

# **Analyzing study designs of clinical studies for the evaluation of diagnostic imaging modalities**

**J.F. Veldhuis**

Master thesis, December 2005

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

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



## Preface

The final phase of the study Business Mathematics and Informatics (BMI) at the Vrije Universiteit Amsterdam consists of an obligatory internship. The internship should integrate the three disciplines of the study, Business administration, Mathematics and Computer science.

This master thesis is about my internship that is performed within the Assessment of Radiological Technology (ART) group at the Department of Epidemiology & Biostatistics at the Erasmus Medical Center Rotterdam. This group focuses on the assessment of medical imaging technology, especially related to cardiovascular disease.

Projects are hardly performed by only one person, therefore I would like to thank several people for their support and help during my internship.

First I would like to thank Prof. Myriam Hunink, the leader of the ART group, for giving me the opportunity to perform my internship within the ART group. Thank you for your support and encouragement. Your energy and efficient way of working inspired me to get the best out of myself.

I would also like to thank Dr. G.J. Franx, my supervisor from the Vrije Universiteit. Thank you for your time and great discussions, I enjoyed it a lot.

My colleagues from the ART group were very important for the completion of this study. I would like to thank Majanka, Ineke, Ylian, Marion, Joke, Jan-Jaap, Tien, Sandra, Rody and Marc for the great time (I loved the Wednesdays meetings!) and for their patience while answering all my questions. Taye, thanks for being my roommate!

Finally, I would like to thank Eveline van Kan and Linda Everse for answering all my questions about the financial aspects of a clinical study.

Jolanda Veldhuis  
December 2005



## Abstract

The objective of the study was to evaluate and compare several study designs, which will guide the choice of efficient diagnostic study designs and which will maximize the return on money invested in diagnostic research.

Four discrete event simulation models were developed to analyze the Cohort study, the Before-After study, the Randomized Controlled Trial with randomization of all patients across diagnostic testing strategies (RCT-1) and the Randomized Controlled Trial with randomization between providing results of test A versus providing results of test B (RCT-2).

Data of the Living Renal Donors study, DIPAD trial and the Coronary CTA study was used for the comparison of study designs. For each clinical study the effect of using different study designs were measured. Different scenarios were analyzed as well. These scenarios included realistic external factors that might occur when clinical studies are performed in practice. The study designs were evaluated in terms of the utilization of the test modalities, the number of withdrawn patients, the duration of the study and the total costs.

The results show that the duration, utilization of the test modalities and the total costs of a study are almost similar for the Cohort and RCT-2 study designs. These outcome measures are similar as well for the Before-After and RCT-1 study design. A study is less expensive when it is set up according to the Before-After or RCT-1 study design than when it is set up according to the Cohort or RCT-2 study design.

The results of the sensitivity analysis suggest that it is better to reserve extra capacity for the diagnostic imaging than on average needed. An overcapacity of the number of reserved timeslots of 150% results in a reduction of 7% of the total costs and a reduction of 12% of the total duration of a study.

## Table of Contents

<b>Preface</b> .....	<b>3</b>
<b>Table of Contents</b> .....	<b>6</b>
<b>1 Introduction</b> .....	<b>8</b>
1.1 Erasmus medical Center, Department of Epidemiology & Biostatistics .....	8
1.2 Problem description .....	8
1.3 Objective of the study.....	9
1.4 Contents of this document .....	9
<b>2 Study designs: Background information</b> .....	<b>10</b>
2.1 Clinical studies .....	10
2.2 Cohort study .....	12
2.3 Before-After study .....	13
2.4 Randomized Controlled Trial with randomization of all patients across diagnostic testing strategies .....	14
2.5 Randomized Controlled Trial with randomization between providing results of test A versus providing results of test B .....	15
2.6 Differences between study designs .....	16
<b>3 Methods</b> .....	<b>18</b>
3.1 Discrete Event Simulation .....	18
3.2 Measurement of Costs .....	19
3.3 Simulation model.....	20
3.3.1 Basic Process .....	20
3.3.2 Assumptions .....	22
3.3.3 Outcome Measures.....	24
3.3.4 Cohort simulation model .....	25
3.3.5 Before – After simulation model.....	26
3.3.6 RCT simulation model with randomization of all patients across diagnostic testing strategies .....	27
3.3.7 RCT simulation model with randomization between providing the results of Test A versus providing results of Test B .....	28
3.4 Verification and Validation.....	29
3.4.1 Verification and Validation Living Renal Donors study .....	29
3.4.2 Verification and Validation DIPAD trial .....	31
3.5 Analysis .....	33
3.5.1 Analysis of Living Renal Donors Study.....	33
3.5.2 Analysis of DIPAD Trial.....	34
3.5.3 Analysis of an ongoing study: the Coronary CTA Study.....	34
<b>4 Results</b> .....	<b>36</b>
4.1 Results of Validation .....	36
4.2 Results of analysis Living Renal Donors study .....	37
4.3 Results of analysis DIPAD trial .....	40
4.4 Results of analysis Coronary CTA study .....	43
<b>5 Discussion</b> .....	<b>51</b>
5.1 Results .....	51
5.2 Limitations .....	53
5.3 Conclusions.....	54
<b>References</b> .....	<b>57</b>
<b>Appendix A – Scheduling patients using a probability</b> .....	<b>59</b>

**Appendix B – Screenshots of input screens Cohort Simulation model .....61**

**Appendix C – User manual of Arena simulation models.....65**

**Appendix D – Arena Models .....99**

# 1 Introduction

## 1.1 *Erasmus medical Center, Department of Epidemiology & Biostatistics*

The Erasmus University Medical Center Rotterdam (Erasmus MC) is the largest university medical center in the Netherlands with more than 10,000 employees. The core activities of Erasmus MC are patient care, education and research. The core activity at the hospital is patient care; at the medical faculty it is education and research. Every medical department has an educational purpose and conducts its own 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 in clinical epidemiology and biostatistics for clinicians of the Erasmus Medical Center Rotterdam. The research activities of the department are organized in three clusters; Epidemiology of Diseases, Basic Epidemiology, 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). The ART program comprises a network of researchers who focus on the assessment of medical imaging technology, both diagnostic imaging and image-guided therapies, 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 *Problem description*

Decision making in clinical medicine and health-care has become a very complicated process [2]. Trade-offs need to be made between risks, benefits, costs and preferences. In the early days a physician had a few possible diagnoses, several simple tests, and a few, mostly ineffective, treatments to choose from. Nowadays physicians can choose from a broad range of diagnoses, new technologies and treatments. This certainly improves the potential to help the patients, but it also increases costs and it makes decision making more complex and difficult.

The assessment of medical imaging technology comprises the evaluation and comparison of diagnostic test strategies. The benefits of one diagnostic strategy over another diagnostic strategy are investigated in clinical studies.

The evaluation and comparison of diagnostic test strategies entails a long process of clinical studies, data collection and decision modeling. At the end of this process the evaluation commonly demonstrates only very small benefits from replacing one diagnostic strategy with another, leading to the typical conclusion "more research is needed". This calls into question whether elaborate large diagnostic research studies are necessary, whether the chosen outcomes measures are relevant and whether the currently used study designs are efficient.



The decision, which study design to use for a study, does not only influence the obtained information while performing the study, it also influences the operational and financial control. Incorrectly planned studies can lead to excessive overrun of the assigned budget and might lead to insufficient information.

### **1.3 Objective of the study**

The **overall objective** of this study is to evaluate and compare several study designs, which will guide the choice of efficient diagnostic study designs and which will maximize the return on money invested in diagnostic research.

The **specific aims** are to determine the following for each study design and for several scenarios (i.e. several different clinical problems):

- What will be the expected duration of the study?
- What are the expected total research costs of a particular study?
- What is the utilization of the imaging equipment when performing the study?
- What is the bottleneck in performing the study?

Subsequently we will address the following specific research questions:

- How are the duration of the study, the research costs, the utilization of the imaging equipment, and the bottleneck of the study influenced by varying the number of reserved timeslots for the diagnostic imaging modalities?
- How are the duration of the study, the research costs, the utilization of the imaging equipment, and the bottleneck of the study influenced by the choice of study design?
- How are the duration of the study, the research costs, the utilization of the imaging equipment, and the bottleneck of the study influenced by a particular clinical study?

Each study design provides different information, for this reason we will determine for several clinical studies which study design is the most efficient.

### **1.4 Contents of this document**

The document starts with providing background information of clinical studies and different study designs in Chapter 2. In Chapter 3 we will discuss the research methodology. A detailed description of the simulation models is provided. The methods for verifying and validating the models are described as well. The last section of the chapter contains a description of the performed analysis. The results of the verification, validation and analysis are presented in Chapter 4. Finally, conclusions are given in Chapter 5.

## 2 Study designs: Background information

In this chapter we will provide some background information about the organization of clinical studies.

### 2.1 Clinical studies

To diagnose whether a patient has a certain disease or not, physicians perform diagnostic tests. A diagnostic test strategy contains the test to diagnose a certain disease and the probability that a certain treatment is performed. Figure 1 visualizes this concept:

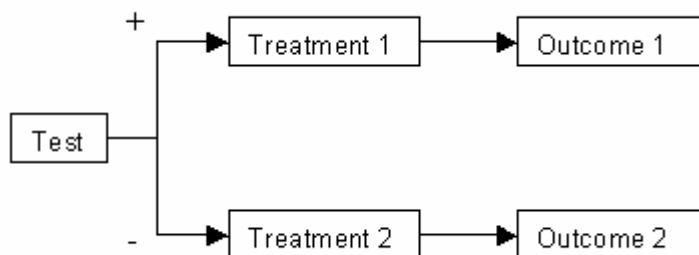


Figure 1: Diagnostic test strategy

In figure 1 we assumed that a test result is dichotomous, “+” stands for a positive test result and “-“ stands for a negative test result. The treatment of the illness will be based on the result of the test; treatment 1 will be performed when the test result is positive, treatment 2 when it is negative. Possible treatments are medication and surgery. A diagnostic test strategy for a test with multi-category results will have as many branches as the number of results that are possible.

There are different tests that can yield the same diagnosis. For example a computed tomographic angiography (CTA) scan and a magnetic resonance angiography (MRA) scan can be used to evaluate peripheral arterial disease (PAD).

Several diagnostic test strategies exist for every disease. It is interesting to know which test strategy is the most efficient to diagnose a certain disease. Therefore clinical studies are performed. Clinical studies are research studies to determine whether new or known therapies, treatments and tests are both safe and effective. In many clinical studies patients are divided into two groups, one group of patients will be given an experimental drug, treatment or test while the control group is given either a standard treatment for the illness, a placebo or the reference test. The two groups are required to make a meaningful comparison between the effectiveness of the new treatment and the standard treatment or placebo. To avoid selection bias it is required to randomize the patients between the experimental and the control treatment.

Patients are selected into a clinical study according to inclusion and exclusion criteria. The inclusion and exclusion criteria give a precise definition of which patients are eligible for the study. The main objective of these criteria is to ensure the selection of a representative group

of patients that are most likely to benefit from the new drug, treatment or test under investigation. However, the inclusion and exclusion criteria should not be too restrictive. Otherwise the clinical study will be too small and its findings will lack generality.

We can distinguish between an open clinical study and a blinded clinical study. In an open clinical study both the physician and the patient knows which treatment is being used. This might lead to distortion of the comparison of treatments, for example, there is a psychological benefit when a patient knows that he or she is receiving the new treatment.

In a single-blind study the patient does not know which intervention is provided. In a double blind study neither the physician nor the patient is aware of which intervention the patients receives. Double blinding in a clinical study is preferred since the influence of external factors on the results is minimized and the bias of the results is reduced. It is not always possible to perform double blinding in a clinical study. Double blinding in a study where diagnostic imaging technology is evaluated is very complicated because an average patient and the radiologist both know the difference between a CTA and a MRA. Whether double blinding should be applied depends on the individual circumstances of each clinical study.

Each clinical study is organized through a certain study design. A study design influences the way the study is carried out and the final outcomes that will be measured. There are many different study designs that can be used to set up the research process of diagnostic test strategies, for example:

- a. Cohort studies
- b. Before-after studies
- c. Randomized control trial (RCT) with randomization of all patients across diagnostic testing strategies
- d. Randomized trial with randomization between providing the results of Test A versus providing the results of Test B

In this project we compared and evaluated the mentioned study designs for several different types of clinical studies using simulation methods. We focused our study on clinical studies for the evaluation of diagnostic imaging technology.

In the next sections we will describe the process of each of the study designs in the scope of diagnostic imaging and several differences between the study designs.

## 2.2 Cohort study

In a cohort study for the evaluation of diagnostic imaging technology the researcher will evaluate and compare the outcomes of both the new test and the current used test that all patients underwent in the study.

The process of a study organized through a cohort study design can be represented by the scheme in figure 2.



**Figure 2: Scheme of Cohort study design**

According to this study design, patients enter the system and will then undergo both tests, first test A and then test B. After that, a reference test will be done to determine whether the outcomes of both tests are valid. The reference test is seen as “the gold standard”, which means that the result of the reference test is assumed to be the correct one. In practice it is not very common that all patients undergo the reference test, except when the reference test is one of the test modalities that is evaluated. To keep the explanation of each study design as simple as possible, we assume for now that all patients will undergo the reference test. When the reference test is carried out a treatment will be performed based on the results of this test and subsequently the patient can be followed over time prospectively to measure the development of different outcomes.

In the cases that the reference test is performed it is very likely to assume that the treatment will be based on the outcomes of the reference test, because this is seen as the gold standard. If all patients should undergo the reference test it is likely to assume that all treatments within the study will be based on the results of the reference test. In that case it might not be necessary to monitor the patient during the treatment and follow-up period. As already mentioned, we assume that all patients will undergo the reference test for the simplicity of the explanation of the study designs. In practice, the reference test is mostly used as an additional test when the other test results are not clear enough and therefore, the reference test will only be performed in a small percentage of the participating patients.

An example of a study where the reference test is one of the test modalities that is evaluated is the study for living renal donors performed at the Erasmus MC. The study for living renal donors was a cohort study performed by M.C.J.M Kock [4]. In this particular study the researcher evaluated, for example, the agreement on decision making of the outcomes of the MRA and Digital Subtraction Angiography (DSA), which is the reference test. In this study only the test results and the decision which kidney should be used for transplantation based on each of the test results was of interest. For this reason the researcher did not monitor the patient during the follow-up period.

When a clinical study is organized through a certain study design it does not mean that the total process is performed in that particular study. It is also possible to implement the process partially, like was done for the study of the living renal donors. To keep the explanation of the

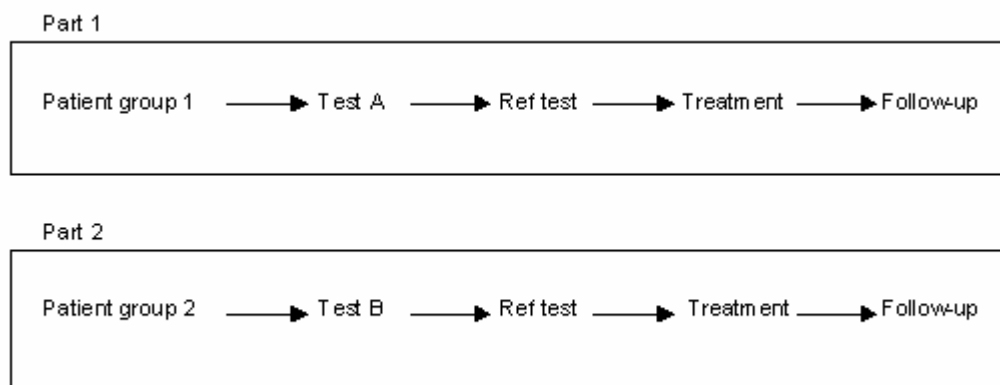
different types of study designs as clear as possible we will describe the complete process of each study design.

Using a Cohort design for a clinical study, the sensitivity and specificity of test A and test B compared to the reference test can be precisely calculated and a head-to-head comparison can be made. This sensitivity and specificity gives information about the diagnostic accuracy of the test at performance. It is important to get to know the sensitivity and specificity as precisely as possible. Disadvantages of this study design are the duration, logistics and the costs of the study incurred by performing both tests on all patients. Furthermore, it is not possible to measure the effect of each test on further management of the patient.

### 2.3 Before-After study

In a Before-After study a researcher compares the outcomes of the performance of the standard imaging test in a group of individuals with the outcomes of a new diagnostic imaging test that subsequently will be performed on a comparable group of individuals.

A study organized through the Before-After design can be divided into two parts as can be seen in figure 3.



**Figure 3: Scheme of Before-After study design**

In the first part of the study, the standard diagnostic imaging test (Test A) will be performed on a group of individuals. After a patient has undergone test A, a reference test will be done and the patient will be treated and followed.

In the second part of the study a comparable group of individuals will undergo test B. After performing test B a reference test is done and patients are treated and followed.

The two parts of the study are carried out at different moments in time; the second part occurs later in time than the first part, it will only start when the inclusion period of the first part of the study has ended.

With the use of this study design the sensitivity and specificity of test A compared to the reference test can be calculated. It is also possible to calculate the sensitivity and specificity of test B compared to the reference test. However, these values of the sensitivity and specificity are not determined concurrently and a head-to-head comparison cannot be made as is done with a Cohort Study.

The impact of introducing test B into clinical practice as a replacement of test A can be determined, but the analysis may be biased because the comparison is not concurrent.

A study using the Before-After study design is easy to integrate into routine clinical practice, since this study design is organizationally and logistically simpler than other study designs.

## 2.4 Randomized Controlled Trial with randomization of all patients across diagnostic testing strategies

A Randomized Controlled Trial (RCT) is a clinical study where the division of patients into two comparable groups is based on randomization. In this RCT patients are divided into two groups based on the diagnostic imaging test that they are allocated to. Once a patient is allocated to either test A or test B it will follow the process of the trial as denoted in the branches of figure 4. The randomized allocation of patients to the tests is denoted with a cube.

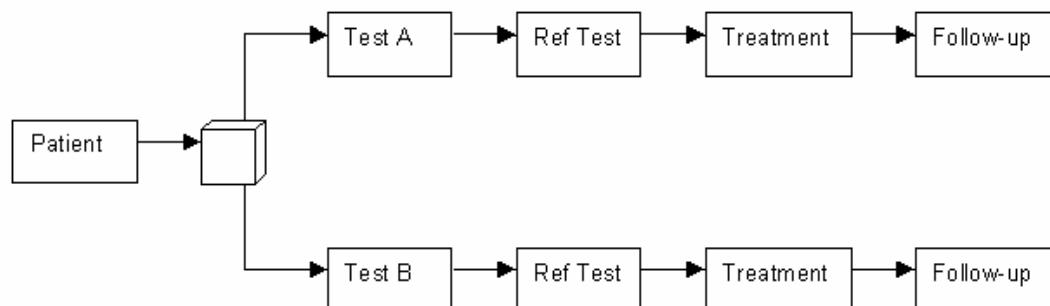


Figure 4: Randomized Control Trial with randomization of all patients across diagnostic testing strategies

An example of a RCT with randomization of patients across diagnostic testing strategies is the DIPAD trial performed by R. Ouwendijk [9]. The objective of the DIPAD trial was to evaluate clinical utility, patient outcomes and costs of MRA compared to CTA for initial imaging in the diagnostic work-up of patients with peripheral arterial disease (PAD). In PAD, fatty deposits (plaque) build up in the wall of an artery. The plaque will narrow the artery and restrict the blood circulation. In its early stages a common symptom is cramping or fatigue in the legs and buttocks during activity. With CTA or MRA the narrowing of the lumen of the arteries can be visualized.

The patients were randomly assigned to undergo either MRA or CTA as the initial imaging examination. Block randomization was used with a block size of eight patients to obtain equal numbers for both strategies. Several researchers that were unaware of the randomization sequence enrolled eligible patients. After randomization, patients and clinicians were not blinded to the imaging strategy because this would have been impossible and inconsistent with the pragmatic study design. The reference test was only performed in a small number of patients for whom a decision for treatment could not be made based on MRA or CTA. The treatments of the patients were based on the results of the diagnostic imaging modality they underwent. If an additional reference test was made, the treatment was based on the outcome of this test.

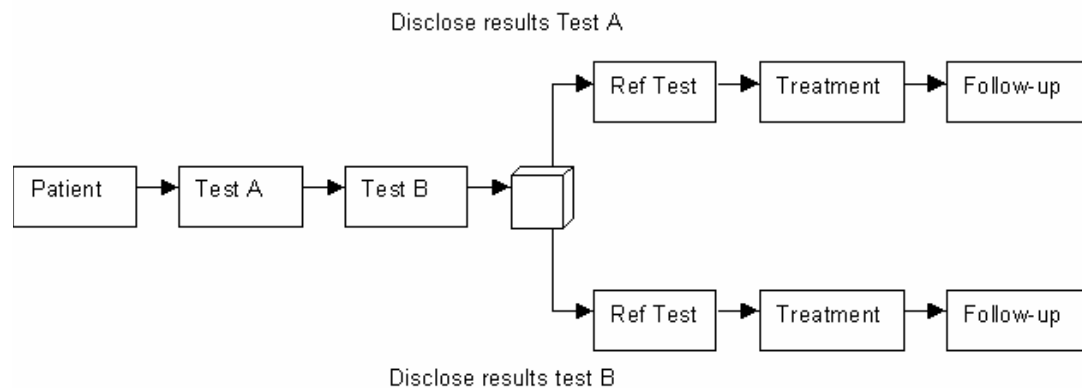
With the use of this study design it is possible to determine the sensitivity and specificity of test A and test B with less bias than in the Before-After study, but a head-to-head comparison as in

the Cohort study is not possible. Moreover, if the treatment and follow-up are based on the outcomes of test A and test B, a researcher is able to determine the diagnostic, therapeutic and prognostic effect of introducing test A and B into clinical practice.

If timeslots for tests are limited, this study design will not take as much time as the other study designs because only one test is performed and not both. Consequently, this study design might save time and money compared with the other study designs mentioned.

## 2.5 Randomized Controlled Trial with randomization between providing results of test A versus providing results of test B

The disclosure of the results of test A and B to the treating physician is randomized in this RCT. The treatment of the patient will only be based on the test result that is provided to the physician the result of the other test is discarded. If the test results are not clear enough for the diagnosis a reference test will be performed. The treatment of the patient will then be based on both the results of the reference test and the previously performed test. Figure 5 denotes the process of this type of RCT.



**Figure 5: RCT with randomization between providing the results of test A versus providing results of test B**

According to this randomization strategy, all patients undergo all diagnostic tests but only one test result is disclosed to the treating physician. The assignment of which test result is disclosed to the physician is randomized.

An example of this type of RCT is the “RCT of magnetic-resonance pelvimetry in breech presentation at term” performed by A.J. van Loon [7]. In this trial the clinical value of magnetic-resonance (MR) pelvimetry in women with breech presentation at term was evaluated. The study design of this trial differs slightly from the study design we just discussed, but another example was not directly available.

One of the objectives of this study was to assess whether use of MR pelvimetry in breech presentation at term affects the obstetric management policy, like the selection of the delivery route. In this trial the effect of providing test results versus not providing test results was measured. This means that there is no test B to evaluate. In this example test A was equal to the MR pelvimetry and test B was the current clinical practice where no diagnostic imaging was used. In other words test B is in this case a dummy.

Each of the women underwent a MR pelvimetry (test A). For only one group of the women the test results were reported to the obstetrician. For the other group (the control group) the current clinical practice was used and the results of the MR were not provided. In this way the

researchers were able to evaluate the effect of MR pelvimetry for women with breech presentation at term.

Using this study design, physicians are able to precisely determine the sensitivity and specificity of test A and test B compared to the reference test in a head-to-head comparison. This study design determines the sensitivity and specificity just as well as the Cohort study does. The impact of implementing test A and test B into clinical practice can also be determined.

A disadvantage of this study design is that the physician does not receive all available diagnostic information, which may sometimes lead to ethical problems. Furthermore, this study design may be very expensive and logistically cumbersome.

## **2.6 Differences between study designs**

The RCT with randomization between providing test results of test A versus providing test results of test B delivers the most specific information about the evaluated diagnostic tests. In theory the use of this study design will lead to the best results, namely the determination of the sensitivity and specificity in a head-to-head comparison between the index tests and a determination of the impact of test A and test B into clinical practice by a concurrent comparison. But in practice this study design will be hard to implement. Just like the Cohort study all patients undergo all diagnostic tests. If the time between the two diagnostic tests is too large the patient is not eligible anymore for the study. Therefore, during scheduling, the researcher should consider the maximum time allowed between the two diagnostic tests. Another problem of this RCT study design is the assignment of the test results to the physician. In theory each physician only receives one of the two test results. In practice however, physicians can easily request the result of the other test on ethical grounds because both tests have been performed. This affects the treatment and therefore the outcome, which may lead to biased results.

When performing the RCT with randomization of all patients across diagnostic test strategies a researcher is able to determine the impact of introducing the diagnostic tests into clinical practice. Unfortunately a researcher is not able to determine the sensitivity and specificity of two tests in a head-to-head comparison when using this study design. However performing this study design will probably take less time, might be less expensive and is easier to integrate in clinical practice than the RCT with randomization between providing the results of test A versus providing results of test B.

A Cohort study has the advantage of performing a head-to-head comparison of the sensitivity and specificity of the tests. Nevertheless, it is not possible to determine the impact of introducing a new test into clinical practice. Besides this, the study design may be time consuming and expensive because all patients will undergo all tests, just like the RCT with randomization between providing the results of test A versus providing results of test B.

The Before-After study design will probably be the least expensive study design and the easiest to integrate in routine clinical practice. However, this study will take a long time before the researchers can come to any conclusion. At that time the research might be out-of-date. Another disadvantage of the Before-After study is that comparing the sensitivity and specificity



of the tests and evaluating the impact of introducing test B into clinical practice as a replacement of test A, cannot be accurately determined.

Table 1 gives an overview of the general advantages and disadvantages.

Study design	Advantages	Disadvantages
Cohort Study	Head-to-head comparison concerning the sensitivity and specificity  Ethical	Expensive Time consuming Difficult to integrate in routine clinical practice Cannot determine the impact of introducing the tests in clinical practice
Before-After Study	The impact of introducing test B into clinical practice, as a replacement of test A can be determined.  Easy to integrate study in routine clinical practice Less expensive Ethical	A head-to-head comparison concerning the sensitivity and specificity is not possible  Time consuming
RCT across Patients	Concurrent controls for the sensitivity and specificity  The impact of introducing the tests into clinical practice can be precisely determined  Less expensive Less time consuming Ethical	A head-to-head comparison concerning the sensitivity and specificity is not possible
RCT providing results test A vs. providing results of test B	Head-to-head comparison and concurrent controls concerning the sensitivity and specificity  The impact of introducing the tests into clinical practice can be determined	May cause ethical dilemma's Time consuming Cross-overs can cause interpretation problems Expensive Logistically cumbersome Results might be biased

**Table 1: An overview of the advantages and disadvantages of the different study designs**

### 3 Methods

In our study we developed simulation models to evaluate and compare several outcome measures for the Cohort, Before-After, RCT with randomization of patients across diagnostic imaging tests and the RCT with randomization between providing test results of test A versus providing test results of test B. In this chapter we describe the methods we used for the development of the models. In the first section we give an introduction of Discrete Event Simulation. Next, information about the costs measurement is given. Section 3.3 provides a detailed description about the simulation models. In the last two sections we describe the methods for verifying and validating the models and the analysis we performed with the simulation models.

#### 3.1 Discrete Event Simulation

A model is a (simplified) description of operations of various kinds of real-world facilities or processes. The term “system” is often used to refer to the object that is to be modeled, for example a manufacturing plant with machines or a personal-service operation like a bank or call center. Through modeling we are able to get a better understanding of the behavior of the system for a given set of conditions. Simulation is one of the most widely used operations research techniques. The results of the “what-if” analysis obtained from the simulation models gain insight into a system and form a basis for management decisions.

Simulation models can be classified along three different dimensions; static vs. dynamic, deterministic vs. stochastic and continuous vs. discrete simulation models [6]. Discrete-Event Simulation (DES) is the generic term for simulation models that are classified as discrete, dynamic and stochastic.

In discrete-event simulation, the modeling of a system is represented by state variables that change at discrete points in time. An example of a state variable is the state of a physician during consulting hours; he can be idle or busy. An event is an occurrence at one point in time that may change the state of the system. The arrival of a new patient and the departure of a treated patient are examples of two different events.

The key idea of DES is that all events occur at discrete points in time. The behavior of the number of patients in the waiting room at a particular moment in time can be modeled using discrete-event simulation, because this number only changes when a patient enters the room or leaves the room. The behavior of an airplane moving through the air cannot be modeled using discrete-event simulation, since a state variable such as the position of the plane changes continuously over time.

The simulation models change in a dynamic way; therefore we must keep track of the current value of simulated time. The simulation clock gives the current value of simulated time and needs to be advanced from one value to another. Most simulation models use the next-event time-advance approach to advance the simulated time. The simulated clock is initialized to zero and the time of future events are determined. The clock is then advanced to the first of these future events. Before jumping to the next event the state of the system is updated. The process of advancing the simulation clock from one event time to another is continued until a pre-defined stopping condition is satisfied. Jumping the simulation clock from event time

to event time using the next-event time-advance approach skips the periods of inactivity between different event times.

The models we developed can be classified as Discrete Event Simulation models, since they are discrete, dynamic and stochastic in nature.

For the development of the simulation models we used the Arena Rockwell Simulation software package, version 9.0 Research Edition, which is a popular general-purpose simulation package in the field of operations research. In section 3.3 we will discuss the simulation model in more detail.

### **3.2 Measurement of Costs**

For the cost analysis, we performed a survey by interviewing the research coordinator and the financial administration staff of the radiology department. All costs were computed from the radiological perspective. According to the radiological perspective the cost prices for the imaging modalities, materials, personnel, administration and overhead are included. The costs per study can be divided into directly and non-directly assignable costs.

Directly assignable costs of a study include personnel costs like the costs for the researcher, and the costs per diagnostic imaging modality. Personnel costs per hour were computed using the highest wage scale for each involved personnel category. Next a social security of 38% of the wage was added. The costs per diagnostic imaging modality are based on the material costs, equipment costs such as the initial investment of equipment and maintenance costs. The costs for the diagnostic imaging modalities also include the personnel costs for the assistant performing the test, which were computed using the measured time spent on a diagnostic imaging test times the personnel costs per hour.

Non-directly assignable costs include overhead costs and the costs of supporting divisions such as costs for the trial office. The overhead costs for the radiology department were estimated to be 16% of directly assignable costs. The overhead costs required by the board of directors were estimated to be 16% of the directly assignable costs plus the overhead costs for the radiology department. We assume that the 16% overhead for the board of directors includes housing costs.

We calculated the costs per study separately from the Arena simulation models. The expectation is that running the simulation models will take a large amount of time. If the cost price of the material used is the only difference that is evaluated, it is a waste of time to re-run the models. Therefore we chose to build an Excel-file in which the user is able to import the results of the simulation models. Next, the user is able to perform a sensitivity analysis for the cost prices while the output of the simulation model does not change.

### 3.3 Simulation model

We performed the evaluation and comparison of the several study designs using four simulation models developed with Arena Rockwell Software and Visual Basic for Applications. All simulation models contain the same overall basic process. First we will discuss this basic process and the assumptions that are made during the development. Subsequently, we will discuss the specific features and assumptions per model.

#### 3.3.1 Basic Process

Each of the four study designs can be denoted by the basic process as shown in figure 1. Before discussing the characteristics of each study design we will give an overview of the basic process and how it is implemented.

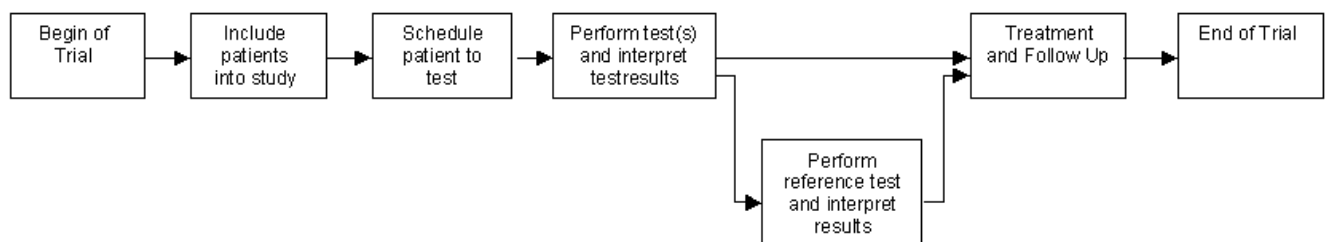


Figure 6: Basic process of simulation models

#### Inclusion of patients in the study

After arriving in the system only a certain percentage of patients turn out to be eligible for the study. In the literature this effect is called Lasagna's Law; As soon as one starts a study on a disease, the incidence of the disease appears to reduce compared to initial estimates [5], [8]. A proportion of the eligible patients will not participate in the study and leaves the system. After a predefined number of patients have entered the study, all new arriving patients are defined automatically to be non-participants. Non-participants immediately leave the system.

#### Schedule patient to test

When patients enter the study they are scheduled to the diagnostic imaging test. In practice, patients are not always available on the first free timeslot, for example, due to appointments for work. Therefore we take into account that there is a probability that a patient will reject the first available options to undergo the test. See Appendix A for an explanation of the usage of this probability.

After scheduling the patient according to their preferences the total waiting time can be determined. Subsequently, we will determine whether the patient withdraws from the study; long waiting times are assumed to have a negative effect on the willingness of the treating physician and the patient to participate in the study. The withdrawal of patients will be determined as a probability increasing linearly over time, till a maximum of 100%. Withdrawn patients immediately leave the system.

Most of the time the limitation of availability occurs for patients that have a job. Therefore we make a distinction between working people and not-working people. Hereby we assume that

patients without a job are always available on every offered timeslot. Only for the patients with a job we assume that they might have other obligations on the first available timeslots.

### **Perform test(s) and interpret test results**

Next, the patients will undergo the test on the scheduled appointment day and time. After the test is performed, the images need to be reconstructed and are then evaluated by the specialist. The interpretation of the test results includes scoring and reporting.

For the evaluation of the test results for a study the test interpreters need to score and report the aspects of the disease in detail. The interpreter normally reserves a part of the day in the week (morning or afternoon) for the evaluation of the test results for the study. During this time the interpreter will evaluate as much test results as possible. The total number of evaluated test results depends on the total evaluation time that is needed per test result. It is possible that the interpreter is not able to evaluate all test results on the same day; the remaining test results will be evaluated on another day.

### **Perform reference test and interpret test results**

The test result of the diagnostic imaging test is in several cases not clear enough. This means that an extra reference test is needed before the specialist can diagnose the patient. Since the reference test is performed in order of the specialists' request, we assume that the patients will undergo the reference test based on First Come First Serve (FCFS). After performing the reference test, the test result will be evaluated.

### **Treatment and Follow-up**

After the interpretation of the initial test results, follow-up is started. During follow-up patients undergo the advised treatment and are examined in several examination rounds. All patients will be followed during the follow-up period even if the reference test is performed.

The patient will not leave the system until the follow-up time has expired. When the patients test results are negative, which might indicate that the patient is healthy, he or she still participates in the follow-up. During the examination rounds the researcher is able to find possible false-negative values of the test results, which says something about the performance of the diagnostic imaging test. If the patient with the (false-) negative test result is excluded from the follow-up, the performance will be overestimated; therefore all patients undergo the follow-up.

When the pre-defined number of patients has reached the end of the follow-up period, the study is finished.

The treatment and follow-up period of our simulation model is implemented as a deterministic process. The treatment and follow-up period is very specific for each type of disease, which makes it hard to implement it in detail for a simulation model that should be general applicable. To give a complete overview of the total process of the study design we decided to include the treatment and follow-up period in the model.

### 3.3.2 Assumptions

The arrival of patients in the system is modeled as a stochastic Poisson process. The patients are scheduled to undergo a test at least one day after arrival with a maximum of one year in advance. Long waiting times are assumed to have a negative effect on the willingness to participate in the study. Non-participants will immediately leave the system.

We assume that patients might not show up for their appointment. In case of a no-show the reserved timeslot cannot be used for another patient that participates in the study. Therefore we have an unused timeslot the moment a patient does not show up. These no-shows influence the utilization of the test modalities, the duration of the study and the waiting times per patient. Whether a patient shows up for an appointment depends on a probability that is given by the end-user. We assume that no-shows occur at most one time per patient with a certain probability. In practice it might occur that the same patient does not show up numerous times. Because it is rather difficult to implement the possibility of not showing up for the second time we decided to assume that a patient does not show up at most for one time.

In our simulation models we assume that no patients will drop out of the study once they have undergone the diagnostic imaging tests. In practice, it might happen that a patient will leave the study before it has ended, for example due to death or due to clinical decisions concerning the treatment of the patient. When this happens during the treatment and follow-up period of the study no new patients will be included, because the inclusion period has been ended at that time. The researcher will use the gathered information of the rest of the participating patients for further research.

Because the effect of dropouts during the treatment and follow-up period of the study is minimal, we assume that no patients will drop out of the study once they underwent the diagnostic imaging tests.

Next we assume that patients always arrive exactly on their appointment time. Each timeslot starts and ends exactly at the given start and end time, the scan time is constant per test and none of the visits will extend beyond the available time. The time to evaluate a test result is fixed, but might be different per imaging test.

The interpreter that evaluates the test results may once in a while be absent due to other obligations, like a congress, seminar or meeting. The absence of the interpreter is modeled as a stochastic Poisson process.

The patients will undergo the reference test in order of their arrival, First Come, First Served (FCFS). All patients are treated equally; there is no difference between patients that are hospitalized (inpatients) or patients that are not hospitalized (outpatients). The Follow-up starts when the (last) test result is evaluated. Patients will leave the system as soon as the follow-up ends.

All simulation models are user-friendly, which means that the end user is able to perform sensitivity analysis by changing several parameter values. The end-user is able to modify the following input variables at the start of each simulation:

1. Sample size of the study
2. Number of patients expected to arrive per week
3. The length of the follow-up period
4. The percentage of patients that is eligible for the study
5. The percentage of the eligible patients that will participate in the study
6. The percentage of participating patients that should undergo the reference test
7. The probability that denotes the availability of the patients
8. The probability that denotes the willingness of the patients to participate in the study given the waiting time
9. The maximum waiting time
10. The percentage of the patients that have a job
11. The shift properties of the test interpreters, like:
  - The time per day that the interpreter has time to interpret the test results
  - The days in the week that the interpreter is available for the evaluation of the test results
  - The time that is needed to evaluate one test result
  - The absence of the interpreter; the user is able to enter the frequency and the mean number of days that an interpreter is unavailable for the study.
12. The timeslot properties for all diagnostic imaging tests, like:
  - The time per day that the test modality is available for the study
  - The time that is needed to perform one single test
  - The total number of timeslots per week
  - The days in the week that the test modality is available for the study

Appendix B contains screenshots of the input screens of the Cohort simulation model.

### 3.3.3 Outcome Measures

While performing the simulation we collected various output data. Table 2 gives an overview of the collected outcome measures. For each performance measure, the average values, the standard deviation and the minimum and maximum were collected.

Measure	Explanation
Duration of the study	The time between the start and end of the study measured in weeks
Duration of the inclusion period	The time between the start of the study and the last test performance measured in weeks
Utilization per test	The usage of the test modality as a percentage of the total time available
Utilization per test interpreter	The percentage of the planned time available that the physician spends on the study
Used time of the interpreter in hours	The total time in hours that the interpreter participates in the study.
Number of withdrawn patients	The total number of patients that withdraw from the study because of the long waiting times
Number of patients that underwent the reference test	The number of patients that underwent the reference test
Waiting time per patient per test	The time between the arrival in the system and the scheduled appointment
Number of unused timeslots per test	The total number of timeslots that were not filled during the study

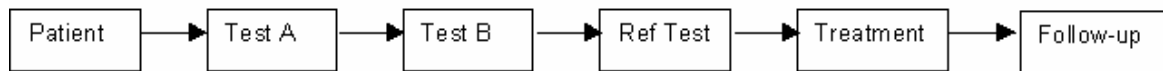
**Table 2: Outcome measures Arena simulation models**

The difference between the four different types of study designs is mainly caused by the number of tests each patient will undergo, the assignment of patients to the tests and the assignment of test results to the treating physician. In the next subsections we will discuss the specific assumptions for each study design individually.



### 3.3.4 Cohort simulation model

As we already saw in Section 2.2, all participating patients will undergo two diagnostic imaging tests and a reference test when the Cohort study design is used for the study.



**Figure 7: Scheme of Cohort study design**

The only difference between the Cohort model and the basic process described in Section 3.3.1 is the order in which the tests will be performed.

There are several options for the assignment of patients to the two diagnostic imaging tests:

1. The sequence of the imaging tests is fixed, the patients will always undergo test A first, followed by test B
2. The patients will be assigned randomly to the imaging tests, using block randomization
3. The first imaging test will be assigned alternately

According to the first option, patients will always undergo test A first and when this test is performed they will undergo test B.

Using block randomization, the diagnostic imaging tests are assigned to the patients using a randomization scheme. This randomization scheme is a chance procedure that assigns tests to patients. The randomization scheme determines the first test the patient will be assigned to. The sequence of which of the two imaging tests will be assigned to the patients will be determined for each patient individually and will be based on the chronological time in which patients enter the system. Neither the researcher nor the patient knows the first imaging test to be assigned at the time a patient is included in the study.

The block size indicates after which number of patients the allocation of the first test should be equal. For example, if we have a block randomization scheme with a block size of 6 it means that after each 6 included patients the total patients assigned to each tests should be equal. Thus after 6 patients, 3 of them are first assigned to test A and 3 to test B, after 12 patients, 6 patients will undergo test A first and 6 of them will first undergo test B, etcetera.

The block size is variable and can be changed by the end user. For a more detailed description of block randomization we refer to M. Zelen [10].

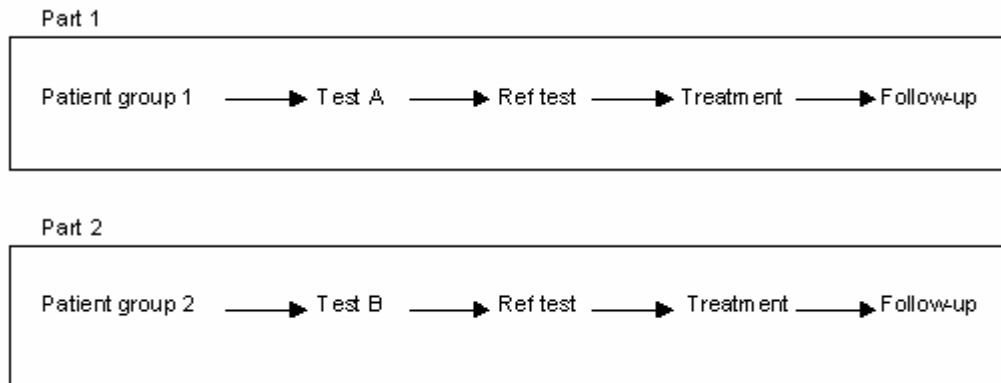
Finally we can assign patients alternately to the tests, which means that the first participating patient will undergo test A and after that test B, the second participating patient will then first undergo test B followed by test A, the third participating patient will then again be assigned to test A followed by test B, and so on.

The different options of assigning patients to tests will give us insight in how we can optimize the utilization of the diagnostic imaging modalities.

Because patients undergo two tests in this study design, the decision whether the patient will withdraw from the study depends on the total waiting time till the day the second test will be performed.

### 3.3.5 Before – After simulation model

As presented in Section 2.3 the Before-After study design can be globally represented by the next figure.



**Figure 8: Scheme of the Before-After study design**

The figure shows us that a study that is set up according to the Before-After study design can be divided into two parts. In the first part of the study the outcomes of the performance of the currently used imaging test are examined. In the second part the outcomes of the performance of a new diagnostic imaging test are investigated. The two parts of the study are carried out at different moments in time; the second part occurs later in time than the first part, it will not start before the inclusion period of the first part of the study is ended. For this reason we can represent each part of the Before-After study design as a basic process that is described in Section 3.3.1.

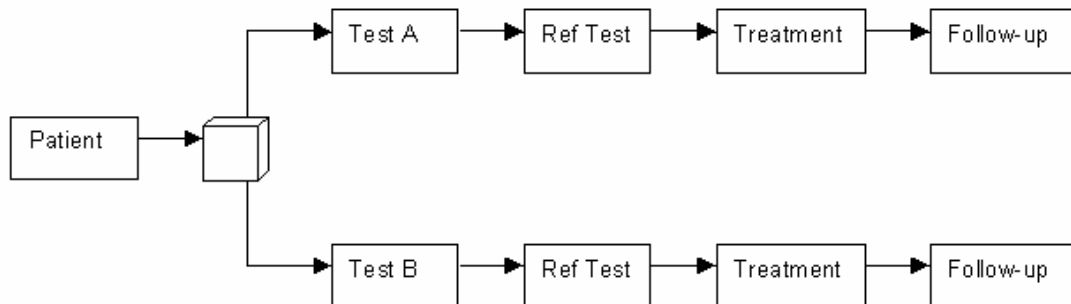
The final Before-After simulation model consists of two times the basic process; each process indicates one part of the study. The simulation starts with the first part of the Before-After model. The second part of the study might begin when the inclusion period of the first part is ended; this is when the pre-defined number of patients (sample size) is included in the study. Thus the start of the second part of the study is not earlier than the end of the inclusion period of the first part. The inclusion periods of the different parts do not overlap in time. The follow-up periods of the different parts might partially overlap during the study, this depends on the total length of the follow-up. The follow-up period of the first part and the inclusion period of the second part might also overlap; this depends as well on the length of the follow-up period. It depends on the settings of the study whether the two parts overlap in time.

Because the second part of the study occurs at a different time than the first part it is very likely that the arriving and participating characteristics of the patients change. Therefore, the end user is able to modify extra parameter values for each part of the study; we refer to the user manual of the Before-After simulation model for more details.

Because this study design consists of two parts, there are a few specific characteristics of each part that might be interesting for the researcher, like the duration of each part separately, instead of the duration of the whole study. Therefore we implemented extra outcome measures

like the duration of each part and the inclusion periods per part of the study. These extra outcome measures do not influence the other outcome measures or the cost calculations.

### 3.3.6 RCT simulation model with randomization of all patients across diagnostic testing strategies



**Figure 9: Scheme of RCT with randomization of all patients across diagnostic testing strategies**

In a trial using this study design (lets call it RCT-1 for simplicity) each patient undergoes one test; test A or test B. The assignment of the patients to one of the tests occurs according to block randomization. A randomization scheme determines for each patient individually which test they should undergo. The randomization scheme is formulated using the block randomization concept with a pre-specified block size, just like the randomization scheme that is used in the Cohort simulation model.

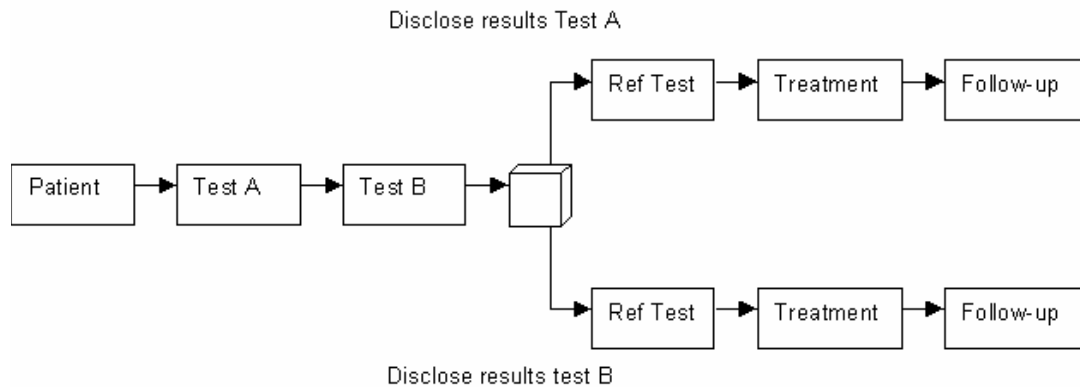
In practice it is essential that all participating employees of the study are blinded for the randomization scheme that assigns the patients to the tests. Before a patient is included in the study the participating employees do not know to which test the particular patient will be assigned. The moment that the patient agrees to participate in the study and it is time to make an appointment for one of the tests, the researcher gets to know to which test the patient will be assigned. Until that moment nobody of the study (including the patient) knows which test will be performed.

It might occur that a patient insists on undergoing one of the tests, for example test A. When this occurs, the patient cannot be included in the study anymore since the patient decided beforehand which test to undergo, so the allocation of this particular patient to the test will not be random.

In our simulation model of the RCT-1 study design we assume that the percentage participating patients covers the occurrence of excluding patients based on their preference to undergo one of the two tests. The percentage participating patients denotes the percentage of eligible patients that is willing to participate in the study. As soon as a patient has a preference for one of the two tests we assume that the patient is not willing to participate in the study anymore and will therefore not be included in the study.

The randomization of patients across the diagnostic testing modalities is the only feature of the RCT-1 simulation that differs from the basic process as it is described in Section 3.3.1

### 3.3.7 RCT simulation model with randomization between providing the results of Test A versus providing results of Test B



**Figure 10: Scheme of RCT with randomization between providing the results of test A versus providing results of test B**

In a trial organized through this RCT design (lets call it RCT-2 for simplicity), all patients undergo all tests. Just like the Cohort study design there are several options for the assignment of patients to the two diagnostic imaging tests:

1. The sequence of the imaging tests is fixed, the patients will always undergo test A first, followed by test B
2. The patients will be assigned randomly to the imaging tests, using block randomization
3. The first imaging test will be assigned alternately

After a patient undergoes both diagnostic imaging tests, the test results will be evaluated. Next, a randomization scheme determines which test result will be provided to the treating physician. Consequently, the physician is blinded for one of the test results. The physician will subsequently base the treatment of the patient on the provided test result.

The randomization between providing the results of test A versus providing the results of test B is based on a block randomization scheme as described in Section 3.3.4.

### **3.4 Verification and Validation**

Model verification and validation are the two steps of determining whether a model is a good representation of the actual system. Model verification is intended to ensure that the model does what it should do; it is like debugging the computer program. Model validation is used to demonstrate that the model is a good representation of the system, for example by comparing the outcome measures of the model with data obtained from reality.

To verify our simulation models we used various techniques like debugging, step-by-step analysis (structured walk-through), animation and tracing. The key idea of these methods is that the developer will look at the model from a different perspective and may therefore discover problems with the implementation of the model. The animation and tracing techniques are most useful for isolating an error, while debugging and structured walkthrough are mainly used to find out whether there is a problem in the model. For a complete description of these methods see Hillston [1] and Law & Kelton [6].

To validate the simulation models during the development we discussed them several times during the weekly meeting with the members of the Assessment of Radiological Technology (ART) group of the Department of Epidemiology and Biostatistics and Department of Radiology at the Erasmus Medical Center Rotterdam. Furthermore the models were discussed numerous times with the supervisor from the Vrije Universiteit in Amsterdam.

Once we finished the development of the models we validated the outcomes using clinical studies performed in the past by the ART group by R. Ouwendijk and M.C.J.M Kock. In the next subsections we will give a detailed description of the validation process for these two clinical studies.

#### **3.4.1 Verification and Validation Living Renal Donors study**

The first clinical study we used for the validation of our simulation model is the study *“Contrast-Enhanced MR Angiography and Digital Subtraction Angiography in Living Renal Donors: Diagnostic Agreement, Impact on Decision Making and Costs”* performed by M.C.J.M. Kock [2].

The purpose of this study was to evaluate the diagnostic agreement, the impact on decision making, and the costs from the hospital’s perspective of contrast-enhanced MR angiography (MRA) and digital subtraction angiography (DSA) in the preoperative diagnostic imaging of potential living renal donors.

The study was organized through a Cohort study design. The two tests that were compared during this study were the MRA and the DSA. In the period May 2000 through September 2001, all potential living renal donors were informed about the study and asked to participate. All participating patients underwent a MRA and DSA scan. Separate evaluations of the two scan results took place for decision making. First the decision of which kidney should be used for transplantation was based on the results of the MRA scans. At least two months later the reviewers based their decision on which kidney to use for transplantation on the results of the DSA scan. The reviewers were blinded at that time for the findings and outcomes of the MRA scans.

The study can be simulated using the Cohort simulation model. We were able to obtain the following data from the article.

In the period from May 2000 till September 2001 52 potential living renal donors were informed about the study, 42 of them were enrolled in the study. The sample size is therefore equal to 42.

The percentage of eligible patients was set to 100 percent. The number of patients per week is set to 0.702, which is equal to the mean number of patients per week when 52 patients are approached during the 17 months of the inclusion period. Nine of the 52 potential donors did not participate in the study because they refused to undergo one of the tests or because of the canceling of the transplantation. One patient is not included in the study because the equipment turned out to be defective at the appointment time. Rescheduling the examination was impractical for the donor. This results into a percentage of participating patients that is equal to 83 percent.

The MRA scan (Test A) was performed on Wednesday from 10:00 till 10:45. It took 45 minutes to perform this test and there was one timeslot per week reserved for the study. Every Friday a radiologist evaluated the test results. The mean evaluation time of the test results was 30 minutes per test.

The DSA scan (Test B) was performed on Thursday from 15:00 till 16:00. It took 60 minutes to perform a DSA scan and one timeslot per week was reserved for the study. Every Tuesday a radiologist evaluated the test results. The mean evaluation time of the test results took 5 minutes per test.

The entire donor procedure took several months to complete; therefore we assume that it is not very common that potential donors will withdraw from the study when the total waiting time for the tests is less than six months. Once the total waiting time will be longer than six months patients start to withdraw from the study;  $33\frac{1}{3}$  percent of the eligible patients will not participate when the waiting time exceeds six months. The maximum waiting time is set at one year. The percentage of eligible patients that will not participate in the study increases linearly over time till the maximum waiting time of one year is reached.

We let the Cohort simulation model run for 1000 replications, which means that we simulated the clinical study for 1000 times.

Chapter 4 presents the results of the validation of the Cohort simulation model. This chapter shows the results obtained by performing extra analysis for the Living Renal Donors study as well. In the next subsection we will first discuss the validation of the RCT-1 simulation model for another clinical study, the DIPAD trial.

### 3.4.2 Verification and Validation DIPAD trial

In this section we will discuss the validation process of the RCT-1 simulation model using the data from the DIPAD trial. The data was obtained from the article *"Imaging Peripheral Arterial Disease: A Randomized Controlled Trial Comparing Contrast-enhanced MR Angiography and Multi-Detector Row CT Angiography"* written by R. Ouwendijk [9] and from the results of the internship project of A. Hutzschenreuter [3]

The objective of the DIPAD trial was to evaluate clinical utility, patient outcomes and costs of MRA compared to CTA for initial imaging in the diagnostic work-up of patients with peripheral arterial disease (PAD). The trial was performed to evaluate two diagnostic strategies, either of which could become routine clinical practice.

The trial was organized through the RCT-1 study design; patients were randomized across the MRA or CTA. The inclusion period of the trial took place from December 2001 to September 2003. During this period 264 patients were assessed for eligibility, 103 patients were excluded because they did not fulfill inclusion criteria. This results in an eligibility percentage of 61%. Four of the eligible patients refused to participate. The percentage of participating patients is therefore 97.5%. Finally, 157 patients enrolled in the study, so the sample size is set to 157.

The number of patients per week that could be eligible for the study is set to 2.8, which is equal to the mean number of patients per week when 264 patients are approached during the 22 months of the inclusion period.

The CTA (Test A) was performed on Tuesday from 10:00 till 10:30. It took 15 minutes to perform this test and there were two timeslots per week reserved for the study. Every Tuesday the test results were evaluated during the vascular meeting. The mean evaluation time of the test results was 30 minutes per test.

The MRA (Test B) was performed on Wednesday from 14:00 till 15:30. It took 45 minutes to perform a MRA scan and two timeslots per week were reserved for the study. Every Tuesday the test results were evaluated during the vascular meeting. The mean evaluation time of the test results took 15 minutes per test.

In 8 % of the cases a reference test was performed. The reference test for this trial was the DSA. There were four timeslots a week available for the study. The reference test could be performed on Monday, Wednesday till Friday from 10:00 till 11:00 and took in total 60 minutes per test.

Because PAD is a disease that typically occurs in elderly people we set the percentage of working people at zero percent; it is not very likely that the elderly are not available for the free timeslots. If the waiting time for the diagnostic imaging test exceeds six weeks patients start to withdraw from the study, 30% of the patients will not participate. The maximum waiting time is set to three months.

We simulated the clinical trial 1000 times by running the RCT-1 simulation for 1000 replications. For the validation of the simulation models we mainly examined the expected

duration of the clinical study. Unfortunately there was no data available to validate several other output measures like the utilization of the test modalities and test interpreters. Therefore we presented these measures of the Living Renal Donors study and the DIPAD trial to the members of the ART group during the weekly meeting and they validated the output measures based on their experience for these specific clinical studies in the past.

The results of the validation of both clinical studies can be found in chapter 4, section 4.1. In section 3.5 we will first describe the extra analyses we performed for these clinical studies. Information about the analysis of another clinical study is provided as well.



### **3.5 Analysis**

In order to compare the various outcome measures listed in table 2 we decided to evaluate three different clinical studies. Two of the clinical studies are performed in the past. The data of these studies is used for the validation of the simulation models (Sections 3.4.1, 3.4.2). At that time we decided to perform extra analyses for these studies. The analyses we performed for these studies are discussed in more detail in Section 3.5.1 and 3.5.2.

Our main analysis is performed for a currently ongoing study at the Erasmus MC. Section 3.5.3 provides the information about the analysis of the Coronary CTA study.

#### **3.5.1 Analysis of Living Renal Donors Study**

The analysis of the Living Renal Donors study can be split in two cases. The first case evaluates the outcome measures of the study for each type of study design. The main focus was on the evaluation and comparison of the mean total costs and the mean total costs including a penalty for each unused timeslot. In this case we did not consider the influence of the behavior of patients and the influence of the absence of the interpreter. We assumed that each patient was available on the first offered timeslot of a test. We did not consider the possibility that a patient was not available on the offered timeslot due to obligations at work.

In the second case we considered the influence of the behavior of patients and the influence of the absence of the interpreter on the mean total costs for each type of study design.

Because 90 percent of the included patients had a job and the entire donor procedure took several months to complete it is very likely that there was a restriction on the availability of a patient when an appointment had to be made. For this reason we performed extra simulation runs for the situation where patients have a limited availability in the first two weeks after calling for an appointment. We assumed that 50% of the included working patients was available to undergo a test on the first day after arriving in the system and that patients were available with a chance of 100% to undergo a test when the appointment was 14 days after arriving in the system. This means that a patient with a job is available with a probability of 0.75 if the offered appointment day is seven days after the first time this patient contacted the hospital.

We considered the possibility that a patient might not show up for an appointment as well. We assumed that 15% of the patients did not show up for their appointment.

Finally we considered the effect on the total costs of the study when the test interpreter was absent for evaluating the test results. We considered the case where the test interpreter was not present for about once per month.

For the simulation of these analyses we used the input parameters as denoted in Section 3.4.1 and as mentioned in the description above.

### 3.5.2 Analysis of DIPAD Trial

The second study we evaluated was based on the data of the DIPAD trial. The DIPAD trial differs somewhat of the Living Renal Donors trial, the detailed information about the trial and the values of the input parameters can be found in Section 3.4.2.

For the DIPAD trial we performed a similar kind of analysis as we did for the Living Renal Donors study. First we evaluated the mean total costs and the mean total costs including a penalty for each unused timeslots for each type of study design.

Next, we also performed extra simulation runs concerning the behavior of patients and the absence of the test interpreters. We performed two extra simulation runs. In the first run of 1000 replications we examined the effect on the total costs when 15% of the patients did not show up for their appointment. In the second run of 1000 replications we analyzed the influence of the absence of the test interpreters. We assumed that each test interpreter was not available once a month for about a week. We did not perform extra simulation runs to examine the effect of the limited availability of the patients on the mean total costs. Because PAD is a disease that typically occurs in elderly people we assumed that none of the participating patients had a job and that it is therefore not very likely that the patients were unavailable on the free timeslots due to obligations at work.

### 3.5.3 Analysis of an ongoing study: the Coronary CTA Study

The objective of the Coronary CTA study is to evaluate the costs, effectiveness, and cost-effectiveness of replacing diagnostic invasive coronary angiography with non-invasive coronary angiography using multisection-CT (MS-CT) in patients with stable or unstable angina pectoris. Angina pectoris is a term for chest pain caused by an inadequate supply of blood and oxygen to the heart.

The study is set up according to the Cohort study design and is multi-centric. For the analysis we will only look at the part that is performed at the Erasmus MC. At the Erasmus MC 120 patients will be enrolled. About 12 patients per week might be eligible for the study. The researchers estimate that about 16.7% (2 of the 12) patients are really eligible for the study. The expectation is that 95% of the eligible patients will participate in the study.

The test modalities that are used for this study are especially purchased for research purposes. Therefore there are no timeslot restrictions, a patient can undergo a test at any time the patient would like it. Because it is not very common that there are not any timeslot restrictions, we will assume as baseline that each test has two timeslots on Tuesday morning.

It will take 15 minutes to perform a CTA (Test A). The mean evaluation time of the test result is 30 minutes per test. The time to perform a DSA (Test B) is 60 minutes. The evaluation of the test result can be done in 5 minutes.

Angina pectoris is a disease that mainly occurs in men around the age of 55. As baseline we assume that 25% of the working patients is not available on a timeslot that is one day after contacting the hospital. All patients are available with a probability of 1 when the offered appointment day is more than two weeks after contacting the hospital for the first time.

The maximum waiting time for the diagnostic imaging tests is set to 6 weeks because of the serious nature of the disease. When the total waiting time for the patients exceeds 3 weeks 30% of the patients will start to withdraw from the study. The percentage of patients that will withdraw from the study will increase linearly over time till a maximum of 100% when the waiting time is 6 weeks.

Finally we assume as baseline that 15% of the patient will not show up at their appointment and we also assume that the test interpreters are not available for the study about once a month for approximately seven days.

The performed simulation analysis can be divided into three parts. First we analyzed different simulation scenarios for the Cohort study design, next we analyzed the influence of the number of timeslots for the baseline scenario for the Cohort study design and finally we evaluated the baseline scenario for each type of study design.

#### **Varying the simulation scenarios for the Cohort study design**

In the baseline situation we assume that patient availability is limited during the first two weeks after contacting the hospital, that the willingness to participate decreases when the total waiting time exceeds three weeks, that 15% of the patients will not show up at their appointment and that the test interpreters are absent about once a month. The assumptions we made are typical problems that occur in clinical practice. In our analysis we analyzed the influence on the outcome measures when each of these types of problems are solved. The ideal situation, that is that none of these problems occur, is analyzed as well.

The simulations were performed for the situation that the study is set up according to the Cohort study design. We evaluated the mean utilization of the test modalities, the mean number of patients that withdraws during the study, the mean duration of the study and the mean total costs of the study.

#### **Varying the number of timeslots of the baseline scenario for the Cohort study design**

In the second simulation scenario for the Coronary CTA study we evaluated the outcome measures for the baseline scenario for varying number of timeslots per test per week. For different number of available timeslots per test per week we evaluated the mean utilization of the test modalities, the mean number of patients that withdraws from the study, the mean duration of the study and the mean total costs. We only evaluated the outcome measures for the case where the study is organized through the Cohort study design.

#### **Varying the type of study design for the baseline scenario**

In the final simulation analysis for the Coronary CTA study we evaluated the outcome measures for the baseline scenario for the four different types of study designs. For each type of study design we evaluated the mean utilization of the test modalities, the mean number of withdrawn patients, the mean duration of the study and the mean total costs.

## 4 Results

In this chapter we will first present the results of the validation of the simulation models. These results can be found in Section 4.1. The results of the analysis of the Living Renal Donors study and the DIPAD trial are provided in Sections 4.2 and 4.3. Section 4.4 presents the results of the analysis of the Coronary CTA study.

### 4.1 Results of Validation

For the validation of our simulation models we used two clinical studies that are performed in the past by the ART group. We described the validation methods in detail in section 3.4.1 and 3.4.2.

#### Living Renal Donors study

For the validation of the simulation model we looked at the mean duration of the inclusion period and the mean total duration of the study. The inclusion period of the Living Renal Donors study was 17 months ( $73\frac{2}{3}$  weeks) and the total duration of the study was  $82\frac{1}{3}$  weeks. This information is obtained from the article about the study [2].

After simulating the clinical study for 1000 times using the Cohort simulation model we found a mean duration of the inclusion period of 71.4 weeks and a mean duration of the study of the simulation runs of 80.8 weeks (standard deviation is 10.8).

A simulation model is a good representation of reality if the values obtained from reality does not differ more than two times the standard deviation of the calculated mean. In this case the calculated mean value is 80.8. The real value for the duration of the trial should be in the interval  $80.8 \pm 2 \cdot 10.8$ . We see that the real duration of the study was  $82\frac{1}{3}$ , which is a value in the given interval. For this reason we can accept the model as a good representation of the real system.

#### DIPAD Trial

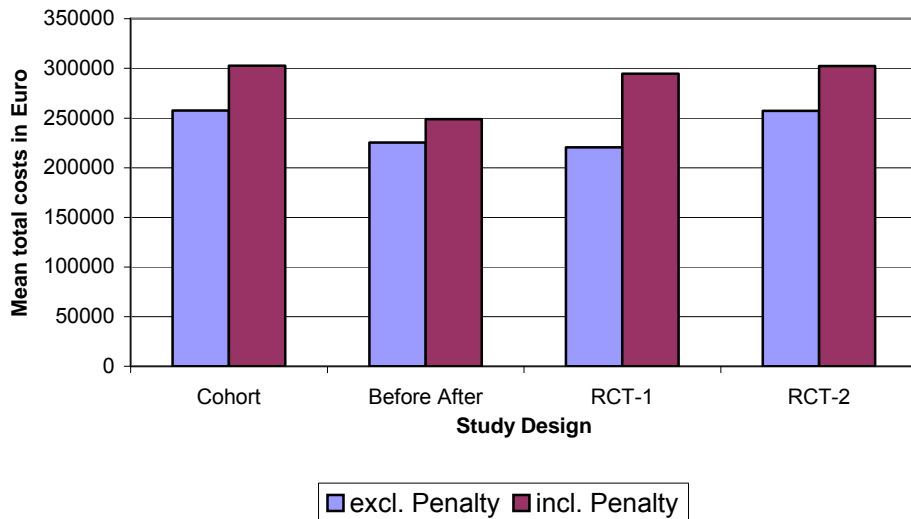
To validate the simulation model using the data from the DIPAD trial we also looked at the mean duration of the inclusion period and the mean total duration of the trial. The duration of the inclusion period of the DIPAD trial was 22 months ( $95\frac{2}{3}$  weeks) and the total duration if the trial was  $121\frac{1}{3}$  weeks [9].

We simulated the clinical trial 1000 times by running the RCT-1 simulation for 1000 replications. The mean duration of the inclusion period as it was obtained by the simulation run is 94.3 weeks. The mean duration of the study of the simulation runs is 120.7 weeks the standard deviation is 7.3.

The values for the duration of the inclusion period and the total duration of the trial are both in the interval  $94.3 \pm 2 \cdot 7.3$  and  $120.7 \pm 2 \cdot 7.3$ . Therefore we are allowed to assume that the simulation model is a good representation of the system.

## 4.2 Results of analysis Living Renal Donors study

For the Living Renal Donors study we first evaluated the mean total costs and the mean total costs including a penalty for the different types of study design. In Figure 11 both the mean total costs for the study excluding and including the penalty for the unused timeslots are shown.



**Figure 11: Mean total costs in euro for the living renal donor study when organized through different types of study designs.**

The mean total costs of the Living Renal Donors study, which is organized through a Cohort study design, are about 257,000 Euro. The mean total costs of the study can be divided into the total costs for the test modalities, the total costs for support of the study and the personnel costs. The personnel costs cover about 71.3% of the mean total costs; the costs for the test modalities 23.7% and the cost for support of the study by the trial bureau cover about 5.0%. The personnel costs include the costs for the researcher, principal investigator, and the test interpreter. The costs for the test modalities include the material costs, equipment costs, maintenance costs and the personnel costs that are made while performing a test.

The figure shows us that the mean total costs would have been almost equal to the total costs of the Cohort study if the study was organized through an RCT-2 study design. The difference would only have been a few hundred Euros.

The total costs for the study set up according to the Before-After and RCT-1 study design are almost equal as well; the RCT-1 study design is approximately 5,000 Euro less expensive than the Before-After study design. Both the RCT-1 study design and the Before-After study design are less expensive than the Cohort study design or RCT-2 study design. The difference between the mean total costs of these study designs is about 37,000 Euro.

The total number of tests a patient undergoes per study design causes the difference in mean total costs. Within the Cohort and RCT-2 study design each patient underwent two diagnostic imaging tests, while patients within the Before-After and RCT-1 study design only underwent one diagnostic imaging test.

	Cohort	Before After	RCT-1	RCT-2
<b>Utilization A</b>	59%	58%	30%	60%
<b>Utilization B</b>	59%	59%	30%	60%
<b>Mean total waiting time in days</b>	10	8	12	9
<b>Duration of study in weeks</b>	81	83	81	81
<b>Mean total costs</b>	€ 257,620	€ 225,431	€ 220,547	€ 257,367
<b>Mean total costs incl. Penalty</b>	€ 302,601	€ 248,994	€ 294,548	€ 302,348

**Table 3: Results of mean utilization, mean total waiting time, mean duration of the study and the mean total costs for different types of study designs for the Living Renal Donors study**

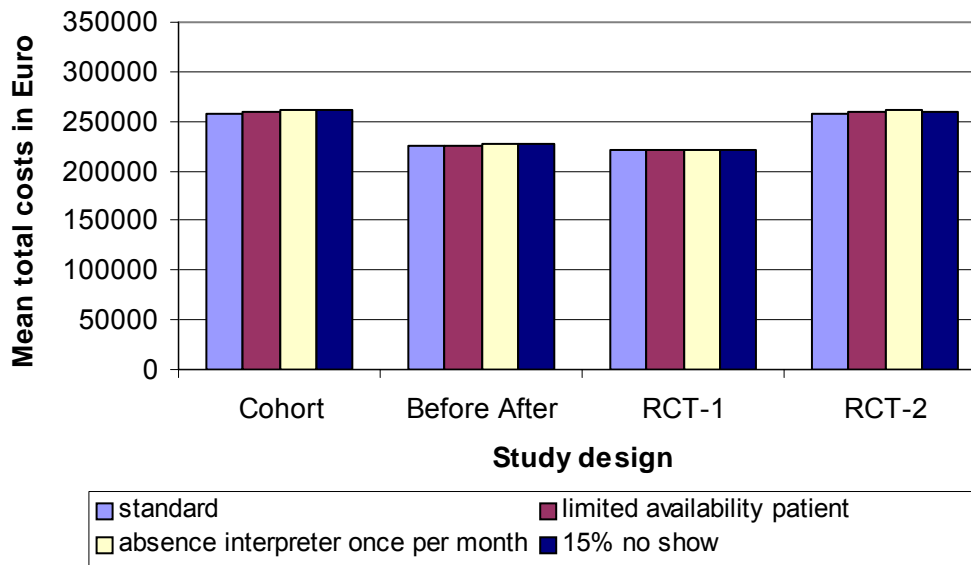
The mean duration of the study was almost stable for the type of study design that is being used. The study will take about 2 weeks longer when it is set-up according to the Before-After study design. For the other study designs the mean duration of the study will be around 81 weeks (Table 3). The extra time that is needed for the Before-After study design results in higher personnel costs. The personnel costs of the Before-After study design are about 3,000 Euro higher than the personnel costs of the Cohort and RCT-2 study design. However, due to the less diagnostic imaging tests that were performed in the Before-After study, the study design is still less expensive than the other two.

The utilization of the diagnostic test modalities of the Cohort, Before-After and RCT-2 study design is approximately 59%. The utilization of the test modalities of the RCT-1 study design is about 30%. This means that the total number of unused timeslots within the RCT-1 study design is much higher than the total number of unused timeslots for the other study designs. Unused timeslots involve great expenses for the hospital. Each time a timeslot is not used another patient could have been helped. Therefore we decided to charge the cost price for the test modalities each time a timeslot is not used. This resulted in the mean total costs including a penalty as displayed in Figure 11. As we can see in the figure, the mean total costs of the study set up according to the RCT-1 study design increases rapidly, but it still remains less expensive than the Cohort or RCT-2 study design.

Although the utilization of the tests of the RCT-1 study design is only 30%, the mean waiting time per patient is very high compared to the waiting times that occur with the other study designs. We would expect that the waiting time for patients in the study set up according to the RCT-1 study design would have been about 3 to 4 days, but not 12. We do not really have an explanation for this relative high waiting time.

The calculation of the mean total costs per study design in Figure 11 is based on the simulation runs using the standard input as presented in Section 3.4.1. The behavior of the patient based on their agenda and the possible absence of the interpreter due to conferences were not considered for these simulation runs.

In Figure 12 we can see that the behavior of the patient and the absence of the test interpreters does hardly influence the mean total costs for the study.



**Figure 12: Effect of limited availability of the patient, 15% no show and the absence of the test interpreter on the mean total costs per type of study design for the living renal donor study.**

If we look closely, we can find a small increase in the mean total costs as a result of a limited availability of the patients and the occurrence of no-shows. This increase is negligible when we look at the overall costs of the study.

We would have expected that the behavior of the patients and the absence of the test interpreters were of more influence. The fact that it hardly influences the mean total costs can be explained by the sample size of the study. For the living renal donors study only 42 patients were included. The sample size might be too small to find any association of the behavior of patients and the absence of the test interpreters on the mean total costs.

### 4.3 Results of analysis DIPAD trial

The DIPAD trial was evaluated in terms of total costs, total costs including a penalty for each unused timeslot, and the total costs if we include some uncertainty about the behavior of the patients and the presence of the test interpreters.

In Figure 13 we can see that the mean total costs of the DIPAD trial are much higher than the mean total costs for the Living Renal Donors study. The total number of patients that is included in the study causes this. The sample size of the Living Renal Donors study was 42, while the sample size of the DIPAD trial was 157.

Nevertheless we can see the same trend in Figure 13 as we saw in Figure 11; the mean total costs for a Cohort study design are nearly equal to the mean total costs for the RCT-2 study design. The mean total costs for the Before-After study design and the RCT-1 study design are again comparable. However, the difference between the total costs is for this scenario larger than it was for the previous scenario. The RCT-1 study design is 12,000 Euro less expensive than the Before-After study. The difference can be explained by the mean duration of the study and thus the extra personnel costs that were made. The mean duration of the trial set up according to the RCT-1 study design was 121 weeks, the trial set up according to the Before-After study design took 4 weeks longer; the mean duration was 125 weeks (table 4)

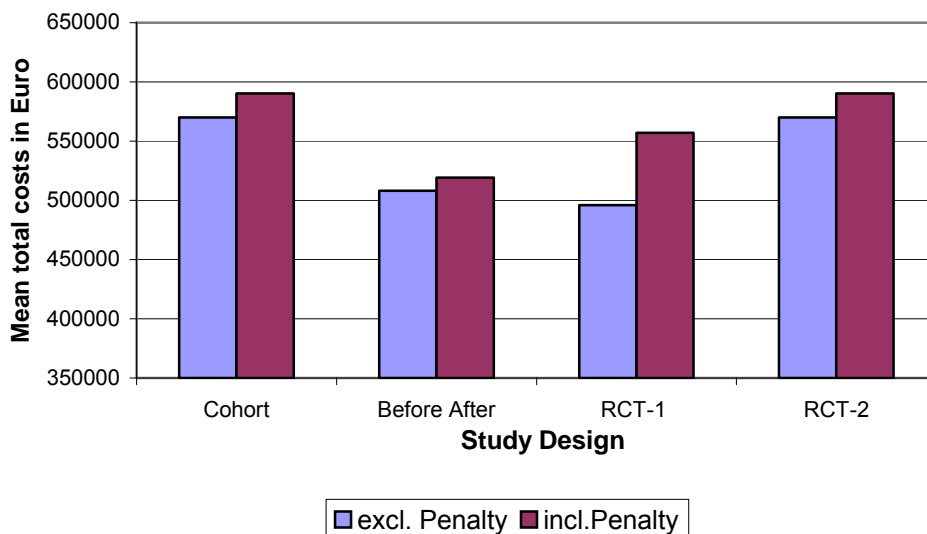


Figure 13: Mean total costs for the DIPAD trial set up according to different types of study designs.

The Cohort study design was again more expensive than the RCT-1 study design and the Before-After study design. The difference between the mean total costs of the Cohort study design and the RCT-1 study design is about 74,000 Euro. The proportion of the mean total costs of the Cohort study design and the RCT-1 study design are almost equal for the DIPAD trial and the Living Renal Donors study. For the Living Renal Donors study the mean total costs of the Cohort are 16.8% higher than the mean total costs of the RCT-1 study design, for the DIPAD trial they are 14.9% higher.



	Cohort	Before After	RCT-1	RCT-2
<b>Utilization A</b>	82%	80%	42%	82%
<b>Utilization B</b>	82%	81%	42%	82%
<b>Mean total waiting time in days</b>	12	10	4	12
<b>Duration of study in weeks</b>	123	125	121	123
<b>Mean total costs</b>	€ 569,986	€ 508,039	€ 495,989	€ 569,998
<b>Mean total costs incl. Penalty</b>	€ 590,299	€ 519,221	€ 556,928	€ 590,311

Table 4: Results of mean utilization, mean total waiting time, mean duration of the study and the mean total costs for different types of study designs for the DIPAD trial.

The timeslots of the CTA and the MRA for the Cohort, Before-After and RCT-2 study design are approximately 82% of the time filled (Table 4). Because of the high utilization of the test modalities the mean total costs did not increase as much as the costs for the Living Renal Donors study when a penalty is counted.

Within the RCT-1 study design the timeslots for the diagnostic imaging tests were only filled for 42% of the time. The low utilization of CTA and MRA causes a considerable increase of the mean total costs when a penalty for the unused timeslots is counted. But the RCT-1 study design is still less expensive than the Cohort and the RCT-2 study design.

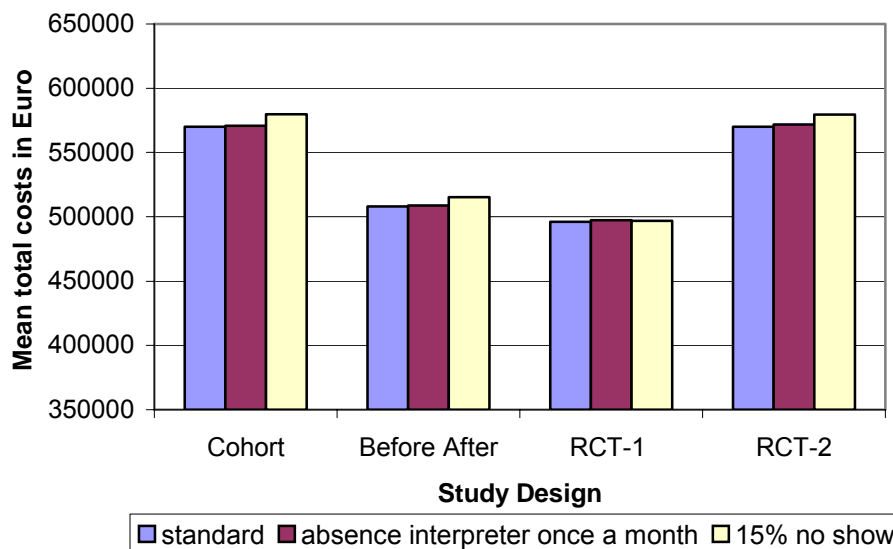


Figure 14: Effect of 15% no show and the absence of the test interpreters on the mean total costs per type of study design for the DIPAD trial.

The absence of the test interpreters hardly influences the mean total costs of the trial (Fig.14). The increase of the mean total costs is in the range of 0.14% to 0.31%, which is totally negligible with overall costs of the trial.

The number of no shows that occur during a trial has a small effect on the mean total costs of the trial. For the Cohort study design and the RCT-2 study design the mean total costs increased with 1.7% ( $\approx 10,000$  Euro). The mean total costs for the Before-After study design increased with 1.4% ( $\approx 7,000$  Euro) and the costs for the RCT-1 study design increased with about 0.18% ( $\approx 1,000$  Euro).

For the DIPAD trial and the Living Renal Donors study we mainly examined the effect of changes in the input parameters on the mean total costs of the trial. In the next section we will discuss the results of the analysis of the influence of the number of timeslots, the availability of the test interpreters and the behavior of patients on the utilization of the diagnostic imaging tests, the total mean waiting time per patient, the number of withdrawn patients and the total duration of the trial.

#### 4.4 Results of analysis Coronary CTA study

The last study we evaluated using our simulation models is the Coronary CTA study that is currently going on at the Erasmus MC.

As already mentioned in Section 3.5.3 we can divide the performed simulation analysis into three parts. In our first analysis we varied the simulation scenarios for the Cohort study design, in the second we analyzed the influence of the number of timeslots for the baseline scenario for the Cohort study design and finally, in the third, we evaluated the influence of varying the type of study design for the baseline scenario.

For each of the three analyses we evaluated the mean utilization of the test modalities, the mean number of patients that withdraws during the study, the mean duration of the study and the mean total costs of the study.

##### Varying the simulation scenarios for the Cohort study design

In the baseline situation we assume that patient availability is limited during the first two weeks after contacting the hospital, that the willingness to participate decreases when the total waiting time exceeds three weeks, that 15% of the patients will not show up at their appointment and that the test interpreters are absent about once a month.

In our analysis we analyzed the influence on the outcome measures when each of these types of problems are solved. The ideal situation, that is that none of these problems occur, is analyzed as well.

The simulations were performed for the situation that the study is set up according to the Cohort study design.

Table 5 shows the numerical results of this simulation analysis. Figure 15 gives the graphical representation of the results.

	Most ideal situation	Baseline	Baseline excl. limited avail. patient	Baseline excl. 15% no show	Baseline excl. absence interpreter	Baseline excl. departure patient
Utilization A	89%	75%	77%	82%	75%	81%
Utilization B	88%	74%	76%	81%	74%	80%
Total mean waiting time in days	21	19	18	17	19	30
Nr of Withdrawn Patients	3	27	23	16	27	13
Duration of trial in weeks	68	81	79	74	81	75
Mean total costs	€ 341,488	€ 370,486	€ 365,206	€ 355,096	€ 369,792	€ 356,737

Table 5: Numerical results for different simulation scenarios for the Coronary CTA study that is set up according to the Cohort study design

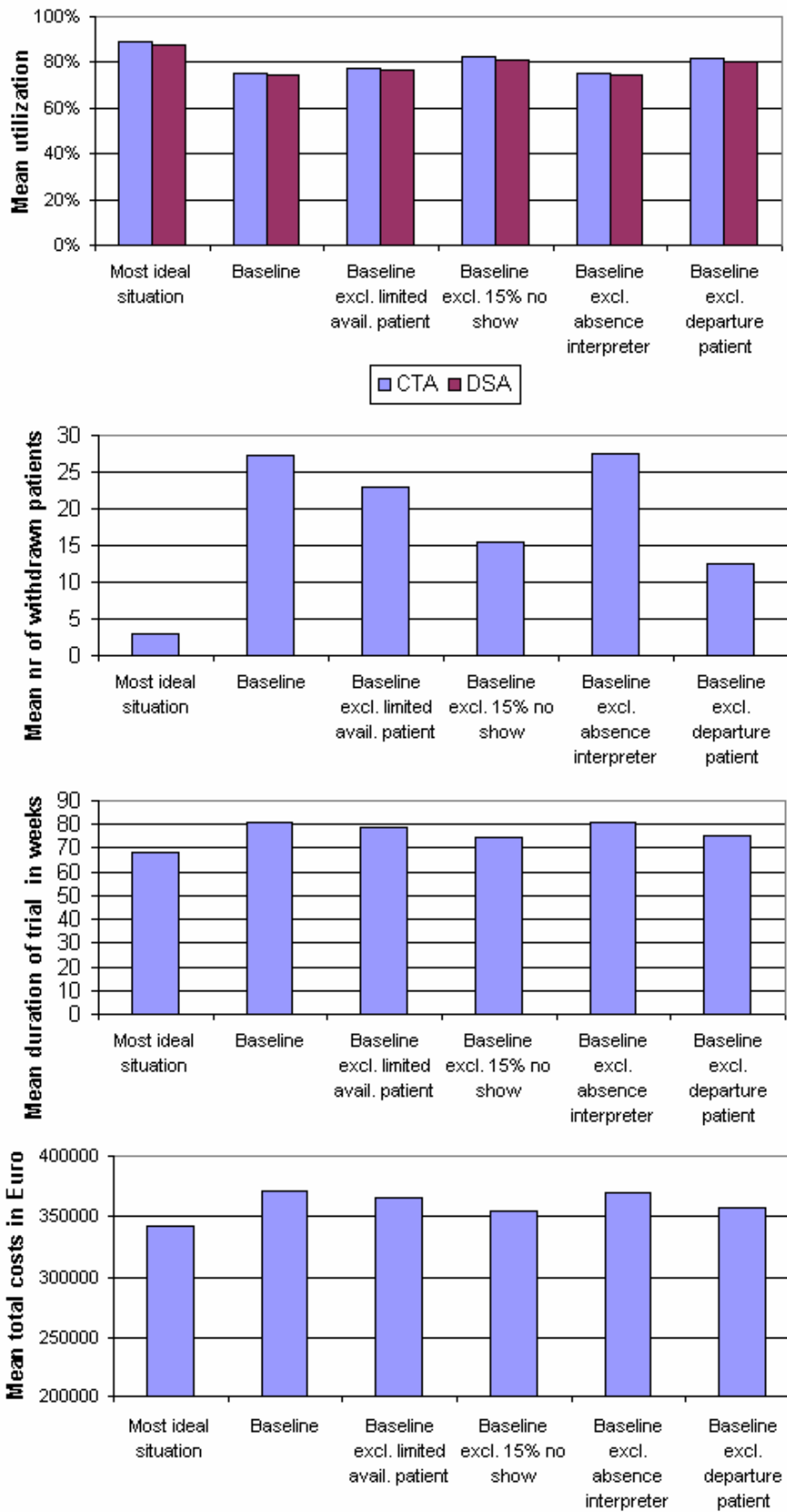


Figure 15: Results varying situations

In case the study is performed in the most ideal situation we see that the utilization of the CTA and DSA are about 89%. In the most ideal situation patients are always available on the first offered appointment day, the patients will not withdraw from the study before the maximum waiting time is reached, all patients will show up exactly on time for their appointment and the test interpreters are always present for the test evaluations. The baseline comprises all these kind of problems and we see in Figure 15 that the utilization of the CTA and DSA decreases to approximately 75%.

The other four bars in Figure 15 show what happens to the utilization of the diagnostic imaging modalities if one of the problems is solved. For example the “Baseline excl. 15% no-show” denotes the scenario where all types of problems occur except of the 15% no-shows. Thus the patients have limited availability concerning the appointment day, the willingness to participate in the study decreases when the waiting time exceeds three weeks and the interpreter is not always available to interpret the test results. We see that if the problem of no-shows is solved in practice that the utilization of the CTA and DSA increase to approximately 81%. When the problem of the willingness to participate is solved the utilization of the test modalities increases to 81% as well.

The absence of the test interpreters and the limited availability of the patients hardly influence the utilization of the equipment. Solving these problems results in an average increase of the utilization of two percentage points compared to the baseline situation.

The mean number of withdrawn patients in the most ideal situation is nine times smaller than the mean number of patients that withdraw from the study when the baseline situation is examined (second graph, figure 15). If the patients are more willing to participate in the study when the waiting time increases then the number of withdrawals will decrease to half the number of patients compared to the baseline situation.

If all patients show up on time for their appointment the mean number of withdrawals will decrease as well, to 17 per study. The mean number of patients that will withdraw from the study will decrease when the restriction of the limited availability of the patients is not counted.

The mean study duration for the most ideal situation is 68 weeks, for the baseline it is 81 weeks. We can find a similar trend for the mean study duration as we saw for the mean utilization and the mean number of withdrawn patients; the duration of the study decreases when no-shows occur and when the patients are not bothered with a waiting time longer than three weeks.

The mean total costs of the study organized through a Cohort study design varies between 341,500 Euro and 371,500 Euro. The lowest costs are obtained in the most ideal situation. The study is the most expensive for the baseline situation. The situation where none of the patients will depart before the maximum waiting time is reached is about 13,700 Euro less expensive than the baseline situation. The total costs for the study decreased about 5,000 Euro and 1,000 Euro for the situation where no restriction exists on the availability of the patient and the presence of the interpreters. The situation where no-shows do not occur has, apart from the most ideal situation, the lowest costs; approximately 355,000 Euro.

### Varying the number of timeslots of the baseline scenario for the Cohort study design

In the second simulation scenario for the Coronary CTA study we evaluated the outcome measures for the baseline scenario for varying number of timeslots per test per week. For different number of available timeslots per test per week we evaluated the mean utilization of the test modalities, the mean number of patients that withdraws from the study, the mean duration of the study and the mean total costs. We only evaluate the outcome measures for the case where the study is organized through the Cohort study design.

In Table 6 the numerical results of the analysis are presented. The results are displayed in graphs in Figure 16.

	1 TS per test	2 TS per test	3 TS per test	4 TS per test
<b>Utilization A</b>	84%	75%	60%	46%
<b>Utilization B</b>	83%	74%	59%	46%
<b>Mean total waiting time in days</b>	27	19	15	14
<b>Nr of Withdrawn Patients</b>	146	27	4	1
<b>Duration of trial in weeks</b>	144	81	71	69
<b>Mean total costs</b>	€ 514,912	€ 370,486	€ 346,427	€ 343,881
<b>Mean total costs incl. penalty</b>	€ 547,360	€ 424,934	€ 458,137	€ 532,443

Table 6: Numerical results for the baseline scenario where the number of timeslots per test per week are varied.

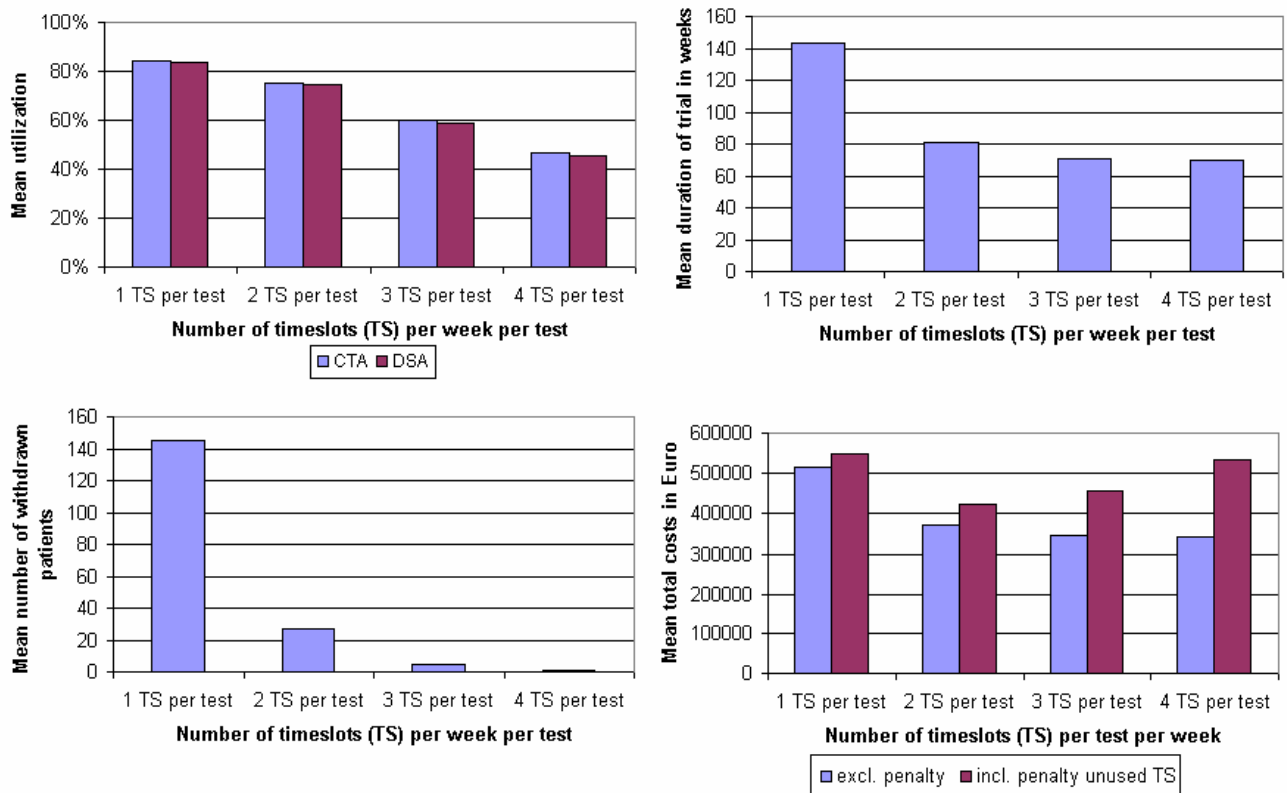


Figure 16: Graphical representation of the results for the baseline scenario where the number of timeslots per test per week is varied.

Table 6 and the first graph of Figure 16 show us that there is a slight decrease of the mean utilization of the test modalities if the number of timeslots per week is doubled from one timeslot per test to two. However, if one decides to reserve more than two timeslots the mean utilization of the test modalities decreases considerably to approximately 46% when four timeslots per test per week are reserved.

The mean duration of the study will explode to 144 weeks if only one timeslot per test is reserved for the study. The mean duration of the study will decrease considerably as soon as two timeslots per week are available. This can be explained by the fact that there are more patients a week available for the study than there are timeslots when only one timeslot per test is reserved for the study. With two or more timeslots per test available there is enough capacity to satisfy the needs.

We see this trend as well for the number of withdrawn patients per study. When only one timeslot per test per week is reserved, we see that the number of withdrawn patients increases to 145. Because there is not enough capacity, the mean waiting time per patient will increase sharply. This leads to withdrawals because patients are not willing to wait longer than three weeks before undergoing a test.

When enough capacity is reserved for the study (two timeslots or more per test), we see that the waiting time decreases and that the number of withdrawn patients decreases as well.

There is a considerable decrease of the mean duration when three timeslots per test per week are available. The duration decreases with more than 10 weeks compared to the situation where two timeslots are reserved for each test. The availability of four timeslots a week hardly makes a difference in the mean total duration of a study, it will last two weeks shorter than when three timeslots per test per week are available.

The mean total costs of the study are the lowest if for each test three or four timeslots per week are reserved. The costs for the study increases when two timeslots per test are reserved weekly for the study. The costs will increase by approximately 7% up to 370,500 Euro. When one timeslot is reserved for the study the costs rise up to 515,000 Euro.

If we consider a penalty cost for each timeslot that remains unfilled we see that it is better to reserve two timeslots per test per week for the study. The mean total costs including a penalty for the study with four timeslots per tests are almost as high as the costs for the study with one timeslot per test. The difference is only 15,000 Euro while without penalty costs it is more than 171,000 Euro.

### Varying the type of study design for the baseline scenario

In the final simulation scenario for the Coronary CTA study we evaluated the outcome measures for the baseline scenario for the four different types of study designs. For each type of study design we evaluated the mean utilization of the test modalities, the mean number of withdrawn patients, the mean duration of the study and the mean total costs.

Table 7 shows the numerical results for varying the type of study designs for the baseline scenario. Figure 17 gives the graphical representation of these results.

	Cohort	Before After	RCT-1	RCT-2
<b>Utilization A</b>	75%	78%	47%	74%
<b>Utilization B</b>	74%	79%	47%	75%
<b>Total mean waiting time in days</b>	19	16	8	19
<b>Nr of Withdrawn Patients</b>	27	16	0	28
<b>Duration of trial in weeks</b>	81	76	65	81
<b>Mean total costs</b>	€ 370,486	€ 267,893	€ 241,185	€ 371,327
<b>Mean total costs incl. penalty</b>	€ 424,934	€ 290,297	€ 332,403	€ 425,277

Table 7: Numerical results for the baseline scenario of the Coronary CTA study where the type of study design is varied.

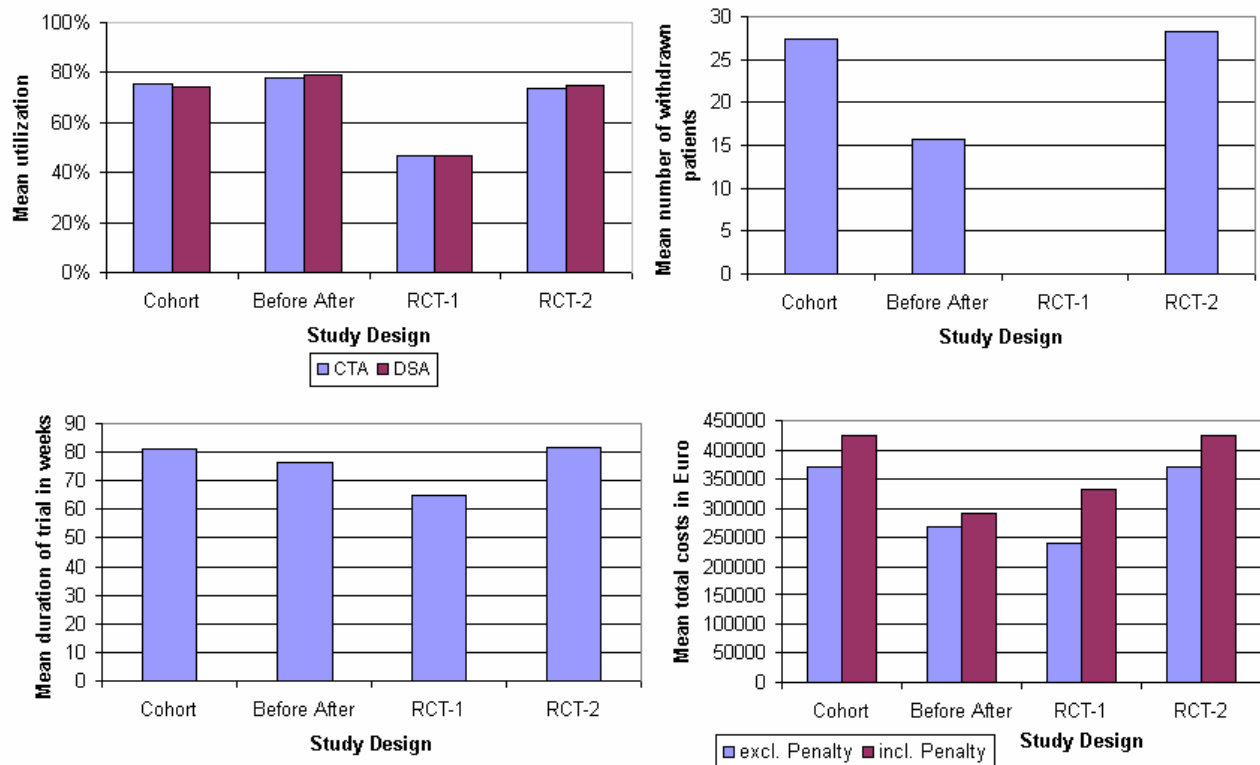


Figure 17: The graphical representation of the results for the baseline scenario of the Coronary CTA study for the different types of study designs.

From the figure and the table we can see that the highest utilization is obtained by the Before-After study design. The utilization of the Cohort and RCT-2 study design are almost equal to each other. The lowest utilization is obtained by the RCT-1 study design, the utilization for the



diagnostic test modalities is approximately half the utilization obtained with the other study designs.

The mean duration of the study set up according to the Cohort study design is equal to the mean duration of the study organized through the RCT-2 study design, 81 weeks. If the study is organized through the Before-After study design it will last 5 weeks shorter. The study can be performed the fastest when the RCT-1 study design is used, then it will take approximately 65 weeks to complete the trial.

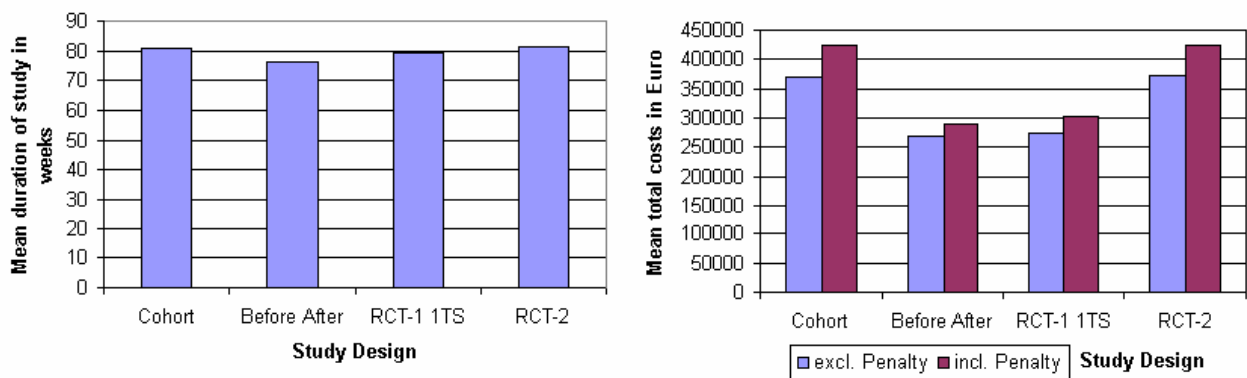
The mean waiting time and the mean number of withdrawn patients show the same trend as the mean duration of the study. The Cohort and the RCT-1 study design obtain the highest values for the outcome measures. The mean waiting time and the number of withdrawn patients decreases when the Before-After study design is used for the study, but the main difference is obtained by the RCT-1 study design. The total mean waiting time for the patients is less by half than the total mean waiting time for the Cohort study design, namely 8 days against 19 days. The difference of the number of withdrawn patients is even bigger; within the Cohort study design 27 patients will leave the study while when the study is set up according to the RCT-1 study design no withdrawals will occur.

The trend of the mean total costs for the study for each different study design is similar to the trend of the mean total costs of Figure 13 in Section 4.3. The costs for the Cohort and RCT-2 study design are the highest and equal to each other. The RCT-1 study design is the less expensive study design of all. Finally, the mean total costs for the RCT-1 study and the Before-After study hardly differ.

As we already saw from Figure 17 and Table 7 the RCT-1 study design will lead to the lowest utilization of the test modalities, the shortest mean duration of the trial and the lowest mean total costs. To obtain the same utilization for the test modalities as the other study designs did we decided to perform an extra analysis for the RCT-1 study design where we changed the number of reserved timeslots per week from two to one. Table 8 and Figure 18 show the results if we compare the outcomes of the RCT-1 study design with one timeslot per test with the outcomes of the other study designs where two timeslots per test per week are reserved.

	<b>Cohort</b>	<b>Before After</b>	<b>RCT-1</b>	<b>RCT-1 1TS</b>	<b>RCT-2</b>
<b>Utilization A</b>	75%	78%	47%	76%	74%
<b>Utilization B</b>	74%	79%	47%	76%	75%
<b>Total mean waiting time in days</b>	19	16	8	21	19
<b>Nr of Withdrawn Patients</b>	27	16	0	25	28
<b>Duration of trial in weeks</b>	81	76	65	80	81
<b>Mean total costs</b>	€ 370,486	€ 267,893	€ 241,185	€ 275,426	€ 371,327
<b>Mean total costs incl. Penalty</b>	€ 424,934	€ 290,297	€ 332,403	€ 301,344	€ 425,277

Table 8: Numerical results for the baseline scenario of the Coronary CTA study where the type of study design is varied.



**Figure 18: The mean duration of the study (left) and the mean total costs of the study for different types of study designs. One timeslot per test is reserved for the RCT-1 study design, for the other study designs are two timeslots reserved.**

The results of the extra simulation analysis show us that the outcomes of the RCT study design and the Cohort and RCT-2 study design are nearly the same when we decrease the number of timeslots for the RCT-1 study design while we leave the number of timeslots for the other study designs unchanged. But, the RCT-1 study design is still less expensive compared to the Cohort and RCT-2 study design. The difference in the mean total costs is about 95,000 Euro. The total number of tests each patient will undergo causes the main difference in the mean total costs. In the Cohort and RCT-2 study design each patient undergoes two tests, while within the RCT-1 study design each patient undergoes one test. This leads to a difference of approximately 78,000 Euro. The rest of the difference in the mean total costs is caused by the extra personnel costs for the extra week that the study takes when it is set up according to the Cohort or RCT-2 study design.

## 5 Discussion

### 5.1 Results

In this study we evaluated four types of study designs to find guidelines for the choice of study design for future research. The most important performance measures of this study are the mean utilization of the diagnostic test modalities, the mean duration of the study and the mean total costs of the study. The four study designs we investigated are:

- Cohort
- Before-After
- RCT with randomization of patients across diagnostic imaging tests (RCT-1)
- RCT with randomization between providing the test results of test A versus providing test results of test B (RCT-2)

In the study we developed four discrete event simulation models to study the complete process of each type of study design. The most striking results of comparing the mean utilization, mean duration and mean total costs of the different study designs are:

- Using the Cohort study design or the RCT-2 study design result in almost the same outcome measures. The mean duration, mean utilization and the mean total costs are almost similar for both types of study design.
- The Before-After study design and the RCT-1 study design are both less expensive than the Cohort and RCT-2 study designs. The outcome measures of the Before-After and the RCT-1 study design were also almost equal to each other. The difference between the mean total costs is only a few thousand euros.
- The utilization of the diagnostic test modalities is for the RCT-1 study design approximately half of the utilization of the other study designs. When a penalty cost is counted for the unfilled timeslots within the RCT-1 study design, the mean total costs of this study design are still less expensive than the mean total costs of the Cohort and RCT-2 study design.
- The mean duration of the trial and the mean utilization of the test modalities will be equal to the outcome measures of the Cohort and RCT-2 study design when the RCT-1 study design is implemented with half the capacity as the other study design. The mean total costs of the RCT-1 study design are much lower than the mean total costs of the Cohort and RCT-2 study design; the difference is about 100,000 Euro.
- A slight overcapacity of the number of reserved timeslots per week results in a considerable reduction in duration of the study and expected total costs (excluding penalty). The results of Section 4.4 showed that the reservation of three timeslots per week instead of two, with an average of two patients per week, reduces the duration of the study with 10 weeks and the total costs with approximately 24,000 Euro.

In our study we also considered the effect of external factors on the simulation outcomes. The external factors we considered are the occurrence that patients do not show up for their appointment, the possible limited availability of patients due to obligations at work, the decrease of the willingness of patients to participate in the study when the waiting time increases and the possibility that a test interpreter is not available for the study due to meetings and seminars. The most striking results for these analyses were:

- The occurrences of patients that will not show up for the study increase the mean duration and mean total costs of the study. The mean utilization of the diagnostic imaging tests decreases when patients do not show up for their appointment.
- The influence of the number of no shows on the mean total costs of a study increases when the total sample size of a study increases.
- The mean duration of the trial and the mean total costs of the study increase when the willingness of the patients decreases as a result of an increasing total waiting time for diagnostic imaging.
- The outcome measures are hardly influenced when the availability of patients is limited due to obligations at work.
- The limited availability of the test interpreter for the study also hardly influences the outcome measures of the simulation models.

The question may arise why we decided to simulate the implementation of the different study designs rather than using common sense and solving the problem by hand, what currently is done when a prediction of the mean duration and costs is made when a new study is organized.

By simulating a clinical study we are able to include realistic external factors that will not be included when the prediction of the outcome measures is calculated by hand. The external factors that we considered are the possibility that patients might not show up for their appointment, the decrease of the willingness to participate in the study when the waiting time increases, the limited availability of patients due to obligations at work and the possible absence of the test interpreter due to seminars and meetings.

In a simulation study we are able to examine the effect of each of these factors individually on the outcome measures. The power of simulation however is in the ability of examining the combination of these external factors. It is possible that the effect of an external factor will be more powerful when another external factor is implemented. With simulation we are able to perform sensitivity analysis where the influence of combining the existence of external factors is examined. For example, the user is able to give the percentage of eligible patients that might not show up for their appointment time. We expect that the total duration of the trial will increase when the number of no-shows increases as well. As long as we know the average number of patients per week that will be included in the study we are able to estimate by hand the extra time needed to perform the study. However, when using common sense we do not consider the effect on the average waiting time for the patients. If a no-show occurs a timeslot will be unfilled and the patient that should have undergone a test on that day has to be rescheduled to another timeslot and the waiting time for the other patients increases. The

occurrence of a no-show might result in a higher number of withdrawn patients, because the willingness to participate in the study decreases when the waiting time increases. No-shows do not only influence the number of unfilled timeslots and thus the duration of a study, they influence the waiting times and the willingness to participate in a study as well. These effects are very logical but it is hard to quantify them, especially by solving the problem by hand. Simulation makes it very easy to gain insight in these effects.

Using simulation the investigator is also able to obtain more information like the expected utilization of the test modalities, the expected waiting time for the patients and the expected number of patients that might withdraw from the study. These outcome measures are not considered when simulation is not used for the organization of a new study.

The results that are presented in Section 4.4 show the usefulness of simulating the complete process of the study designs. In this section we evaluated the most ideal situation and the baseline scenario for the Coronary CTA study. For the most ideal situation we assumed that none of the external factors occur. The baseline scenario implements all of the external factors. Thus the influence of no-shows, the willingness to participate in the study in relation with the waiting time, the limited availability of patients due to obligations at work and the possible absence of the test interpreter due to seminars and meetings are all implemented in the baseline scenario and not in the most ideal situation. The most ideal situation can therefore be compared with common sense and solving the problem by hand that is currently done by the organization of a new study. The results of our analysis show that we would have underestimated the costs for the study with 30,000 Euro and the mean duration of the study with 13 weeks if we had calculated it by hand. The difference of the utilization of the test modalities is about 15 percentage points; the utilization of the equipment would have been overestimated when common sense was used.

Simulation is useful for quantifying the effects of factors of which we know they exist but are not able to estimate them. Beside this, the sensitivity analysis that can be performed with simulation is very useful for the evaluation of the influence of combining different factors on the outcome measures of a study. The obtained information can hardly be estimated by solving the problem by hand.

## **5.2 Limitations**

The main limitation of our study is that the information obtained by a type of study design is not included in the evaluation. In chapter 2 we discussed the advantages and disadvantages of each type of study design and the information that is obtained. Each type of study design provides different information and we did not weigh the additional costs that are needed to obtain more information. For further research we recommend performing a value of information analysis and combining the outcomes with the results presented in this study.

Another limitation might be the perspective of the cost calculations. We measured the costs from a radiological perspective; we calculated the total costs of a study for the Department of Radiology. In general cost effectiveness analysis it is more common to take a broader perspective like the hospital, healthcare system or even societal perspective. To obtain the costs from these perspectives we would need detailed information about the treatment, total number of days a patient stayed in the hospital, etcetera. These costs are all very specific per

disease and not general applicable. For the same reason we did not implement the treatment and follow-up part in detail.

In our simulation models we only simulated the total waiting times for the patients that participated in the study. We did not look at the effect of the waiting times for regular patients when a study is performed. It is very likely that if the waiting time for the patients that participate in a study is minimized that the waiting time for regular patients that do not participate in a study will increase. The prioritization of patients that participate in a study may lead to extra hospitalization of regular patients who have to wait for their scan, which results in extra costs for the hospital.

However, the unfilled timeslots of a study are used to scan the patients that are at that moment hospitalized. The Department of Radiology is able to schedule inpatients in the unfilled timeslots of the study. The researcher of the study should inform the department on time, for example a few days ahead, that the timeslots are not going to be used for the study.

### **5.3 Conclusions**

Many factors play a role in the choice of the most efficient study design. Each type of study design provides its own type of information. Before setting up a new study the investigators have to consider what kind of information they would like to obtain and the available capacity for the study.

The results show that the duration, utilization of the test modalities and the total costs of a study are almost similar for the Cohort and RCT-2 study designs. These outcome measures are similar as well for the Before-After and RCT-1 study design. A study is less expensive when it is set up according to the Before-After or RCT-1 study design than when it is set up according to the Cohort or RCT-2 study design.

When we look at the availability of the test modalities we saw that the RCT-1 study design is the most efficient when there is not enough capacity. When enough capacity is available for the study it is better to reserve extra capacity of approximately 150% of needed on average. It is better to reserve three timeslots per tests when two patients a week are expected to be eligible for the study instead of reserving two timeslots per week. The 150% overcapacity ensures that there is almost no obstruction of the patient flow and will therefore minimize the duration of the study and thus the expected total costs.

Reserving extra capacity is preferable when the unfilled timeslots are easy to fill by the Department of Radiology. One should think of an agreement with the department, for instance, the researcher should inform the department in time whether the timeslots are filled or not. If a timeslot is not filled, the department is able to schedule inpatients.

Furthermore we saw that the mean duration and the mean total costs are negatively influenced by the occurrence of no shows and the decreasing willingness of the patients to participate when the waiting time increases. The influence of no shows is difficult to manage. The willingness to participate in a study can be managed by reducing the waiting times by reserving enough timeslots per test per week. The waiting times will hardly increase when there is enough capacity to satisfy the needs. Another way to manage the willingness to participate in a study is to pay enough personal attention to the patients. The patients do not

mind waiting a week or two longer when they feel appreciated and if they feel comfortable with the personnel. Personal attention to the patient is therefore very important.

The researchers that start a new clinical study have to consider the advantages and disadvantages of the types of study design and make the decision which study design fits the best within their study. The results of our study can be used as a guideline for their decision.





## References

- [1] Hillston, J, "Model Validation and Verification", 2003, Sep, <http://www.inf.ed.ac.uk/teaching/courses/ms/notes/note14.pdf>
- [2] Hunink, M.G.M., P.P. Glasziou, "Decision making in health and medicine: Integrating evidence and values", Cambridge University Press, 2001
- [3] Hutzschenreuter, A, "Comparison of Study Designs of Randomized Controlled Trials for the Evaluation of New Diagnostic Imaging Technology", 2004, <http://www.few.vu.nl/stagebureau/stage/stageverslagen/stageverslagen-nl.html>
- [4] Kock, M.C.J.M. et al., "Contrast-Enhanced MR Angiography and Digital Substraction Angiography in Living Renal Donors: Diagnostic Agreement, Impact on Decision Making, and Costs", AJR 2005; 185:448-456
- [5] Lasagna, L., "Problems in publication of clinical trial methodology", Clin. Pharmacol. Ther. 1979, Vol. 25, pp 751-753
- [6] Law, A.M., W.D.Kelton, "Simulation and Modeling Analysis", McGraw-Hill, third edition 2000
- [7] Loon, van, A.J. et al. "Randomised Controlled trial of magnetic-resonance pelvimetry in breech presentation at term", The Lancet 1997; 350:1799-804
- [8] Mires, G., F. Williams, P.Howie, "Randomised controlled trial of cardiotocography versus Doppler auscultation of fetal heart at admission in labour in low risk obstetric population", BMJ. 2001 Jun, 322(7300), pp 1457-1460, discussion 1460-1462
- [9] Ouwendijk, R, "Imaging Peripheral Arterial Disease: A Randomized Controlled Trial Comparing Contrast-enhanced MR Angiography and Multi-Detector Row CT Angiography", Radiology 2005; 236:1094-1103
- [10] Zelen, M, "The randomization and stratification of patients to clinical trials", J Chronic Dis. 1974, Vol. 27, pp 365-375.



## Appendix A – Scheduling patients using a probability

The participating patient has to be scheduled to an available timeslot of the test. We assume that patients are not always available on the offered time and date. For instance, patients with a job are not always able to rearrange or move appointments when the waiting time is very short, for example less than two weeks. For this reason it might occur that patients are not available on the first free timeslot.

Therefore we simulate the availability of patients according to a *probability of availability table*. The idea is that the system finds the first free timeslot of the diagnostic test that has to be performed. Next it will check whether the patient is available on that timeslot. If the patient is available then the timeslot will be reserved, if not, the system will look for another timeslot, until the patient is available. After finding the first timeslot, the system will look for a free timeslot for the second test, which will be later in time than the timeslot for the first test. The moment a timeslot is found according to the patients agenda we are able to define the waiting time between now and the day of the second test.

The probability that the patient eventually will not participate in the study increases when the waiting time increases. As soon as is decided that the waiting time is not too long and the patient will still participate in the study, the found timeslots are booked.

The availability of the patient and the decision whether or not to participate according to the length of the waiting time are both based on a probability. When the simulation starts the user has to enter several input parameters. The end-user enters, for instance, the probability that the patient is not available on the first day after arriving in the system and how many days after arriving this probability will be equal to one. Between these two days the probability of the availability of the patient will increase linearly over time.

For example, let assume that the patient is available with a chance of 40% on the first day after arriving. Next, assume that the patient is availability with a probability of 1 after four days after arrival. Then the probability of availability table will be as follows:

Number of days after arrival	Probability of availability of patient
1	0.4
2	0.6
3	0.8
4	1

**Table 9: Example of probability of availability table**

Assume that the first available timeslot for the scan is one day after the arrival of the patient. The chance that the patient will undergo the test on this offered free timeslot is 40%. There is a chance of 60% that the patient is not able to cancel his or hers appointment. When a patient is not available on the first offered free timeslot, which is determined by a sample of the homogeneous distribution, the system will look for another timeslot. Assume that the next free timeslot is 3 days after the arrival of the patient in the system. Then the probability that the patient is available on that day is increased to 0.8. With a chance of 80% we are able to schedule the patient on that particular timeslot.

The patient is available with a probability of 1 on the offered day when the first free timeslot is more than 4 days after the arrival of the patient in the system.



## Appendix B – Screenshots of input screens Cohort Simulation model

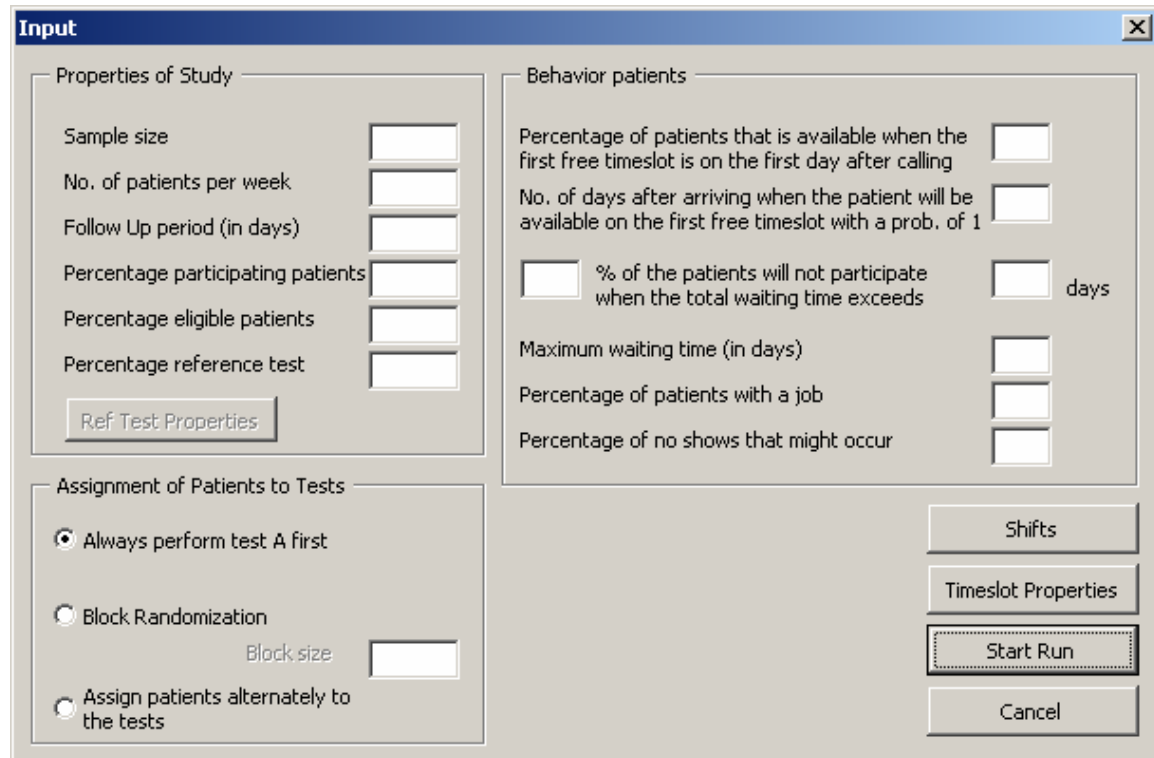


Figure 19: The Input screen of the Cohort simulation model

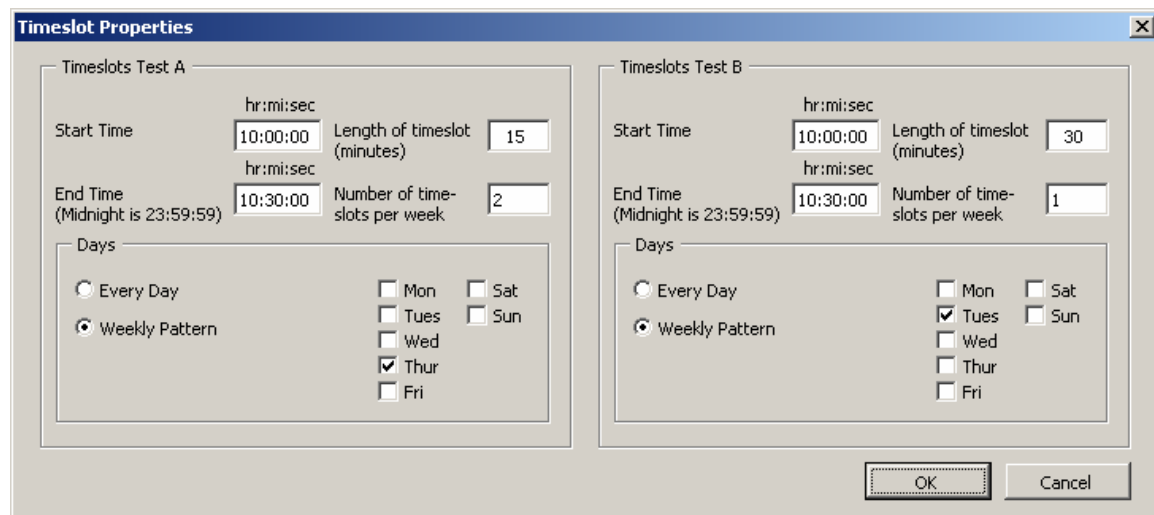


Figure 20: The Timeslot Properties screen of the Cohort simulation model

Figure 21: The Shift properties of test interpretation screen of the Cohort simulation model

Figure 22: The Timeslot properties Reference Test screen of the Cohort simulation model

**Shift of Ref Test Interpretation**

Shift properties of Interpreter Reference Test

Start Time

End Time (Midnight is 23:59:59)

Time needed to evaluate test result (minutes)

Days

Every Day  Mon  Sat

Weekly Pattern  Tues  Sun

Wed

Thur

Fri

Absence Interpreter

Mean number of days present between each absence

Mean period of absence (in days)

Figure 23: The Shift properties of Ref Test Interpretation screen of the Cohort simulation model





## Appendix C – User manual of Arena simulation models



# 1 Introduction

This user manual is written for the Department of Epidemiology & Biostatistics at the Erasmus Medical Center Rotterdam. The document can be used as a support when simulation runs are performed using the simulation models Cohort.doe, Before-After.doe, RCT-1.doe or RCT-2.doe.

After reading this manual the user is able to perform simulation runs and to interpret the results of these runs.

In this document we linked four user manuals, namely the user manuals of each of the four simulation models. The input that is needed from the user is almost similar for each model. The user manuals of the simulation models are for this reason quite the same. Therefore we decided to include the complete user manual of the Cohort simulation model in chapter 3. Chapter 4 till 6 consist of descriptions of the differences in input of each simulation model compared to the input of the Cohort simulation model. In chapter 7 we discuss the output of the simulation models, and finally, in chapter 8 we describe how the output of the simulation models can be used for costs calculations.

Enjoy reading this user manual and have fun simulating!

Jolanda Veldhuis  
December 2005

## 2 Arena

For the development of the simulation model we used the Arena Rockwell Simulation software package, version 9.0 Research Edition. Arena is a popular general-purpose simulation package in the field of operations research where people are able to analyze complicated systems.

In this chapter we will shortly describe the features that are important for the user to know before getting started with Arena. We will describe how a simulation run of the Before-After model can be started, stopped and how the end-user is able to change the number of replications performed in one simulation run. The information provided in this chapter is obtained from the book “Simulation with Arena” written by Kelton, Sadowski & Sadowski. For more details about Arena and its features we refer to this book. It is also very helpful to check the Arena online Help, which can be started by the Help menu or pressing de F1 function key.

### 2.1 Getting started

Arena can be started in two different ways. The user can start Arena through the short cut on the desktop of his or hers computer or by opening an existing model. When Arena is started by the short cut on the desktop it opens with a brand new model named Model1.doe. The user is then able to define and make its own simulation model. Because we are not developing a new model we will start the Arena program by opening the Before-After model. First go to the folder where the file Before-After.doe is located. The Arena program will start the moment the user double-clicks on the Before-After.doe file. After a few seconds Arena is started with the Before-After simulation model. On top of the screen, just where the menu bar is, we can define several toolbars of Arena, figure 24. It depends on the settings of the program whether all toolbars are shown or not.

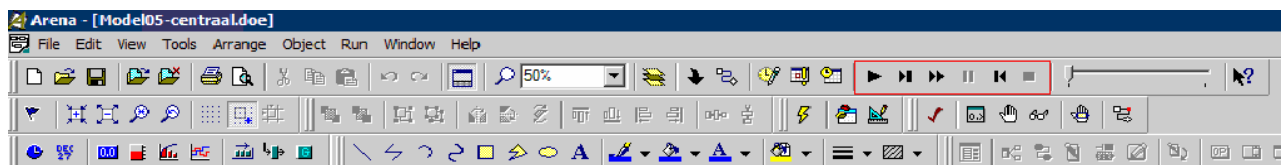









Figure 24: The toolbars in Arena.


We are only interested in running the model and analyzing the results. The Run toolbar is therefore the only toolbar that is of interest simulating the Before-After model. All the other toolbars are mostly used for designing and developing new models, which is beyond the scope of this user manual. The Run toolbar is marked with a red square in figure 24. The options displayed in this toolbar are also available from the Run menu.


The  button from the Run toolbar starts running the model. The model can be started as well by pressing Go from the Run menu or by the F5 function key.

The  button (or Step from the Run menu, or the F10 function key) executes the model one action at a time so the user can see in detail what is going on. This is primarily used as a debugging or demonstration tool.

The  button (or *Fast-Forward* from the Run menu) disables the animation and executes the run at a much faster rate. Because the animation of the Before-After simulation model is already turned of to obtain a much faster execution of the model this option will not influence anything of the speed of the simulation run.

The  button (or *Pause* from the Run menu or Esc) interrupts the run. Hit ,  or  to resume the simulation.

The  button (or *Start Over* from the Run menu or Shift+F5) goes back to the beginning and reruns the model. It is recommended not to use this option for the Before-After simulation model, because this model needs a lot of input from the end-user (chapter 3). When the start over button is pressed the model does not remember which input is entered by the user and will therefore has problems starting over again. When the user would like to start over the simulation run we recommend to end the run properly and restart it by pressing the start button again. The input screens still contain the same input as the user just had entered.

The  button (or *End* or Alt+F5) will end the simulation run. The moment a simulation run is started the model gets in *run mode*. To get out of the run mode the user needs to select the end button. For our Before-After simulation model the user only has to end the model when he or she decides to stop the simulation run before it is finished. Else, the Before-After model will automatically end the run mode.

## 2.2 Changing the number of replications

A simulation run consists of several replications. In one replication the process of performing a study is simulated one time. Thus, if a simulation run consists of 1000 replications, then the study is simulated for 1000 times. In the Before-After simulation model the number of replications is standard set to 1000. Executing the simulation run takes a lot of time; the average time to complete one simulation run of 1000 replications is about 30 to 60 minutes. If the user would like to change the number of replications then this can be done via *Set up...* in the Run menu. A box like the one in figure 25 appears on the screen. As we can see it consists of several tabs, like Run Speed, Run Control, Reports etcetera. To change the number of replications the user has to go to the Replication Parameters tab. The number of replications can be changed in the textbox that has the label Number of Replications; this one is selected in figure 25. The user can change the number of replications in any number larger than zero. We recommend to enter a number larger than 500, the results are the most reliable when more than 500 replications are done.

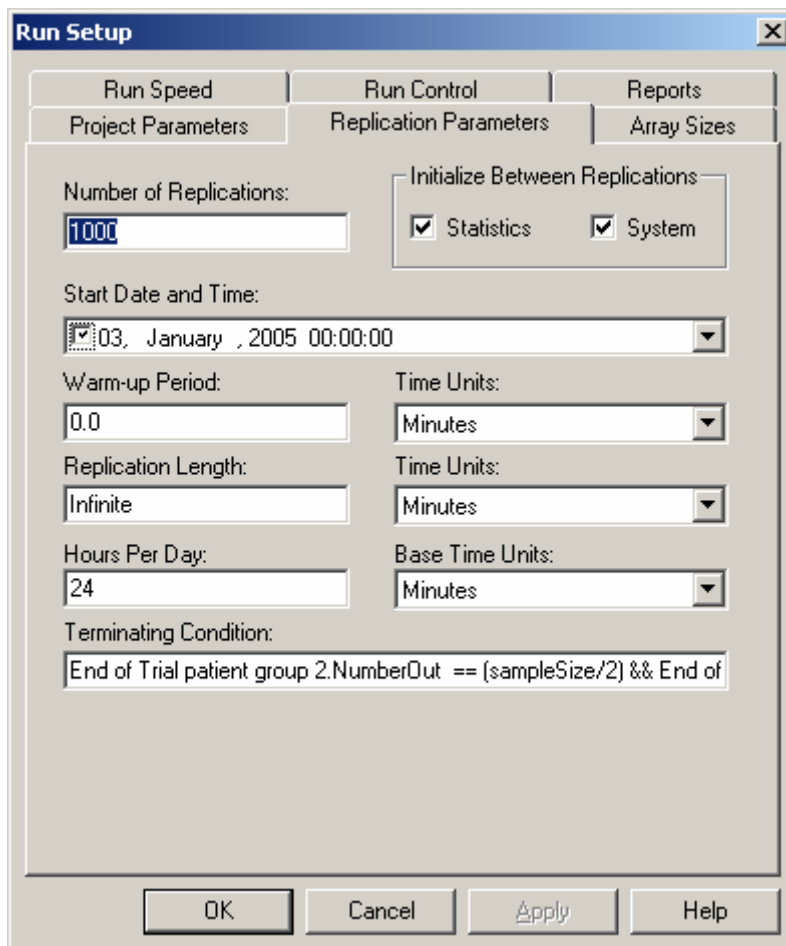


Figure 25: Set up box of the Run menu.

### 3 Cohort simulation model

The user is able to enter the values for each input parameter at the beginning of the simulation. These values need to be entered through the “Input screen”, the “Timeslot properties Reference Test screen”, the “Shift properties of Ref Test Interpretation screen”, the “Timeslot properties screen” and the “Shift properties of test interpretation screen”. In this chapter we will discuss the features and parameters of each screen.

#### 3.1 Input screen

The input screen is the first screen that appears when a simulation run is started. In this screen all parameter values for the simulation will be entered. The input screen is also the navigator to all other screens. From the input screen we can get to the “Timeslot properties Reference Test screen”, “Timeslot Properties screen” and the “Shift properties of test interpretation screen”.

The next figure shows the input screen of the Cohort simulation model.

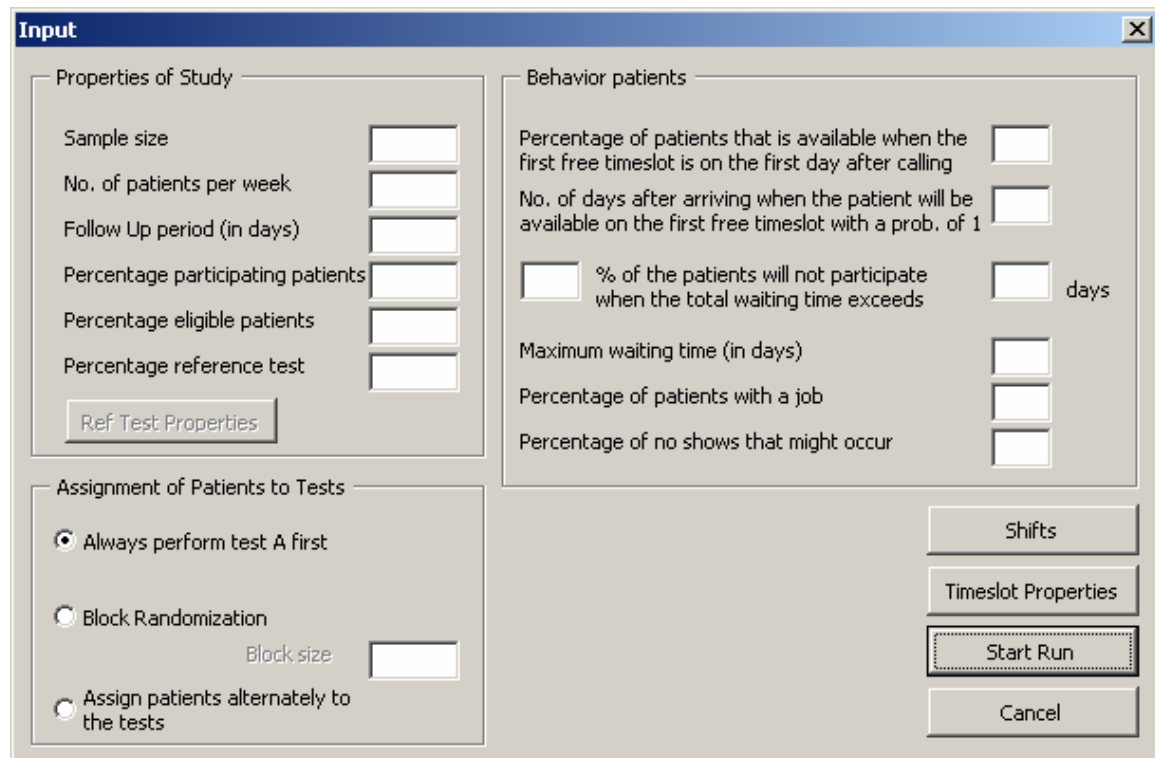


Figure 26: The Input screen of the Cohort simulation model

#### Sample size

The first parameter value that should be entered is the value for the sample size. With the sample size we mean the sample size of the study. In the textbox you can fill in any value larger than zero. The sample size is the number of included patients that will reach the end of the trial. The simulation run is ended as soon as this number of patients has reached the end of the trial.

### **No. of patients per week**

This parameter value indicates the number of patients that is expected to arrive per week. In the textbox you can fill in any (decimal) value larger than zero. If the expectation is that there will arrive only one patient each two weeks, just fill in the inverse value, 0.5. One patient arriving in two weeks is the same as a half patient arriving in one week.

### **Follow Up period (in days)**

For this parameter value we just fill in the total length of the follow-up period represented in days. The value of this parameter should be larger or equal to zero.

### **Percentage participating patients**

The percentage participating patients denotes the average percentage of eligible patients that is willing to participate into the study. The value of this parameter should be in the range [0,100].

### **Percentage Eligible patients**

The percentage eligible patients denote the percentage of patients that is eligible for the study. The patients that arrive in the system are not all eligible for the study. This value denotes the percentage of the patients that will be included in the trial. The value of this parameter should be in the range [0,100].

Before continuing the explanation of the other input parameters we first give an example of the last two parameters.

**Example.** Assume that we expect four patients to arrive in the hospital per week with the same symptoms of disease. Two of the patients turn out to be eligible for participation in the study. Only one of the patients is willing to participate in the study. For this particular example the percentages of eligible patients and participating patients will be:

Percentage Eligible patients = 50%, namely 2 of the 4 patients turn out to be eligible.

Percentage participating patients = 50%, namely 1 of the 2 patients is willing to participate in the study.

### **Percentage Reference test**

The percentage Reference test parameter denotes the percentage of included patients that will eventually undergo the reference test. The value of this parameter should be in the range of [0, 100].

In our simulation model we assume that clinical studies evaluate two new diagnostic imaging tests and that an extra reference test can be performed in several or all patients. If the intension of the user is to simulate a study that indeed evaluates two new diagnostic tests, test A and test B, and that an additional reference test can be performed in several cases, then the end-user should specify the percentage of included patients that should undergo the test. Three diagnostic imaging tests are considered in these kind of clinical studies, namely, test A, test B and the reference test.

In clinical practice there are also studies that evaluates one new diagnostic imaging test with the reference test. In this case only two diagnostic imaging tests are considered. If the intension of the end-user is to simulate a study that evaluates a new diagnostic imaging test with the reference test, then the end-user should assume that the new diagnostic imaging test



is test A and that the reference test is test B. The percentage of patients that undergoes the reference test should then be set to zero, because only two diagnostic tests are considered and test B is assigned as the reference test.

**Example.** Assume that three tests will be evaluated in a future study, the Magnetic Resonance Angiography (MRA), Computed Tomographic Angiography (CTA) and Digital Subtraction Angiography (DSA). The MRA (Test A) and the CTA (Test B) are the new diagnostic imaging modalities and the DSA is the current gold standard, the reference test. Then the end-user should specify the percentage of patients that will undergo the reference test on the input screen.

Now assume that in a future study only two tests will be evaluated, the MRA and the DSA. The MRA (Test A) is the new diagnostic imaging test and the DSA is the current gold standard, the reference test. Because there are only two tests evaluated in this future study we set DSA equal to test B and set the percentage of patients that undergo the reference test to zero. We consider the DSA just as another diagnostic imaging test that will be evaluated.

### **"Assignment of patients to test"-frame**

Each patient has to undergo two tests, test A and test B. The user is able to decide in what order the patients should undergo the tests. In the "assignment of patients to tests"-frame the user can choose between the following options:

- Always perform test A first
- Block randomization
- Assign patients alternately to the test

If the end-user chooses the first option, all patients will first undergo test A. When test A is performed the patients will undergo test B.

The second option denotes that the assignment of tests to patients will be according a randomization scheme with a certain block size. As soon as this option is chosen the end-user has to fill in the block size. The value of the block size should be larger than zero. If this option is chosen the assignment of tests will be randomized per patient.

The third and final option denotes that the assignment of tests to patients will be alternated. This means that the first patient will first undergo test A and then test B. The second patient will be first assigned to test B and will then undergo test A. The third patient will undergo test A and after that test B. This process continues until the last patient is included into the study.

### **Percentage of patients that is available when the first free timeslot is on the first day after calling**

The value that is given to this variable denotes the percentage of working patients that is available on a free timeslot that is one day after contacting the hospital for the first time.

The idea is that patients with a job have difficulties with changing their agenda's on the short run. Therefore we give a percentage that denotes the percentage of working people that are able to change their agenda's to undergo the timeslot on the first day after calling.

The value of this parameter should be in the range of [0, 100].

**No. of days after arriving when the patient will be available on the first free timeslot with a prob. of 1.**

This parameter value also gives insight in the behavior of patients with a job. Just like the previous parameter we assume that a certain percentage of patients with a job are not able to change their agenda on the short run. The previous variable denotes the percentage of patients that are available on the first day after calling for an appointment.

This variable denotes the number of days after the first call to the hospital that a patient is available for a timeslot with a chance of 100%. The value should be in the range of [1, 100].

The key idea of the previous two parameters is to build in a feature in the model that considers the behavior of patients with a job. These patients are not always available on the short run and will therefore schedule their appointments later in time. When waiting times are very short, for example less than two weeks, it might be possible that a timeslot will be unfilled due to the obstruction of working patients that were not able to change their agendas. The two parameters indicate the behavior of the patients for a small period of time. The first parameter asks for the probability that a working patient is not able to undergo a test on the first day after contacting the hospital. The second parameter asks for the period that a patient is limited available. With these two parameters the model is able to calculate a probability table which denotes the chances that a patient is able to change his or hers agenda. These probabilities indirectly represent the probability that a timeslot will be filled or not.

Before continuing the explanation of the other input parameters we first give an example of this feature.

**Example.** Assume that the patient is available with a chance of 40% on the first day after arriving. Next, assume that the patient is available with a probability of 1 after four days after arrival. Then the probability of availability table will be as follows:

Number of days after arrival	Probability of availability of patient
1	0.4
2	0.6
3	0.8
4	1

**Table 10: Example of probability of availability table**

Assume that the first available timeslot for the scan is one day after the arrival of the patient. The chance that the patient will undergo the test on this offered free timeslot is 40%. There is a chance of 60% that the patient is not able to cancel his or hers appointments. When a patient is not available on the first offered free timeslot, which is determined by a sample of the homogeneous distribution, the system will look for another timeslot. Assume that the next free timeslot is 3 days after the arrival of the patient in the system. Then the probability that the patient is available on that day is increased to 0.8. With a chance of 80% we are able to schedule the patient on that particular timeslot.

The patient is available with a probability of 1 on the offered day when the first free timeslot is more than 4 days after contacting the hospital for the first time.

For this particular example the end-user should enter the following values for each parameter:

Percentage of patients that is available when the first free timeslot is on the first day after calling = 40.

No. of days after arriving when the patient will be available on the first free timeslot with a prob. of 1 = 4

We set the percentage of patients that is available to 100% and the no. of days after arriving to 1 if we do not want to include the influence of the behavior of patients in the simulation.

### **...% of the patients will not participate when the total waiting time exceeds ... days**

In our simulation model, we make the assumption that long waiting times have a negative effect on the willingness of the patient to participate in the study. If the waiting time per patient increases, the chance that a patient and the treating physician are not willing to participate in the study will increase as well.

The "... % of the patients will not participate when the total waiting time exceeds ... days"-variable denotes the turning point of the acceptance of a long waiting time. For example, if patients and their physicians know that in another hospital the waiting time for a certain test is less than three months and that the waiting time when they participate in this study will be more than 4 months, it is very likely that the patient and the physician decides not to participate in the study and will therefore withdraw.

The percentage given in this variable represents the percentage of patients that might withdraw when the total waiting time exceeds a certain amount of days. For example, we fill in 30 and 10 for this variable, if 30% will not participate in the study when the total waiting time is more than 10 days.

The value of the percentage can be any decimal value in the range of [0,100]. The value for the number of days should be larger than one.

### **Maximum waiting time (in days)**

The maximum waiting time (in days) represents the total waiting time that is acceptable for the patients and their physicians to wait for a diagnostic imaging test. The patient will withdraw from the study if the total waiting time exceeds this value. The value for this variable should be equal or larger than one.

**IMPORTANT:** The value of the maximum waiting time should ALWAYS be one day larger than the value of the number of days of the variable "... % of the patients will not participate when the total waiting time exceeds ... days".

For example, the value of the maximum waiting time should at least be 11 days, if 30% of the patients will not participate in the study when the total waiting time is more than 10 days.

It might be possible that we do not want to take into account the influence of long waiting times in the simulation run. Assume that the maximum waiting time is 42 days and that we do not want to include the influence of the long waiting times in our simulation model. Then we get; 100% of the patients will not participate when the total waiting time exceeds 42 days. The maximum waiting time is then set to 43, because this value needs to be one day larger than the number of days of the variable "... % of the patients will not participate when the total waiting time exceeds ... days".

---

### **Percentage of patients with a job**

The percentage of patients with a job is of interest when we are examining the availability of the patient. Because we assume that patients that do not have a job will be available on all offered timeslots, we need to keep track of patients that are not always available, which are the patients that have a job. The value for this parameter should be in the range of  $[0, 100]$ .

When the expectation is that all participating patients are older than 65 or if it is expected that none of the participating patients have a job we simply give this variable the value 0.

### **Percentage of no-shows that might occur**

In practice it might occur that several patients are not able to get on time in the hospital for their test, some patients might not get in the hospital at all. This can be caused by traffic jam, illness or any other reason. Unfortunately we are not able to perform a test for that particular patient and a timeslot will get lost. To get insight in the total costs of unused timeslots we enabled a “percentage of no-shows that might occur”. This percentage denotes the percentage of no-shows that might occur during the trial. The value of this parameter should be between zero and 100.

The user is able to navigate directly from the “Input”-screen to the “Timeslot properties”-screen and the “Shift properties of test interpretation”- screen. In the next two subsections a description of the input parameters of these screens will be given. After that we will discuss the last two screens of the simulation model.

### 3.2 Timeslot Properties screen

In the “Timeslot Properties screen” the end-user is able to specify on which day and time the test modalities are available. For each test modality the end-user is able to specify the start time, end time, the duration of a test, the number of timeslots per week and the days on which the test modalities are available for the research.

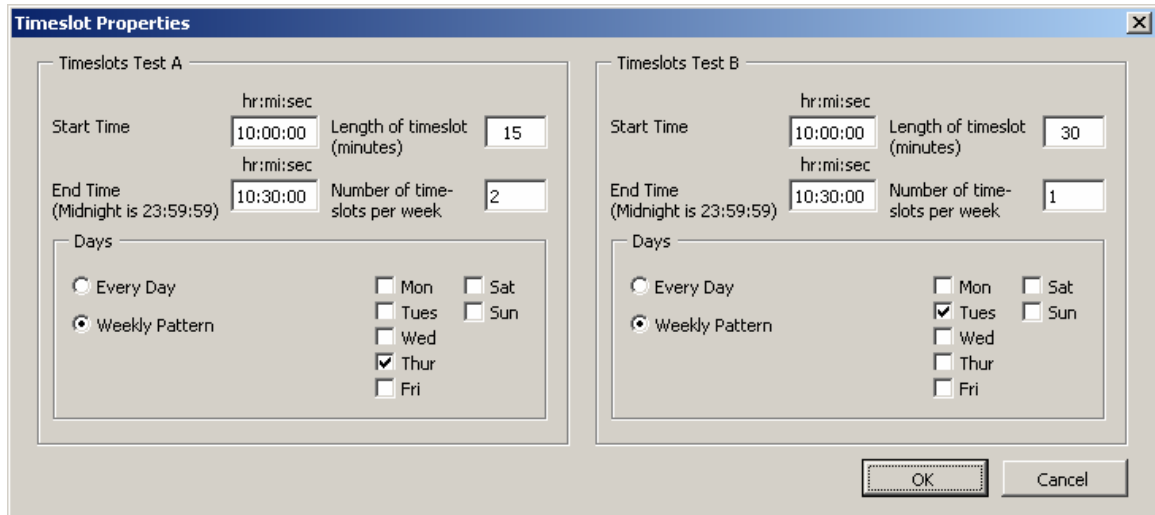


Figure 27: The Timeslot Properties screen

The Timeslot Properties screen consists of two frames; the “Timeslots Test A”- and the “Timeslots Test B”-frame. Both frames are identical. The input parameters in the “Timeslots Test A”-frame represent the timeslot properties for the test modality that is referred to as Test A and the “Timeslots Test B”-frame represents the timeslot properties of the diagnostic imaging test that is referred to as Test B.

#### Start time

The start time denotes the beginning of the timeslot session for the research. The start time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### End time

The end time denotes the end of a timeslot session for the research. The end time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### Length of timeslot (in minutes)

The length of the timeslot denotes the total time that is reserved for the performance of one test, this time includes the scanning time, the entrance and departure time of the patients as well.

#### Number of timeslots per week

This parameter value denotes the total number of timeslots per week that are reserved for the research. This parameter value is used to check the input that is given by the end-user.

#### “Days”-frame

In the “Days”-frame the end-user is able to specify the days on which a timeslot is reserved for the study. There are two options; “Every Day” and “Weekly Pattern”. If the every day option is

selected it means that there are timeslots reserved on each day of the week between the given start and end time. If the “Weekly Pattern” option is selected the end-user is able to specify the days on which timeslots are reserved.

### 3.3 Shift properties of test interpretation screen

In the “Shift properties of test interpretation screen” the end-user is able to specify the start and end time of the test interpreters that interpret the results of test A and test B. The end-user is also able to specify the average evaluation time that a test interpreter might need to evaluate the test results. Finally, the end-user is able to specify the days on which the test interpreter is available for the study and how many times the test interpreter might be absent due to other obligations for work.

Figure 28: The Shift properties of test interpretation screen

#### Start time

The “start time”-variable denotes the moment from which the interpreter is available for the evaluation of test results. As soon as the start time is reached the test results will be evaluated by the interpreter in order of arrival, thus First-Come-First-Served.

The start time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### End time

The “end time”-variable denotes the moment from which the interpreter is no longer available to evaluate and score the test results. As soon as the end time is reached the test results will be stored till the moment the interpreter is available again to evaluate the results.

The end time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

---

### **Time needed to evaluate test result (minutes)**

The value given to this parameter represents the total time that is needed to evaluate the test result. This value should be larger than zero.

### **“Days”-frame**

In the “Days”-frame the end-user is able to specify on which day the test interpreter is available. The interpreter might be available on “Every Day” or available on a “Weekly Pattern”. The last option indicates that the interpreter will be available between the start and end time on the selected days.

### **“Absence interpreter”- frame**

In the “Absence interpreter”-frame the end-user specifies the availability of the interpreter. It might occur that an interpreter is not able to interpret test results due to absence for seminars, conferences, holiday and so on. In the “Absence interpreter”-frame the end-user is able to specify the frequency of absence over time.

The *mean number of days present between each absence* indicates the mean total number of days the interpreter will be available before his or her absence. The *mean period of absence (in days)* is the mean number of days that an interpreter will be absent if he or she is absent. For example, suppose that an interpreter is not present for the research during one week per month. Then the *mean number of days present between each absence* will be 30 and the *mean period of absence (in days)* will then be equal to 7, which indicate that the interpreter is not present for one week.

The Arena software will block the availability of the interpreter based on this parameter values. The user should fill in zero for the mean period of absence when the test interpreter is always present for the evaluation of the test results.

### 3.4 Timeslot properties Reference Test screen

The “Timeslot properties Reference Test screen” can be accessed through the button “Ref Test Properties” on the input screen. This button enables the moment a value larger than zero is filled for the percentage reference test.

In the “Timeslot properties Reference Test screen” the end-user is able to enter the properties of the reference test timeslots (figure 29). The end-user is able to give the start time, end time, length of the timeslot, number of timeslots per week and the days on which a timeslot is reserved for the study.

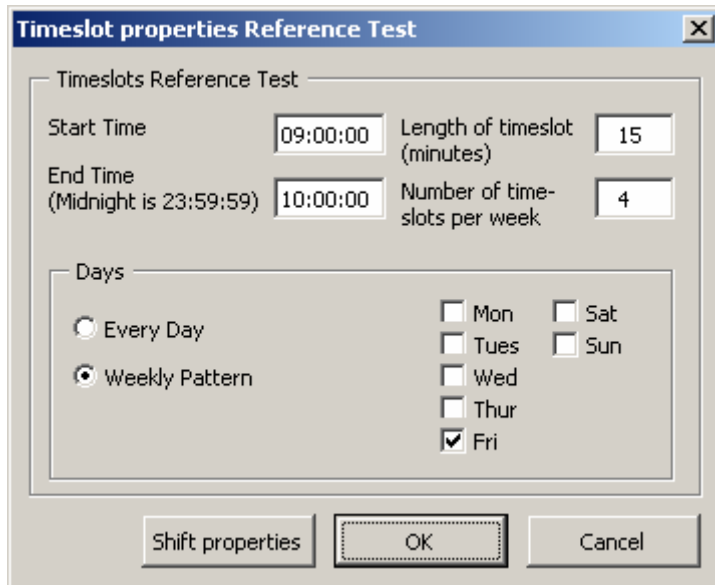


Figure 29: The Timeslot properties Reference Test screen

#### Start time

The start time denotes the beginning of the timeslot session for the research. The start time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### End time

The end time denotes the end of a timeslot session for the research. The end time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### Length of timeslot (in minutes)

The length of the timeslot denotes the total time that is reserved for the performance of one test, this time includes the scanning time and the entrance, departure time of the patients as well. The value for the length of timeslot should be larger than zero.

#### Number of timeslots per week

This parameter value denotes the total timeslots per week that are reserved for the research. The value should be larger than zero and should not contain any decimals.

#### “Days”-frame

In the “Days”-frame the end-user is able to specify the days on which a timeslot is reserved for the study. There are two options; “Every Day” and “Weekly Pattern”. If the every day option is selected it means that there are timeslots reserved on each day of the week between the given



start and end time. If the “Weekly Pattern” option is selected the end-user is able to specify the days on which timeslots are reserved.

### 3.5 Shift properties of Ref Test Interpretation screen

The “Shift properties of Ref Test Interpretation screen” can be accessed through the “Shift Properties” button on the “Timeslot properties Reference Test screen”.

The “Shift properties of Ref Test Interpretation screen” shows the shift properties of the interpreter that evaluates the results of the reference test. Most of the time an interpreter has limited availability to evaluate and score the test results for the study. For this reason the end-user is able to specify the availability of the interpreter. The end-user is able to specify the start time, end time of a shift, the evaluation time of a test result, the days on which an interpreter is available and the degree of absence (if it occurs) of the interpreter.

Figure 30: The Shift properties of Ref Test Interpretation screen

#### Start time

The “start time”-variable denotes the moment from which the interpreter is available for the evaluation of test results. The interpreter will evaluate and score the test results in order of arrival, First-Come-First-Served. The start time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### End time

The “end time”-variable denotes the moment from which the interpreter is no longer available to evaluate and score the test results. As soon as the end time is reached the test results will be stored till the moment the interpreter is available again. The end time should be given in a 24 hour format, from 00:00:00 to 23:59:59.

#### Time needed to evaluate test result (minutes)

The value given to this parameter represents the average total time that is needed to evaluate one test result. This value should be larger than zero.

---

### **“Days”-frame**

In the “Days”-frame the end-user is able to specify on which day the researcher is available. The interpreter might be available on “Every Day” or available on a “Weekly Pattern”. If the “Every Day” option is selected the interpreter is available to evaluate and score the test results on each day of the week between the given start and end time. If the “Weekly Pattern” option is selected the end-user is able to specify the days on which the interpreter is available. The interpreter is available to evaluate and score the test results on the selected days during the given start time and end time.

### **“Absence interpreter”- frame**

It might occur that an interpreter is not able to evaluate and score test results at the given time and days due to suddenly absence for seminars, conferences, holiday and so on. In the “Absence interpreter”-frame the end-user is able to specify the frequency of this absence over time.

The *mean number of days present between each absence* indicates the average number of days that the interpreter will be available before his or her absence. The *mean period of absence (in days)* is the average number of days that an interpreter will be absent if he or she is absent. For example, suppose that an interpreter is not present for the research during one week per month. Then the “mean number of days present between each absence” will be 30. The “mean period of absence (in days)” will then be equal to 7; which indicates that the interpreter is not present for one week.

If the interpreter is always available for the evaluation of the test results then the user should fill in zero for the *mean period of absence*.

## 4 Before-After simulation model

The input needed from the user to perform a simulation run of the Before-After study differs slightly from the input needed for the Cohort simulation model. The difference between the input of the models is in the first screen that appears when the simulation is started, the input screen. The input for all the other screens is just the same as it is for the Cohort simulation model. For a description of the other screens we refer to chapter 3.

The next figure shows the input screen of the Before-After simulation model.

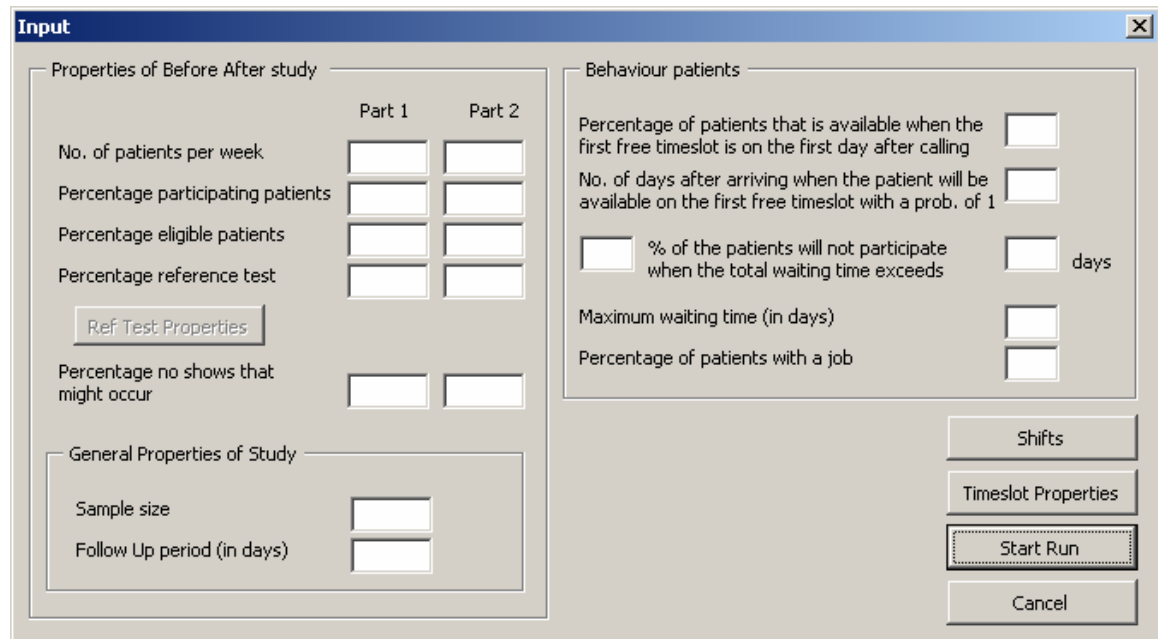


Figure 31: The Input screen of the Before-After simulation model

The Before-After study can be split in two parts. In the first part the current diagnostic imaging test is evaluated, in the second a new diagnostic imaging modality will be evaluated. The two parts of the study are subsequent in time. It might occur that several factors of the study are changed over time. For this reason the end-user is able to enter two values for several input parameters.

### No. of patients per week

This parameter value indicates the number of patients that is expected to arrive per week. In the textbox you can fill in any (decimal) value larger than zero. If the expectation is that there will arrive only one patient each two weeks, just fill in the inverse value, 0.5. One patient arriving in two weeks is the same as a half patient arriving in one week. The value of this parameter might be different when the second part of the study is started. The user is therefore able to enter the values for the first part of the study and the second.

---

### **Percentage participating patients**

The percentage participating patients denotes the average percentage of eligible patients that is willing to participate into the study. The value of this parameter should be in the range [0,100].

The percentage of participating patients might be different as well when the second part of the study is started. The user is therefore able to enter the expected percentage of participating patients for the first part of the study and the second.

### **Percentage Eligible patients**

The percentage eligible patients denote the percentage of patients that is eligible for the study. The patients that arrive in the system are not all eligible for the study. This value denotes the percentage of the patients that will be included in the trial. The value of this parameter should be in the range [0,100]. The value of this parameter might be different when the second part of the study is started. The user is therefore able to enter the values for the first part of the study and the second.

### **Percentage Reference test**

The percentage Reference test parameter denotes the percentage of included patients that will eventually undergo the reference test. The value of this parameter should be in the range of [0, 100]. The expected number of patients that has to undergo the reference test might be different for the second part of the study. The user is able to enter the expected percentage of reference test for the first part of the study and the second part.

### **Percentage of no-shows that might occur**

In practice it might occur that several patients are not able to get on time in the hospital for their test, some patients might not get in the hospital at all. This can be caused by traffic jam, illness or any other reason. Unfortunately we are not able to perform a test for that particular patient and a timeslot will get lost. To get insight in the total costs of unused timeslots we enabled a "percentage of no-shows that might occur". This percentage denotes the percentage of no-shows that might occur during the trial. The value of this parameter should be between zero and 100.

### **Sample size**

The first parameter value that should be entered is the value for the sample size. With the sample size we mean the sample size of the study. In the textbox you can fill in any value larger than zero. The sample size is the number of included patients that will reach the end of the trial. The simulation run is ended as soon as this number of patients has reached the end of the trial.

The sample size should be larger than zero and an even number. If the end-user enters an odd number the simulation model will not work.

### **Follow Up period (in days)**

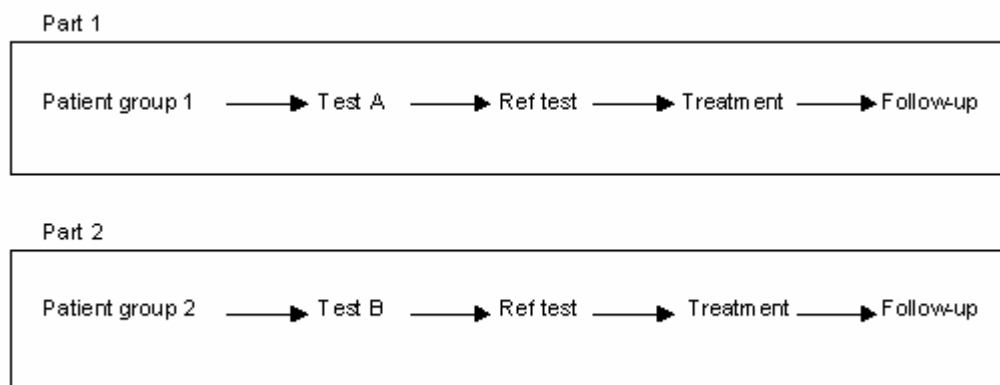
For this parameter value we just fill in the total length of the follow-up period represented in days. The value of this parameter should be larger or equal to zero.

The user is able to navigate directly from the “Input”-screen to the “Timeslot properties”-screen and the “Shift properties of test interpretation”- screen. A description of these screens can be found in chapter 3.

Before starting a simulation run of the Before-After simulation model the end-user should specify the start time of the second part of the Before-After study in the Arena model. The next section shows how the user should do this.

#### 4.1 Changing the start time of the second part of the study

The Before-After study consists of two parts (figure 32). In each part of the study a diagnostic imaging modality is evaluated. The two parts of the study are carried out at different moments in time; the second part occurs later in time than the first part, it will only start when the inclusion period of the first part of the study has ended.



**Figure 32: Process of the Before-After study**

To ensure that the second part of the simulation starts at the correct moment in time the end-user needs to perform an extra simulation run. The first simulation run the end-user will perform is to determine the duration of the inclusion period of the first part. The output of the model gives the duration of the inclusion period in weeks. The end-user is able to enter the start time of the second part of the study. To do this, the end-user should double click on the module surrounded by the red square (figure 33). A pop-up will appear on the screen. There is a textbox on the right lower corner named “First Creation”. In this textbox the user should fill in the start time of the second part of the study in MINUTES. Thus if the total duration of the first part of the study is 50 weeks than the start time of the second part is  $50 \times 7$  (days)  $\times 24$  (hours)  $\times 60$  (minutes) = 504000. This number should be entered in the textbox named “First Creation”.

As already mentioned in the previous section the end-user is able to change the number of replications of a simulation run. We advise the user to change the number of replications of the experimental run to at least 300. The experimental simulation is performed to find out what the duration is of the first part, the other results of the experimental simulation run can not be used for analysis, therefore we advise a simulation run with less replications because this will take less time. Keep in mind that it is still necessary to perform enough replications for the simulation run, else the results are not reliable.

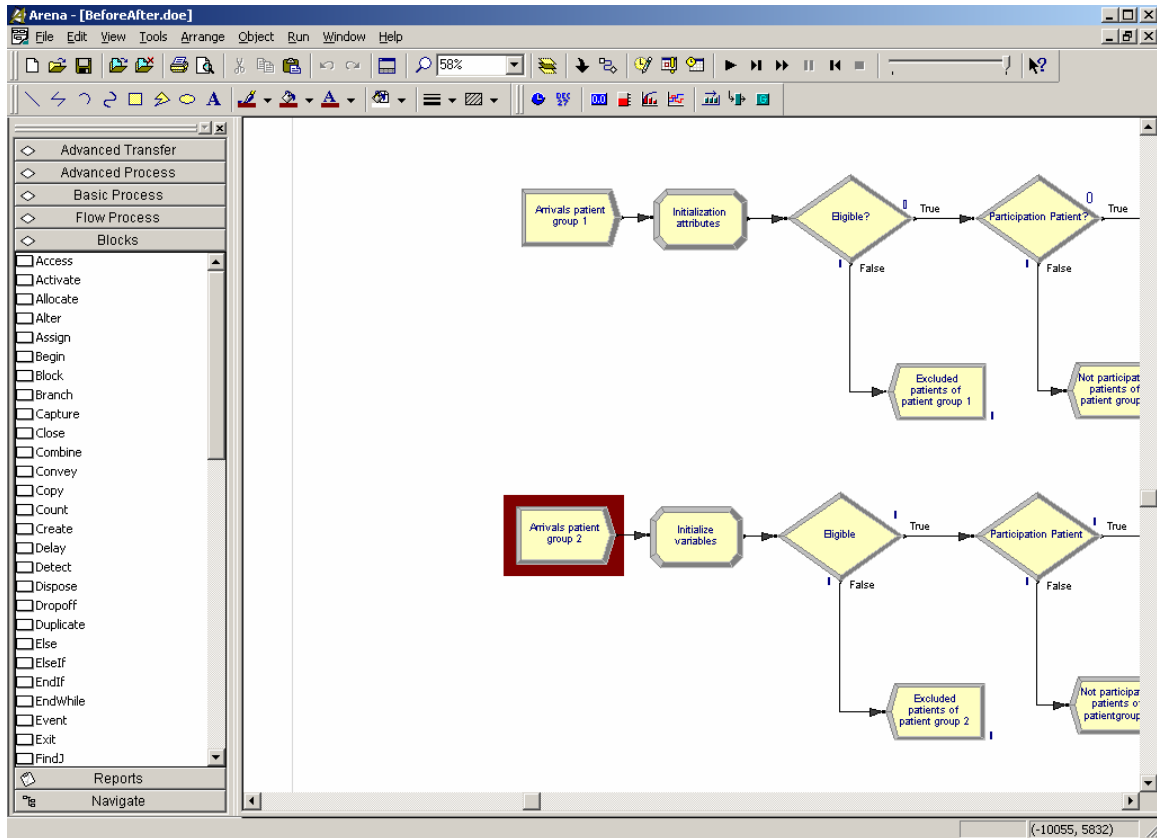


Figure 33: Screenshot of the Before-After model. The user is able to change the start time of the second part by double clicking on the module named "Arrivals patient group 2". The module is marked with a red square.

## 5 RCT-1 simulation model

The only difference between the input of the RCT-1 simulation model and the Cohort simulation model can be found in the input screen. The main difference between the two input screens is that the input screen of the RCT-1 simulation model does not contain an “Assignment of patients to tests”-frame and that it has one additional input parameter, the *block size*. The rest of the input screen is just the same.

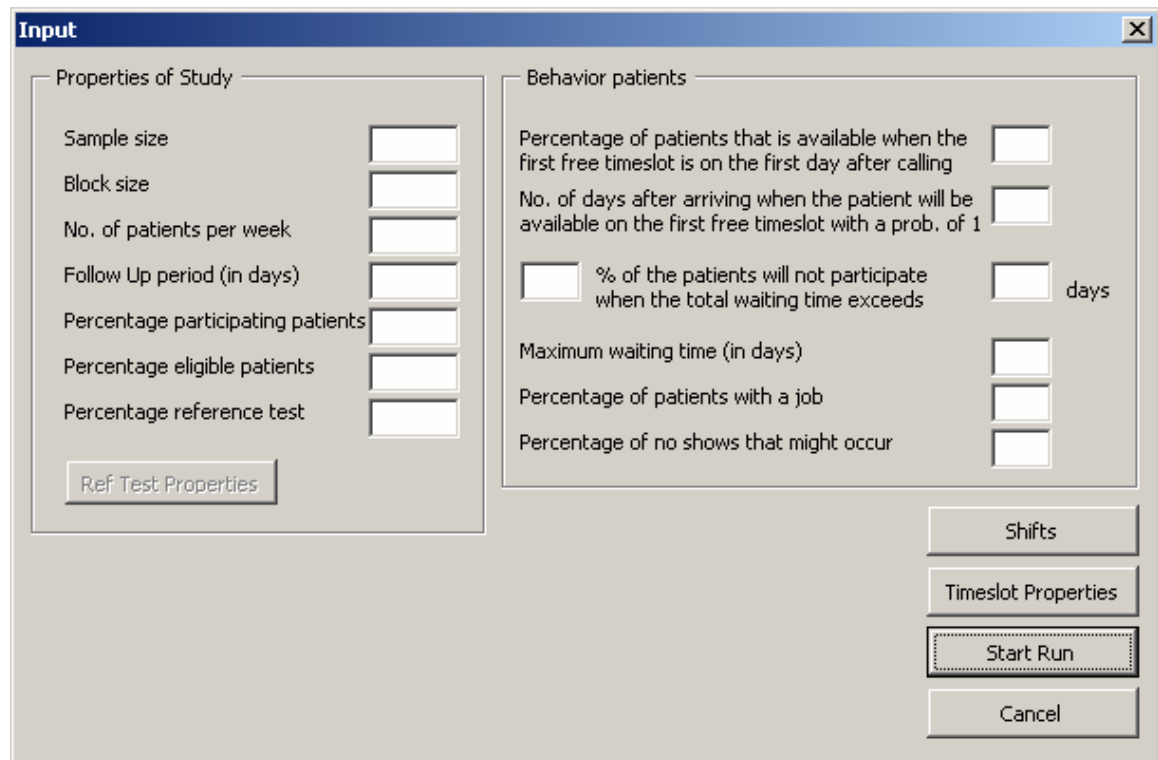


Figure 34: The Input screen of the RCT-1 simulation model

### Block size

In the RCT-1 study design patients are randomized across the diagnostic imaging modalities. A randomization scheme will be developed using a block size. The patients should be equally assigned to the diagnostic imaging tests. The block size indicates after how many patients the number of times a test is performed should be equal for both tests. For example, consider a block size of 8. This means that each test is performed exactly four times when the eighth patient has undergone the test that he or she should undergo. After the sixteenth patient, each test is performed 8 times.

The value for the block size should be larger than zero.

## 6 RCT-2 simulation model

The input for the RCT-2 simulation model is just like the input for the Cohort simulation model. The input screen of the RCT-2 simulation model contains one additional parameter; the *block size for randomization test results*. The rest of the input screens are just the same as those of the Cohort simulation model.

The next figure shows the input screen of the RCT-2 simulation model.

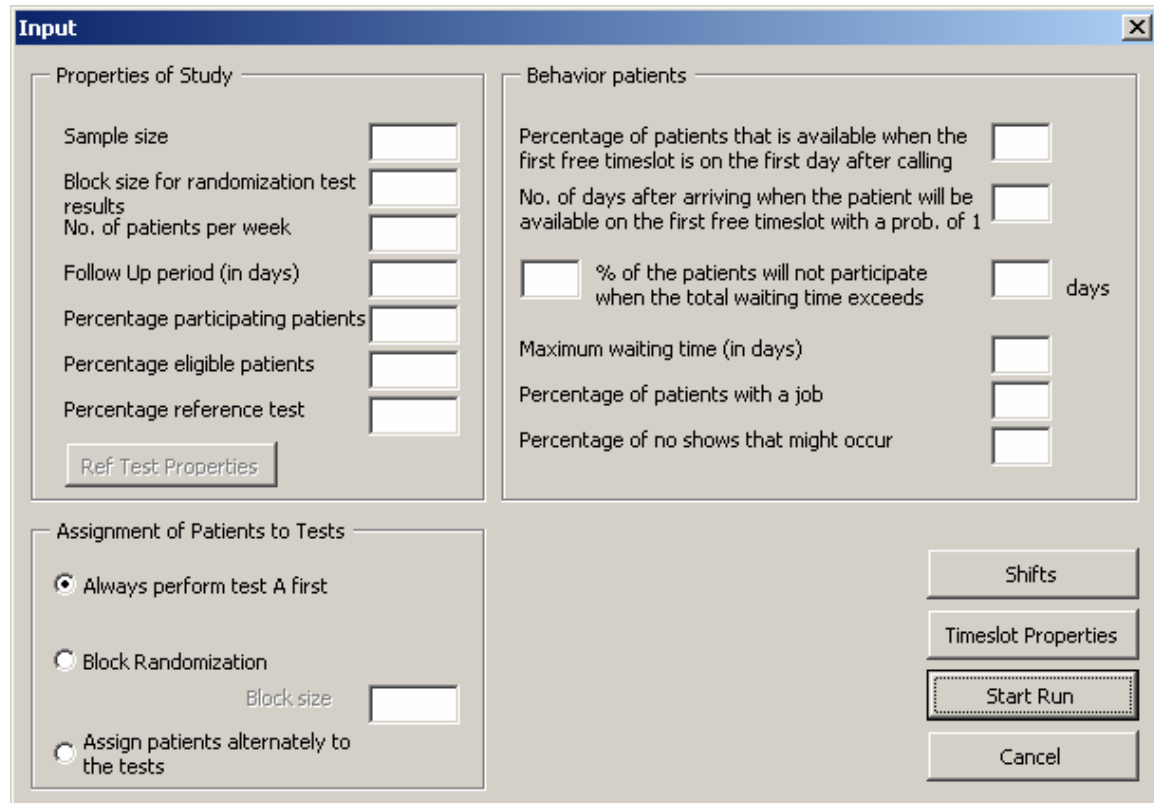


Figure 35: The Input screen of the RCT-2 simulation model

### Block size for randomization test results

The disclosure of test results is randomized in the RCT-2 study design. A randomization scheme will be developed to determine which test result should be provided to the physician. The randomization scheme will be developed using a block size. The test results should be equally assigned to the treating physician. The block size indicates after how many test results the number of disclosed test results of each type should be equal. The value for the block size should be larger than zero.



## 7 Output of simulation models

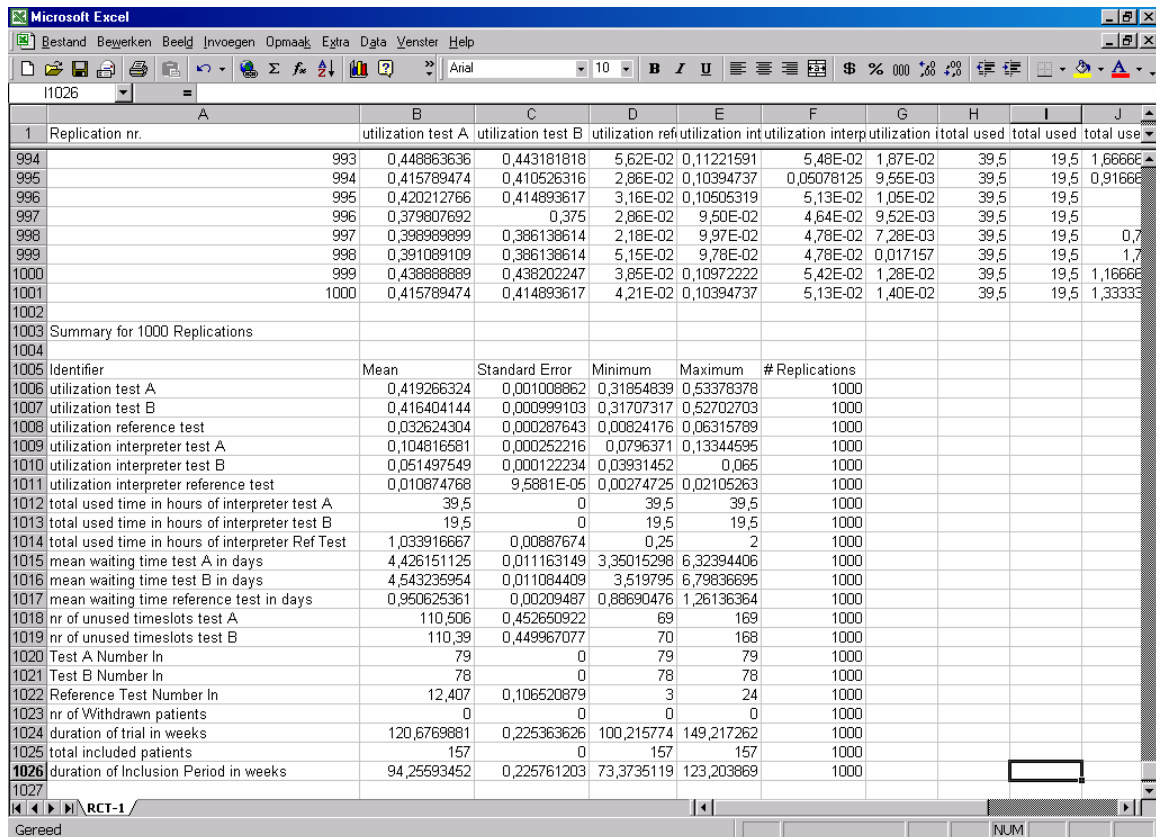
Arena writes the output of the simulation runs directly to a comma separated file. The name of the file is equal to the simulation model that is performed, for example the output file for the Cohort simulation model is Cohort.csv. The file is created at the beginning of the simulation process and is placed in the same folder where the program is located. When the simulation run is finished a message box will pop-up. This message box gives the path name of the folder where the output file is saved.

An example of output of the simulation model is given in figure 36. In one simulation run the model performs several replications. The number of replications that is performed is standard set to 1000, the end-user can change this value, as explained in Section 2.2. The output file consists of two parts. In the first part the values for the output parameters of each individual replication are listed. In the second part a summary of these output parameters is given.

The first part of the output starts in row 1 of the output file. The first row of the Excel sheet shows the names of the output parameters. The next following rows give the results of each of the output parameters for each replication that is performed. Thus the second row presents the results of replication 1, the third presents the results of replication 2 etc. The first column denotes the replication number.

For each replication the results of the following output parameters are given:

- Utilization test A
- Utilization test B
- Utilization reference test
- Utilization interpreter test A
- Utilization interpreter test B
- Utilization interpreter reference test
- Total used time in hours of interpreter test A
- Total used time in hours of interpreter test B
- Total used time in hours of interpreter Ref Test
- Waiting time test A in days
- Waiting time test B in days
- Waiting time reference test in days
- Nr of unused timeslots test A
- Nr of unused timeslots test B
- Test A Number In
- Test B Number In
- Reference Test Number In
- Nr of Withdrawn patients
- Duration of trial in weeks
- Total included patients
- Duration of Inclusion Period in weeks



1	Replication nr.	utilization test A	utilization test B	utilization ref	utilization int	utilization interp	utilization (total used)	total used	total use
994	993	0,448863636	0,443181818	5,62E-02	0,11221591	5,48E-02	1,87E-02	39,5	19,5
995	994	0,415789474	0,410526316	2,86E-02	0,10394737	0,05078125	9,55E-03	39,5	19,5
996	995	0,420212766	0,414893617	3,16E-02	0,10505319	5,13E-02	1,05E-02	39,5	19,5
997	996	0,379807692	0,375	2,86E-02	9,50E-02	4,64E-02	9,52E-03	39,5	19,5
998	997	0,398989899	0,386138614	2,18E-02	9,97E-02	4,78E-02	7,28E-03	39,5	19,5
999	998	0,391089109	0,386138614	5,15E-02	9,78E-02	4,78E-02	0,017157	39,5	19,5
1000	999	0,438888889	0,438202247	3,85E-02	0,10972222	5,42E-02	1,28E-02	39,5	19,5
1001	1000	0,415789474	0,414893617	4,21E-02	0,10394737	5,13E-02	1,40E-02	39,5	19,5
1003	Summary for 1000 Replications								
1005	Identifier	Mean	Standard Error	Minimum	Maximum	# Replications			
1006	utilization test A	0,419266324	0,001008862	0,31854839	0,53378378	1000			
1007	utilization test B	0,416404144	0,000999103	0,31707317	0,52702703	1000			
1008	utilization reference test	0,032624304	0,000287643	0,00824176	0,06315789	1000			
1009	utilization interpreter test A	0,104816581	0,000252216	0,0796371	0,13344595	1000			
1010	utilization interpreter test B	0,051497549	0,000122234	0,03931452	0,065	1000			
1011	utilization interpreter reference test	0,010874768	9,5881E-05	0,00274725	0,02105263	1000			
1012	total used time in hours of interpreter test A	39,5	0	39,5	39,5	1000			
1013	total used time in hours of interpreter test B	19,5	0	19,5	19,5	1000			
1014	total used time in hours of interpreter Ref Test	1,033916667	0,00887674	0,25	2	1000			
1015	mean waiting time test A in days	4,426151125	0,011163149	3,35015298	6,32394406	1000			
1016	mean waiting time test B in days	4,543235954	0,011084409	3,519795	6,79836695	1000			
1017	mean waiting time reference test in days	0,950625361	0,00209487	0,88690476	1,26136364	1000			
1018	nr of unused timeslots test A	110,506	0,452650922	69	169	1000			
1019	nr of unused timeslots test B	110,39	0,449967077	70	168	1000			
1020	Test A Number In	79	0	79	79	1000			
1021	Test B Number In	78	0	78	78	1000			
1022	Reference Test Number In	12,407	0,106520879	3	24	1000			
1023	nr of Withdrawn patients	0	0	0	0	1000			
1024	duration of trial in weeks	120,6769881	0,225363626	100,215774	149,217262	1000			
1025	total included patients	157	0	157	157	1000			
1026	duration of Inclusion Period in weeks	94,25593452	0,225761203	73,3735119	123,203869	1000			

Figure 36: Example of possible output of the simulation models

The second part of the output file consists of a summary of the total number of replications that is performed. In our case we performed 1000 replications.

In the summary the mean value, standard deviation, minimum and maximum value of each output parameter is presented. The total number of performed replications is also denoted; this is just for reminding the user of how many replications are done.

The summary part of the output is used for the cost calculations. More information about the costs calculations is provided in the Costs User Manual.

## 7.1 Output parameters

In this section we give a short description of each of the output parameters. The descriptions of several of the output parameters are brought together because these are almost equal despite the fact that another test modality or test interpreter is considered.

### Utilization test A, Utilization test B, Utilization reference test

The values for the utilization of the test modalities denote the percentages of the total reserved timeslots that is used for the study. These outcome measures give the user insight in the total capacity that is used by each type of test modality and whether it is better to reserve extra capacity.

---

**Utilization interpreter test A, Utilization interpreter test B, Utilization interpreter reference test**

The utilization of a test interpreter denotes the percentage of the planned time that the test interpreter spends on the study. The value for this output parameter gives the end-user an impression of the total time a test interpreter should reserve for the study.

**Total used time in hours of interpreter test A, Total used time in hours of interpreter test B, and Total used time in hours of interpreter Ref Test**

These output parameters denote the total time in hours that the interpreter of the test results participates in the study. The value of these outcome measures gains insight in the total extra work for the test interpreters as soon as the study is started.

**Waiting time test A in days, Waiting time test B in days, Waiting time reference test in days**

The waiting time per test represents the number of days a patient has to wait on average before it can undergo one of the tests. These outcome measures give the user insight in the total capacity that should be reserved to perform the study. The waiting time of the patients decrease when the number of reserved timeslots per test increases.

**Nr of unused timeslots test A, Nr of unused timeslots test B**

The nr of unused timeslots denotes the total number of reserved timeslots of test A and test B that were not filled during the inclusion period of the study. These output parameters are mainly used for the costs calculations.

**Test A Number In, Test B Number In, Reference Test Number In**

These output parameters represent the total number of patients that underwent each type of test.

**Nr of Withdrawn patients**

The nr of withdrawn patients denotes the total number of patients that withdraws from the study because of the long waiting times. Together with the utilization per test and the mean waiting time per patient this output parameter gains insight in the number of timeslots to reserve for each test to include as much patients in the study as possible.

**Duration of trial in weeks**

The duration of trial in weeks gives the total time in weeks between the start and end of the study. This parameter presents the expected duration of the study.

**Total included patients**

The total included patients denote the total number of patients that participated in the study. This value should be equal to the sample size that is given at the start of the simulation. This parameter value is mainly used for the costs calculations.

**Duration of Inclusion Period in weeks**

The duration of inclusion period denotes the total time in weeks that is needed to include the required number of patients in the study.

## 8 Costs calculations

In this final chapter we discuss the costs calculations that can be performed with the output of the Arena simulation models using the Excel file costs.xls.

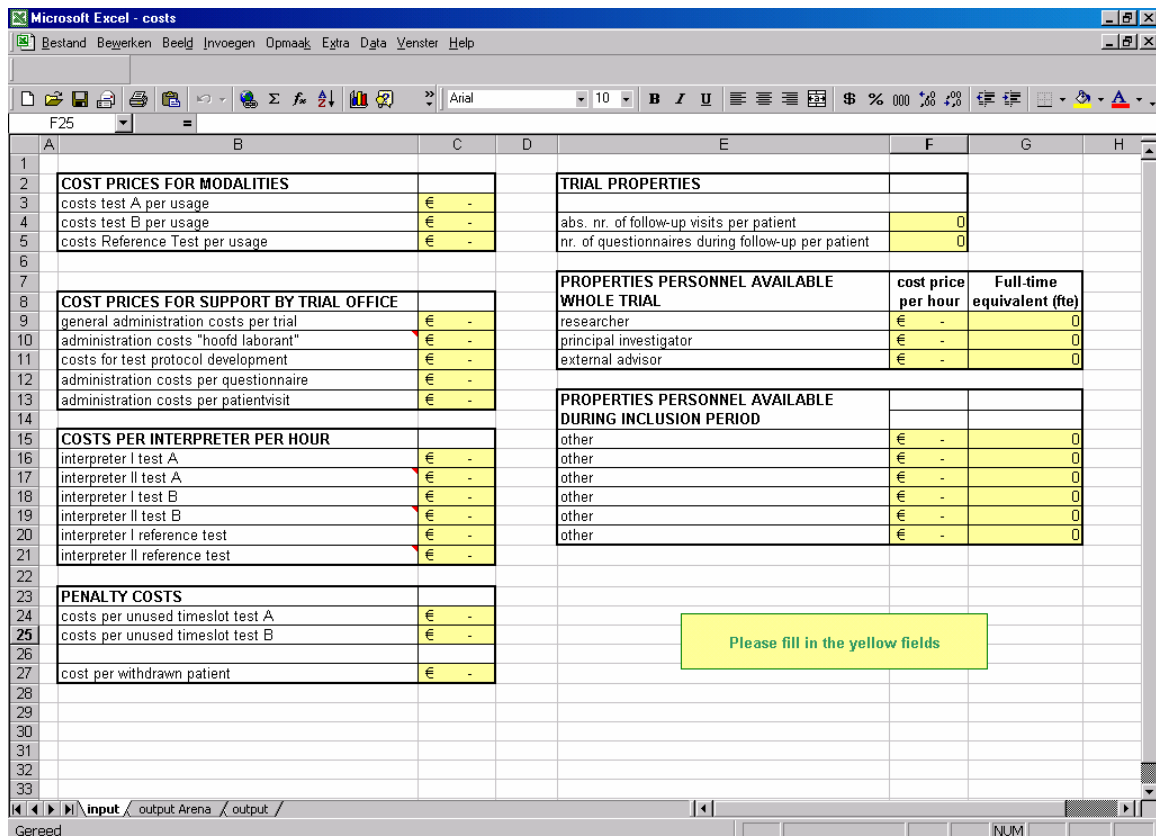
The costs.xls file consists of three worksheets; the input, output Arena and the output sheet. In the input and output Arena sheet the user is asked to enter the required data. The actual cost calculations are done in the output worksheet of the costs.xls file. We start with the description of the input worksheets followed by a description of the output.

### 8.1 Input worksheet

In the input worksheet the user is asked to fill in the yellow fields (figure 37). The data the user should provide can be categorized. We can distinct between the following categories:

- Cost prices for modalities
- Cost prices for support by trial office
- Costs per interpreter per hour
- Penalty costs
- Trial properties
- Properties personnel available whole trial
- Properties personnel available during inclusion period

We will discuss the input parameters per category. Most of the cost prices that should be entered in this worksheet can be found in the "Tarieven Radiologie 2005" file.



The screenshot shows the 'input' worksheet of the 'costs.xls' file in Microsoft Excel. The worksheet is organized into several sections with yellow input fields. A green callout box with the text 'Please fill in the yellow fields' is positioned over the bottom right area of the input fields.

A	B	C	D	E	F	G	H
1							
2	<b>COST PRICES FOR MODALITIES</b>			<b>TRIAL PROPERTIES</b>			
3	costs test A per usage	€ -		abs. nr. of follow-up visits per patient	0		
4	costs test B per usage	€ -		nr. of questionnaires during follow-up per patient	0		
5	costs Reference Test per usage	€ -					
6							
7							
8	<b>COST PRICES FOR SUPPORT BY TRIAL OFFICE</b>			<b>PROPERTIES PERSONNEL AVAILABLE WHOLE TRIAL</b>		<b>cost price per hour</b>	<b>Full-time equivalent (fte)</b>
9	general administration costs per trial	€ -		researcher	€ -	0	
10	administration costs "hoofd laborant"	€ -		principal investigator	€ -	0	
11	costs for test protocol development	€ -		external advisor	€ -	0	
12	administration costs per questionnaire	€ -					
13	administration costs per patientvisit	€ -					
14							
15	<b>COSTS PER INTERPRETER PER HOUR</b>			<b>PROPERTIES PERSONNEL AVAILABLE DURING INCLUSION PERIOD</b>			
16	interpreter I test A	€ -		other	€ -	0	
17	interpreter II test A	€ -		other	€ -	0	
18	interpreter I test B	€ -		other	€ -	0	
19	interpreter II test B	€ -		other	€ -	0	
20	interpreter I reference test	€ -		other	€ -	0	
21	interpreter II reference test	€ -		other	€ -	0	
22							
23	<b>PENALTY COSTS</b>						
24	costs per unused timeslot test A	€ -					
25	costs per unused timeslot test B	€ -					
26							
27	cost per withdrawn patient	€ -					
28							
29							
30							
31							
32							
33							

Figure 37: The input worksheet of the costs.xls file

### **Cost prices for modalities**

In the cost prices for modalities box the end-user is asked to enter the cost prices for each diagnostic imaging test. For general purposes we refer to the tests as test A, test B and the reference test, just like is done in the Arena simulation models. The user itself should keep track on which test is assigned as test A, B or the reference test.

### **Cost prices for support by the Trial office**

#### *General administration costs per trial*

General administrations costs are the administration costs that are made for each new study or trial that is started.

#### *Administration costs "hoofd laborant"*

The administration costs for the "hoofd laborant" are the personnel costs that are made by the "hoofd laborant" for the test protocol development. The test protocol development contains testing the settings of the diagnostic imaging modalities.

These administration costs can be found in the "Tarieven Radiologie 2005" file, named as "Administratie algemeen toestelhoofd".

#### *Costs for test protocol development*

The costs for testing the settings of the diagnostic imaging modalities can be split up in personnel costs and standard costs. The costs for test protocol development are the standard costs. These costs consist of the regular cost price of the diagnostic imaging modalities and an additional charge. The additional charge can be found in the "Tarieven Radiologie 2005" and is named "testscan".

#### *Administration costs per questionnaire*

The administration costs per questionnaire consist of the costs that are made for the development of a questionnaire that will be provided to the patients during the follow-up period of the study.

#### *Administration costs per patient visit*

These costs denote the administration costs that are made each time patients are in the hospital to undergo a test. The same administration costs are made when patients need to come to the hospital for the follow-up.

### **Costs per interpreter per hour**

In this box the user should give the personnel costs per hour of each interpreter. If one person interprets all the test results, thus the test results of test A, test B and the reference test, than the user should fill the personnel costs per hour of this particular person in the cells of "Interpreter I test A", "Interpreter I test B" and "Interpreter I reference test". It may be possible that more than one person interprets the results of a test. The end-user is also able to give the cost price of a second test interpreter. If there is only one test interpreter the end-user should enter 0.

### **Penalty costs**

The user is able to bring in charge extra costs when reserved timeslots remain unfilled.

### **Trial properties**

During follow-up patients sometimes need to come to the hospital or need to fill in questionnaires. These activities result in extra administration costs of the Trial office. Therefore we keep track of the number of patient visits and the number of questionnaires that should be filled in by the patient.

### **Properties personnel available whole trial**

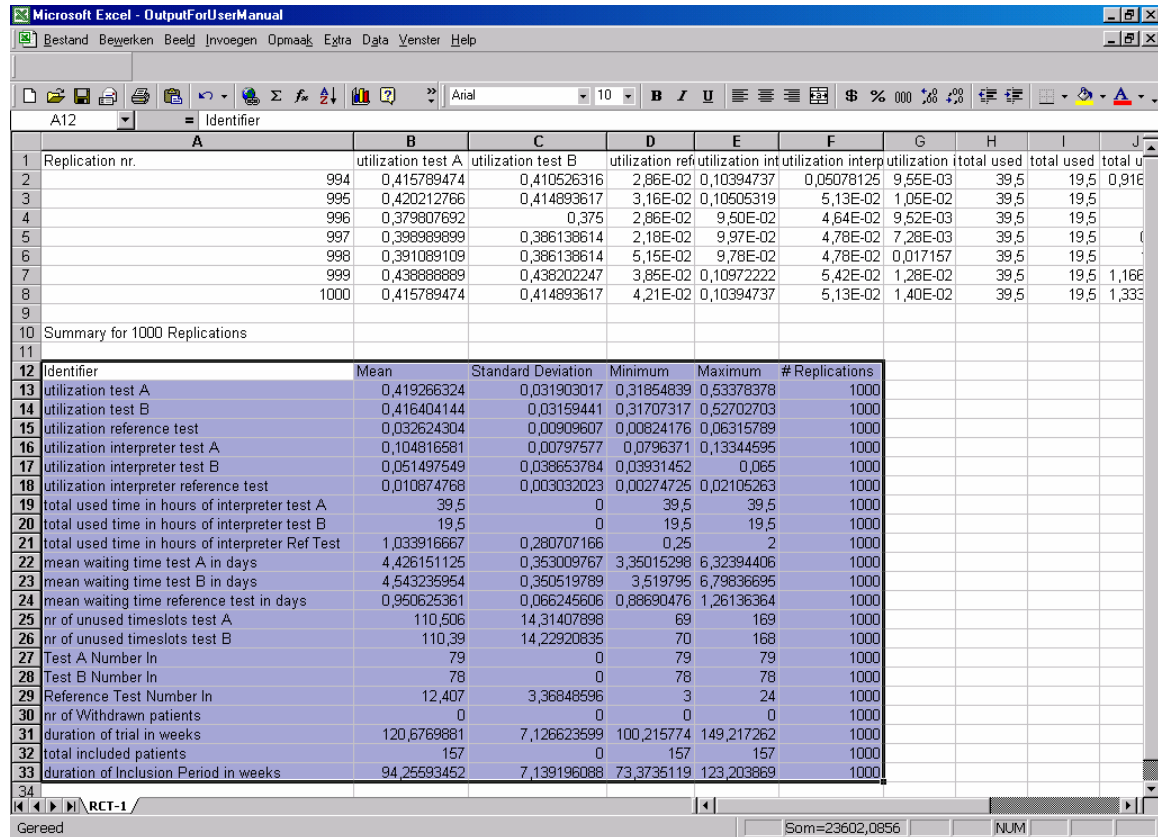
In this box the user enters the cost price per hour and the expected full time equivalent (fte) of the personnel that is involved in the study from the start to the end, like the principal investigator, the researcher and an external advisor.

### **Properties personnel available during inclusion period**

It might be possible that the researcher is assisted with the inclusion of patients in the study. The cost prices of the personnel that assist the researcher during the inclusion period can be filled in this box. The user should also fill in the full time equivalent (fte) of these personnel. The full time equivalent is 0.1 when a person is involved in the study for 10% of the time of a full time workweek.

## 8.2 Output Arena Worksheet

In the output Arena worksheet the user only has to copy the summary of the output file of the Arena simulation. Figure 38 gives an example of output from the Arena simulation model. The selected part is the part that should be copied in cell A1 of the output Arena sheet of the costs.xls file.



12	Identifier	Mean	Standard Deviation	Minimum	Maximum	# Replications
13	utilization test A	0,419266324	0,031903017	0,31854839	0,53378378	1000
14	utilization test B	0,416404144	0,03159441	0,31707317	0,52702703	1000
15	utilization reference test	0,032624304	0,00909607	0,00824176	0,06315789	1000
16	utilization interpreter test A	0,104816581	0,00797577	0,0796371	0,13344595	1000
17	utilization interpreter test B	0,051497549	0,036653784	0,03931452	0,065	1000
18	utilization interpreter reference test	0,010874768	0,003032023	0,00274725	0,02105263	1000
19	total used time in hours of interpreter test A	39,5	0	39,5	39,5	1000
20	total used time in hours of interpreter test B	19,5	0	19,5	19,5	1000
21	total used time in hours of interpreter Ref Test	1,033916667	0,280707166	0,25	2	1000
22	mean waiting time test A in days	4,426151125	0,353009767	3,35015298	6,32394406	1000
23	mean waiting time test B in days	4,543235954	0,350519789	3,519795	6,79836695	1000
24	mean waiting time reference test in days	0,950625361	0,066245606	0,88690476	1,26136364	1000
25	nr of unused timeslots test A	110,506	14,31407898	69	169	1000
26	nr of unused timeslots test B	110,39	14,22920835	70	168	1000
27	Test A Number In	79	0	79	79	1000
28	Test B Number In	78	0	78	78	1000
29	Reference Test Number In	12,407	3,36848596	3	24	1000
30	nr of Withdrawn patients	0	0	0	0	1000
31	duration of trial in weeks	120,6769881	7,126623599	100,215774	149,217262	1000
32	total included patients	157	0	157	157	1000
33	duration of Inclusion Period in weeks	94,25593452	7,139196088	73,3735119	123,203869	1000

Figure 38: Output of Arena simulation model. The selected part of the output is used as input in the Output Arena worksheet of the file costs.xls

Figure 39 shows the output Arena worksheet after the selected data of figure 38 is copied in cell A1. The only thing the end-user had to do is copy the results of the Arena simulation in the worksheet. It is very important that the heading is copied as well, thus the first row of the output Arena worksheet consists of the following data: Identifier, Mean, Standard Deviation, Minimum, Maximum, # Replications.

The copying is very strict. The results of the simulation model should exactly be copied starting in cell A1, if we do not to this the cost calculation will not be performed properly and the output of the costs file is not correct.

Microsoft Excel - costs

Bestand Bewerken Beeld Invoegen Opmaak Extra Data Venster Help

B29

1	Identifier	Mean	Standard Devi	Minimum	Maximum	# Replications	G	H	I	J
2	utilization test A	0,419266324	0,031903017	0,318548387	0,533783784	1000				
3	utilization test B	0,416404144	0,03159441	0,317073171	0,527027027	1000				
4	utilization reference test	0,032624304	0,00909607	0,008241758	0,063157895	1000				
5	utilization interpreter test A	0,104816581	0,00797577	0,079637097	0,133445946	1000				
6	utilization interpreter test B	0,051497549	0,038653784	0,039314516	0,065	1000				
7	utilization interpreter reference test	0,010874768	0,003032023	0,002747253	0,021052632	1000				
8	total used time in hours of interpreter test A	39,5	0	39,5	39,5	1000				
9	total used time in hours of interpreter test B	19,5	0	19,5	19,5	1000				
10	total used time in hours of interpreter Ref Test	1,033916667	0,280707166	0,25	2	1000				
11	mean waiting time test A in days	4,426151125	0,353009767	3,350152983	6,323944059	1000				
12	mean waiting time test B in days	4,543235954	0,350519789	3,519795004	6,798366947	1000				
13	mean waiting time reference test in days	0,950625361	0,066245606	0,886904762	1,261363636	1000				
14	nr of unused timeslots test A	110,506	14,31407898	69	169	1000				
15	nr of unused timeslots test B	110,39	14,22920835	70	168	1000				
16	Test A Number In	79	0	79	79	1000				
17	Test B Number In	78	0	78	78	1000				
18	Reference Test Number In	12,407	3,36848596	3	24	1000				
19	nr of Withdrawn patients	0	0	0	0	1000				
20	duration of trial in weeks	120,6769881	7,126623599	100,2157738	149,2172619	1000				
21	total included patients	157	0	157	157	1000				
22	duration of Inclusion Period in weeks	94,25593452	7,139196088	73,3735119	123,203869	1000				
23										
24										
25										
26										
27										
28										
29	The summary output of the Arena simulation should be copied into this sheet, starting in cell A1. The headings of the summary should be copied too									
30										
31										
32										
33										
34										

input output Arena output / Gereed NUM

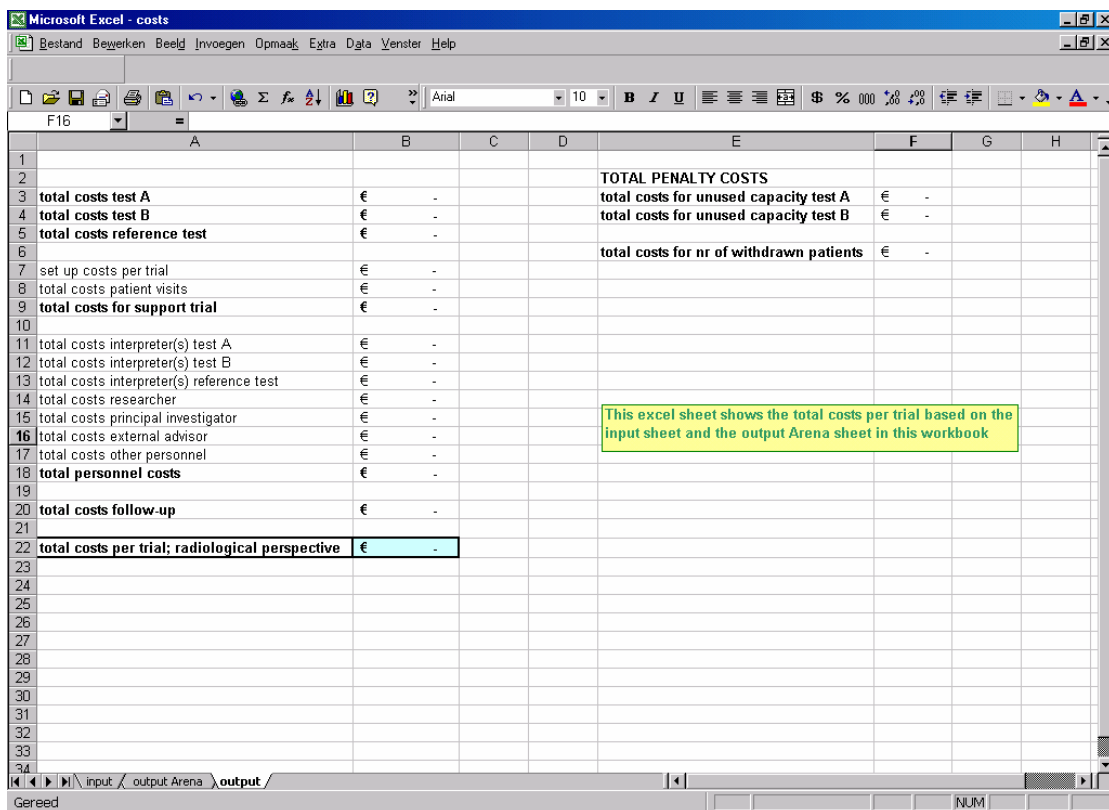
Figure 39: The output Arena worksheet with the summary of the output as selected in figure 38.



### 8.3 Output

The results of the cost calculations are given in the output sheet of the costs.xls file. The output presents the following costs (fig. 40):

- Total costs test A
- Total costs test B
- Total costs reference test
- Total costs for support trial
- Total personnel costs
- Total costs follow-up
- Total costs for unused capacity of the test
- Total costs for nr of withdrawn patients
- Total costs per trial: radiological perspective



	A	B	C	D	E	F	G	H
1								
2					TOTAL PENALTY COSTS			
3	total costs test A	€ -			total costs for unused capacity test A	€ -		
4	total costs test B	€ -			total costs for unused capacity test B	€ -		
5	total costs reference test	€ -						
6					total costs for nr of withdrawn patients	€ -		
7	set up costs per trial	€ -						
8	total costs patient visits	€ -						
9	total costs for support trial	€ -						
10								
11	total costs interpreter(s) test A	€ -						
12	total costs interpreter(s) test B	€ -						
13	total costs interpreter(s) reference test	€ -						
14	total costs researcher	€ -						
15	total costs principal investigator	€ -						
16	total costs external advisor	€ -						
17	total costs other personnel	€ -						
18	total personnel costs	€ -						
19								
20	total costs follow-up	€ -						
21								
22	total costs per trial; radiological perspective	€ -						
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								

Figure 40: Screenshot of the output worksheet of the costs.xls file

#### Total costs test A

The total costs test A denote the total costs that are made to let the patients undergo test A.

#### Total costs test B

The total costs test B denote the total costs that are made to let the patients undergo test B.

#### Total costs reference test

The total costs reference test denote the total costs that are made to let the patients undergo the reference test.

---

### **Total costs for support trial**

The total costs for support trial consist of the set up costs for the trial and the trial costs of all patient visits. The set up costs contain the personnel and administration costs for the protocol development and the general administration costs that are made for each study.

The trial costs of all patient visits contain the administration costs for each time a patient visits the hospital to undergo a test.

### **Total personnel costs**

The total personnel costs consist of all the personnel costs that are made during the trial. The personnel costs include the costs for the test interpreters, the researcher, the principal investigator, the external advisor and all the other personnel that spend time on the study.

### **Total costs follow-up**

The costs for the follow-up period consist of the administration costs for the trial office for processing all the questionnaires filled in by the patients and the administration costs for each patient visit that takes place during the follow-up period.

### **Total costs for unused capacity**

Each time a reserved timeslot remains unfilled costs the Department of Radiology an amount of money. Charging an amount of money for each time a timeslot is not filled can visualize the losses the department might have. The total costs for unused capacity denote for each test the total costs if an additional charge is done for each unused timeslot.

### **Total costs for nr of withdrawn patients**

Each patient that withdraws from the study because of long waiting times is bad publicity for the hospital. To visualize the impact of withdrawn patients the user is able to charge each time a patient withdraws. The total costs for nr of withdrawn patients denotes the total costs that are made when a certain number of withdrawn patients occur. The number of withdrawn patients is given by the Arena output.

### **Total costs per trial: radiological perspective**

This is the most important parameter of the sheet. This parameter value gives the total costs that are expected to be made if a study is performed.

## Appendix D – Arena Models