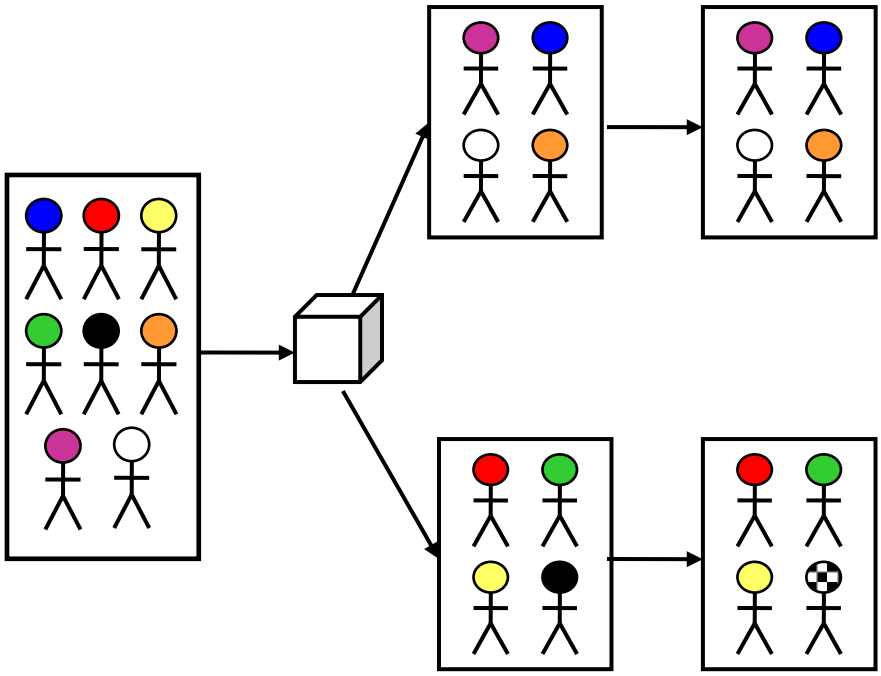
Comparison of Study Designs of Randomized Controlled Trials for the Evaluation of New Diagnostic Imaging Technology



Anke Hutzschenreuter

Vrije Universiteit Amsterdam
Faculty of Science
De Boelelaan 1081a
1081 HV Amsterdam

Erasmus Medical Center
Department of Epidemiology and Biostatistics
Dr. Molenwaterplein 50
3015 GE Rotterdam

Preface

The Swiss artist and designer Paul Klee stated in his creative credo in 1920:

"Art does not reproduce the visible; rather, it makes visible."

This also holds true for the research group for the assessment of radiological technology (ART) at the Department of Epidemiology and Biostatistics of the Erasmus Medical Center in Rotterdam. The group strives to improve decision making in health-care, especially related to radiology and to cardiovascular disease. A systematic and quantitative approach is applied to describe and analyze decision problems. An explicit way of making decisions helps to illuminate the main points of controversy, clarifies important aspects of the problem and identifies the need of further research.

The leader of this research group, Prof. Hunink, gave me the opportunity to perform this study within an internship where I could combine the knowledge I obtained during my studies of Business Mathematics and Informatics (BMI) at the Vrije Universiteit in Amsterdam with my interests in medicine and medical decision making. The paper you hold in your hands is the final report of the project.

I hope the reader will enjoy this paper.

Rotterdam, July 2004

Anke Hutzschenreuter

Abstract

The goal of this paper was to evaluate and compare different designs of randomized controlled trials under a realistic study setting. Peripheral Arterial Disease was chosen as scenario for the analysis of two randomization strategies. According to the first study design, patients are randomized across diagnostic imaging strategies. According to the second design, all patients undergo all diagnostic imaging strategies and are then randomized between providing test results versus not providing test results. Based on data obtained from a recently completed trial on the initial diagnostic work-up of peripheral arterial disease, a clinically based discrete event simulation model was developed.

The randomization strategies were analyzed for different study sizes, proportions of participating patients and varying capacity for the imaging procedures. The performance was measured in terms of research costs, duration of the trial and utilization of the imaging modalities.

The results suggest that the research costs form an important part of the total costs of a study (about 33%). The second study design appeared to be more expensive in all scenarios considered but also showed a higher utilization of the imaging modalities. For a baseline eligibility of 50% the difference in the mean duration of the two designs is a few weeks, increasing to almost 4 months if 70% of the patients are eligible. Reducing the weekly number of reserved time slots had a stronger impact on the performance of the second randomization strategy than on the first one. The results of the sensitivity analysis emphasize that the proportion of patients that are eligible is a crucial factor when setting up and performing a randomized controlled trial.

Table of Contents

PREFACE 3

ABSTRACT 4

TABLE OF CONTENTS 6

1 INTRODUCTION 8

1.1 ERASMUS MEDICAL CENTER ROTTERDAM 8
1.2 OBJECTIVE OF THE STUDY WITHIN THE SCOPE OF MEDICAL DECISION MAKING 9

2 MEDICAL AND METHODOLOGICAL BACKGROUND 11

2.1 RANDOMIZED CONTROLLED TRIALS 11
2.1.1 RANDOMIZATION STRATEGIES TO EVALUATE DIAGNOSTIC IMAGING TESTS 13
2.1.1.1 Randomization across diagnostic testing strategies 14
2.1.1.2 Randomization between Providing Test Results versus Not Providing Test Results 15
2.1.1.3 Randomization of Discordant Test Results 16
2.2 PERIPHERAL ARTERIAL DISEASE (PAD) 17
2.2.1 WHAT IS PERIPHERAL ARTERIAL DISEASE? 17
2.2.2 HOW IS PAD DIAGNOSED? 18
2.2.3 HOW IS PAD TREATED? 20

3 MATERIALS AND METHODS 21

3.1 DIPAD TRIAL 21
3.2 INTRODUCTION TO DISCRETE-EVENT SIMULATION (DES) 23
3.3 SIMULATION MODEL 24
3.3.1 STRATEGY 1 – RANDOMIZATION ACROSS DIAGNOSTIC IMAGING STRATEGIES 25
3.3.2 STRATEGY 2 – RANDOMIZATION BETWEEN PROVIDING TEST RESULTS VERSUS NOT PROVIDING TEST RESULTS 25
3.4 MODELING ASSUMPTIONS 26
3.5 VALIDATION OF THE MODEL 29

4 RESULTS 31

4.1 RESULTS FOR THE FIRST RANDOMIZATION STRATEGY 31
4.1.1 RESULTS FOR VARYING NUMBER OF INCLUDED PATIENTS AND PROPORTION OF ELIGIBLE PATIENTS 31
4.1.2 RESULTS FOR VARYING NUMBER OF DIAGNOSTIC IMAGING MODALITIES 35

4.2 RESULTS OF THE SECOND RANDOMIZATION STRATEGY **38**
4.2.1 RESULTS FOR VARYING NUMBER OF INCLUDED PATIENTS AND PROPORTION OF ELIGIBLE PATIENTS FOR RANDOM SEQUENCE OF DIAGNOSTIC IMAGING TESTS 39
4.2.2 RESULTS FOR VARYING NUMBER OF DIAGNOSTIC IMAGING MODALITIES 43
4.2.3 RESULTS FOR DIFFERENT SEQUENCES OF DIAGNOSTIC IMAGING TESTS 44
4.3 COMPARING RANDOMIZATION STRATEGIES **45**
4.3.1 RESULTS FOR VARYING NUMBER OF INCLUDED PATIENTS AND PROPORTION OF ELIGIBLE PATIENTS FOR RANDOM SEQUENCE OF DIAGNOSTIC IMAGING TESTS 45
4.3.2 RESULTS FOR VARYING NUMBER OF DIAGNOSTIC IMAGING MODALITIES 49

5 DISCUSSION **52**

6 CONCLUSIONS **59**

ACKNOWLEDGEMENTS **62**

REFERENCES **63**

APPENDICES **64**

APPENDIX A - COSTS INCLUDED IN DES MODELS **64**
APPENDIX B – SCREENSHOTS OF DES MODELS **66**
APPENDIX C – SUMMARY PROGRAMMING CODE OF SIMULATION MODELS **68**
APPENDIX D – STATISTICAL DATA ANALYSIS OF INPUT DATA **72**
APPENDIX E – TABULATED RESULTS OF SIMULATION EXPERIMENTS **77**

1 Introduction

In this first chapter, the context of this project will be outlined which includes a description of the organization and the objective of the project within the scope of medical decision making.

1.1 Erasmus Medical Center Rotterdam

The Erasmus Medical Center is the Netherlands' largest academic medical center with more than 9,000 employees, over 1000 clinical beds and 53 scientific departments and institutes. The core activities of Erasmus MC are patient care, education and research. At the hospital, patient care is the core activity; the core activities at the medical faculty are education and research. Every medical department has an educational purpose and also conducts scientific research. Erasmus MC's research covers the entire spectrum from fundamental non-clinical research to patient related research.

The Department of Epidemiology & Biostatistics offers research consultancy facilities for clinicians of the Erasmus Medical Center Rotterdam in clinical epidemiology and biostatistics. The research activities of the department are divided into the following clusters: epidemiology of diseases, basic and clinical epidemiology.
The clinical epidemiology group collaborates with the department of radiology in a joint research program for the Assessment of Radiological Technology (ART program) where this project was carried out. The ART program comprises a network of researchers who focus on the assessment of medical imaging technology, especially related to cardiovascular disease. The research performed is based on methods from clinical epidemiology, decision sciences, and medical technology assessment. Methodological research focuses on developing the methods and study design for evaluating diagnostic and therapeutic imaging procedures.

1.2 Objective of the study within the scope of medical decision making

In our daily life we have to face difficult decisions day by day. For instance, we have to decide whether to use our savings to buy a new computer or to go on vacation. Also in the health care sector decisions need to be made. It becomes increasingly important how to use the limited health care budget in order to ensure that as many patients as possible benefit from optimal care.

Medical decision making is a science that investigates how decisions are made in clinical practice and how they can be improved. Making decisions we have to make trade-offs between risks, costs, patient preferences, etc., and take into account the rapidly increasing evidence available.

During the last years the number of medical publications increased rapidly. 20,000 articles on randomized controlled trials (RCTs) were published in the year 2003. RCTs provide the best type of information on which to base medical decisions. They are widely accepted as the gold standard for comparing different therapeutic and diagnostic options.

Apart from the evidence obtained by a RCT, also the operational and financial control while performing such a trial are important aspects. Incorrectly planned studies can lead to excessive overrun of the assigned budget and/or poor (delayed) evidence. In many cases this overrun is caused by unexpected fluctuations of the patient inclusion process. But also the study design itself plays an important role. Study designs can differ greatly both in a qualitative and a quantitative way.

The thesis will address the methodological issue of the planning of medical trials using descriptive and simulation methods.
The paper will proceed as follows. First a description of the medical and methodological background of the project will be provided in section 2, including

descriptions of the randomization strategies under consideration. This is followed by a depiction of the research methodology and the results from the simulation study in sections 3 and 4 respectively. Finally the results will be discussed in combination with the conclusions we could draw from the results.

2 Medical and methodological background

In this chapter some background information will be provided in the field of the medical application and the design of medical trials.

2.1 Randomized controlled trials

A randomized controlled trial (RCT) is a special form of a clinical trial, a scientific study in order to assess the effect of a new procedure or drug [1].

One can distinguish between RCTs for the evaluation of a new treatment, which will be outlined first, and of a diagnostic imaging test.

Before a new drug is established in the market, clinical trials need to be done. Patients are typically divided into two groups, one treated with the new drug and the other not in order to study the effect of treatment. A carefully chosen *control group* is required to make a meaningful comparison. A control group undergoes the same routine (seeing a doctor, taking pills at the same time, etc.) but does not receive the treatment. This control group should receive either no treatment (e.g. placebo, which is sugar pills) or receive the current standard treatment (if, for example, it would be unethical not to treat their disease at all).
Randomization between the experimental and the control treatment is required to avoid selection bias. For example, in the treatment of coronary artery disease (a disease process by which the coronary arteries become narrowed or completely occluded) elderly and frail patients may preferentially be selected for stent placement (insertion of a metal grid to prevent occlusion of a vessel after balloon dilatation) as opposed to coronary artery bypass surgery. This would lead to groups that are not balanced for age and other diseases and would give one group an unfair disadvantage. Therefore, without randomization the results could be biased. Of course this is undesirable and a random allocation of the groups could avoid this bias.

Patients are selected for a clinical trial according to *inclusion/exclusion criteria* that should guarantee the selection of a representative group of patients. The inclusion criteria should not be too restrictive, e. g. all patients with suspected PAD older than 40 years of age. The exclusion criteria are used to filter out the exceptions, e. g. patients with a pacemaker are excluded from a study on PAD where MRA (magnetic resonance angiography) is one of the diagnostic imaging strategies evaluated. For establishing the inclusion and exclusion criteria, a trade-off is needed between a homogenous group and a sufficiently large study population with generalizable results.

In an *open* clinical trial both doctor and patient knows which treatment is provided. For this study design a certain bias is unavoidable because a doctor is likely to see what he wants to see. Alternatively, *blinding* can be used which means that the doctor (and/or the patient) is blinded to information about the treatment. In single-blinded trials the patient is unaware which treatment he/she receives. In double-blinded trials neither the doctor nor the patient knows which treatment is provided. Ideally RCTs are performed with double-blinding as the influence of external factors is minimized and thus the bias of the results is reduced. In RCTs for a new treatment, blinding can be achieved by using a so-called *placebo*, a sugar-pill without clinical effect, of the same size, color and taste, and given in the same frequency and amount as the “real” drug.

As for the assessment of new diagnostic imaging technology, a hierarchical approach to the development, the assessment and the implementation has traditionally been used. The design is as follows. After the technical development, the diagnostic performance is studied on a limited number of cases. This usually involves performing a cohort study, which is a longitudinal study where patients (the cohort) undergo the imaging tests and a reference standard test. The cohort may also be followed over time prospectively to measure the development of different outcomes (disease state, etc.).

In the next step, the diagnostic, therapeutic and prognostic impact is studied and the effectiveness is evaluated. For example, the impact on the diagnosis could be that the old technique should not be replaced by the new one or that the new technique should only be used in addition to the old one. As for the effectiveness, outcomes related to patient and society like maximum walking distance or quality of life could be studied. A possible study design for this step could be a RCT.

Finally, before the new technology is implemented the effect on cost and effectiveness outcomes is evaluated. This is usually done with a decision analytic model that integrates all the available evidence and explores the influence of uncertainty on the decision.

This process is time-consuming and often cannot keep up with the rapid technological advances. By the time the hierarchical approach is completed, the new technology could already be implemented because of the “beautiful images” it produces or out of date due to even newer techniques.

In order to perform a reliable study that evaluates the diagnostic strategy in a faster way, integration of research and clinical practice is needed. Hunink and Krestin [2] advocate an empirically based RCT design whenever feasible and ethical. They further propose that the current clinical practice should be used as control strategy.

Blinding and especially double-blinding is more complicated in a diagnostic RCT than in a therapeutic RCT. For example, an average patient knows the difference between a CT and a MRI and will therefore not be blinded to the procedure he or she has to undergo. The same holds true for the doctor as the resulting images are different and a radiologist will recognize them immediately.

2.1.1 Randomization strategies to evaluate diagnostic imaging tests

The strategies that will be studied in this project are based on the article by Hunink and Krestin [2]. Here only dichotomous test results are considered, but of course these strategies can also be applied to tests with multicategory results.

2.1.1.1 Randomization across diagnostic testing strategies

The following strategy (Fig. 3) can be considered as the classical way of performing a RCT.

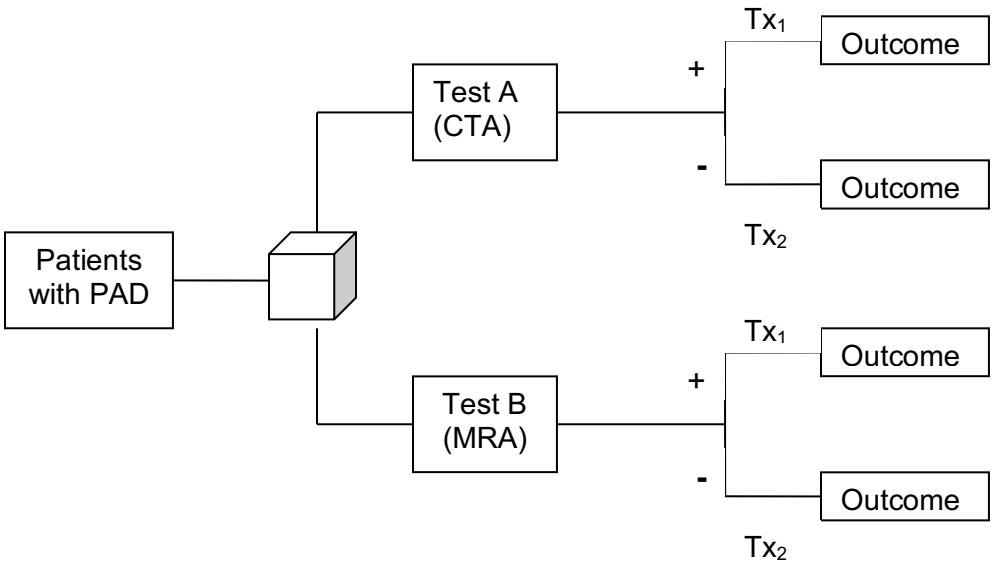


Figure 3 - Randomization across diagnostic imaging strategies

This design uses random allocation of the consenting patients to either test A or test B. The randomization is denoted with a cube. The test result is assumed to be dichotomous, "+" stands for a positive and "-" for a negative test result. Based on the information obtained by the test, further management is determined. To simplify matters, there are only two treatment options resulting from the test result, denoted "Tx_1" and "Tx_2". For example, "Tx_1" could be surgery and "Tx_2" medication. The effectiveness of the tests is obtained by comparing the outcomes of the respective groups.

According to Bossuyt, Lijmer et al. this design is not always efficient [3]. The outcome of patients with concordant test results would only be influenced by the choice of treatment, not by the choice of test. Only patients with a negative test

result on one test and a positive result on the other test can contribute to a difference in the mean outcome between the two groups.

Another disadvantage of this design is that the information obtained by the trial can only be used to compare the diagnostic imaging strategies. It cannot provide information about the agreement between the test results nor about agreement between sensitivity and the specificity of each test. The test results can only be determined if the results of both tests are available for every patient and sensitivity and specificity can only be determined if in addition a reference standard has been performed in all patients.

2.1.1.2 Randomization between Providing Test Results versus Not Providing Test Results

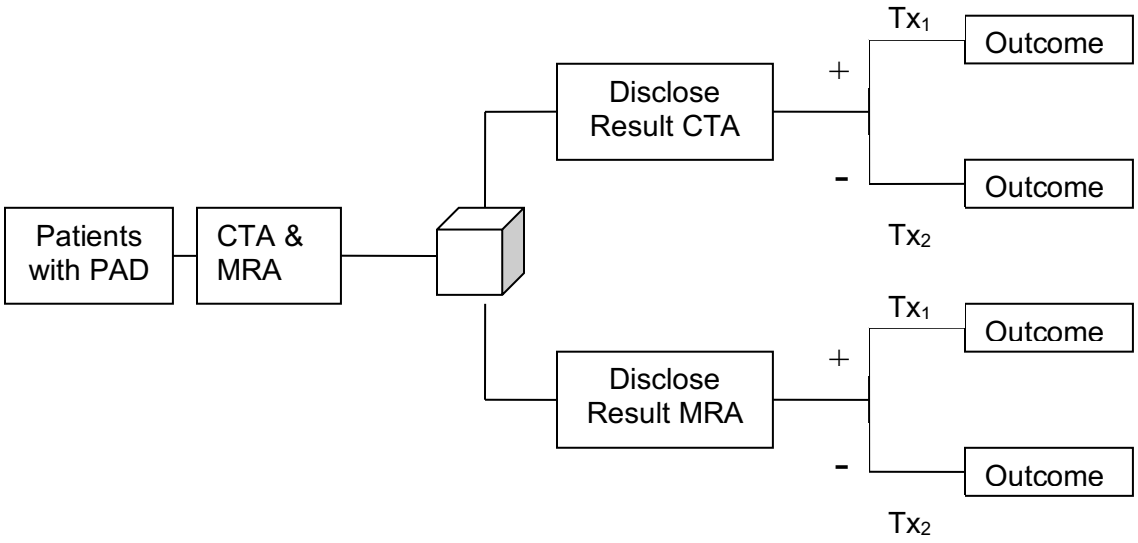


Figure 4 - Randomization between Providing Test Results versus Not Providing Test Results

According to this randomization strategy, all patients undergo all diagnostic tests but only one test result is disclosed to the treating physician. All patients are monitored during follow-up and their respective outcome is included in the evaluation of the testing strategies.

As with the previous design, this design is not always efficient. Only if the subsequent management is different the outcome will be affected. In many fields of interest it is uncommon that every patient undergoes two diagnostic imaging tests. Therefore, the study would incur additional testing and the costs are higher. Another disadvantage is that physicians are more likely to request the result of the other test as both tests have been performed. This, of course, affects the treatment and therefore the outcome.

2.1.1.3 Randomization of Discordant Test Results

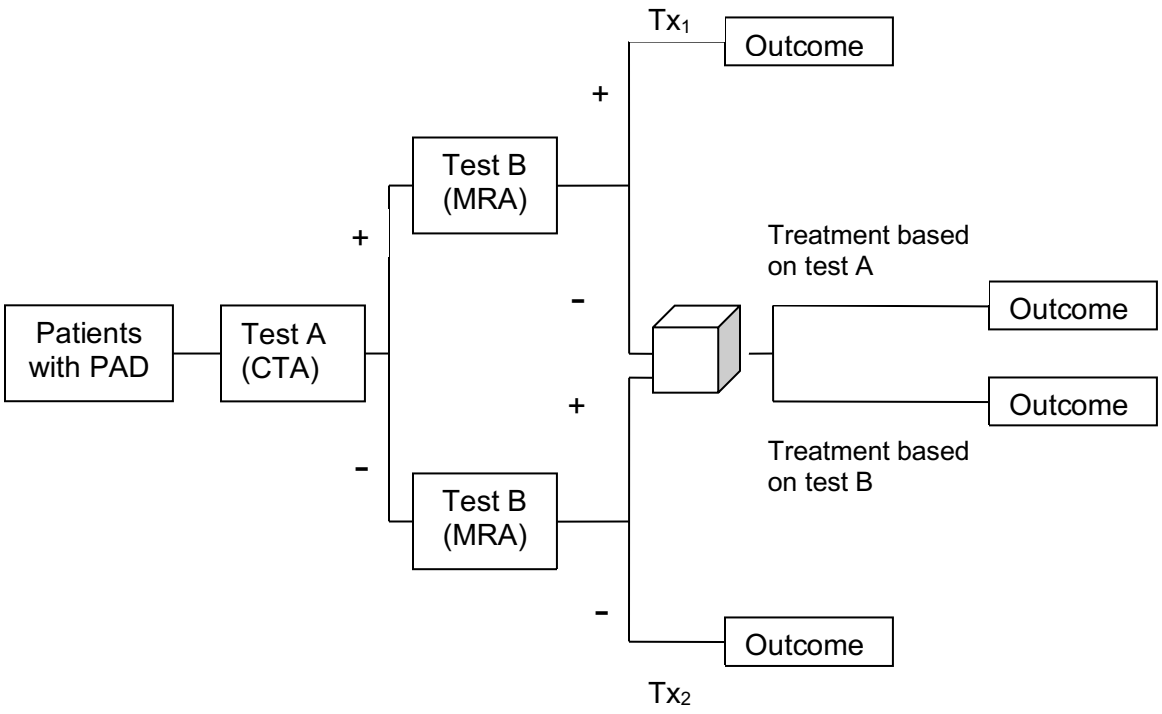


Figure 5 - Randomization of discordant test results

In the study design described in Fig. 5, all patients undergo both diagnostic tests and patients with concordant test results (i.e. same test result on both tests) are treated according to the test result. Their follow-up can be monitored but they will not contribute to the comparison of the tests. The choice of treatment for the patients with discrepant test results is randomly made on the basis of test A or test B.

In this design the knowledge that the test A suggested another treatment than test B could affect the decision of the treating physician which may introduce a bias.

Another disadvantage of this design is that it does not depict reality in the daily health care of patients with PAD. Usually, each patient would undergo only one test.

2.2 Peripheral arterial disease (PAD)

Peripheral arterial disease (PAD) was chosen as scenario for the evaluation of different randomization strategies of a randomized controlled trial (RCT) of diagnostic imaging technology.

2.2.1 What is peripheral arterial disease?

In peripheral arterial disease (PAD), fatty deposits build up along artery walls and affect blood circulation, mainly in arteries leading to the legs and feet. It is a manifestation of atherosclerosis in the lower extremities. In its early stages a common symptom is cramping or fatigue in the legs and buttocks while walking or climbing stairs. Such cramping pain stops when the person stands still. This is called "intermittent claudication". When the disease progresses severe symptoms as critical ischemia can appear. In general, patients with PAD have a higher risk to die from stroke and heart attack.

2.2.2 How is PAD diagnosed?

Diagnostic imaging work-up is necessary to localize the narrowed arterial segment (also called stenosis) for revascularization procedures. It provides a "map" of the arterial system for the radiologist and/or the surgeon for treatment. Techniques used to diagnose PAD include Doppler ultrasound, X-ray angiography (also called intraarterial digital subtraction angiography (DSA)), computed tomographic angiography (CTA) and magnetic resonance angiography (MRA).

Traditionally, X-ray angiography has been used as diagnostic imaging modality. It is still considered to be the reference standard (with assumed 100% sensitivity and specificity). During this test a contrast agent is injected into the artery and X-rays are taken to show arteries of the legs and any blockages that may be present. A disadvantage of DSA is that a risk is associated with performing the test (mortality of 3 out of 10,000 and morbidity of 3 out of 100) because of the intraarterial injection of iodinated contrast media. Compared with other imaging modalities it is expensive and requires a period of bed rest and observation after the procedure.

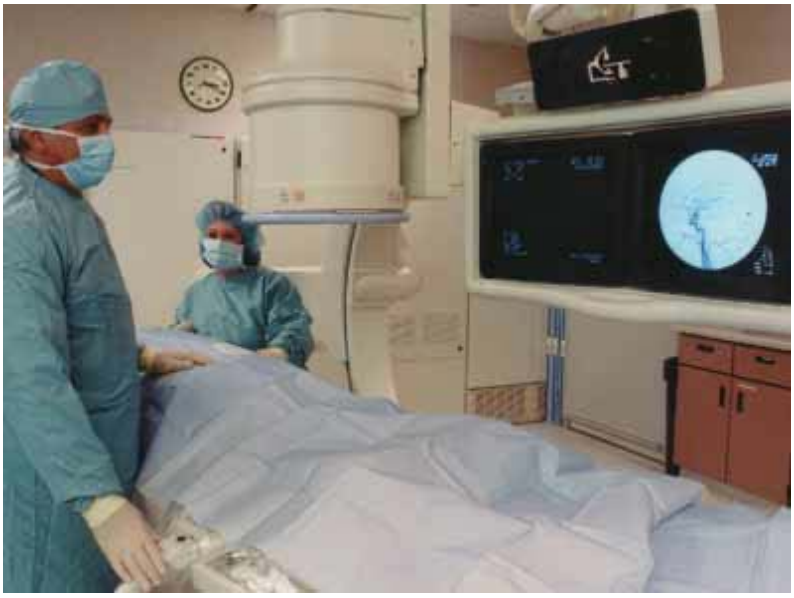


Figure 1 - Picture of an angiography suite

Computed tomography angiography (CTA) is a cross-sectional X-ray imaging technique that can be used to image the arteries of the abdomen, the pelvis and the legs at moderate costs. There are low short- and long-term risks associated with the exposure to radiation and allergic reactions to the contrast material.

Figure 2 – Pictures of a CT (left) and a MR scanner (right)

Magnetic resonance angiography (MRA) is a relatively new non-invasive diagnostic technique that gives similar information to that of a CT without the use of X-rays. It uses non-iodinated contrast agents and can produce high-quality images without any known hazards. Disadvantages are the relatively high costs, the occurrence of uninterpretable images and contra-indications like a pacemaker and claustrophobia.

2.2.3 How is PAD treated?

Percutaneous transluminal angioplasty (PTA) is an intervention to enlarge a narrowed blood vessel. A thin tube called a catheter with a deflated balloon on its tip is passed into the narrowed region. The balloon is inflated at the site of the stenosis in order to open the stenosis and is subsequently deflated. This procedure may include positioning of a stent – a tiny metal cylinder– to keep the vessel open and prevent the recurrence of stenosis.

Surgical options for PAD include an aorta-bifurcation prosthesis, bypassing (BP) below the groin, or amputation of the diseased part of the leg. Bypassing means that a vein or a synthetic blood vessel is grafted to reroute the blood flow around the narrowed part of the arteries.

The third option for treating PAD is exercise therapy to improve the symptoms.

All patients receive medication for the disease. This involves aspirin, medical treatment of the associated risk factors of PAD such as high cholesterol levels or hypertension and lifestyle interventions such as smoking cessation and exercise in order to reduce the risk of myocardial infarction or stroke.

3 Materials and Methods

In this study we developed a discrete-event simulation model for two study designs of a RCT for PAD. Specific issues to be addressed by the simulation study are:

- How much would a certain study design cost? What part of the total costs form the costs related to research?
- How long would a study take for a given number of patients to be included in the study? And how expensive would this trial be?
- What is the effect of varying the proportion of eligible patients on the costs and on the duration?
- Which logistical environment is needed for efficiently performing a RCT including a given number of patients?

The simulation model was based on the data obtained by the recently finished DIPAD (Diagnostic Imaging for PAD) trial carried out as a multi-center study by R. Ouwendijk at the Department of Radiology and the Department of Vascular Surgery at the Erasmus Medical Centre, Rotterdam, in cooperation with the Department of Radiology at the University Hospital of Maastricht. For the evaluation of the randomization strategies only the data obtained at the ErasmusMC were used.

3.1 DIPAD trial

The objective of this trial was to evaluate the clinical utility, patient outcomes and costs of MRA compared to CTA as the initial imaging test in the diagnostic work-up of patients with peripheral arterial disease (PAD).

According to the study design, patients were randomized across the different imaging strategies; see section 2.2.1.1 for a description. In order to resemble daily practice no blinding to clinical information was used for this trial.

Clinical utility was measured as the therapeutic confidence of the physicians involved on a ten-point rating scale and as the need of additional imaging 60 days after the initial test.

Patient outcomes included medical measurements like the ankle-brachial index (ABI), the maximum walking distance and the change in clinical status but also the health related quality of life using different quality of life questionnaires.

According to the hospital perspective, costs related to the diagnostics, the therapies/treatments, outpatient visits, etc. were measured. Also directly and non-directly assignable costs were included as personnel, material, equipment costs and supporting department, housing and overhead costs respectively. Most of the costs were assessed by a cost analysis. For the other costs national estimates were used according to the Dutch guidelines for cost calculations.

Costs (Euro)	**Mean Costs (SD)**	
	MRA group (n=77)	**CTA group (n=79)**
Total diagnostic unit costs	676 (477)	317 (477)
Therapeutic unit costs		
Percutaneous interventions	1379 (1834)	1078 (1636)
Surgical procedures	4594 (8987)	2476 (4694)
Outpatient visits	198 (57)	192 (57)
Total	**6848 (8957)**	**4064 (4521)**

Table 1 - Cost results from DIPAD trial for the ErasmusMC; SD – standard deviation

For a detailed composition of the costs see Appendix A. These costs were included in the discrete-event simulation model both as time-dependant and fixed unit costs.

As for the patient related logistics, the potential participants were identified by the treating physicians using the provided inclusion and exclusion criteria. During the weekly outpatient clinic R. Ouwendijk assessed their eligibility. Then the patients were randomly allocated to the two groups. After entering the trial, there was a minimum time period of one week until the patients underwent the diagnostic imaging test. According to the operational planning, two time slots per week for the CTA and two for the MRA were assigned to the participants of the DIPAD trial. If the initial diagnostic imaging test did not provide sufficient information, further evaluation with other imaging technology was allowed. For example, if the image obtained by MRA was not possible due to claustrophobia of the patient, an invasive test could be done in order to obtain better information about the disease. The final decision about the appropriate treatment of the patient was achieved by consensus during the weekly vascular conference.

3.2 Introduction to Discrete-event simulation (DES)

In many application areas such as production planning, mathematical models need to be built as they offer the possibility to study an encountered phenomenon and to analyze a real-world situation, to forecast and to optimize under certain criteria. This can be achieved by using a number of tools and techniques, one of which is simulation. The Oxford English Dictionary describes simulation as:

> *"The technique of imitating the behavior of some situation or system (economic, mechanical, etc.) by means of an analogous model, situation, or apparatus, either to gain information more conveniently or to train personnel."*

Simulation represents a good alternative to direct experimentation when this is not feasible or expensive. By running “what if” experiments cause and effect relationships of the studied system can be approximated and used for further analysis.

Discrete-event simulation (DES) is a special simulation technique that enables us to observe the time based behavior of a system [4]. For the following description of DES, we will use the example of a physician's consulting hours.

In a DES model, a system is represented by state variables that change over time. An example for a state variable could be the status of the physician, he can be idle or busy. The points of time at which the state of the system may change are called events. An event could be the arrival of a new patient or the departure of a treated patient. If the doctor is busy with a patient and another one arrives, the state of the doctor is unaffected, he is still busy.

The key assumption of this simulation technique is that events happen at discrete points in time. The behavior of a river, for example, could not be modeled using discrete-event simulation. If we defined the amount of water in a specific part of the river as the state of the system, the state changes continuously.

Because the models change in a dynamic way, we must keep track of the current value of the simulated time. This is done using a simulation clock as a variable in our model. This variable is also used to advance the simulated time. Most of the simulation software available uses the next-time advance approach. According to this approach, the simulation clock is initialized as zero and the times of future events are determined. The clock is then advanced to the first of the future events. The state of the system is updated and the clock jumps to the time of the next event. This is done until a predefined stopping criterion is met. Using the next-time advance approach, inactive periods of the system are skipped.

3.3 Simulation model

We performed the evaluation of randomization strategies using a simulation model developed on the Simul8 software platform (Simul8 Cooperation) [5]. Simul8 was chosen based on the availability of this software program. See Appendix B for screenshots of the simulation models developed for the first and second randomization strategy.

3.3.1 Strategy 1 – Randomization across diagnostic imaging strategies

After the patients arrive in the system, a certain proportion of patients does not participate in the trial and leaves the system. When patients enter the trial they are randomly assigned to the two imaging strategies of the study. After a predefined number of patients have entered the trial, all arriving patients are routed to the non-participants. This is done by simple self-written programming code and makes sure that only the specified number of patients is included in the cost, duration and utilization calculations, see Appendix C. Both tests have a minimum waiting time of a week before the patients undergo the test. If the waiting time for the initial diagnostic imaging test exceeds eight weeks patients withdraw from the study. A long waiting time also influences the eligibility of the patients, in other words surgeons are likely to include fewer patients if an early diagnostic work-up is not guaranteed.

After the test, the images need to be post-processed and are then interpreted by a radiologist. This period is modeled as a minimum waiting time between one and a half and two weeks. The interpretation of the images includes scoring and dictating. Angiography can be interpreted at once and only needs dictating the report. After the initial test, the result is discussed during the weekly vascular conference and a follow-up period of six months starts. During follow-up the patients undergo the advised treatment and two outpatient visits are scheduled, one visit before the treatment and one afterwards. When the defined number of patients has reached the end of the follow-up, the stopping criterion of the model is met.

3.3.2 Strategy 2 – Randomization between providing test results versus not providing test results

The second randomization scheme differs from the first with respect to the tests to be undergone by the included patients. In this strategy, each patient undergoes both a CTA and a MRA. There are several options for the order of the tests:

(1) The sequence of the imaging tests is fixed (first CTA or MRA)

(2) The sequence of the imaging tests is assigned randomly

(3) The first imaging test is assigned alternately: CTA – MRA – CTA …

(4) Patients are assigned to the shortest queue

Both tests have a minimum waiting time of a week before the patients undergo the test. Patients who already underwent one test are given priority over other patients in the queue. If the waiting time for the entire diagnostic imaging work-up exceeds eight weeks patients withdraw from the study prematurely. As in the first strategy, there is a negative influence of long waiting times on the eligibility of the patients. Only one test result is provided to the treating physician according to which the therapeutic plan is established. If the information of the test is not sufficient, DSA is performed as additional work-up. The follow-up period starts after the test result of the second test is interpreted.

3.4 Modeling Assumptions

A model is a description of a (part of a) real-world system such that it allows an analysis of the aspects the model builder is interested in. In order to keep the model simple, assumptions about the system need to be made.

In this simulation study the arrival of patients suffering from PAD was modeled as a stochastic process. To fit a stochastic model we used the inclusion dates from the DIPAD trial. This results in an exponential distribution of the interarrival times with a parameter equal to 2.96 (unit: weeks), see Appendix D for details.

A certain proportion of the potential participants may not be eligible for the study. In the literature this effect is called Lasagna's Law, for a description see Commentary by Nesheim [6]. In our analysis we conducted a sensitivity analysis where we varied this proportion between 50, 60 and 70%. A percentage of 70% of eligible patients would be a relatively optimistic assumption whereas 50% is quite pessimistic for this disease. 60% of the patients participated in the DIPAD trial. The study size was set to 100, 150 and 500 respectively.

Long waiting times are assumed to have a negative effect on the willingness of the treating physicians to include patients. Unfortunately, the available data was too limited to demonstrate a consistent relation between waiting times and

patient eligibility. Therefore we assumed that the effect is represented in the following scheme:

Waiting time	**Effect on the proportion of eligible patients**
Waiting time > 5 wks	Eligibility reduced by 5 percentage points compared to the baseline eligibility
Waiting time > 6 wks	Eligibility reduced by 10 percentage points compared to the baseline eligibility
Waiting time > 7 wks	Eligibility reduced by 15 percentage points compared to the baseline eligibility
Waiting time > 8 wks	Eligibility reduced by 20 percentage points compared to the baseline eligibility

Table 2 – Effect of waiting time on patient eligibility

In the case of the first randomization strategy, the eligibility is decreased if the waiting time for at least one of the tests exceeds the limits given above. For the second strategy, the total waiting time for CTA and MRA is considered.
Non-participants immediately leave the system and are not included in the cost, duration and utilization calculations.
Furthermore, we made assumptions on the duration of the test itself, see Appendix D. At the radiology department of the Erasmus Medical Center, appointments for the CTA are scheduled every 15 minutes including the patients entering and leaving the examination room. MRA tests are planned every 45 minutes, DSA every two hours. Employing the Kolmogorov-Smirnov test at a confidence level of 95%, we could not find a distribution that fitted the data on the CTA, MRA and DSA duration. Therefore, we chose to model these durations as constants of 15, 45 and 120 minutes respectively. Using the Kolmogorov-Smirnov test with confidence level equal to 95%, we also saw that the durations of the interpretations of CTA and MRA were significantly different. A triangular

distribution with a minimum duration of 9 minutes, a mean of 22 minutes and a maximum duration of 44 minutes fitted the data best. As for the MRA, the duration was no significantly different to a lognormal random variable with the parameters 2.4428 and 0.6113.

The waiting time before the test result is interpreted, is assumed to follow a uniform distribution on the interval [1.5 weeks, 2 weeks].

As for the personnel required for the tests, we assumed that for both CTA and MRA two technologists were scheduled. During a diagnostic DSA one radiologist and one technologist are present. The costs associated with personnel can be found in Appendix A. DSA was assumed to be performed in 8% of cases after MRA and in 6% after CTA. These estimates were obtained from the DIPAD trial data.

During the weekly vascular conference, subsequent patient management was discussed and the final decision was achieved by consensus. This resulted in the following proportions:

- PTA with selective stent placement - 36.62% after CTA, 41.43% after MRA
- Surgery (67.3 %bypass surgery, 23.6% aorta-bifurcation prosthesis surgery, 9.1% amputation surgery) – 33.8% after CTA, 40% after MRA
- Conservative treatment (drug, advice to do exercise) – 29.6% after CTA, 18.6% after MRA

Directly after undergoing the diagnostic imaging test the follow-up period of half a year started including at least one outpatient visit, two questionnaires about the patients' quality of life after three and six months and the measurement of the ankle-brachial index and the maximum walking distance also after three and six months.

Apart from the scheduled outpatient visits also additional outpatient visits may be necessary due to the treatment or aggravations of the disease state. The number of additional visits was generated at the beginning of the follow-up period according to a probability profile obtained from the DIPAD patients.

We analyzed the two randomization strategy also for varying number of time slots available per week. The simulated scenarios are described in the following table. A middle-sized trial (150 patients included) with a moderate proportion of participating patients (60%) was chosen as baseline setting for the evaluation.

Name	*Arrangement*
2 CTA, 2 MRA	Two CTA and two MRA time slots per week reserved for trial (time slots per test directly after each other) – baseline scenario
1 CTA, 1 MRA	One CTA and one MRA time slots per week reserved for trial
½ CTA, ½ MRA	One CTA and one MRA time slots every two weeks reserved for trial

Table 3 - Description of the simulation scenarios

As performance measures we used the utilization of the test capacities, the costs for unused capacity, the total study costs, including the costs for the regular health care and research costs, and total duration of the trial. The costs for unused capacity are calculated on the basis of the costs for housing, equipment and personnel.

All results are based on 15 independent simulation runs of randomized controlled trials.

3.5 Validation of the model

Our model underwent various stages of validation. In the course of this project the author discussed the model numerous times with the members of the Assessment of Radiological Technology (ART) group at the Department of

Epidemiology and Biostatistics at the Erasmus Medical Center Rotterdam and the assigned supervisors from the Vrije Universiteit in Amsterdam. This was done to ensure the validity of the simulation model before and during the development of the model.

Once the model was completed, we validated the model by comparing the simulated output measures with the same types of output measures found in the data of the DIPAD trial by R. Ouwendijk. Measures were the number of reference tests performed and the health care unit costs. Furthermore, a sensitivity analysis was performed to validate the effect of long waiting times on the eligibility.

4 Results

The experimental results of the performance of different study designs are tabulated in Appendix E, Table 8 and 9.

4.1 Results for the first randomization strategy

The first randomization strategy was evaluated in terms of total costs and duration of the trial and utilization of the testing modalities.

4.1.1 Results for varying number of included patients and proportion of eligible patients

In this simulation scenario two time slots per week for CTA and MRA were reserved for performing the trial.

In the following figure, the mean total costs and the mean research costs of the trial (Fig. 6) are shown.

Generally speaking, the total costs for the trial and the research costs tend to increase when the proportion of eligible patients decreases. There are, however, differences between the different combinations of the number of included patients and proportion of eligible patients.

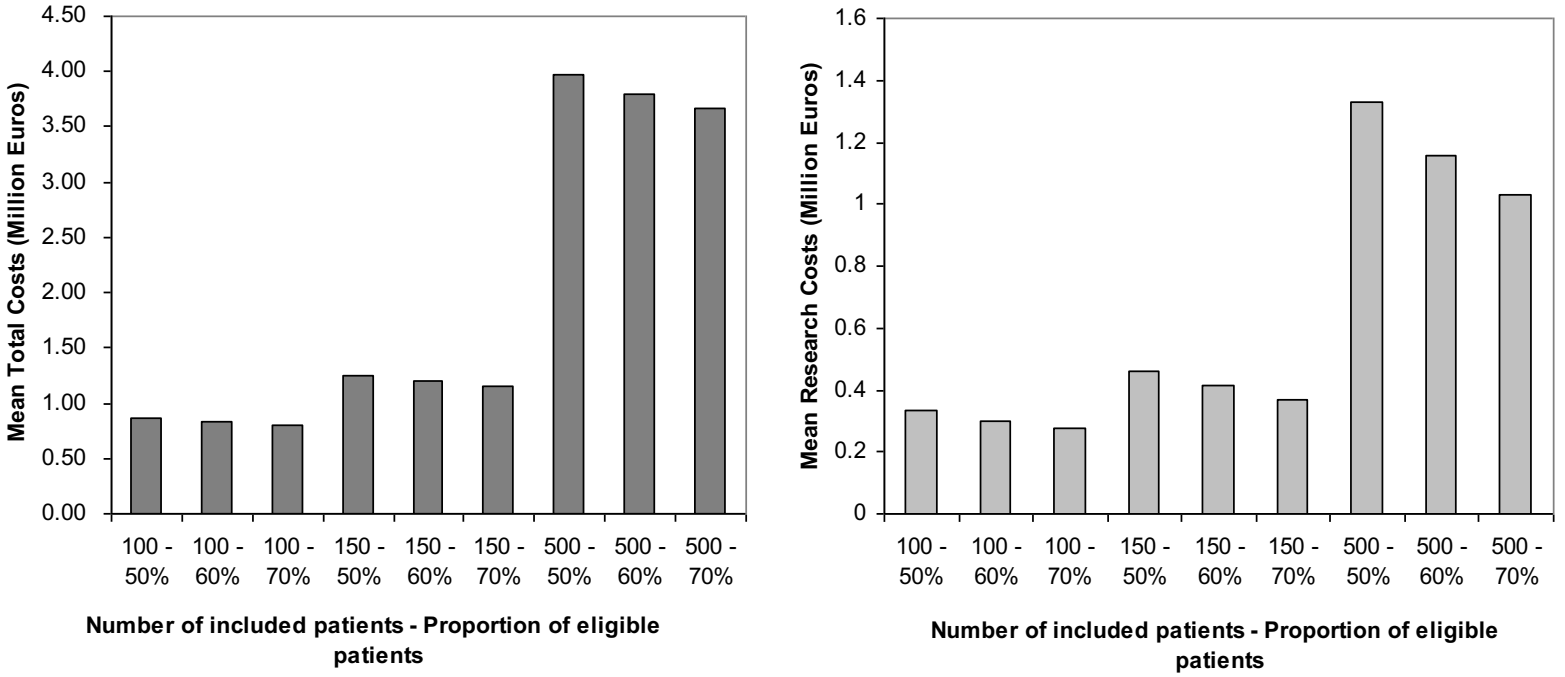


Figure 6 - Mean total costs (left) and mean research costs (right) for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

The total costs of a study set up according to the first randomization strategy, where patients are randomized across the radiological test strategies, can differ greatly depending on the study size. A small trial where 100 patients participate costs about 830,000 Euro of which the costs for research activities are about 300,000 Euro. A trial with 150 patients costs about 1.2 Million Euro, of which about 415,000 Euro for research, and a large study sample of 500 patients would result in total costs of about 3.8 Million Euro including almost 1.2 Million Euro for research costs.

The health care costs are about 5,200 Euro per patient. These costs are not affected by the eligibility of the patients. The research costs, on the contrary, vary under different assumptions on the proportion of eligible patients. If 50% instead of 60%of the patients are assumed eligible and a large trial is performed (500 patients), this results in a total difference of about 170,000 Euro (per patient: ≈

2,700€ instead of ≈ 2,300€). If the rate of eligibility 70%, the difference is about 130,000 Euro (costs per patient about 2000 Euro). For middle-sized (150 patients) and small-sized trials (100 patients) the impact is smaller and results in a difference of about 90,000 Euro (600€ per patient) and 55,000 Euro, respectively, between 50% and 70% eligible patients.
Research costs form an important part of the study costs, on average about one third of the total costs. For small trials where only 100 patients are included, the research costs make up about 36% of the total costs whereas for middle and large trials this proportion is smaller, approximately 34% and 30% respectively. These costs are strongly related to the duration of the trial.

From the following figure, Fig. 7, we can see that the trial duration increases considerably with increasing study size. A small trial takes on average about 85 (standard deviation (SD): 5.58) weeks (ranging between 77 and 94 weeks) whereas the inclusion of 50 more patients takes on average about seven months longer (range: 101 – 130). If a large trial is to be performed, the duration almost triples and it takes approximately 6 years (SD: 17 weeks) over a proportion of eligible patients ranging between seven years for 50% eligibility and almost five and a half years for 70% eligible patients.
Especially for a large study size, there is a big difference in the total study duration. If 60% of the patients participate in the study instead of 50% the trial takes about one year less whereas the study duration is reduced by almost ten months if 70% of the patients are eligible for the study. For trials with 100 patients included, the difference is about 17 weeks. Trials with 150 patients show a fluctuation of about 29 weeks.

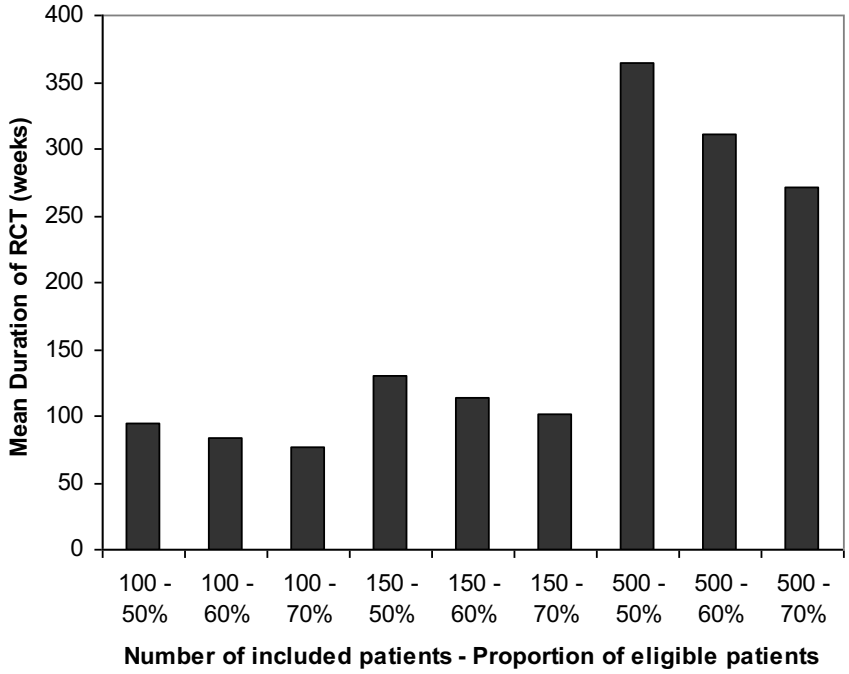


Figure 7 - Mean duration of the trial for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

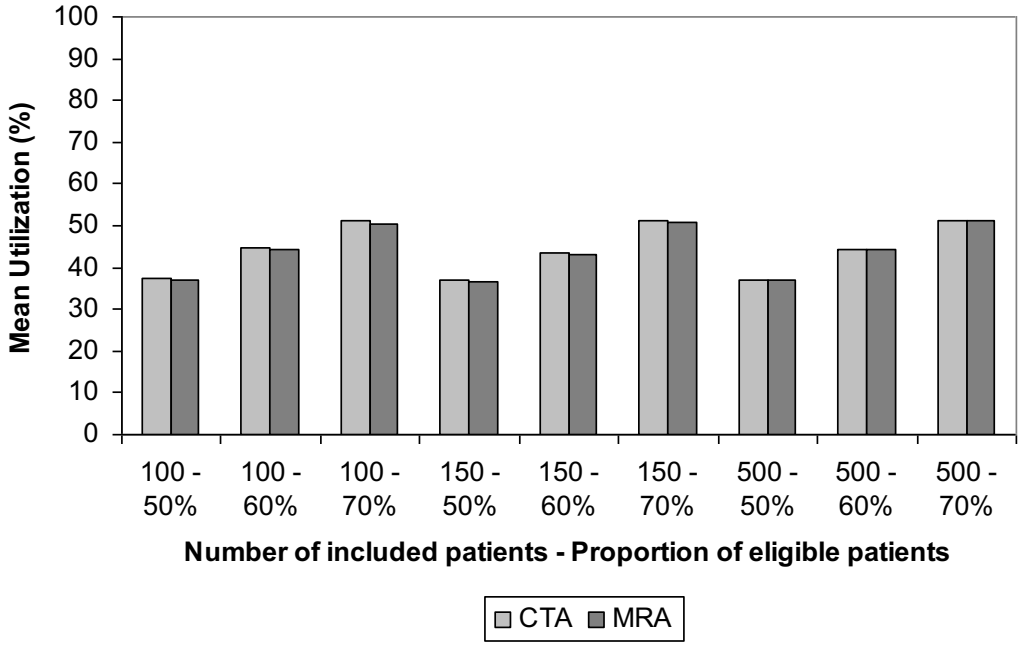


Figure 8 - Mean utilization of the imaging modalities for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

From Figure 8 we can see that the utilization remains almost equal for increasing study size. For the CTA and MRA, this results in an average workload of about 44% and 43% (SD: 3.5 and 3.42) respectively.

As for the varying proportion of eligible patients, an increasing proportion leads to an increased utilization of CTA and MRA. If 50% of the patients are participating in the trial, only about 36% (SD: 2.88) of the capacity is used (almost 7% less than for 60% eligibility). If the eligibility is increased to 70%, the utilization is increased by 6.2% to 51% (SD: 4.04). These numbers remain almost the same for all study sizes considered.

Besides the utilization, the queuing time is also of interest. The mean queuing time of the patients for CTA is about 1.7 workweeks with a standard deviation of about 5 work days including the minimum waiting time of one week. Patients have to wait on average two weeks for the MRA. The queuing times remain almost constant over different study sizes and proportions of eligible patients. CTA queuing times range between 1.8 and 1.6 weeks whereas the time between inclusion and test takes between 2.1 and 2 weeks for the MRA. The proportion of participating patients appears to have little effect on this performance measure. As for the standard deviation this results in a range between 4 and 5 work days for small trials, 3 and 4 work days for middle trials and about 1 and 3 work weeks for large trials.

4.1.2 Results for varying number of diagnostic imaging modalities

In figure 9, the mean research costs of the trial are given for the scenarios described in Table 2, see section 3.4.

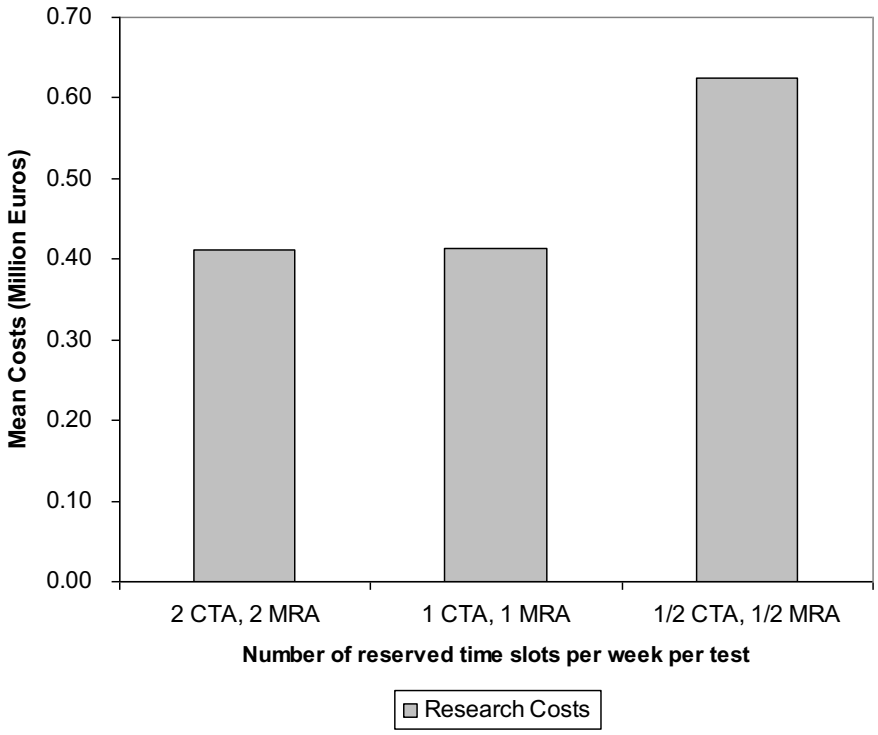


Figure 9 - Mean research costs for varying number of time slots reserved for the trial (150 patients, 60% eligible patients)

As the costs for the health care remain constant for all scenarios, we only show the research costs associated with the RCT. We can see that there is only a slight increase if the time slots reserved for the trial are reduced to one CTA and one MRA per week. The costs differ about 2000 Euro. If, however, one decides to perform only one CTA and MRA in two weeks the costs increase with about 200,000 Euro. This means that the research costs are increased by 150%.

The trend described above can also be seen in the following figure (Fig. 10) in which the mean duration of the RCT is presented for the different allocation scenarios.

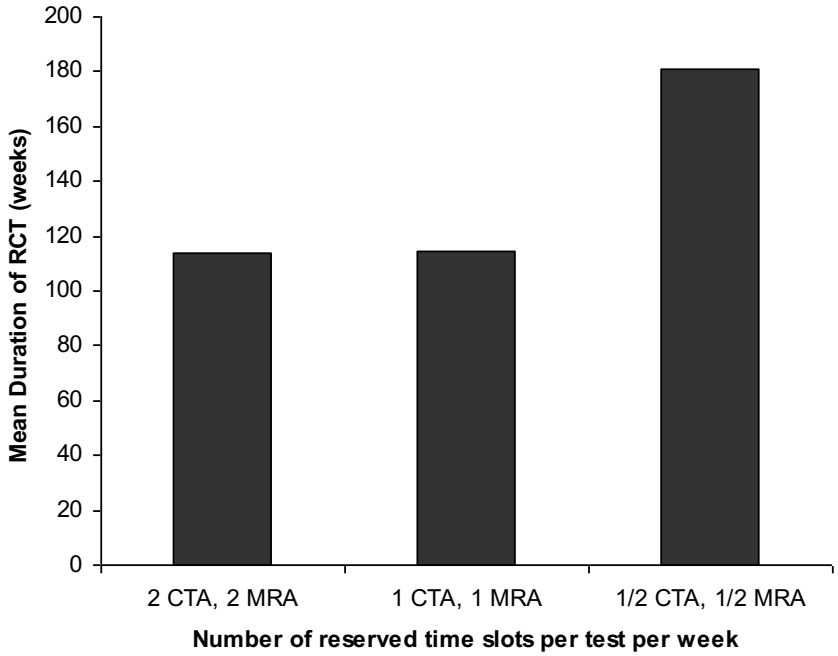


Figure 10 - Mean duration of the trial (150 patients, 60% eligible patients) for varying number of time slots reserved for the trial

As we saw in the previous figure, the trials take almost equally long for the first two scenarios. The trial with one time slot for both CTA and MRA takes about one week longer than the trial performed with two CTA and MRA appointments per week (mean duration 114.2, SD 6.81 weeks). If the imaging capacities are limited to one CTA and MRA in two weeks it takes more than a year longer to perform the trial (mean duration 180.3, SD 1.44 weeks). This holds true if other patients are allowed to be tested during the time slots reserved for the trial.

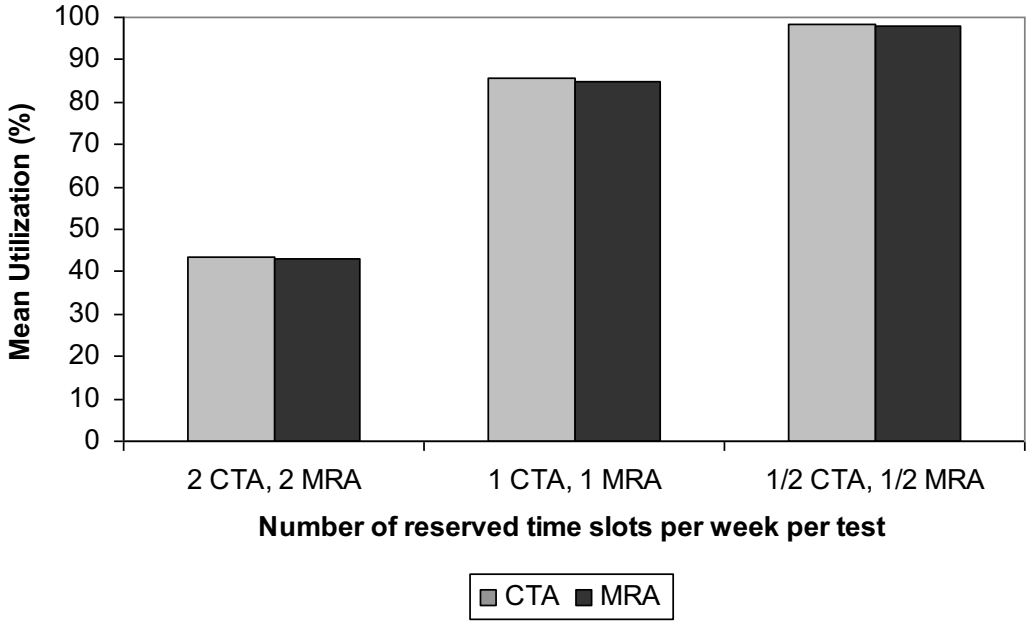


Figure 11 - Mean utilization of the imaging modalities (CTA and MRA) for varying number of time slots reserved for the trial (150 patients, 60% eligible patients)

As we can see in the Figure 11, there is a clear increase of the utilization of the modalities with decreased availability of time slots. The highest utilization of about 98% (SD: 0.63) for both CTA and MRA is obtained by the last allocation scheme. If the time slots for both tests are doubled, the utilization decreases by about 13 percentage points and results in an average utilization of about 85% (SD: 6.04). If two time slots are reserved for CTA and MRA for the participants in the trial, the workload is reduced by a factor of 2.2 and the reserved time slots are only filled in about 43% of the cases.

4.2 Results of the second randomization strategy

In the cost calculations for this randomization strategy, the research costs include the costs for personnel, administration and the material costs for the additional imaging tests.

4.2.1 Results for varying number of included patients and proportion of eligible patients for random sequence of diagnostic imaging tests

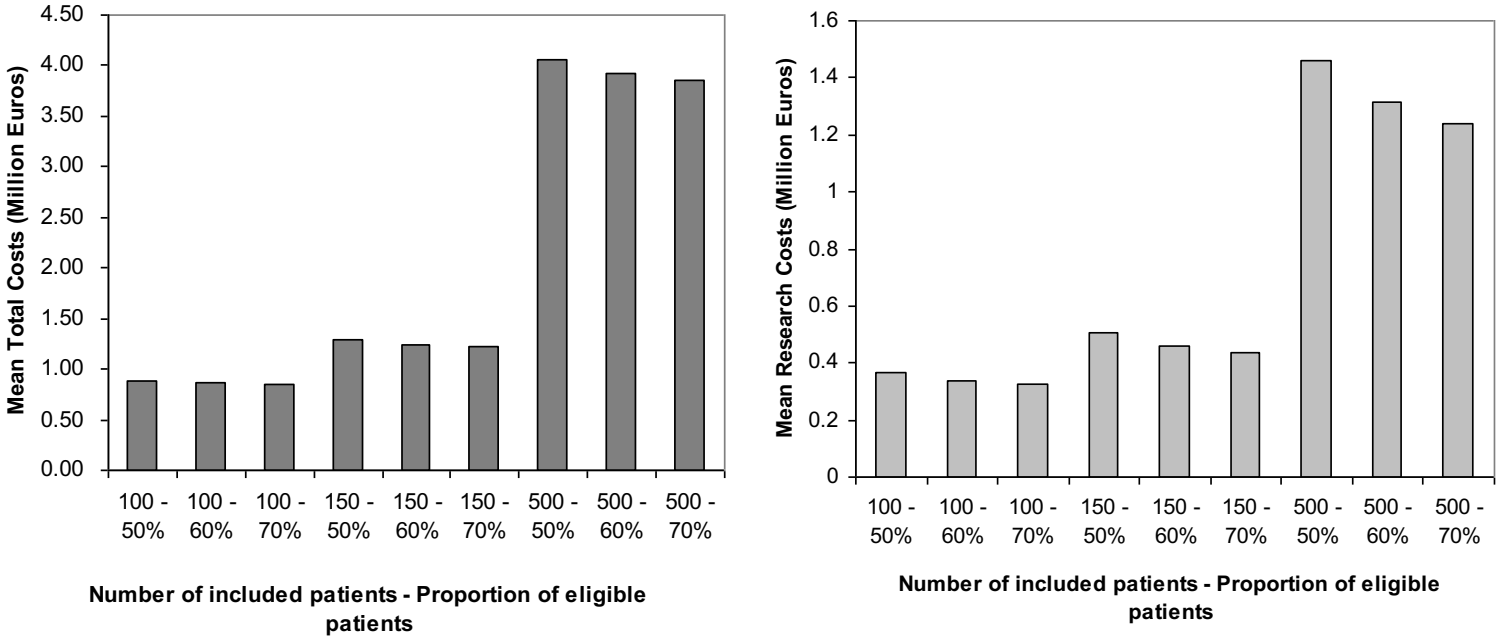


Figure 12 - Mean total costs (left) and mean research costs (right) for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

As in the first strategy, the health care costs remain constant while the research costs tend to increase when the proportion of eligible patients decreases.

The total costs of a study set up according to the second randomization strategy can differ greatly depending on the study size. In this strategy, patients are randomized between providing and not providing information from the diagnostic imaging tests. A small trial where 100 patients participate costs about 860,000 Euro of which the costs for research activities make up about 342,000 Euro. A trial with 150 patients costs about 1.25 Million Euro, of which about 470,000 Euro for research activities. A large study sample (500 patients) would result in total

costs of about 3.9 Million Euro including almost 1.3 Million Euro of research costs.

In addition to this, the research costs vary under different assumptions on the proportion of eligible patients at baseline. If 50% instead of 60% of patients are assumed eligible and a large sample size is performed (500 patients), this results in a difference of about 146,000 Euro (almost 300€ per patient). If the rate of eligibility is 70%, the difference is about 219,000 Euro (almost 450€ per patient). For middle-sized (150 patients) and small-sized trials (100 patients) the impact is smaller and results in a difference of about 111,000 Euro (740€ per patient) and 67,000 Euro respectively.

Also for the second study design, research costs form an important part of the study costs. On average this is more than one third of the total study costs. For small trials where only 100 patients are included, the research costs make up about 40% of the total costs whereas for middle and large trials this proportion is smaller, approximately 37% and 34% respectively. These costs are included as variable costs.

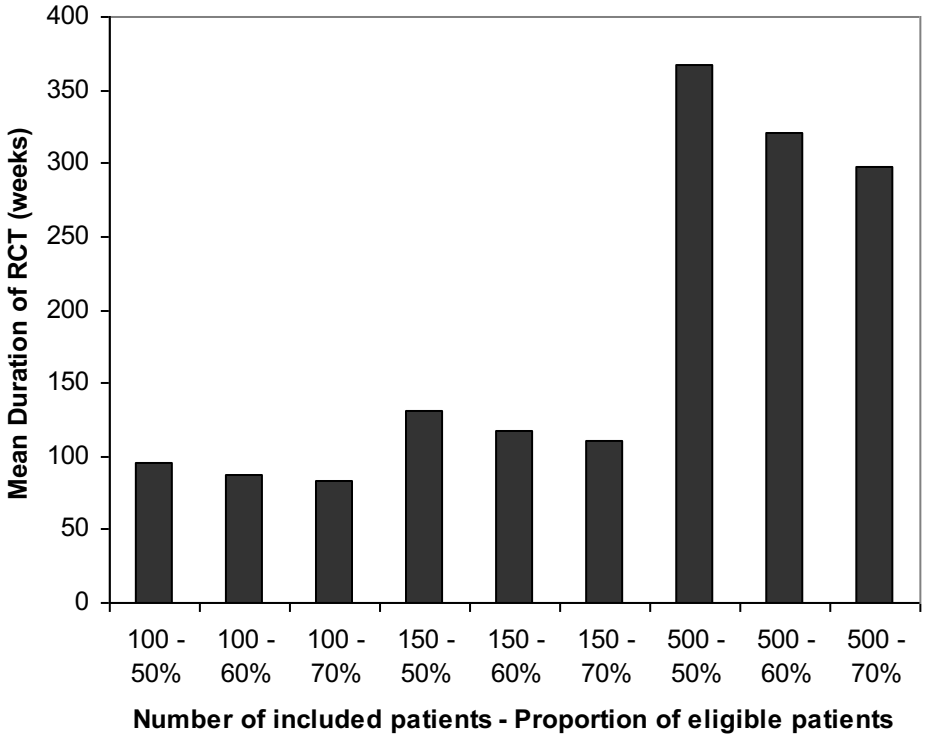


Figure 13 - Mean duration of trial for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

Figure 13 shows the mean duration for combinations of the number of included patients and patient eligibility. Similar to the results for the first strategy, we can see that the trial duration decreases with decreasing study size. A small trial takes on average about 89 weeks with a standard deviation of four weeks (mean duration ranging between 84 and 96 weeks) whereas the inclusion of 50 more patients takes on average more than seven and a half months longer (SD: 5.6 weeks, range of mean duration: 2 – 2.5 years). If a large trial is to be performed at a single center, it takes approximately six and a half years (SD: 3.5 months) over a proportion of eligible patients ranging between more than seven years for 50% eligibility and more than five and a half years for 70% eligible patients.

Especially for a large study size, there is a big difference in the total study duration. If 60% of the patients participate in the study instead of 50% the trial

takes about one year less whereas the study duration is reduced by more than six months if 70% of the patients are eligible for the study. This effect is also to be seen for middle- and small-sized trials, though less significant.

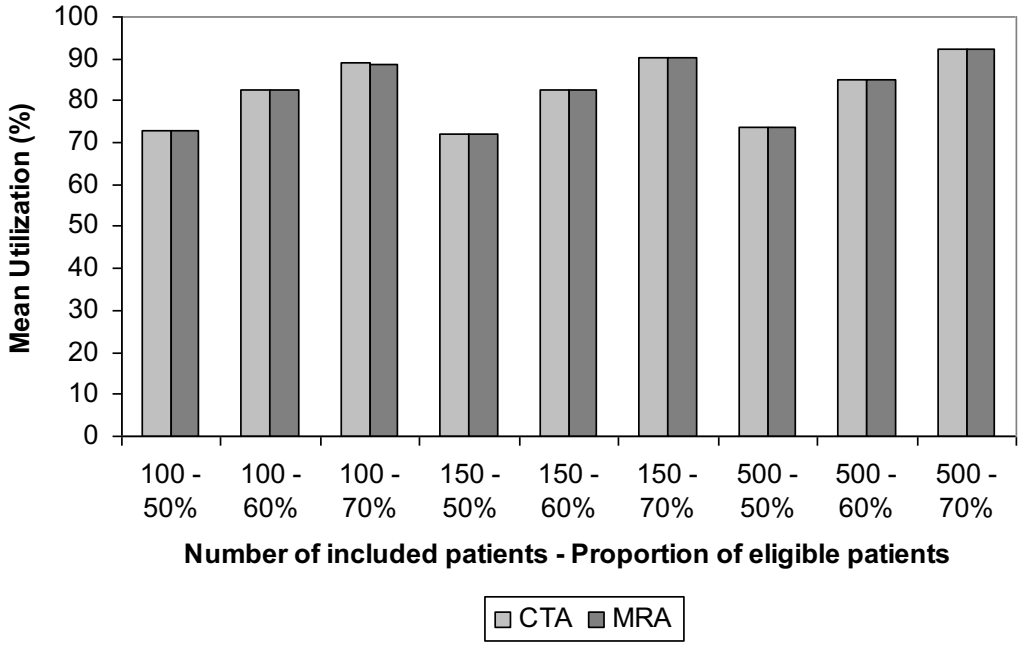


Figure 14 - Mean utilization of the imaging modalities for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

From Figure 14 we can see that the utilization remains almost equal for increasing study size. For both CTA and MRA, this results in an average workload of about 83% (SD: 4.4%).

As for the varying proportion of eligible patients, an increasing proportion leads to an increased utilization of CTA and MRA. If 50% of the patients are participating in the trial, only about 73% (SD: 5.35%) of the capacity is used (11.3 percentage points less than for 60% eligibility). If the eligibility is increased to 70%, the utilization is increased to almost 92% (SD: 2.9%). These results remain almost equal for all study sizes considered.

Besides the utilization, the queuing time of the patients can be of interest. The mean queuing time of the patients for CTA and MRA is about 3 work weeks with

a standard deviation of about 6 work days including the minimum waiting time of one week. The queuing times remain rather constant over different study sizes and proportions of eligible patients. CTA queuing times range between 2.1 and 3.8 weeks whereas the time between inclusion and test takes between 2.2 and 3.6 weeks for the MRA. The proportion of participating patients appears to have little effect on this performance measure.

If the trial is performed without prioritizing the patients who already underwent one test, the results do not show much change. The trial would take on average one week less which results in a difference in the research costs of about 3,000 Euro. As for the utilization of the equipment, the time slots for CTA and MRA would be filled in 84% of the cases in comparison to 82% if priority queuing is applied.

4.2.2 Results for varying number of diagnostic imaging modalities

For this randomization strategy, we only considered the scenario when the imaging capacity is reduced to one time slots for both CTA and MRA per week. The mean duration is increased by 71 weeks to more than three and a half year for the 1CTA, 1 MRA scenario (SD: 0.62 months). This increase is equivalent to almost 225,000 Euro of research costs. As for the utilization of the imaging equipment, the time slots are filled in more than 97% of the cases (SD: 1.52%) which is an increase of about 15 percentage points.

If the number of time slots is reduced by factor 2, the trial duration is increased to more than 450 weeks (more than eight and a half years) which is certainly not desirable for a single-center trial. The proportion of patients waiting less than eight weeks drops to 9%, so it seemed irrelevant to consider this study setting.

4.2.3 Results for different sequences of diagnostic imaging tests

We can see from Fig. 15 and 16, that there is only a slight improvement of the duration and the utilization if patients are assigned alternately to the imaging tests, always to the CTA first or to the tests with the shortest queue instead of a random sequence. In place of 118 weeks, it takes about 116 weeks (SD: 6.25) if CTA is performed first, alternate and shortest queue routing takes slightly longer (about half a week). Also the mean utilization is slightly increased if another routing rather than random sequencing of the tests is used. If the patients are assigned alternately to CTA and MRA or to the shortest queue, the utilization of the imaging modalities is about 84% (SD: 5.5%). This is an average increase of about two percentage points compared to the random sequence. If the sequence is fixed, the utilization that can be achieved is about 0.6 percentage points higher.

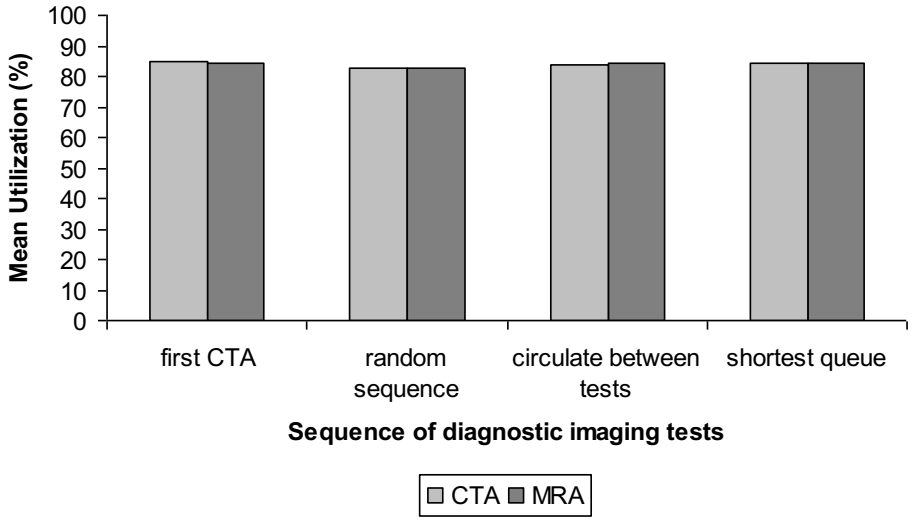


Fig. 15 - Mean utilization of the imaging modalities for different sequences of diagnostic imaging tests

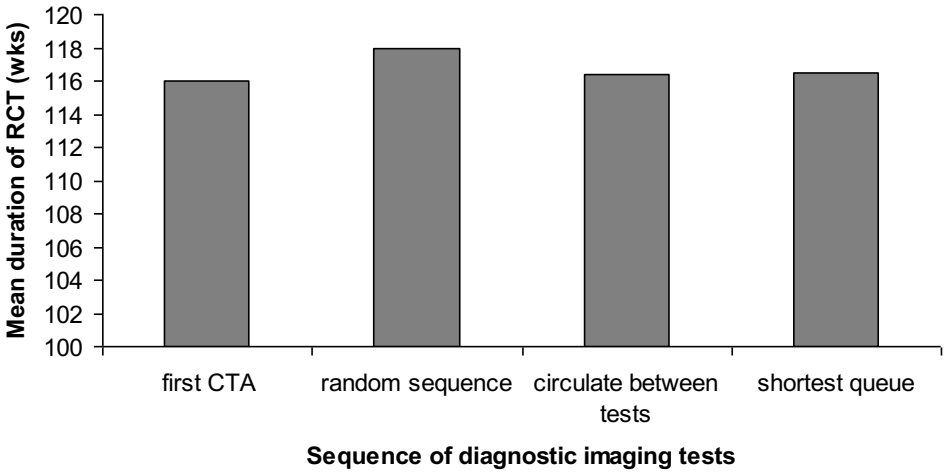


Fig. 16 - Mean duration of the trial for different sequences of diagnostic imaging tests

In relation to the total utilization of the equipment and the total duration of the trial, however, the improvement that can be achieved by changing the sequence of the first diagnostic imaging test is only very little.

4.3 Comparing randomization strategies

In order to compare the different randomization strategies for RCTs, we will focus on the outcome measures research costs, cost for unfilled time slots and duration. The costs for unfilled time slots are based on the costs for housing, equipment and personnel. The costs related to health care are equal for both strategies (see sections 4.1 and 4.2), so they will not be included in the comparison.

4.3.1 Results for varying number of included patients and proportion of eligible patients for random sequence of diagnostic imaging tests

In the following figure, Fig. 17, the mean durations of RCTs are given for both randomization strategies (Rand. Strategy 1 and Rand. Strategy 2).

In general, we can conclude that the strategy where all patients undergo both tests results in a longer duration in comparison to the strategy where patients are only tested once. The difference in the mean duration is influenced by the study size. For a large study size (500 patients), the second randomization strategy takes on average 13 weeks longer than strategy 1. Small and middle-sized trials would take on average four and five weeks longer if CTA and MRA are performed on all patients instead of one of the tests.

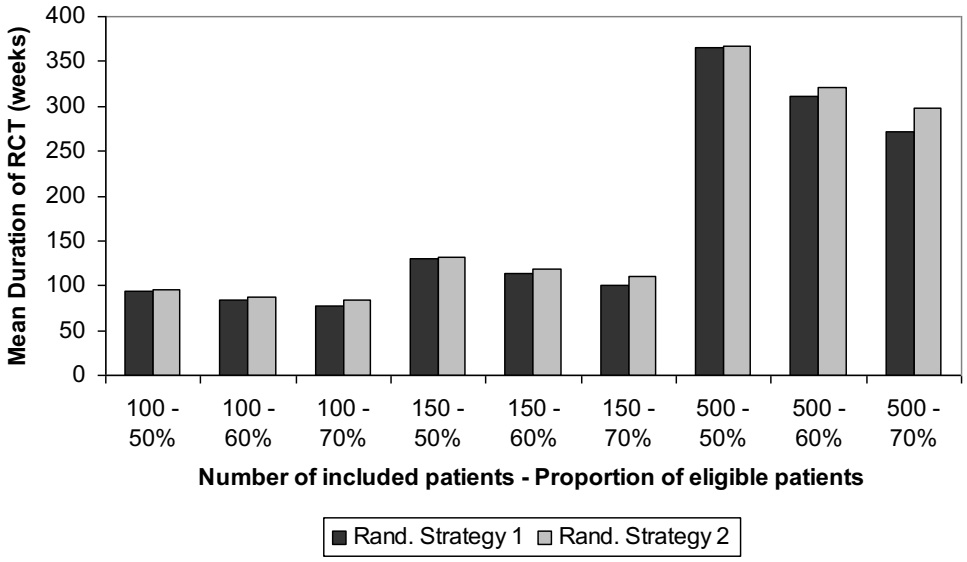


Figure 17 - Mean duration of RCTs set up according to strategy 1 (dark grey) and 2 (light grey) for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

The difference in the mean duration increases for increasing proportion of eligible patients. For 50% of participating patients the difference is over all study sized relatively small. Performing both tests on all patients would take on average two weeks longer than only one test. If the eligibility is 10% higher, the trial set up according to strategy 2 would take about six weeks longer than for strategy 1. For 70% participating patients the difference is almost four months.

The mean difference of the duration of a small-sized trial ranges between two and seven weeks. The fluctuation for middle-sized trials is about the same, the difference in the mean duration is between two and nine weeks. For large-sized trials it ranges between two and seven months.

The difference in the mean duration can be seen again in the mean research costs given in Fig. 18. As we saw in the previous figure, implementing a trial according to the second randomization strategy would lead to considerably higher costs than for strategy 1. This holds true over all study sizes and proportions of eligibility. For a large study size (500 patients), the research activities for a RCT set up according to the second randomization strategy costs on average about 165,000€ more than a trial after strategy 1. Small and middle-sized trials would cost on average 39,000€ and 53,000€ more if CTA and MRA are performed on all patients instead of only one of the tests.
The difference in the mean research costs increases for increasing proportion of eligible patients. If an eligibility of 60% is reached instead of 50%, the trial set up according to strategy 2 would cost about 80,000€ more than for strategy 1. For 70% participating patients the difference is more than 100,000€.

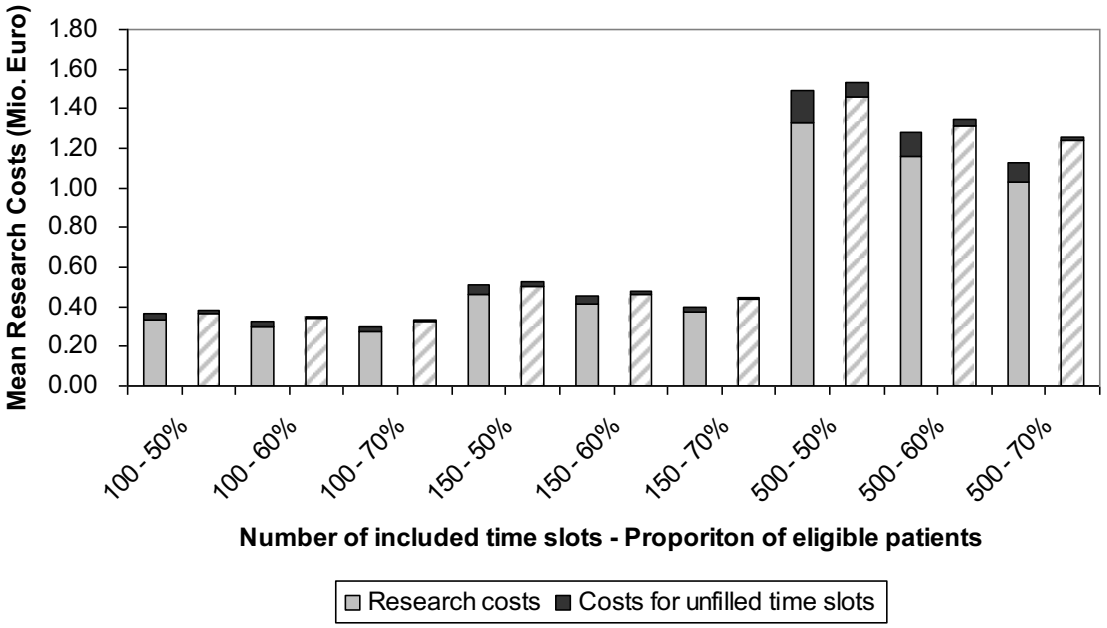


Figure 18 - Mean research costs and costs for unfilled time slots for both randomization strategies (strategy 1 – grey filled bars and strategy 2 – grey striped bars) for combinations of the number of included patients (100,150,500) and the proportion of eligible patients (50%, 60%, 70%)

The effect of the difference in the mean duration is alleviated when the costs for unfilled time slots are added to the research costs. We can conclude that the costs for unfilled time slots are higher for the first randomization strategy than for the second. This is due to the higher utilization of the imaging equipment for strategy 2, see section 4.1 and 4.2. The costs for unused capacity increase with increasing study size. For a small study size, overcapacity for strategy 2 results in costs of about 10,000€ (ranging between 5,000€ and 15,000€). This is about 17,000€ less than for strategy 1. Performing a middle-sized RCT according to strategy 2 where unused capacity is charged would lead on average to 14,000€ of additional research costs (ranging between 6,500€ and 23,000€). A RCT after strategy 1 would result in 26,000€ higher research cost. A large RCT with 500 patients included would result in about 40,000€ of costs for unfilled time slots, almost 90,000€ less than for strategy 1. As the utilization of the imaging

equipment increases with increasing patient eligibility for both randomization strategies (see sections 4.1 and 4.2), the costs for unused capacity decrease with increasing proportion of participating patients.

Including the costs for unused capacity, the research costs for strategy 1 are about 330,000€ for a small-sized RCT, 455,000€ for a middle-sized trial and a large trial would result in about 1.3 Million € for research activities. This is a difference of about 22,000€, 27,000€ and 78,000€ for the respective study sizes in comparison to the second randomization strategy. The total difference in the research costs increases with increasing participation of the patients. For a large study size, a high participation of the patients (70% baseline eligibility instead of 50%) results in a difference of about 100,000€ between the strategies. This effect is smaller for smaller study size and lower baseline eligibility.

4.3.2 Results for varying number of diagnostic imaging modalities

In this section we will only compare the results for two reserved time slots per week for the initial imaging test and for one time slot per week. This is due to the fact that the second randomization strategy would lead to excessive trial duration if only one time slot in two weeks was reserved for the trial.

From the following figure, Fig. 19, we can see that the number of the reserved time slots has a bigger influence on the results of strategy 2 than of strategy 1.

On average, a trial set up after randomization strategy 1 takes about two years if two time slots are reserved for the participants of the trial. The second randomization strategy would prolong the mean duration by about a month. If, however, the time slots are reduced to one time slot per week, a trial after the second strategy takes more than three and a half year, almost one and a half year longer than for strategy 1.

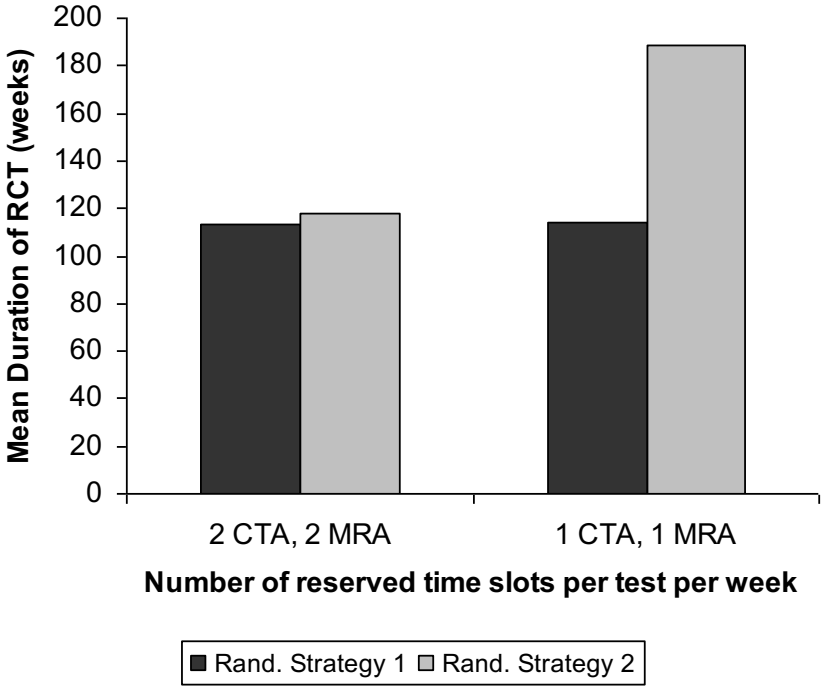


Figure 19 - Mean duration for both randomization strategies (150 patients, 60% eligible patients) for varying number of time slots reserved for the trial

The impact of the number of time slots is also reflected in the research costs, given in the following figure. Exclusive of the costs for overcapacity, the research costs differ about 50,000€ if two time slots are reserved per week. Reducing the time slots to one per week per test leads to a difference in research costs of about 270,000€.

Including the costs for overcapacity, the difference in the research costs is almost reduced by a factor of 2 (a RCT after strategy 1 costs about 24,000€ less than a RCT after strategy 2). The difference for one time slot per week per test remains rather constant with about 268,000€ higher costs for strategy 2.

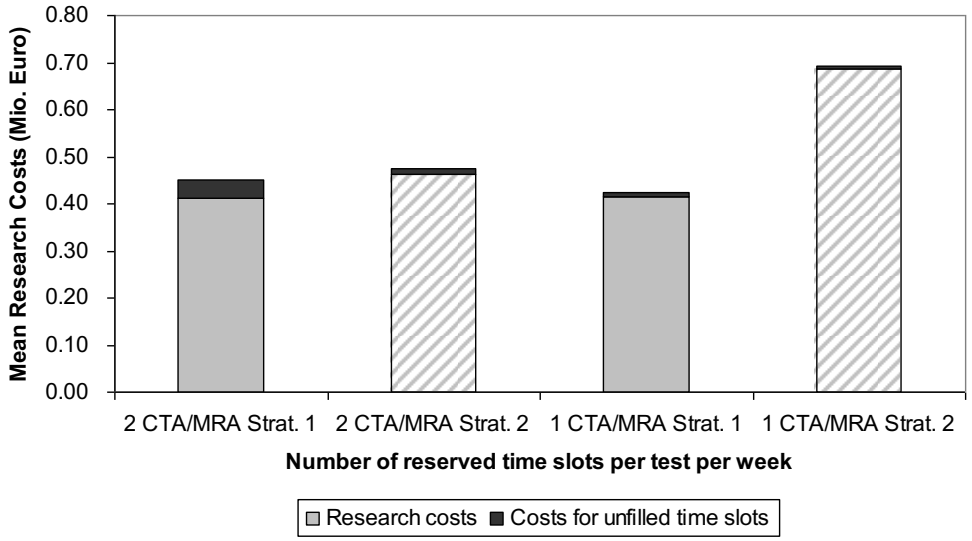


Fig. 20 - Mean research costs and costs for unfilled time slots for both randomization strategies (150 patients, 60% eligible patients) for varying number of time slots reserved for the trial

Due to the high utilization of the equipment achieved by both randomization strategies, see sections 4.1 and 4.2, only few time slots remain unfilled and therefore the costs for unused capacity are relatively low. For the first strategy unfilled time slots cost almost 40,000€ if two time slots are reserved per week. If one time slot is reserved the costs are reduced to about 10,000€. The second strategy causes less idle time with costs of about 13,000€ for unused capacity for two reserved time slots and only 4,000€ if one time slot is reserved per week per test.

5 Discussion

The implementation of a randomized controlled trial can be modeled as a stochastic queuing system in which patients arrive with a specific disease and their participation in the trial is determined by the disease state of the patient and the willingness to cooperate of the physicians and the patients. Long waiting times for diagnostic work-up have a negative effect on the proportion of participating patients. We considered two options of study designs. The first randomization strategy is designed such that patients are randomized between the two imaging strategies under study. Applying the second design, all patients undergo both diagnostic imaging tests and are subsequently randomized between disclosing one of the two test results. We built two clinically based discrete event simulation models to compare the performance of the two strategies.

Conceptually, when setting up a trial one should consider the fact that the utilization of the diagnostic imaging capacity would be (close to) optimal if patients were added to the regular waiting list instead of reserving time slots for the patients participating in the trial. By giving the participants priority over the non-participating patients, the waiting time for the initial diagnostic work-up could be limited within a certain range acceptable for patients and physicians. Besides, the impact of fluctuations in the arrival process of the patients could be regulated. Periods during which few patients are included in the study would not lead to a suboptimal utilization of the imaging equipment and the overall waiting time of patients could be reduced.

From a mathematical point of view, the two randomization strategies differ greatly in the queuing process for the initial diagnostic imaging test. The first randomization strategy can be seen as two parallel queues with one common arrival process. The second strategy could be modeled as a tandem queue

where two queues are scheduled after each other. The arrival process of the first queue consists of the arrival process of the patients entering the trial and the patients who already underwent the other test; the second queue receives patients in the same way. In the scenario where patients are assigned alternately to the two tests when entering the trial the queuing system is kept in a stable condition. In equilibrium, about two patients enter the queue of each test every week, one after a test is performed and the other one arrives as a 'new' patient in the trial. Then the two time slots reserved for the trial would be used in an efficient way and therefore the duration of the trial would be minimized. In this equilibrium situation, alternate routing is equivalent to assigning the patients to the shortest queue, both queues have the same length. If the routing of the patients is done randomly, this equilibrium can not be reached and the performance of the study design would deteriorate. However, routing patients alternately to the two tests could introduce a selection bias to the study. When performing a trial according to this study setting, the recruiting physician would know what the first test would be and could therefore manipulate the randomization. On the other hand, the disclosure of the test results is still randomized so the effect of this bias should/will be limited. Furthermore, if the sequence of the tests is fixed, for example CTA is performed prior to MRA, a bias could be introduced by the fact that information from the first test may be available before the second test is interpreted. Then this information could influence the interpretation of the MRA so the results would not be independent of each other. Therefore, it is not advisable to set up a study with a fixed sequence although it would lead to a slightly shorter duration and a higher utilization.

Apart from the bias mentioned above, the choice of the study design can induce other threats to the validity of the study.

In section 2.1.1.2 the second randomization strategy is described in detail. One of the disadvantages mentioned is that physicians are more likely to request the result of the undisclosed test as both tests have been performed. For some

diagnostic imaging tests and diseases, it could be possible that test A produces "better" images than test B. Then the physicians to whom test B is disclosed are more likely to request the other test result than the physicians to whom test result B is disclosed in the first place. Accordingly, the treatment of the patient based on test B would be dependent on the result of test A and the comparison of the two patient groups would be biased. Also, composition of the group of physicians may influence the outcomes. If a trial according to this study design is performed at a hospital where the general attitude is favoring one of the tests under consideration, the physicians might request more additional test results of one test and therefore bias the interpretation of the test results and the treatment decision. As result of these two problems, the number of additionally requested test results should be included in outcome measures of the RCT set up according to the second randomization strategy.
The attitude of the hospital might also pose a problem for the first strategy. In this case, the outcomes of the trial might not reflect an objective assessment of the diagnostic imaging techniques.

The question may arise why we decided to simulate the implementation of different designs of randomized controlled trials rather than calculating solutions using queuing mathematical techniques. Our aim was to model a RCT under realistic environmental assumptions. In our case, the realistic external factors included the effect of the waiting time on the eligibility of the patients and the maximum time patients were willing to wait for their diagnostic imaging work-up. We assumed that a long waiting time had a negative effect on the cooperativeness of the treating physicians to include their patients in the study. If patients had to wait longer than eight weeks for the diagnostic imaging test(s), they were assumed to withdraw from the study. To the current knowledge of the author, there are no queuing models that reflect these properties of the model, so we decided against an analytical approach to this problem. On the basis of the availability of simulation software, we decided to perform a simulation study using the Simul8 software platform.

Our simulation results for the first randomization strategy (baseline scenario with 150 patients and 60% eligibility) are close to the findings from a recently completed RCT on PAD performed by R. Ouwendijk from December 2001 until March 2004. See section 3.1 for a description of the DIPAD trial.

The mean duration of our simulated trials is about three weeks less than the DIPAD trial. This difference can be explained by the fact that there were some logistical problems at the beginning of the DIPAD trial. No time slots were reserved for the imaging tests of the patients participating in the trial and the patients were scheduled during the regular outpatient hours on the radiology department. Therefore relatively long waiting times for the patients occurred at the beginning of the trial which prolonged the inclusion period and therefore the total duration of the RCT. This also explains why the utilization of the imaging capacities is slightly lower in the DIPAD trial than in our simulation results. The difference, however, with about two percentage points is rather small.

Also the mean health care costs of our simulation results agree with the findings from the DIPAD trial. The mean costs for health care in our model are about 790,000€, about 30,000€ less than the actual DIPAD trial. This can be explained by the fact that we did not include other personnel than the two technologists for the CTA and MRA and one radiologist for the DSA and the evaluation of the test results of CTA and MRA. In the actual implementation of the DIPAD trial, the allocation of personnel to the tests fluctuated. Sometimes a radiologist, a third technologist or an apprentice were present during CTA and MRA. The additional personnel costs were included in the cost calculations for the DIPAD trial but could not be included in the simulation model developed for this study. Based on the available data from the DIPAD trial, the distribution of the test durations, CTA, MRA and DSA, were simplified to constant duration variables which also contributes to the difference in the health care costs.

Further comparison with results in the literature was difficult because the chosen performance measures are usually not evaluated or described in publications on RCTs.

Our study is a synthesis of the medical and methodological knowledge of the ART group at the Department of Epidemiology and Biostatistics at the Erasmus Medical Center Rotterdam and the mathematical knowledge of the Vrije Universiteit in Amsterdam. We recognize that the current version of the model is based on several simplifying assumptions which were necessary in order to keep the problem manageable for the model builder within the given time horizon of this project. The following limitations could serve as interesting aspects for further research.

The main limitation of this study is that the information obtained by trial is not included in our evaluation. When performing the second randomization design one obtains additional information by performing all tests on all patients. From a cost perspective, the second randomization strategy will never be optimal, but there may be situations in which the additional information could outweigh the additional costs. A possible approach would be to use value of information theory and combine a decision analytic model with the discrete event model presented in this study. The decision analytic model could then be used for the calculation of the expected value of sample information and in combination with the costs of the trial the net benefit of performing a trial can be determined [7].

Another limitation of the project is that this study focused on two study designs for medical trials. When setting up a clinical trial, there are several possibilities to obtain evidence on the medical decision problem. Study designs that could be considered include a cohort-study in combination with a medical decision model, a cohort study which evaluates the impact on the clinical practice or other designs for randomized controlled trials [1, 2]. A third design for a RCT is described in section 2.1.1.3 where all patients undergo all diagnostic tests and only patients with discordant test results are randomized and followed.

Furthermore, the estimates for the input parameters of the simulation model are based on only one completed diagnostic trial. Therefore, the results may not be representative for other research institutions in the area of the assessment of radiological technology.

Another limitation of our study is that we did not include the time costs, transportation costs and lost productivity of the participating patients in our cost calculations. The time costs reflect the time required for transportation, for waiting in the clinician's office and for undergoing the diagnostic test and treatment, etc. The majority of the patient group of the DIPAD trial was older than 65, so it seemed reasonable to assume that these costs do not change the results significantly. For other areas of application, for instance diseases in younger patients, this may be an important issue.

A further limitation may be that we did not include late arriving patients and no-show patients. In the DIPAD trial of R. Ouwendijk, patients came on time for the appointment so this issue was not relevant for the successful implementation of the trial. In the reader's opinion, this is mostly due to the elderly patient group and the good organization of the trial. In other patient groups, however, late arrivals and no shows could pose a problem.

We recognize that the assignment of the patients to the different treatment options was done rather roughly, see section 3.4 for a description. In a further elaborated version of the simulation model, specific information about the patient could be used like age, severity of the disease, etc. This can be done using several patient labels in Simul8 or by combining the discrete event simulation model with a decision tree. Our simulation model did not take into account waiting times for treatment or other logistical problems related to the medical therapy of the patients. We assumed that the patients participating in a clinical trial were treated during the follow-up period of six months. For some fields of application this may be a relevant factor to be included in the study.

Furthermore, we did not include the possibility that other patients could fill empty time slots reserved for the participants of the trial. A realistic representation would include the arrival process of the other patients and the period of "availability". We considered this issue on a preliminary stage, but due to little information we decided to restrict our analyses on the scenarios without other patients.

Another limitation of our study may be that the model did not include utilities. As the main focus laid on the costs and logistics related to a RCT, we modeled mainly the part of the trial where diagnostic imaging is performed. According to the calculations of quality of life [1], a few minutes of discomfort during the imaging test do not have much influence on the quality of life of the patient if she lives the following three years in good health after a successful treatment based on the results from the diagnostic imaging test(s).

Finally, the information needed to fill the model could pose a limitation to the practical use of the model for the planning of studies. At the stage of planning a trial, information on the duration of the test, scoring and dictating the test results, waiting times according to the logistics of the radiology department and personnel costs may not be available. Practical experience and information from previous trials could, however, be used as estimates, for instance the test duration and the post-processing period for a CTA.

6 Conclusions

In our simulation study, we analyzed quantitative aspects of the planning of a medical trial. The results suggest that the research costs form about one third of the total study costs. Therefore, the duration of the trial is an important aspect to be considered when deciding about the design for a diagnostic study. The mean research costs for both randomization strategies increased with increasing study size and decreasing proportion of eligible patients. The strategy where all patients undergo both diagnostic imaging tests appeared to be more expensive in all scenarios considered but also showed a higher utilization of the imaging modalities. For a baseline eligibility of 50% the difference in the mean duration of the two designs is only two weeks, increasing to almost 4 months for a high proportion of eligible patients. The difference of the research costs can be interpreted as the costs of the additional information when performing all tests on all patients. Adding the costs for unused diagnostic capacity to the costs related to research activities, decreased the differences in research costs but did not change the results significantly. Reducing the weekly number of reserved time slots had a stronger impact on the performance of the second randomization strategy than the first one. Changing the sequence of the tests for strategy two appeared not to influence the results tremendously.

The results in the sensitivity analysis emphasize that the proportion of eligible patients should be a crucial factor when setting up and performing a randomized controlled trial. Encountering a smaller proportion of eligible patients at baseline than anticipated can have a considerable impact on the duration, the costs and the utilization of the imaging devices. For the first randomization strategy, the difference per patient in the research costs is almost 600€. The effect of a smaller eligibility was smaller for the second randomization strategy. Here, the

mean research costs per patient showed a range of about 400€ between 50% and 70% eligibility at baseline.

Every solution breeds new problems.

Murphy's Law #13

Acknowledgements

Several people contributed to this study and I would like to mention them in this place.

First of all, I would like to thank my supervisor at the Erasmus Medical Center in Rotterdam, Prof. Hunink, for her support and encouragement. I very much enjoyed working in her research group and learned a lot from her helpful comments and remarks with respect to medical and non-medical decision making.

I would also like to thank my supervisors from the Vrije Universiteit, Dr. Franx and Dr. Jongbloed, for making the contact with Prof. Hunink and our discussions concerning my final project.

My colleagues from the ART group made an important contribution to this study and I would like to point out Rody, Joke, Majanka, Ylian and Bas. Thank you for your patience with me and my medical questions, your support, the tea and the great time.

Finally I would like to mention Christian, thank you for everything.

References

1. Hunink, M.G. and G.P. P., *Decision Making in Health and Medicine: Integrating Evidence and Values*. 2001, Cambridge, UK: Cambridge University Press.
2. Hunink, M.G. and G.P. Krestin, *Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology.* Radiology, 2002. **222**(3): p. 604-14.
3. Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, *Randomised comparisons of medical tests: sometimes invalid, not always efficient.* Lancet, 2000. **356**(9244): p. 1844-7.
4. Law, A.M. and W.D. Kelton, *Simulation Modeling and Analysis*. 3rd ed. 2000: McGraw-Hill.
5. Hauge, J.W. and K.N. Paige, *Learning SIMUL8 The Complete Guide*. 2001, Bellingham, WA: PlainVu Publishers.
6. Mires, G., F. Williams, and P. Howie, *Randomised controlled trial of cardiotocography versus Doppler auscultation of fetal heart at admission in labour in low risk obstetric population.* Bmj, 2001. **322**(7300): p. 1457-60; discussion 1460-2.
7. Ades, A.E., G. Lu, and K. Claxton, *Expected value of sample information calculations in medical decision modeling.* Med Decis Making, 2004. **24**(2): p. 207-27.

Appendices

Appendix A - Costs included in DES models

The following costs were included in the discrete event simulation models for the first and second randomization strategy described in sections 3.3.1 and 3.3.2. Most of these costs were assessed by a cost analysis at the ErasmusMC Rotterdam. For the other costs national estimates were used according to the Dutch guidelines for cost calculations. The costs are slit up into health care costs and research costs. The health care costs include costs related to the procedures and immediate health care. Research costs include costs for research staff from the departments concerned, costs for additional diagnostic imaging tests (relevant for the second study design) and administration costs for data collection during the trial.

1. Health care costs

- Personnel

Staff	***Costs (per minute)***
Radiologist	0.94€
Laboratory assistant	0.37€
Specialized laboratory assistant	0.42€

Table 4 - Personnel costs

- Costs per procedure

Procedure	***Type of costs***	***Costs***
MRA	*Equipment (per minute)*	5.29 €
	Materials (unit costs)	103.83 €
	Housing and housekeeping (per minute)	0.21 €
CTA	*Equipment (per minute)*	2.42 €
	Materials (unit costs)	52.47 €
	Housing and housekeeping (per minute)	0.24 €
DSA	*Equipment (per minute)*	2.47 €
	Materials (unit costs)	651.90 €
	Housing and housekeeping (per minute)	0.28 €
PTA	*Average total costs*	2,869.99 €
Surgery	*Average total costs Bypass*	4,848.62 €
	Average total costs Aorta-bifurcation prosthesis	15,501.83 €
	Average total costs Amputation	11,271.56 €
Conservative therapy	*Average total costs*	0 €

Follow-up	*Per outpatient visit (unit costs)*	70 €
	Maximal walking distance measurement (unit costs)	33 €
	Brachial-Arm Index measurement (unit costs)	61 €

Table 5 - Costs of procedures

- Overhead – 15% of personnel, material and equipment costs

2. Research costs of trial

- Costs for personnel: 165,000 Euro per year

Staff	***Full-time equivalent (fte)***
Ph.D. student	1
Post-Doc	0.4
Staff member Radiology department	0.1
Staff member Epidemiology department	0.05
Consultation Biostatistics	0.05

Table 6 - Costs of research staff

- Material costs per patient:
  Health care costs for additional diagnostic imaging test(s)

- Administration costs:
  50 Euro for gathering data at baseline
  50 Euro per test (incl. data collection)
  50 Euro per data collection in follow-up (after 12 and after 24 weeks)

Appendix B – Screenshots of DES models

1. Randomization across diagnostic testing strategies

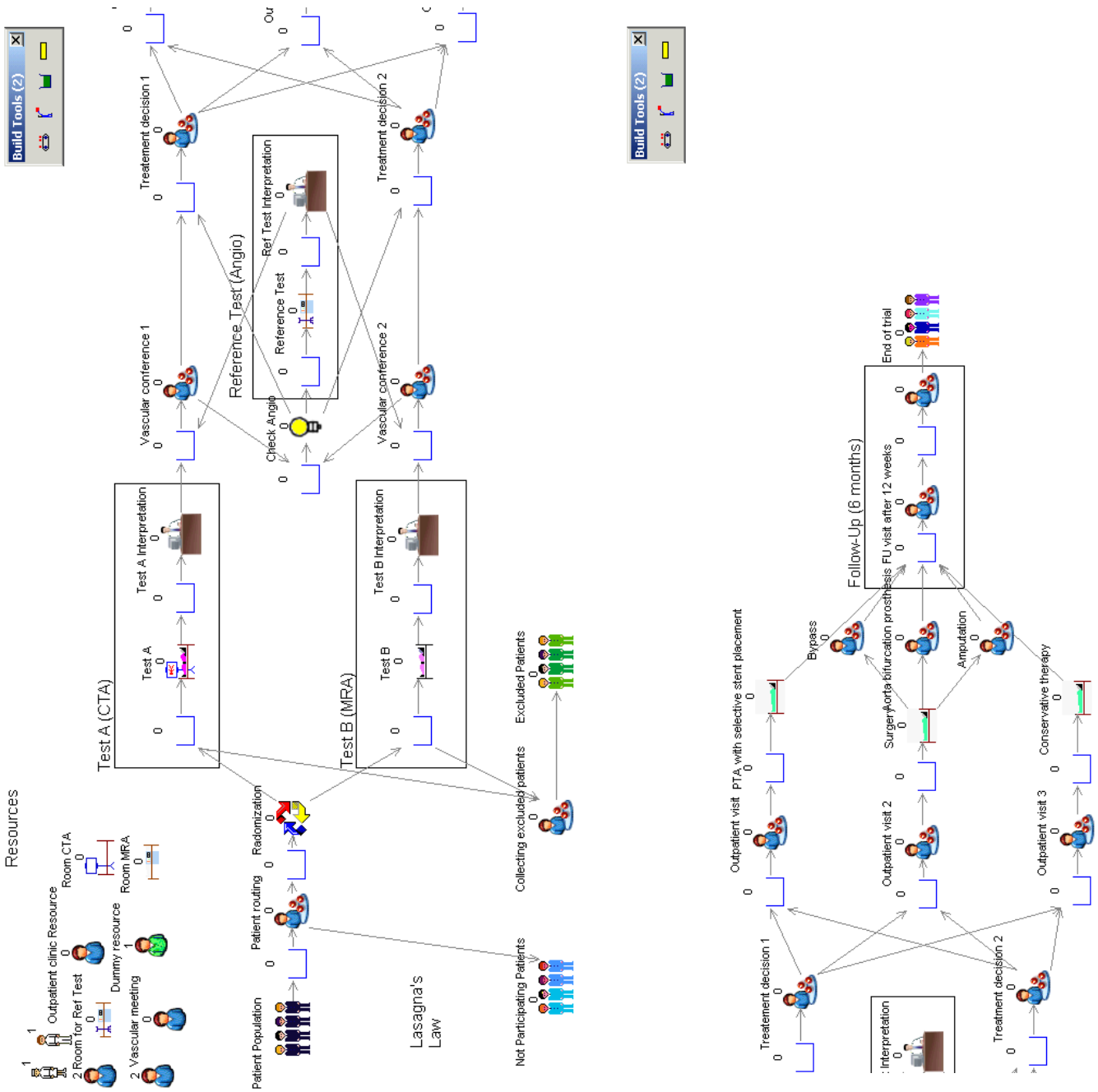


Fig. 21 – Screenshot of the simulation model developed for the first randomization strategy (first part of the model – left, second part of the model – right)

2. Randomization between providing test results versus not providing test results

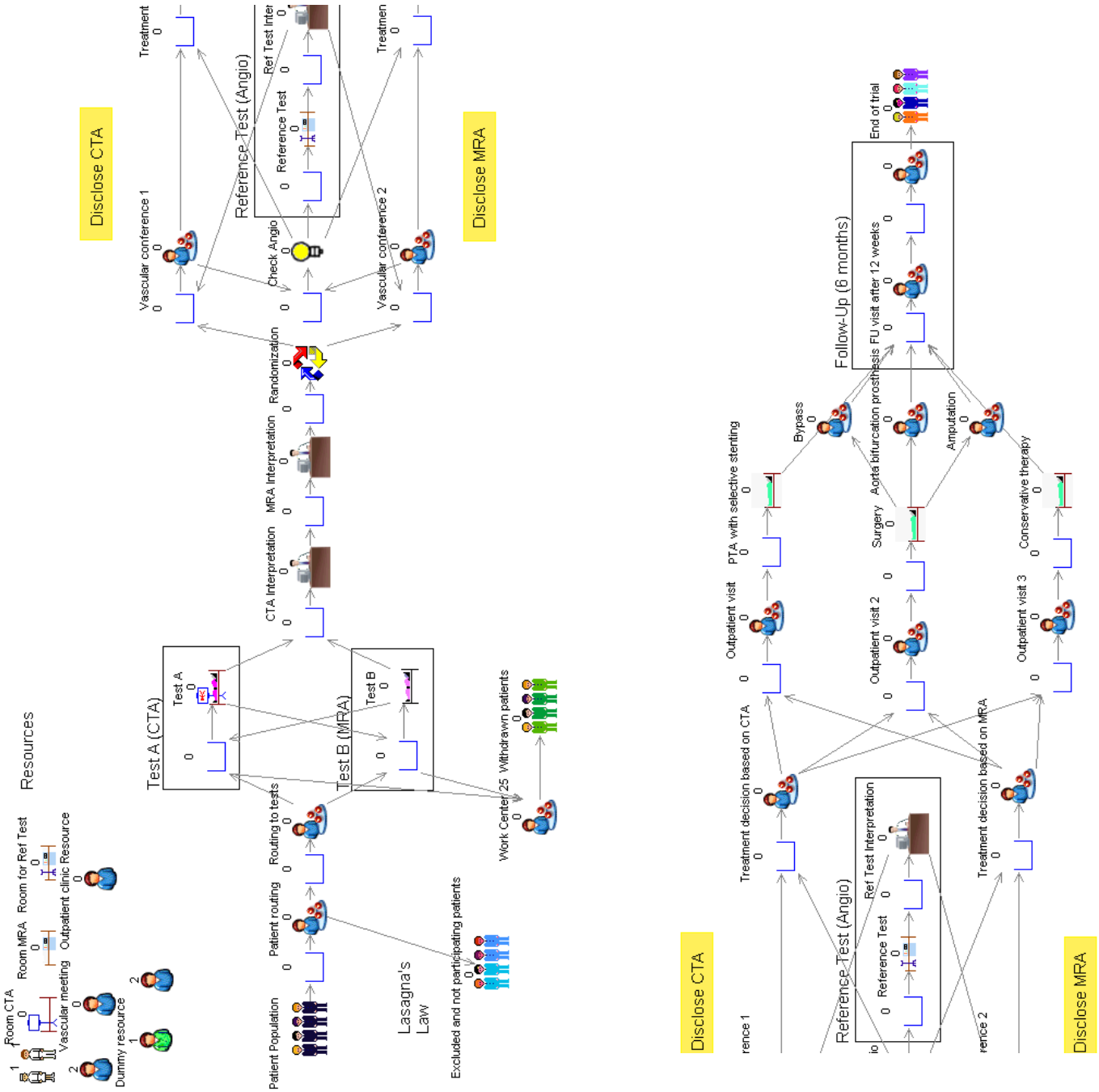


Fig. 22 – Screenshot of the simulation model developed for the second randomization strategy (first part of the model – left, second part of the model – right)

Appendix C – Summary programming code of simulation models

1. First randomization strategy (baseline scenario 150 patients, 60% eligibility)

List of patient labels used in simulation model:

- Number of outpatient visits (value in [2;6]; assigned at Work Center 24; the number of additional outpatient visits can be obtained by reducing the value of this label by 2)
- Ref Test done (value 1 or 2; 1: no reference test done, 2: reference test done; assigned at work center “Reference Test”, initialized at work entry point “Patient Population”)
- Test (value 1 or 2; 1: patient randomized to CTA group, 2: patient randomized to MRA group; assigned at work center Test A and Test B)

List of resource shifts in simulation model:

We used shift patterns of the resources for CTA, MRA, DSA, vascular meeting, the staff and dummy resources for the outpatient clinics. These shift patterns were used to constrain the availability of the resources mentioned above.

SECTION: Radomization Action Logic

```
IF [Test A.Completed+Test B.Completed]  >  150
  Set Route Out Percent     Patient routing ,  100 ,  Excluded and not participating patients
  Set Route Out Percent    Patient routing ,  0 ,  Queue for Radomization
  Percent Route Out Adjust to 100    Patient routing
  Set Shift Resources    Examination room A ,  Room CTA ,  0
  Set Shift Resources    Examination room B ,  Room MRA ,  0
ELSE
  IF Queue for Test A.Count Contents  >=  8
    Set Route Out Percent     Patient routing ,  45 ,  Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  55 ,  Queue for Radomization
    Percent Route Out Adjust to 100    Patient routing
    IF Queue for Test B.Count Contents  >=  10
      Set Route Out Percent     Patient routing ,  50 ,  Excluded and not participating patients
      Set Route Out Percent    Patient routing ,  50 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
    IF Queue for Test B.Count Contents  >=  12
      Set Route Out Percent     Patient routing ,  55 ,  Excluded and not participating patients
      Set Route Out Percent    Patient routing ,  45 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
    IF Queue for Test B.Count Contents  >=  14
```

```
      Set Route Out Percent     Patient routing ,  60 ,  Excluded and not participating
patients
      Set Route Out Percent    Patient routing ,  40 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
   ELSE
     Set Route Out Percent    Patient routing ,  60 ,  Queue for Radomization
     Set Route Out Percent     Patient routing ,  40 ,  Excluded and not participating
patients
     Percent Route Out Adjust to 100    Patient routing
   IF Queue for Test A.Count Contents  >=  10
     Set Route Out Percent     Patient routing ,  50 ,  Excluded and not participating
patients
     Set Route Out Percent    Patient routing ,  50 ,  Queue for Radomization
     Percent Route Out Adjust to 100    Patient routing
     IF Queue for Test B.Count Contents  >=  12
      Set Route Out Percent     Patient routing ,  55 ,  Excluded and not participating
patients
      Set Route Out Percent    Patient routing ,  45 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
     IF Queue for Test B.Count Contents  >=  14
      Set Route Out Percent     Patient routing ,  60 ,  Excluded and not participating
patients
      Set Route Out Percent    Patient routing ,  40 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
   ELSE IF Queue for Test B.Count Contents  >=  12
     Set Route Out Percent    Patient routing ,  45 ,  Queue for Radomization
     Set Route Out Percent     Patient routing ,  55 ,  Excluded and not participating
patients
     Percent Route Out Adjust to 100    Patient routing
     IF Queue for Test B.Count Contents  >=  14
      Set Route Out Percent     Patient routing ,  60 ,  Excluded and not participating
patients
      Set Route Out Percent    Patient routing ,  40 ,  Queue for Radomization
      Percent Route Out Adjust to 100    Patient routing
   ELSE IF Queue for Test B.Count Contents  >=  14
     Set Route Out Percent     Patient routing ,  60 ,  Excluded and not participating
patients
     Set Route Out Percent    Patient routing ,  40 ,  Queue for Radomization
     Percent Route Out Adjust to 100    Patient routing
   ELSE
     Set Route Out Percent    Patient routing ,  60 ,  Queue for Radomization
     Set Route Out Percent     Patient routing ,  40 ,  Excluded and not participating
patients
     Percent Route Out Adjust to 100    Patient routing
```

SECTION: Check Angio Action Logic

```
 IF Ref Test done  =  2
   IF Test  =  1
     Move Work Item To    Queue for Treatement decision 1 ,  -1
   ELSE
```

```
    Move Work Item To    Queue for Treatment decision 2 ,  -1
  ELSE
    Move Work Item To    Queue for Ref Test ,  -1
```

2. Second randomization strategy (baseline scenario 150 patients, 60% eligibility)

With respect to the first randomization strategy, the second strategy differs in the routing to the initial diagnostic tests, the programming code for routing between the tests and the patient labels used in the model.

SECTION: Routing to tests Action Logic

```
 IF CTA Interpretation.Completed  >=  150
   Set Route Out Percent      Patient routing ,   100 ,   Excluded and not participating patients
   Set Route Out Percent    Patient routing ,  0 ,  Queue for Routing to test
   Percent Route Out Adjust to 100    Patient routing
   Set Shift Resources    Examination room A ,  Room CTA ,  0
   Set Shift Resources    Examination room B ,  Room MRA ,  0
 ELSE
   IF Queue for Test A.Count Contents+Queue for Test B.Count Contents  >=  12
    Set Route Out Percent      Patient routing ,   45 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  55 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
   ELSE
    Set Route Out Percent      Patient routing ,   40 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  60 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
   IF Queue for Test A.Count Contents+Queue for Test B.Count Contents  >=  16
    Set Route Out Percent      Patient routing ,   50 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  50 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
   ELSE IF Queue for Test A.Count Contents+Queue for Test B.Count Contents  >=  20
    Set Route Out Percent      Patient routing ,   65 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  45 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
   ELSE IF Queue for Test A.Count Contents+Queue for Test B.Count Contents  >=  24
    Set Route Out Percent      Patient routing ,   60 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  40 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
   ELSE
    Set Route Out Percent      Patient routing ,   40 ,   Excluded and not participating patients
    Set Route Out Percent    Patient routing ,  60 ,  Queue for Routing to test
    Percent Route Out Adjust to 100    Patient routing
```

SECTION: Test A Work Complete Logic

```
IF MRA done  =  2
  Move Work Item To    Queue for CTA Interpretation ,  -1
ELSE
  Move Work Item To    Queue for Test B ,  -1
  IF Simulation Time-Patient arrival time  >=  19200
    SET Patient shell life  =  0
  ELSE
    SET Patient shell life  =  [19200-Simulation Time]+Patient arrival time
```

The programming code for test B is similar to test A.

List of additional patient labels used in simulation model:

- CTA done/MRA done (value 1 or 2; 1: CTA/MRA not yet done, 2: CTA/MRA done; assigned at work center “Test A” and “Test B”, initialized at work entry point “Patient Population” with value 1)
- Patient shell life (value initialized at work entry point “Patient Population” with value 19200 minutes, subsequently reduced by waiting time of patient for initial tests; if label value ≤ 0 patients leave system → withdrawal)
- Patient arrival time (value initialized at work entry point “Patient Population”, set the value of the label to the current simulation time)

Appendix D – Statistical data analysis of input data

The input data of the simulation model was obtained from a recently completed randomized controlled trial performed at the ErasmusMC Rotterdam by R. Ouwendijk, the DIPAD trial is described in section 3.1
The arrival process of patients suffering from PAD was modeled as a stochastic process. To fit a stochastic model we used the inclusion dates of the DIPAD trial. We transformed the data into a time series of the number of included patients per week represented in Fig. 23.

In order to fit a model, we used the Kolmogorov-Smirnoff test that requires no assumptions about the underlying statistical distribution of the data. Employing the Kolmogorov-Smirnoff test, we could not reject the hypothesis that the underlying model is a Poisson distribution with a mean equal to 1.77381 at a confidence level of 95% (p-value=0.864). From the theory of stochastic processes it follows (together with the independence of the arrivals) that the interarrival times of the patients follow an exponential distribution with parameter 1.77381 (in the unit of weeks). We furthermore calculated that the percentage of participating patients in our data was 60%, so we calculated the “real” interarrival times of the patients. Our calculations resulted in an exponential distribution of the interarrival times of the patient with a parameter equal to 2.96 (unit: weeks).

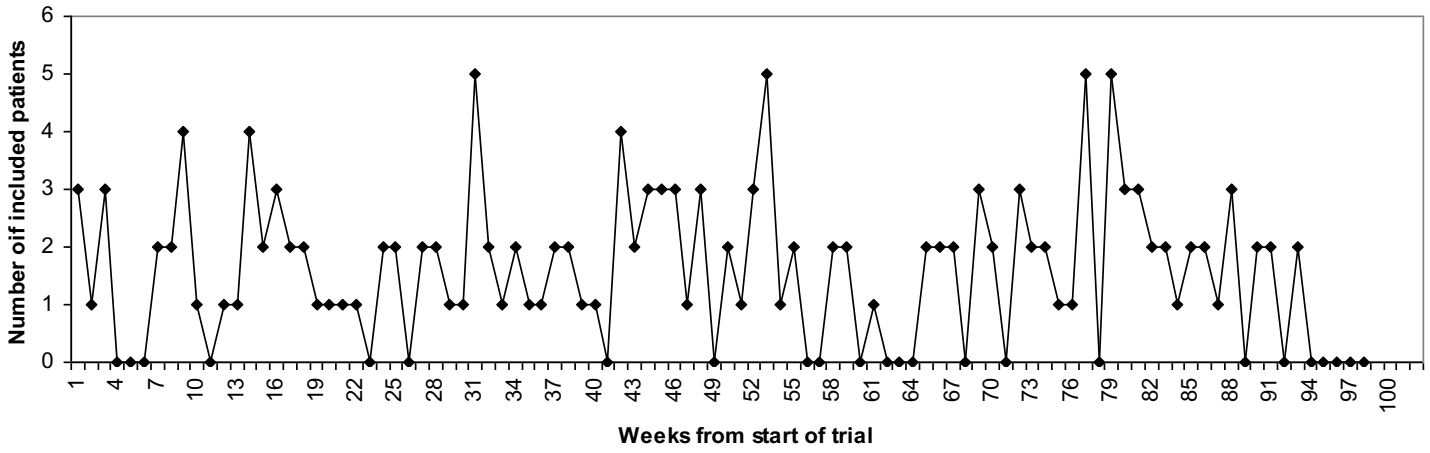


Fig. 23 - Time series of the inclusion dates of the DIPAD trial

When analyzing the data as a time series of the number of included patients, we saw a strong weekly pattern in the data. The outpatient clinic for patients with PAD was usually held one day every week so during that day the inclusion usually took place. However, a small fraction of patients was included outside these hours. For the sake of simplicity, we used the mean number of patients included per week to determine the arrival process of the system.

For further insight into the underlying stochastic process we used descriptive statistical and data analytical methods.

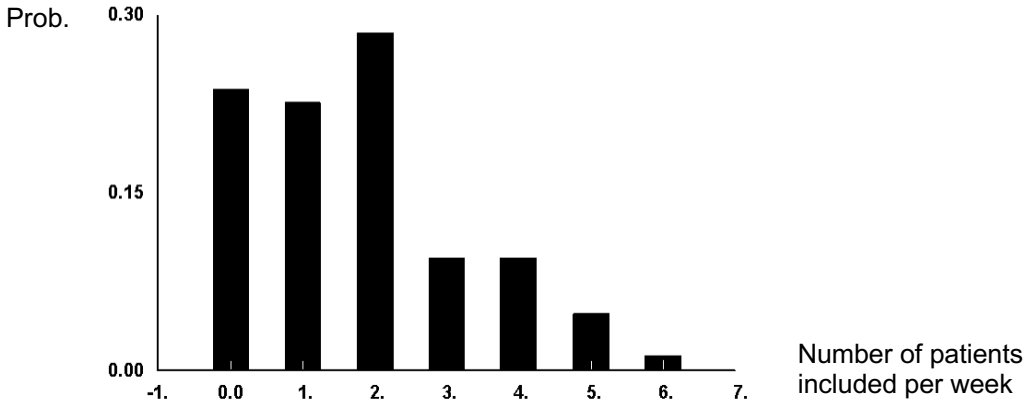


Fig. 24 - Histogram of the number of patients included per week

From Fig. 24 we see that during more than 20% of the weeks no patient was included in the study. Most frequently, two patients were included per week. In only about 30% of the weeks the number of included patients exceeds this value.

Furthermore, we made assumptions on the duration of the test itself, but also on the duration of the post-processing and the interpretation of the test results. In the following table the descriptive statistics of the test durations for CTA, MRA and DSA are given as well as the duration of the interpretation of the test results of the respective tests.

	Mean	**Median**	**Std. Dev.**	**Minimum**	**Maximum**	**N**
Duration CTA	11.48	10	2.20	10	20	79
Duration MRA	46.36	40	18.99	10	100	77
Duration diagn. DSA	101.92	90	39.35	45	200	11
Duration Interpret. CTA	23.74	24	7.06	10	43	79
Duration Interpret. MRA	16.69	13.5	9.35	3	55	78

Table 7 – Descriptive statistics of test duration, interpretation of test results of CTA, MRA, PTA and diagn. DSA.

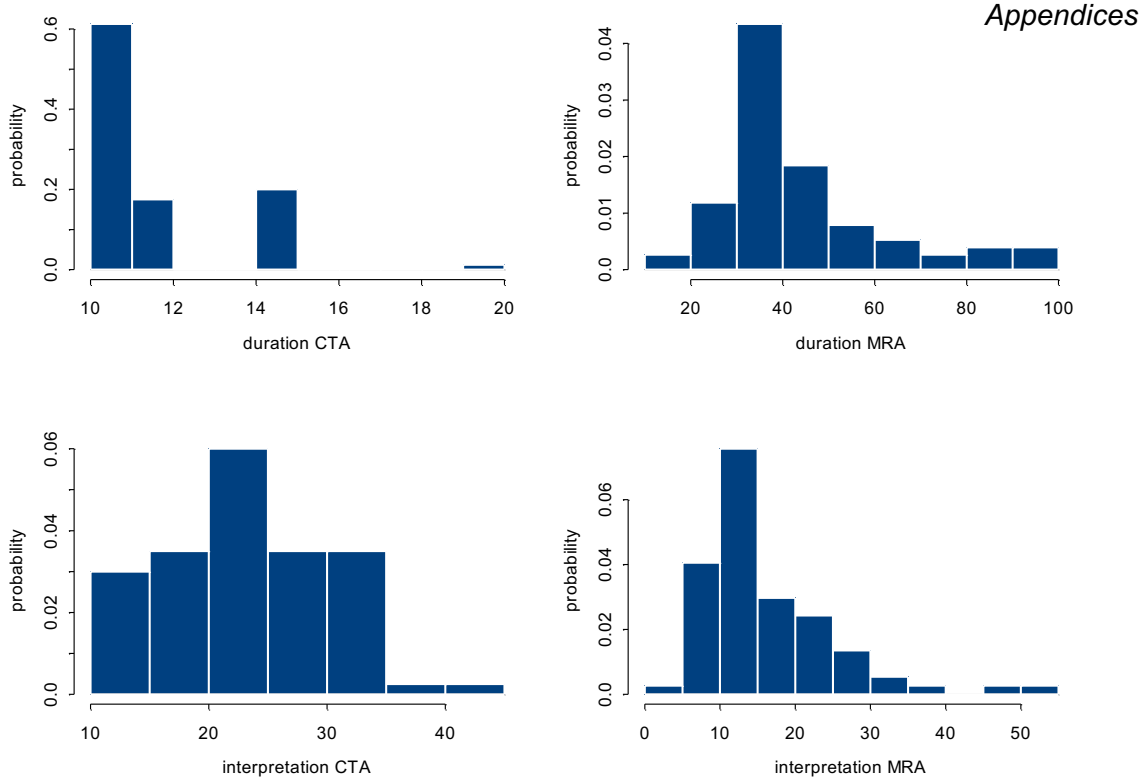


Figure 25 – Histograms of the test duration of CTA/MRA and the duration of interpreting the images obtained by CTA/MRA (incl. dictating), unit: minutes

The figure above shows that the time of perform a CTA is relatively constant, most frequently it takes 10 minutes. At the Radiology department of the ErasmusMC appointments for a CTA are scheduled every 15 minutes. These time slots include the time to lead the patients in the examination room, prepare and perform the diagnostic test and see out the patient after the test is done. Applying the Kolmogorov-Smirnoff test on the data at a confidence level of 95%, we could not determine a good fit for the data. Therefore, we chose to model the duration of undergoing a CTA as a constant of 15 minutes. As for the MRA, patients are scheduled every 45 minutes and we also modeled this variable as a constant service time of 45 minutes as we could not fit a distribution to the data. The duration of a diagnostic DSA shows a range between 45 and 200 minutes. All participants of the DIPAD trial could undergo a diagnostic DSA within a few days, so given the mean number of patients included per week we chose to reserve four time slots for the DSA per week, each taking 2 hours. As a result of the broad range of the DSA durations, a refinement of the scheduling technique

applied by the ErasmusMC may be useful, but this goes beyond the scope of this paper. The interpretation of the MRA and CTA results differ considerably. CTA takes longer to interpret the image and there is less deviation in this process. Using the Kolmogorov-Smirnoff test with confidence level equal to 95%, we could conclude that these durations are significantly different. A triangular distribution with a minimum duration of 9 minutes, a mean of 22 minutes and a maximum duration of 44 minutes fitted the data best. As for the MRA, the duration was no significantly different to a lognormal random variable with the parameters 2.4428 and 0.6113.

Directly after undergoing the diagnostic imaging test the follow-up period of half a year started including two outpatient visits, questionnaires about the patients' quality of life and the measurement of the ankle-brachial index and the maximum walking distance. Apart from the scheduled outpatient visits, additional outpatient visits may be necessary due to the treatment or aggravations of the disease state. These are generated at the beginning of the follow-up according to a probability profile obtained from the DIPAD trial, see Fig. 26.

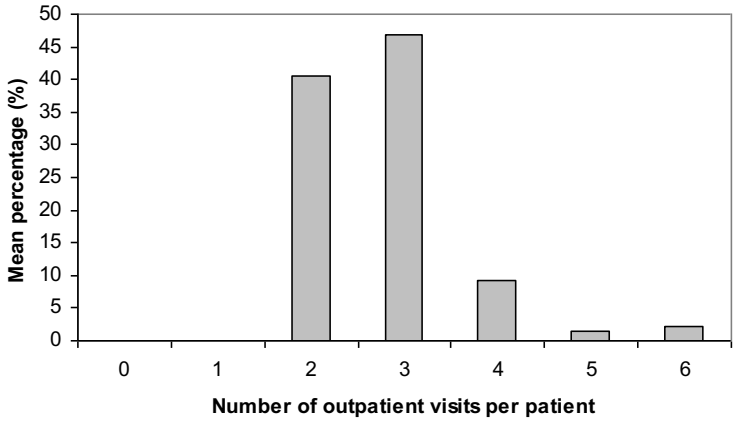


Fig. 26 – Histogram of the number of outpatient visits

Appendix E – Tabulated results of simulation experiments

	Utilization: Mean (SD)		Queuing time: Mean (SD)		Duration: Mean (SD)	Mean research costs	Mean trial costs
	CTA	MRA	CTA	MRA			
			(weeks)	(weeks)	(weeks)	(Euro)	(Euro)
100-50%	37.51607 (3.307)	36.98671 (3.356)	1.8 (0.82)	2.1 (0.9)	94.1 (6.36)	€ 332,660.00	€ 859,320.00
100-60%	44.77728 (4.641)	44.19471 (4.569)	1.8 (0.88)	2.1 (0.94)	83.3 (5.8)	€ 298,390.00	€ 824,750.00
100-70%	51.01827 (5.012)	50.4679 (4.738)	1.9 (1.04)	2.1 (1.02)	76.6 (4.58)	€ 277,220.00	€ 805,000.00
150-50%	36.97145 (3.081)	36.75127 (3.114)	1.7 (0.66)	2 (0.64)	129.7 (8.69)	€ 462,670.00	€ 1,250,960.00
150-60%	43.53973 (3.463)	43.19415 (3.402)	1.7 (0.76)	2 (0.74)	113.7 (7.35)	€ 411,900.00	€ 1,200,900.00
150-70%	51.10655 (4.015)	50.74656 (3.97)	1.8 (0.94)	2.1 (0.9)	100.7 (5.92)	€ 370,850.00	€ 1,154,970.00
500-50%	37.15047 (2.264)	37.06165 (2.256)	1.6 (0.36)	2 (0.4)	364.8 (20.45)	€ 1,328,510.00	€ 3,967,540.00
500-60%	44.22756 (2.623)	44.14471 (2.606)	1.7 (0.48)	2 (0.46)	310.9 (16.94)	€ 1,157,590.00	€ 3,799,920.00
500-70%	51.39002 (3.095)	51.29291 (3.059)	1.7 (0.58)	2 (0.52)	271.1 (14.5)	€ 1,031,090.00	€ 3,673,080.00
150-60%							
1 CTA,MRA	85.57389 (6.236)	84.876 (5.849)	3 (1.44)	3.3 (1.42)	114.2 (6.81)	€ 413,590.00	€ 1,202,590.00
1/2 CTA, MRA	98.37786 (0.518)	98.122 (0.687)	7.3 (1.08)	7.5 (1.04)	180.3 (1.44)	€ 623,220.00	€ 1,412,230.00

Table 8 - Simulation results for the first randomization strategy for combinations of number of included patients (100,150,500) and proportion of eligible patients (50%, 60% and 70%)
SD – Standard deviation

	Utilization: Mean (SD)		Queuing time: Mean (SD)		Duration Mean (SD)	Mean research costs	Mean trial costs
	CTA	MRA	CTA	MRA			
	(%)	(%)	(weeks)	(weeks)	(weeks)	(Euro)	(Euro)
100 - 50%	72.97237 (6.214)	73.0423 (5.932)	1.8 (0.82)	2.2 (0.9)	96.1 (6.01)	€ 363,880.00	€ 881,360.00
100 - 60%	82.67086 (4.491)	82.67299 (4.504)	2.2 (0.89)	2.6 (0.94)	87.6 (3.62)	€ 336,810.00	€ 860,860.00
100 - 70%	88.91586 (2.961)	88.5963 (3.295)	2.8 (2.98)	3.1 (1.02)	83.5 (2.42)	€ 323,920.00	€ 847,000.00
150 - 50%	72.24631 (5.551)	72.21586 (5.555)	1.7 (1.4)	2.1 (1.32)	131.5 (8.54)	€ 505,630.00	€ 1,293,290.00
150 - 60%	82.75391 (5.069)	82.67576 (4.924)	2.1 (1.86)	2.6 (1.84)	117.9 (5.47)	€ 462,490.00	€ 1,245,180.00
150 - 70%	90.1544 (3.123)	90.26667 (3.044)	2.7 (2.34)	3.1 (2.38)	110 (2.75)	€ 437,320.00	€ 1,223,680.00
500 - 50%	73.66957 (4.29)	73.64136 (4.308)	1.6 (1.34)	2.1 (1.22)	367.2 (19.8)	€ 1,459,430.00	€ 4,061,310.00
500 - 60%	85.1254 (4.176)	85.10413 (4.20)	2.1 (1.96)	2.6 (1.8)	321.1 (14.6)	€ 1,313,250.00	€ 3,917,780.00
500 - 70%	92.26704 (2.537)	92.31382 (2.459)	2.7 (2.4)	3.2 (2.36)	297.9 (7.59)	€ 1,239,490.00	€ 3,846,310.00
150 - 60%							
1 CTA,MRA	97.90271 (1.491)	95.77672 (1.548)	6.9 (1.4)	5.5 (1.94)	188.8 (2.47)	€ 687,331.92	€ 1,478,322.81
first CTA	85.04689 (589)	84.18018 (5.893)	3 (1.44)	1.4 (0.6)	116 (6.25)	€ 456,350.00	€ 1,239,040.00
random seq.	82.75391 (5.069)	82.67576 (4.924)	2.1 (1.86)	2.6 (1.84)	117.9 (5.47)	€ 462,490.00	€ 1,245,180.00
circulate	83.98737 (5.465)	84.06081 (5.478)	2 (1.74)	2.4 (1.62)	116.4 (6.05)	€ 457,600.00	€ 1,240,290.00
shortest qu.	84.20248 (5.445)	84.357 (5.612)	2 (1.46)	2.4 (1.28)	116.5 (6.2)	€ 458,040.00	€ 1,240,740.00

Table 9 - Simulation results for the second randomization strategy for combinations of number of included patients (100,150,500) and proportion of eligible patients (50%, 60% and 70%)
SD – Standard deviation