly, the adaptive steering is stopped for either objective if a certain bound has been achieved. This allows the clinical experts to still have enough choice to assess the trade-off for a patient. In this research, the adaptive steering is stopped when the LCI has achieved a value of -0.5 and the LSI a value of -0.05. This is based upon the chosen plans in the clinic.

When optimizing the treatment plans in the current set-up used in the clinic, there is no requirement to speed up the process as it can reach convergence within 180 seconds. However, if computational complexity is increased by one of the proposed adjustments then adaptive steering might be necessary to achieve reasonable run-time results. This is especially true for adding a new objective, since this will increase the objective space.

The pseudo-code of the adaptive steering procedure is shown in Algorithm 4.

Algorithm 4 Adaptive Steering

procedure AdaptiveSteering($LCI_{lb-max}, LCI_{lb}, LCI_{min}, LCI_{max}, LSI_{lb-max}, LSI_{lb}, LSI_{min}, LSI_{max}, S, a_s, a_p, ts$)
// $LCI_{lb-max}, LSI_{lb-max}$ = max lower bound for LCI and LSI
// LCI_{lb}, LSI_{lb} = current lower bound on LCI and LSI
// LCI_{min}, LSI_{min} = current worst LCI and LSI values
// LCI_{max}, LSI_{max} = current best LCI and LSI values
// S = elitist archive
// a_s, a_p = start and strength of adaptive steering
// ts = elitist archive target size
    $LCI_{new-lb} \leftarrow LCI_{lb}$
    $LSI_{new-lb} \leftarrow LSI_{lb}$
    $u \leftarrow 0$ // Boolean to track changes
    if $|S| \geq a_s \cdot ts$ **then** // Check if archive is full enough
        if $LCI_{new-lb} \neq LCI_{lb-max}$ **then** // Check if bound is already at max
            $adj \leftarrow (LCI_{max} - LCI_{min}) \cdot a_p$ // Calculate adjustment
            if $adj \geq LCI_{max} \cdot 0.01$ **then** // Only apply adjustment if significant change
                if $LCI_{new-lb} + adj \leq LCI_{lb-max}$ **then**
                    $LCI_{new-lb} \leftarrow LCI_{new-lb} + adj$ // Save new bound if not max bound
                    $u \leftarrow 1$ // Record change
                else
                    $LCI_{new-lb} \leftarrow LCI_{lb-max}$ // Save new bound if max bound
                    $u \leftarrow 1$
        if $LSI_{new-lb} \neq LSI_{lb-max}$ **then** // Repetition for LSI
            $adj \leftarrow (LSI_{max} - LSI_{min}) \cdot a_p$
            if $adj \geq LSI_{max} \cdot 0.01$ **then**
                if $LSI_{new-lb} + adj \leq LSI_{lb-max}$ **then**
                    $LSI_{new-lb} \leftarrow LSI_{new-lb} + adj$
                    $u \leftarrow 1$
                else
                    $LSI_{new-lb} \leftarrow LSI_{lb-max}$
                    $u \leftarrow 1$
    if $u = 1$ **then** // Only iterate over solutions if bounds changed
        for $s \in S$ **do**
            if $s_{LCI} < LCI_{new-lb} \vee s_{LSI} < LSI_{new-lb}$ **then**
                $S \leftarrow S - s$ // Remove all infeasible solutions given bounds
    return $LCI_{new-lb}, LSI_{new-lb}$ // Return new bounds for plan evaluation

The computational complexity of this method is determined by the iteration over all solutions in the elitist archive and is therefore $\mathcal{O}(|S|)$. To prevent too much overhead of this method, the bounds are

only adjusted if they make meaningful changes. If changes are less than 1% of the original bounds, then changes are not applied and the iteration over the elitist archive is not performed.

5

Preliminary Experiments

In this chapter, the experiments for testing the effectiveness of adaptive steering and for the difference between fixed and variable edge length for HSI will be discussed. This set of experiments is discussed in isolation of the main experiments of this research as it is a precursor to the main experiments. The findings of these experiments will be used in the default set-up of the experiments that will follow in the next chapters.

5.1. Adaptive Steering

5.1.1. Experiment set-up

To test the effectiveness of adaptive steering for both the bi-objective problem as well as a tri-objective problem, several different combinations of input parameters will be tested. The input parameters that the adaptive steering procedure has are the minimum percentage of the elitist archive to be filled and the percentage of the LCI/LSI range in the population to cut-off with the new constraint. For both parameters the values 10%, 25%, 50% and 75% are evaluated. The Cartesian product of both parameter sets are used to find the best combination and to evaluate their impact.

For the experiments with two objectives the normal brachytherapy problem formulation with the LSI and LCI is used. The number of dose-calculation points is 20.000 per organ which results in 100.000 points, the population size is 96, the number of clustering components is 5, the elitist archive size target has been set to 1000 and the time-limit for optimization is set to 180 seconds, according to the found convergence time in Bouter et al., 2019.

For the experiments with three objectives, the third objective is the sum of the extra V indices as described in Section 4.2.2. The motivation to pick this as the third objective rather than the HSI is the required time as the convergence time when using 3 objectives with the sum of extra V indices is shorter than that with HSI. The second reason is the fact that the sum of extra V indices will result in a 3D surface that is spread out over the search space because of its continuous nature. With HSI, since it has a lower bound in the sum, it is not necessarily a broader 3D surface of solutions but can tend to be more shaped like a thicker line when visualized in 2D with the third objective projected onto the plane spanned by the first two objectives. The number of dose-calculation points used per organ is 20.000 points, which results in 140.000 points because of the added normal tissues. The number of clustering components is 12, the population size is 288, the elitist archive size target has been set to 1000 and the time-limit for optimization has been set to 600 seconds, which has been found to be the convergence time for this problem based on visual inspection.

The experiments are run on eleven out of the twelve patients. Due to technical difficulties in data exports patient 8 could not be used in this research. Every experiment is repeated five times to account for the randomness in BRIGHT. The results are aggregated and averaged. The calculated performance metrics are the time it takes to reach the golden corner (if reached), the starting time of adaptive steering, the stopping time of adaptive steering (when it has reached the -0.5 LCI and -0.05 LSI boundaries) and the size of the final approximation set. To make a comparison without adaptive steering, both problems have also been run without adaptive steering with the same settings for five times.

The experiments have been run on a server with 20 Intel Xeon Processor E52630 v4 @ 2.20 GHz with multi-threading, 128 GB RAM and a NVIDIA GeForce GTX TITAN X with 12 GB of GDDR5 VRAM.

5.1.2. Results

In this section only a selection of the results will be shown, the full results are shown in the Appendices B.2 and B.3.

Bi-objective problem

In Figure 5.1, the resulting Pareto approximation fronts are shown for a starting requirement of 75% of the elitist archive and an adaptive strength of 10% of the range for LCI/LSI in the population. This combination has shown the best performance in terms of how far it reached into the golden corner consistently. The blue dots are the plans generated using adaptive steering and the grey dots are the plans generated without using adaptive steering. The plots have been zoomed in to the golden corner to allow for better visual inspection. The grey Pareto approximation front reaches further than the limits of the plots.

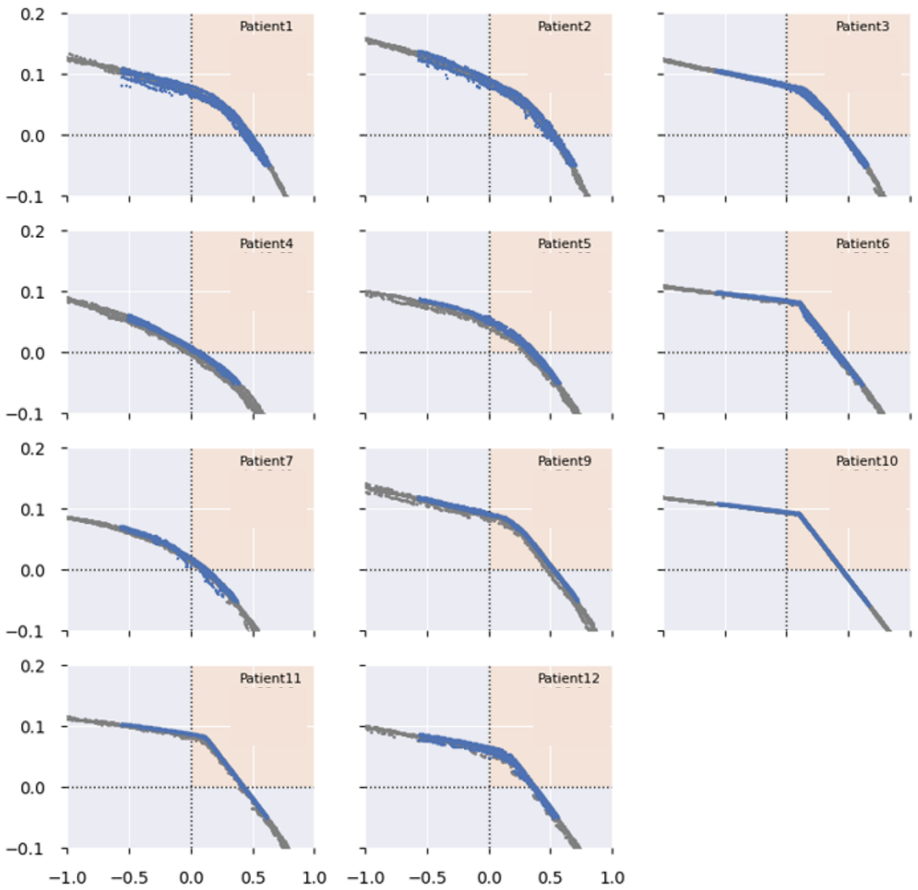

Figure 5.1: Resulting Pareto approximation fronts for the bi-objective problem (LCI and LSI), with adaptive steering strength parameter set to 0.10 and adaptive start parameter set to 0.75. The blue dots show the plans achieved with adaptive steering, the grey dots the plans without adaptive steering.

What can be seen in the figures above is that in the 2D setting, there is no advantage in using adaptive steering when it comes to how far the optimization reaches into the golden corner. For almost all patients the optimization without adaptive steering reach as far as with adaptive steering. Only in patients 4, 5 and 9 it shows a little improvement, but in patients 1, 2 and 7 there is a slight deterioration. These differences are however marginal and could be caused by randomness.

In Table 5.1, the results are shown for a selection of the performance metrics. The 'Time Difference to GC' states the relative time difference between optimizing with and without adaptive steering for

reaching the golden corner. The average column states the average result for all patients, whereas the 'Easiest (14s)' states the results for the patient which was reached the quickest without adaptive steering and 'Hardest (96s)' the slowest. The 'GC size Difference' states the difference in the number of individuals in the Pareto approximation front that are within the Golden Corner.

		Time Difference to GC (less = better)			GC size Difference (more = better)		
start%	strength%	Average (33s)	Easiest (14s)	Hardest(96s)	Average (167)	Smallest (117)	Largest (256)
0.1	0.1	12%	42%	9%	553%	716%	378%
0.1	0.25	10%	35%	-52%	557%	690%	352%
0.1	0.5	14%	31%	-24%	555%	733%	337%
0.1	0.75	17%	40%	-32%	564%	747%	346%
0.25	0.1	4%	29%	-30%	535%	617%	395%
0.25	0.25	3%	20%	-21%	558%	803%	389%
0.25	0.5	3%	19%	-24%	555%	790%	381%
0.25	0.75	7%	27%	-19%	549%	695%	384%
0.5	0.1	8%	24%	-12%	542%	592%	390%
0.5	0.25	9%	24%	15%	556%	741%	374%
0.5	0.5	4%	35%	-36%	552%	730%	386%
0.5	0.75	1%	23%	-28%	558%	702%	377%
0.75	0.1	6%	19%	-2%	559%	805%	391%
0.75	0.25	4%	31%	-14%	543%	712%	388%
0.75	0.5	5%	29%	-20%	556%	644%	384%
0.75	0.75	0%	23%	-49%	555%	753%	389%

Table 5.1: Overview of difference (%) in required time before golden corner is reached and the size (number of plans) of the resulting Pareto approximation front in the golden corner between BRIGHT with and without adaptive steering. Results are shown for the bi-objective problem (LCI and LSI) for different settings for adaptive steering. The column 'average' shows the average results for all patients. The column 'Easiest' shows the results for the 'easiest' patient, the one that reached the golden corner the quickest (14 seconds). The column 'Hardest' shows the results for the most difficult patient. The same column division is made for the golden corner size, where the size is reported in the column names.

What can be seen in the table is that almost all settings of adaptive steering will results in an increase in time to reaches the golden corner on a average for all patients. This can be ascribed to the overhead which is created by the reevaluation of the elitist archive since the time increases when the strength decreases. For patients that easily reach the golden corner the overhead has a higher relative impact. For patients that take longer to reach the golden corner there is an advantage in time visible.

When looking at the number of solutions found in the golden corner a very clear trend can be seen, namely that adaptive steering greatly increases the number of plans found in the golden corner. This however is not unexpected as the region outside the golden corner is rendered infeasible.

Tri-objective problem

In Figure 5.2, the resulting Pareto approximation fronts are shown for a starting requirement of 75% of the elitist archive and an adaptive strength of 10% of the range for LCI/LSI in the population. This combination has shown to be the best performing in terms of how far it reached into the golden corner consistently. In the plots, the blue dots are the plans generated using tri-objective optimization with adaptive steering, the purple dots the plans generated using tri-objective optimization without adaptive steering and the grey dots the plans generated using bi-objective optimization without adaptive steering.

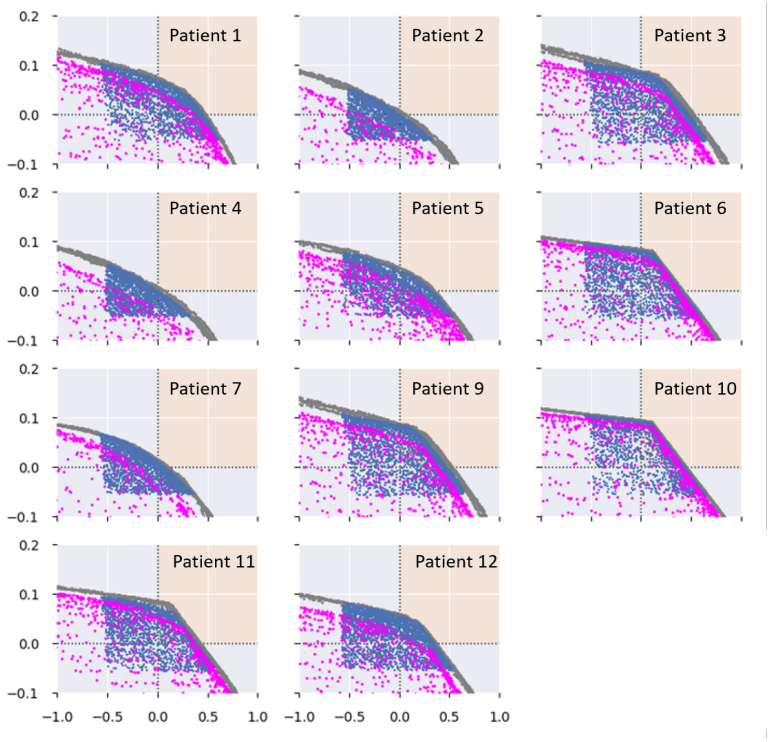

Figure 5.2: Resulting Pareto approximation fronts for the tri-objective problem (LCI, LSI and sum of extra V indices), with adaptive steering strength parameter set to 0.10 and adaptive start parameter set to 0.75. The blue dots show the plans achieved with adaptive steering, the purple dots the plans without adaptive steering and the grey dots the plans without adaptive steering for the bi-objective problem.

What can be seen in the plots is that using adaptive steering will result in reaching further into the golden corner than optimizing without adaptive steering. What can also be seen is that with adaptive steering almost the same LSI and LCI values are achieved as what is done in the bi-objective optimization.

In Table 5.2, the results are shown for a selection of the performance metrics.

		Time Difference to GC (less = better)			GC size Difference (more is better)		
start%	strength%	Average	Easiest(76s)	Hardest(308s)	Average	Smallest (62)	Largest (147)
0.10	0.10	-19%	-12%	-56%	1742%	1872%	730%
0.10	0.25	-15%	-3%	-51%	1570%	1613%	693%
0.10	0.50	-24%	13%	-84%	1392%	1310%	690%
0.10	0.75	-30%	1%	-70%	1479%	1104%	706%
0.25	0.10	-15%	-4%	-61%	1818%	1599%	769%
0.25	0.25	-18%	-1%	-47%	1631%	1728%	736%
0.25	0.50	-26%	-1%	-82%	1582%	1557%	650%
0.25	0.75	-23%	-1%	-55%	1344%	1735%	756%
0.50	0.10	-23%	0%	-62%	1726%	1681%	737%
0.50	0.25	-19%	-3%	-44%	1630%	1520%	752%
0.50	0.50	-25%	1%	-63%	1737%	1313%	706%
0.50	0.75	-25%	1%	-72%	1712%	1423%	741%
0.75	0.10	-21%	-4%	-58%	1720%	1748%	736%
0.75	0.25	-23%	-5%	-56%	1763%	1746%	751%
0.75	0.50	-26%	-7%	-59%	1755%	1687%	740%
0.75	0.75	-23%	-7%	-47%	1725%	1640%	669%

Table 5.2: Overview of difference in required time to golden corner and number of treatment plans found in the golden corner between BRIGHT with and without adaptive steering for the tri-objective problem. Construction of the table is similar to that of Table 5.1.

What can be seen in the table is that using adaptive steering in the tri-objective problem will significantly reduce the amount of time to reach the golden corner and that the amount of solutions found in the golden corner increases significantly.

5.1.3. Discussion

The time to reach the golden corner does not tell everything about the convergence time. Ideally one would take multiple points starting from the golden corner and increasing in both LSI and LCI to the top right corner of the figure. Then using these points one could calculate how long it takes for the optimization to reach those points. That would increase the insight in the convergence time. Unfortunately this idea only came after the experiments had already been performed. Since solid conclusions can also be drawn based upon that single point it was chosen to not repeat the experiments (which took more than 50 hours) to only get the time at those points, both from a time perspective as from a sustainability perspective as it would waist energy.

5.1.4. Conclusion

What can be concluded based upon the results of the experiment is that adaptive steering can significantly improve how far the Pareto approximation front reaches into the golden corner when optimizing for three objectives. Furthermore, it also reduces the time to reach the golden corner. These conclusions can not be drawn for the bi-objective problem. For the bi-objective case there was only an advantage found for the patients that had a longer time to reach the golden corner. Since on average it did not increase the time to reach the golden corner for the best settings but did decrease the time of the hardest patient it is still useful to apply adaptive steering in the bi-objective problem. This conclusion is further strengthened by the fact that for both the bi-objective as the tri-objective problem the number of plans found in the golden corner greatly increases, which are the plans that the clinical experts are interested in. The best settings were based on the visual inspection of the Pareto approximation fronts. Deviating from these settings (increasing adaptive strength, lowering adaptive start requirements) will potentially negatively affect the optimization procedure thereby resulting in worse Pareto approximation fronts. For sake of space, only the plots for the best performing setting are shown, the other plots can be found in Appendices B.2 and B.3.

5.2. HSI edge length

5.2.1. Experiment set-up

To test what the consequences are of optimizing on HSI whilst having a too low number of dose calculation points for 0.5 mm hotspot divisibility, some experiments have been performed.

The number of dose calculations points has been set to 20.000 points per organ/volume. This is the same amount used in the normal set-up of BRIGHT. This results in 140.000 dose calculation points used in total as compared to the 100.000 points in the normal set-up.

For the edge length two different settings have been tested: both the fixed and the variable edge length as described in Section 4.2.2. For the fixed edge length setting all organs get the same edge length of 0.5 mm, the potential downside of this is underestimation of hotspots. This is caused by missing connections between dose calculation points since they are just above this edge length but the received doses does not drop between them. For the variable edge length, the length is set so that, if the assumption is made that all dose calculation points are perfect cubes and uniformly distributed over the volume, then the adjacent dose calculation points are connected. The edge length is calculated per organ by calculating the side length of a cube given the volume per dose calculation point. This edge length should reduce the underestimation of hotspot sizes which could be present in the fixed edge length case. The potential downside is overestimation.

To get insights into the performance of both settings, the results are re-run with the required number of dose calculation points for 0.5 mm hotspot divisibility per organ/volume and a fixed edge length of 0.5 mm. Then the calculated HSI value of the optimization run and the reevaluation run are compared to see the difference and correlation.

To test both edge length settings the settings are tested on all eleven patients. Each experiment is repeated five times to mitigate randomness in the results. All runs are ran until no additional significant improvement are observed. The used hardware is the same as for the adaptive steering experiments.

5.2.2. Results

In Table 5.3, the deviation from HSI with 140.000 points to 700.000+ points are shown for both fixed and variable edge length. For every treatment plan in the final approximation set, the difference in HSI is calculated. Based on these values the minimum, average and maximum are determined for every

run. The results shown in Table 5.3 are the averages of these runs.

	Difference						Average HSI (#hotspots)	
	Min		Mean		Max			
	Fixed	Variable	Fixed	Variable	Fixed	Variable	Fixed	Variable
Patient1	-0.041	0.063	0.032	0.189	0.176	0.399	0.167 (1.15)	0.242 (1.18)
Patient2	-0.037	0.233	0.006	0.405	0.064	0.554	0.462 (0.98)	0.654 (1.00)
Patient3	-0.048	-0.112	0.072	0.077	0.22	0.241	0.182 (1.08)	0.267 (1.35)
Patient4	1.207	-0.09	1.556	-0.003	1.909	0.139	2.040 (2.23)	1.200 (2.98)
Patient5	0	-0.042	0	-0.006	0	0.023	0.000 (0.00)	0.047 (0.27)
Patient6	0.264	-0.09	0.575	0.073	0.924	0.354	0.746 (2.64)	1.141 (2.40)
Patient7	0.25	-0.255	0.758	-0.036	1.426	0.207	0.835 (3.02)	1.566 (3.13)
Patient9	0.798	-0.242	1.136	-0.041	1.903	0.159	1.551 (3.36)	0.857 (2.43)
Patient10	0.798	-0.216	1.392	0.03	2.279	0.448	1.481 (3.63)	2.03 (4.03)
Patient11	0.02	-0.124	0.298	0.027	0.7	0.251	0.443 (2.09)	0.585 (2.13)
Patient12	0.541	-0.093	0.815	0.065	1.452	0.272	0.981 (2.72)	0.711 (2.31)

Table 5.3: Overview of the average, minimum and maximum deviation between registered hotspot sizes in treatment plans evaluated on 140.000 DCPs and the required number of DCPs for 0.5 mm separability for both the fixed and variable edge length hotspot registration. In the two right most columns the average HSI value and the average number of hotspots in treatment plans for both the fixed and variable edge length are shown.

What can be seen in the results is that on average the variable edge length outperforms the fixed edge length, as was expected. What can also be seen is that the average deviation for the variable edge length is less then 0.1 cc (the lower bound volume for a hotspot). The fixed edge length generally underestimates the size of the hotspots. For the adaptive edge length it both underestimates and overestimates the hotspot sizes. The strongest deviation for the variable edge length can be seen for patient 2. When looking at the average HSI value and number of hotspots it can be deduced that it simply misses the one hotspot that is present in those plans.

In Table 5.4, the absolute deviation from HSI with 140.000 points to 700.000+ points are shown for both fixed and variable edge length.

	Absolute Difference						Average HSI (#hotspots)	
	Min		Mean		Max			
	Fixed	Variable	Fixed	Variable	Fixed	Variable	Fixed	Variable
Patient1	0	0.063	0.034	0.189	0.176	0.399	0.167 (1.15)	0.242 (1.18)
Patient2	0	0.233	0.01	0.405	0.08	0.554	0.462 (0.98)	0.654 (1.00)
Patient3	0	0.007	0.08	0.096	0.22	0.244	0.182 (1.08)	0.267 (1.35)
Patient4	1.207	0.015	1.556	0.081	1.909	0.201	2.040 (2.23)	1.200 (2.98)
Patient5	0	0	0	0.011	0	0.064	0.000 (0.00)	0.047 (0.27)
Patient6	0.087	0.002	0.573	0.1	0.807	0.354	0.746 (2.64)	1.141 (2.40)
Patient7	0.25	0.001	0.758	0.085	1.426	0.258	0.835 (3.02)	1.566 (3.13)
Patient9	0.798	0.001	1.136	0.091	1.903	0.276	1.551 (3.36)	0.857 (2.43)
Patient10	0.798	0	1.392	0.107	2.279	0.448	1.481 (3.63)	2.03 (4.03)
Patient11	0.024	0	0.298	0.071	0.7	0.253	0.443 (2.09)	0.585 (2.13)
Patient12	0.541	0.006	0.815	0.1	1.452	0.272	0.981 (2.72)	0.711 (2.31)

Table 5.4: Overview of the absolute average, minimum and maximum deviation between registered hotspot sizes in treatment plans evaluated on 140.000 DCPs and the required number of DCPs for 0.5 mm separability for both the fixed and variable edge length hotspot registration. In the two right most columns the average HSI value and the average number of hotspots in treatment plans for both the fixed and variable edge length are shown.

What can be seen in these results is again that on average the variable edge length outperforms the fixed edge length. Furthermore we see that the absolute deviation are a little more significant, from which can be concluded that both underestimation and overestimation of hotspot sizes is occurring. The average absolute deviations are close to the 0.1 cc lower bound of the hotspot volume, meaning that it will have some problems in matching the actual hotspots.

In Table 5.5, the correlation of the calculated HSI (140.000 points) and HSI (700.000+ points) are shown for both fixed and variable edge length.

	Correlation		Average HSI (#hotspots)	
	Fixed	Variable	Fixed	Variable
Patient1	0.857	0.83	0.167 (1.15)	0.242 (1.18)
Patient2	0.997	0.913	0.462 (0.98)	0.654 (1.00)
Patient3	0.914	0.845	0.182 (1.08)	0.267 (1.35)
Patient4	0.948	0.995	2.040 (2.23)	1.200 (2.98)
Patient5	1	0.971	0.000 (0.00)	0.047 (0.27)
Patient6	0.702	0.989	0.746 (2.64)	1.141 (2.40)
Patient7	0.168	0.997	0.835 (3.02)	1.566 (3.13)
Patient9	0.927	0.991	1.551 (3.36)	0.857 (2.43)
Patient10	0.437	0.996	1.481 (3.63)	2.03 (4.03)
Patient11	0.436	0.979	0.443 (2.09)	0.585 (2.13)
Patient12	0.298	0.983	0.981 (2.72)	0.711 (2.31)

Table 5.5: Overview of the correlation between treatment plan HSI value calculated on 140.000 points with both fixed and variable edge length and treatment plan HSI value calculated on the required DCPs for 0.5 mm separability. In the two right most columns the average HSI value and the average number of hotspots in treatment plans for both the fixed and variable edge length are shown

What can be seen in the correlations is that the fixed edge length shows some low correlations for a number of patients. What can also be seen is that the correlation of the variable edge length is generally high. Two dips in correlation can be observed in patient 1 and patient 3. These dips in correlation can be ascribed to the under and overestimation due to the lower number of dose calculation points. As can be seen for patient 1, 3 and 5 the average HSI values are the lowest, which means the impact of deviation is the greatest. For patient 5, the average HSI is almost 0, this means that correlation is likely to be high as both implementations will not find hotspots.

5.2.3. Discussion

The high correlation between the variable edge length HSI and the full HSI does not necessarily mean that it registers the hotspots sufficiently. There still can be discrepanties between the two metrics. The correlation only shows if the data has the same 'direction', meaning a high HSI for one registration method means a high HSI for the other registration method. It could still be generally under- or overestimating the hotspot sizes. Nonetheless, these over- or underestimations are, based on the seen results, likely to be in the range of 0 and 0.1 cc which might not prove to be problematic for optimization. This over- and underestimation are symptoms of over-fitting, as treatment plans are found that perform well on a lower number of DCPs but their performance reduces when the number of DCPs increases.

5.2.4. Conclusion

From the results of the experiments, it can be concluded that the variable edge length outperforms the fixed edge length. Furthermore, the pitfall of HSI with a lower number of dose calculation points will be the registration of smaller hotspots with a size just above the hotspot volume lower bound. In general, the HSI can be used in a setting with a lower number of DCPs per organ (20.000) and still achieve high correlations with the actual HSI and low deviations. This means that the HSI could potentially be used to improve treatment plans during optimization. In the remainder of this research the variable edge length is used for directly optimizing on the HSI.

6

Methods and Experiment Set-Up

In this chapter the set-up of the experiments of this research will be discussed. This is followed by a description of the methods used for calculating the performance metrics of different BRIGHT adaptations to gain insights in the results.

6.1. Experiment set-up

6.1.1. General

All experiments have been performed on the same set-up as the adaptive steering experiments. This set-up is a server with 20 Intel Xeon Processor E52630 v4 @ 2.20 GHz with multi-threading, 128 GB RAM and a NVIDIA GeForce GTX TITAN X with 12 GB of GDDR5 VRAM.

All experiments, except for the reevaluation of the clinical results, have been performed in five-fold to counter the contribution of randomness in the results.

Before every experiment the convergence time was determined by running the experiment with a run-time in which it could be reasonably assumed that the optimization converged (i.e. long run time). Based on the generational Pareto approximation results, plots were made to visually inspect when convergence had happened, i.e. when no improvements were observed anymore. A single maximum run-time was set for an experiment in which all patients were used. These maximum run-times per experiment were determined in a worst case manner, meaning that the longest convergence time among the different patients was chosen as maximum run-time for the whole group of patients.

After the experiments were conducted all results are reevaluated for hotspot detection using a large number of dose calculation points, as described in Table 4.1. The results from that reevaluation are reported on.

All proposed adjustments to BRIGHT are first tested as a third objective to get insights in how they perform. Based on these results a judgement will be made if they are suited for further exploration in a bi-objective setting as either a constraint or as augmentation of the current objectives and find the range of the metric throughout the patient group..

6.1.2. BRIGHT

Several parameters have to be set for BRIGHT. In this section the two deviations in standard parameters with the basic version of BRIGHT are discussed. The deviations are the number of clusters, and the population size. In the bi-objective problem optimization, the number of clusters is set to 5 and the population size to 96. In the tri-objective problem optimization this is set to 12 clusters and 288 individuals as population size, as explained in Section 4.1.1. Ideally, these two parameters would be determined using experiments for all different third objectives. This would require a multitude of additional experiments and was deemed to be too computationally expensive and time consuming for this thesis..

Another change as compared to the previous version of BRIGHT is the creation of normal-tissue around the target volumes. This is done to be able to register hotspots in the normal-tissue. The normal-tissue is created by sampling DCPs in a margin around the target volumes (i.e. prostate and seminal vesicles). The margin is set to 1 cm. This value was established together with a clinical expert based

on how far dwell positions are activated outside of the target volumes and the normal-tissue hotspots seen in previous patients. The normal-tissue DCPs are sampled after the DCPs for the organs and in a mutually exclusive way. This means that if the 1 cm margin stretches into an organ, then DCPs will only be sampled up until the boundary of the organ, but not within the organ as DCPs are already sampled there.

6.1.3. Patient data

A dataset of 11 patients with intermediate- and high-risk prostate cancer, previously treated at the Amsterdam UMC in Amsterdam, the Netherlands, is used for all experiments in this thesis. These patients were treated between April 2020 and July 2021 with external beam radiotherapy on the prostate and base of the seminal vesicles to a dose of 44 Gy in daily fractions of 2.2 Gy followed by a single dose of 15 Gy HDR brachytherapy on the prostate. A median of 20 (range: 17–24) catheters were implanted, resulting in a median of 333 (range: 216–429) dwell positions. Catheter reconstruction and contouring of Regions Of Interest (ROIs) were done on three orthogonal pelvic T2-weighted turbo spin echo MRI (Ingenia 3 T Philips Healthcare, Best, the Netherlands) scans with a resolution of 0.52 9 0.52 mm, and a slice thickness of 3.0 mm with a 0.3 mm gap. Three interpolated contours were added between each contoured slice of each ROI.

For patient 8 there were technical difficulties in the export of the data and therefore could not be used in this research.

6.2. Performance metrics

Several different performance metrics are used to give insights into the performance of different BRIGHT set-ups. These metrics are motivated both by performance requirements from the clinic, as well as metrics to give more insights in the underlying problem of reducing hotspots.

The metrics can be divided into 4 categories: hotspots, correlations, time, and robustness. Each of these categories will be discussed in the subsections that follow. Next to the quantitative metrics, the Pareto approximation fronts are plotted in the results section for all experiments to visually inspect the performance. These plots are made with all the plans from the 5 runs for every experiment.

Hotspots

One of the most important performance metrics in this research is if there are hotspots in the treatment plans. For every treatment plan every hotspot (according to the set specification of hotspots) is recorded and written to file using its respective volume. Based on these hotspots the sum of hotspot volumes and the number of hotspots is used for plan comparison.

Given that Pareto approximation fronts of a tri-objective problem can be a broader 3D surface, the metrics are calculated in 3 ways. The first one is simply taking the whole front, the second one is only looking at the solutions that are non-dominated in LCI and LSI (i.e. the bi-objective Pareto approximation front) and the third one is the ϵ-approximation front in the 2D Pareto approximation front. The ϵ-approximation front are all plans that are within 5% of both LSI and LCI of at least 1 treatment plan in the bi-objective approximation front. By splitting up the results in these 3 ways more information about the quality of the treatment plans is generated from the aggregate numbers. On top of those 3 subsets, another subset is used for the calculation of the metrics, namely the full Pareto approximation front but with the hotspots smaller than 0.15 cc filtered out. This is done to check the conclusions on robustness. The default lower bound of hotspots is 0.10 cc, which was set in accordance with the clinical experts. To gain more confidence in the conclusions based upon the results, this value is controlled for by checking if increasing this lower bound will show different results.

To assess the performance on LCI and LSI for different HSI upper bound filters on the resulting Pareto approximation fronts, both a visual and a non-visual method is used. The visual method is simply plotting the Pareto approximation fronts on a 2D grid, with on the x-axis the LCI value and on the y-axis the LSI value. The HSI value is the shown by applying a colour gradient on the dots that represent the treatment plans. For the non-visual way of describing the performance on LCI and LSI three different metrics are calculated. The first metric is the best achieved LCI value in the Pareto approximation front given a certain upper bound on HSI, the second is the best achieved LSI value given an HSI upper bound and the last is the L-value for a given HSI upper bound. The L value is calculated in a worst-case manner over the achieved LCI and LSI values. For every treatment plan

the minimum is calculated between LCI and LSI, then from the list of minimums the maximum is taken. This value is the L-value.

Time

In the clinic, time is not unlimited. Therefore, the faster BRIGHT is the better. However, an increase in computation time of BRIGHT could also result in reduced work in the later clinical workflow, for instance the need for manual adjustments.

For the time performance of the different adjustments the following metrics are calculated: the number of evaluations per second, time to reach the golden corner, number of evaluations to reach the golden corner and the number of real evaluations to reach the golden corner. The time to reach convergence is determined visually on the basis of three scatter plots per time step per experiment. Where x, y and z axis combination of the scatter plots are: LSI and LCI, LSI and third objective, and LCI and third objective.

Robustness

As described in Section 3.1.2, one of the aspects of homogeneity is the increased robustness. To fully assess the robustness of the treatment plans one would have to simulate disturbances and measure the changes in outcome. This is out of scope for this research project, nonetheless since it is an important aspect it is interesting to get a hint on the effect the different adaptations of BRIGHT might have on this. To do so, the DTMR values for both the absolute and the relative DTMR (as described in Balvert et al., 2014) will be calculated for all treatment plans.

Correlations

To see which aspects of a treatment plan are predictive for the hotspot size several different correlations are calculated. These correlations will help us gain insights in how well different adjustments are able to capture the creation of hotspots.

The correlation between several features and the resulting HSI value are calculated and reported on. The features used for the correlation metrics are: LSI, LCI, Number of dwell positions used, Maximum dwell time and the sum of extra V indices.

Statistical significance

To assess the significance of the results, a statistical significance test is applied. The two-tailed Mann Whitney U test with p value of 0.05 is used for comparing results between the 5 runs of 2 experiments.

With this explication of the performance metrics the first part of the following research questions can be answered:

Research Question: BRIGHT performance

What are the important performance aspects of BRIGHT for HDR-BT *and how do different enhancements score on these aspects?*

The important performance aspects of BRIGHT for HDR-BT treatment plans are how far the Pareto approximation front has reached into the golden corner, what the resulting hotspot sizes in those plans are and how long it takes to reach these results.

7

Results

In this chapter, the results of the performed experiments will be discussed to give answer to the research questions. This will start with the reevaluation of the Pareto approximation fronts produced in the clinic to get a feel for what the baseline is. Next, the experiments for all different adaptations of BRIGHT will be discussed.

7.1. Results

7.1.1. Re-evaluation of clinical results

Assessing the performance of the proposed BRIGHT adjustments requires a baseline. This baseline is determined in two ways. Firstly, re-evaluation of the Pareto approximation fronts as produced in the clinic. Secondly, running the bi-objective BRIGHT multiple times (five) to assess the performance. Running the bi-objective BRIGHT again is done to mitigate the influence of randomness on the results.

In Figures 7.1 and 7.2 the Pareto approximation fronts of two patients, as produced in the clinic are shown.

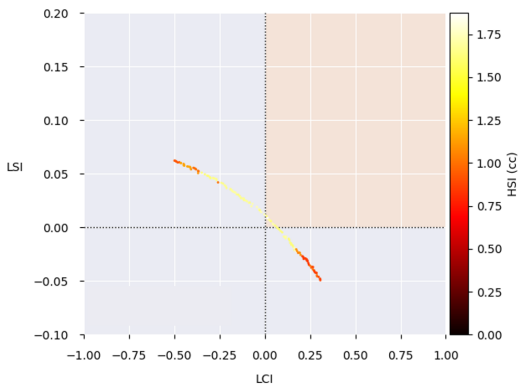

Figure 7.1: Pareto approximation front from the clinic for patient 7 reevaluated for HSI values, zoomed in to the golden corner.

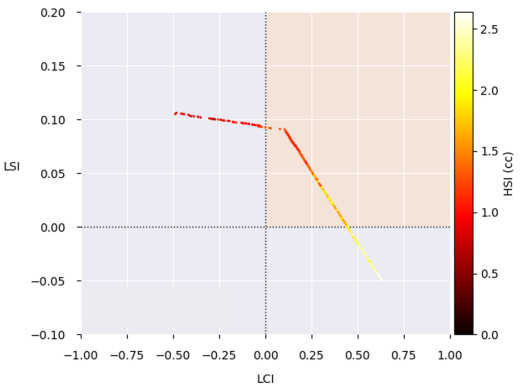

Figure 7.2: Pareto approximation front from the clinic for patient 10 reevaluated for HSI values, zoomed in to the golden corner.

These figures have been zoomed in on the golden corner as that is the region of interest to the clinical experts, all plans with a LCI value lower than -0.5 or LSI lower than -0.05 are discarded. The LCI and LSI values for the treatment plans are shown on the x- and y-axis respectively. On the right side of the figures a colour bar is shown. This colour bar indicates the colour corresponding a HSI value of the treatment plans. What becomes clear from the patient on the left is that the presence of hotspots is not necessarily correlated with the LCI value. Moving to a treatment plan with increased or decreased LCI value can both lead to plans with a higher HSI value. In the figure on the right this behaviour is not visible.

Calculated metrics

These previous observations become even more clear when looking at the reported metrics of the basic run with five repetitions. In Tables 7.1, 7.2, 7.3 and 7.4, the average results of HSI, number of hotspots and correlation of HSI with LCI and LSI in the golden corner are shown.

Average HSI in GC (cc)			
Patient1	0.182	Patient7	1.017
Patient2	0.577	Patient9	1.512
Patient3	0.223	Patient10	1.493
Patient4	2.036	Patient11	0.345
Patient5	0.013	Patient12	0.810
Patient6	0.715		

Table 7.1: Overview of average achieved HSI value for treatment plans in the golden corner for different patients which are the result of the basic BRIGHT configuration.

Average #hotspots in GC			
Patient1	1.218	Patient7	3.605
Patient2	1.006	Patient9	3.043
Patient3	1.078	Patient10	3.405
Patient4	2.609	Patient11	1.835
Patient5	0.103	Patient12	2.156
Patient6	2.755		

Table 7.2: Overview of the average number of hotspots in treatment plans in the golden corner for the different patients which are the result of the basic BRIGHT configuration.

Average Corr(LCI) in GC (cc)			
Patient1	0.676	Patient7	0.367
Patient2	0.827	Patient9	0.802
Patient3	0.762	Patient10	0.860
Patient4	0.958	Patient11	0.722
Patient5	0.459	Patient12	0.759
Patient6	0.718		

Table 7.3: Overview of the average correlation between achieved LCI value and HSI value for treatment plans in the golden corner for different patients which are the result of the basic BRIGHT configuration.

Average Corr(LSI) in GC			
Patient1	-0.696	Patient7	-0.378
Patient2	-0.846	Patient9	-0.896
Patient3	-0.816	Patient10	-0.901
Patient4	-0.965	Patient11	-0.803
Patient5	-0.440	Patient12	-0.816
Patient6	-0.714		

Table 7.4: Overview of the average correlation between achieved LSI value and HSI value for treatment plans in the golden corner for different patients which are the result of the basic BRIGHT configuration.

What can be seen is that the average HSI value is different for each patient, the same holds for the correlation of LCI and LSI with the HSI value. For some patients this correlation is high (patient4), where for other is it low (patient7). On average, one could say that LCI and LSI are both predictors of hotspot sizes in treatment plans but not too strong (on average 0.72 for LCI and -0.75 for LSI). One thing that can also be seen is that the need for hotspot reduction for some patients (patient4) is higher than others (patient5). What is important is that the adjustments in BRIGHT will reduce the hotspots in the patients where it is needed but will not deteriorate the quality of the treatment plans for the other patients.

In Figure 7.5, the remainder of the calculated correlations are shown. Active-dwell is the number of dwell positions used, max-dwell is the maximum dwell time used and v-indices is the sum of extra added V indices.

	Correlation HSI		
	active_dwell	max_dwell	v_indices
Patient1	-0.132	0.261	0.748
Patient2	0.151	0.240	0.944
Patient3	0.228	0.463	0.856
Patient4	-0.302	0.590	0.983
Patient5	0.136	0.451	0.451
Patient6	0.273	0.257	0.814
Patient7	-0.231	0.248	0.676
Patient9	-0.005	0.437	0.913
Patient10	0.013	0.394	0.967
Patient11	-0.297	0.281	0.866
Patient12	-0.110	0.231	0.871

Table 7.5: Overview of correlations between HSI and the number of active dwell positions, the maximum dwell time and the achieved sum of extra V indices in treatment plans in the golden corner for different patients for the bi-objective BRIGHT configuration.

What can be seen from these results is that there is no correlation between the number of dwell positions used and the size of the hotspots. For the maximum dwell time there is some correlation, but not a strong one. The V indices in general do show a higher correlation with the HSI values, this was expected as it indirectly measures the hotspots. However the correlation is not perfect as can be seen in patient 5 and 7. This can be attributed to the fact that it is a more general measure that not necessarily favors a plan with smaller hotspots over a plan with bigger hotspots. For some patients it does show a high correlation, such as patient 4 and 10. This corresponds with the correlation of those patients on LCI as shown in Figure 7.3. These result hint towards a bigger potential for using the extra V indices for these patients.

Time performance

In time performance there are some discrepancies to be seen. The time metrics are shown in Figure 7.6.

	Evaluations per second	Time to GC	Evaluations to gc	Number of dwell positions
Patient1	751.090	17.915	13103.800	300
Patient2	718.941	14.856	9847.800	301
Patient3	501.527	21.374	10677.000	381
Patient4	628.531	95.785	58029.000	305
Patient5	524.062	37.441	19722.200	429
Patient6	621.194	17.881	11337.800	333
Patient7	567.604	65.439	38197.200	334
Patient9	594.213	25.310	15069.600	367
Patient10	704.067	14.762	10157.200	288
Patient11	489.032	23.333	11090.800	355
Patient12	525.095	32.097	17803.400	335

Table 7.6: Overview of the number of evaluations per second, the time needed to reach the golden corner, the number of evaluations to reach the golden corner and the number of dwell positions for the different patients for the basic configuration of BRIGHT.

In general the optimization reaches the golden corner in around 30 seconds, however for some patients (4 and 7) it takes longer. This increased time is not necessarily caused by an increased complexity of calculation, because the number of dwell positions for patient 4 is not higher than others. The same holds for the number of evaluations per second. It simply takes more evaluations to reach the golden corner. Furthermore, for these two patients, the golden corner is just barely reached. This means that for these patients it is more difficult to find clinical protocol satisfying treatment plans, which is due to an unfavorable geometry of the implanted catheter and anatomy of the patient.

Treatment plan robustness

Lastly the robustness metrics for the plans in the golden corner are shown in Figure 7.7. Next to the achieved average value for DTMR-A and DTMR-R the correlation of those number with HSI have also been calculated.

	DTMR_A		DTMR_R	
	Value	Correlation	Value	Correlation
Patient1	1886.02	0.61	552405.80	0.31
Patient2	1740.17	0.63	400749.37	0.55
Patient3	1813.14	0.82	538737.58	0.71
Patient4	2298.85	0.91	830611.99	0.66
Patient5	2572.16	0.59	750281.90	0.24
Patient6	2599.95	0.71	783754.84	0.34
Patient7	1941.52	0.40	755184.10	0.20
Patient9	3765.45	0.88	1124829.63	0.37
Patient10	3233.63	0.89	922296.32	0.65
Patient11	1698.10	0.77	476623.20	0.51
Patient12	3080.14	0.77	860277.28	0.44

Table 7.7: Overview of the average DTMR-A and DTMR-R values and their correlation with HSI value for treatment plans in the golden corner for the different patients for the basic configuration of BRIGHT.

These absolute values are hard to interpret, but are shown here to allow for comparison with adjustments later in the report. When looking at the correlations of the DTMR values and the HSI it can be seen that for some patients high correlations are achieved (patient 4) and for some low correlations (patient 7). What can further be seen is that the relative DTMR has a lower overall correlation with the HSI, which is logical as it allows for larger absolute deviations due to its relative nature. In literature no motivation was given on why to use relative DTMR as compared to absolute DTMR.

Adjusted treatment plan

After optimization of the treatment plans some plans are exported for inspection in OncentraBrachy, which visualizes the iso-dose lines. During this inspection some adjustments can be made if deemed necessary. In this section these applied changes are explored with regards to hotspots and an overview of the chosen treatment plans and the used clinical plans are shown Table 7.8.

	LCI		LSI		HSI		Most significant change	
	Before	After	Before	After	Before	After	Promotes	Cost
Patient1	0.403	0.402	0.014	0.018	0.000	0.000	No changes	No changes
Patient2	0.467	n/a	0.010	n/a	0.700	n/a	...	
Patient3	0.428	0.072	0.009	0.045	0.372	0.236	Decreased hotspot	Prostate coverage
Patient4	0.009	n/a	0.003	n/a	1.991	n/a	...	
Patient5	0.263	0.241	0.006	-0.004	0.000	0.000	Spreading dwell peaks	None
Patient6	0.501	0.431	-0.022	-0.008	0.741	0.889	Sparing bladder	Increased hotspot
Patient7	0.043	-0.035	0.003	0.005	1.723	0.295	Hotspot near rectum	Coverage
Patient9	0.478	0.485	0.010	-0.013	1.699	1.709	No significant changes	No significant changes
Patient10	0.364	0.106	0.020	0.091	2.007	1.108	Decreased hotspot	Prostate coverage
Patient11	0.319	0.460	0.027	-0.014	0.503	0.385	Prostate coverage	Urethra sparing
Patient12	0.215	0.293	0.032	-0.004	0.458	0.612	Bladder sparing, hotspot	Rectum sparing, hotspot

Table 7.8: Overview of the changes applied to exported treatment plans after optimization in BRIGHT as seen in the clinic.

In the table above the changes for patient 2 and patient 4 are omitted. The reason for this is that during optimization for patient 2, a non-existing catheter was used. Due to time considerations, the optimization was not restarted, but changes were made in OncentraBrachy. Comparing the before and after treatment plans would not be a fair comparison and therefore they are omitted. For patient 4 the clinical experts deviated from the used clinical protocol in optimization during adjustment, because the treatment procedure was a re-irradiation, thereby neglecting the coverage aim of the seminal vesicles.

What can be seen in the table above is that the applied changes are not necessarily related to hotspot. For three patients (3, 7 and 10) it could be directly detected that a hotspot was reduced. For patient 6 an existing hotspot was increased in size. For patient 12 the changes were aimed at reducing a hotspot near the bladder, but in applying these changes a hotspot was created near the rectum.

7.1.2. Objective Metrics and Correlation

In this section a description is given on the achieved values for the different adaptions as third objectives per patient and the correlation of different treatment plan features with the HSI value.

In Table 7.9, a summary of the average achieved values for the third objectives is given per adaptation.

	HSI		V indices		DTMR		DLDM	
	min	mean	min	mean	min	mean	min	mean
Patient1	0.000	0.110	1.441	2.089	219.794	525.559	77.168	1364.165
Patient2	0.000	0.263	1.309	1.789	174.871	467.652	16.388	1614.940
Patient3	0.000	0.203	0.702	1.442	91.385	446.625	23.182	1358.117
Patient4	0.891	1.606	2.966	3.852	285.598	614.600	868.925	6852.421
Patient5	0.000	0.050	1.631	2.189	333.924	696.096	63.220	2193.969
Patient6	0.120	1.100	2.443	3.887	301.431	805.809	58.580	3814.813
Patient7	0.087	1.831	1.936	3.674	214.926	546.841	102.288	2914.065
Patient9	0.450	1.270	3.417	5.032	538.024	974.674	308.716	3976.063
Patient10	0.136	2.017	2.518	5.051	396.663	1123.025	79.957	5509.272
Patient11	0.000	0.587	1.778	2.696	66.032	376.239	0.000	689.169
Patient12	0.072	0.668	3.049	4.271	346.092	666.680	269.864	4206.301

Table 7.9: Average and minimum achieved third objective values for the proposed adaptations, for the treatment plans in the 5% Pareto approximation front for the different patients.

There is no need to dive deep into what these numbers mean. The important insight to get from this table is that none of the adaptations show similar values across all patients. For every adaptation there is a difference in how the patients score on them, when optimized on these objectives. The motivation to show this table is to try and see if an aspiration value could be distilled for use in a bi-objective setting.

To see what predictive power of LSI and LCI have for the achieved HSI values, their correlations are calculated. These correlations have been calculated based on the results from optimizing on the HSI value directly. Given that the LSI and LCI values are aggregates of multiple sub-functions and many of these sub-functions are unrelated to hotspots, also the correlation of the sum of extra V indices is calculated.

Optimizing with a third objective results in a 3D approximation surface rather than a line. Given that the interest is in treatment plans close to the 2D Pareto approximation front, only the plans on the approximation front and the plans that are within 5% distance of the Pareto approximation front are used to calculate the correlations. The results of these correlations are shown in Table 7.10.

	LCI		LSI		Vindices	
	Pareto	5%-Pareto	Pareto	5%-Pareto	Pareto	5%-Pareto
patient1	0.755	0.744	-0.815	-0.798	0.861	0.851
patient2	0.835	0.761	-0.811	-0.732	0.953	0.930
patient3	0.552	0.469	-0.544	-0.461	0.742	0.715
patient4	0.750	0.754	-0.749	-0.735	0.935	0.932
patient5	0.581	0.441	-0.575	-0.409	0.698	0.567
patient6	0.262	0.335	-0.276	-0.336	0.866	0.885
patient7	0.441	0.398	-0.394	-0.346	0.982	0.981
patient9	0.460	0.487	-0.475	-0.456	0.779	0.775
patient10	0.339	0.326	-0.364	-0.320	0.949	0.952
patient11	0.454	0.435	-0.494	-0.456	0.804	0.796
patient12	0.460	0.345	-0.518	-0.360	0.867	0.813

Table 7.10: Overview of correlation between HSI with LCI, LSI and sum of extra V indices in treatment plans in the 2D- and 5%-2D-Pareto approximation front for the different patients for the BRIGHT configuration with HSI as third objective.

To increase the confidence in the results, they have been reevaluated when increasing the hotspot volume lower bound to 0.15 cc rather than 0.1 cc. The relative changes in achieved correlations are shown in Table 7.11.

	LCI		LSI		Vindices	
	Pareto	5%-Pareto	Pareto	5%-Pareto	Pareto	5%-Pareto
patient1	-22%	-20%	-13%	-12%	-10%	-9%
patient2	0%	0%	0%	0%	0%	0%
patient3	-26%	-24%	-25%	-25%	-18%	-17%
patient4	2%	2%	2%	2%	1%	1%
patient5	-21%	-28%	-21%	-31%	-15%	-19%
patient6	-11%	20%	-10%	-8%	-2%	-2%
patient7	1%	1%	0%	1%	0%	0%
patient9	1%	1%	0%	2%	-1%	-1%
patient10	-4%	-6%	-4%	-7%	-1%	-1%
patient11	-2%	-6%	-1%	-4%	0%	-1%
patient12	2%	-1%	2%	0%	0%	-1%

Table 7.11: Overview of correlation between HSI with LCI, LSI and sum of extra V indices in treatment plans in the 2D- and 5%-2D-Pareto approximation front for the different patients for the BRIGHT configuration with HSI as third objective after applying a lower bound of 0.15 cc for hotspots instead of 0.1 cc.

What can be seen in these tables is that the correlation between LCI and LSI with the HSI for many patients are low. Only for patient 1, 2 and 4 it could be said that there is some correlation. Increasing the hotspot volume lower bound only worsened the predictive value of LCI and LCI for HSI. When looking at the sum of extra V indices, one can see that they are generally high. However, for some patients this is not the case. These patients are 1, 3, 5 and 11. When increasing the hotspot volume lower bound we observe a decrease in predictive power. This drop is the strongest in patients that already had a lower predictive value. Nonetheless, the extra added V indices could, for a certain degree, be used to predict the HSI value.

7.1.3. Hotspot size

In this section the achieved results relating the hotspots metrics will be discussed. Every proposed adjustment will be discussed in order. The distribution of HSI values over a population does not lend itself to be reasonably explained using a single metric for the whole set of treatment plans in the population. The best found approach to inspect potential improvements has been visually. Therefore this is done in this section. Showing all figures here would deteriorate the readability of this report, therefore a selection to highlight insight is shown. The full set of Pareto approximation front visualisations can be found in Appendices B.5, B.6, B.7 and B.8.

First an overview of the achieved hotspots sizes for the basic BRIGHT configuration is shown in Figure 7.3, for comparison. In this figure only the plans with a HSI value lower than 0.5 are shown. The figures with bound [0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5] are added in Appendices B.5, B.6, B.7 and B.8 for all experiments.

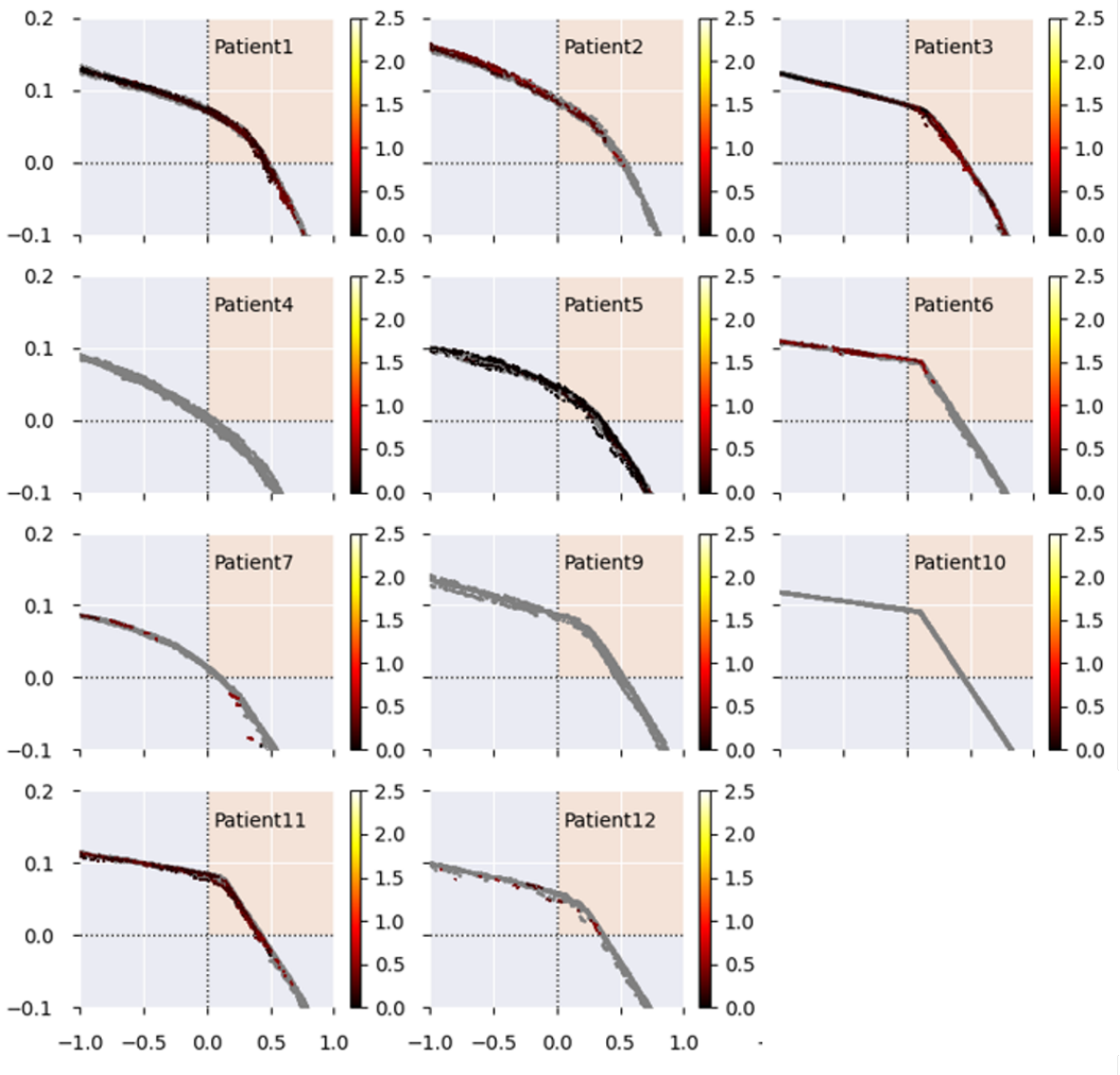

Figure 7.3: Pareto approximation fronts for the different patients resulting from the basic BRIGHT configuration. Only treatment plans with an HSI value lower than 0.5 HSI value are shown. The grey Pareto approximation front is from the bi-objective optimization, which is not filtered on HSI value.

From these Pareto fronts it once again becomes clear that there are differences between patients on how large the hotspots are that are present in the treatment plans when optimized with the standard BRIGHT configuration. For patients 4, 7, 9, 10 no plans were produced without an HSI value of over 0.5 cc. For patient 6 the plans only partly cover the golden corner.

HSI

In this subsection the results regarding the formation of hotspots with the extention of HSI as the third objective will be discussed. In Figure 7.4, the results are shown in a graphical manner.

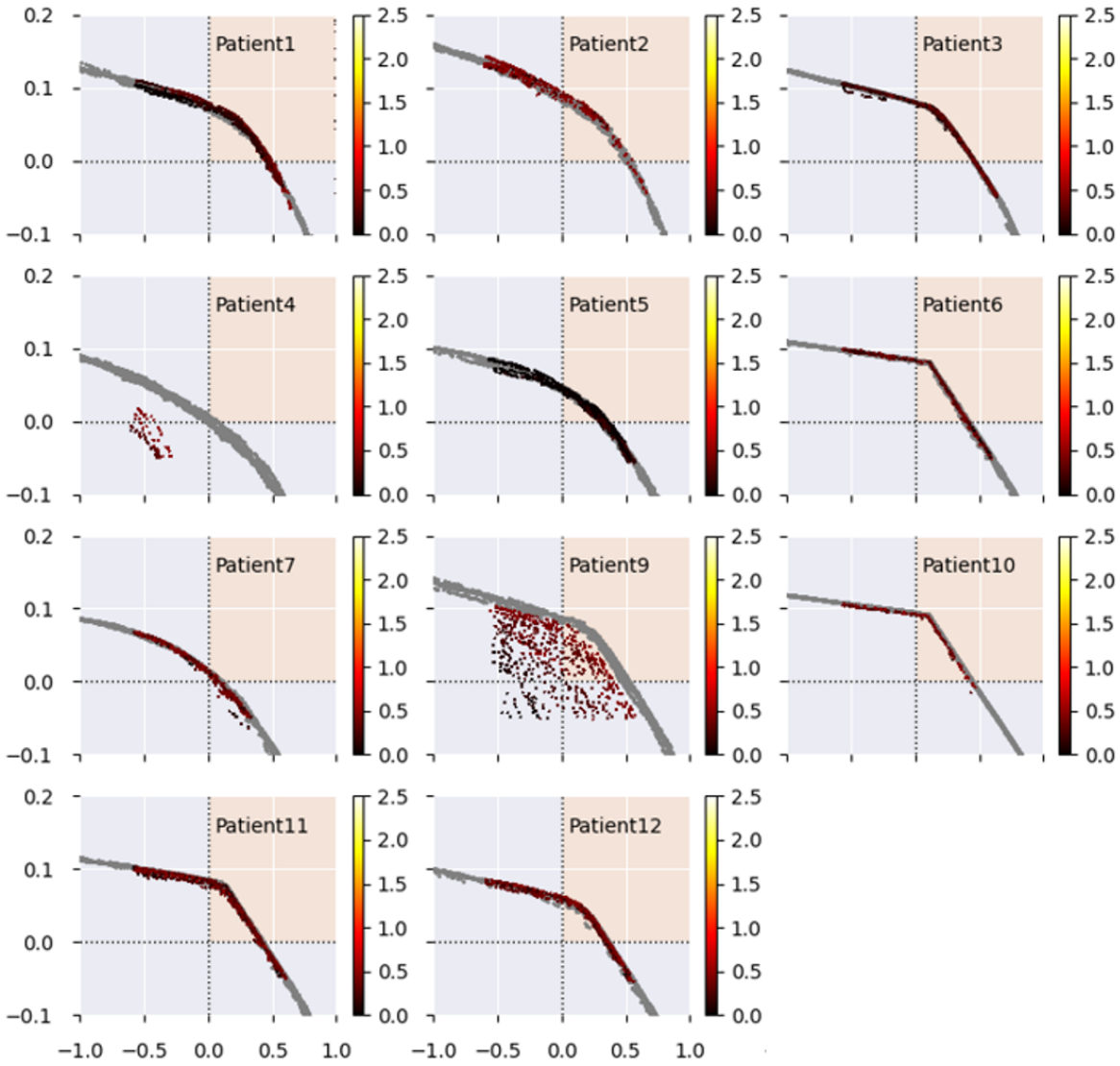

Figure 7.4: Pareto approximation fronts for the different patients resulting from the BRIGHT configuration with HSI as third objective. Only treatment plans with an HSI value lower than 0.5 HSI value are shown. Grey Pareto approximation front is from the bi-objective optimization, which is not filtered on HSI value.

What becomes clear from these figures when compared to the basic results is that it is possible to reduce the formation of hotspots whilst maintaining an equal LSI and LCI value for the treatment plans. This becomes most clear in the plot for patient 10, where no plans were found with an HSI lower than 0.5 for the basic BRIGHT configuration, but now the full golden corner is covered.

Another interesting observation is that for patient 9 there is a broad 3D-surface rather than a thicker line (it also happens for patient 4, this becomes more clear with more slices/upper bounds, which can be found in Appendix B.5). This behaviour occurs due to a new trade-off, i.e. when a slight improvement in one objective results in a deterioration of another objective or vice versa. For this to occur, there must be a high correlation between the objectives. When looking at the correlations reported in Table 7.3 and 7.4, we can see a correlation of 0.8 and -0.9 for LCI and LSI respectively for patient 9 (patient 4 has LCI: 0.96, LSI: -0.97).

What is thus interesting to see for patient 4 and 9, is that when it is allowed to reduce the quality of the treatment plans with regards to LCI and LSI for only a small percentage then it is possible to find treatment plans that perform better for the HSI. This is a new trade-off to take into account for the clinical experts. These patients were also the most difficult to minimize the HSI for whilst retaining similar DVI values.

Sum of extra V indices

In this subsection, the results regarding the formation of hotspots with the extension of the sum of extra V indices as the third objective will be discussed. In Figure 7.5, the results are shown in a graphical manner.

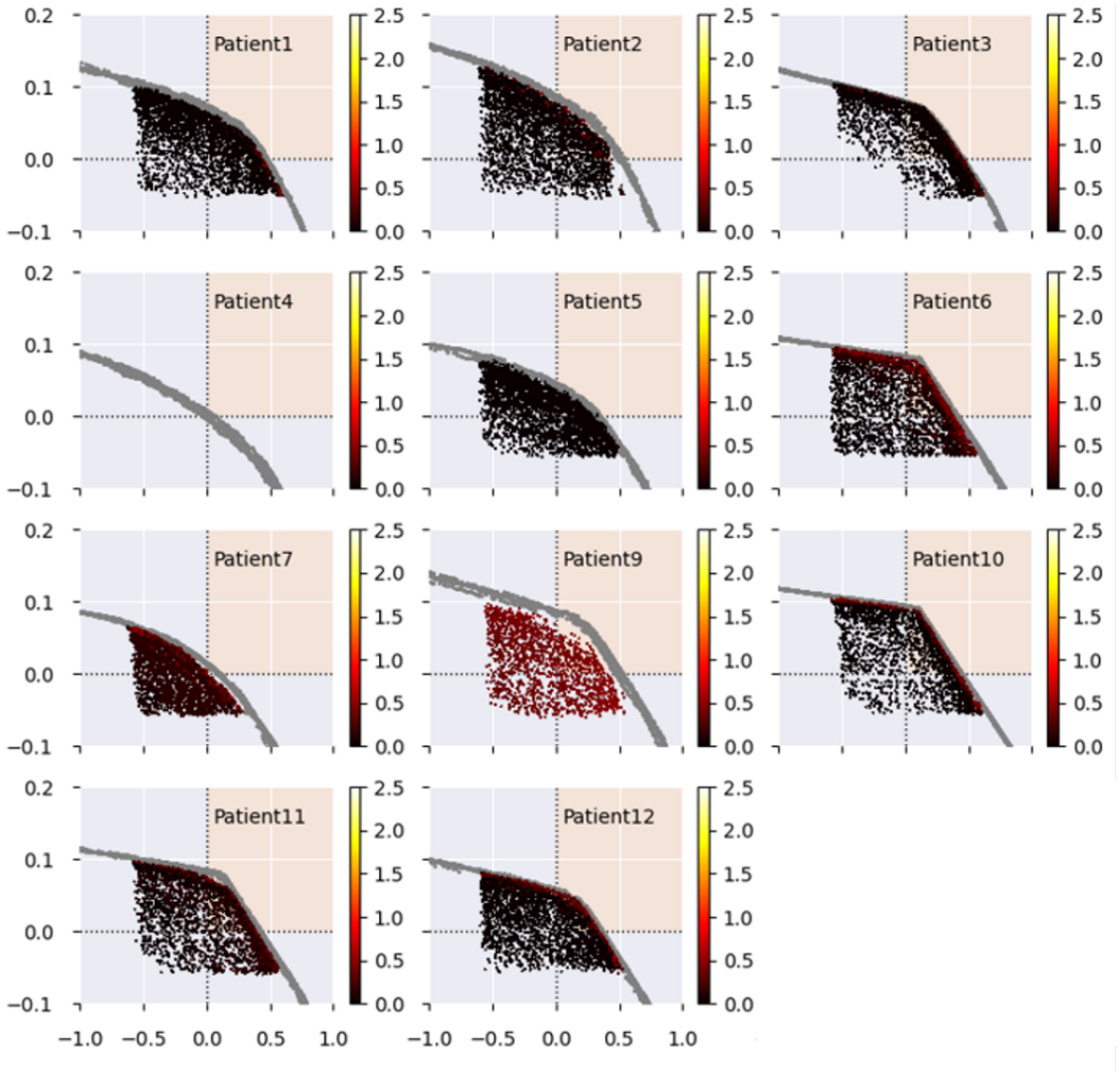

Figure 7.5: Pareto approximation fronts for the different patients resulting from the BRIGHT configuration with the sum of extra V indices as third objective. Only treatment plans with an HSI value lower than 0.5 HSI value are shown. Grey Pareto approximation front is from the bi-objective optimization, which is not filtered on HSI value.

The difference with using HSI and the sum of extra V indices instantly becomes clear when looking at the Pareto approximation fronts. When using the sum of extra V indices, it results in broader 3D surfaces rather than thicker lines. At first glance this looks advantageous as more plans are visible with a low HSI value. This however also has two disadvantages. First of all, it means that there is a reduced correlation between the HSI and the sum of extra V indices, since otherwise the plans with low HSI that are near the top right corner values would dominate the plans in the bottom left corner. This means that it is not steering directly on the HSI, which was not unexpected. The second downside is that the population and clusters are now required to approximate a larger area in 3D space as compared to HSI, which resuls in a reduced pressure towards the top right corner (i.e. the golden corner). The consequence of this is that at convergence this set-up was not able to match all the LCI and LSI values of the bi-objective approach, which was the case with using the HSI as third objective.

The sum of extra V indices is able to reduce the HSI values as compared to the basic BRIGHT

configuration. It outperforms the HSI as third objective for plans with an HSI value below 0.25 cc. For treatment plans with HSI value below 0.5 cc, it is outperformed by the HSI as third objective as that objective pushes further into the golden corner. Furthermore it is not able to find plans with an HSI below 0.5 cc for patient 4.

Dwell Time Modulation

In this subsection the results regarding the formation of hotspots with the extension of the absolute DTMR as the third objective will be discussed. In Figure 7.6, the results are shown in a graphical manner.

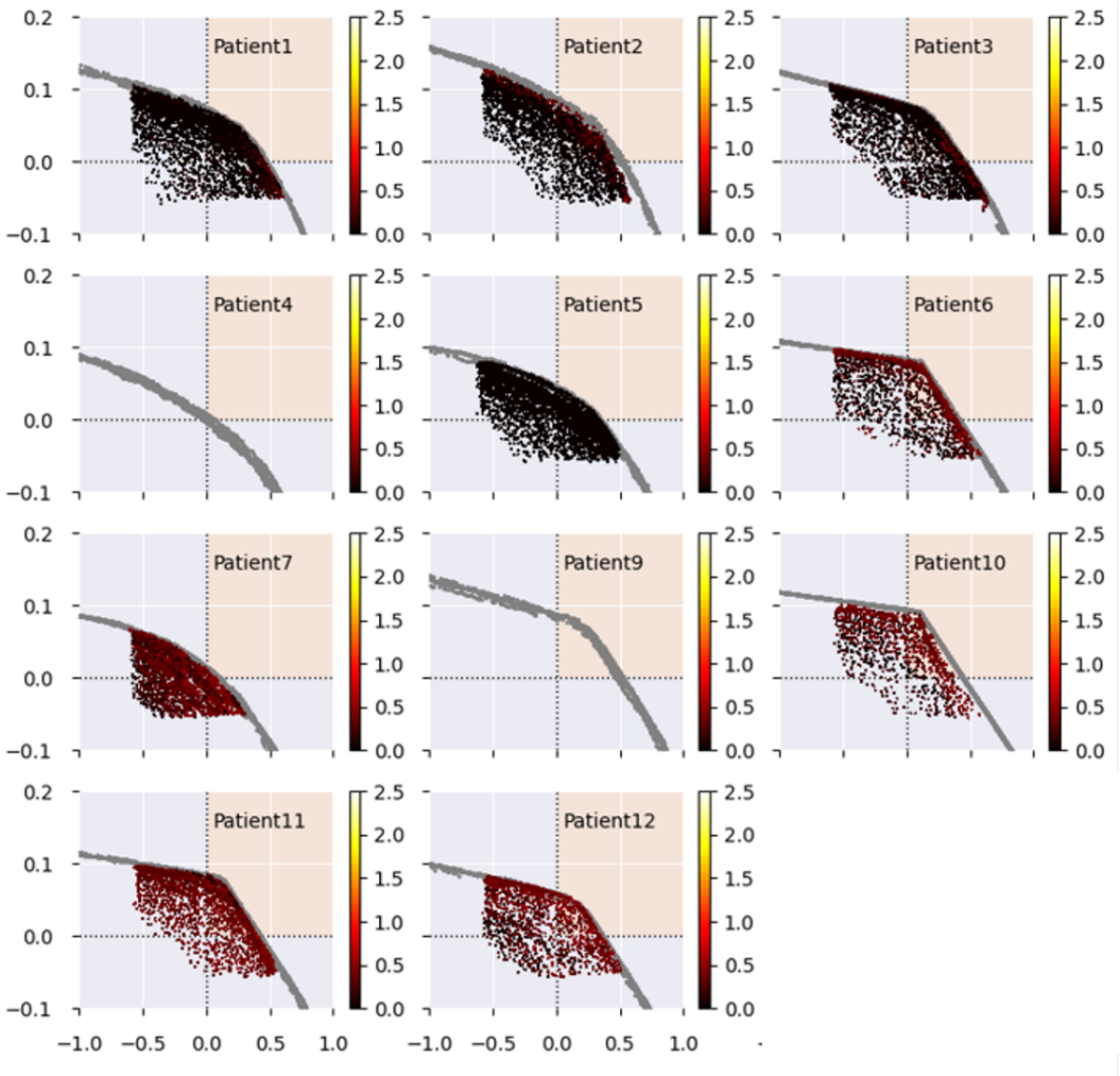

Figure 7.6: Pareto approximation fronts for the different patients resulting from the BRIGHT configuration with DTMR as third objective. Only treatment plans with an HSI value lower than 0.5 HSI value are shown. Grey Pareto approximation front is from the bi-objective optimization, which is not filtered on HSI value.

The DTMR is able to reduce the HSI value of treatment plans compared to the basic configuration of BRIGHT, but not for patients 4 and 9, which can be seen in the figures in Appendix B.7. For the DTMR the same conclusions regarding the broader 3D surface as compared to a thicker line hold as the ones for the sum of extra V indices.

What can be seen however is that the performance of the DTMR is worse compared to that of the sum of extra V indices. For patient 4 and 9 it is not able to produce any treatment plans with a HSI index lower than 0.5 cc. For patients 6, 7, 10, 11 and 12 the plans have a higher HSI index as compared to that of the sum of extra V indices. This behaviour can be explained by the lower correlation that DTMR

has with the HSI as compared to the sum of the extra V indices. The DTMR is able to push further into the golden corner compared to the sum of extra V indices. Compared to the HSI it performs better for creating treatment plans with HSI lower than 0.25 cc, but worse for creating treatment plans with an HSI value of lower than 0.5 cc. It also pushes less far into the golden corner for patient 2 and 10, compared to the HSI as third objective. Furthermore, it is not able to create treament plans with HSI value lower than 0.5 cc for patient 4 and 9.

Dwell Length and Distance Modulation

In this last subsection the results regarding the formation of hotspots with the extension of the DLDM as the third objective will be discussed. In Figure 7.7, the results are shown in a graphical manner.

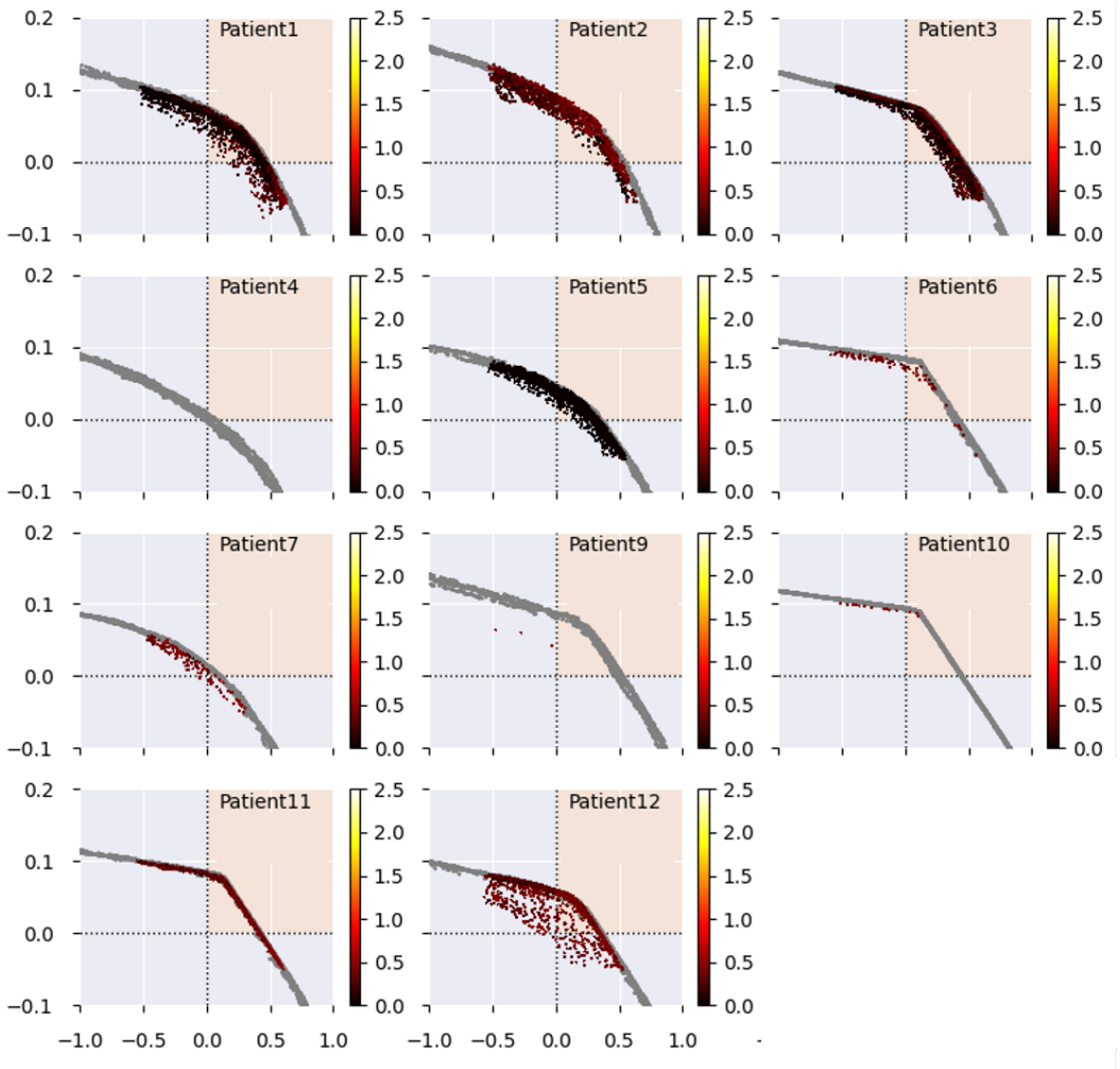

Figure 7.7: Pareto approximation fronts for the different patients resulting from the BRIGHT configuration with DLDM as third objective. Only treatment plans with an HSI value lower than 0.5 HSI value are shown. The grey Pareto approximation front is from the bi-objective optimization, which is not filtered on HSI value.

From the Pareto approximation fronts it can be seen that the DLDM is less able to mitigate hotspots as compared to the other adaptations, although it performs a little better than HSI for treatment plans under the 0.25 cc What can furthermore be seen is that the trade-off between the DLDM and the LCI and LSI is less strong as compared to the sum of extra added V indices and the DTMR. For patients 9 and 10, the DLDM was not able to produce treatment plans with an HSI value lower than 0.5 in the golden corner (save a few in the top for patient 10), whereas for patient 10 this was possible for the other metrics and for patient 9 this was possible for the HSI and the sum of extra added V indices. It

does improve slightly over the basic BRIGHT configuration.

LCI, LSI and L metrics for different hotspot bounds

In Table 7.12, the achieved values are shown for the LCI, LSI and L value per patient for different hotspot upper bounds. Only a selection of the results is shown. For every patient the results are available with bound 0, 0.25, 0.5, 0.75, 1, 1.5, 2 and 2.5, which are included in Appendix B.9. The selection is chosen so to be able to show the most important trends. Furthermore, for every patient, the first upper bound is included for which basic BRIGHT resulted in treatment plans and the first upper bound for which the eventual best LCI, LSI and L value are achieved. The full results can be found in the appendix.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient1	0	N/A[2]	0.095[3]	0.092	0.095	0.100	N/A[2]	-0.237[3]	**0.420**	**0.423**	**0.434**	N/A[2]	-0.237[3]	**0.532**	**0.535**	**0.434**
Patient1	0.25	0.103	0.101	**0.092**	**0.095**	0.100	0.455	0.458	0.428	0.441	0.447	0.535	0.518	0.541	0.548	0.545
Patient2	0	N/A[0]	N/A[0]	**0.117**	**0.116**	**0.121**	N/A[0]	N/A[0]	**0.405**	**0.427**	**0.274**	N/A[0]	N/A[0]	**0.466**	**0.525**	**0.550**
Patient2	0.25	N/A[1]	N/A[2]	0.117	0.116	0.122	N/A[1]	N/A[2]	**0.405**	**0.427**	**0.433**	N/A[1]	N/A[2]	**0.466**	**0.535**	**0.577**
Patient2	1.5	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.544	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	**0.635**
Patient3	0	0.100[3]	0.097	0.101	0.100	0.100	0.145[3]	0.306	**0.424**	**0.420**	**0.422**	0.145[3]	0.441	**0.565**	**0.577**	**0.597**
Patient3	0.25	0.100	0.102	0.101	0.100	0.102	0.347	**0.450**	0.426	**0.436**	0.428	0.592	0.621	0.565	0.581	0.604
Patient3	0.75	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient4	0.5	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]
Patient4	0.75	N/A[0]	**0.028**	0.007[3]	N/A[0]	N/A[0]	N/A[0]	**-0.440**	-0.482[3]	N/A[0]	N/A[0]	N/A[0]	**-0.369**	-0.257[3]	N/A[0]	N/A[0]
Patient4	1	N/A[0]	**0.050**	**0.043[4]**	N/A[0]	N/A[0]	N/A[0]	**-0.124**	-0.204[4]	N/A[0]	N/A[0]	N/A[0]	**-0.029**	**-0.032[4]**	N/A[0]	N/A[0]
Patient4	1.5	N/A[1]	**0.054**	0.048	N/A[1]	0.041[3]	N/A[1]	**-0.023**	**-0.079**	N/A[1]	-0.428[3]	N/A[1]	**0.270**	**0.221**	N/A[1]	-0.428[3]
Patient4	2	0.053	0.055	0.048	0.056	0.046	-0.014	0.002	-0.020	0.052	-0.062	0.032	**0.301**	**0.272**	0.072	0.198
Patient4	100	0.053	0.055	0.048	0.056	0.046	0.003	0.002	-0.020	0.067	-0.053	0.343	**0.304**	**0.272**	0.316	**0.273**
Patient5	0	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.317	0.550	0.504	**0.423**	0.443	0.516
Patient5	0.25	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient6	0	N/A[0]	N/A[1]	**0.085**	0.075[3]	N/A[0]	N/A[0]	N/A[1]	**0.323**	0.068[3]	N/A[0]	N/A[0]	N/A[1]	**0.475**	0.231[3]	N/A[0]
Patient6	0.25	N/A[1]	N/A[2]	**0.090**	0.084[3]	N/A[1]	N/A[1]	N/A[2]	**0.351**	0.324[3]	N/A[1]	N/A[1]	N/A[2]	**0.489**	0.419[3]	N/A[1]
Patient6	0.5	0.095[4]	0.097	0.091	0.095	0.091[4]	0.031[4]	**0.398**	**0.369**	**0.361**	0.215[4]	0.031[4]	**0.582**	**0.490**	**0.523**	0.325[4]
Patient6	0.75	0.096[4]	0.097	0.093	0.095	0.094	0.391[4]	0.398	0.372	0.383	0.359	0.435[4]	0.595	0.490	0.558	0.531
Patient6	1.5	0.096	0.097	**0.093**	0.095	0.095	0.411	0.408	**0.384**	0.397	0.400	0.598	0.595	**0.490**	**0.581**	**0.564**
Patient7	0	N/A[0]	N/A[2]	**0.040**	**0.050[4]**	N/A[0]	N/A[0]	N/A[2]	**-0.108**	**-0.199[4]**	N/A[0]	N/A[0]	N/A[2]	**0.033**	**0.076[4]**	N/A[0]
Patient7	0.25	N/A[0]	**0.062**	**0.058**	**0.055**	N/A[2]	N/A[0]	**-0.098**	**-0.026**	**-0.048**	N/A[2]	N/A[0]	**0.253**	**0.210**	**0.120**	N/A[2]
Patient7	0.5	N/A[2]	0.062	0.058	0.061	0.053	N/A[2]	**0.073**	**-0.026**	**-0.026**	**-0.035**	N/A[2]	**0.306**	0.211	**0.282**	**0.259**
Patient7	0.75	0.053[3]	0.062	0.060	0.062	0.058	-0.378[3]	**0.082**	**-0.012**	**0.021**	**0.032**	-0.378[3]	**0.306**	0.211	**0.282**	**0.302**
Patient7	1	0.062	0.063	0.060	0.062	0.060	-0.007	0.082	0.004	0.034	0.032	0.163	0.313	0.211	0.282	0.308
Patient7	1.5	0.063	0.064	**0.061**	0.062	**0.060**	0.097	0.082	0.004	0.048	0.079	0.314	0.314	0.211	0.283	0.308
Patient7	2	0.063	0.064	**0.061**	0.062	**0.061**	0.097	0.091	0.004	0.048	0.079	0.339	0.314	**0.211**	0.283	0.320
Patient9	0	N/A[0]	N/A[2]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A	N/A	N/A[0]	N/A[0]	N/A[0]	N/A[2]	N/A[0]	N/A[0]	N/A[0]
Patient9	0.25	N/A[0]	0.054[3]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	-0.030[3]	N/A	N/A[0]	N/A[0]	N/A[0]	0.029[3]	N/A[0]	N/A[0]	N/A[0]
Patient9	0.5	N/A[0]	**0.093**	**0.089**	N/A[0]	N/A[1]	N/A[0]	**0.380**	**0.278**	N/A[0]	N/A[1]	N/A[0]	**0.506**	**0.497**	N/A[0]	N/A[1]
Patient9	0.75	N/A[0]	**0.096**	**0.095**	N/A[1]	N/A[2]	N/A[0]	**0.432**	**0.391**	N/A[1]	N/A[2]	N/A[0]	**0.562**	**0.593**	N/A[1]	N/A[2]
Patient9	1	N/A[0]	**0.098**	**0.095**	**0.091**	**0.092**	N/A[0]	**0.432**	**0.418**	**0.172**	N/A	N/A[0]	**0.562**	**0.606**	**0.172**	**0.236**
Patient9	1.5	0.112	**0.100**	**0.097**	0.101	0.105	0.396	0.432	0.440	0.442	0.454	0.396	0.568	**0.607**	**0.602**	0.569
Patient9	2	0.112	**0.100**	**0.097**	0.101	0.105	0.501	**0.436**	**0.440**	0.449	0.463	0.684	**0.597**	**0.607**	**0.623**	**0.571**
Patient10	0	N/A[0]	N/A[0]	**0.097**	N/A[2]	N/A[0]	N/A[0]	N/A[0]	**0.363**	N/A[2]	N/A[0]	N/A[0]	N/A[0]	**0.535**	N/A[2]	N/A[0]
Patient10	0.25	N/A[0]	0.038[3]	**0.100**	**0.091[4]**	N/A[0]	N/A[0]	-0.033[3]	**0.398**	**0.201[4]**	N/A[0]	N/A[0]	-0.033[3]	**0.569**	**0.302[4]**	N/A[0]
Patient10	0.5	N/A[0]	**0.104**	**0.101**	**0.097**	N/A[1]	N/A[0]	**0.379**	**0.408**	**0.345**	N/A[1]	N/A[0]	**0.435**	**0.575**	**0.469**	N/A[1]
Patient10	0.75	N/A[2]	0.104	0.102	0.102	0.097[4]	N/A[2]	**0.420**	**0.408**	**0.363**	0.112	N/A[2]	**0.529**	**0.585**	**0.549**	0.112[4]
Patient10	1	0.106	0.104	0.102	0.102	0.100	0.120	**0.425**	**0.413**	**0.383**	**0.230**	0.120	**0.611**	**0.585**	**0.549**	**0.230**
Patient10	100	0.106	**0.105**	**0.103**	**0.104**	**0.104**	0.440	0.436	0.423	0.433	0.434	0.636	**0.620**	**0.599**	**0.601**	**0.599**
Patient11	0	N/A[2]	0.069[3]	0.091	0.023[3]	N/A[2]	N/A[2]	0.188[3]	**0.359**	0.160[3]	N/A[2]	N/A[2]	0.188[3]	**0.542**	0.160[3]	N/A[2]
Patient11	0.25	0.098	0.099	0.091	0.094	0.097	0.341	0.353	0.366	0.304	0.341	0.469	0.482	**0.549**	0.319	0.351
Patient11	0.75	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.393	0.419	0.598	0.601	**0.549**	**0.536**	0.599
Patient12	0	N/A[0]	N/A[1]	**0.066**	N/A[2]	0.071[3]	N/A[0]	N/A[1]	**0.252**	N/A[2]	0.029[3]	N/A[0]	N/A[1]	**0.459**	N/A[2]	0.029[3]
Patient12	0.25	N/A[1]	0.042[4]	0.068	0.028	**0.078**	N/A[1]	0.324[4]	**0.308**	-0.163	**0.280**	N/A[1]	0.423[4]	**0.469**	**-0.074**	**0.280**
Patient12	0.5	N/A[1]	**0.081**	0.070	0.076	**0.078**	N/A[1]	**0.348**	0.308	0.286	0.336	N/A[1]	**0.527**	**0.469**	**0.424**	**0.473**
Patient12	0.75	0.076[4]	0.082	0.072	0.078	0.078	0.288	**0.359**	0.316	0.335	0.350	0.288	0.530	0.469	0.519	0.515
Patient12	1	0.078	0.083	0.072	0.078	0.078	0.341	**0.359**	**0.316**	0.335	0.350	0.454	0.530	0.478	0.519	0.515

Table 7.12: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

What can be seen in the table above is that for some patients, there is no need for reducing the

hotspots, i.e. patient 3 and 5. What furthermore can be seen is that in general, the addition of HSI and the added V indices perform the best when reducing hotspots. This performance becomes most clear in the best achieved LCI values. What can also be seen is that the HSI addition shows significant changes when going from 0.25 to 0.5 upper bound in HSI value in comparison to the other additions. When looking at the achieved HSI values in combination with LCI and LSI for the base case and the different additions, it can be observed that hotspots can be decreased whilst staying in the golden corner.

7.2. Time

In this section, the time performance of the different potential improvements are discussed. A distinction can be made between the dwell time oriented improvements and the dose calculation point oriented improvements. For latter more dose calculation points had to be sampled. This resulted in 40.000 points more during optimization as compared to the dwell time oriented improvements. This will lead to a reduction in time performance.

In Table 7.13, the amount of treatment plan evaluations per second is shown. In each of the columns, the results are shown for that particular adaptation of BRIGHT.

	Evaluations per second			
	HSI	V indices	DTMR	DLDM
Patient1	228.47	544.60	737.00	753.01
Patient2	237.93	547.49	757.55	740.95
Patient3	203.26	456.02	617.05	621.59
Patient4	164.90	554.59	736.25	724.16
Patient5	177.60	427.56	573.39	575.81
Patient6	160.64	522.41	688.09	682.78
Patient7	147.12	554.97	734.76	736.62
Patient9	59.96	469.05	622.09	625.17
Patient10	196.51	598.10	792.19	795.40
Patient11	146.20	489.10	654.28	662.20
Patient12	181.30	502.76	661.47	678.72

Table 7.13: Overview of the number of evaluations per second per patient per BRIGHT adaption.

What can be concluded from Table 7.13 is that HSI is the most computationally complex adaptation resulting in the lowest number of evaluations per second. This is then followed by the sum of the extra V indices, for which the time increase can be attributed to the increase in dose calculations points. Both dwell time oriented adjustments have the highest number of evaluations per second.

In Table 7.14, the time it took to reach the golden corner is reported. In this table the same trend is shown as with the number of evaluations per second with regards to increased time for the dose calculation point oriented adaptations. The only changes that can be seen is that the ratio between evaluations per second and the time to the golden corner is not equal for each patient. This is due to the fact that some patients require more evaluations than others to reach the golden corner. The required number of evaluations is shown in Table 7.15.

	Time to golden corner			
	HSI	V indices	DTMR	DLDM
Patient1	221.25	80.55	64.36	62.06
Patient2	267.91	73.67	50.04	64.70
Patient3	298.98	78.02	63.19	74.79
Patient4	1148.46	345.90	288.18	342.58
Patient5	424.00	153.50	112.89	138.98
Patient6	234.52	83.93	63.53	54.09
Patient7	663.28	262.54	279.54	256.02
Patient9	874.80	106.16	86.07	88.18
Patient10	184.79	70.97	48.97	48.04
Patient11	250.31	86.26	66.14	65.63
Patient12	362.57	128.80	105.53	102.41

Table 7.14: Overview of the time to golden corner per patient per BRIGHT adaption.

	Evaluations to golden corner			
	HSI	V indices	DTMR	DLDM
Patient1	49954.60	39431.80	45011.80	47167.20
Patient2	64175.80	35011.00	35375.80	49833.60
Patient3	61097.00	32402.80	37655.40	47511.80
Patient4	188961.40	198458.40	214514.80	251834.60
Patient5	75592.60	66578.40	63000.40	81981.40
Patient6	39059.60	41670.00	41342.40	37301.60
Patient7	98793.80	149564.20	208653.60	187688.40
Patient9	53560.20	49600.60	51472.40	54033.00
Patient10	37184.00	40580.80	36150.60	38861.80
Patient11	37041.60	40630.60	40950.00	46195.40
Patient12	66045.40	65763.20	69035.20	70549.00

Table 7.15: Overview of the number of evaluations to golden corner per patient per BRIGHT adaption.

The last time metric is the 'convergence' times per implementation per patient. The convergence times have been determined visually by inspecting the 2D plots of the x- and y-axis combination of LSI

and LCI, LSI and third objective, and LCI and third objective. An optimization run is deemed 'converged' when no meaningful improvement is observed within a period of time. In Table 7.16, the convergence times per patient are shown. The times are rounded up to the nearest 50 seconds, as the Pareto approximation fronts are outputted irregularly and since visual inspection is not an exact method so a conservative estimation is chosen. For the HSI, the times are rounded to the nearest 100 seconds as the interval between outputted Pareto approximation fronts is longer.

	Time to convergence			
	HSI	V-indices	DTMR	DLDM
Patient1	800	350	300	200
Patient2	1200	350	300	250
Patient3	800	400	300	200
Patient4	1200	500	300	300
Patient5	1400	600	500	350
Patient6	1100	350	300	200
Patient7	1000	250	550	300
Patient9	2600	300	350	250
Patient10	800	250	300	150
Patient11	1000	350	400	200
Patient12	1500	300	400	250

Table 7.16: Overview of the convergence time in seconds per patient per BRIGHT adaption.

What can be seen in the table is that the HSI is, as expected, the slowest implementation, followed by the V-indices, DTMR and DLDM in that order. What should be noted here is that for the HSI and V-indices extra DCPs were added (40.000) during optimization to capture the normal-tissue values.

For patient 9, the run time was significantly longer than that of the other patients. This is caused by the significantly lower number of evaluations per second for patient 9. The explanation for this lower number of evaluations per second is the significantly higher maximum degree in the graph of DCPs. The maximum degree for patient 9 is on average 171, where the average degree of the other patients is 34. Having a higher degree requires more rounds in the hotspot registration algorithm for reducing labels. The higher degree is caused by the fact that patient 9 has a larger than average delineated prostate (64 cc) and smaller than average delineated seminal vesicles (1.3 cc). The ratio between them is 50, where the average rating for the other patients is 9. This ratio is of importance for the construction of the graph, as the maximum edge length for two DCPs from different organs/volumes is determined by the maximum of both organ maximum edge lengths. The maximum edge length is dependent on the volume per DCP, which is determined by the volume of the organ and the amount of DCPs per organ (which is the same for each organ). If one of them is significantly larger (i.e. 50 times), then a lot of edges are formed between a single DCP from the larger organ to DCPs from the smaller organ.

7.2.1. Robustness

Homogeneity does not only include the presence of hotspots. The other motivation for improving homogeneity is the robustness of treatment plans. In this research, this has not been the main motivation. A complete study using disturbance simulation would be required to draw conclusions regarding the improvement or deterioration of robustness by the different adaptions of BRIGHT. Nonetheless, in this section a small insight will be provided by calculating the sum of DTMR absolute value of the treatment plans, as described in Balvert et al., 2014. This DTMR absolute value is the absolute deviation between dwell times of 2 neighbouring neighbouring dwell positions. In their papers, the authors showed a correlation between the value of DTMR absolute and the robustness of the treatment plan, although using the DTMR absolute did deteriorate other treatment plan features.

In Table 7.17, the results are shown for the DTMR absolute values for the 2D Pareto approximation front for the basic BRIGHT settings are presented:

	min	mean
Patient1	1744.2	1874.1
Patient2	1593.4	1593.4
Patient3	1616.2	1813.7
Patient4	2154.2	2302.0
Patient5	2285.4	2570.9
Patient6	2363.1	2609.4
Patient7	1792.2	1932.1
Patient9	3421.5	3777.9
Patient10	2922.4	3240.5
Patient11	1546.5	1703.4
Patient12	2867.2	3094.6

Table 7.17: Overview of the average and minimum DTMR-A values of the different BRIGHT adjustments for the 2D Pareto approximation front of the basic BRIGHT configuration.

In Table 7.18, the results for the DTMR absolute values for the 2D Pareto approximation fronts of the different BRIGHT adaptations:

	HSI		V indices		DTMR		DLDM	
	min	mean	min	mean	min	mean	min	mean
Patient1	1844.9	2021.8	1647.5	1843.7	1480.6	1768.5	1710.8	1896.2
Patient2	1642.0	1801.6	1330.0	1479.3	1344.5	1553.3	1520.1	1790.9
Patient3	1597.6	1784.1	1481.7	1694.1	1354.8	1684.4	1546.3	1842.1
Patient4	2062.3	2221.6	2037.0	2147.4	2029.4	2190.6	2113.0	2272.8
Patient5	2438.6	2663.0	2066.4	2314.4	1913.8	2343.8	2321.8	2608.6
Patient6	2402.2	2642.6	2140.1	2417.8	2210.7	2551.9	2361.1	2647.4
Patient7	1836.1	2028.0	1741.5	1906.1	1640.1	1852.2	1819.4	2010.5
Patient9	3257.8	3552.0	3017.2	3358.0	2995.1	3498.8	3262.4	3574.8
Patient10	2927.6	3293.4	2698.2	3086.9	2631.4	3229.3	2861.5	3354.4
Patient11	1546.5	1702.1	1416.2	1579.8	1290.8	1585.9	1490.7	1743.4
Patient12	2884.7	3130.2	2691.5	2882.7	2676.8	2981.5	2801.0	3047.1

Table 7.18: Overview of the achieved minimum and average DTMR-A values of the treatment plans resulting from the different BRIGHT adjustments for the 2D Pareto approximation front.

Next to the 2D Pareto approximation front, the 5% Pareto approximation fronts are calculated and the deviation with the 2D Pareto approximation fronts are shown in Table 7.19.

	HSI		V indices		DTMR		DLDM	
	min	mean	min	mean	min	mean	min	mean
Patient1	1842.4	2034.2	1639.2	1837.7	1365.4	1728.3	1658.4	1879.0
Patient2	1638.9	1811.4	1287.6	1457.0	1212.9	1512.1	1464.7	1791.8
Patient3	1597.6	1787.7	1410.8	1667.6	1158.1	1576.2	1453.7	1811.5
Patient4	2054.8	2213.4	2001.5	2135.8	1991.5	2180.1	2106.7	2259.5
Patient5	2433.1	2674.8	2045.9	2290.6	1813.6	2290.0	2234.6	2623.9
Patient6	2344.0	2623.8	2078.2	2374.7	1977.8	2424.5	2239.6	2586.6
Patient7	1802.7	2007.2	1672.8	1880.6	1468.8	1802.7	1753.6	1985.4
Patient9	3153.3	3522.6	2971.6	3344.2	2857.5	3403.8	3109.6	3545.1
Patient10	2872.2	3258.2	2531.2	3013.1	2344.9	3026.3	2654.2	3286.6
Patient11	1531.6	1699.1	1385.0	1565.9	1140.6	1509.7	1477.8	1732.0
Patient12	2862.5	3118.8	2595.9	2860.5	2540.1	2913.5	2764.8	3035.6

Table 7.19: Overview of the achieved minimum and average DTMR-A values of the treatment plans resulting from the different BRIGHT adjustments for the 2D-5% Pareto approximation front.

Based on these tables, the following can be observed. Optimizing on the HSI does not deteriorate the performance on the DTMR-A value and thereby potentially the overall robustness. When optimizing on the extra added V indices, the DTMR-A value can even be improved, the same holds for the DLDM. Optimizing directly upon the DTMR obviously results in the best performance on the DTMR-A value.

What can furthermore be seen is that, when optimizing on the extra added V indices, the DTMR value only deviates slightly.

When looking at the 5% Pareto approximation front, it can be seen that the DTMR value overall does not increase or decrease significantly from the normal Pareto approximation front.

7.2.2. Key outcomes

With this explication of the experiment results the second part of the following research questions can be answered by the general observations in the experiments:

Research Question: BRIGHT performance

What are the important performance aspects of BRIGHT for HDR-BT ***and how do different enhancements score on these aspects?***

What can be seen in the results is that optimizing the HSI does not introduce a new trade-off with the LCI and LSI for most patients. Which means that hotspots can be reduced without deteriorating the DVIs. This observation is strengthened by the non-perfect correlation of the LCI and LSI with the HSI. Furthermore, using the HSI or the sum of extra V indices used as third objective show best performance when reducing hotspots to at most 0.5 cc. The sum of extra V indices outperforms the HSI when looking at reducing the HSI to under 0.25 cc. The sum of extra V indices does introduce a new trade-off with the LCI and LSI. The DTMR also showed a reduction in hotspot sizes but less than the HSI and the extra added V indices and also more inconsistent. For treatment plans with HSI below 0.25 cc, the DTMR outperforms the HSI for all patients, except patient 4 and 9. The performance of the DLDM was inferior to the other metrics. The DLDM did introduce a trade-off with the LCI and LSI, but less compared to the DTMR and sum of extra V indices. The sum of extra added V indices shows a slight deterioration in achieved DVI values, where this is not the case for the HSI. Lastly, the required time for HSI is currently significantly longer than that of the extra added V indices. However, it must be noted that the HSI can still be optimized further for computational time. The most important finding is that BRIGHT is able to reduce or even remove hotspots, without deteriorating the other objectives, when configured to do so.

8

Discussion

In this chapter, the results from Chapter 7 and their implications will be discussed, and the limitations of this study will be explicated.

8.1. Improving homogeneity

8.1.1. Required changes to treatment plans

From the changes applied to the exported BRIGHT plans in the clinic (shown in Table 7.8), it can be seen that BRIGHT does not fully satisfy the preferences of the clinical experts yet. This inability can be explained by the fact that BRIGHT simply was not configured to optimize on the specific requirements for a particular patient, but simply finds the best treatment plans given the clinical protocol. In this clinical protocol, there is no explicit formulation of the desire to reduce hotspots. Under the clinical protocol several different treatment plans can be constructed that adhere to the protocol in the same way, but show different dose distributions. These different dose distributions can show differences in the presence of hotspots as has been shown in the experiments (i.e. Figure 7.4). If configured to focus on reducing these hotspot, BRIGHT is able to produce treatment plans that have reduced hotspots but similar LCI and LSI values.

8.1.2. BRIGHT adaptation types

In the changes applied in the clinic and in discussions with the clinical experts it became clear that the reduction of hotspots is not the main priority when constructing a treatment plan. Adhering to the clinical protocol is the foremost objective, after which the dose distribution itself is inspected to focus on secondary treatment plan features such as the hotspots. Given this inherent order of preferences and the diversity of achievable hotspot sizes between patients it was decided not to use hotspots as a constraint or as a sub-function of either LCI or LSI. If it was added as a sub-function or constraint it could either hamper the optimization of the clinical protocol or fail to address the reduction of hotspots properly. The reason for this is the diversity in achievable reduction in hotspot sizes between patients. To add a sub-function or constraint, an aspiration value or cut-off value has to be set. If this value is set too ambitious, then BRIGHT will put too much focus on adhering to this constraint or to reduce the achieved sub-function value. This will lead to BRIGHT not being able to reach the golden corner and/or BRIGHT neglecting the DVIs from the clinical protocol. If the aspiration or cut-off value is set too unambitious, then BRIGHT can easily adhere to them and fail to focus on reducing the hotspots further whereas that would have been possible. Given the diversity in achievable hotpot sizes between patients, that have been shown in the experiments, there is no single aspiration or cut-off value that would suit every patient. A potential solution to this could be to have an adaptive aspiration or cut-off value in a similar fashion as to how adaptive steering was constructed. This however is left for future research.

8.1.3. Treatment plan features to steer on

HSI

When looking at the different potential metrics for BRIGHT to steer on, the results have shown that both the approximate HSI and the extra V indices are best suited. The HSI on 140.000 DCPs showed high correlation with the HSI on +700.000 DCPs. Optimizing HSI during optimization resulted in the best overall improvement of HSI without LCI or LSI deterioration seen in the experiments. Compared to the sum of extra V indices, the ability to eradicate all hotspots is slightly worse. This behaviour could potentially be explained in two ways. The first one is that in the HSI objective there is a lower bound on the hotspot volume, which makes the function non-continuous. During optimization, if the presence of a small hotspot is advantageous for certain DVI values, it can be beneficial for BRIGHT to keep them. If these hotspots are just below the lower bound volume, then they offer no contribution to the HSI objective. As was seen in the comparison between HSI on 140.000 points and 700.000+ points, there was some discrepancy. This could mean that hotspots with a volume below the lower bound in optimization are accounted for in the reevaluation on 700.000+ points. This can be seen as over-fitting on the lower resolution of DCPs. This problem can be mitigated by setting a lower lower bound during optimization then during reevaluation. The second explanation why BRIGHT with the HSI as third objective has more difficulty to reduce the hotspots completely is that the presented problem is more difficult to explore using a population based EA. Due to the introduced lower bound, the search space is non-continuous and therefore the trajectories of improvement are non-smooth. Due to this non-smooth trajectory it is more difficult to learn for an population based EA. This is caused by the fact that perturbations of solutions with hotspots lower than the hotspot lower bound will not lead to a change in HSI value. This leads to an inability to learn. This second explanation is strengthened by the observation that the problem at hand does not come into play for patients 4 and 9, where the Pareto approximation front for HSI was found to be a broader 3D surface rather than a thicker line in 2D space. The advantage of having the lower bound on hotspots volume in comparison to the extra added V indices is that a 'safe haven' is created for the hotspots. One of the strong-suits of brachytherapy is the use of small hotspots. The lower bound allows for these small hotspots, whereas the extra added V indices will also register small hotspots.

Sum of extra V indices

The extra added V indices showed high correlations with the HSI value but for some patients there was a sudden drop in correlation. This drop in correlation can be explained by the fact that the extra added V indices do not directly capture the presence of hotspots. Having a higher volume that receives 300% of the prescribed dose (or 200% for the normal-tissue) does not mean that this dose is clumped together. For the patients where a low correlation was observed, the volumes that received 200/300% of the prescribed dose were likely more scattered. This shows that to capture the hotspots, there is a need for a more sophisticated metric than the sum of randomly scattered points that receive a high dosage. Nonetheless, by using the added V indices as a third objective to steer on, treatment plans were improved in terms of HSI. The downside to the added V indices as compared to HSI is the increased spread of treatment plans over the objective space in terms of LCI and LSI. This is caused by the conflicting nature of the objectives, and therefore a newly introduced trade-off. This increase in Pareto approximation front size/dimensionality resulted in a decrease in pressure towards the Pareto approximation front which lead to a slightly reduced performance in achieved LCI and LSI as compared to adding the HSI as third objective. This is a problem that should be able to be mitigated through different settings of BRIGHT and introducing adaptive steering for the third objective. The extra added V indices as third objective might have a potential benefit over the HSI when it comes to robustness. For the extra added V indices it is more beneficial to have a lot of shorter dwell times than fewer longer ones. This is due to the non-linear dose rate fall-off, as shown in Figure 4.4. This could potentially mean that dwell times will be more spread out. In the results, it was shown that the V indices achieved lower values on the absolute DMTR as compared to the HSI. For assessment of robustness, more is required than simply the absolute DTMR. Nonetheless, this result might hint towards a potential robustness benefit.

DTMR

Adding the DTMR as third objective did show improvements in HSI value for some patients, but the performance was inconsistent over the patient set. This can be explained by the fact that it does not directly focus on the hotspots. The experiments showed a low correlation between DTMR and HSI value.

DLDM

Using the DLDM as metric to mitigate the formation of hotspots showed a similar inconsistent performance as the DTMR, but with worse performance compared to the DTMR. This means that the DLDM did not accurately capture the formation reasons of hotspots. An explanation for this can potentially be found in the hotspots that are formed in the normal-tissue and more affected by higher dwell times. The upper bound on dwell time in the DLDM should potentially be dependent on the distance to the normal-tissue.

8.1.4. Cost of adaptations

From a practical point of view, the time requirements should also be taken into account. The computational advantage of using dwell times rather than DCPs clearly shows in the time required to reach the golden corner. Although the number of evaluations required to reach the golden corner did not differ much between the extra added V indices and the DTMR, the DTMR outperformed the added V indices in the amount of evaluations per second and therefore in the time to reach the golden corner. This computational advantage can be explained by the fact that the dwell time oriented adaptations did not require the sampling of DCPs in the normal tissue surrounding the target organs. These extra added DCPs decrease the performance of BRIGHT. From the DCP oriented adaptation, the HSI was the slowest implementation. This can be explained by the fact that for the extra added V indices, the only computational overhead is the sorting of the DCPs on received dose. This was already done for the DCPs in the target volumes and thus only the sorting of the dose calculations points in the normal tissue added to the computational complexity. Next to that, the received dose of these extra DCPs also needed to be calculated. The computational complexity for the HSI was larger, since it required multiple union finds per DCP which are done sequentially, which is dependant on the number of edges it had and the length of the path it needed to traverse. Furthermore, the number of computational steps in the HSI is larger than that of the other implementations.

8.1.5. Robustness

To study the robustness of the treatment plans, the DTMR values were calculated as they have been shown to have some correlation with the robustness (Balvert et al., 2014). Further research has to be performed to draw significant and meaningful conclusions about the robustness of the treatment plans. Nonetheless, what was observed in the DTMR scores was that by applying the added V indices as the third objective, the DTMR values were also reduced, almost to similar DTMR values when optimized on them with a third objective. This meant that using the extra added V indices improved upon the original treatment plans for the DTMR values. The introduction of the HSI as a third objective did not necessarily increase or decrease the DTMR values as compared to the treatment plans from the basic BRIGHT settings.

8.2. Limitations

This research has been conducted based on a data set of 11 patients. This number is not sufficient to confidently say that the performance of the created adaptations reasonably generalizes to all potential patients. To be able to do this a larger number of patients would be required. Nonetheless, statements like 'treatment plans with equal DVI values but different dose distributions which either do or do not show hotspots can be constructed' are valid to make on the basis of this data set.

In this master thesis, the number of experiments that could be conducted was limited due to time constraints. Therefore decisions had to be made on which experiments to conduct and for which questions an assumption had to be made or a default value had to be chosen. This resulted in for instance choosing default values for the population size, the number of cluster components and the combination of adaptive steering parameters tested. It could very well be the case that performance could be improved when these values were chosen differently. Nevertheless, based upon the achieved results it could already be seen that improved performance is achieved. This means that changing these values would not result in different conclusions, but could potentially only improve and fortify the current ones.

In this research, a definition for hotspots has been set after discussion with the clinical experts. This definition is not based upon clinical research, but only on the experience of the experts. This means that other institutes can have different views on when a hotspot is undesirable. Furthermore, this also means that hard bounds have been set on when a region is considered a hotspot. If a region just

barely stays below these bounds, then it is considered equally desirable to when this same regions would stay far from these bounds. From a clinical perspective this is of course not the case. However, some bounds had to be set. To combat this potential shortcoming the results have also been evaluated when the lower bound for hotspot volume was increased from 0.1 cc to 0.15 cc as this could easily be done using the already produced results. The downside to this approach is that the results were not explicitly optimized for this bound and therefore the conclusions using this increased lower bound are only partly valid. Reducing the lower bound was not deemed necessary as the lower bound is already low from the clinical perspective.

The locations of hotspots were not explicitly accounted for in this research. Given the fact that hotspots are now registered, it could be possible to take the location of the hotspots into account. In the discussion with the clinical experts, they expressed that this could be a good idea as they have preferences regarding the location of the hotspot. This however has been left open for future research. The consequence of this is that in the presented research, treatment plans with equally sized hotspots are deemed equally desirable irrespective of the locations of their hotspots. The location of the hotspot however, can cause one treatment plan to be more preferable than the other.

9

Conclusion and Future Work

In this research a number of questions regarding the homogeneity of treatment plans produced by BRIGHT were addressed. Several different adaptations of BRIGHT were constructed and tested to evaluate their potential improvements. In this chapter, conclusions will be drawn based on these experiments.

9.1. Conclusion

In the experiments it has been shown that hotspot sizes can be decreased whilst maintaining equal LCI and LSI values for most patients. This means that the correlations of both LCI and LSI with the hotspots are low. This blind spot in the clinical protocol leads to a less informed choice when selecting treatment plans from the Pareto approximation front with regards to the hotspots when the clinical experts are only presented with the LSI, LCI, and the achieved values for their sub-functions. The HSI, as described in this research, could help the clinical experts in making a more informed choice on which treatment plan to pick for further inspection.

When optimizing for hotspots, the best pick from a quality point of view would be to steer directly on the HSI. However, the HSI had the worst time complexity and could greatly increase the required run-time in the clinic. Further optimization of the code would be required to make the implementation faster. Discussions with the clinical experts would be required to get a time requirement based upon the potential achieved improvements of the treatment plans. This would also require a clinical observer study.

The other option would be to add the sum of extra added V indices as a third objective, as this has also shown to reduce the hotspots. The sum of extra V indices did have a drop in performance for 2 patient. The other downside of this method is that is does not reach as deep into the golden corner as the HSI. Further improvements would be required to overcome this. A potential benefit could be that steering upon the extra added V indices would also increase the robustness of the treatment plans. This does however require further research to make this claim.

The tested dwell time oriented adaptations yielded less consistent improvements from a hotspot point of view.

With these acquired insights the problem statement of this thesis can now be answered:

Research Question: Problem statement

How can the homogeneity of the resulting treatment plans from BRIGHT for high dose rate brachytherapy for prostate cancer be improved?

The homogeneity of the resulting treatment plans can be improved by adding a third objective to BRIGHT. Two options for this objective that lead to the best decrease in hotspot volume were a graph-based hotspot registration method and sum of the $V_{300\%}$ for target volumes and $V_{200\%}$ for the normal-tissue. Improving the homogeneity can be done without compromising the DVIs. Improving homogeneity does come at the cost of increased run-time. For the sum of volume indices the required time was around 600 seconds, and for the hotspot registration this was around 1800 seconds. Further optimization of the hotspot registration method is required to definitively assess run-time impact.

9.2. Future Work

9.2.1. Publications

Based on this master thesis a conference abstract has been written and submitted for ESTRO 2022. The ESTRO conference is focused on advancements in radiation oncology. The abstract that was submitted can be found in Appendix A.

Besides the conference abstract, a journal paper is in the making. This journal paper will be written for Medical Physics, a global journal on imaging science and engineering research that is focused on therapy and patient diagnosis.

9.2.2. BRIGHT improvements

ϵ-Pareto approximation front

The first idea for potential BRIGHT improvements is to reduce the search space in the third objective. This idea came to mind when looking at the results for using the HSI as a third objective. For patient 9, there was a broader 3D surface visible rather than a thicker line for the Pareto approximation front (see Figure 7.4). The consequence of this is a reduced pressure towards the golden corner. In practice, it is shown that clinical experts are foremost interested in the LCI and LSI values and are only interested in reducing hotspots once these values have been deemed sufficient. Given this fact, an idea would be to only consider treatment plans with improved hotspot sizes when they are not too far from the 2D Pareto approximation front. This could potentially be realized with a non-linear version of adaptive steering, where the adaptive bounds are a non-linear combination of LCI and LSI.

Dwell time precision

Another idea for improvement is to generally reduce the search space. Currently, the dwell times are stored (and sampled) as floating point numbers. This means that a dwell time has a time precision of 6 to 7 decimal places (depending on the compiler). These slight changes in dwell times could also mean a slight change in the trade-off being made between LCI, LSI and HSI. This could result in a non-dominating Pareto relationship between treatment plans. The question is whether that is meaningful since plans will have almost identical dose distributions.

A solution to this could be to change the precision with which either the dwell times are sampled or with which the DVIs and HSI values are stored. If the DVI values are stored with a lower precision than in the Pareto dominance relation, the two aforementioned treatment plans would be considered equal and only one of them would be stored. By decreasing the precision of the dwell times or DVIs and HSI values one would then reduce the amount of solutions used to approximate the Pareto approximation front. This way the memory usage will go down, the required population size can potentially be decreased and consequently BRIGHTs performance might be increased.

The downside of this approach could be the loss in continuity of the solution space. This would be more difficult for BRIGHT to traverse as perturbations of the treatment plans will be larger, which could lead to a decrease in performance.

Postponed third objective

The idea of restarting the optimization with a third objective is focused on improving run-time. With the current set-up for the third objective, it takes significantly longer to reach the golden corner than with the bi-objective problem formulation. Given that clinical experts are only interested in the plans in and surrounding the golden corner the idea would be to first solve the problem as a bi-objective problem with LCI and LSI as the objectives. Once the golden corner is sufficiently reached, only then start looking for plans with better HSI values. A potential problem of this approach would be that the optimization can get stuck in local optima found in bi-objective optimization. This would result in a restart not being able to find better solutions for the HSI since it would be unable to escape from these local optima. To overcome this problem, the idea would be to (intelligently) generate some artificial noise over the dwell times to get out of these local optima. By first performing the optimization as a bi-objective problem, the optimization itself could be sped up, since the population size can be smaller which reduces the computational complexity. It would also mean that no hotspot related features have to be calculated when trying to push towards the golden corner, since this does not aid in reaching it. Some experiments have been performed during this thesis which did show a lot of potential, but this would need to be investigated more. It was deemed out of scope for this thesis and therefore discontinued.

Improving time complexity of hotspot detection

In this research a state-of-the-art method for hotspot detection has been used. Nonetheless, the implementation itself can still be sped up. Apart from the GPU memory optimization, there is also the possibility of partial evaluations. In the current set-up, the whole hotspot detection procedure is performed for every change of FOS-element. However, when only changing a few dwell times, it is not necessary to completely start the hotspot detection from scratch. The results from before the change could be used to calculate the new hotspots. This can be done by checking which dose calculation points have changed from potentially being in a hotspot to not being in a hotspot or vice versa. Then, in the hotspot detection, only for these dose calculation points and their direct neighbours, the labels are updated. This could potentially speed up the hotspot detection procedure.

9.2.3. Additional research

Clinical observer study

The results from this research have shown that it is possible to improve upon the treatment plans with regards to the HSI whilst minimizing the loss in DVI values, or even keeping them the same. However, in similar fashion as to the inability of BRIGHT to focus on the hotspots before this study, a similar dose distribution feature might also be overlooked. In the worst case, it might even be that by focusing on the HSI, another important aspect of the dose distribution is deteriorating. To see if the treatment plans that have been optimized for the HSI are significantly better than the previous treatments plans, a clinical observer study must be performed. It is also interesting to see if a slight deterioration of the DVIs is allowed to remove hotspots. This would then fortify the idea of adapting BRIGHT using ϵ-dominance, as discussed in Section 9.2.2

In this observer study, the clinical experts are shown treatment plans, from which they do not know which version of BRIGHT has produced them, and are asked which one is most preferable. By doing this, potential negative side effects of focusing on the HSI can be identified and the need for focusing on HSI can be confirmed.

Tumor control probability and cold spots

Steering hotspots to a specific location is possible now that the registration of hotspots has been established. This control can lead to increased tumor control probability. If the Gross Tumor Volume (GTV), is delineated in the medical images it should be possible to create hotspots in it. Hotspots in the GTV are linked to increased tumor control probability (Tomé and Fowler, 2002), i.e. the probability that the tumor is eradicated. A similar but opposite problem occurs for cold spots, contiguous low-dose sub-volumes in the GTV or total target volumes. If these cold spots are present that will lead to a reduced tumor control probability. Cold spots can be detected in a similar fashion as hotspots, and thereby their presence reduced in the GTV and CTV.

Robustness evaluation

As described in this research, there are two important aspects when it comes to treatment plan homogeneity. The reduction of hotspots and the increase of robustness. The main focus has been on the former. In the evaluation of the results, some metrics have been calculated and reported on, which are related to the robustness. However, to confidently draw conclusions on this, one would need to do simulations in which disturbances are added. A similar approach to this research could be adopted, namely finding proxy features that have a high correlation with robustness. This could then be added to the BRIGHT set-up.

Long-shot: eliminating the dose calculation points

One of the more daring ideas that came to mind when doing this research was eliminating all dose calculation points from the optimization. Since the dose values for the dose calculation points are the direct result of the dwell times it could potentially be possible to create a set of objectives based on the dwell times, their distance to each other and their distance to the boundaries of the involved organs. This set of objectives could then be adjusted to comply with the clinical protocols of the different clinics around the world. This is, however, a very difficult problem to tackle, hence why it is has not been done before. However, if successful, the optimization can be sped up significantly. The achieved DVIs can then afterwards always be calculated using dose calculation points if required.

Bibliography

Alderliesten, T., Bosman, P. A. N., & Bel, A. (2015). Getting the most out of additional guidance information in deformable image registration by leveraging multi-objective optimization. *Medical Imaging 2015: Image Processing*, *9413*, 94131R.

Allegretti, S., Bolelli, F., & Grana, C. (2019). Optimized block-based algorithms to label connected components on GPUs. *IEEE Transactions on Parallel and Distributed Systems*, *31*(2), 423–438.

Andrews, S., Hamarneh, G., & Saad, A. (2010). Fast random walker with priors using precomputation for interactive medical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9–16.

Balvert, M., Gorissen, B. L., Den Hertog, D., & Hoffmann, A. L. (2014). Dwell time modulation restrictions do not necessarily improve treatment plan quality for prostate hdr brachytherapy. *Physics in Medicine & Biology*, *60*(2), 537.

Barten, D. L. J., Bouter, A., van Wieringen, N., Pieters, B. R., Hinnen, K. A., Westerveld, H., Maree, S. C., van der Meer, M. C., Alderliesten, T., Bosman, P. A. N., Niatsetski, Y., & Bel, A. (2021). Artificial intelligence based planning of HDR prostate brachytherapy: First clinical experience. *Radiotherapy and Oncology*, *161*, S653–S655.

Bel, A., Niatsetski, Y., Alderliesten, T., Pieters, B. R., Bosman, P. A. N., et al. (2020). Robust evolutionary bi-objective optimization for prostate cancer treatment with high-dose-rate brachytherapy.

Bouter, A., Alderliesten, T., Bel, A., Witteveen, C., & Bosman, P. A. N. (2018). Large-scale parallelization of partial evaluations in evolutionary algorithms for real-world problems. *Proceedings of the Genetic and Evolutionary Computation Conference*, 1199–1206.

Bouter, A., Alderliesten, T., Pieters, B. R., Bel, A., Niatsetski, Y., & Bosman, P. A. N. (2019). GPU-accelerated bi-objective treatment planning for prostate high-dose-rate brachytherapy. *Medical physics*, *46*(9), 3776–3787.

Bouter, A., Luong, N. H., Witteveen, C., Alderliesten, T., & Bosman, P. A. N. (2017). The multi-objective real-valued gene-pool optimal mixing evolutionary algorithm. *Proceedings of the Genetic and Evolutionary Computation Conference*, 537–544.

De Jong, K. A. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* University of Michigan.

Deb, K. (2014). Multi-objective optimization. *Search methodologies* (pp. 403–449). Springer.

Deb, K., Saxena, D. et al. (2006). Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. *Proceedings of the world congress on computational intelligence (WCCI-2006)*, 3352–3360.

DiBiase, S. J., Wallner, K., Tralins, K., & Sutlief, S. (2000). Brachytherapy radiation doses to the neurovascular bundles. *International Journal of Radiation Oncology* Biology* Physics*, *46*(5), 1301–1307.

Elekta. (2021). *Brachytherapy applicator guide*.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International journal of cancer*, *144*(8), 1941–1953.

Golshan, M. A., Spadinger, I., Morris, W. J., Hamm, J., & Keyes, M. (2014). Regional dosimetric analysis of LDR prostate brachytherapy patients with confirmed local failure: A case-control study. *Brachytherapy*, *13*, S38.

Holland, J. H. et al. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.

Holm, Å., Larsson, T., & Carlsson Tedgren, Å. (2012). Impact of using linear optimization models in dose planning for HDR brachytherapy. *Medical physics*, *39*(2), 1021–1028.

Hong, C., Dhulipala, L., & Shun, J. (2020). Exploring the design space of static and incremental graph connectivity algorithms on GPUs. *Proceedings of the ACM International Conference on Parallel Architectures and Compilation Techniques*, 55–69.

Hoskin, P. J., Colombo, A., Henry, A., Niehoff, P., Hellebust, T. P., Siebert, F.-A., & Kovacs, G. (2013). GEC/ESTRO recommendations on high dose rate afterloading brachytherapy for localised prostate cancer: An update. *Radiotherapy and Oncology*, *107*(3), 325–332.

Karabis, A., Giannouli, S., & Baltas, D. (2005). 40 HIPO: A hybrid inverse treatment planning optimization algorithm in HDR brachytherapy. *Radiotherapy and Oncology*, *76*(2), S29.

Kataria, T., Sharma, K., Subramani, V., Karrthick, K., & Bisht, S. S. (2012). Homogeneity Index: An objective tool for assessment of conformal radiation treatments. *Journal of medical physicsm, Association of Medical Physicists of India*, *37*(4), 207.

Kim, D., Cunha, J. A. M., Beaulieu, L., & Hsu, I.-C. (2016). A topological analysis of brachytherapy dose distributions: The Bubble Index. *Brachytherapy*, *15*, S158.

Lahanas, M., Baltas, D., & Zamboglou, N. (2003). A hybrid evolutionary algorithm for multi-objective anatomy-based dose optimization in high-dose-rate brachytherapy. *Physics in Medicine & Biology*, *48*(3), 399.

Lemaitre, F., Hennequin, A., & Lacassagne, L. (2021). Taming voting algorithms on GPUs for an efficient connected component analysis algorithm. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7903–7907.

Lessard, E., & Pouliot, J. (2001). Inverse planning anatomy-based dose optimization for HDR brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function. *Medical physics*, *28*(5), 773–779.

Luong, N. H., Alderliesten, T., Bel, A., Niatsetski, Y., & Bosman, P. A. N. (2018). Application and benchmarking of multi-objective evolutionary algorithms on high-dose-rate brachytherapy planning for prostate cancer treatment. *Swarm and Evolutionary Computation*, *40*, 37–52.

Luong, N. H., Bouter, A., Van Der Meer, M. C., Niatsetski, Y., Witteveen, C., Bel, A., Alderliesten, T., & Bosman, P. A. N. (2017). Efficient, effective, and insightful tackling of the high-dose-rate brachytherapy treatment planning problem for prostate cancer using evolutionary multi-objective optimization algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 1372–1379.

Maree, S. C., Luong, N. H., Kooreman, E. S., van Wieringen, N., Bel, A., Hinnen, K. A., Westerveld, H., Pieters, B. R., Bosman, P. A. N., & Alderliesten, T. (2019). Evaluation of bi-objective treatment planning for high-dose-rate prostate brachytherapy—a retrospective observer study. *Brachytherapy*, *18*(3), 396–403.

Morén, B., Larsson, T., & Tedgren, Å. C. (2018a). Mathematical optimization of high dose-rate brachytherapy, derivation of a linear penalty model from a dose-volume model. *Physics in Medicine & Biology*, *63*(6), 065011.

Morén, B., Larsson, T., & Tedgren, Å. C. (2018b). Preventing hot spots in high dose-rate brachytherapy. *Operations research proceedings 2017* (pp. 369–375). Springer.

Niemierko, A., & Goitein, M. (1990). Random sampling for evaluating treatment plans. *Medical physics*, *17*(5), 753–762.

Patel, G., Mandal, A., Choudhary, S., Mishra, R., & Shende, R. (2020). Plan evaluation indices: A journey of evolution. *Reports of Practical Oncology and Radiotherapy*, *25*(3), 336–344.

Rivard, M. J., Coursey, B. M., DeWerd, L. A., Hanson, W. F., Saiful Huq, M., Ibbott, G. S., Mitch, M. G., Nath, R., & Williamson, J. F. (2004). Update of AAPM task group no. 43 report: A revised AAPM protocol for brachytherapy dose calculations. *Medical physics*, *31*(3), 633–674.

Schröder, F. H., Hugosson, J., Roobol, M. J., Tammela, T. L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., et al. (2012). Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine*, *366*(11), 981–990.

Shiloach, Y., & Vishkin, U. (1980). *An O (log n) parallel connectivity algorithm* (tech. rep.). Computer Science Department, Technion.

Sutton, M., Ben-Nun, T., & Barak, A. (2018). Optimizing parallel graph connectivity computation via subgraph sampling. *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 12–21.

Thomas, C. W., Kruk, A., McGahan, C. E., Spadinger, I., & Morris, W. J. (2007). Prostate brachytherapy post-implant dosimetry: A comparison between higher and lower source density. *Radiotherapy and oncology*, *83*(1), 18–24.

Tomé, W. A., & Fowler, J. F. (2002). On cold spots in tumor subvolumes. *Medical Physics*, *29*(7), 1590–1598.

Yoshioka, Y., Yoshida, K., Yamazaki, H., Nonomura, N., & Ogawa, K. (2013). The emerging role of high-dose-rate (HDR) brachytherapy as monotherapy for prostate cancer. *Journal of radiation research*, *54*(5), 781–788.

Zakariaee, R., Hamarneh, G., Brown, C. J., Gaudet, M., Aquino-Parsons, C., & Spadinger, I. (2016). Bladder accumulated dose in image-guided high-dose-rate brachytherapy for locally advanced cervical cancer and its relation to urinary toxicity. *Physics in Medicine & Biology*, *61*(24), 8408.

Zakariaee, R., Hamarneh, G., Brown, C. J., Gaudet, M., Aquino-Parsons, C., & Spadinger, I. (2017). Association of bladder dose with late urinary side effects in cervical cancer high-dose-rate brachytherapy. *Brachytherapy*, *16*(6), 1175–1183.

A

ESTRO Abstract

Based on the findings in this master thesis a conference abstract has been written. This abstract has been submitted for the ESTRO conference 2022. ESTRO is annual meeting of the European Society for Radiotherapy and Oncology.

Incorporating control of contiguous high-dose volumes in automated optimization for prostate BT (96)

J.L.P. Commandeur[1,2], A. Bouter[2], L.R.M. Dickhoff[3], D.L.J. Barten[4], H. Westerveld[4], B.R. Pieters[4], T. Alderliesten[3], P.A.N. Bosman[1,2]

[1]Technische Universiteit Delft, Delft, the Netherlands; [2]Centrum Wiskunde & Informatica; Life Sciences and Health, Amsterdam, the Netherlands; [3]Leiden University Medical Center, Radiation Oncology, Leiden, the Netherlands; [4]Amsterdam UMC-University of Amsterdam, Radiation Oncology, Amsterdam, the Netherlands

Purpose (756):

In 2020, 'BRachytherapy via artificially Intelligent GOMEA-Heuristic based Treatment planning' (BRIGHT) for prostate HDR BT was clinically introduced. BRIGHT is a bi-objective treatment planning method that finds a set of high-quality, patient-specific treatment plans (TPs) with different trade-offs between clinical target coverage and organ sparing, by directly optimizing the dose volume indices (DVIs) in the clinical protocol. However, in the clinic, manual adjustments of BRIGHT TPs are still done to meet additional patient specific aims. Particularly, this includes minimization of contiguous high-dose volumes, i.e., hotspots (HSs). We therefore aim to incorporate control of HS volumes in BRIGHT, while minimally impacting obtainable DVI values.

Methods (1085):

We augment BRIGHT with a third objective to minimize HSs. For this, we define an HS as 'a contiguous volume of >0.1 mL outside catheters receiving >300% in target volumes: prostate and seminal vesicles, or >200% in normal-tissue around target volumes of the prescribed dose'. We tailored a graph-based method, which uses a connected component algorithm (Afforest), to determine HSs. The graph consists of dose calculation points (DCPs) as nodes and edges between close (≤0.5 mm) neighbouring DCPs. DCPs are randomly sampled locations where the dose is calculated (to compute the DVIs).

The third objective in tri-objective BRIGHT is the sum of HS volumes (metric 1). For comparison, we also consider as third objective a more efficiently computable metric, which however ignores whether the high-dose volume is contiguous: the sum of $V_{300\%}$ of the target volumes and $V_{200\%}$ of normal tissue (metric 2).

We compare bi-objective BRIGHT with both tri-objective BRIGHT versions on a data set of 11 prostate cancer patients by retrospectively planning single-dose HDR BT with $\mathrm{D}_{90\%}^{\mathrm{Prostate}} > 15\mathrm{Gy}$.

Results (892):

Figure 1 shows for patient 9 that, both tri-objective BRIGHT versions result in a clear improvement in control of HSs; TPs with HSs ≤0.5 mL are only found using the tri-objective versions. Due to the nature of metric 2 and using a 2D plot for a 3D front, the trade-off between existing DVIs and metric 2 culminates in a larger covered area with TPs with low HS volumes.

Table 1 shows that when using metric 1, for 10 out of 11 patients, total HS volume could be reduced to ≤0.5 mL while satisfying the clinical protocol, versus 6 out of 11 patients for bi-objective BRIGHT. Adding metric 1 does not result in worsening of DVIs. Adding metric 2 does cause slight worsening of DVIs but results in more plans satisfying the clinical protocol without HSs.

Currently, using metric 1 and 2 takes 1800s and 600s, respectively. Metric 1 needs further optimization to definitively assess runtime impact.

Conclusion (252):

We successfully adapted BRIGHT to reduce HSs without compromising obtainable DVI values for most patients, by explicitly computing HSs and minimizing their volume through a third objective. This could potentially render manual HS adjustments redundant.

Keywords: prostate, hotspots, automated-planning.

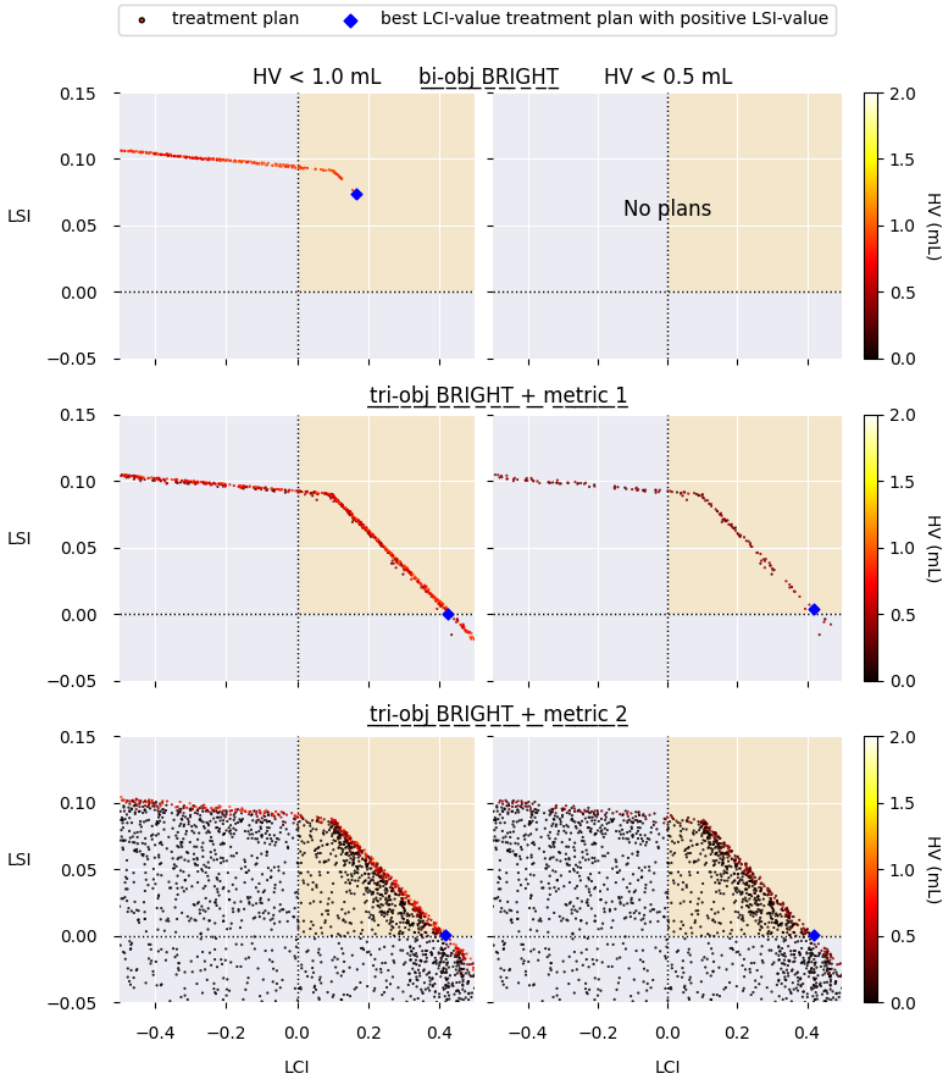

Figure 1: Set of treatment plans automatically generated for patient 9 by bi-objective BRIGHT and both versions of tri-objective BRIGHT. Two different total HS volume (HV) upper bound filters (1.0 mL (left column) and 0.5 mL (right column)) are applied to the resulting plans (post-processing) to give more insight in achieved HV values in the 3D fronts of tri-objective BRIGHT. In the 2D plots the trade-off between coverage and sparing is shown. The LCI (x-axis) is constructed by combining the coverage aims (DVIs) from the clinical protocol, i.e., $V_{100\%}^{\text{Prostate}} > 95\%$, $V_{11\text{Gy}}^{\text{Base Seminal Vesicles}} > 95\%$ and $D_{90\%}^{\text{Prostate}} > 15\text{Gy}$ in a worst-case manner. The LSI (y-axis) is constructed by combining sparing aims (DVIs), i.e. $V_{150\%}^{\text{Prostate}} < 40\%$, $V_{200\%}^{\text{Prostate}} < 15\%$, $D_{1\text{ mL}}^{\text{Bladder}} < 13\text{ Gy}$,

$D_{2\,mL}^{Bladder} < 12$ Gy, $D_{1\,mL}^{Rectum} < 11$ Gy, $D_{2\,mL}^{Rectum} < 9.5$ Gy, $D_{30\%}^{Urethra} < 16.5$ Gy and $D_{0.1\,mL}^{Urethra} < 18$ Gy. As example an LCI of 2.0% means that all targets are covered at least 2.0% more than their planning-aim. The upper-right region in yellow represents the golden corner, where all treatment plans have a positive LSI and LCI, and thereby satisfy all treatment planning aims (DVIs). The treatment plan with the best LCI-value with a positive LSI-value is indicated by a blue diamond. The HV in a treatment plan is shown using a color gradient (right-hand side).

Total Hotspot Volume	bi-obj BRIGHT				tri-obj BRIGHT + metric 1			tri-obj BRIGHT + metric 2		
	<=∞	<=1.0	<=0.5	<=0.0	<=1	<=0.5	<=0.0	<=1.0	<=0.5	<=0.0
P01	0.46	0.46	0.46	0.00	0.47	0.47	-0.12	0.44	0.44	0.42
P02	0.53	0.53	0.15	N/A	0.54	0.42	N/A	0.43	0.41	0.32
P03	0.44	0.45	0.45	0.27	0.46	0.46	0.31	0.44	0.42	0.41
P04	0.00	N/A	N/A	N/A	-0.12	-0.43	N/A	-0.16	N/A	N/A
P05	0.34	0.32	0.32	0.32	0.31	0.31	0.31	0.29	0.28	0.28
P06	0.41	0.38	-0.01	N/A	0.40	0.39	N/A	0.37	0.36	0.31
P07	0.10	0.02	-0.43	N/A	0.08	0.07	-0.22	0.05	0.04	-0.11
P08	0.50	N/A	N/A	N/A	0.43	0.38	-0.23	0.44	0.37	N/A
P09	0.44	0.11	N/A	N/A	0.43	0.38	N/A	0.41	0.40	0.36
P10	0.41	0.40	0.39	-0.26	0.42	0.41	0.36	0.39	0.39	0.37
P11	0.34	0.34	0.34	N/A	0.36	0.35	0.27	0.32	0.32	0.31
# golden corner	11	9	6	3	10	10	4	10	10	8
# worsening	-	3	6	10	2	5	10	10	10	11

GC + no worsening
GC + worsening
No GC + no worsening
No GC + worsening
No plans

Table 1: Summary of the resulting best LCI-value plan given positive LSI-value (LCI values of the blue diamond plans in Figure 1) of treatment planning results (median of 5 runs, to mitigate randomness in BRIGHT) in bi-objective BRIGHT and both versions of tri-objective BRIGHT, given multiple upper bounds on maximum total HS volume (e.g., ≤ 0.5 mL) in the TPs. *Abbreviations: <=X, for plans with HS upper bound X; P01, patient 1, GC = golden corner.* A positive LCI-value means that all clinical aims are satisfied. If no plans had been produced that satisfy the total HS volume upper bound for a patient, then N/A is reported. If the reported median is significantly worse (Mann–Whitney U test, p-value = 0.05) than the found unconstrainted (total HS volume <= ∞) bi-objective BRIGHT value, it is considered as a worsening. In the bottom row the number of patients is reported for which the GC has been reached (# golden corner) and for which number of patients there was a worsening in best LCI value given a positive LSI value (# worsening). For patient 4 and 8, bi-objective BRIGHT resulted in GC treatment plans with a total HS volume of at least 2 mL and 1.5 mL, respectively. Due to unfavourable implant geometry for patient 4, tri-objective BRIGHT was not able to create GC treatment plans with HV <= 1 mL.

B

Results

B.1. Pareto approximation fronts from the clinic

The resulting Pareto approximation fronts from the clinic for the patients treated using BRIGHT. The blue dots represent different treatment plans and the yellow dots represent the plans exported for further inspection in Oncentra Brachy. What can be seen in the different plots is that the long tails of the approximation front are never used. Only the plans close to or in the golden corner (positive LCI and LSI) are selected for further inspection. What also becomes clear from these plots is that for all patients in this patient set the golden corner was (almost) reached.

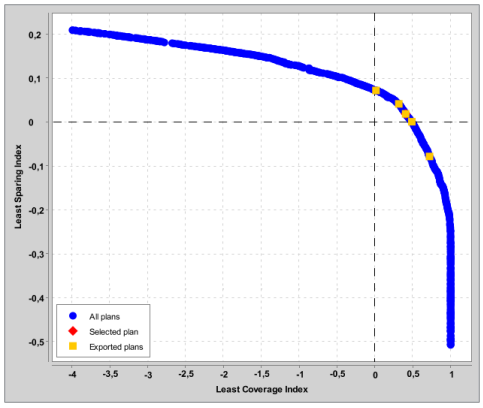

Figure B.1: Patient 1 - clinical approximation front

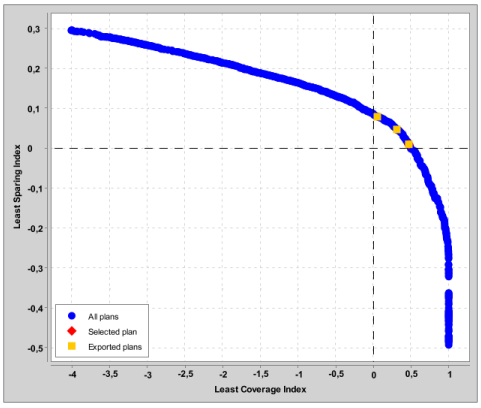

Figure B.2: Patient 2 - clinical approximation front

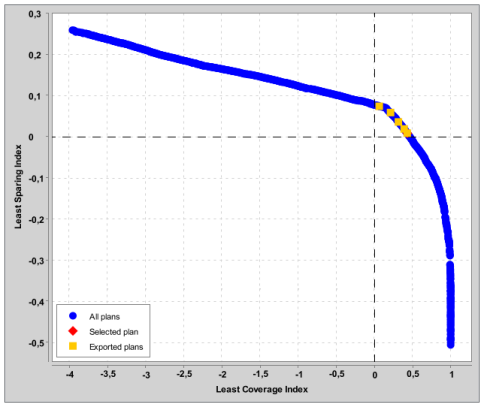

Figure B.3: Patient 3 - clinical approximation front

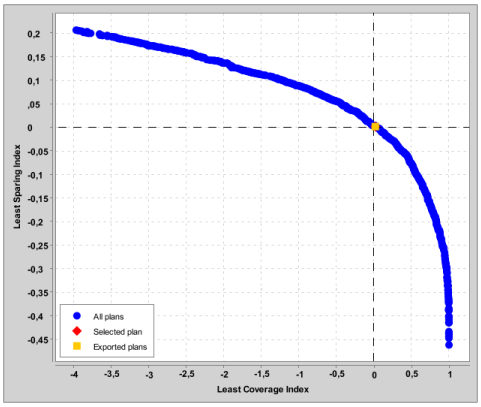

Figure B.4: Patient 4 - clinical approximation front

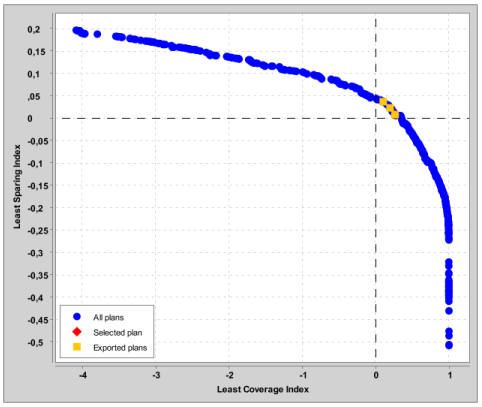

Figure B.5: Patient 5 - clinical approximation front

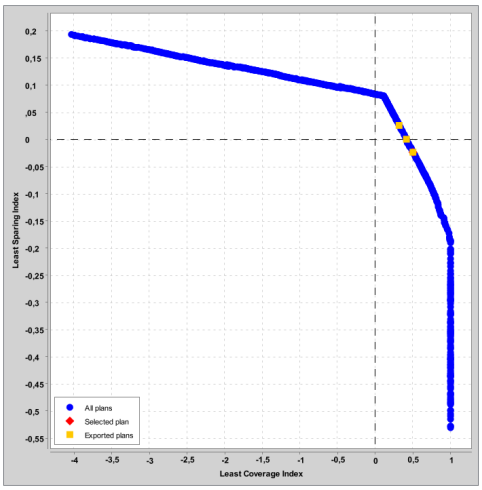

Figure B.6: Patient 6 - clinical approximation front

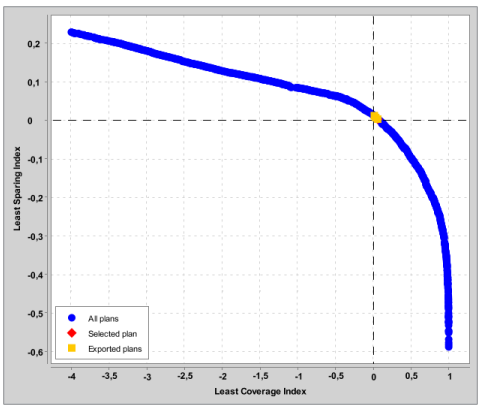

Figure B.7: Patient 7 - clinical approximation front

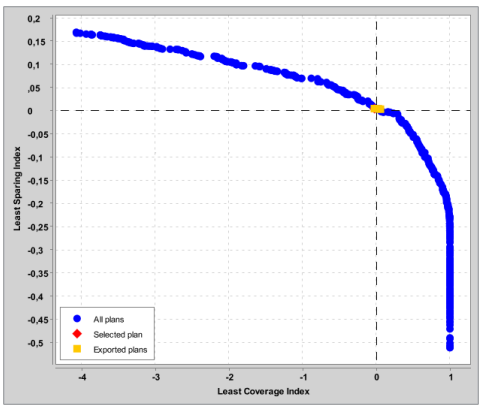

Figure B.8: Patient 8 - clinical approximation front

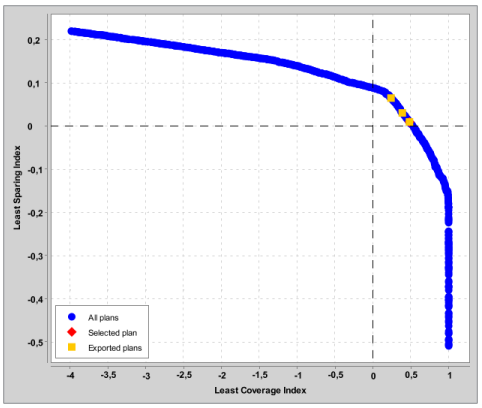

Figure B.9: Patient 9 - clinical approximation front

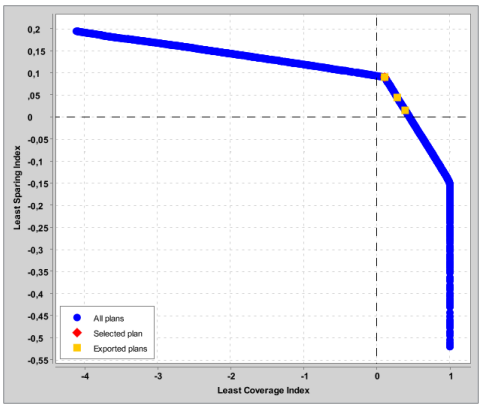

Figure B.10: Patient 10 - clinical approximation front

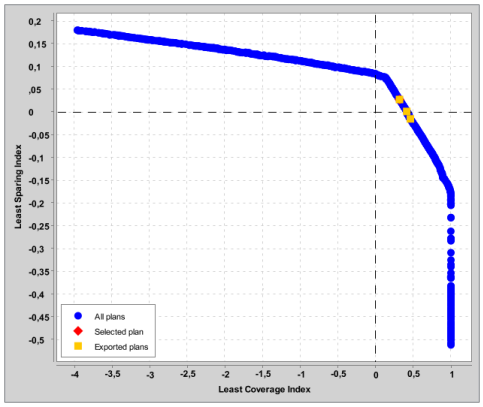

Figure B.11: Patient 11 - clinical approximation front

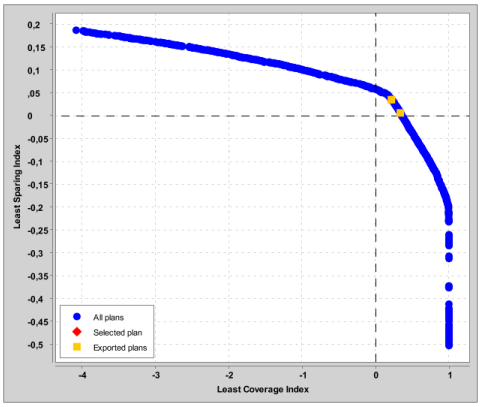

Figure B.12: Patient 12 - clinical approximation front

B.2. Adaptive steering in 2D

In this appendix section all Pareto approximation fronts are shown for the adaptive steering experiment in 2D.

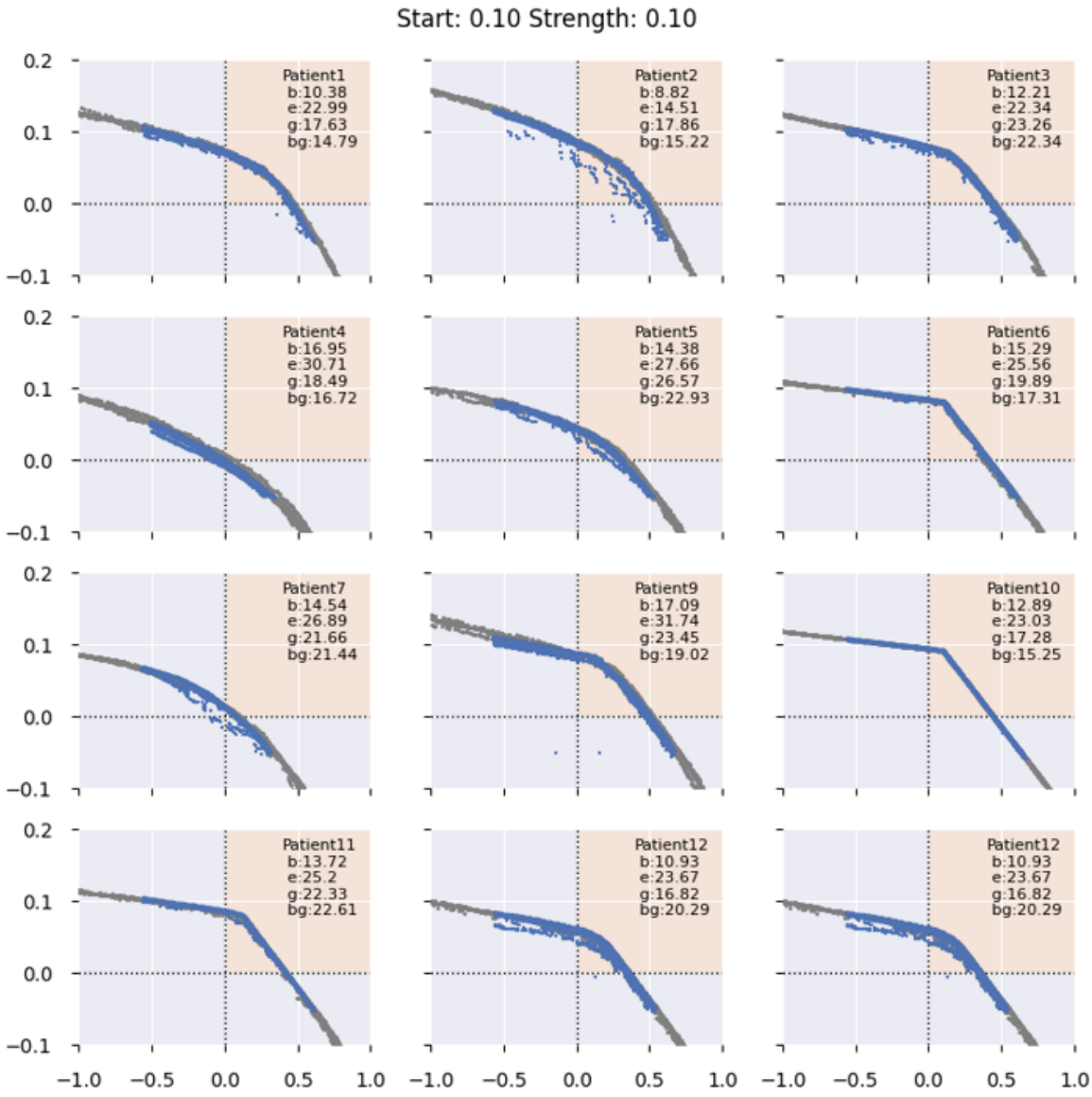

Figure B.13: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

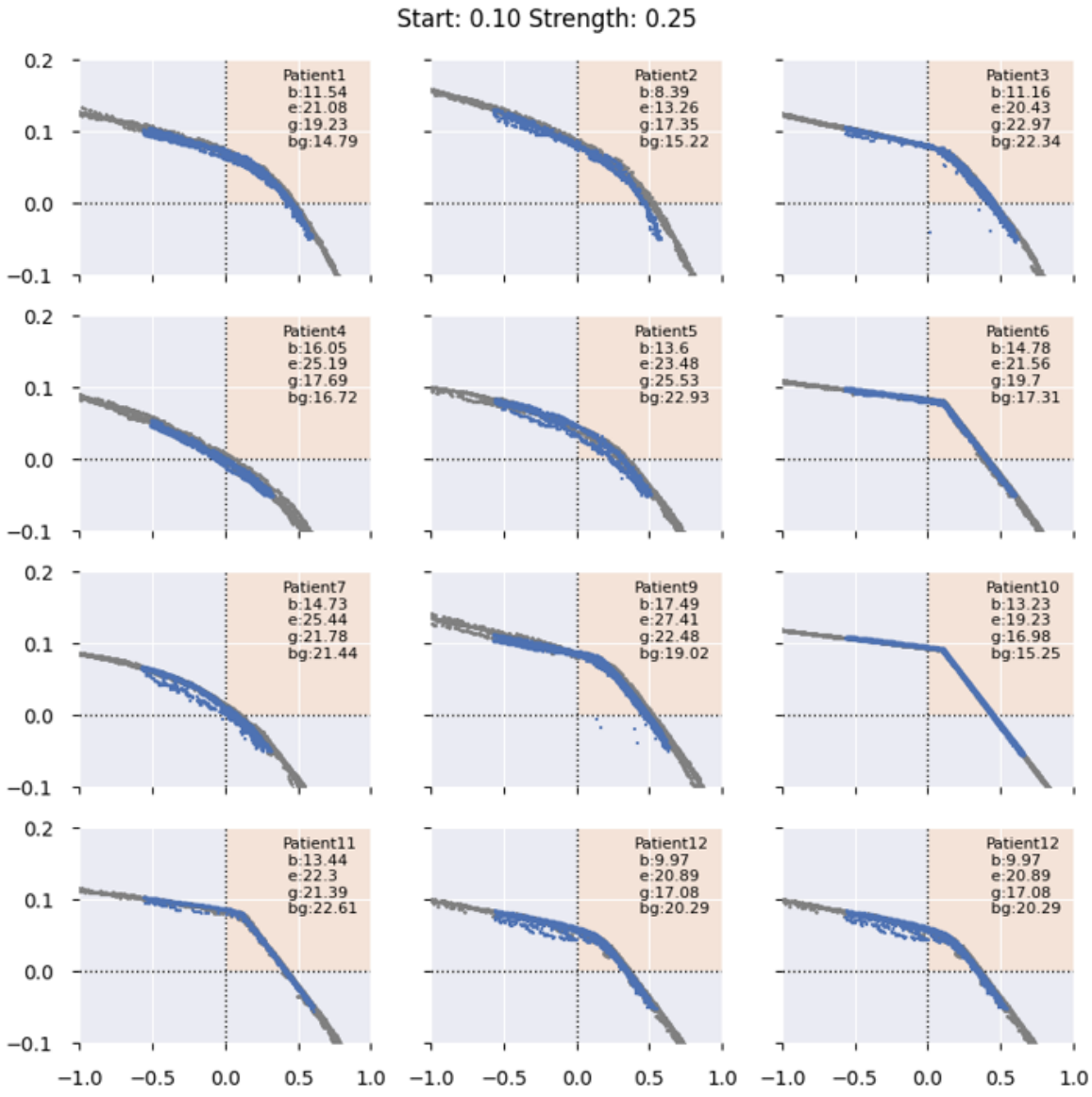

Figure B.14: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

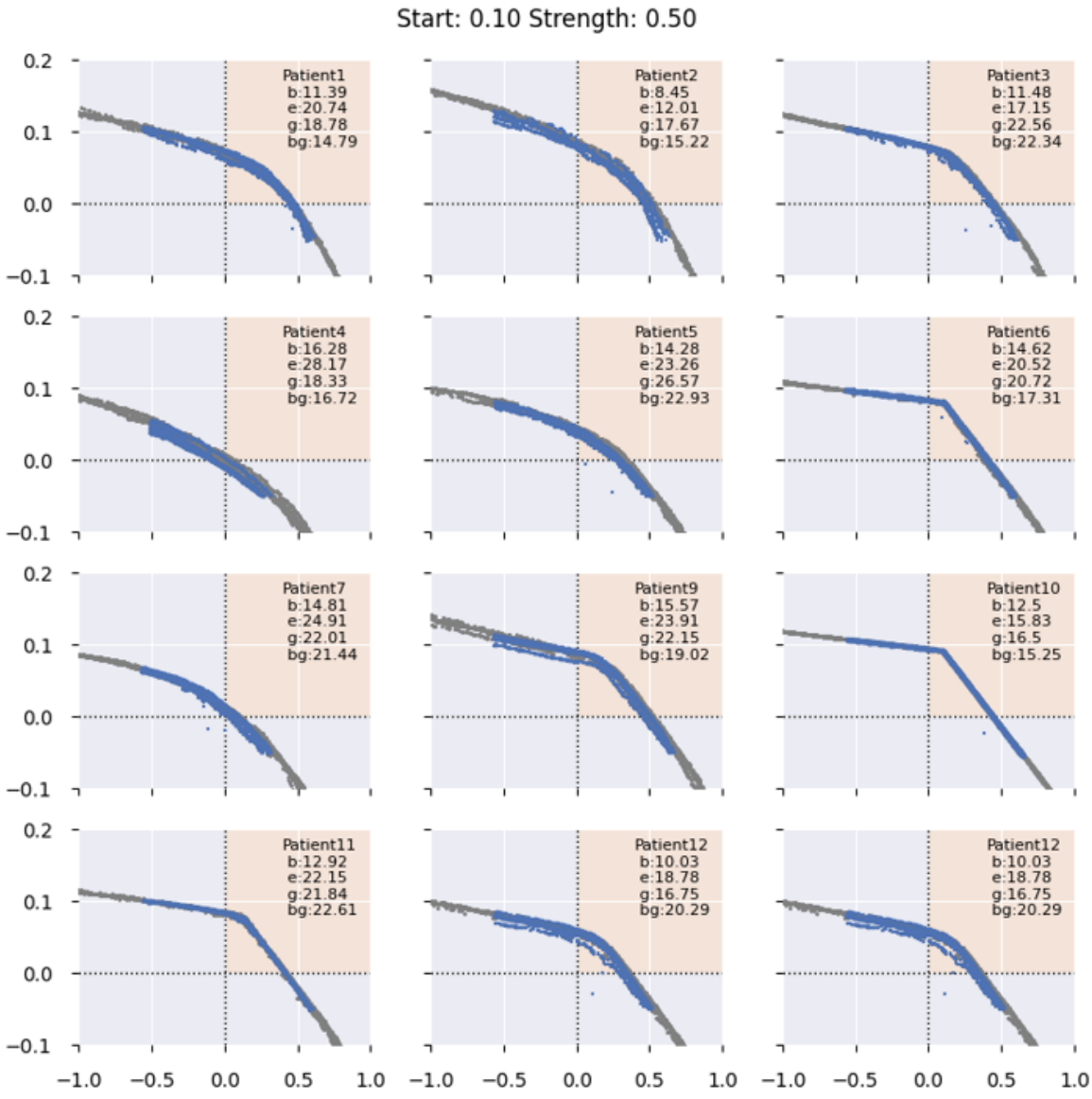

Figure B.15: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

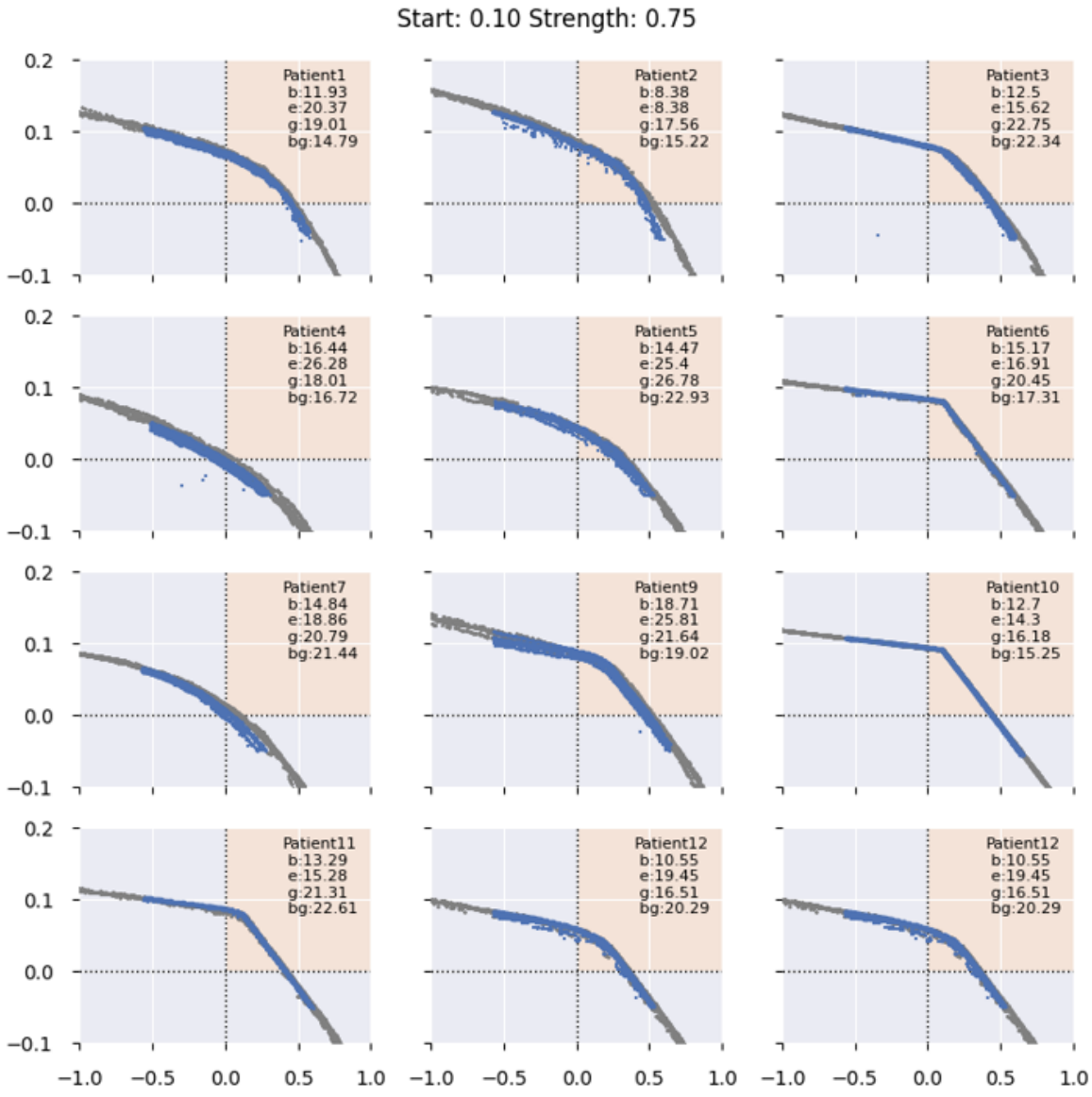

Figure B.16: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

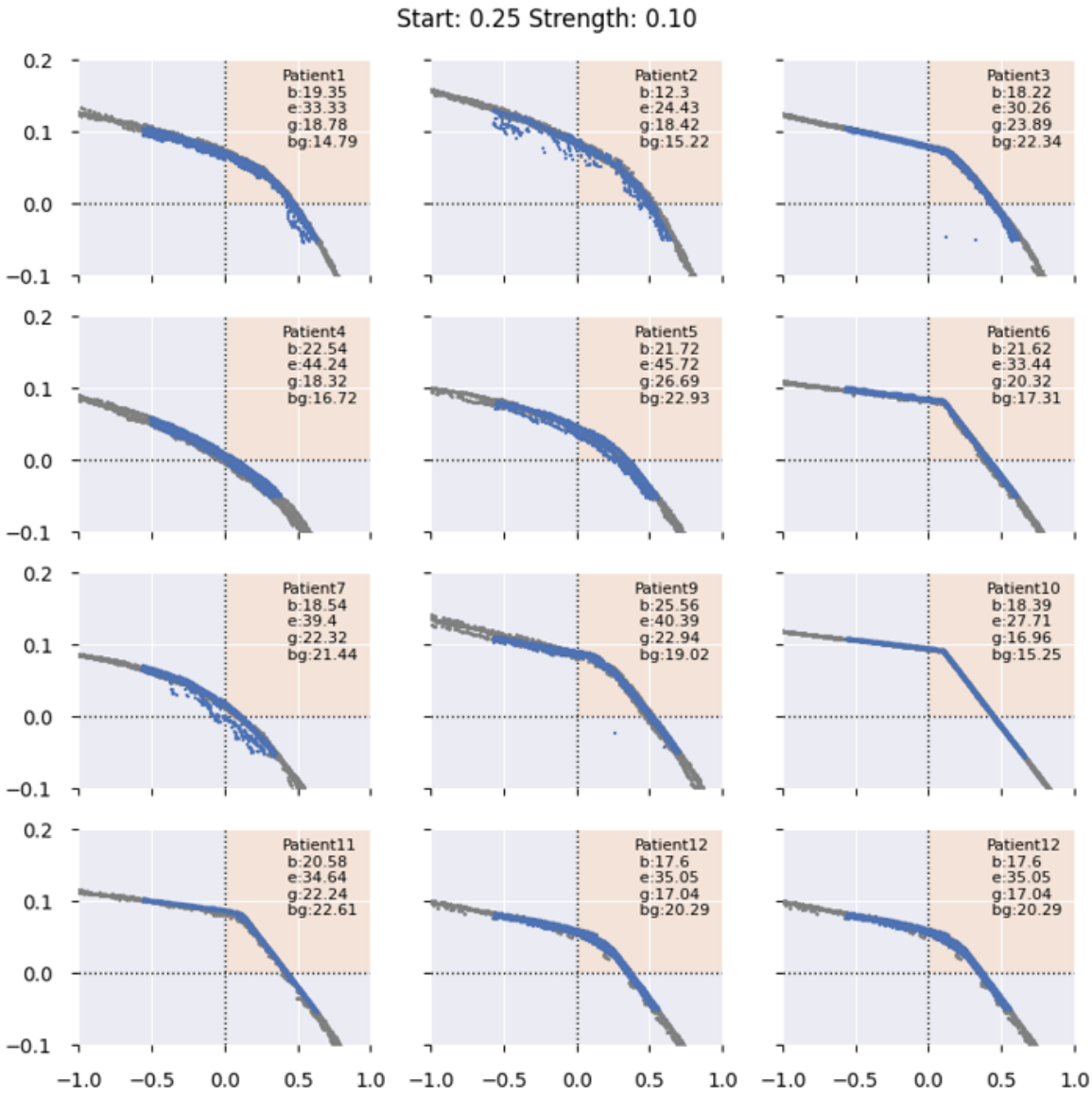

Figure B.17: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

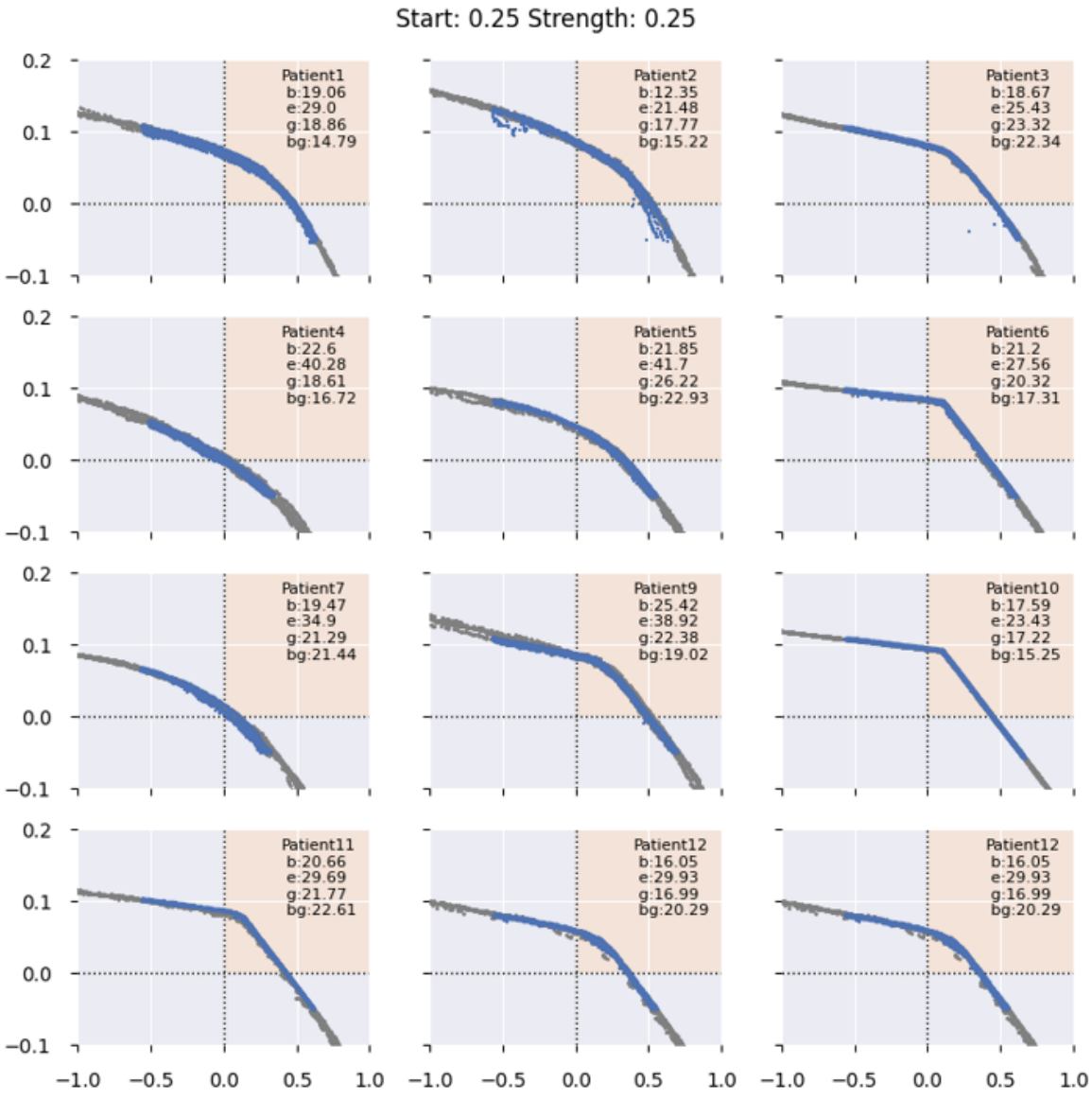

Figure B.18: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

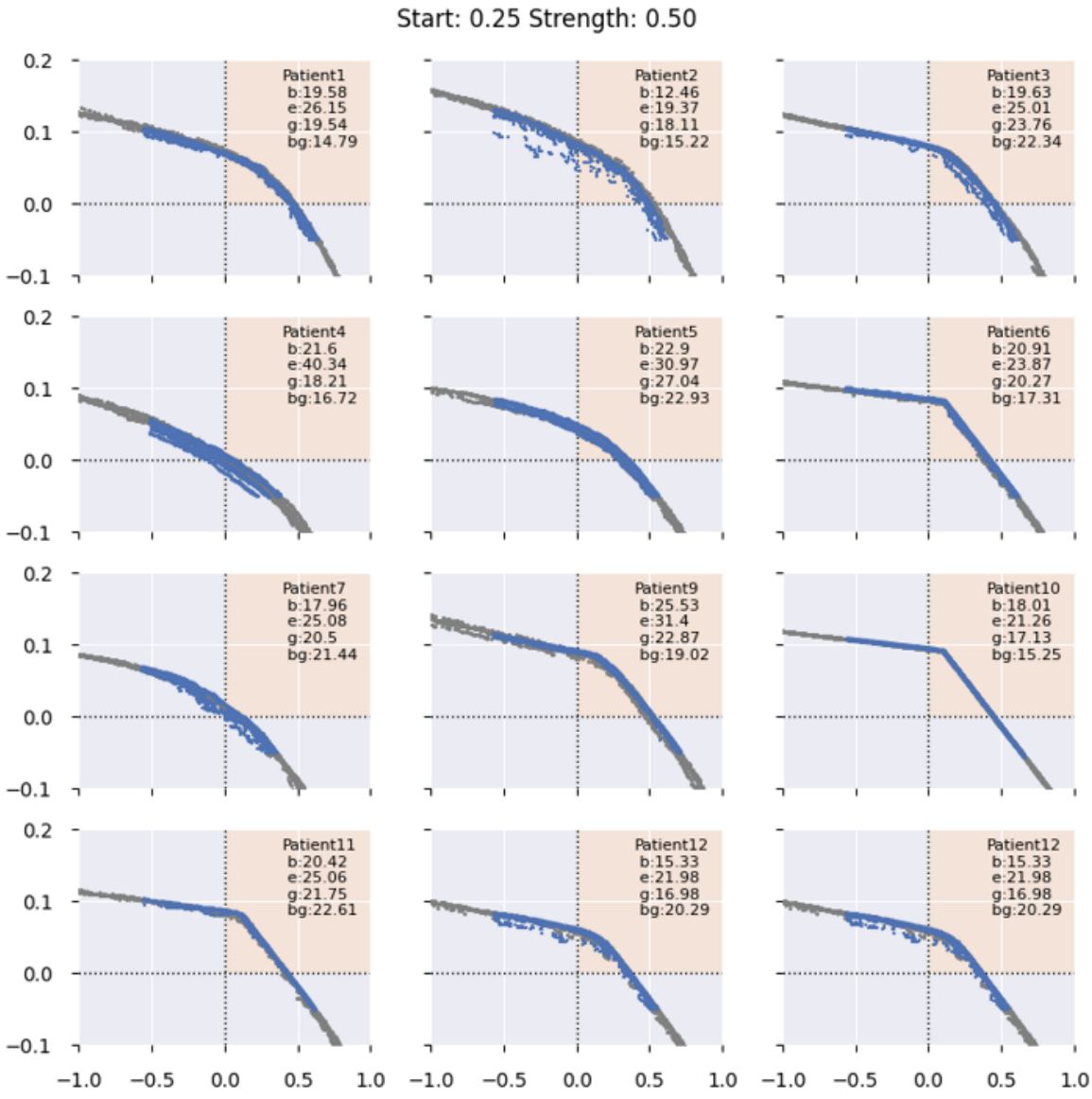

Figure B.19: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

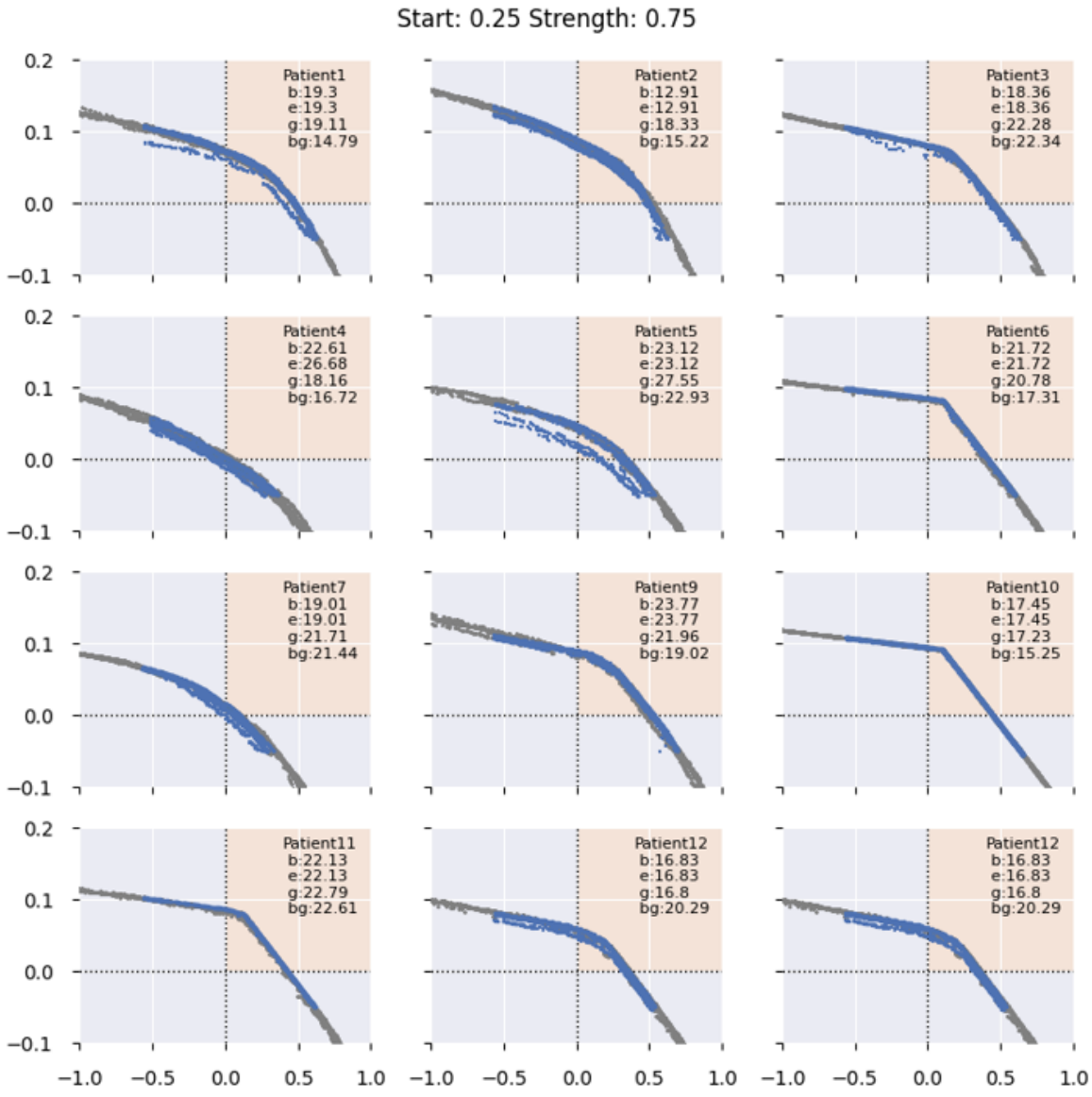

Figure B.20: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

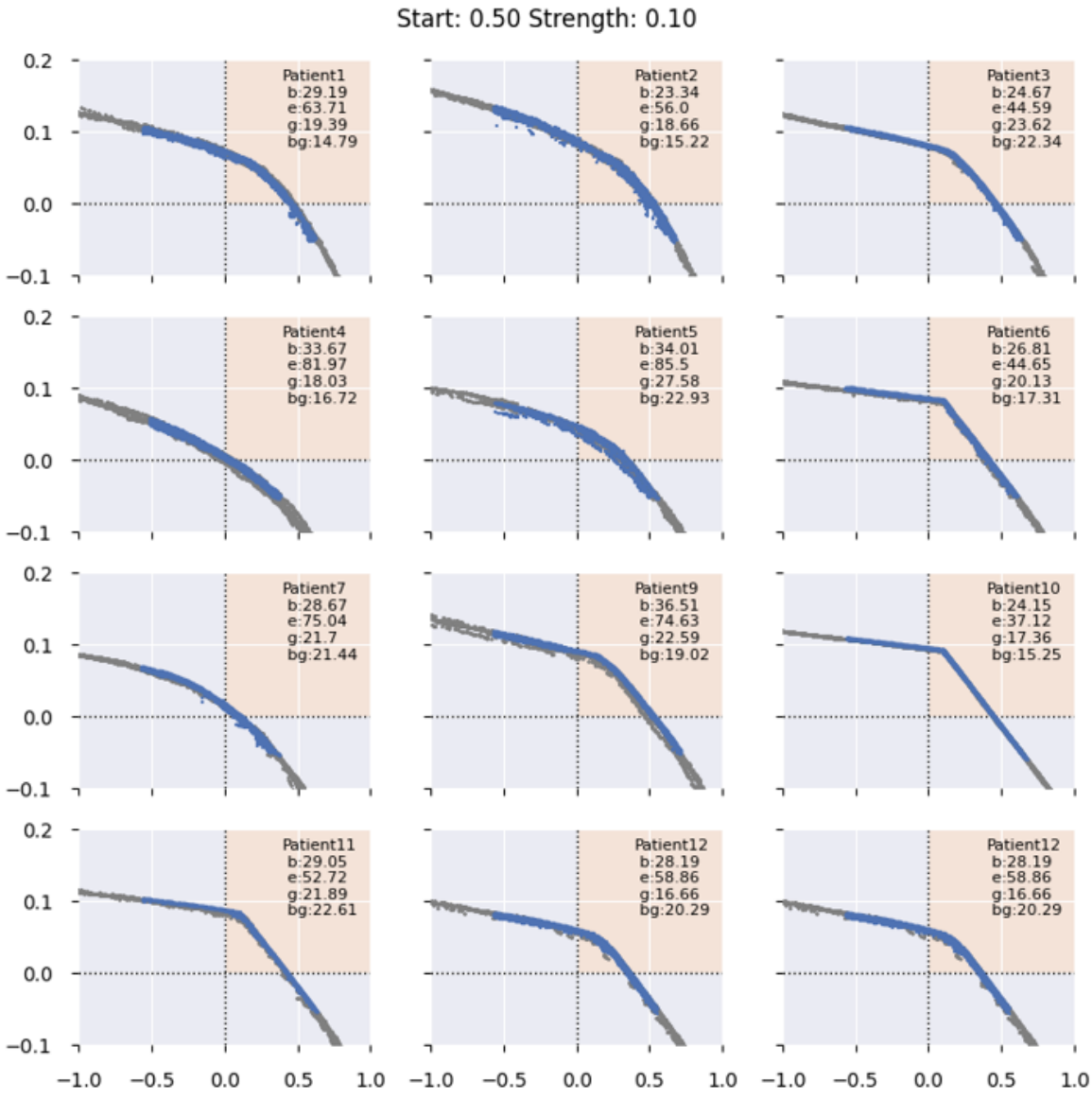

Figure B.21: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

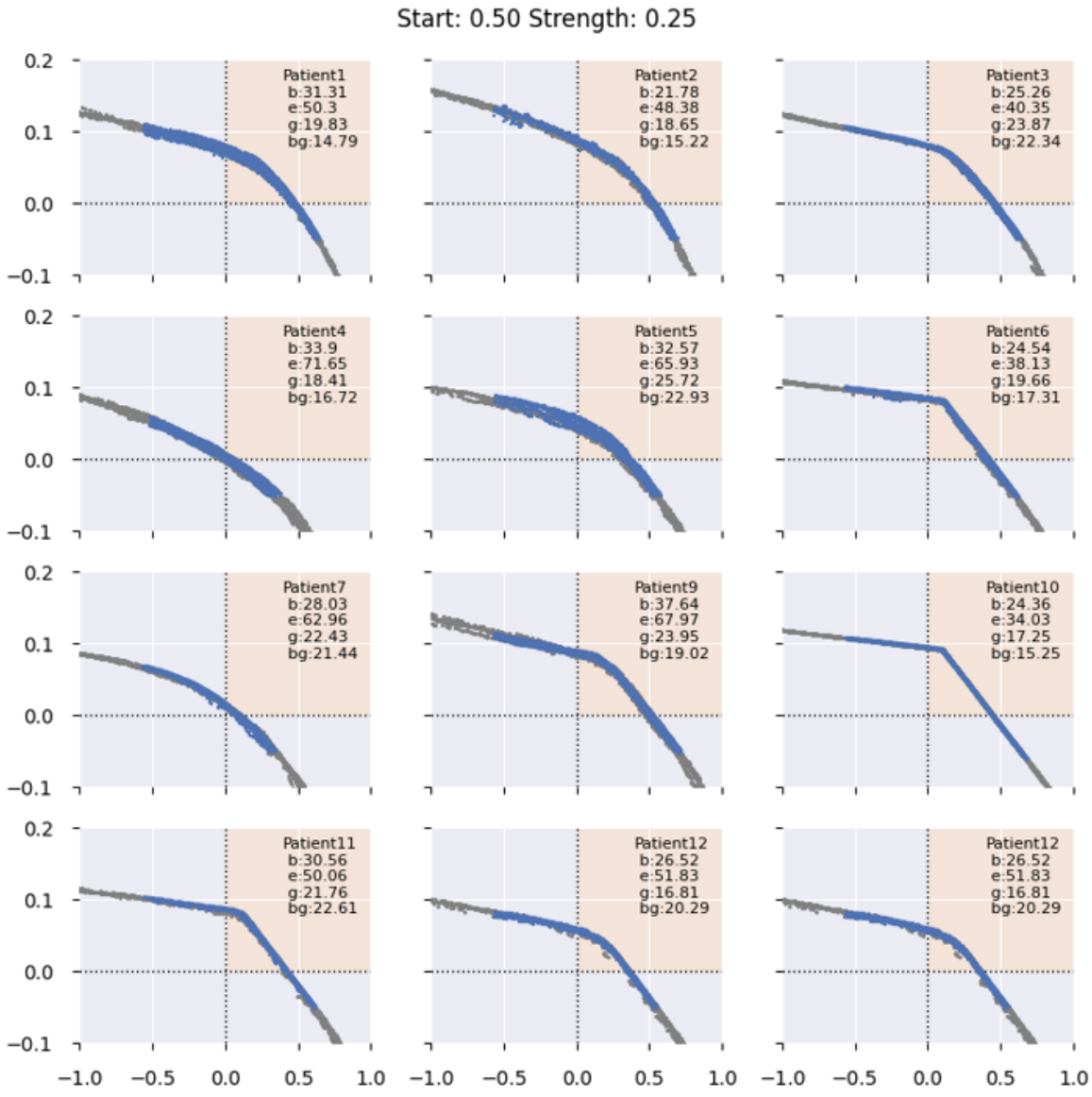

Figure B.22: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

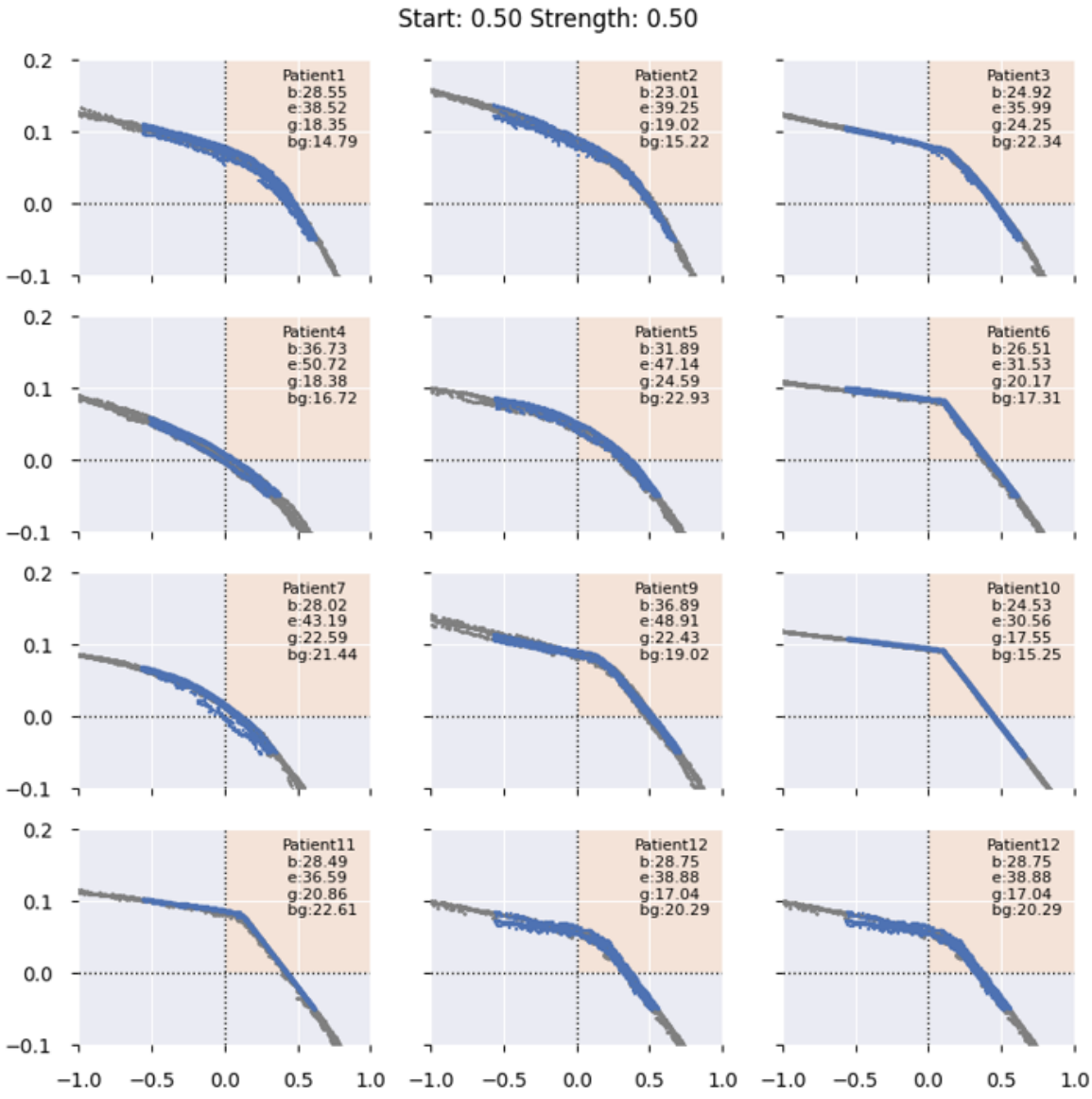

Figure B.23: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

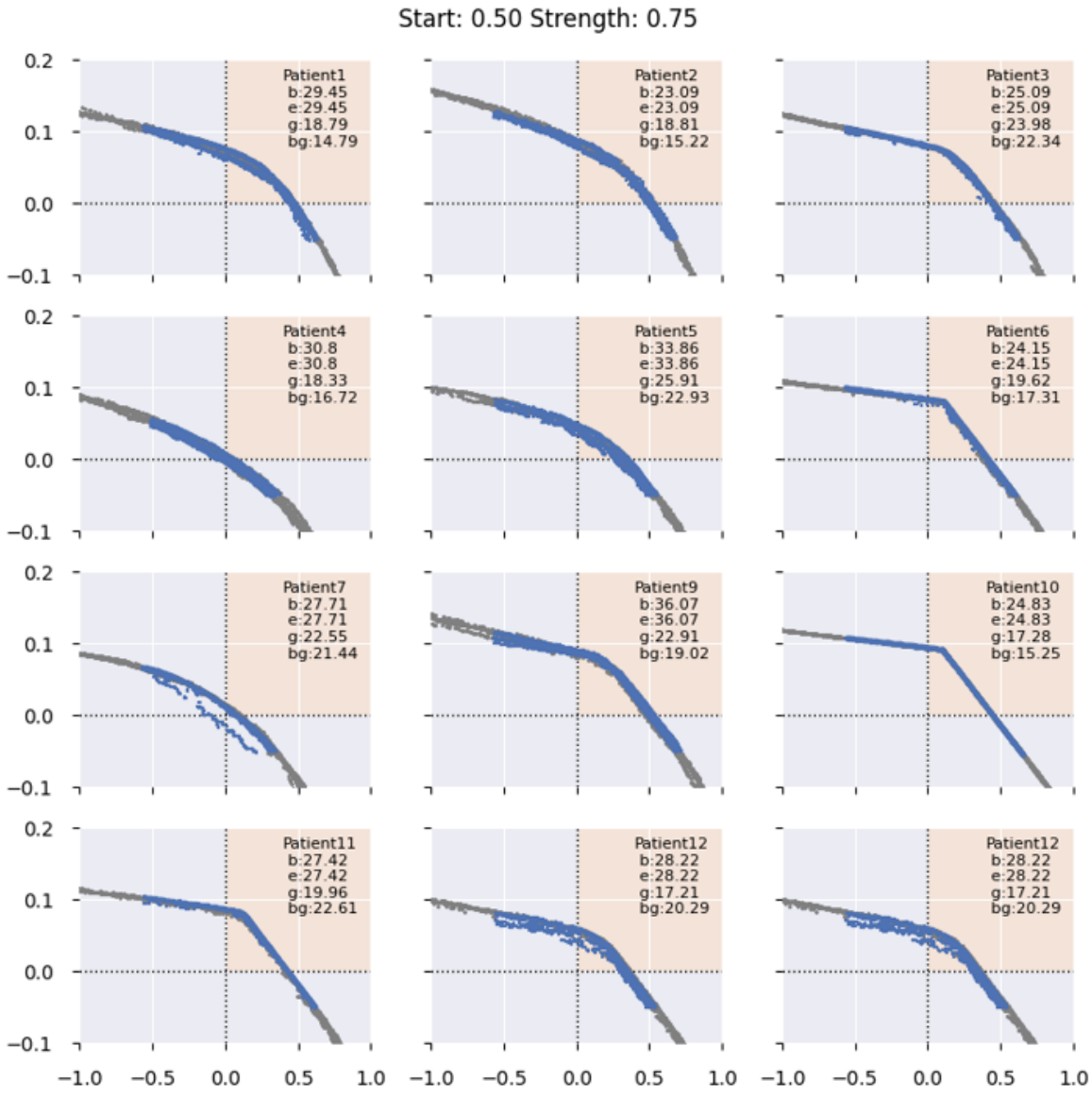

Figure B.24: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

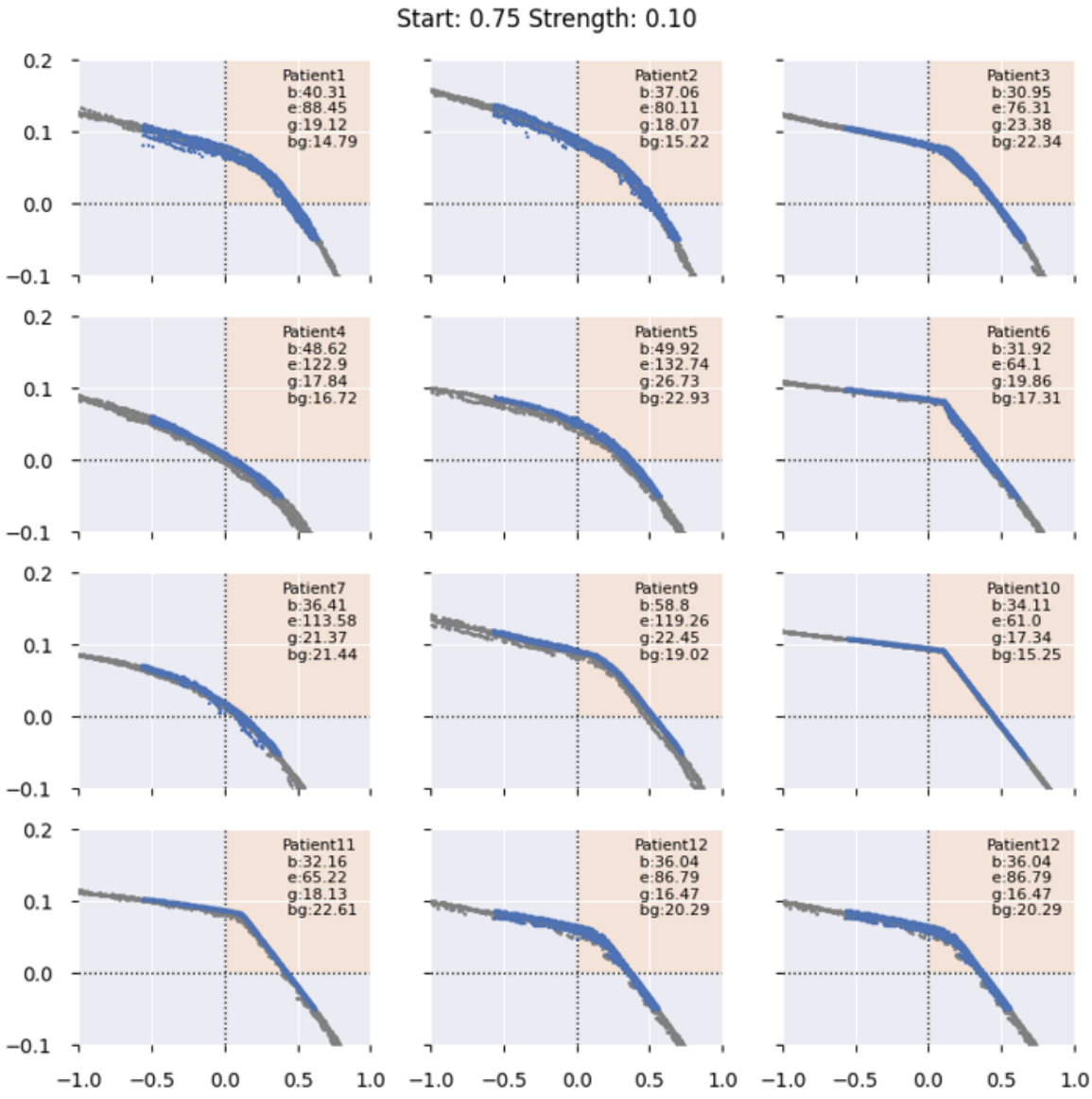

Figure B.25: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

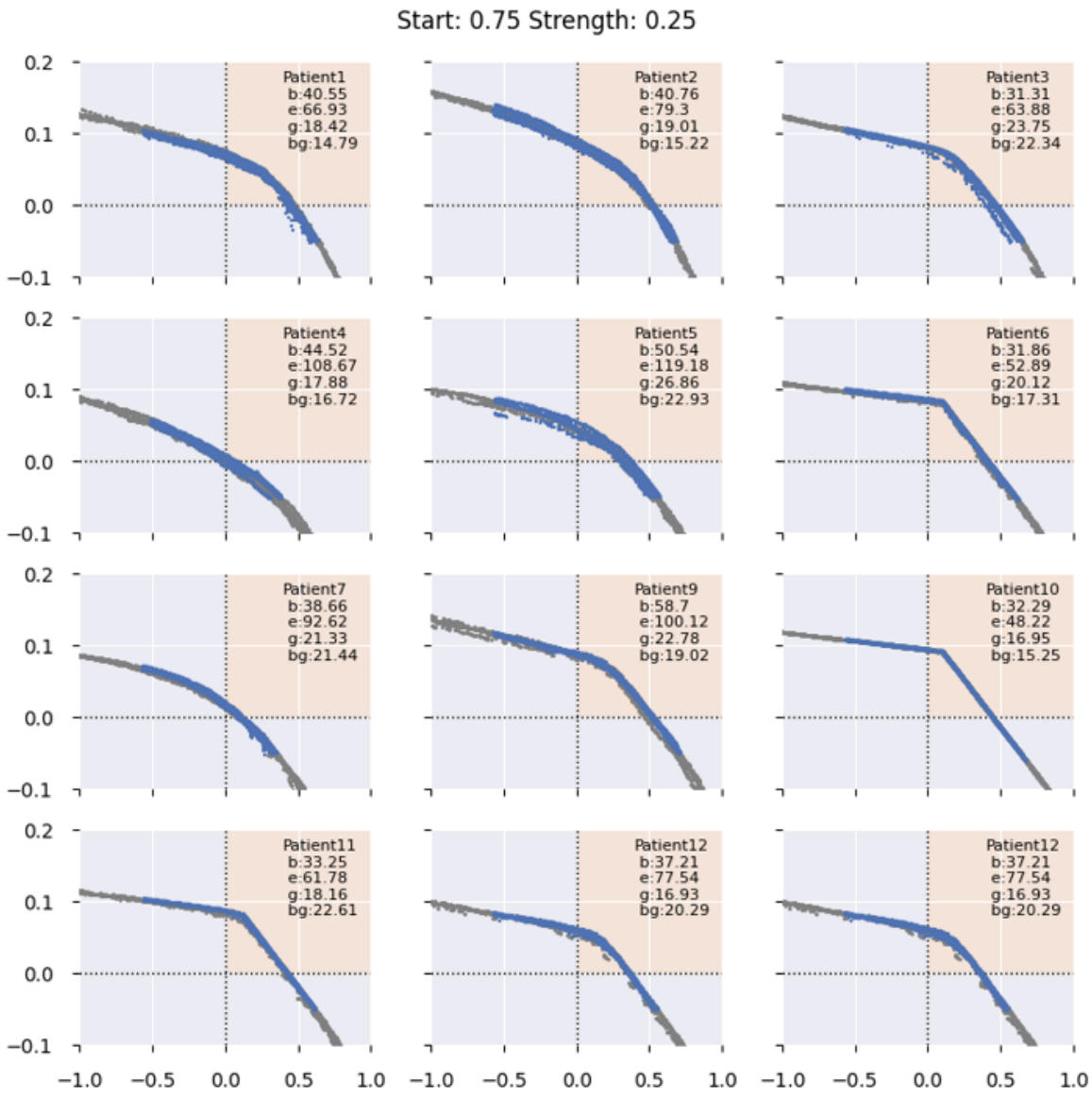

Figure B.26: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

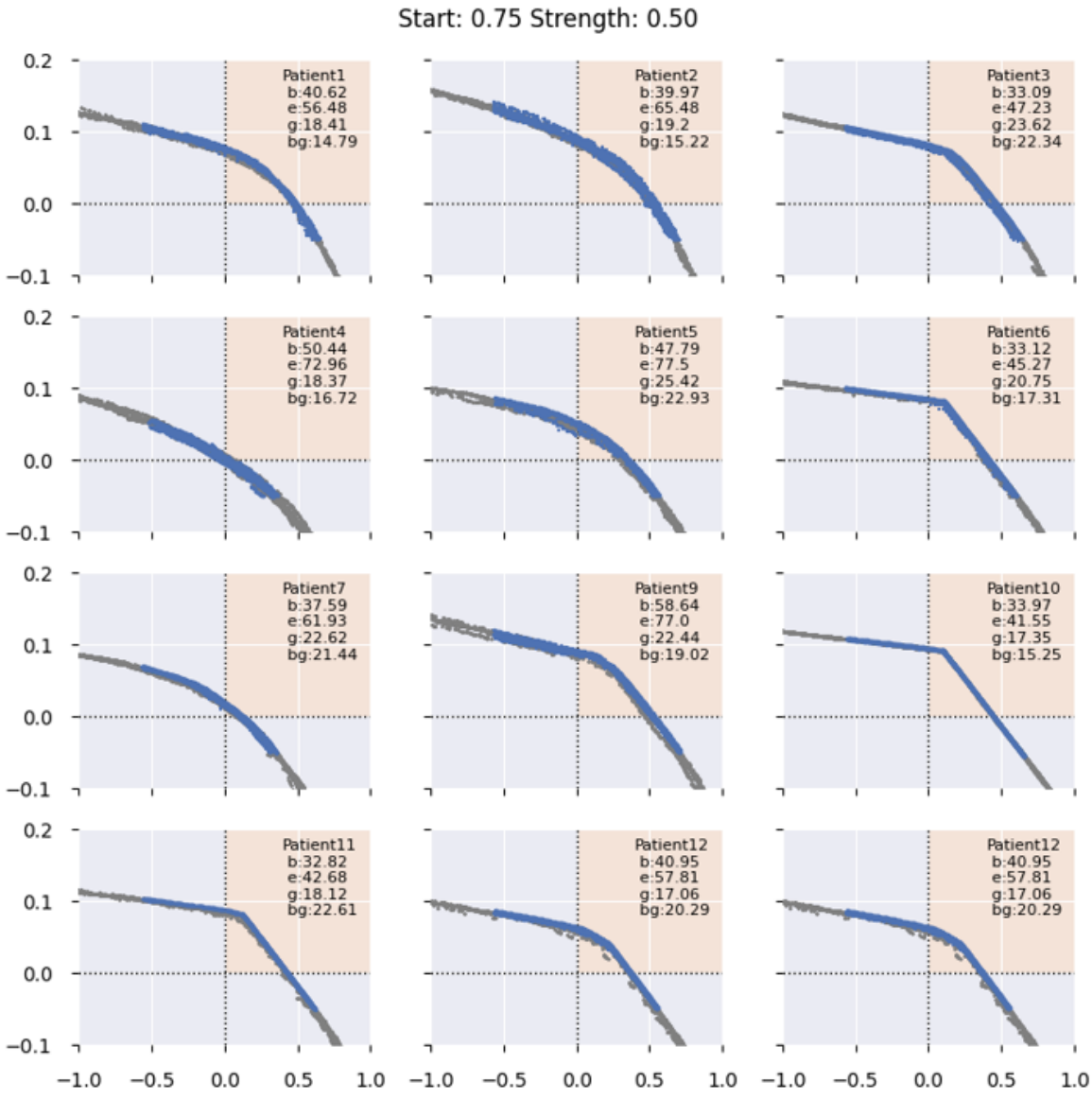

Figure B.27: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

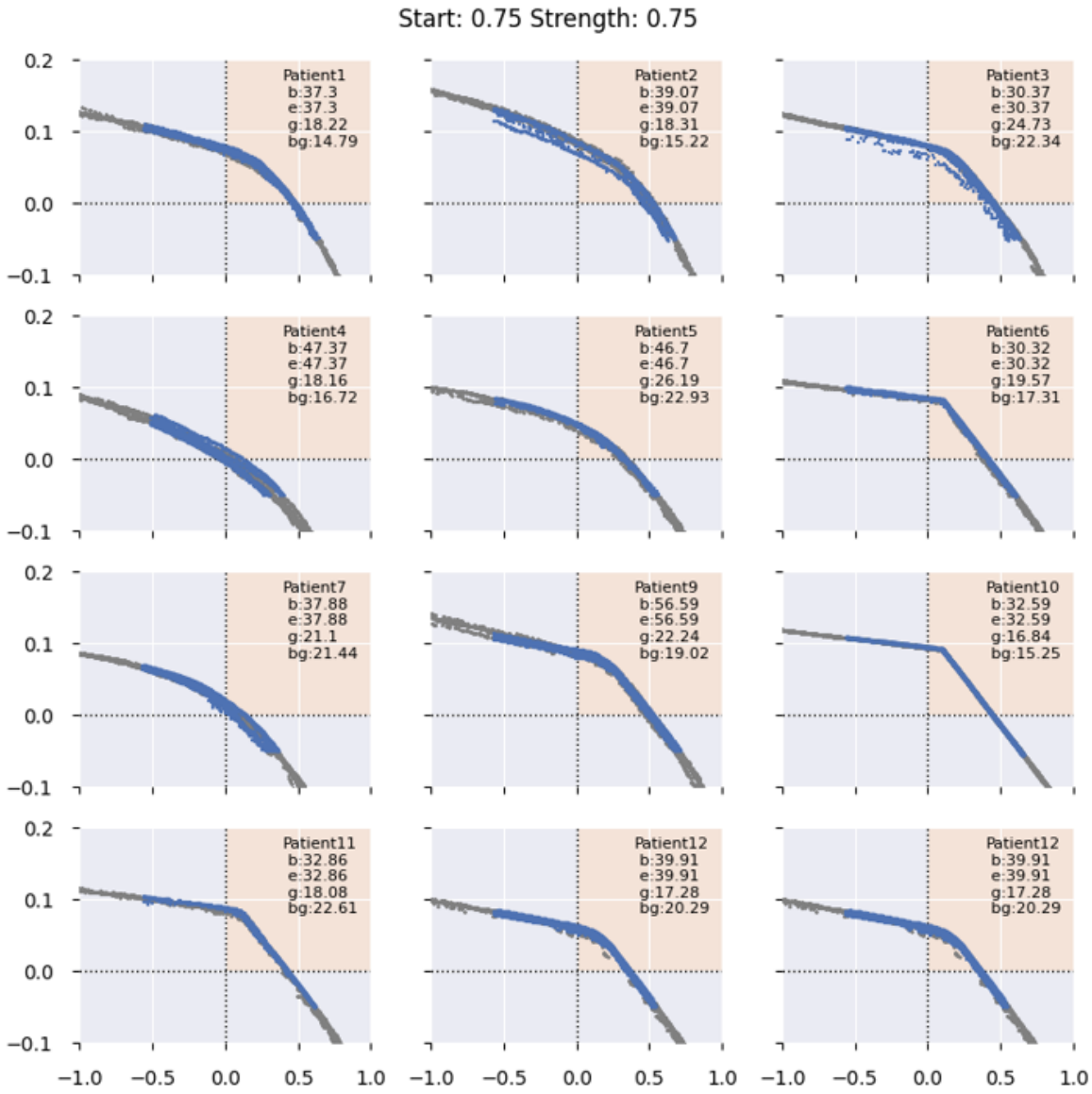

Figure B.28: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic BRIGHT configuration without adaptive steering. In the title of the figure it states what the used setting for adaptive steering were. In the top right corner it states when adaptive steering began (b), when it ended (e), when the golden corner was reached (g) and when the golden corner was reached without adaptive steering (bg).

B.3. Adaptive steering in 3D

In this appendix section all Pareto approximation fronts are shown for the adaptive steering experiment in 3D.

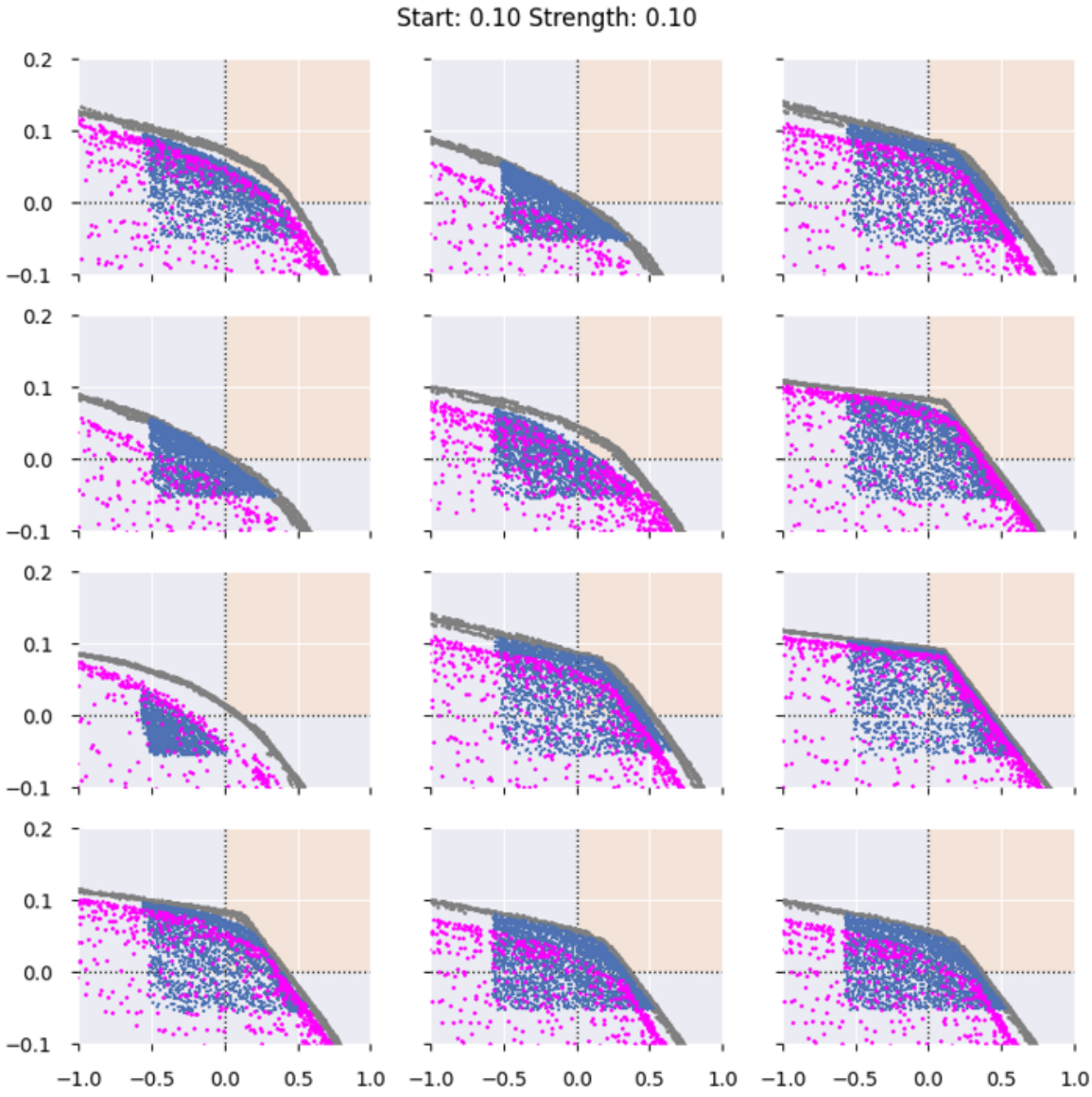

Figure B.29: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

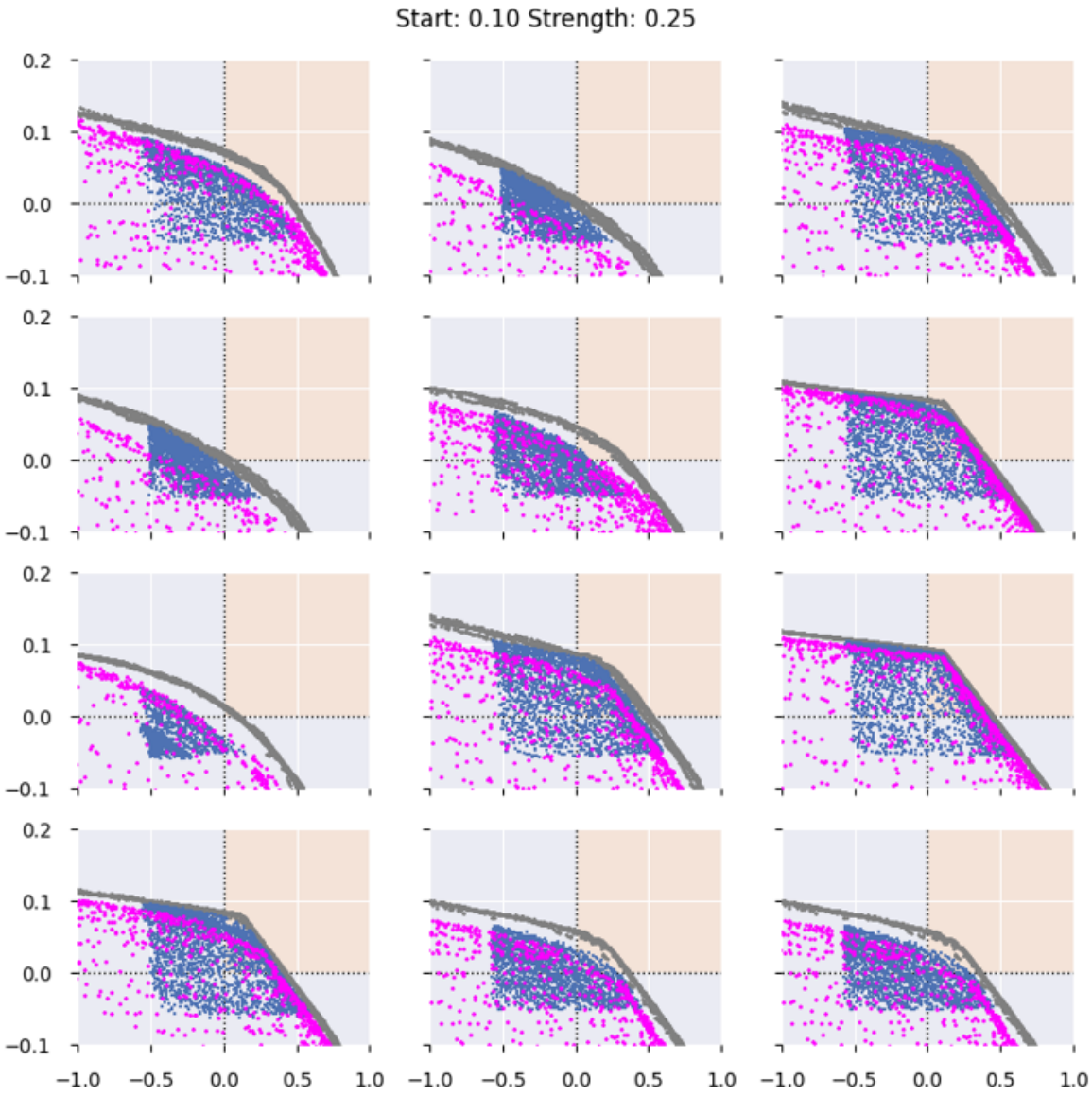

Figure B.30: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

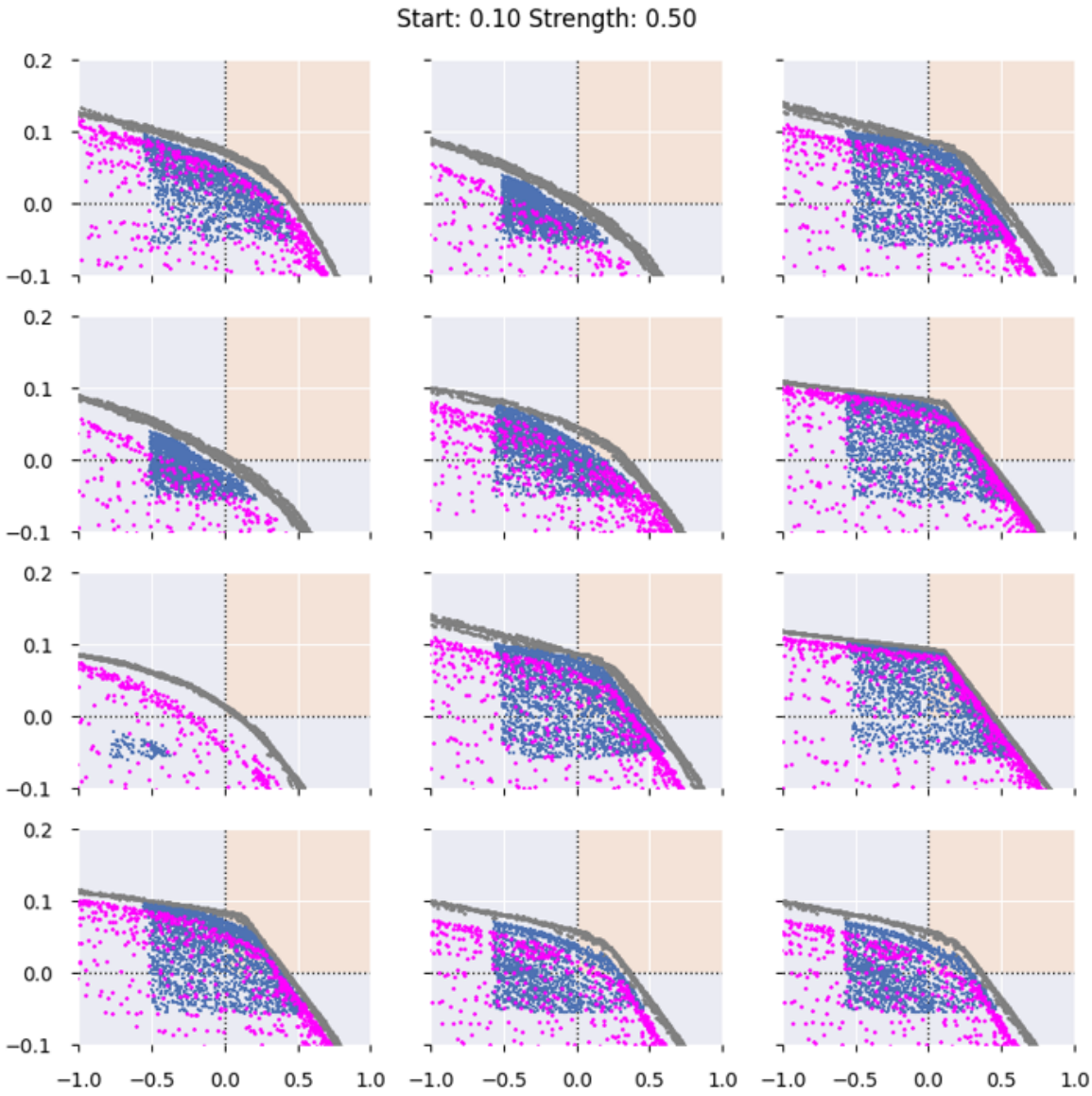

Figure B.31: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

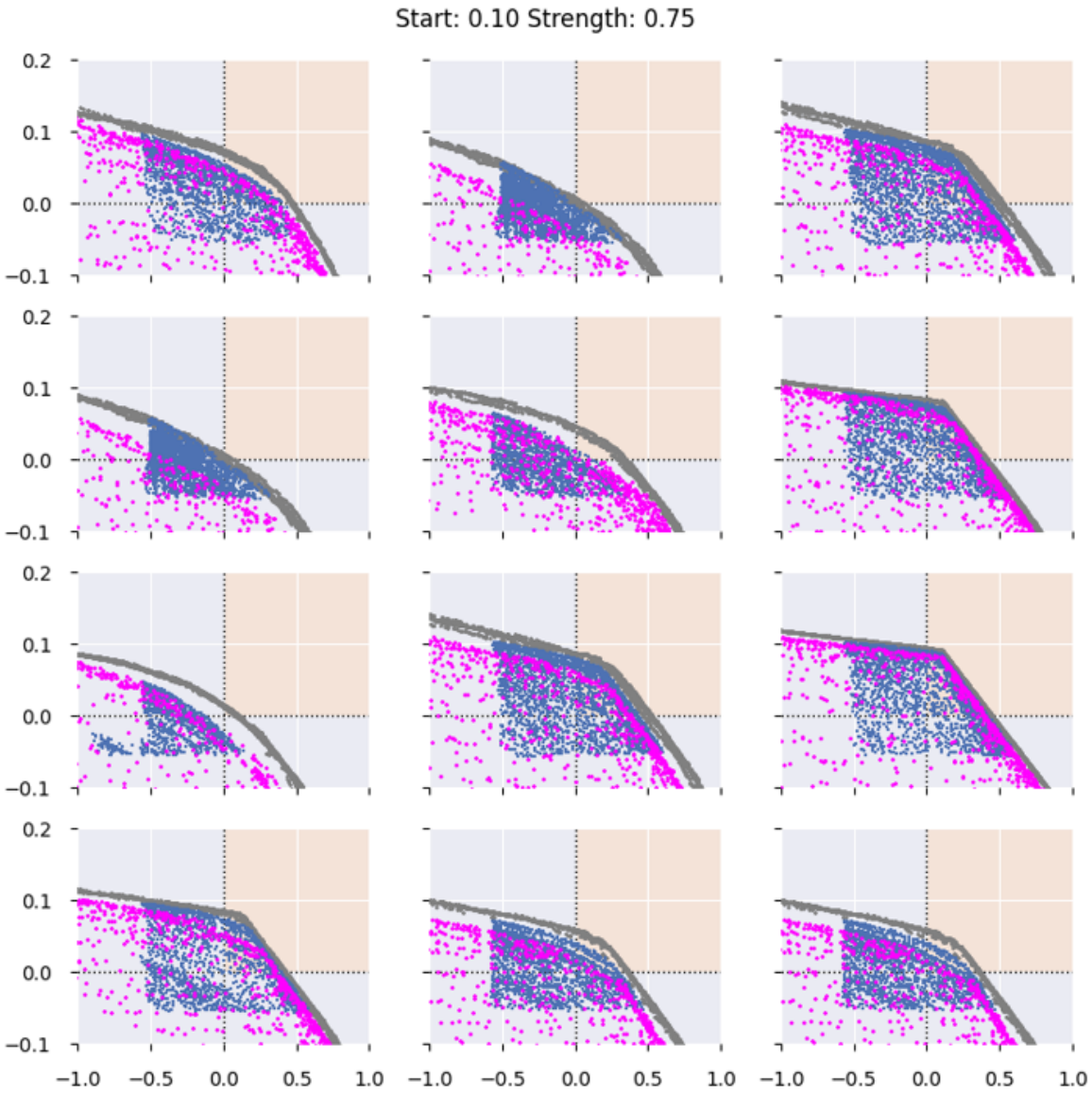

Figure B.32: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

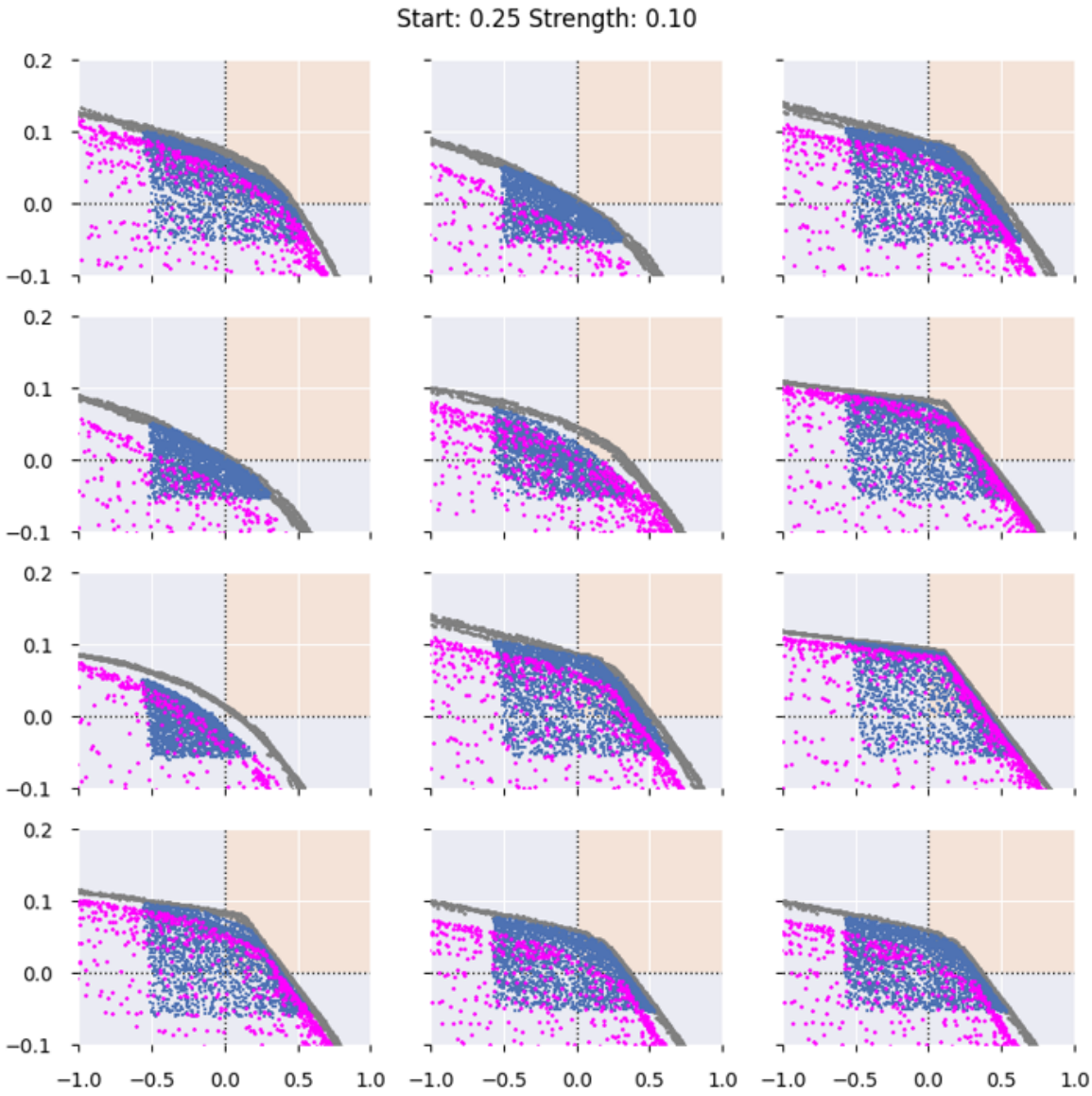

Figure B.33: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

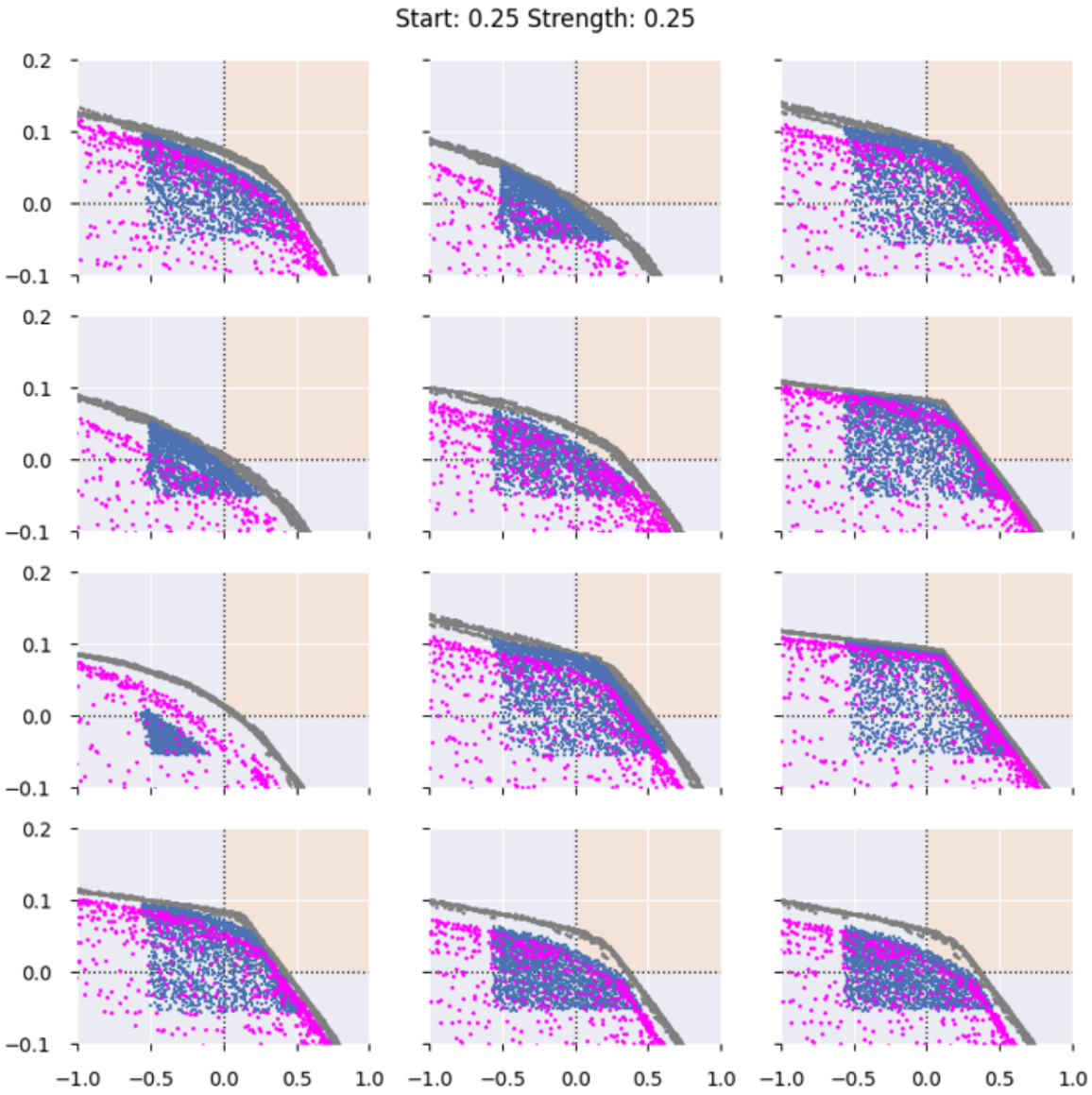

Figure B.34: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

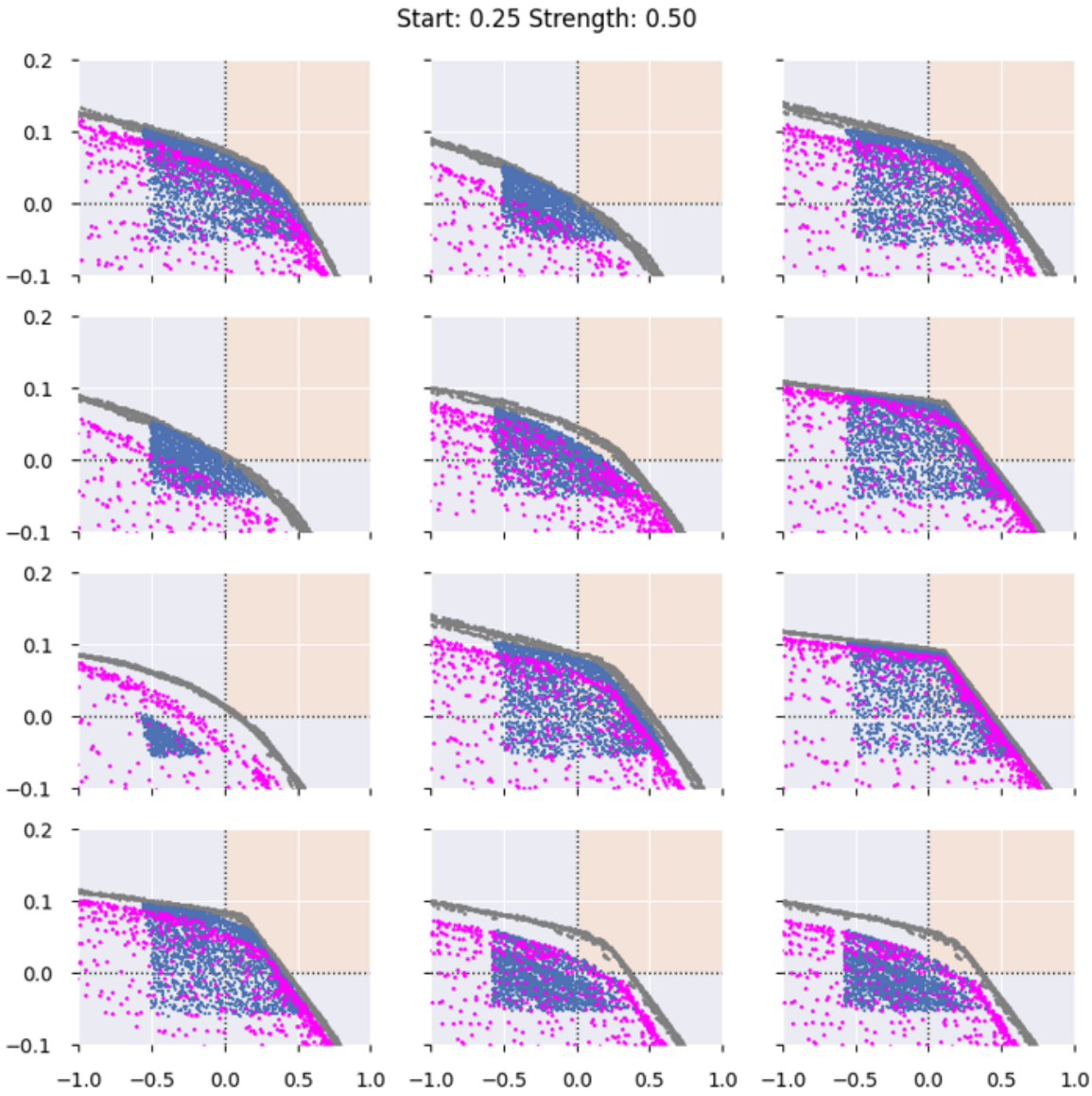

Figure B.35: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

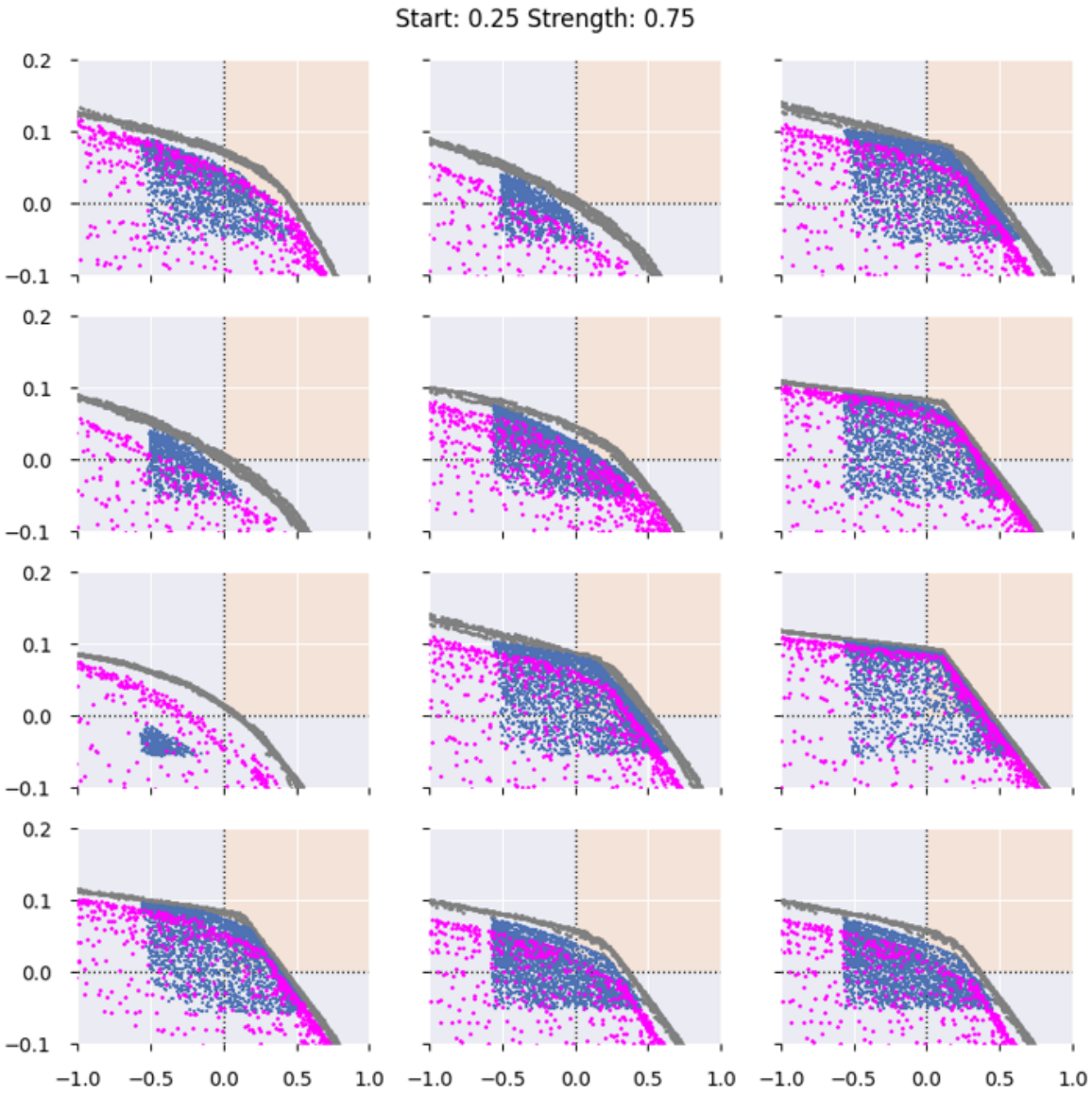

Figure B.36: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

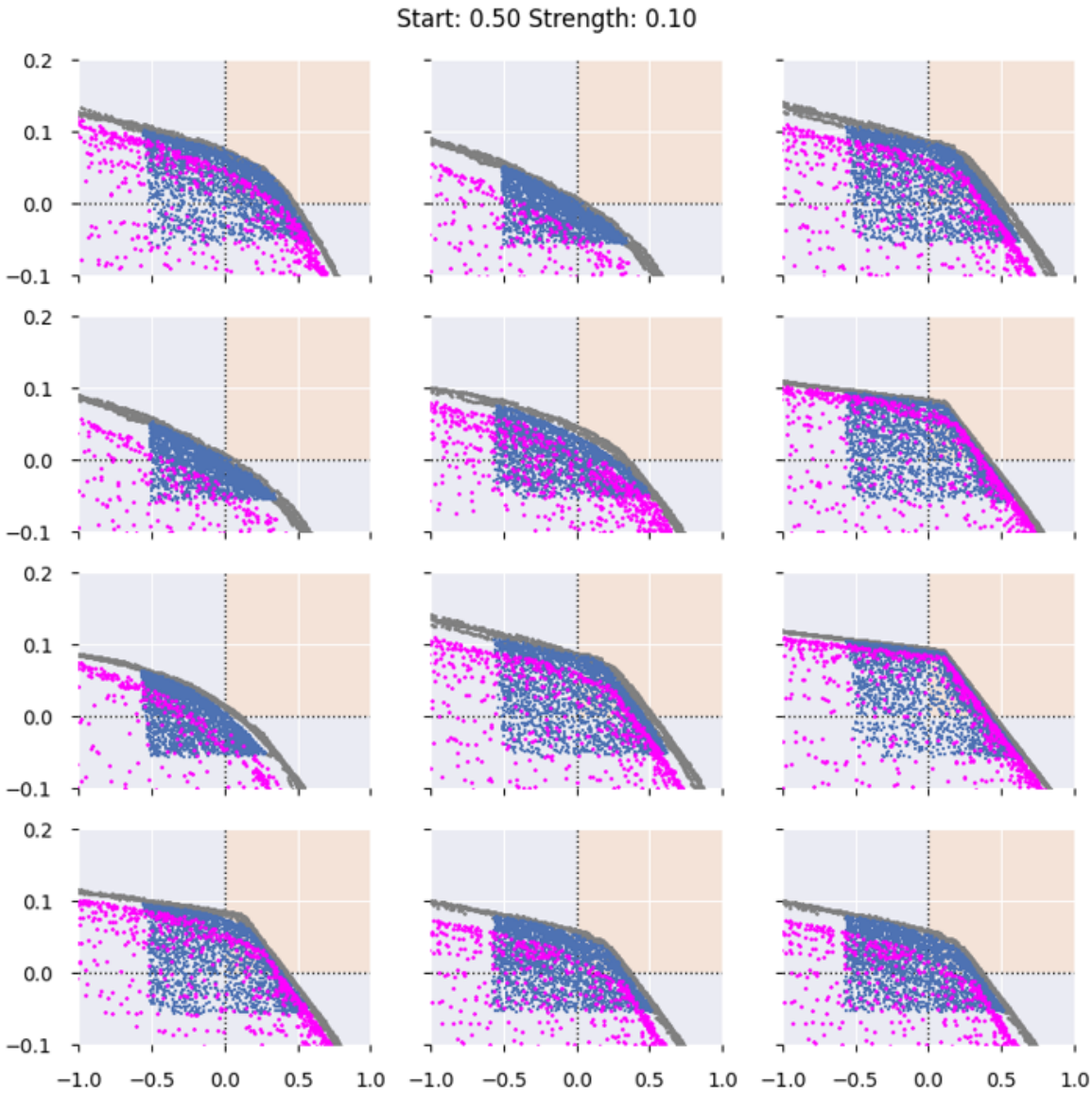

Figure B.37: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

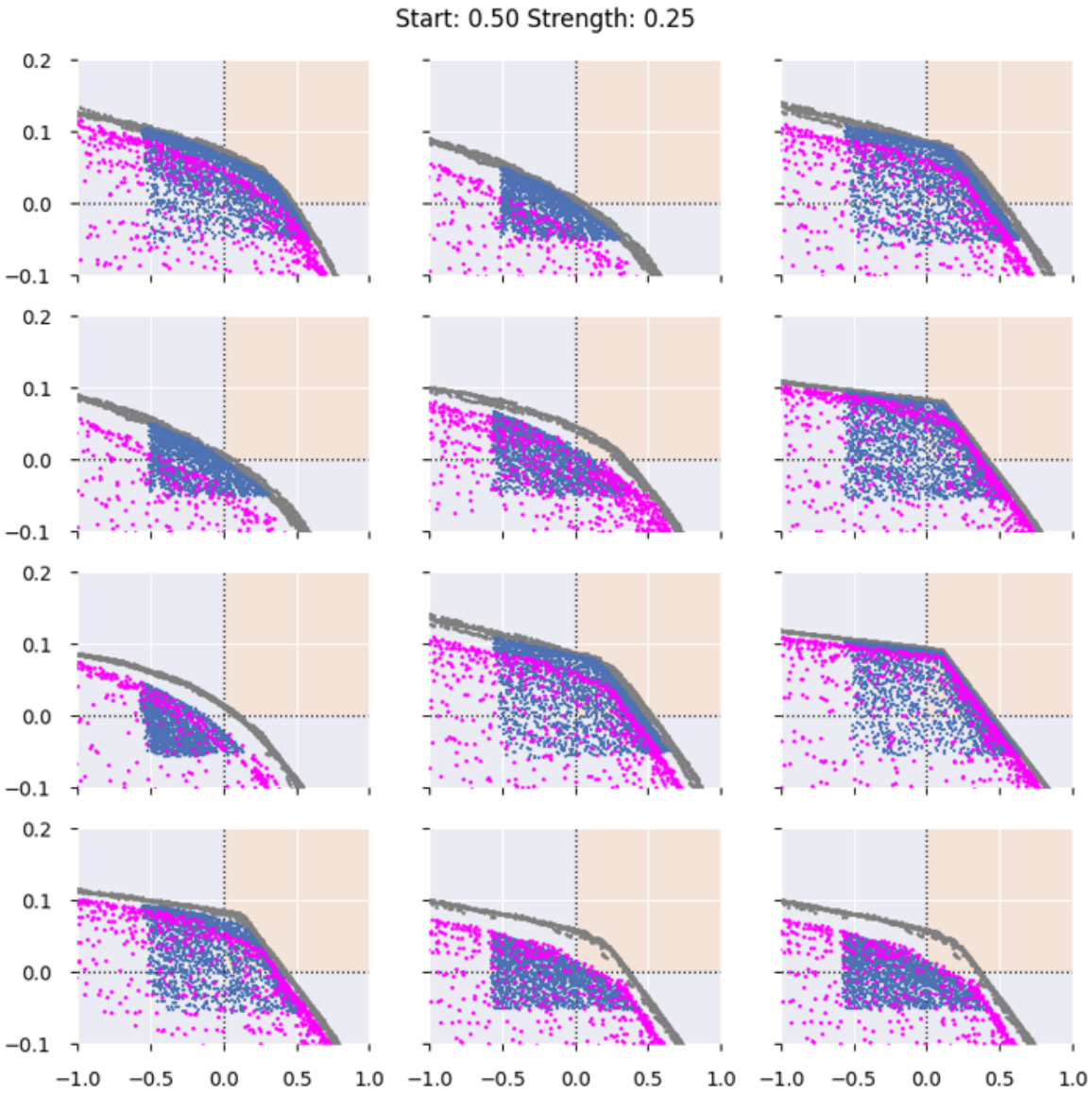

Figure B.38: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

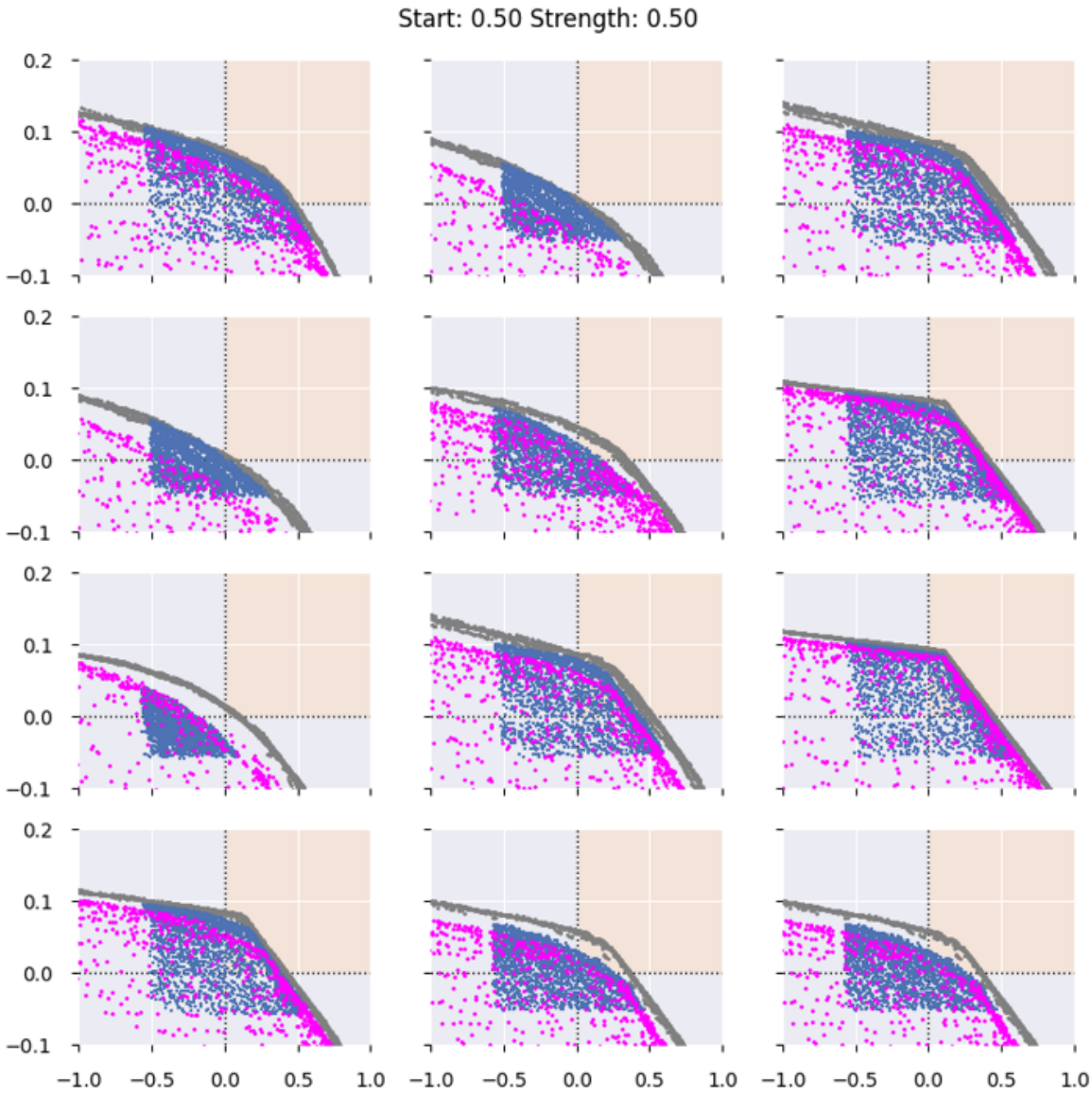

Figure B.39: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

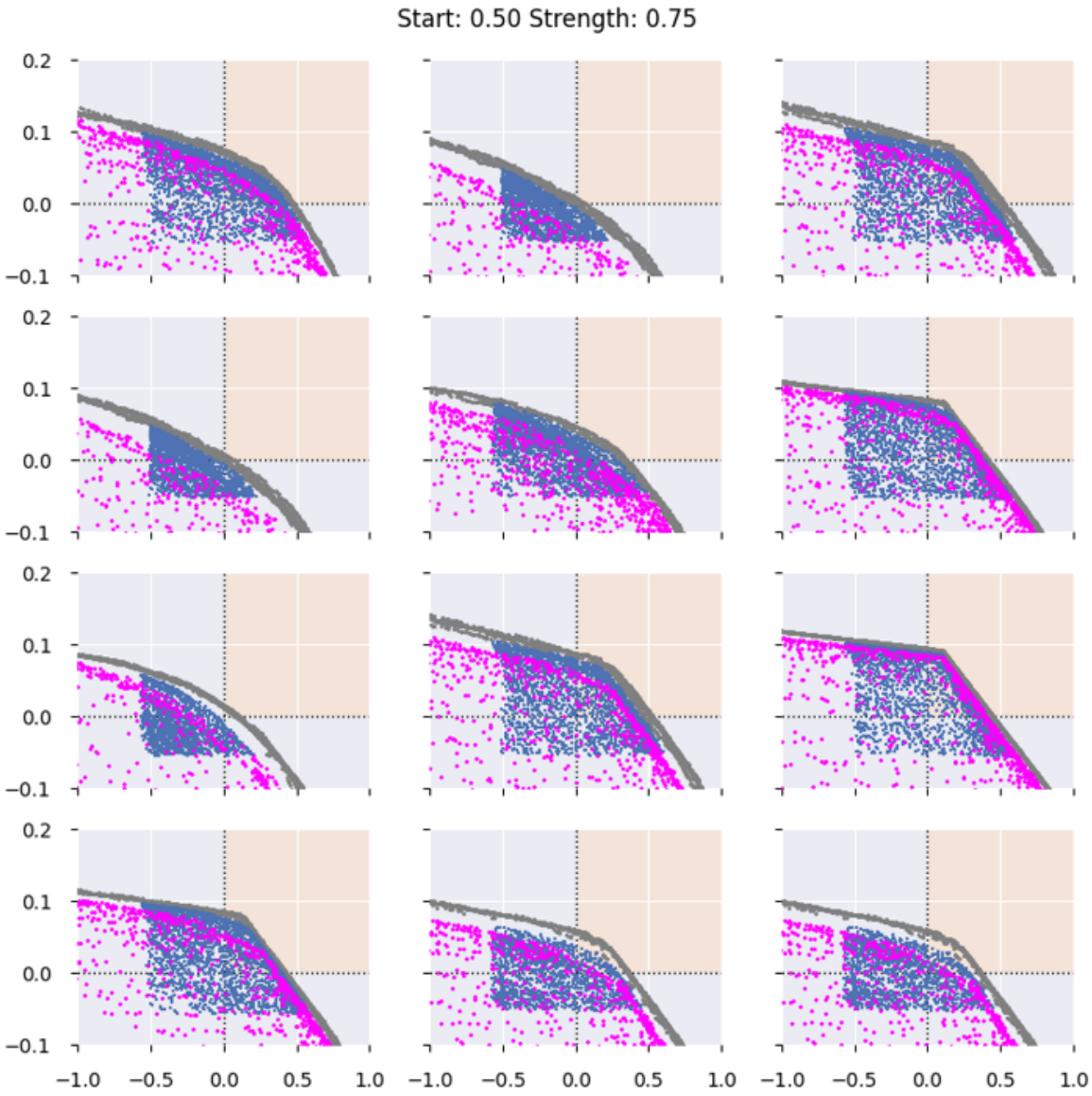

Figure B.40: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

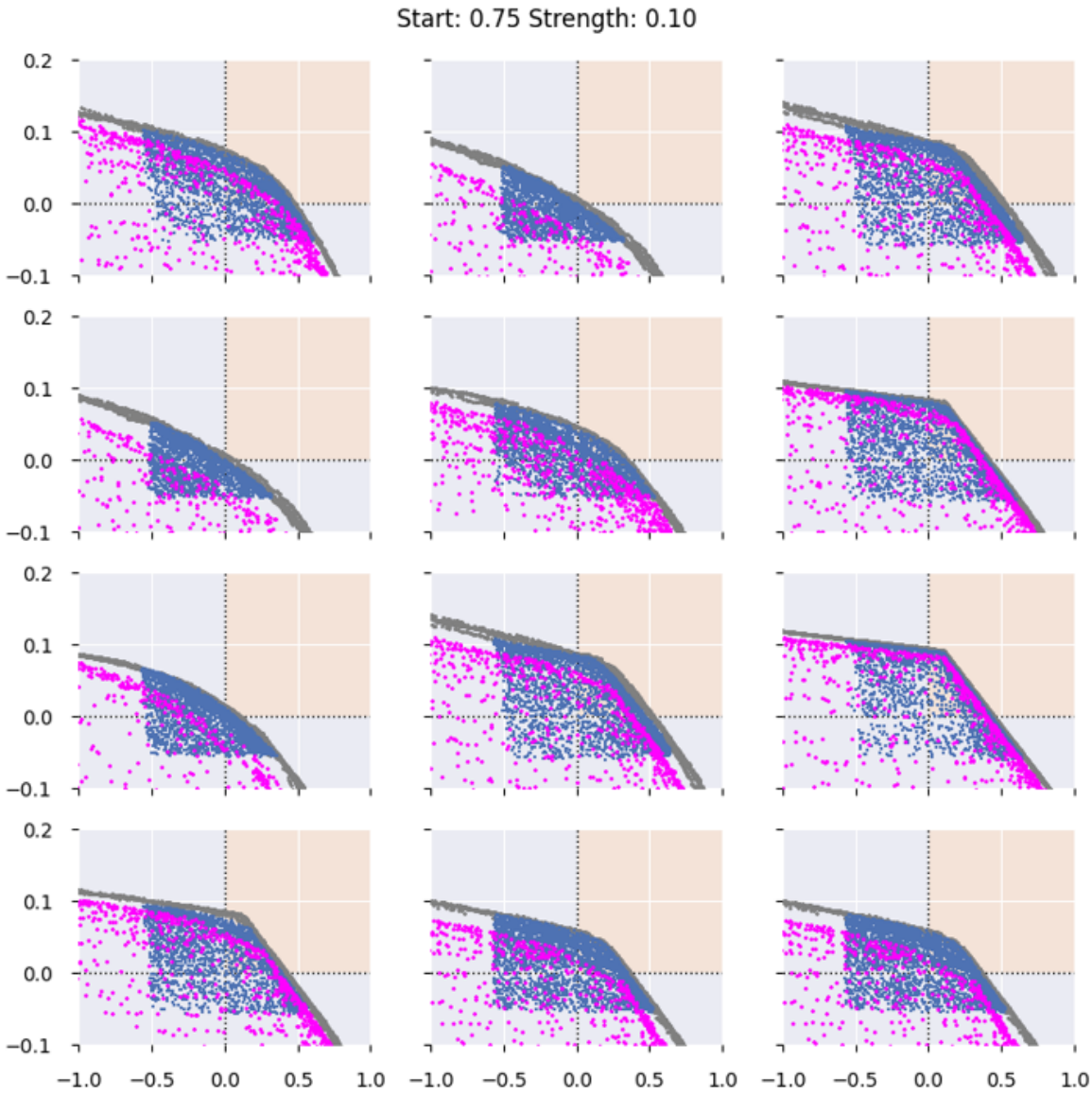

Figure B.41: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

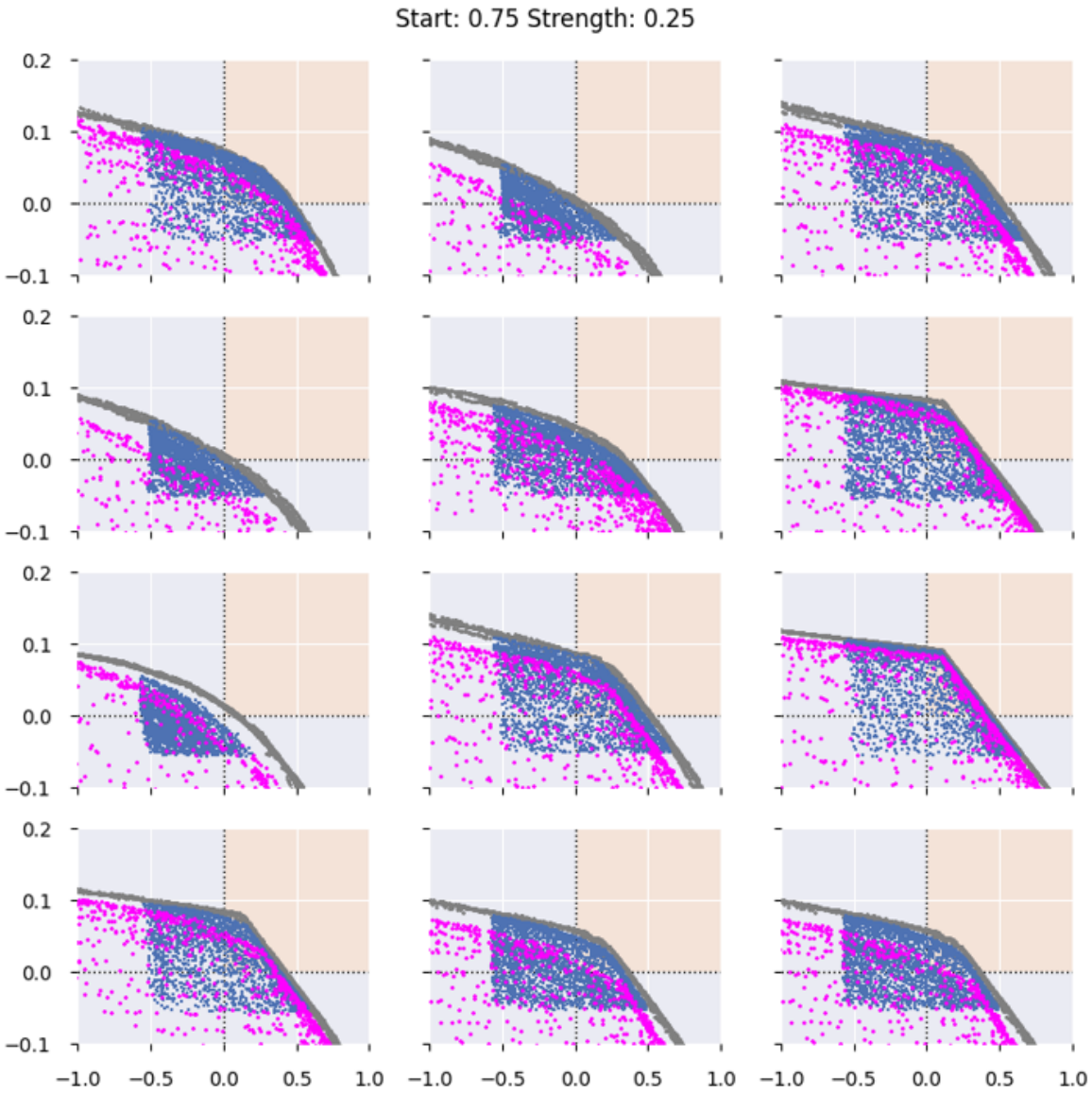

Figure B.42: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

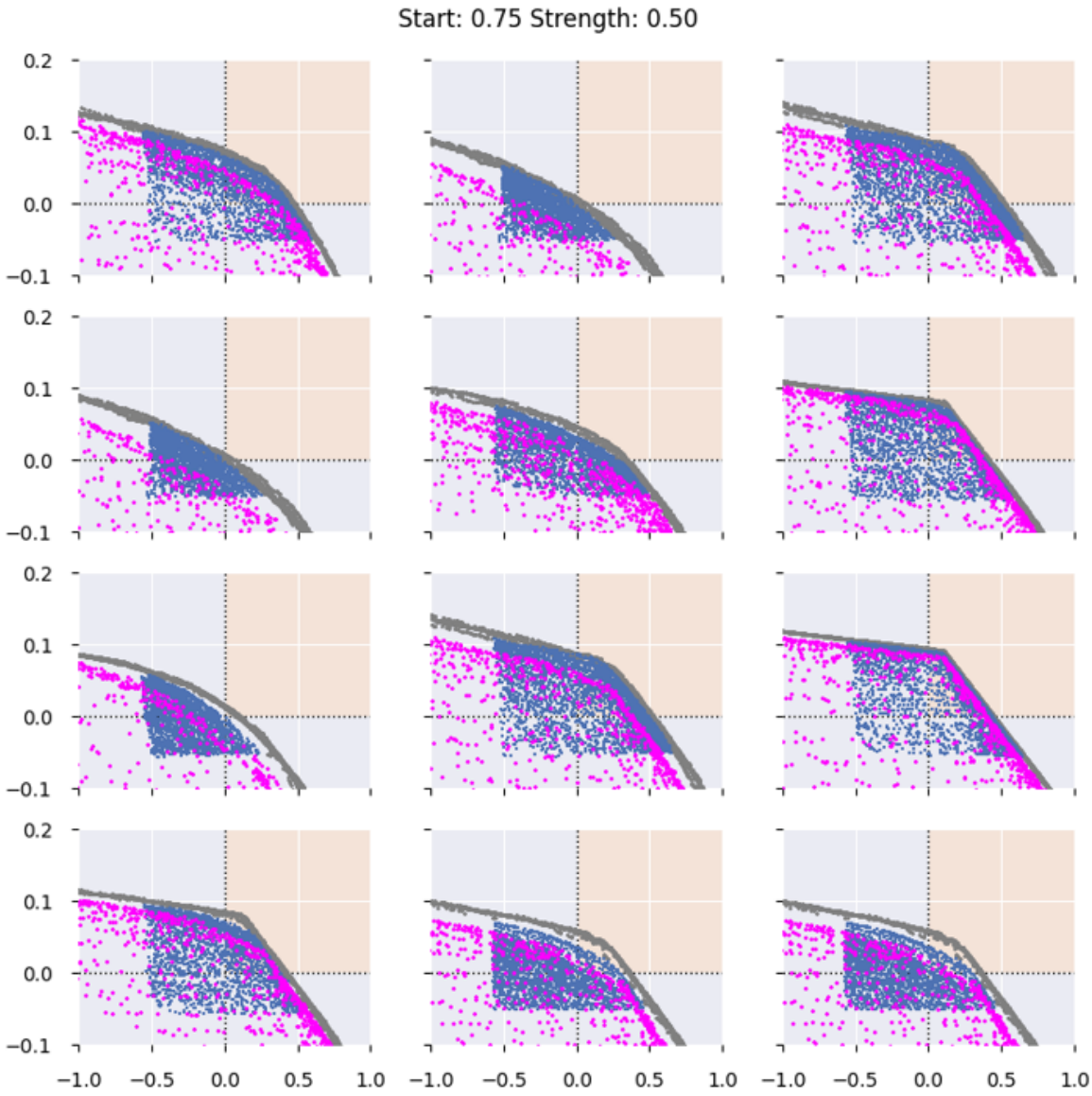

Figure B.43: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

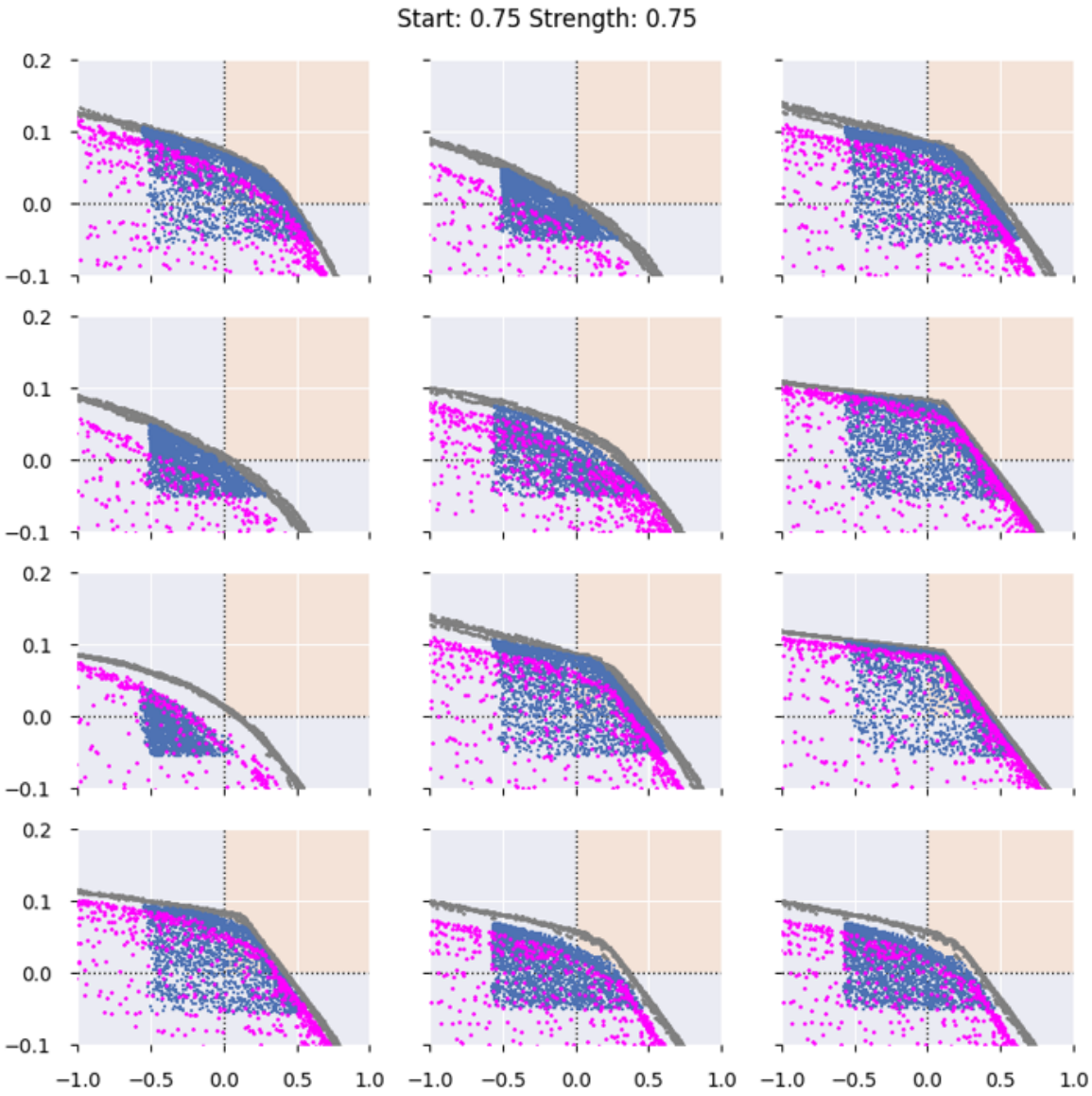

Figure B.44: Blue dots represent treatment plans from the experiment. Grey dots the treatment plans for the basic bi-objective BRIGHT configuration without adaptive steering. Pink dots the tri-objective BRIGHT results without adaptive steering In the title of the figure it states what the used setting for adaptive steering were. The order of the patients is similar as in other figures.

B.4. Basic BRIGHT configuration results

In this appendix section all Pareto approximation fronts are shown for the basic BRIGHT configuration.

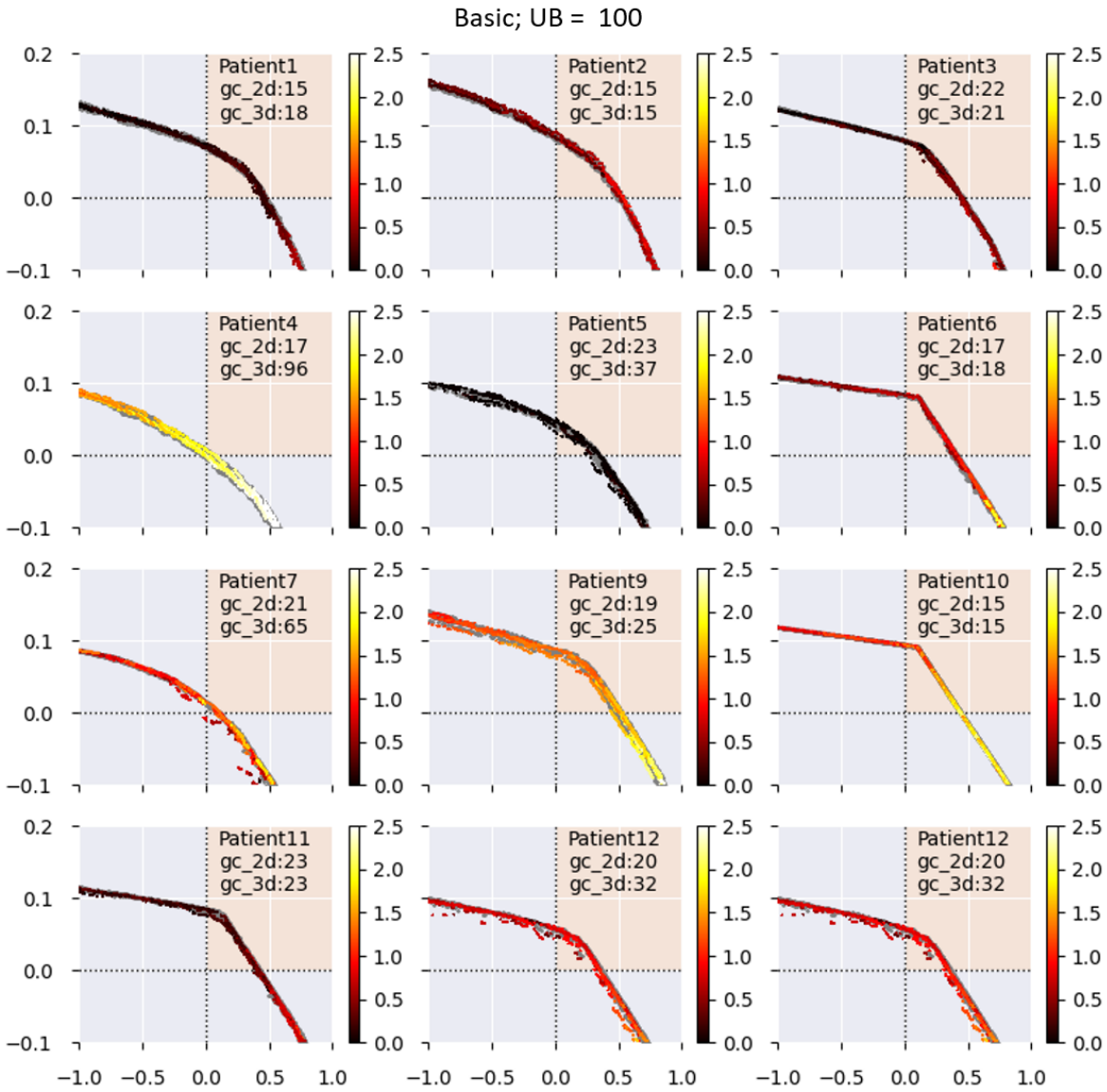

Figure B.45: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

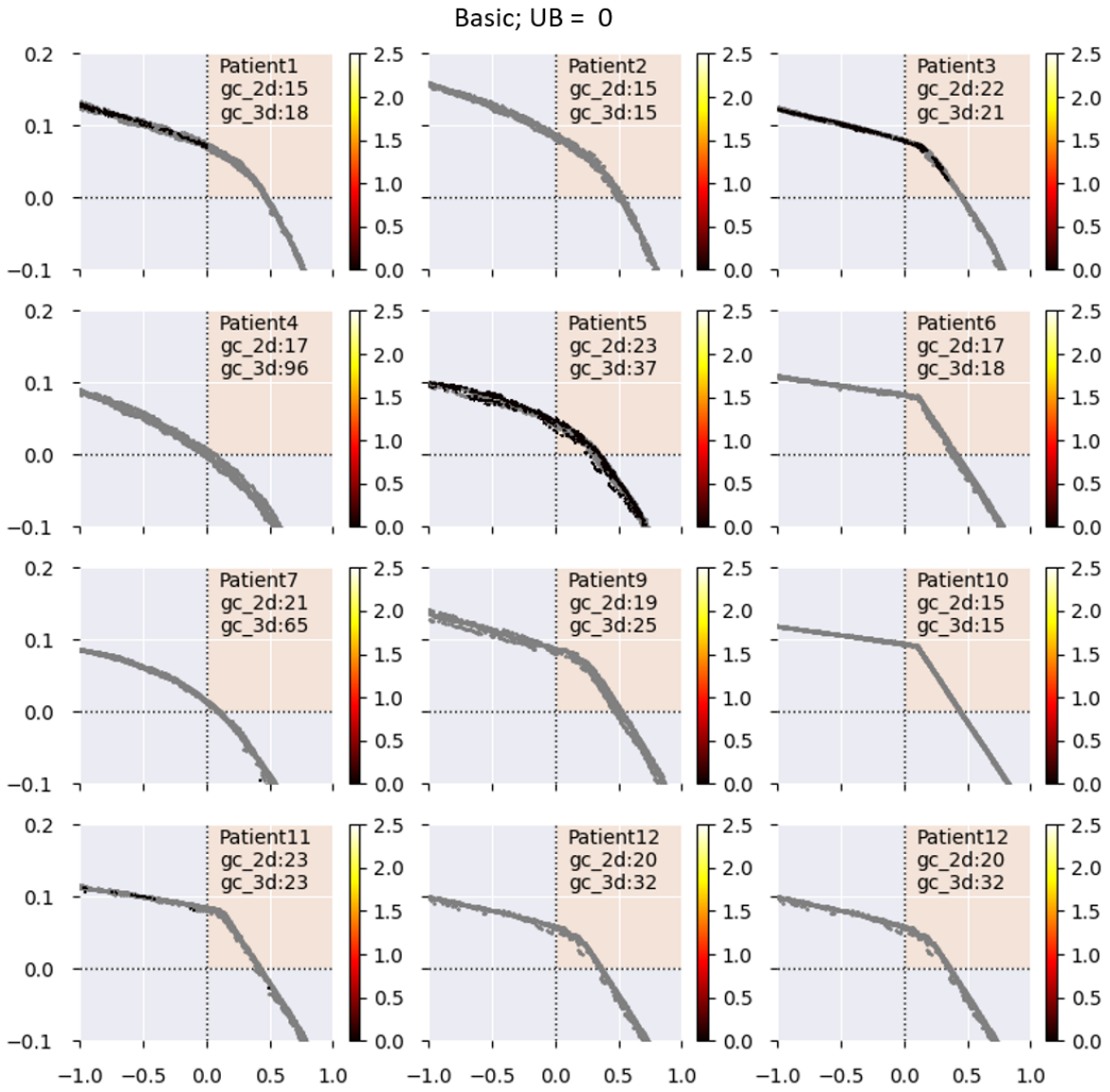

Figure B.46: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

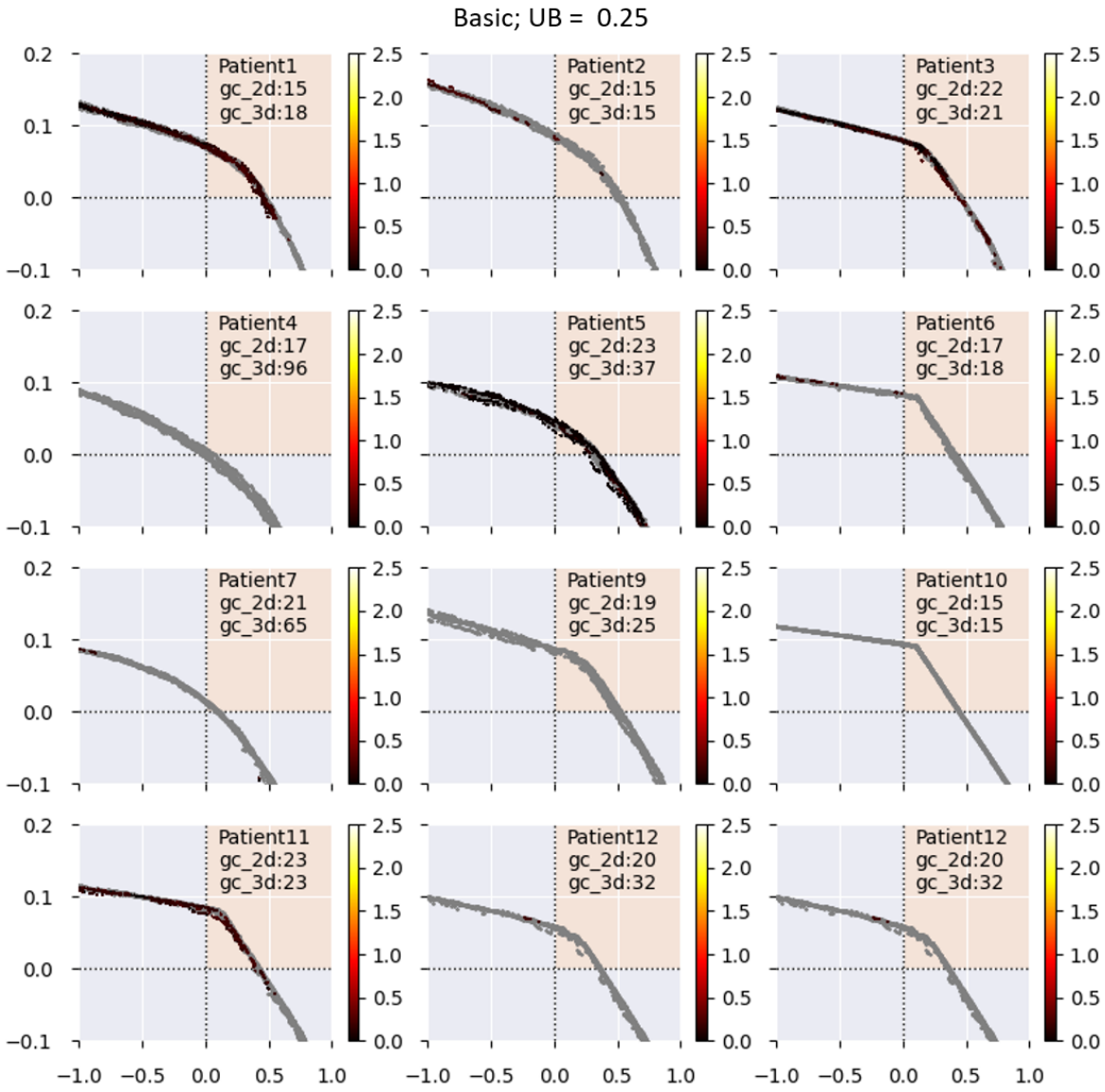

Figure B.47: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

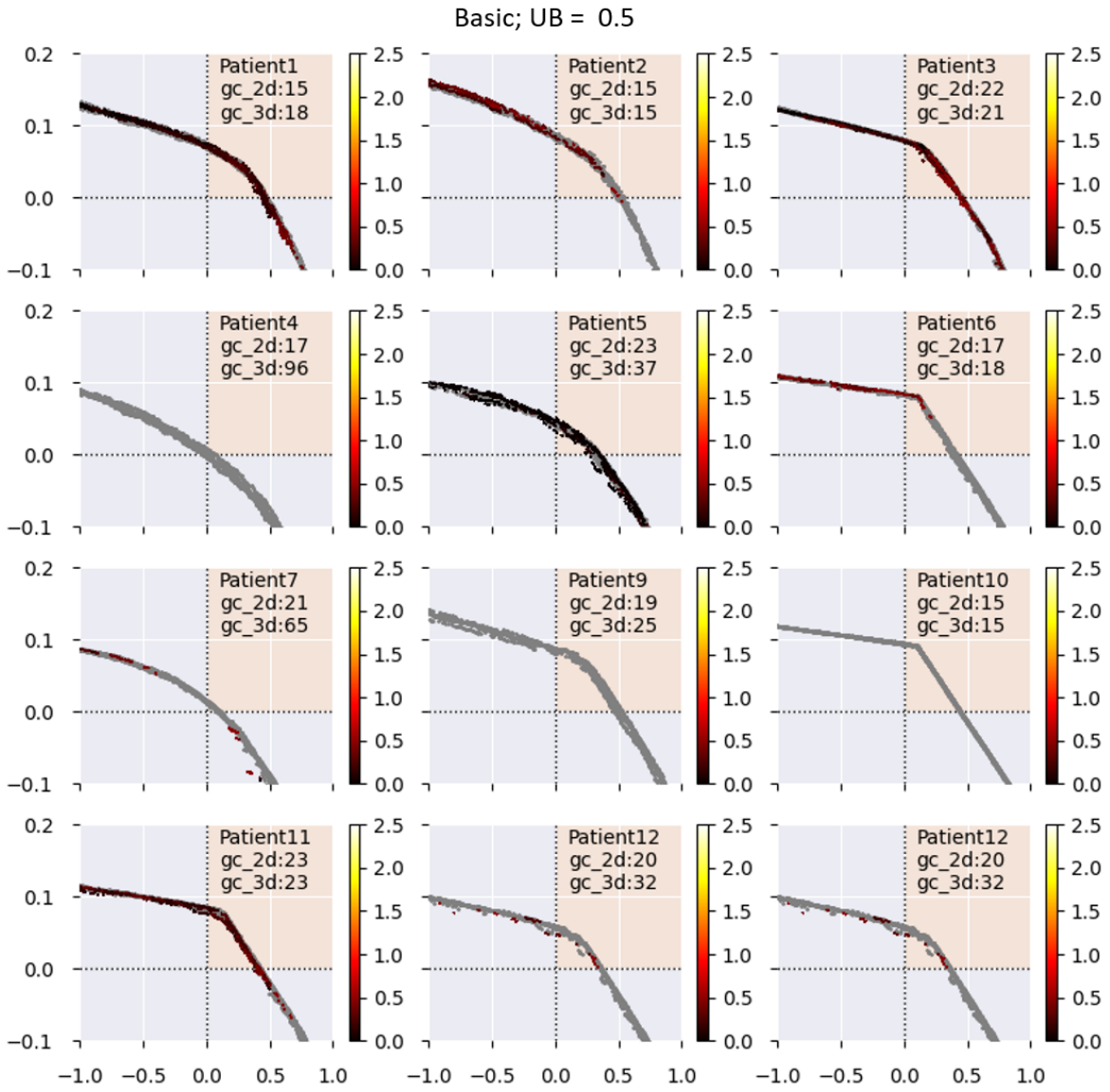

Figure B.48: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

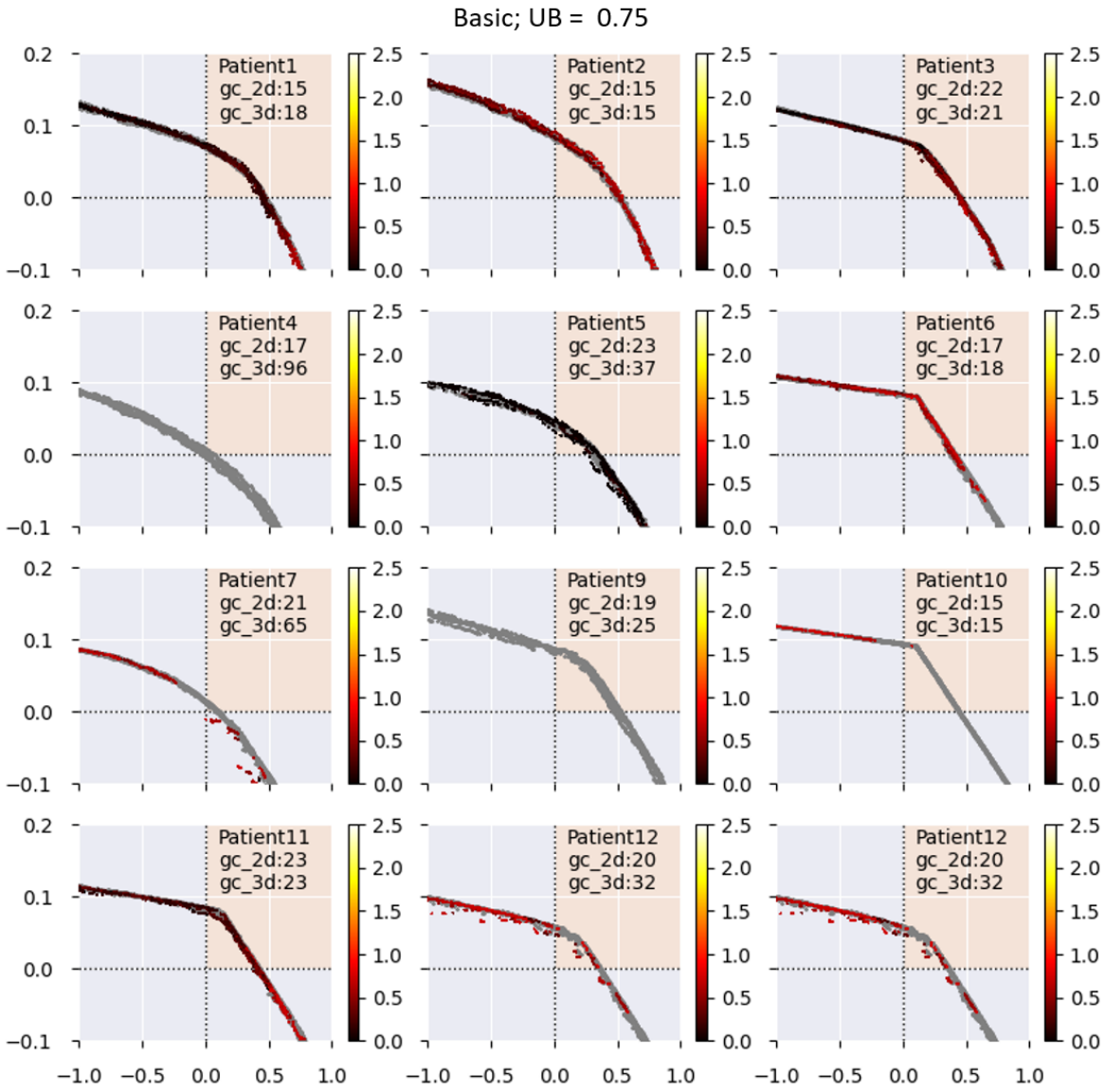

Figure B.49: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

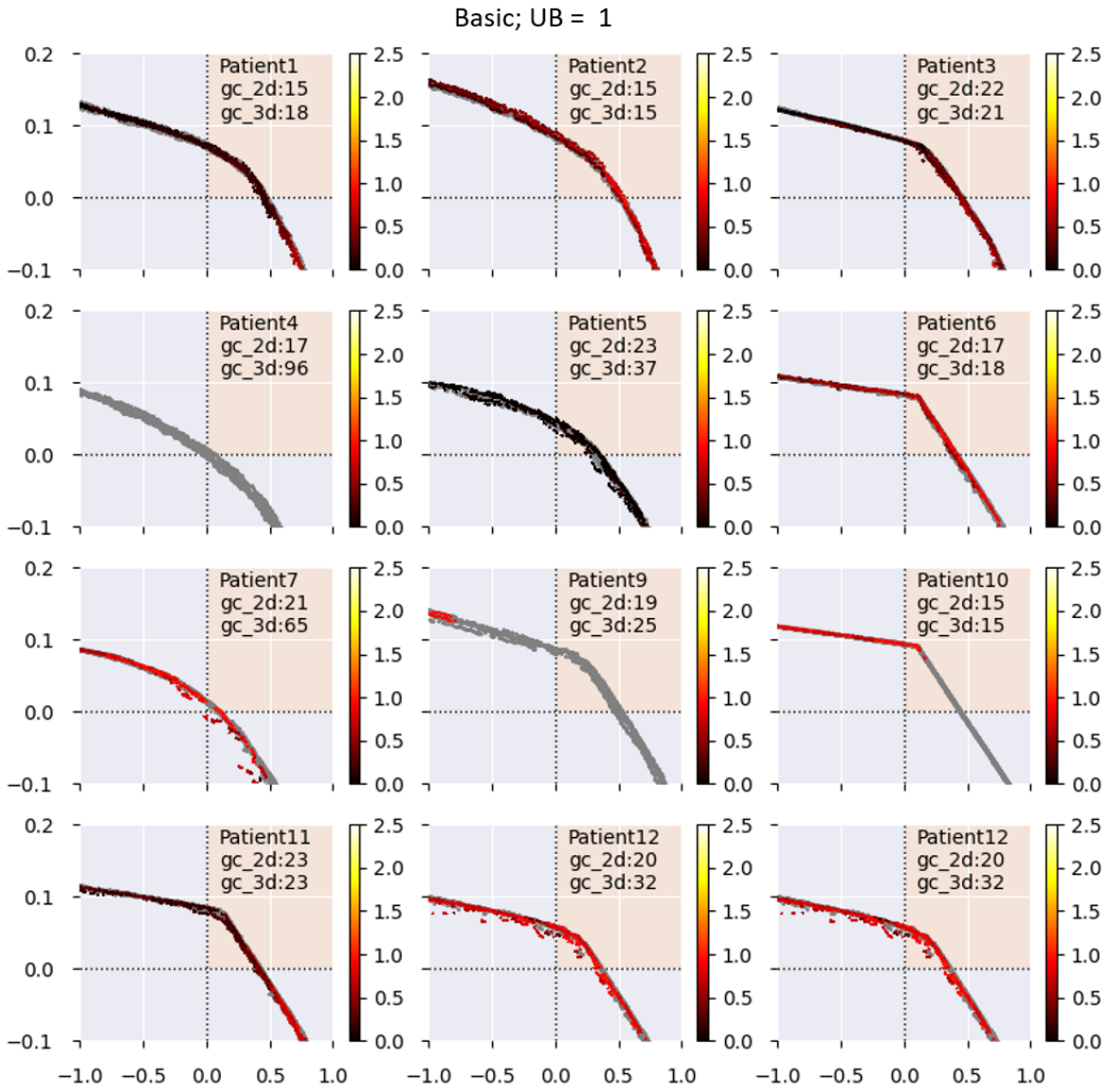

Figure B.50: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

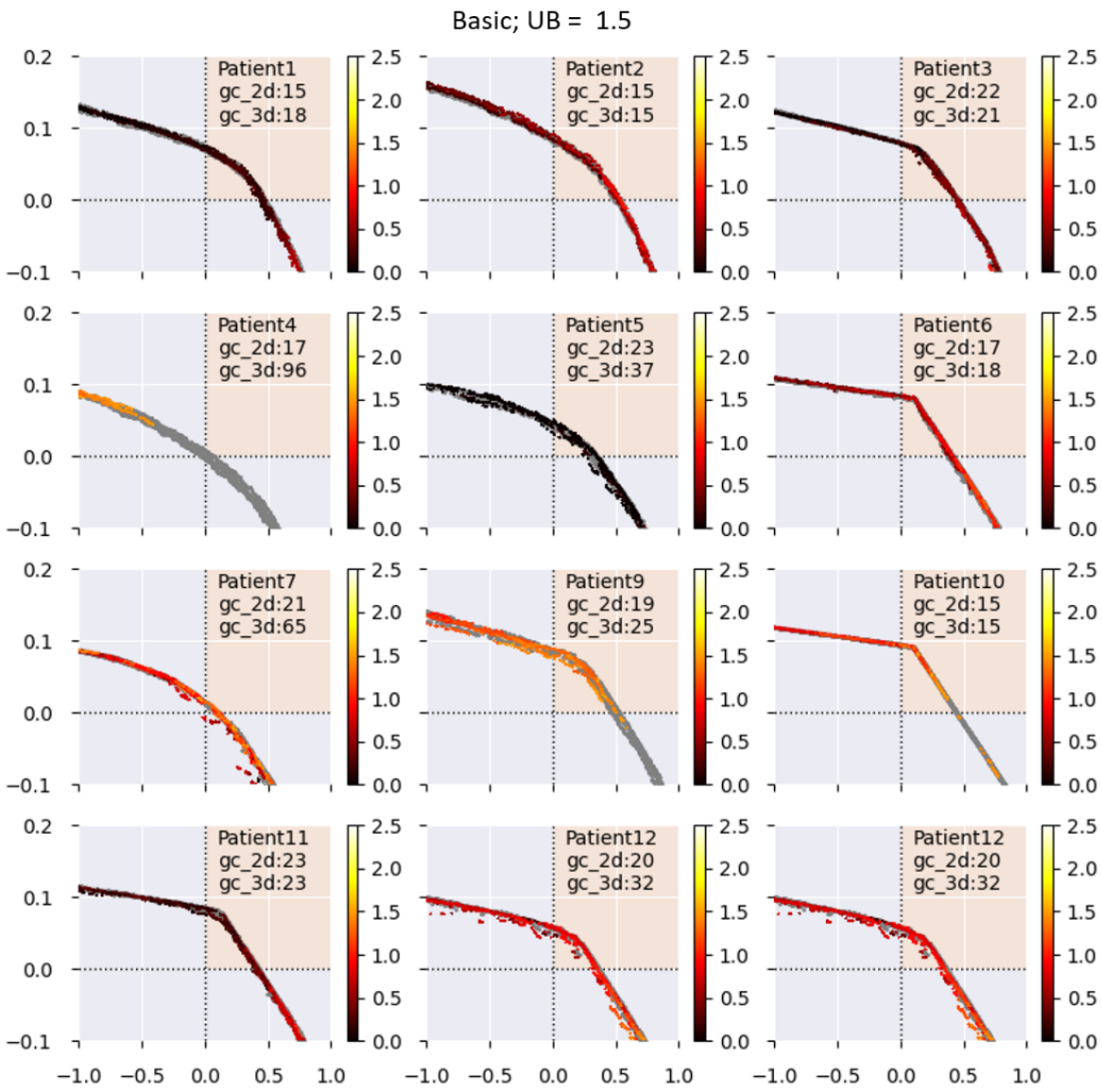

Figure B.51: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

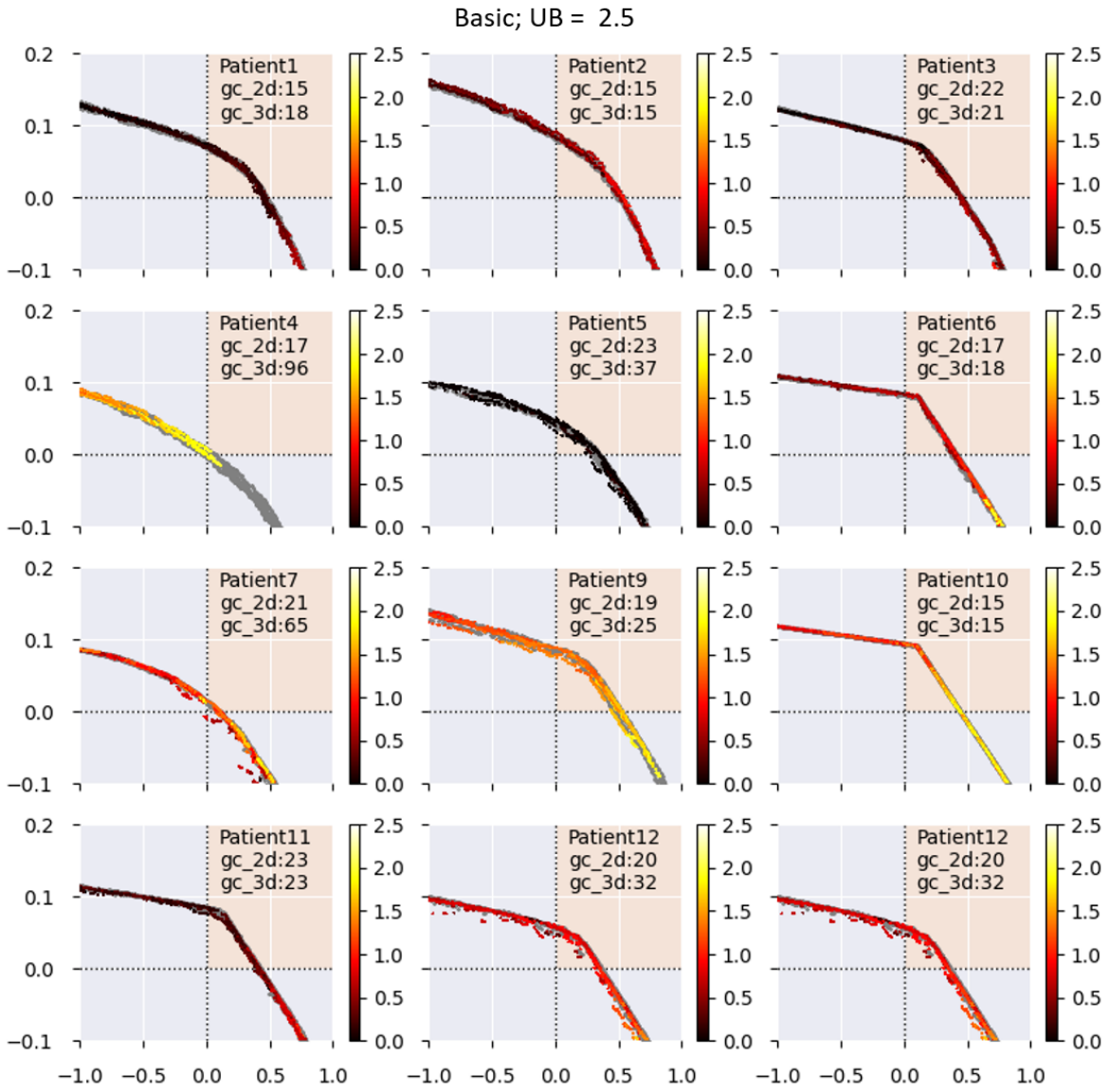

Figure B.52: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

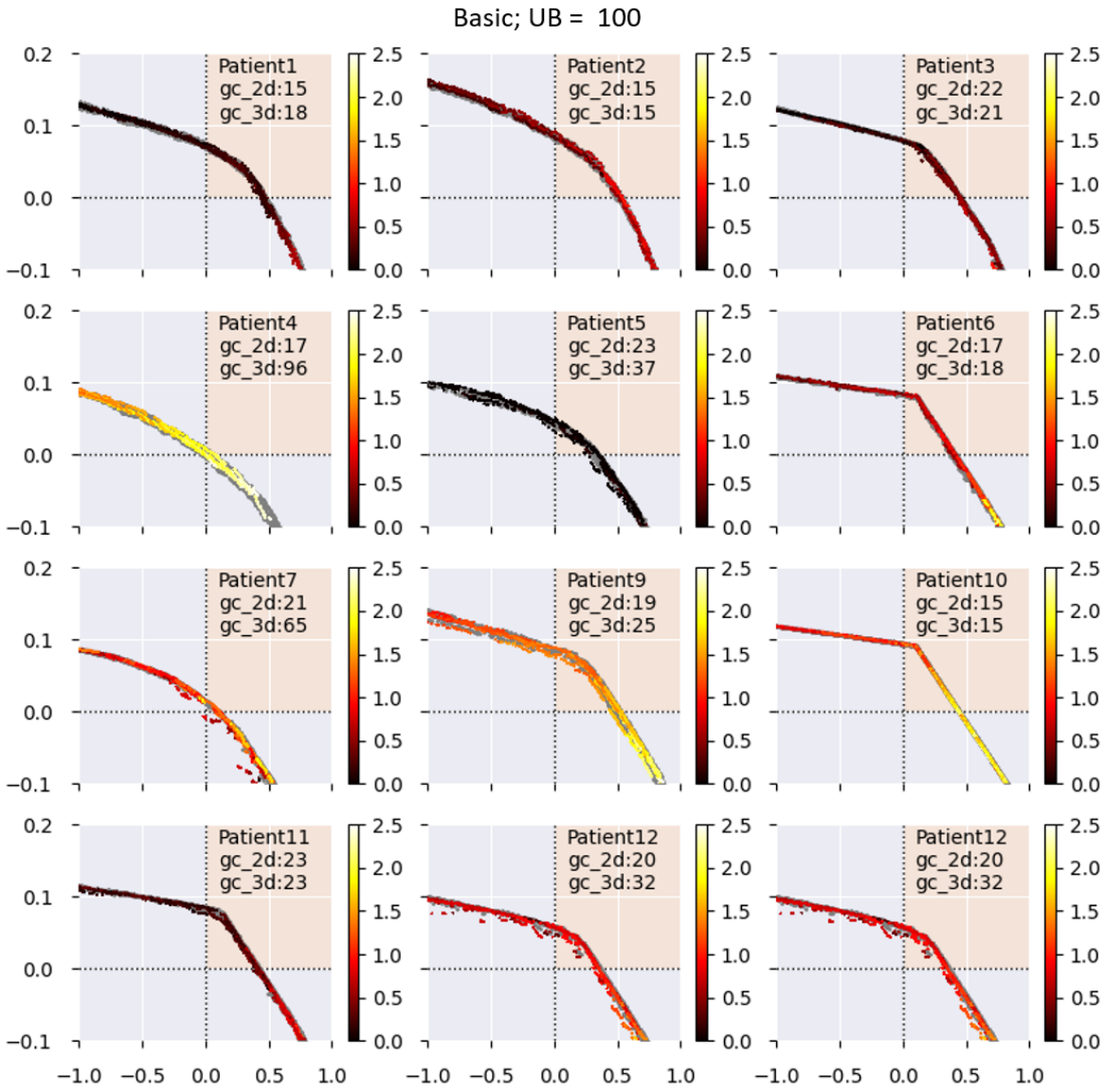

Figure B.53: Pareto approximation fronts resulting from bi-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT without adaptive steering. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and the value for 3D can be ignored.

B.5. HSI results

In this appendix section all Pareto approximation fronts are shown for the HSI as third objective.

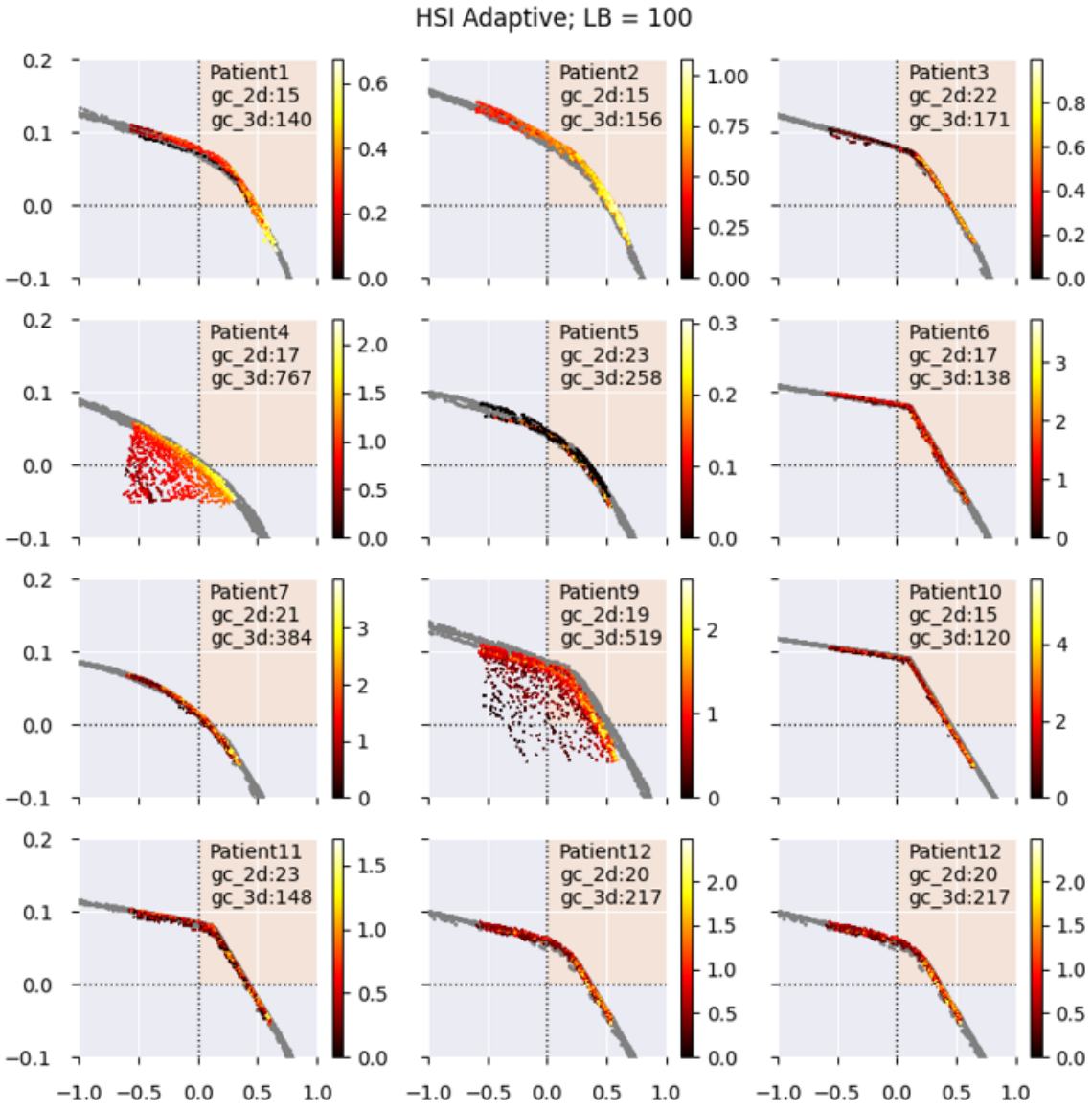

Figure B.54: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

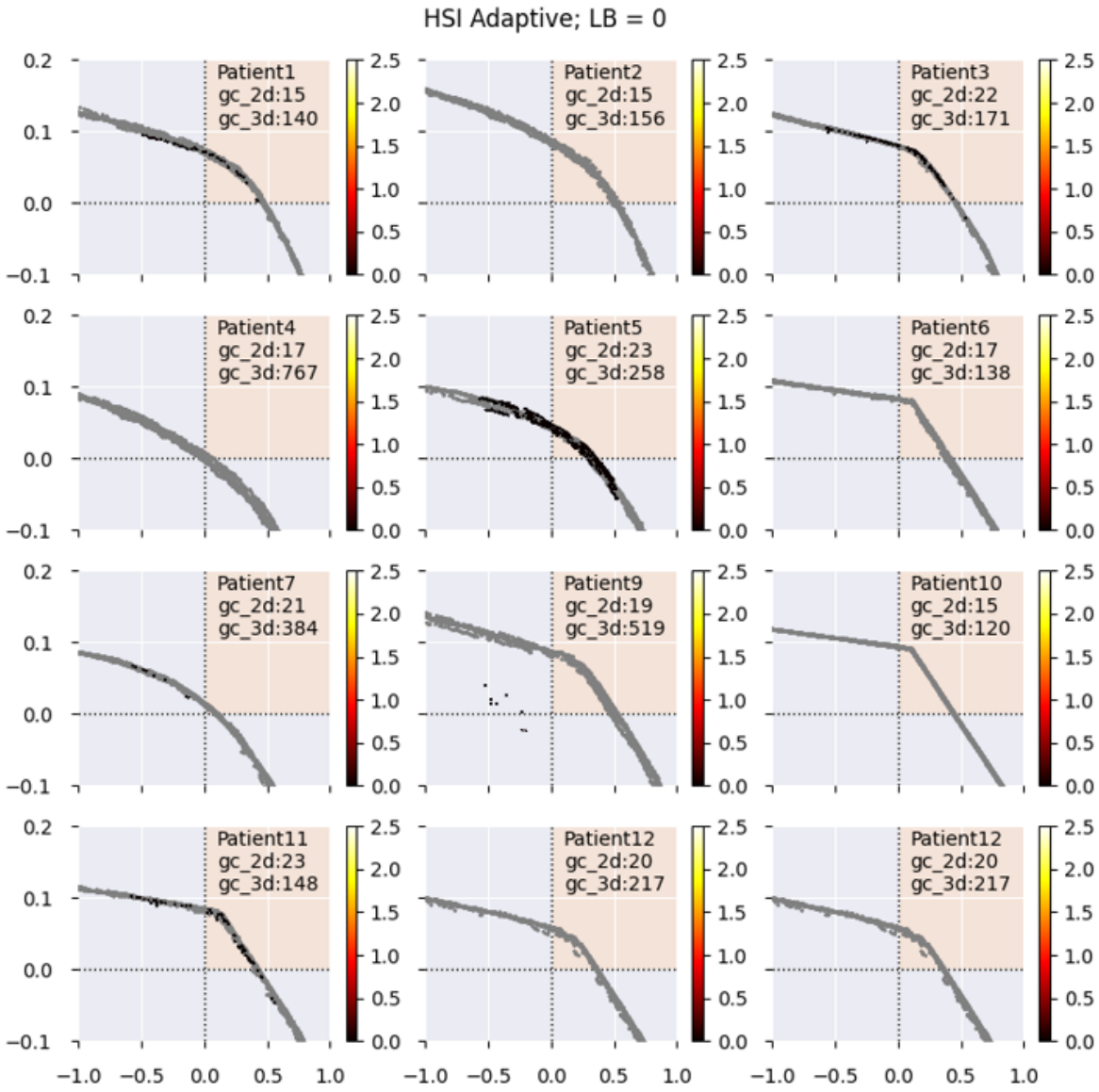

Figure B.55: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

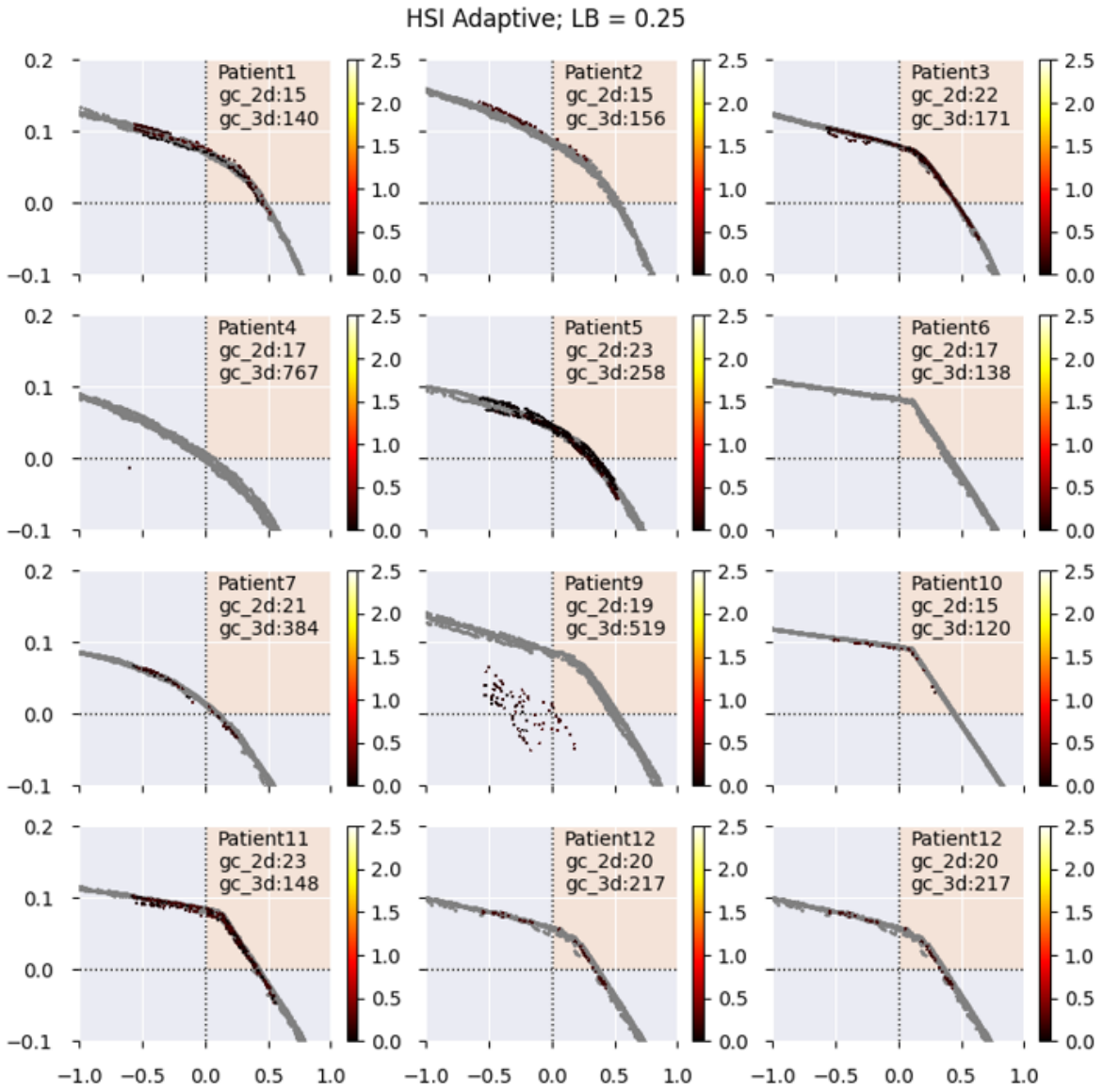

Figure B.56: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

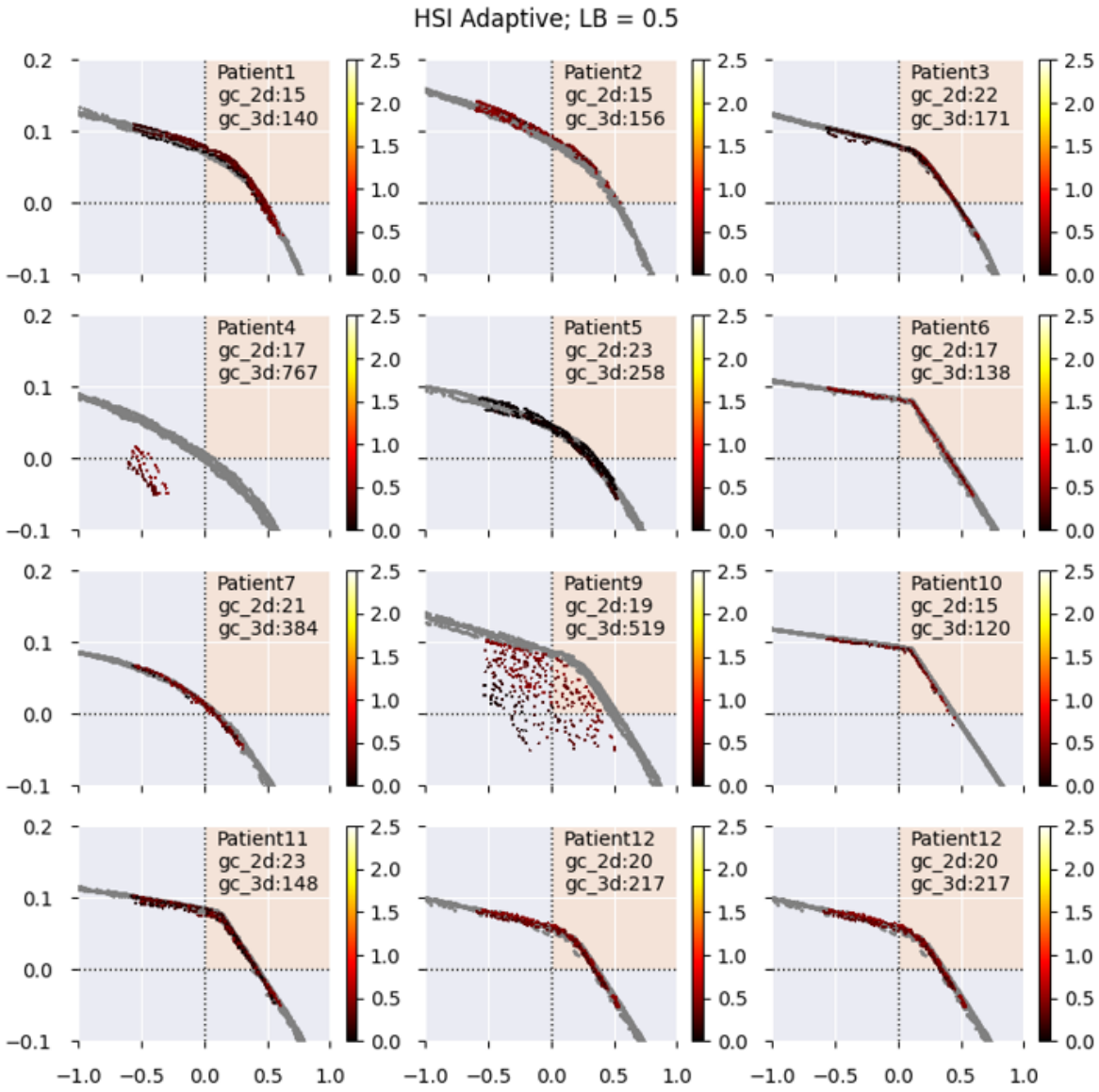

Figure B.57: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

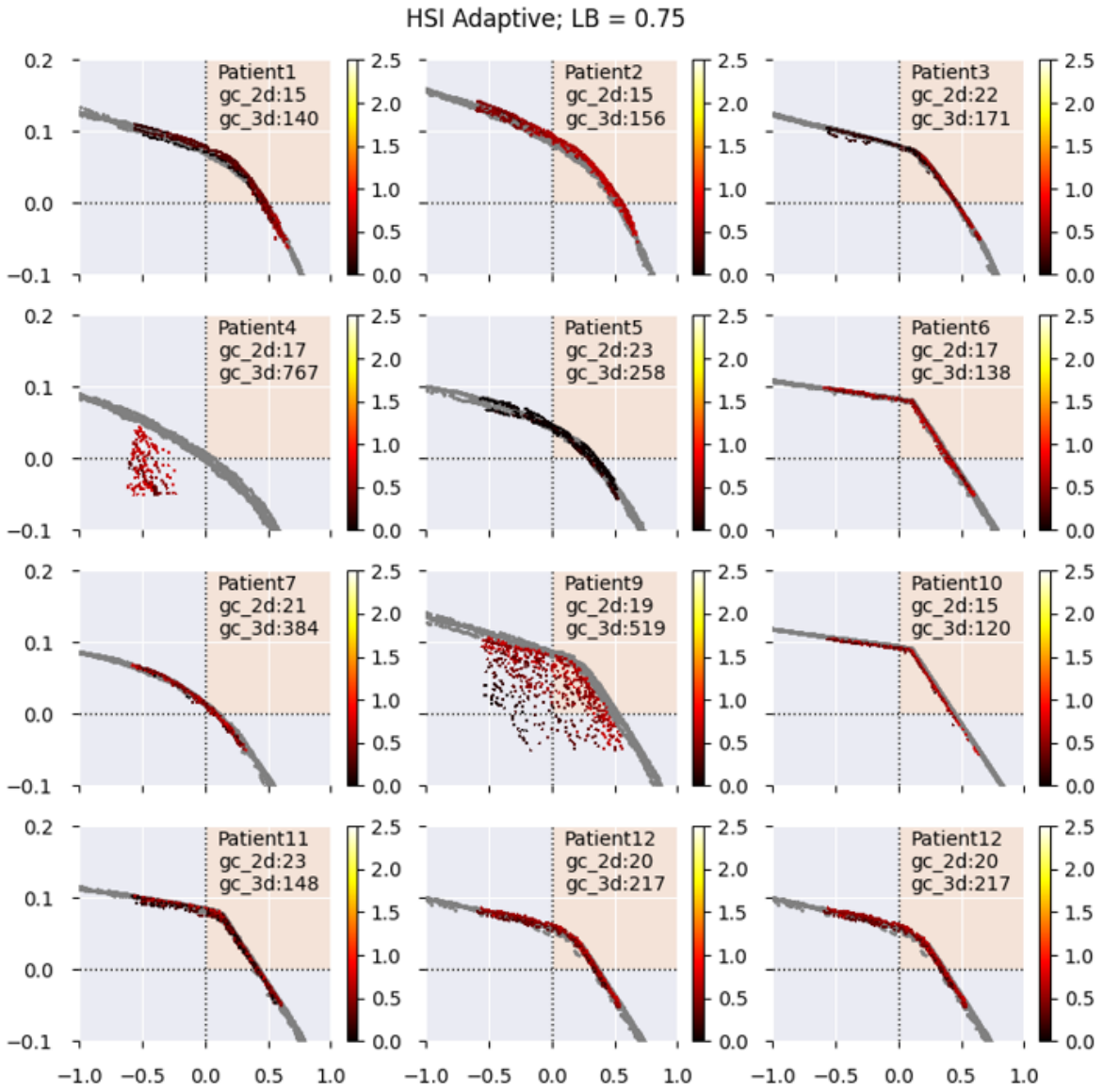

Figure B.58: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

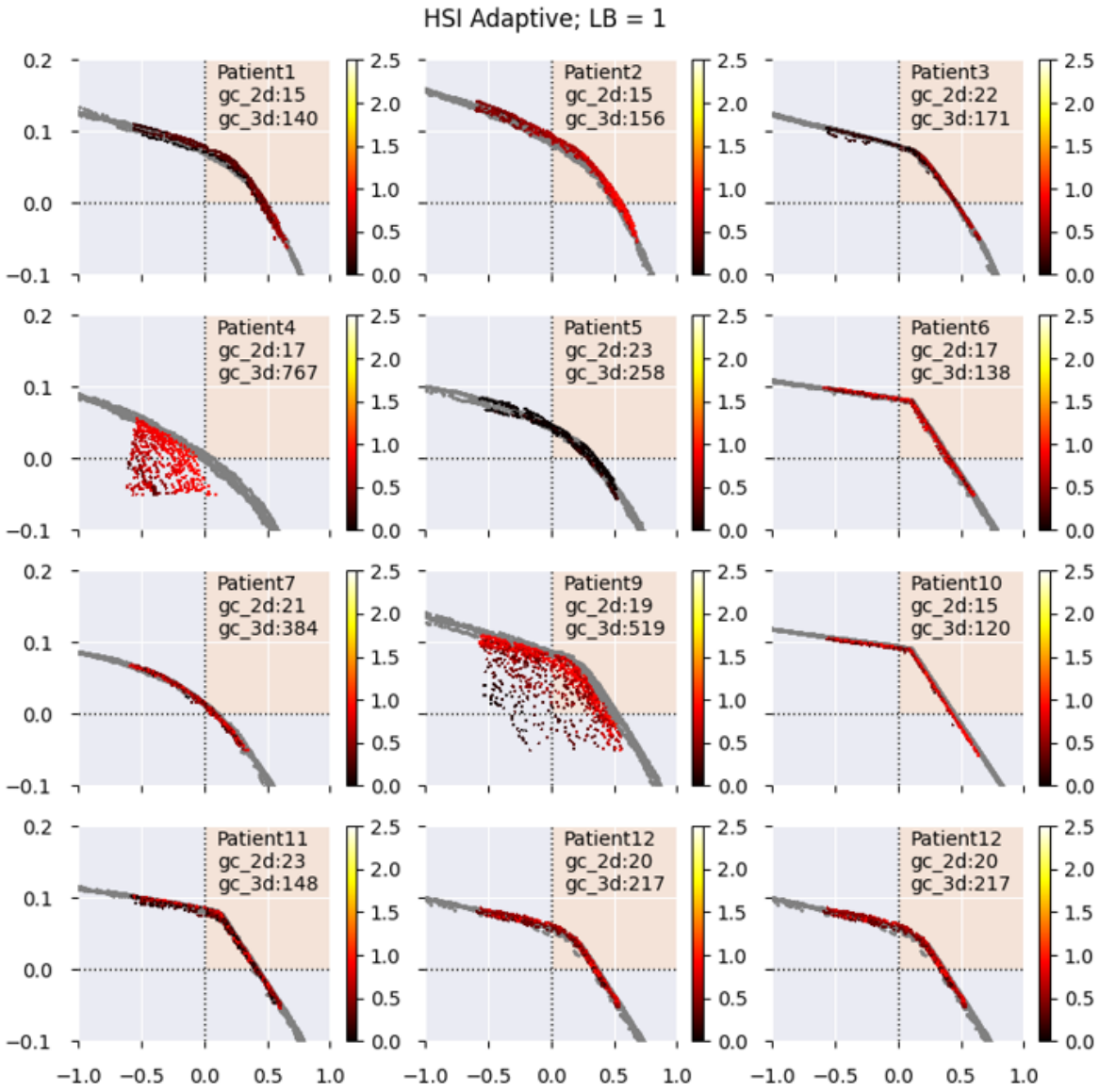

Figure B.59: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

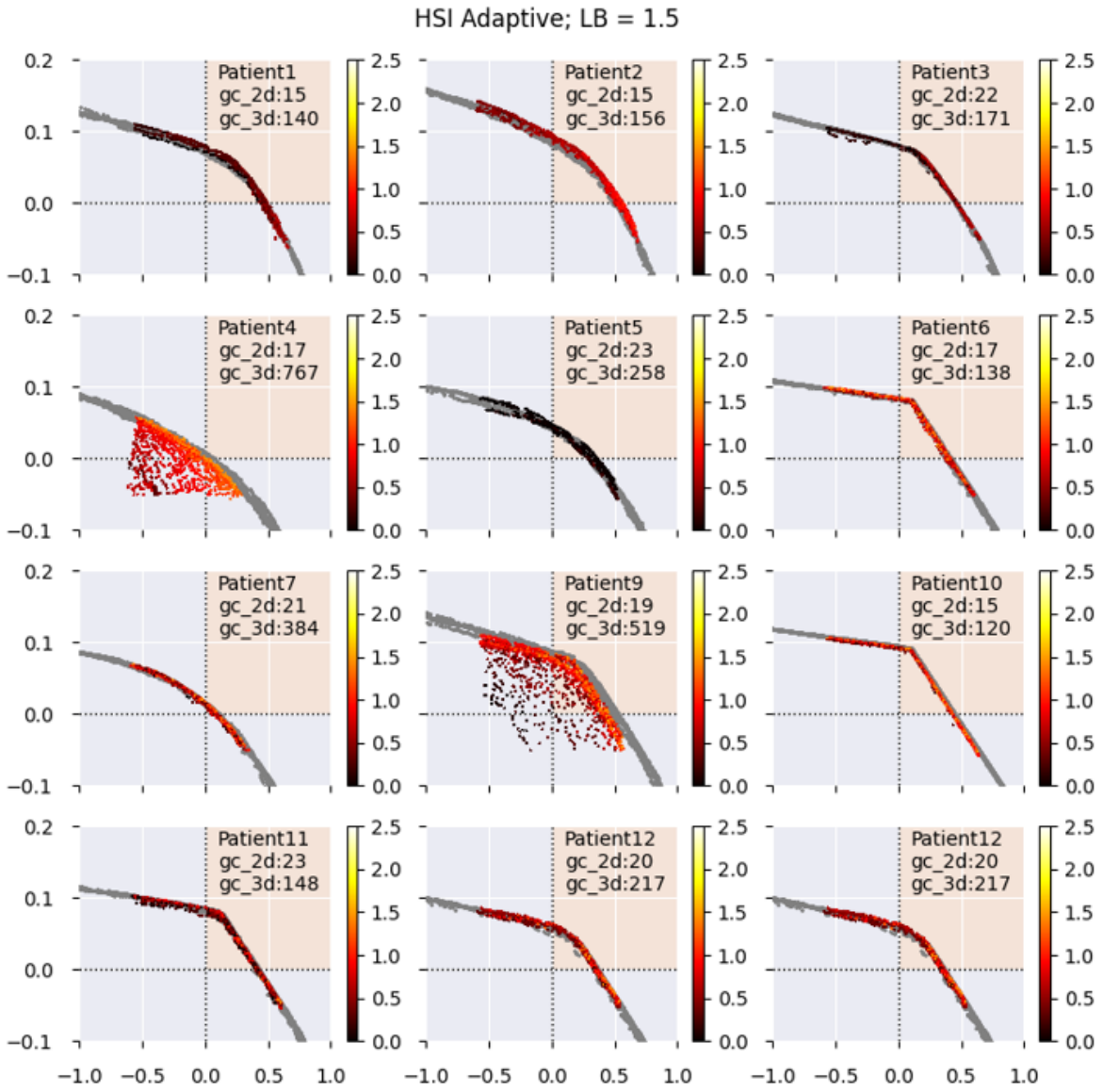

Figure B.60: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

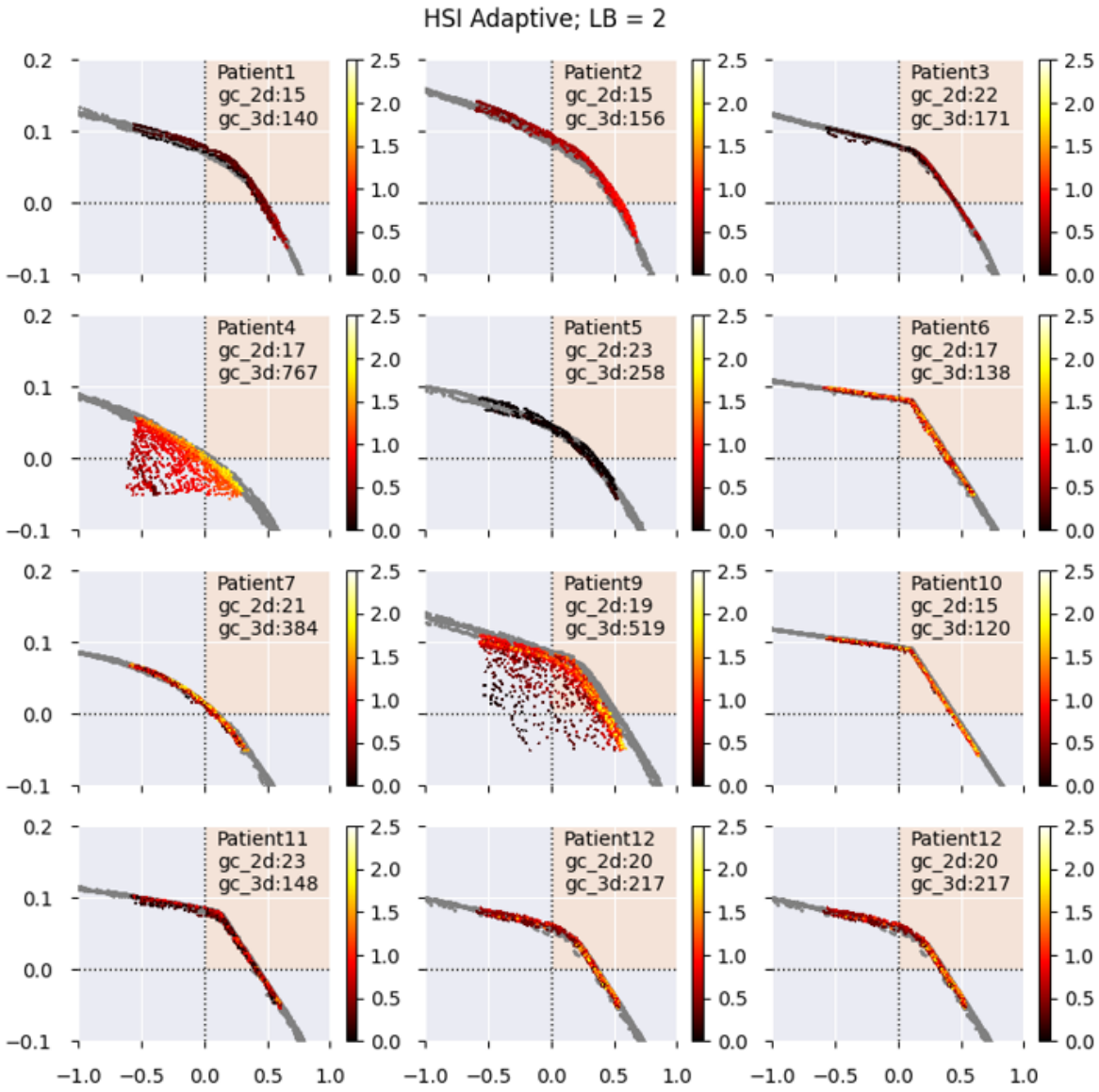

Figure B.61: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

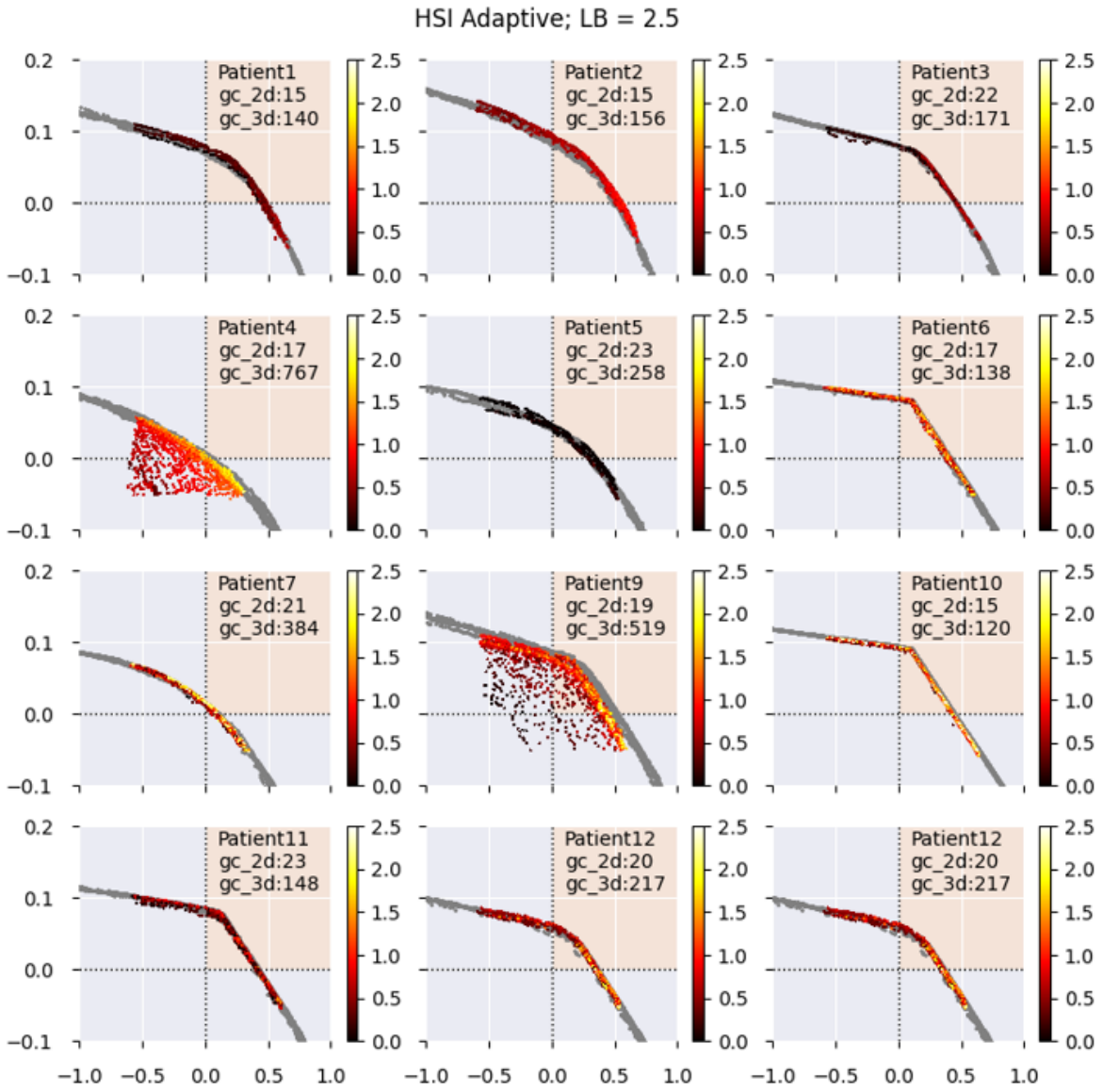

Figure B.62: Pareto approximation fronts resulting from tri-objective BRIGHT with HSI as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

B.6. Sum of extra V indices results

In this appendix section all Pareto approximation fronts are shown for the sum of extra V indices as third objective.

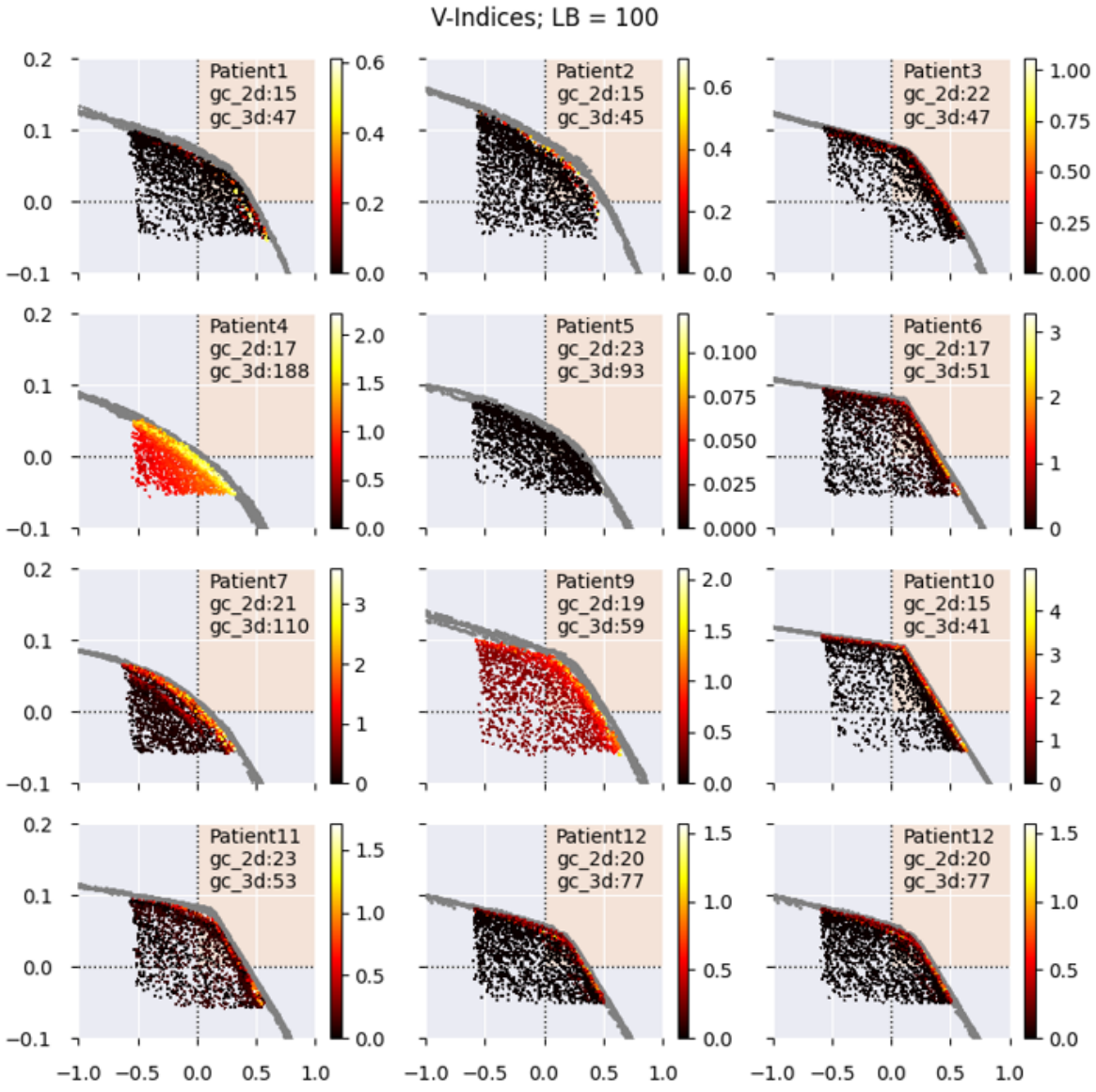

Figure B.63: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

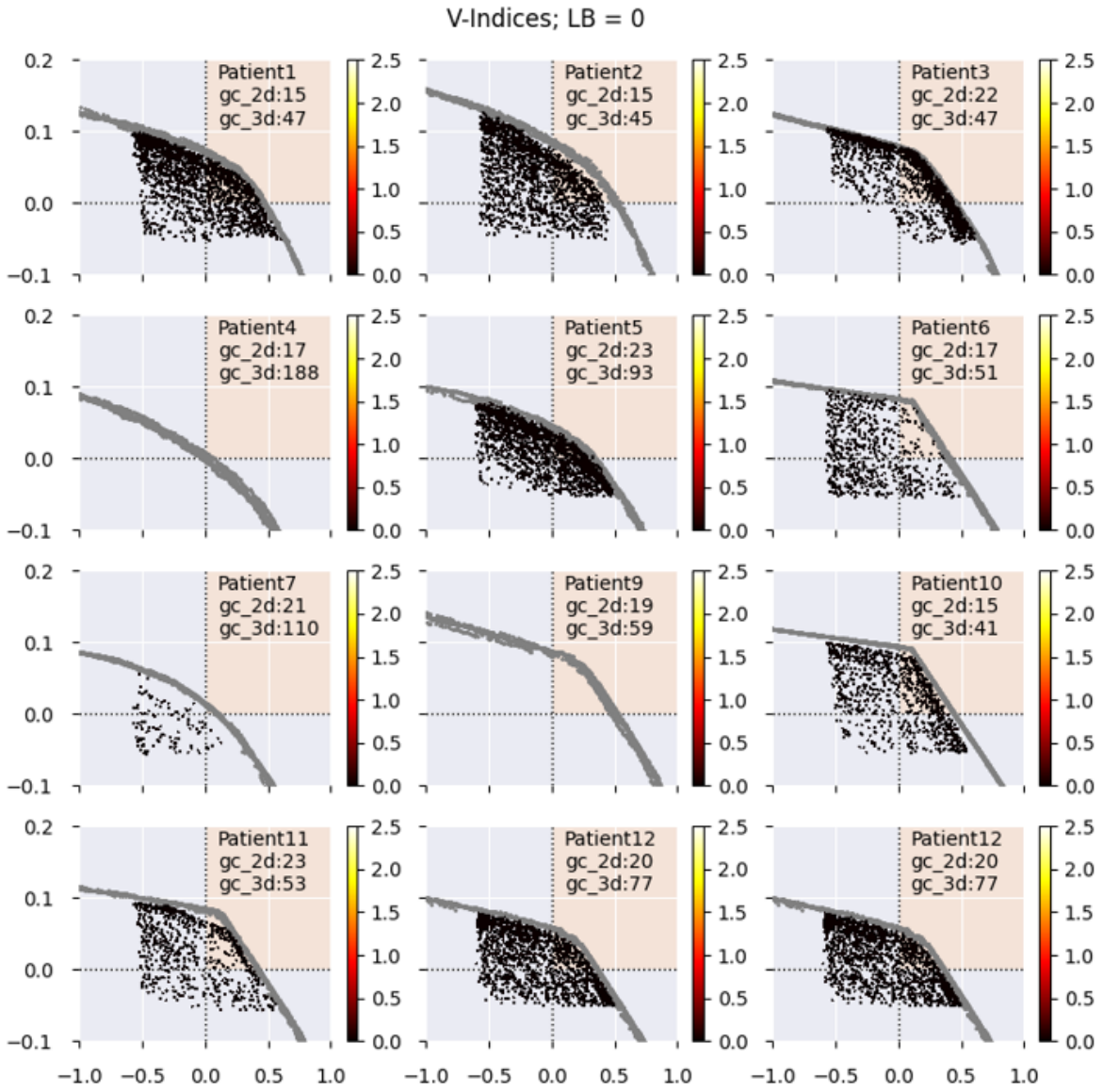

Figure B.64: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

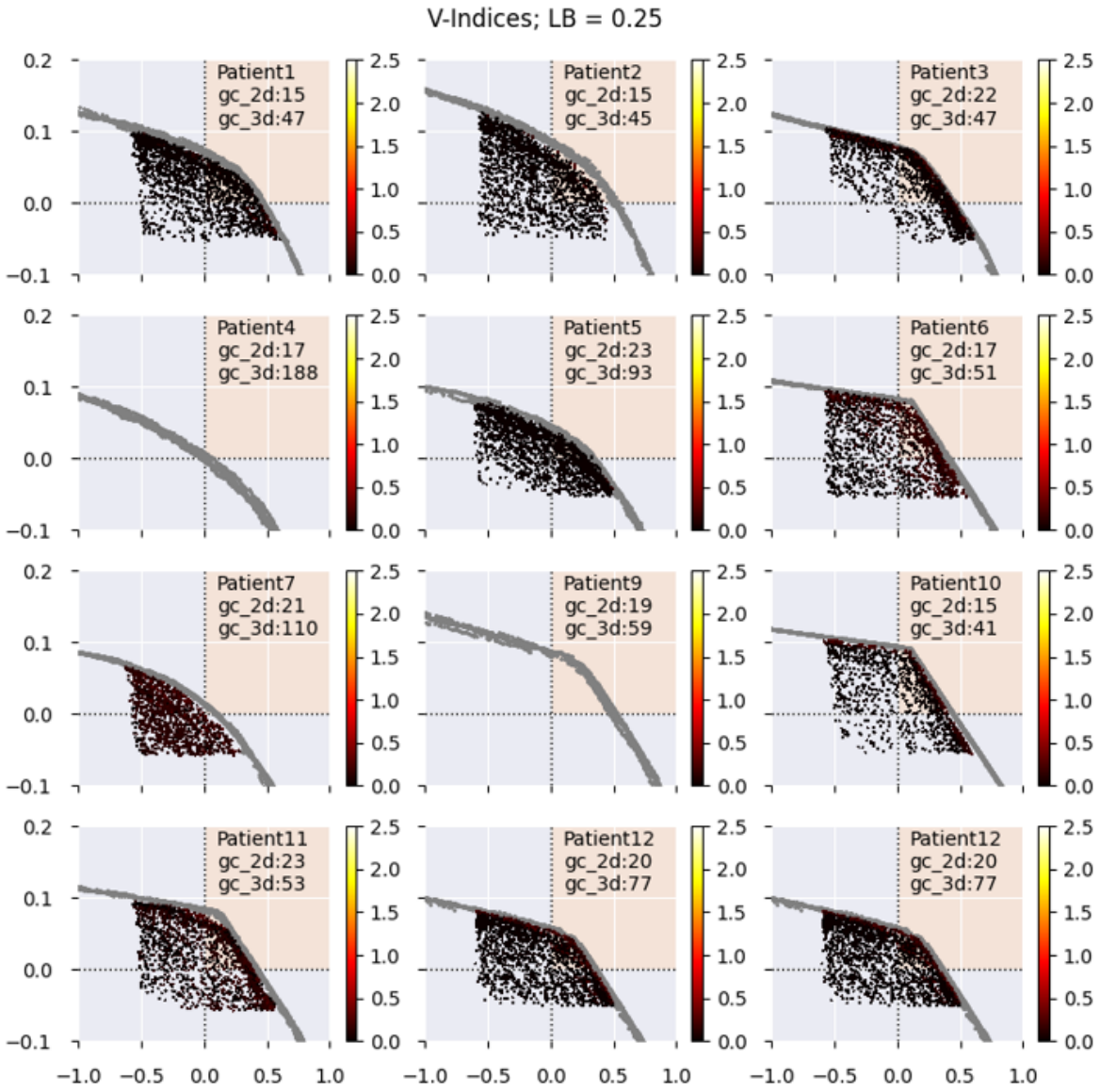

Figure B.65: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

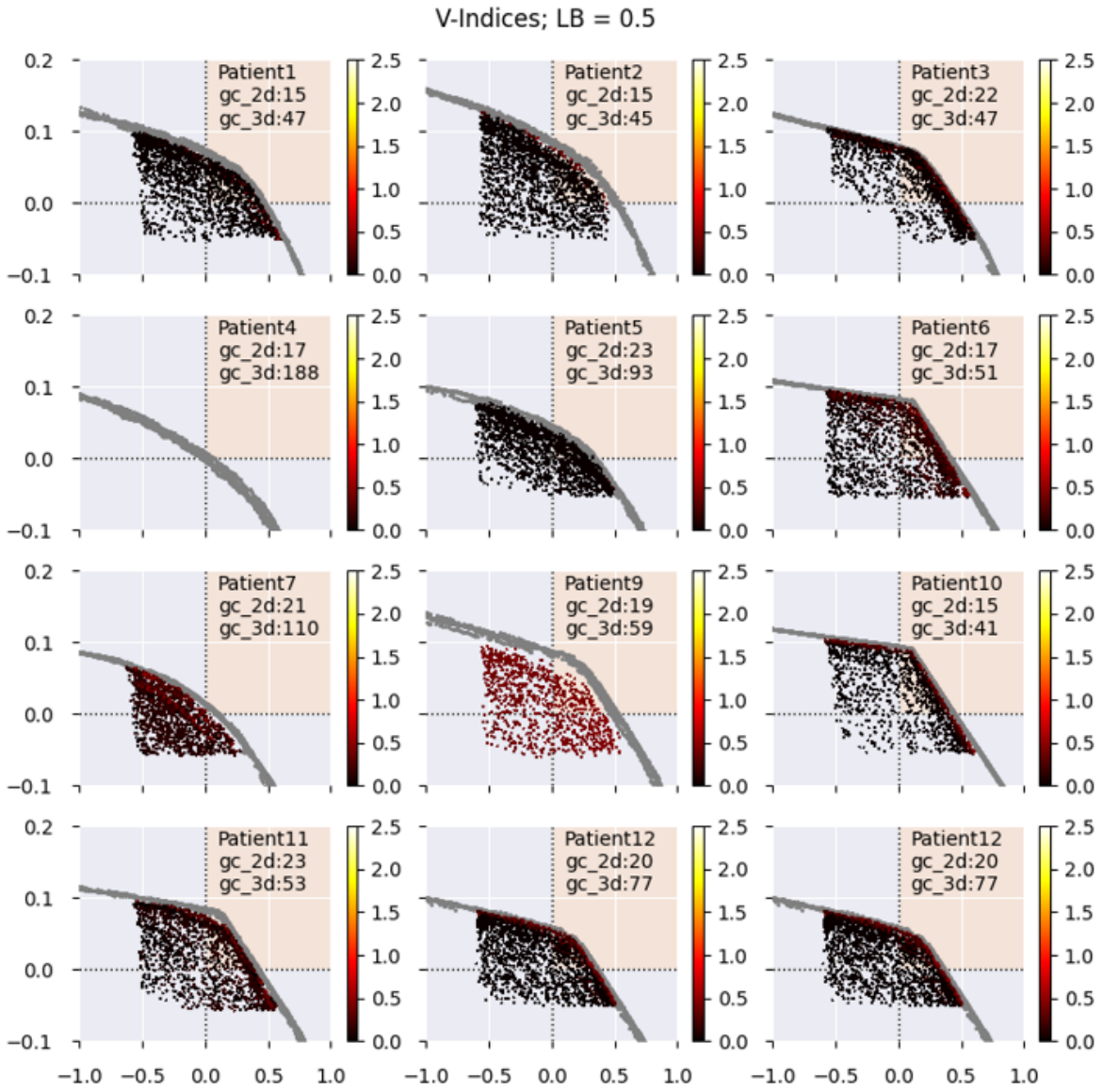

Figure B.66: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

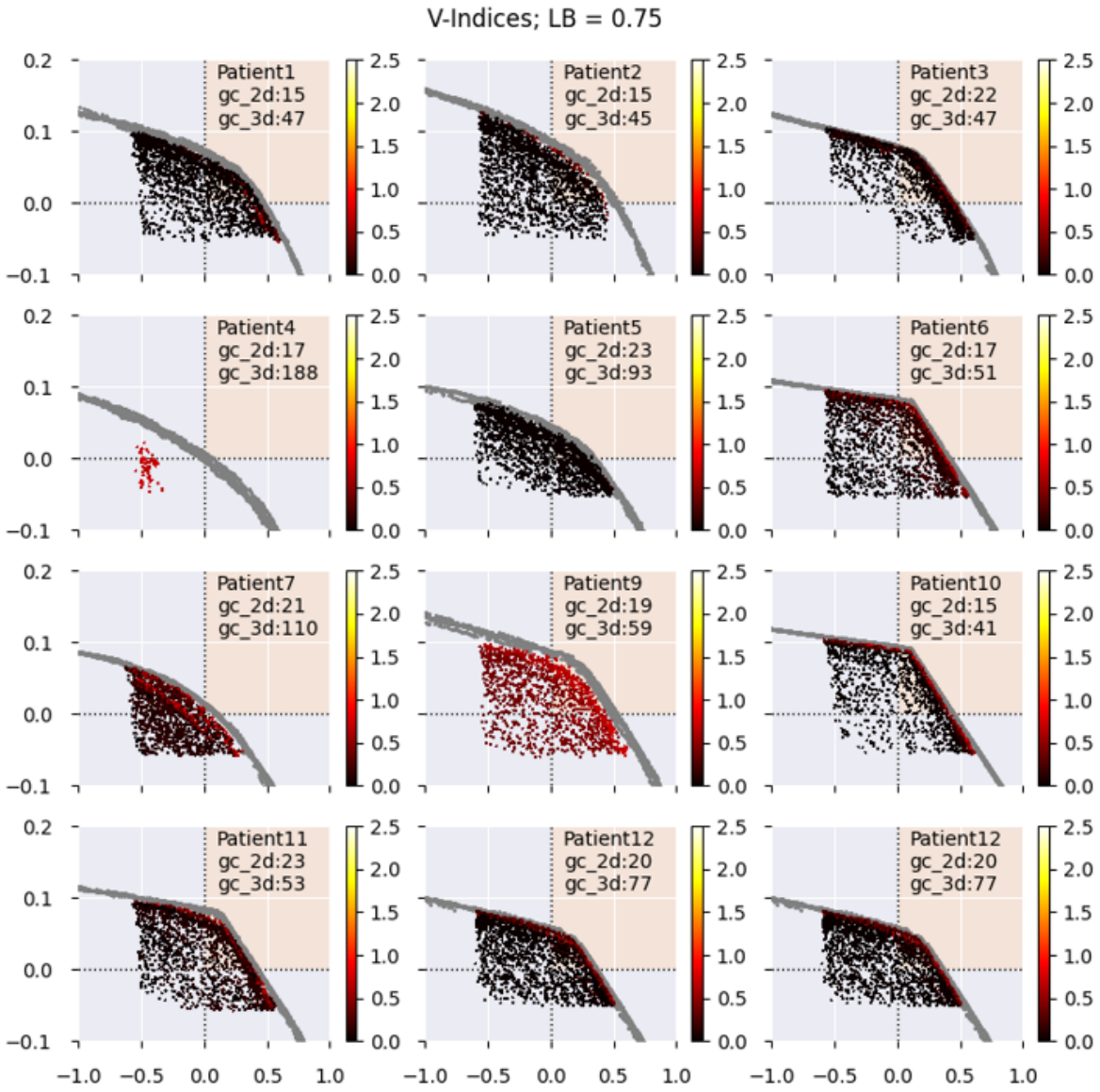

Figure B.67: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

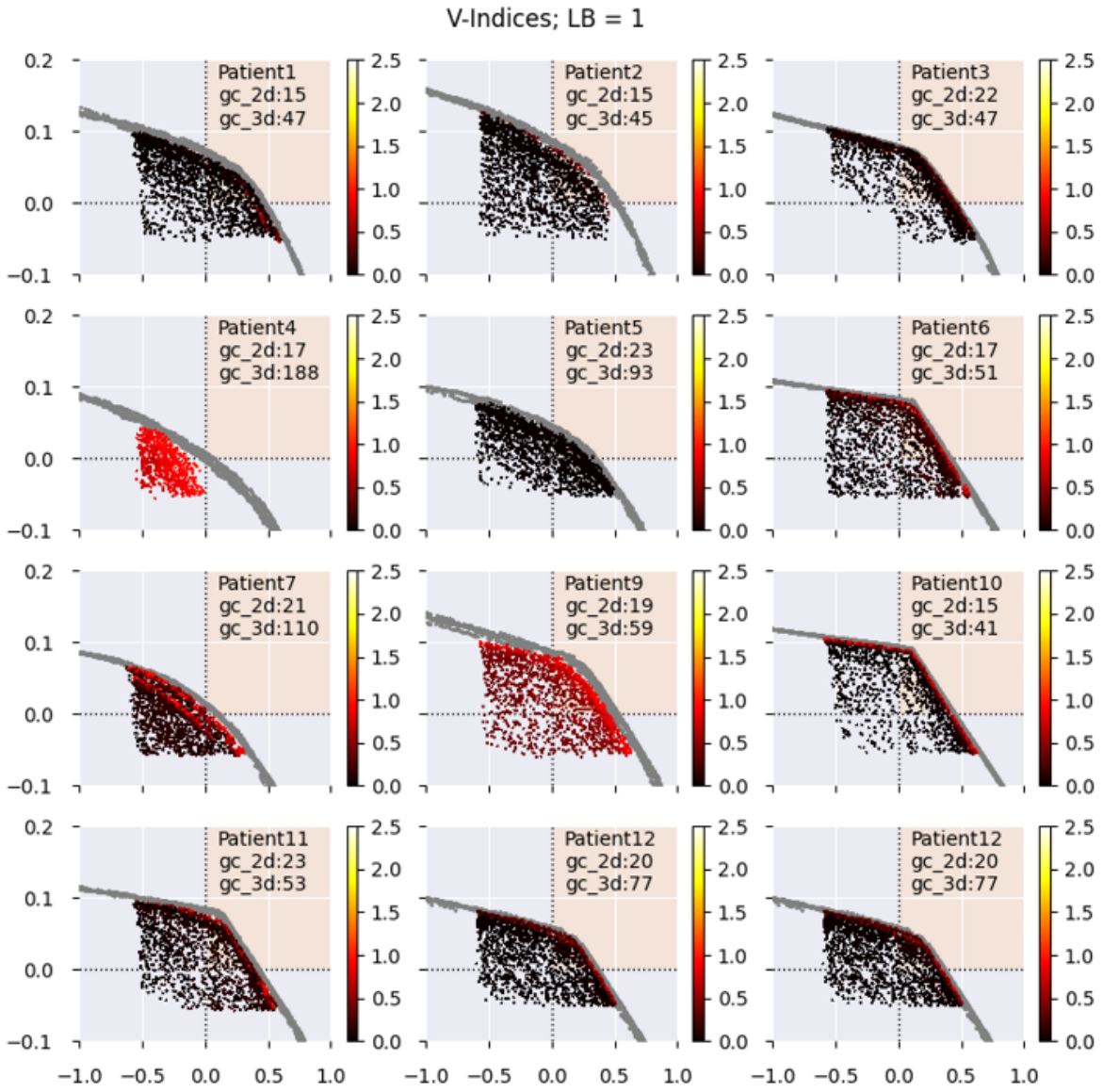

Figure B.68: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

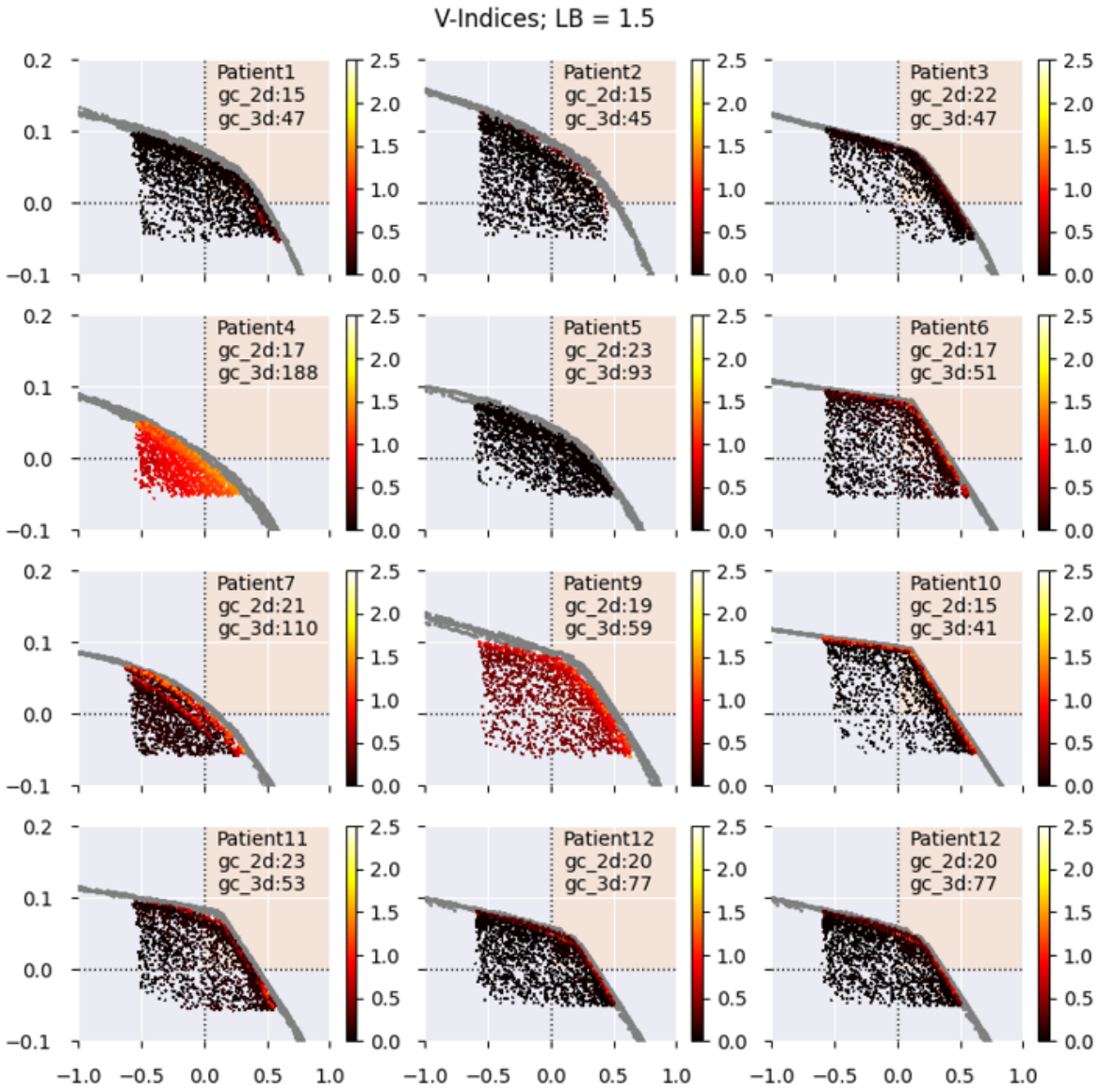

Figure B.69: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

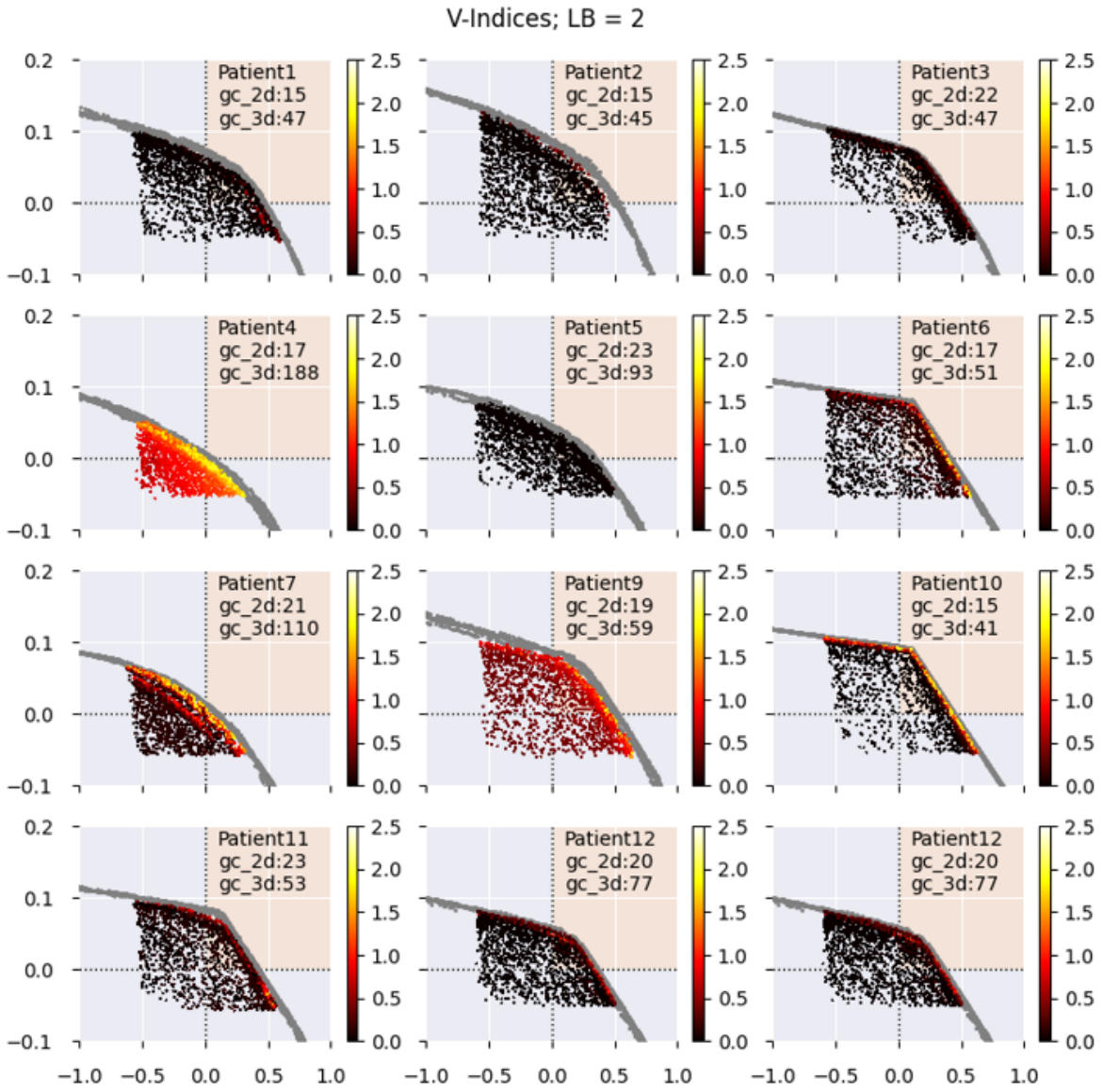

Figure B.70: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

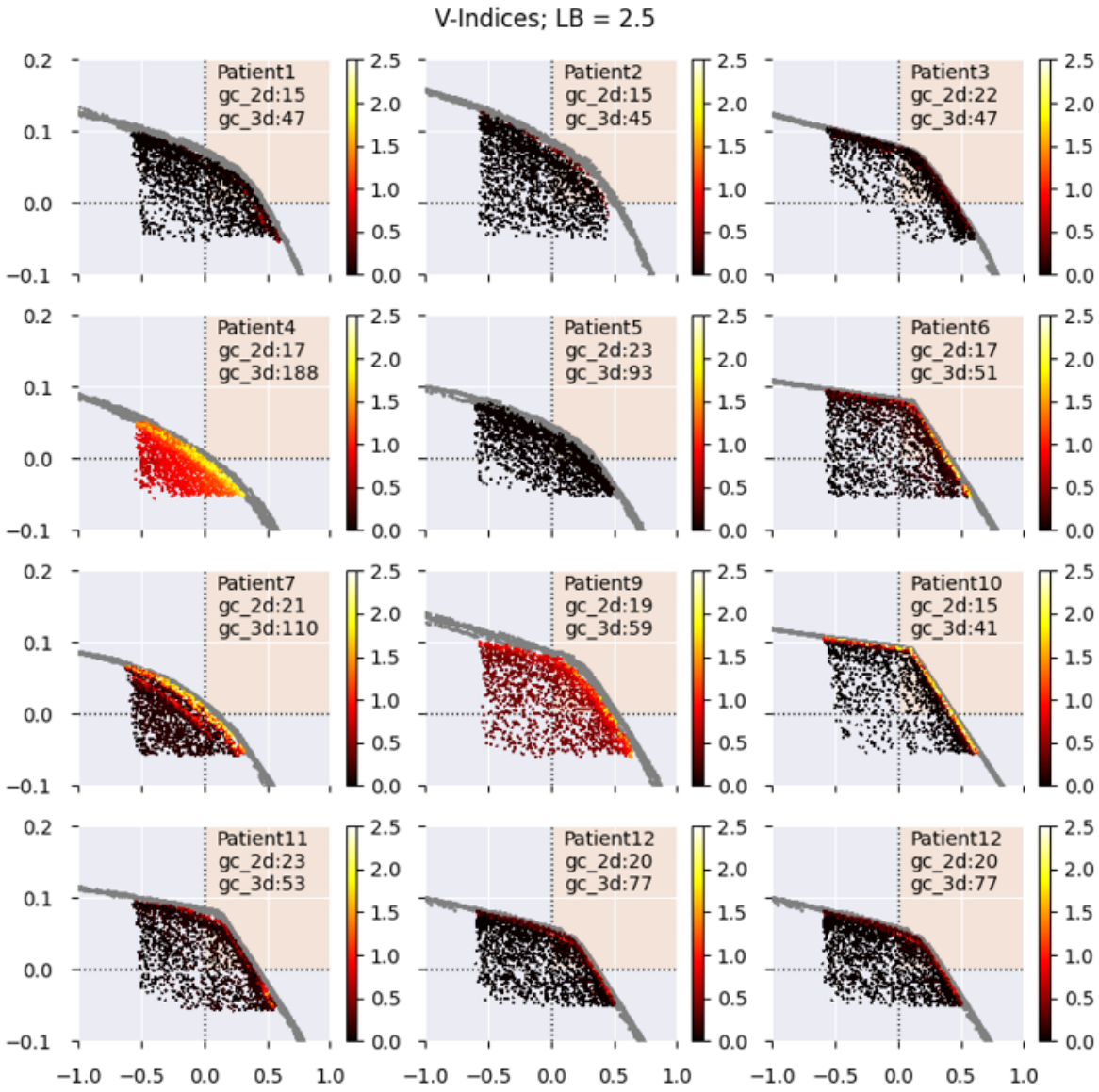

Figure B.71: Pareto approximation fronts resulting from tri-objective BRIGHT with the sum of extra V indices as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

B.7. DTMR results

In this appendix section all Pareto approximation fronts are shown for the DTMR as third objective.

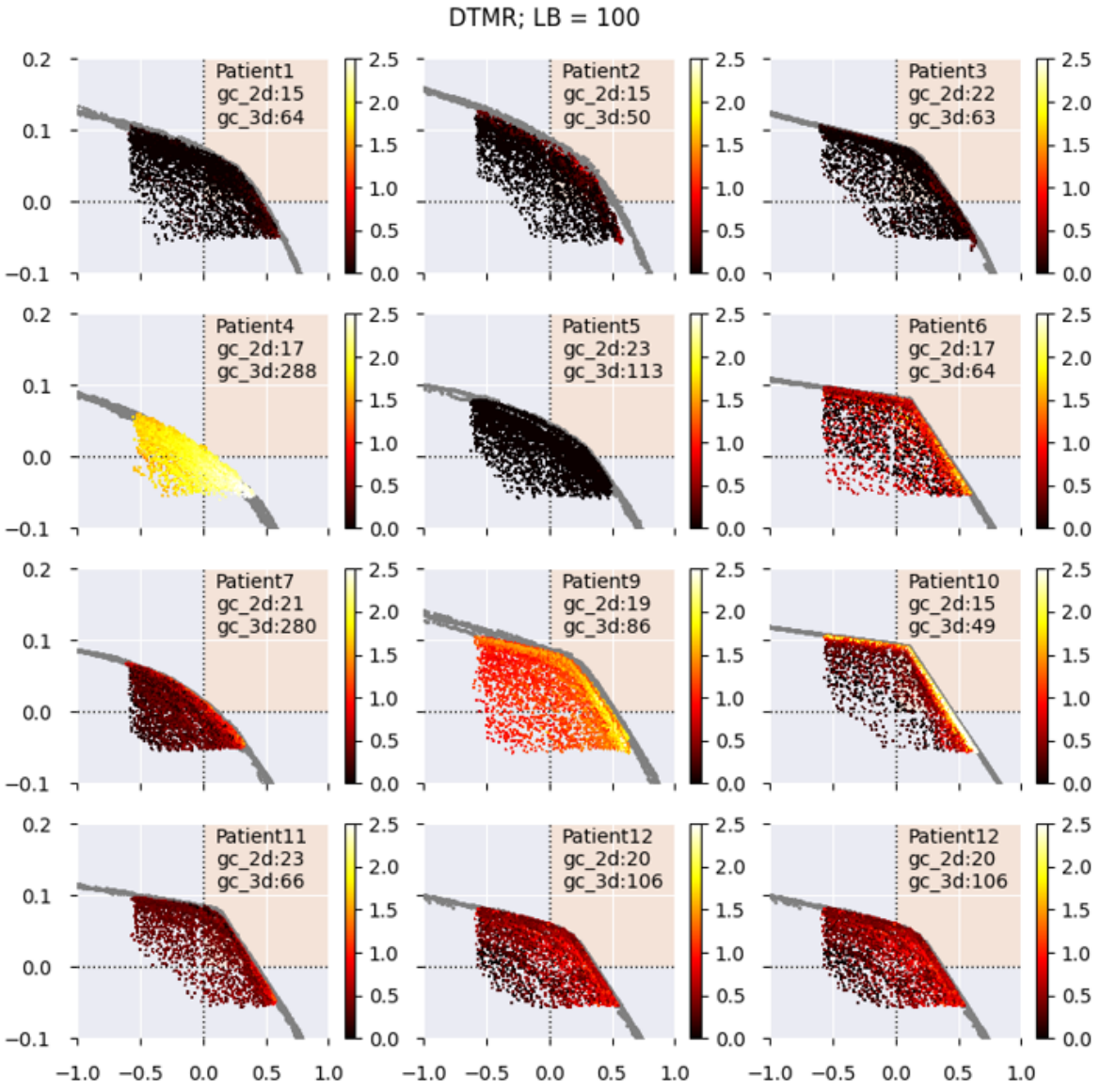

Figure B.72: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

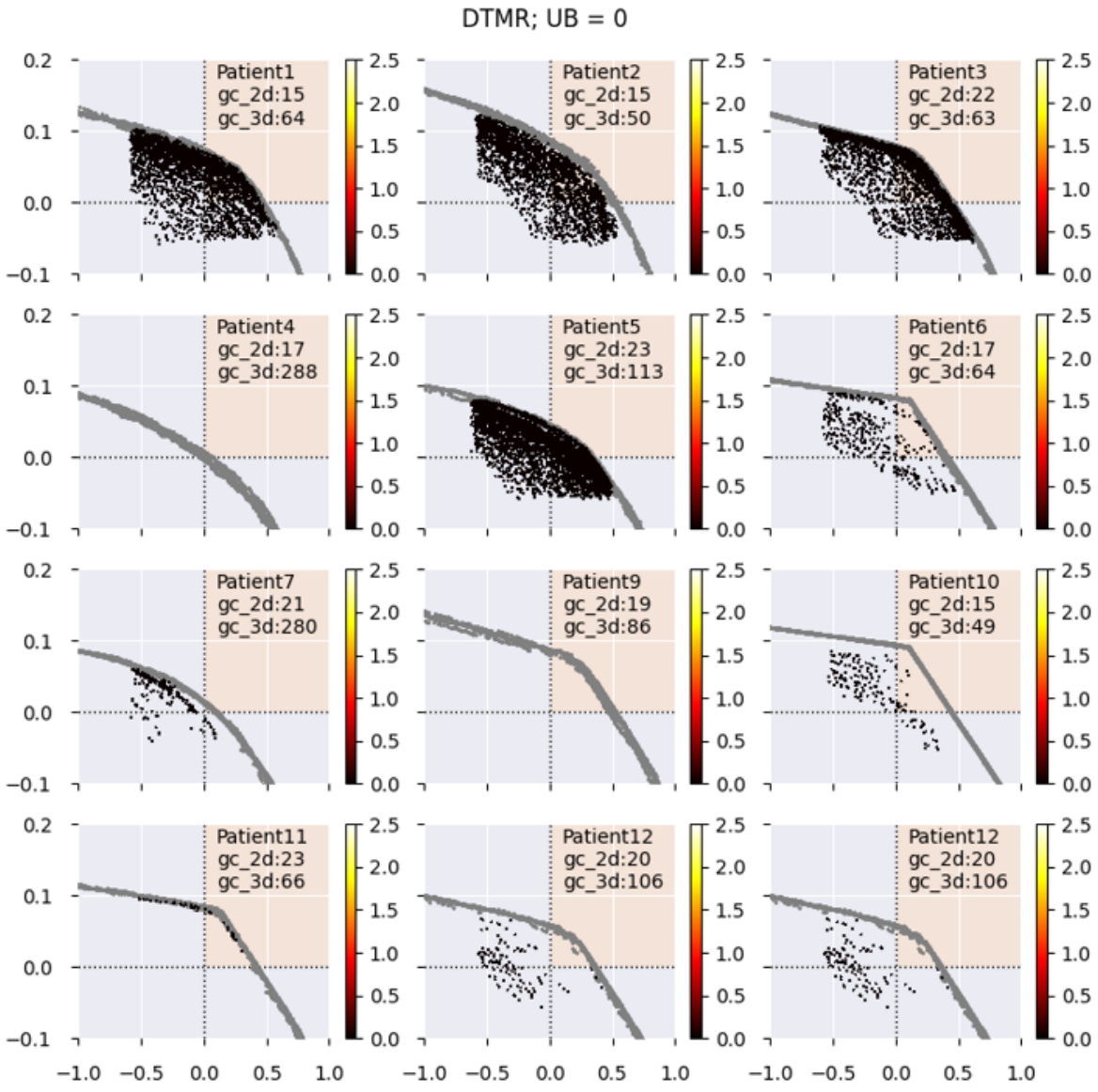

Figure B.73: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

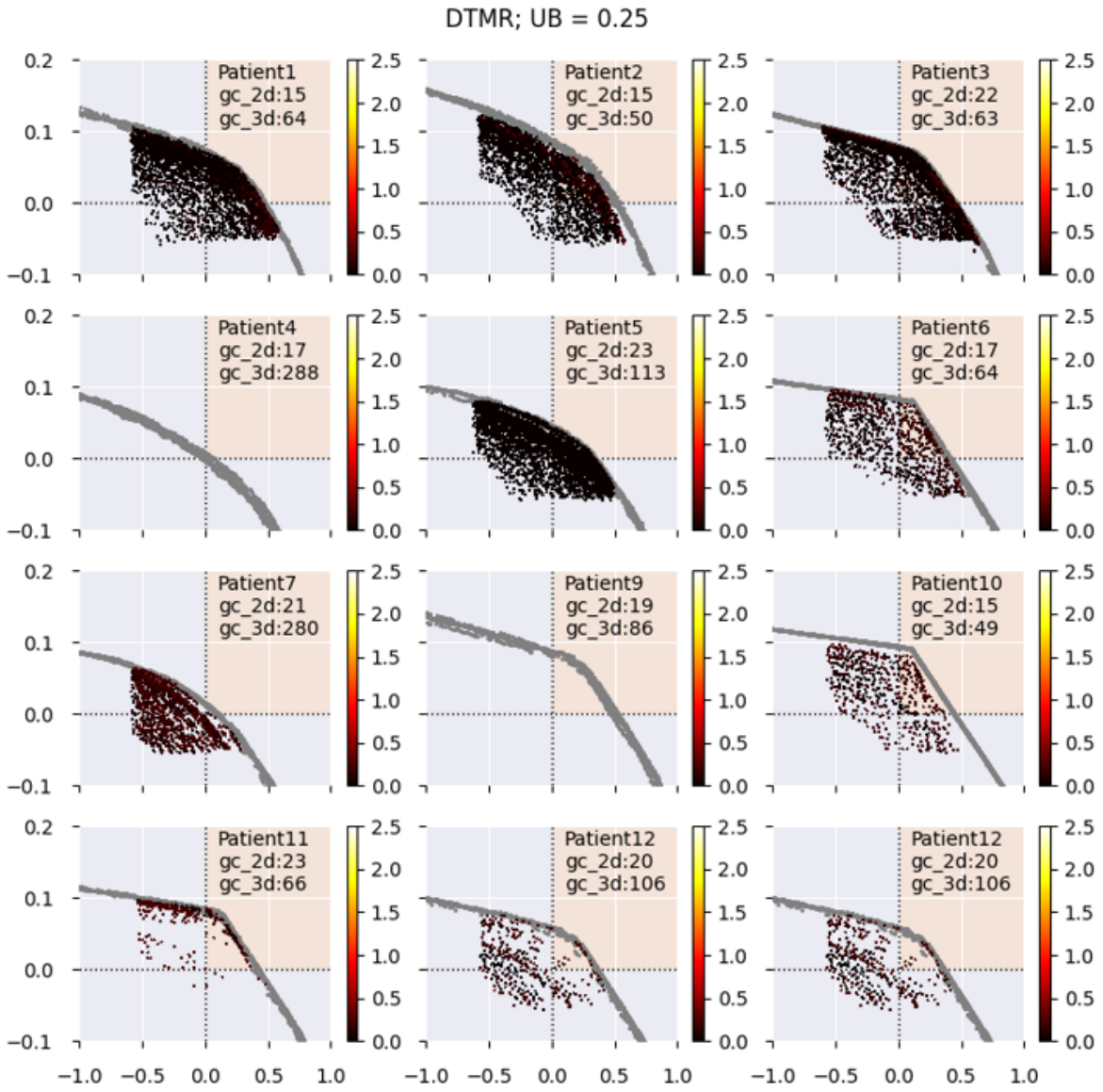

Figure B.74: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

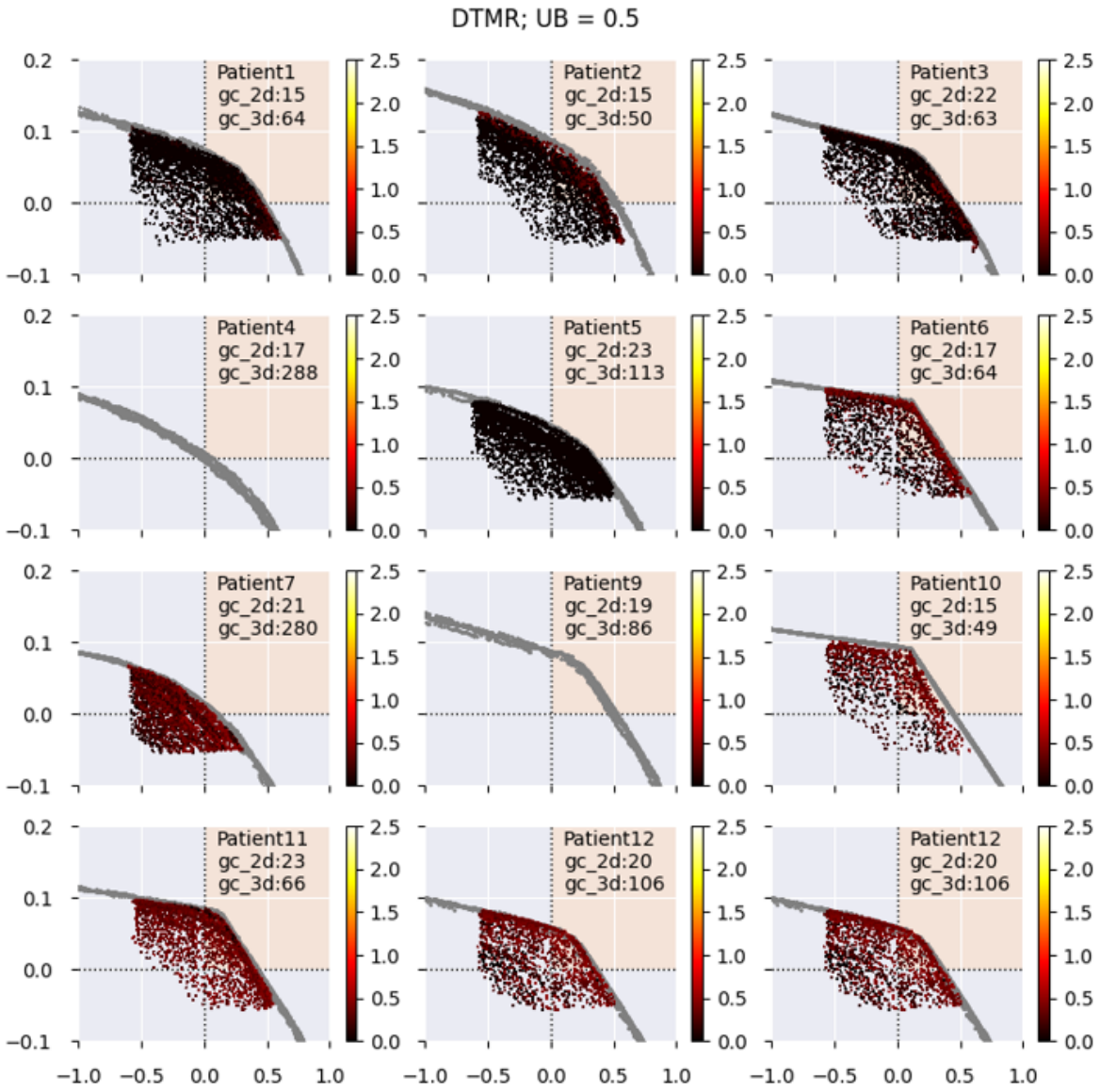

Figure B.75: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

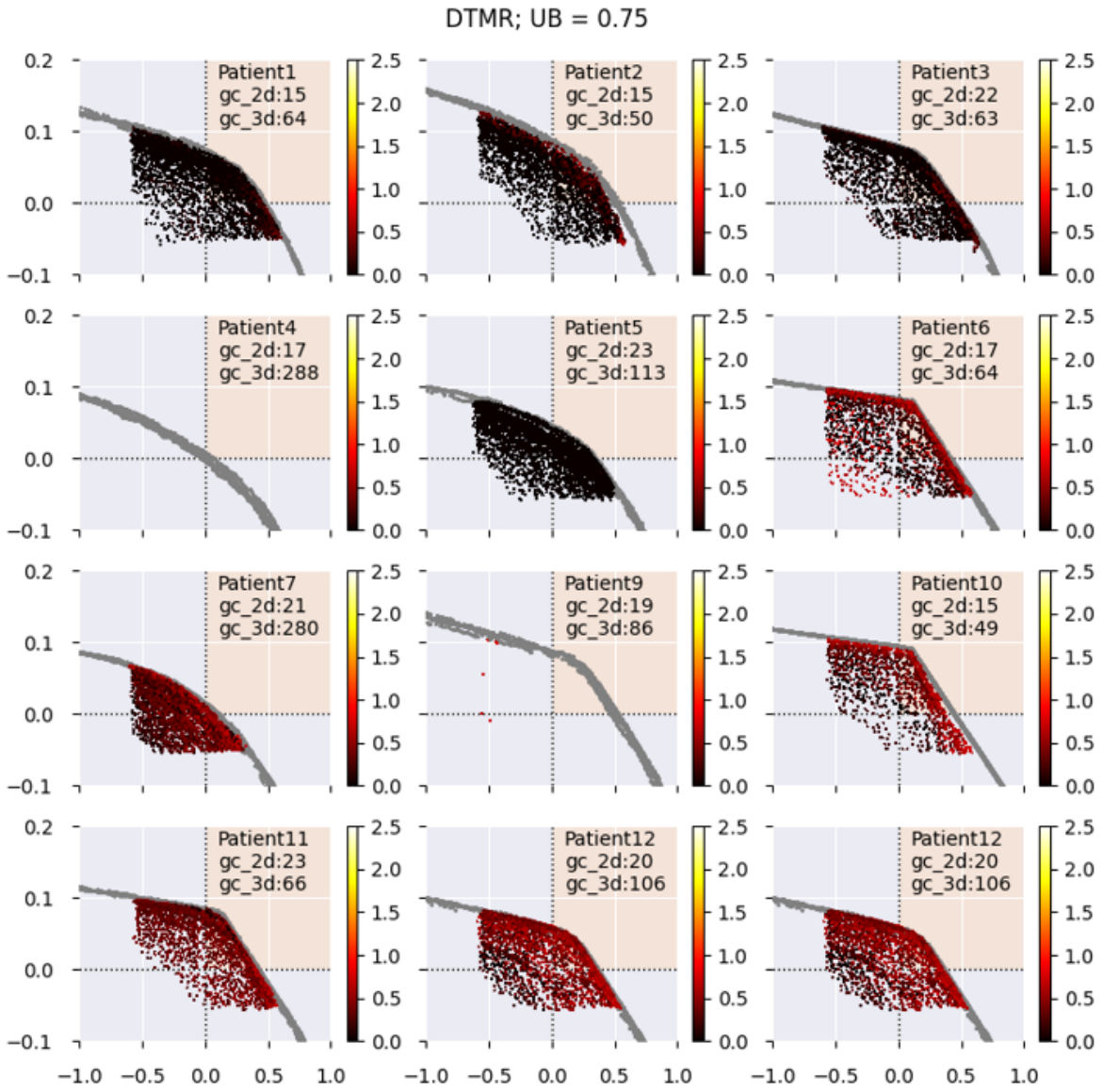

Figure B.76: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

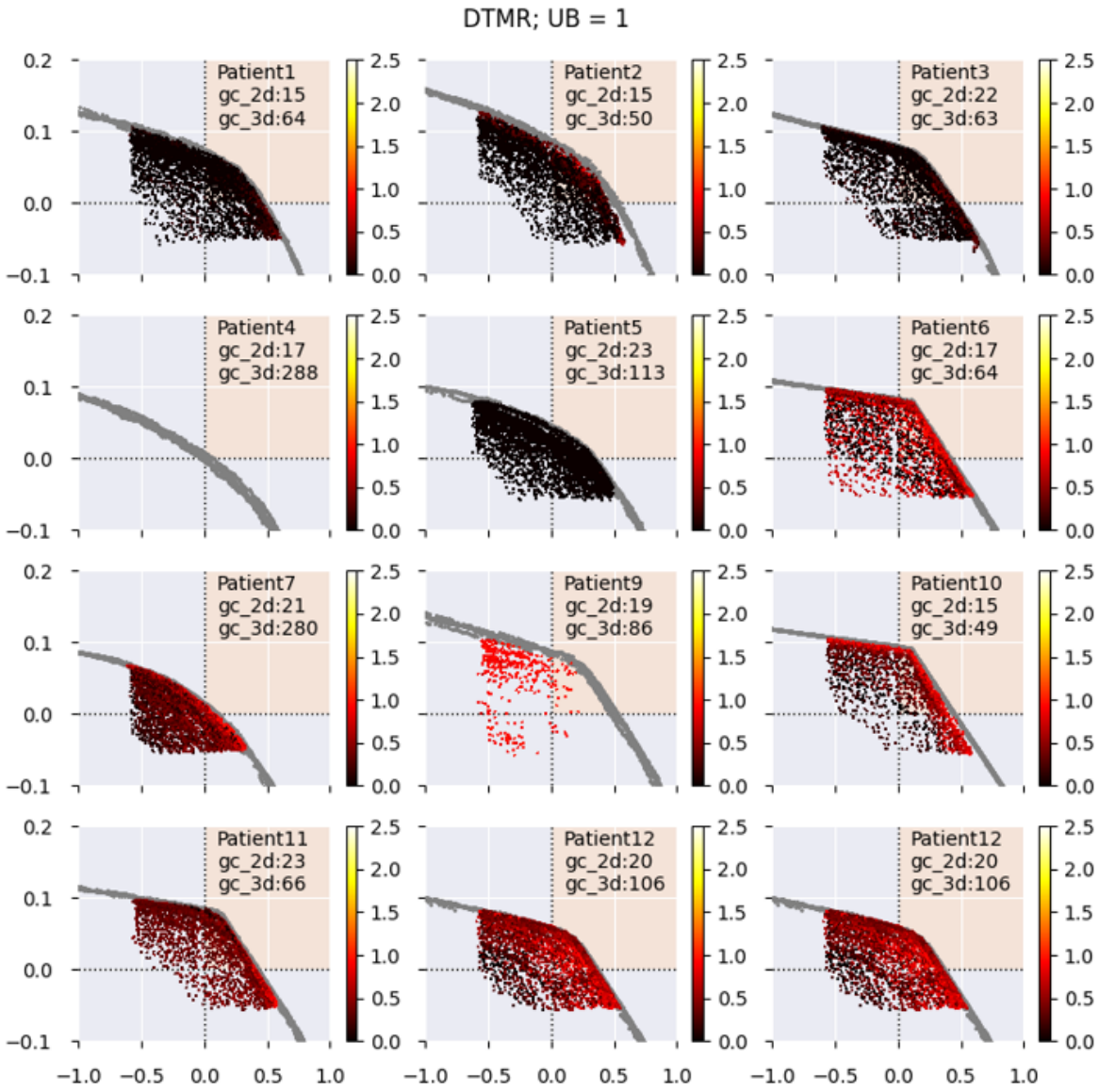

Figure B.77: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

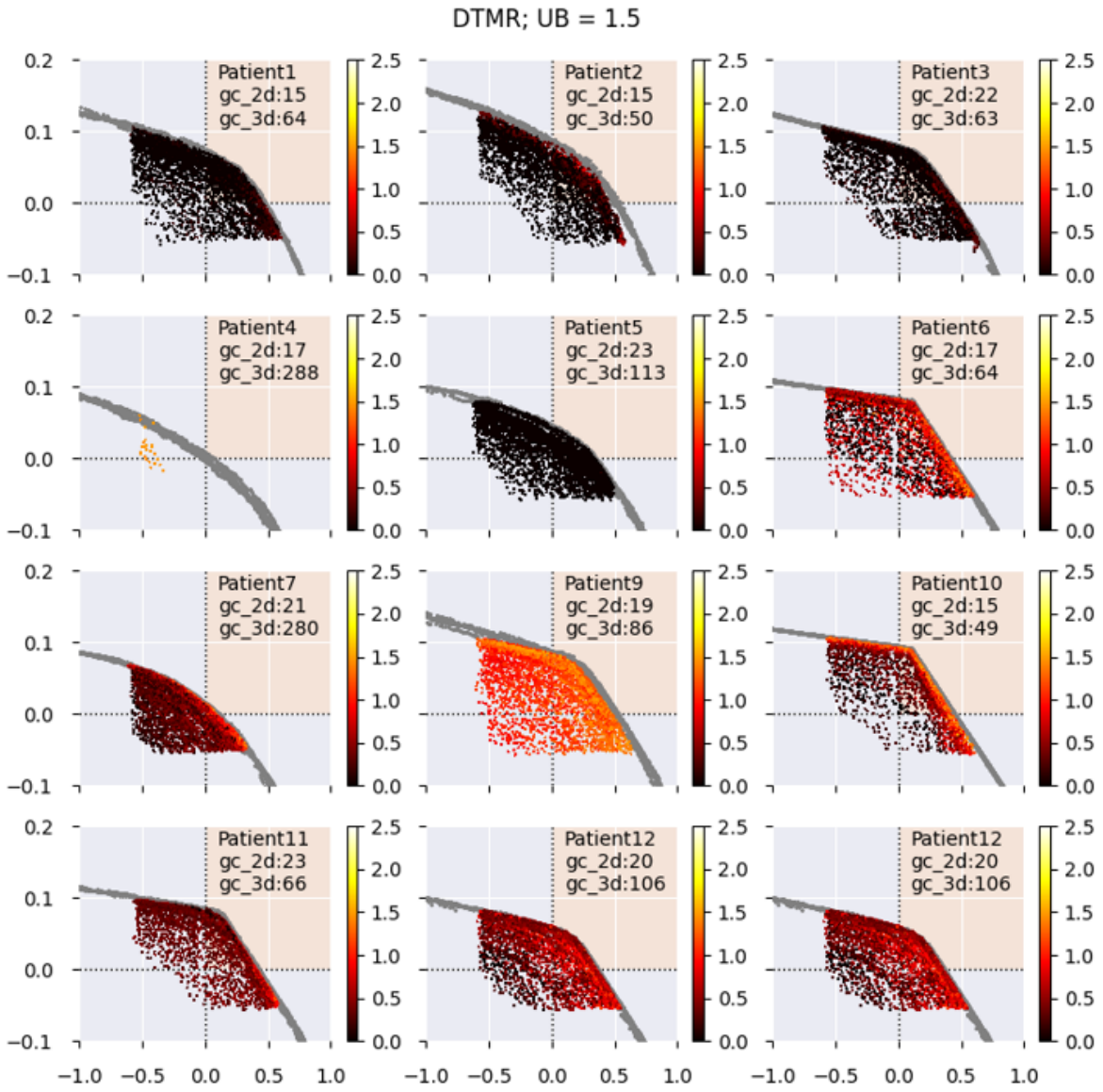

Figure B.78: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

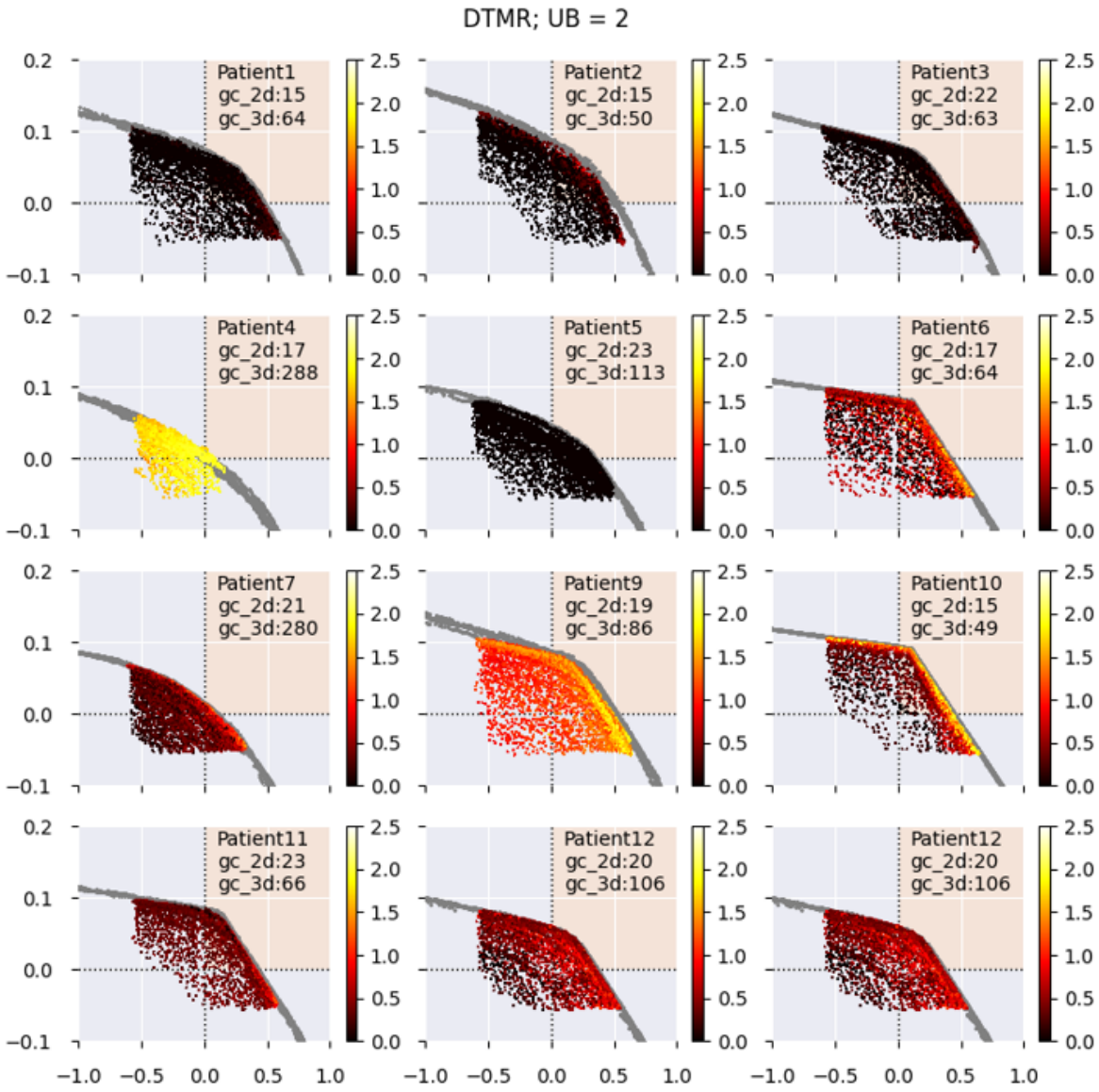

Figure B.79: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

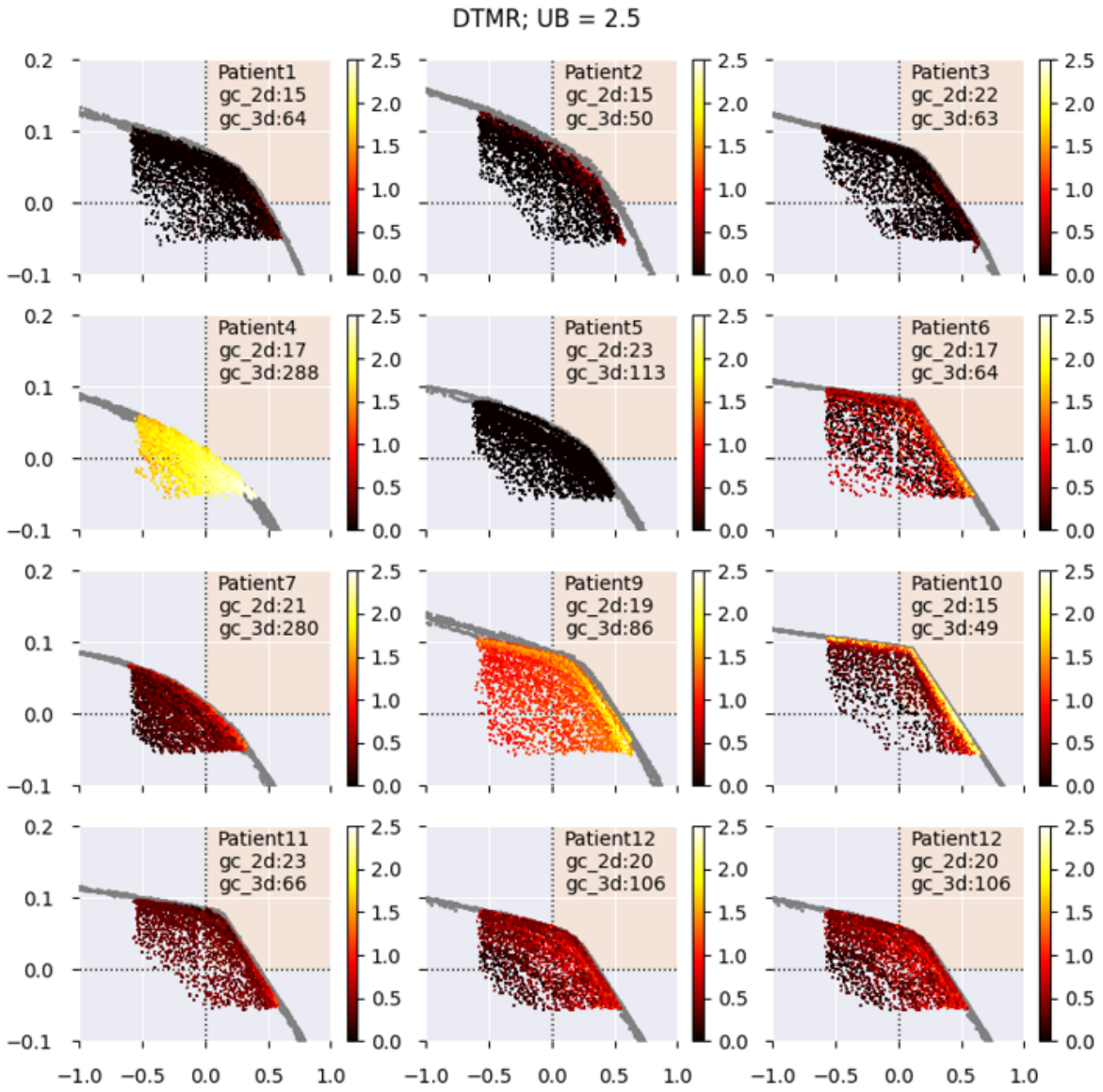

Figure B.80: Pareto approximation fronts resulting from tri-objective BRIGHT with DTMR as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value, although wrongfully indicated with LB instead of UB. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

B.8. DLDM results

In this appendix section all Pareto approximation fronts are shown for the DLDM as third objective.

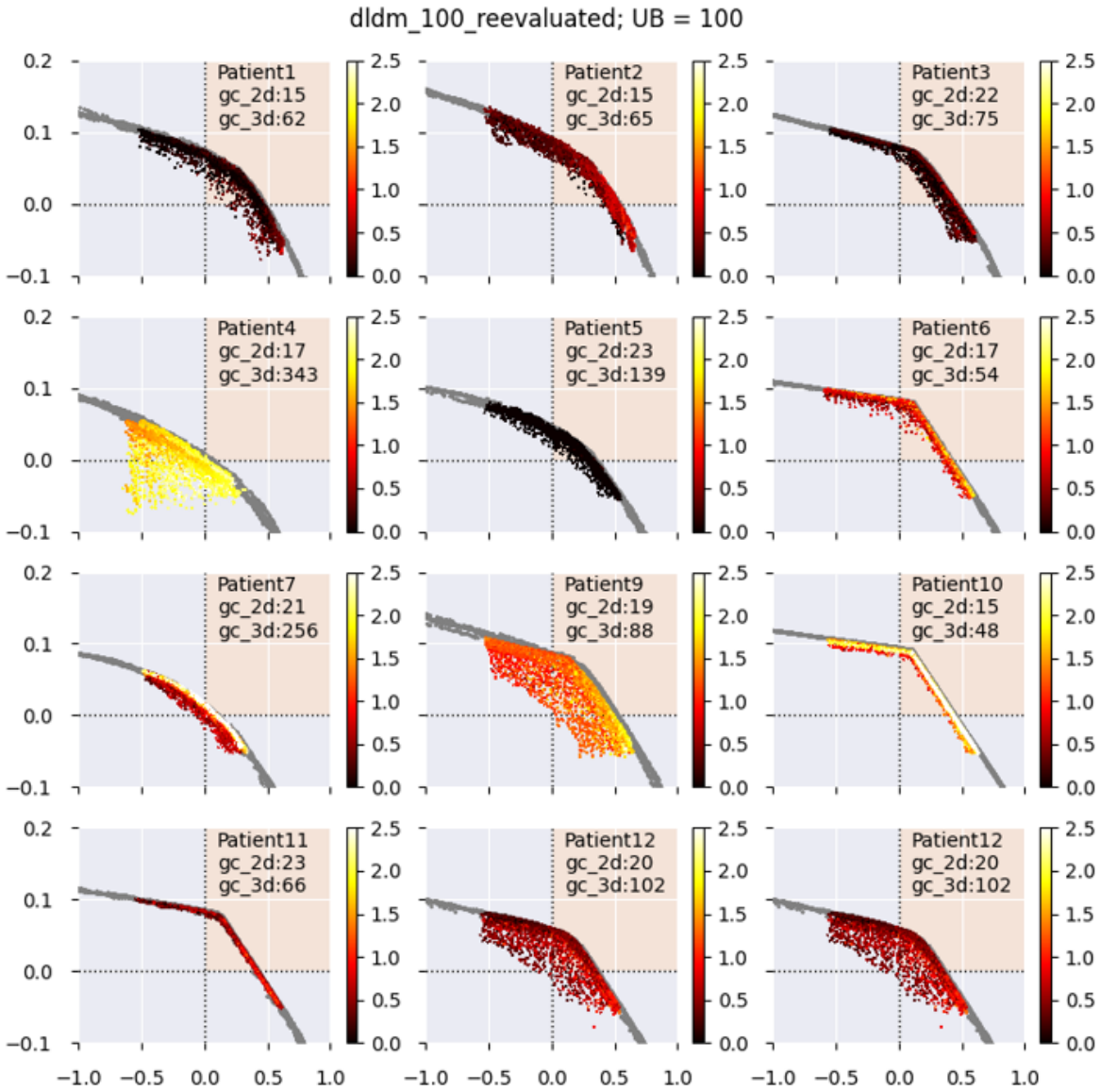

Figure B.81: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

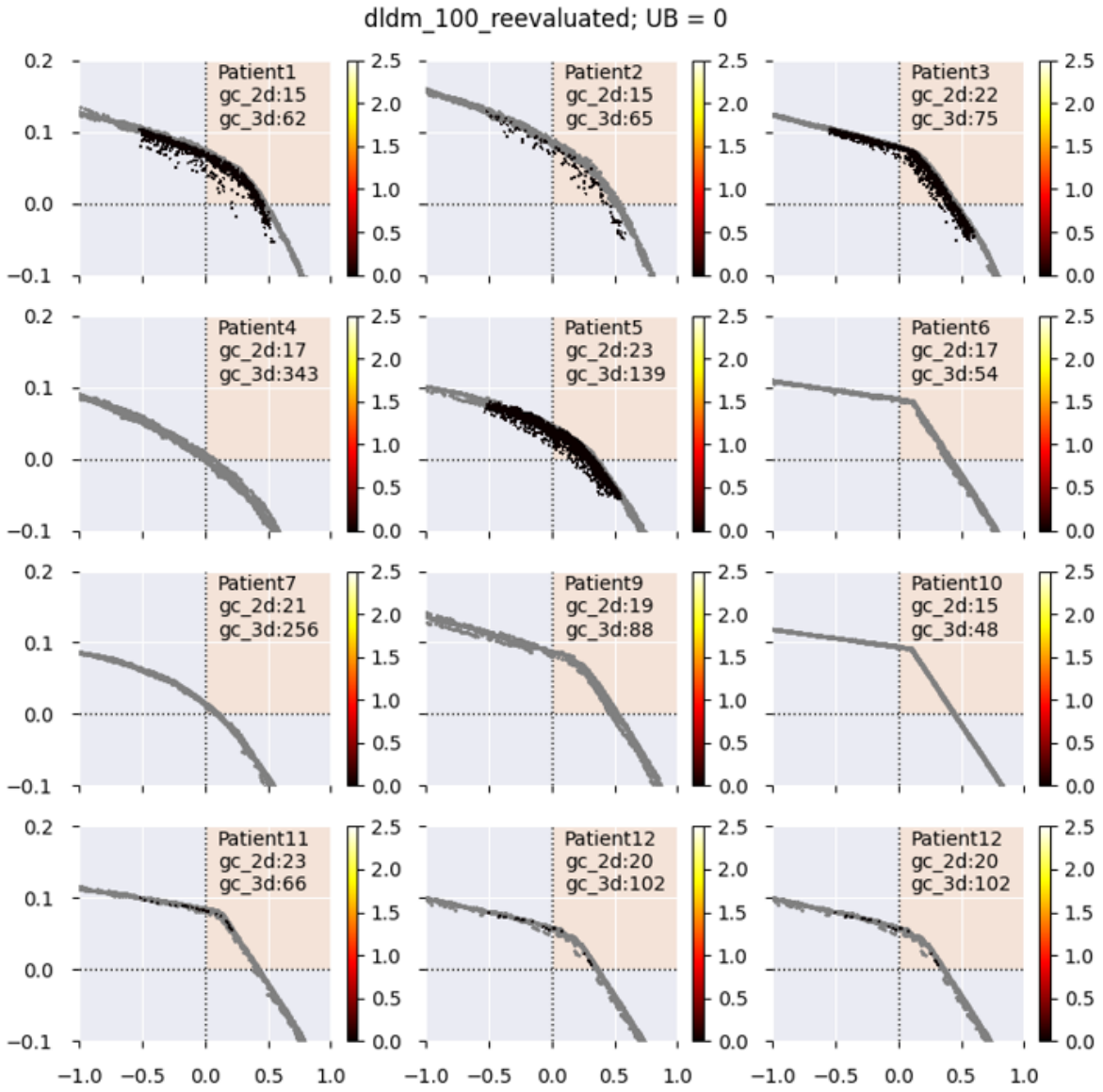

Figure B.82: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

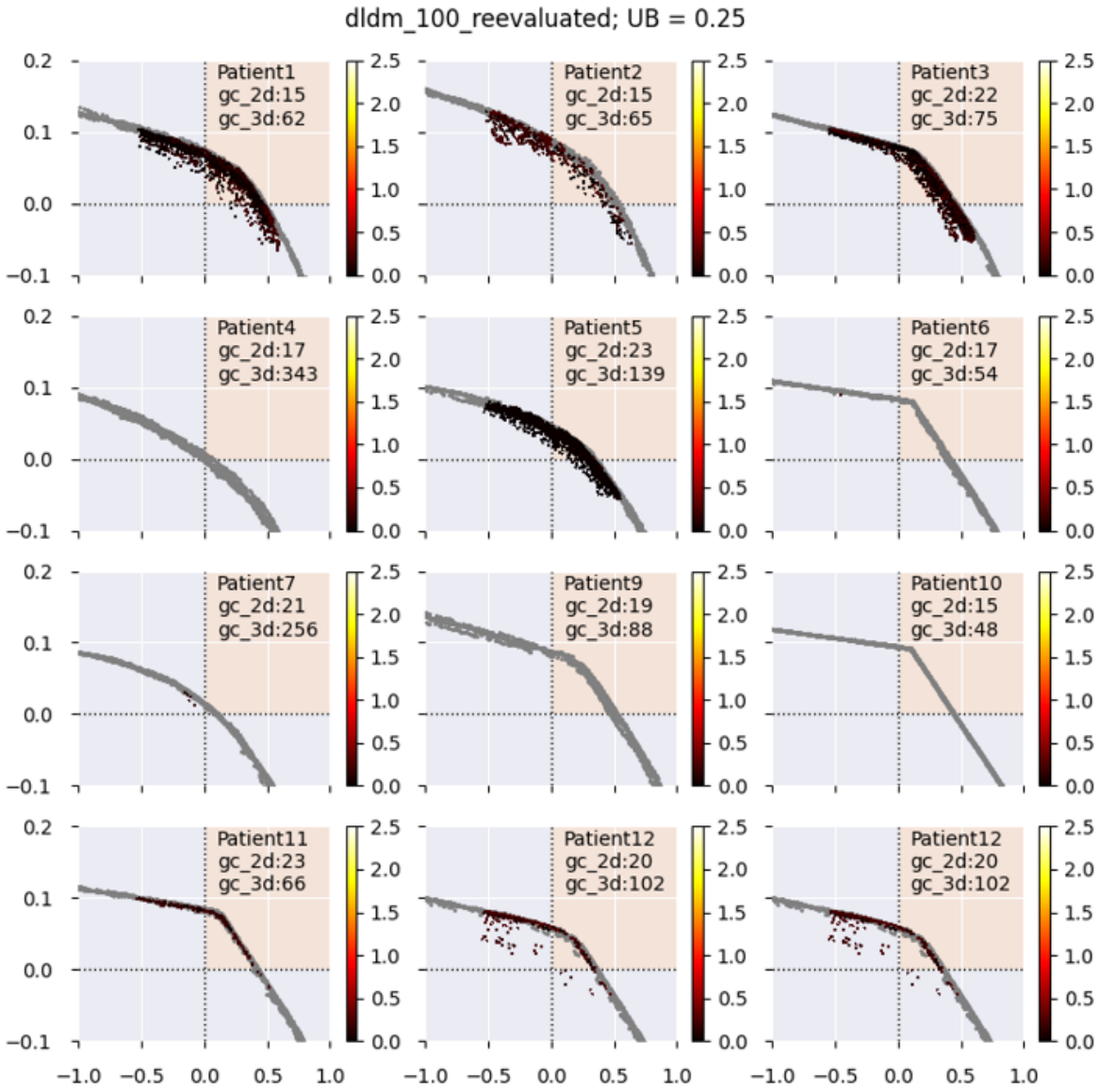

Figure B.83: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

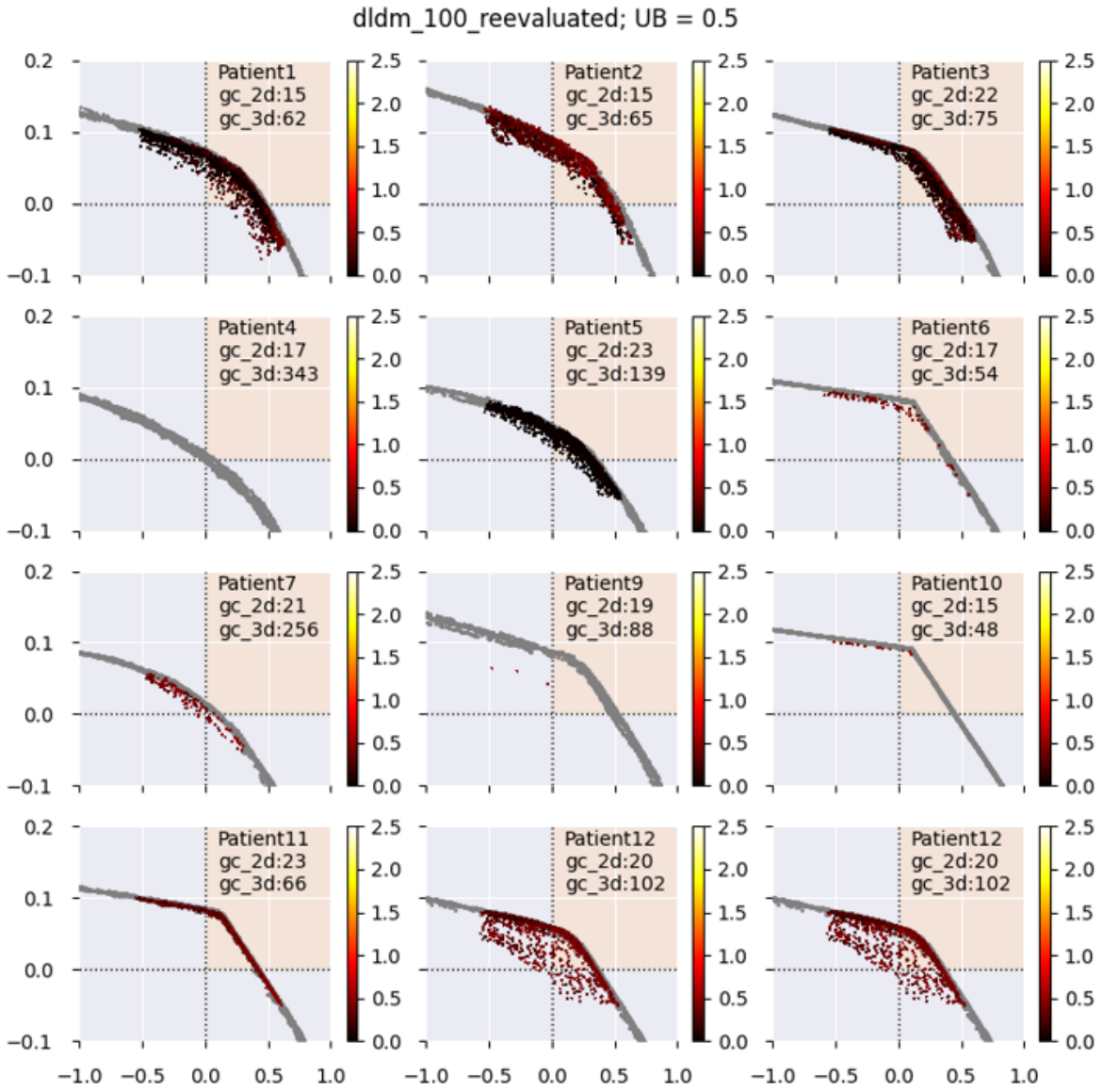

Figure B.84: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

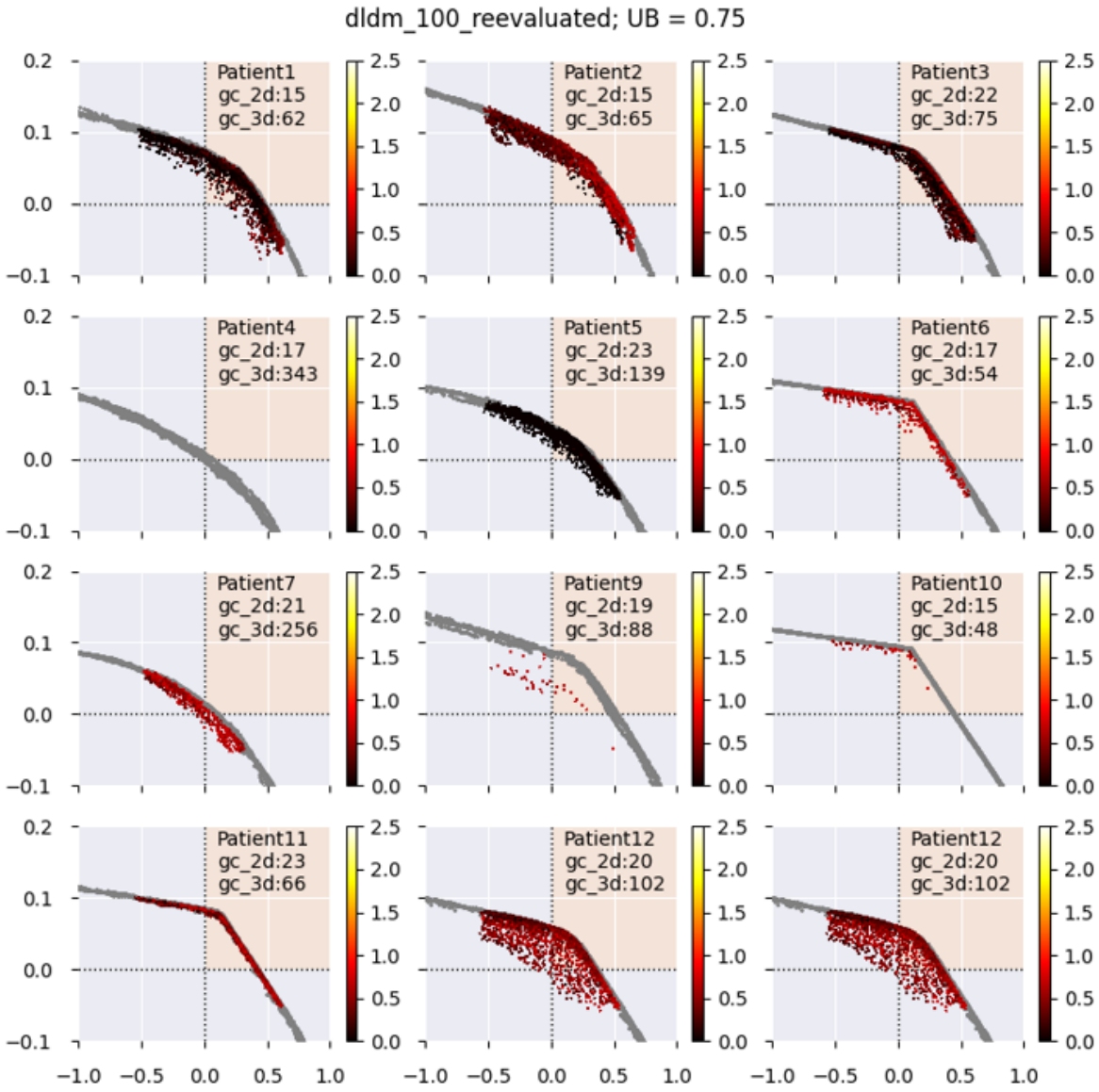

Figure B.85: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

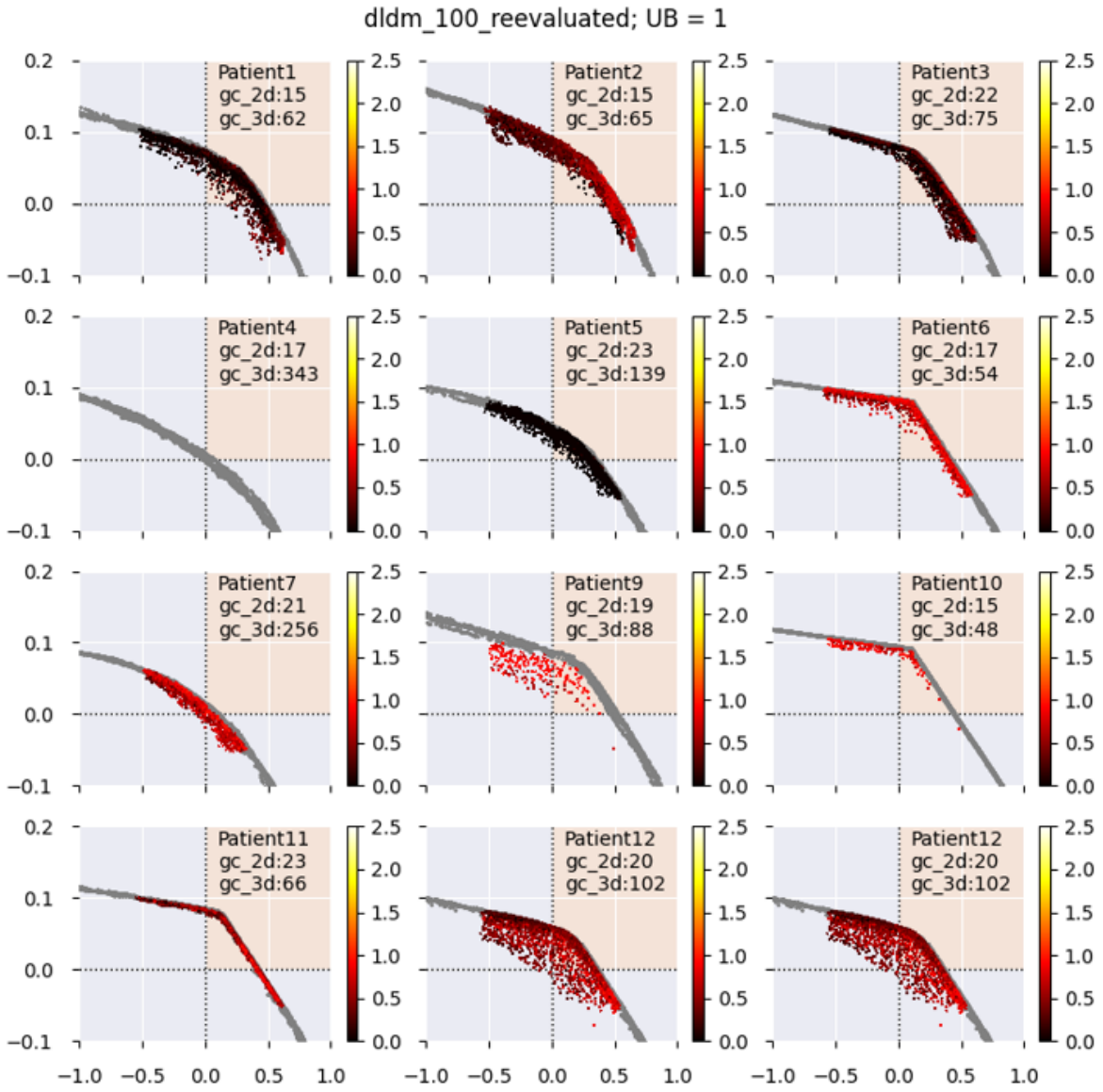

Figure B.86: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

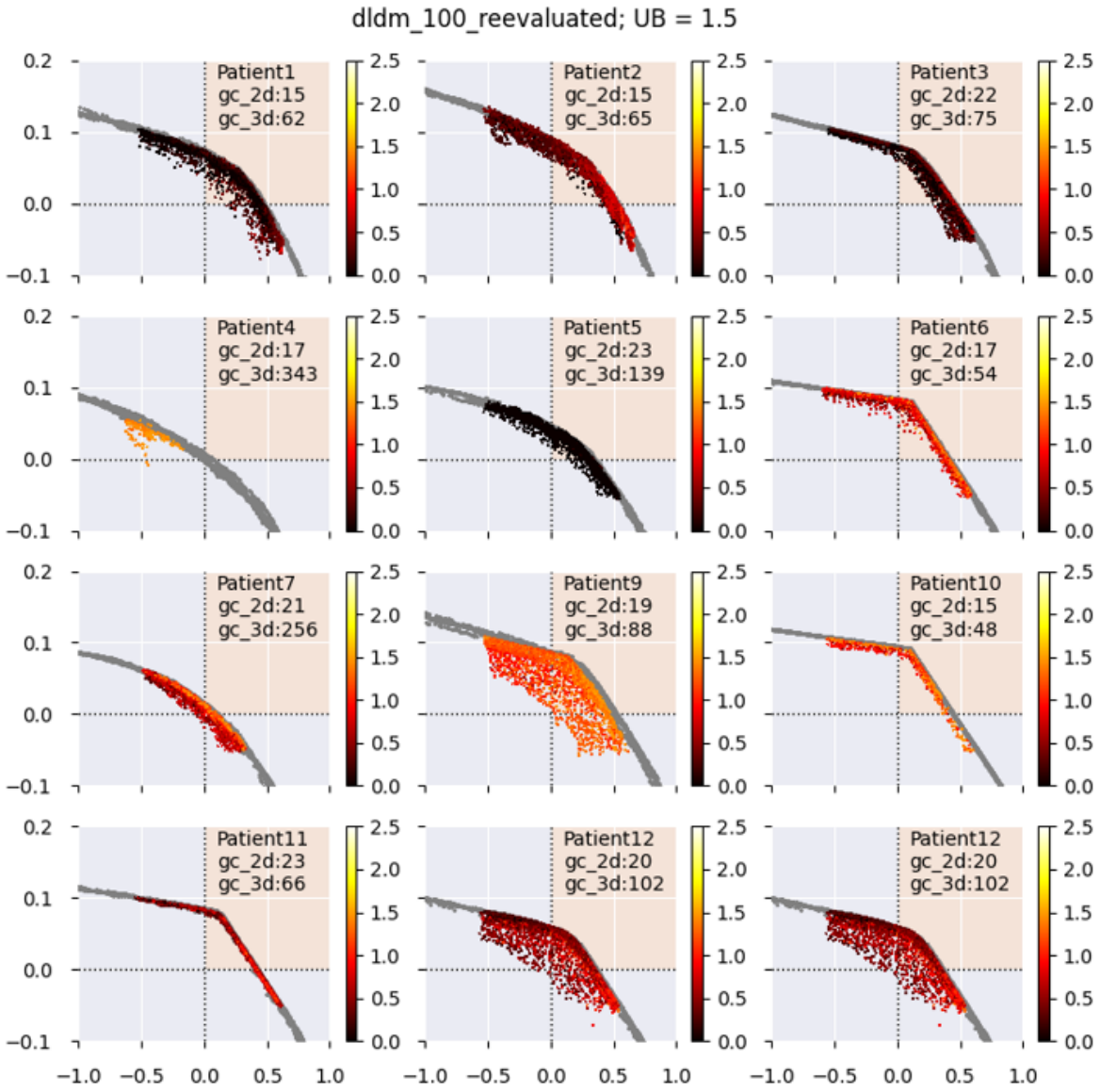

Figure B.87: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

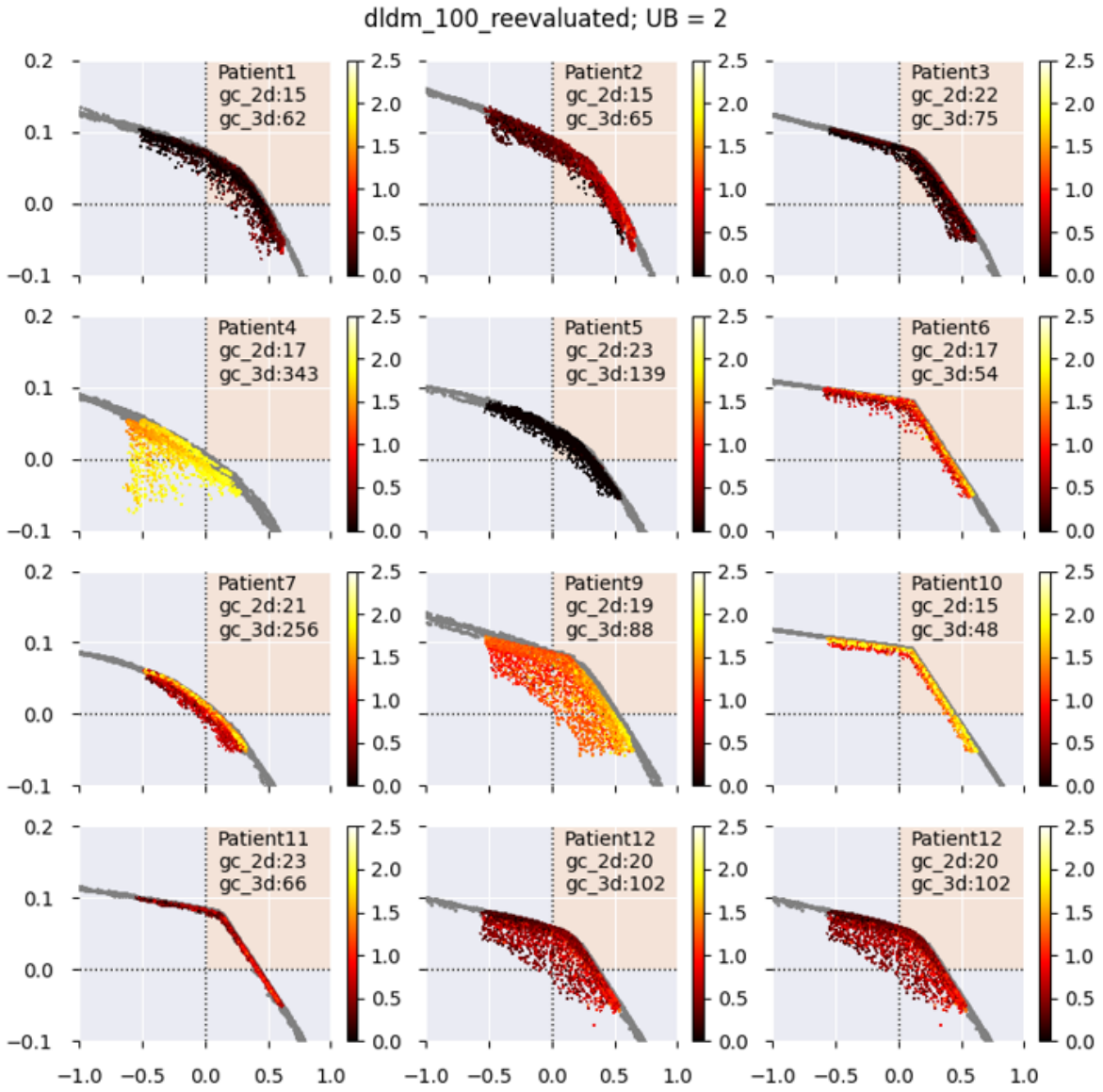

Figure B.88: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

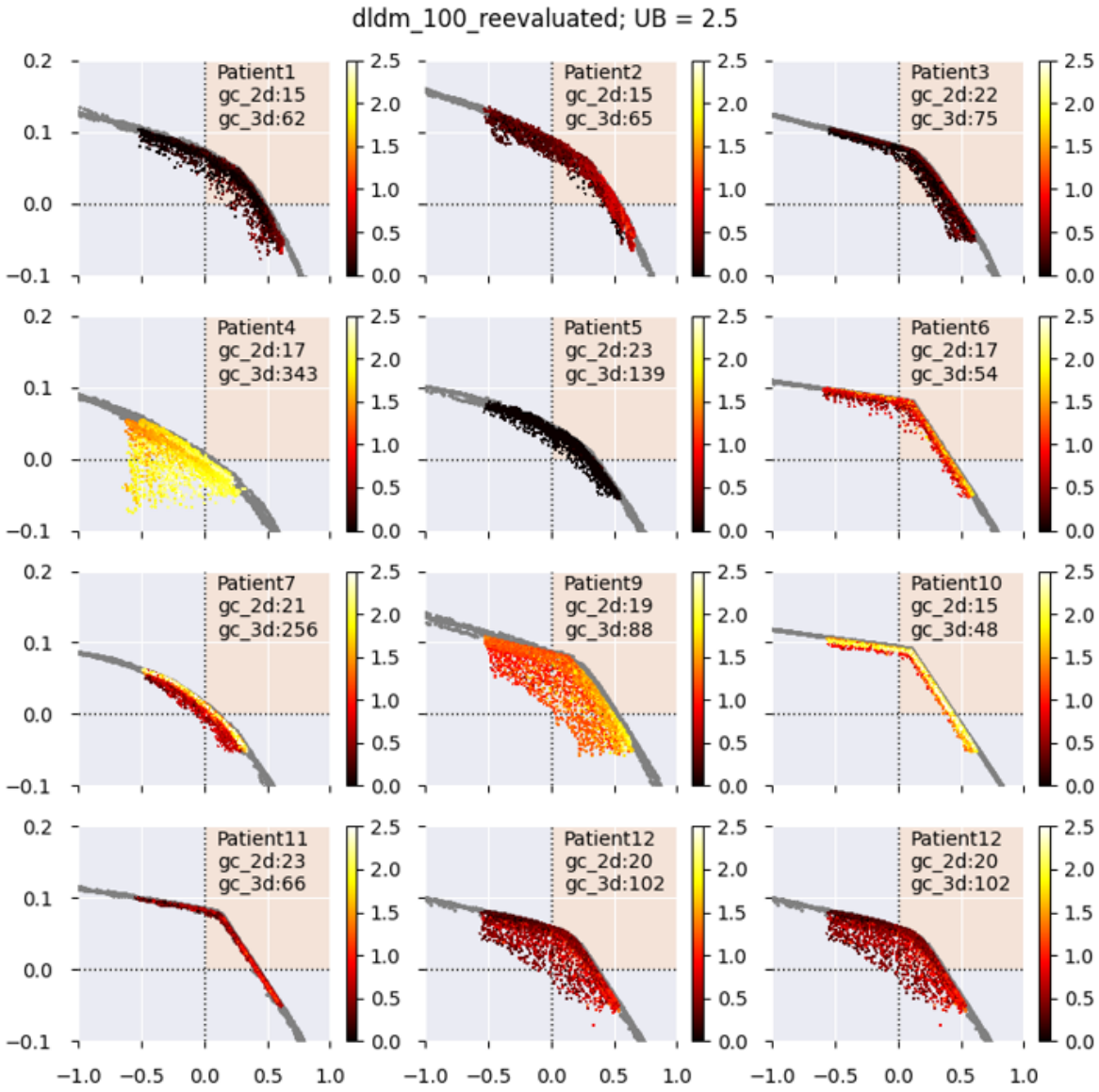

Figure B.89: Pareto approximation fronts resulting from tri-objective BRIGHT with DLDM as third objective. On the right hand side the colour bar is shown with colour corresponding to HSI value. The gray dots are treatment plans generated by bi-objective BRIGHT. In the title it states the applied upper bound on HSI value. In the top right corner it states how long it has taken to reach the golden corner in bi-objective BRIGHT and how long it has taken in tri-objective BRIGHT.

B.9. Slicing the Pareto approximation fronts

In this appendix section all Pareto approximation fronts are shown for the DLDM as third objective.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient1	0	N/A[2]	0.095[3]	0.092	0.095	0.100	N/A[2]	-0.237[3]	**0.420**	**0.423**	**0.434**	N/A[2]	-0.237[3]	**0.532**	**0.535**	**0.434**
Patient1	0.25	0.103	0.101	**0.092**	**0.095**	0.100	0.455	0.458	0.428	0.441	0.447	0.535	0.518	0.541	0.548	0.545
Patient1	0.5	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.613	0.593	**0.541**	0.549	0.609
Patient1	0.75	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.550	0.609
Patient1	1	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.559	0.609
Patient1	1.5	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.559	0.609
Patient1	2	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.559	0.609
Patient1	2.5	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.559	0.609
Patient1	none	0.103	0.101	**0.095**	**0.095**	0.100	0.455	0.466	**0.428**	0.448	0.451	0.627	0.604	**0.541**	0.559	0.609

Figure B.90: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient2	0	N/A[0]	N/A[0]	**0.117**	**0.116**	**0.121**	N/A[0]	N/A[0]	**0.405**	**0.427**	**0.274**	N/A[0]	N/A[0]	**0.466**	**0.525**	**0.550**
Patient2	0.25	N/A[1]	N/A[2]	0.117	0.116	0.122	N/A[1]	N/A[2]	**0.405**	**0.427**	**0.433**	N/A[1]	N/A[2]	**0.466**	**0.535**	**0.577**
Patient2	0.5	0.130	0.129	**0.117**	**0.116**	0.129	0.179	0.423	0.413	0.427	0.489	0.179	0.423	0.466	**0.535**	**0.602**
Patient2	0.75	0.132	0.129	**0.117**	**0.116**	0.129	0.513	0.518	0.413	0.435	0.522	0.663	0.654	0.466	0.536	0.624
Patient2	1	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.543	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	0.635
Patient2	1.5	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.544	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	**0.635**
Patient2	2	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.544	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	**0.635**
Patient2	2.5	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.544	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	**0.635**
Patient2	none	0.132	0.129	**0.117**	**0.116**	0.129	0.526	0.544	**0.413**	**0.435**	0.522	0.664	0.663	**0.466**	**0.536**	**0.635**

Figure B.91: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient3	0	0.100[3]	0.097	0.101	0.100	0.100	0.145[3]	0.306	**0.424**	**0.420**	**0.422**	0.145[3]	0.441	**0.565**	**0.577**	**0.597**
Patient3	0.25	0.100	0.102	0.101	0.100	0.102	0.347	**0.450**	0.426	**0.436**	0.428	0.592	0.621	0.565	0.581	0.604
Patient3	0.5	0.101	0.102	0.101	0.100	0.102	0.441	0.456	0.427	0.443	0.439	0.639	0.625	0.565	0.581	0.604
Patient3	0.75	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient3	1	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient3	1.5	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient3	2	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient3	2.5	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**
Patient3	100	0.101	0.102	0.101	0.100	0.102	0.443	0.460	0.435	0.443	0.439	0.640	0.625	**0.565**	**0.581**	**0.604**

Figure B.92: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient4	0	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]
Patient4	0.25	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[0]
Patient4	0.5	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A[1]	N/A[0]	N/A[0]	N/A[0]
Patient4	0.75	N/A[0]	**0.028**	0.007[3]	N/A[0]	N/A[0]	N/A[0]	**-0.440**	-0.482[3]	N/A[0]	N/A[0]	N/A[0]	**-0.369**	-0.257[3]	N/A[0]	N/A[0]
Patient4	1	N/A[0]	**0.050**	**0.043[4]**	N/A[0]	N/A[0]	N/A[0]	**-0.124**	-0.204[4]	N/A[0]	N/A[0]	N/A[0]	**-0.029**	**-0.032[4]**	N/A[0]	N/A[0]
Patient4	1.5	N/A[1]	**0.054**	0.048	N/A[1]	0.041[3]	N/A[1]	**-0.023**	**-0.079**	N/A[1]	-0.428[3]	N/A[1]	**0.270**	**0.221**	N/A[1]	-0.428[3]
Patient4	2	0.053	0.055	0.048	0.056	0.046	-0.014	0.002	-0.020	0.052	-0.062	0.032	**0.301**	**0.272**	0.072	0.198
Patient4	2.5	0.053	0.055	0.048	0.056	0.046	0.003	0.002	-0.020	0.067	-0.053	0.343	0.304	**0.272**	0.310	**0.273**
Patient4	100	0.053	0.055	0.048	0.056	0.046	0.003	0.002	-0.020	0.067	-0.053	0.343	**0.304**	**0.272**	0.316	**0.273**

Figure B.93: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient5	0	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.317	0.550	0.504	**0.423**	0.443	0.516
Patient5	0.25	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	0.5	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	0.75	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	1	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	1.5	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	2	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	2.5	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516
Patient5	100	0.081	0.082	0.070	0.070	0.075	0.344	0.314	**0.206**	0.276	0.332	0.550	0.519	**0.423**	**0.443**	0.516

Figure B.94: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient6	0	N/A[0]	N/A[1]	**0.085**	0.075[3]	N/A[0]	N/A[0]	N/A[1]	**0.323**	0.068[3]	N/A[0]	N/A[0]	N/A[1]	**0.475**	0.231[3]	N/A[0]
Patient6	0.25	N/A[1]	N/A[2]	**0.090**	0.084[3]	N/A[1]	N/A[1]	N/A[2]	**0.351**	0.324[3]	N/A[1]	N/A[1]	N/A[2]	**0.489**	0.419[3]	N/A[1]
Patient6	0.5	0.095[4]	0.097	0.091	0.095	0.091[4]	0.031[4]	**0.398**	**0.369**	**0.361**	0.215[4]	0.031[4]	**0.582**	**0.490**	**0.523**	0.325[4]
Patient6	0.75	0.096[4]	0.097	0.093	0.095	0.094	0.391[4]	0.398	0.372	0.383	0.359	0.435[4]	0.595	0.490	0.558	0.531
Patient6	1	0.096	0.097	**0.093**	0.095	0.095	0.406	0.405	0.372	0.388	0.388	0.594	0.595	0.490	0.579	0.557
Patient6	1.5	0.096	0.097	**0.093**	0.095	0.095	0.411	0.408	**0.384**	0.397	0.400	0.598	0.595	**0.490**	**0.581**	**0.564**
Patient6	2	0.096	0.098	**0.093**	0.095	0.095	0.411	0.409	**0.384**	0.397	0.401	0.598	0.595	**0.490**	**0.581**	**0.566**
Patient6	2.5	0.096	0.098	**0.093**	0.095	0.095	0.411	0.409	**0.384**	0.397	0.401	0.598	0.597	**0.490**	**0.581**	**0.566**
Patient6	100	0.096	0.098	**0.093**	0.095	0.095	0.411	0.412	**0.384**	0.397	0.401	0.598	0.597	**0.490**	**0.581**	**0.566**

Figure B.95: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient7	0	N/A[0]	N/A[2]	**0.040**	**0.050[4]**	N/A[0]	N/A[0]	N/A[2]	**-0.108**	**-0.199[4]**	N/A[0]	N/A[0]	N/A[2]	**0.033**	**0.076[4]**	N/A[0]
Patient7	0.25	N/A[0]	**0.062**	**0.058**	**0.055**	N/A[2]	N/A[0]	**-0.098**	**-0.026**	**-0.048**	N/A[2]	N/A[0]	**0.253**	**0.210**	**0.120**	N/A[2]
Patient7	0.5	N/A[2]	0.062	0.058	0.061	0.053	N/A[2]	**0.073**	**-0.026**	**-0.026**	**-0.035**	N/A[2]	**0.306**	0.211	**0.282**	**0.259**
Patient7	0.75	0.053[3]	0.062	0.060	0.062	0.058	-0.378[3]	**0.082**	**-0.012**	**0.021**	**0.032**	-0.378[3]	**0.306**	0.211	**0.282**	**0.302**
Patient7	1	0.062	0.063	0.060	0.062	0.060	-0.007	0.082	0.004	0.034	0.032	0.163	0.313	0.211	0.282	0.308
Patient7	1.5	0.063	0.064	**0.061**	0.062	**0.060**	0.097	0.082	0.004	0.048	0.079	0.314	0.314	0.211	0.283	0.308
Patient7	2	0.063	0.064	**0.061**	0.062	**0.061**	0.097	0.091	0.004	0.048	0.079	0.339	0.314	**0.211**	0.283	0.320
Patient7	2.5	0.063	0.064	**0.061**	0.062	**0.061**	0.097	0.091	0.008	0.048	0.079	0.339	0.319	**0.211**	0.283	0.328
Patient7	100	0.063	0.064	**0.061**	0.062	**0.061**	0.097	0.093	0.008	0.048	0.079	0.339	0.328	**0.211**	0.283	0.328

Figure B.96: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient9	0	N/A[0]	N/A[2]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	N/A	N/A	N/A[0]	N/A[0]	N/A[0]	N/A[2]	N/A[0]	N/A[0]	N/A[0]
Patient9	0.25	N/A[0]	0.054[3]	N/A[0]	N/A[0]	N/A[0]	N/A[0]	-0.030[3]	N/A	N/A[0]	N/A[0]	N/A[0]	0.029[3]	N/A[0]	N/A[0]	N/A[0]
Patient9	0.5	N/A[0]	**0.093**	**0.089**	N/A[0]	N/A[1]	N/A[0]	**0.380**	**0.278**	N/A[0]	N/A[1]	N/A[0]	**0.506**	**0.497**	N/A[0]	N/A[1]
Patient9	0.75	N/A[0]	**0.096**	**0.095**	N/A[1]	N/A[2]	N/A[0]	**0.432**	**0.391**	N/A[1]	N/A[2]	N/A[0]	**0.562**	**0.593**	N/A[1]	N/A[2]
Patient9	1	N/A[0]	**0.098**	**0.095**	**0.091**	**0.092**	N/A[0]	**0.432**	**0.418**	**0.172**	N/A	N/A[0]	**0.562**	**0.606**	**0.172**	**0.236**
Patient9	1.5	0.112	**0.100**	**0.097**	0.101	0.105	0.396	0.432	0.440	0.442	0.454	0.396	0.568	**0.607**	**0.602**	0.569
Patient9	2	0.112	**0.100**	**0.097**	0.101	0.105	0.501	**0.436**	**0.440**	0.449	0.463	0.684	**0.597**	**0.607**	**0.623**	**0.571**
Patient9	2.5	0.112	**0.100**	**0.097**	0.101	0.105	0.501	**0.443**	**0.440**	0.449	0.463	0.684	**0.597**	**0.607**	**0.623**	**0.589**
Patient9	100	0.112	**0.100**	**0.097**	0.101	0.105	0.501	**0.443**	**0.440**	0.449	0.463	0.684	**0.597**	**0.607**	**0.623**	**0.589**

Figure B.97: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient10	0	N/A[0]	N/A[0]	**0.097**	N/A[2]	N/A[0]	N/A[0]	N/A[0]	**0.363**	N/A[2]	N/A[0]	N/A[0]	N/A[0]	**0.535**	N/A[2]	N/A[0]
Patient10	0.25	N/A[0]	0.038[3]	**0.100**	**0.091[4]**	N/A[0]	N/A[0]	-0.033[3]	**0.398**	**0.201[4]**	N/A[0]	N/A[0]	-0.033[3]	**0.569**	**0.302[4]**	N/A[0]
Patient10	0.5	N/A[0]	**0.104**	**0.101**	**0.097**	N/A[1]	N/A[0]	**0.379**	**0.408**	**0.345**	N/A[1]	N/A[0]	**0.435**	**0.575**	**0.469**	N/A[1]
Patient10	0.75	N/A[2]	0.104	0.102	0.102	0.097[4]	N/A[2]	**0.420**	**0.408**	**0.363**	0.112	N/A[2]	**0.529**	**0.585**	**0.549**	0.112[4]
Patient10	1	0.106	0.104	0.102	0.102	0.100	0.120	**0.425**	**0.413**	**0.383**	**0.230**	0.120	**0.611**	**0.585**	**0.549**	**0.230**
Patient10	1.5	0.106	**0.104**	**0.103**	**0.104**	**0.103**	0.317	**0.428**	**0.417**	**0.405**	0.381	0.361	**0.615**	**0.589**	**0.580**	**0.561**
Patient10	2	0.106	**0.104**	**0.103**	**0.104**	**0.104**	0.423	0.433	0.417	0.424	0.392	0.630	0.615	0.596	0.587	0.589
Patient10	2.5	0.106	**0.105**	**0.103**	**0.104**	**0.104**	0.440	0.433	0.417	0.432	0.428	0.636	0.615	**0.597**	**0.601**	**0.589**
Patient10	100	0.106	**0.105**	**0.103**	**0.104**	**0.104**	0.440	0.436	0.423	0.433	0.434	0.636	**0.620**	**0.599**	**0.601**	**0.599**

Figure B.98: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient11	0	N/A[2]	0.069[3]	0.091	0.023[3]	N/A[2]	N/A[2]	0.188[3]	**0.359**	0.160[3]	N/A[2]	N/A[2]	0.188[3]	**0.542**	0.160[3]	N/A[2]
Patient11	0.25	0.098	0.099	0.091	0.094	0.097	0.341	0.353	0.366	0.304	0.341	0.469	0.482	**0.549**	0.319	0.351
Patient11	0.5	0.100	0.100	**0.091**	0.096	0.098	0.400	0.406	0.370	0.386	0.415	0.499	0.599	0.549	0.532	0.598
Patient11	0.75	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.393	0.419	0.598	0.601	**0.549**	**0.536**	0.599
Patient11	1	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.398	0.419	0.604	0.601	**0.549**	**0.536**	0.599
Patient11	1.5	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.398	0.419	0.604	0.601	**0.551**	**0.536**	0.599
Patient11	2	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.398	0.419	0.604	0.601	**0.551**	**0.536**	0.599
Patient11	2.5	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.398	0.419	0.604	0.601	**0.551**	**0.536**	0.599
Patient11	100	0.100	0.100	**0.091**	0.096	0.098	0.408	0.416	**0.370**	0.398	0.419	0.604	0.601	**0.551**	**0.536**	0.599

Figure B.99: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.

		Best LSI					Best L					Best LCI				
	HSI UB	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM	basic	HSI	V indices	DTMR	DLDM
Patient12	0	N/A[0]	N/A[1]	**0.066**	N/A[2]	0.071[3]	N/A[0]	N/A[1]	**0.252**	N/A[2]	0.029[3]	N/A[0]	N/A[1]	**0.459**	N/A[2]	0.029[3]
Patient12	0.25	N/A[1]	0.042[4]	0.068	0.028	**0.078**	N/A[1]	0.324[4]	**0.308**	-0.163	**0.280**	N/A[1]	0.423[4]	**0.469**	**-0.074**	**0.280**
Patient12	0.5	N/A[1]	**0.081**	0.070	0.076	**0.078**	N/A[1]	**0.348**	0.308	0.286	0.336	N/A[1]	**0.527**	**0.469**	**0.424**	**0.473**
Patient12	0.75	0.076[4]	0.082	0.072	0.078	0.078	0.288	**0.359**	0.316	0.335	0.350	0.288	0.530	0.469	0.519	0.515
Patient12	1	0.078	0.083	0.072	0.078	0.078	0.341	**0.359**	**0.316**	0.335	0.350	0.454	0.530	0.478	0.519	0.515
Patient12	1.5	0.078	0.083	0.072	0.078	0.078	0.344	0.359	**0.316**	0.343	0.350	0.541	0.530	0.478	0.520	0.515
Patient12	2	0.078	0.083	0.072	0.078	0.078	0.344	**0.359**	**0.316**	0.343	0.350	0.541	0.530	0.478	0.520	0.515
Patient12	2.5	0.078	0.083	0.072	0.078	0.078	0.344	**0.359**	**0.316**	0.343	0.350	0.541	0.530	0.478	0.520	0.515
Patient12	100	0.078	0.083	0.072	0.078	0.078	0.344	**0.359**	**0.316**	0.343	0.350	0.541	0.530	0.478	0.520	0.515

Figure B.100: Overview of median best achieved LCI and LSI values and median achieved L value for different HSI upper bounds. If the reported average is significantly different from the basic configuration of BRIGHT, then the result is shown in bold and underscored. If N/A is reported, then no treatment plans that adhere to the shown bound were found. If a result is shown in green, then it significantly improves upon the base configuration of BRIGHT, if it deteriorates then it it shown in orange. If not all runs resulted in treatment plans that adhered to the upper bound then the number of runs for which it did find plans is reported in the square brackets behind the reported value.