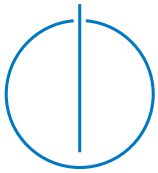
FAKULTÄT FÜR INFORMATIK

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Engineering Science

Design and Evaluation of a Vascular Simulator for Medical Education

Liesa Weigert



FAKULTÄT FÜR INFORMATIK

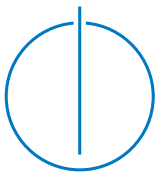
TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Engineering Science

Design and Evaluation of a Vascular Simulator for Medical Education

Design und Evaluation eines Vaskulären Simulators für Medizinische Ausbildung

Author:	Liesa Weigert
Supervisor:	Prof. Dr. Nassir Navab
Advisor:	Philipp Stefan
Submission Date:	September 30, 2015



I assure the single handed composition of this bachelor's thesis in engineering science only supported by declared resources.

Munich, September 30, 2015 Liesa Weigert

Acknowledgments

I would like to thank Dr. Reza Ghotbi, head of vascular surgery at Helios Kliniken, for dedicating his time and expertise to this project.

Furthermore, I would like to thank Dr. Martin Greif and Dr. Nikolaus Sarafoff of the Medizinische Klinik und Poliklinik in Munich for their help.

Special thanks also goes to Loïse Ulrich for the collaboration on this project.

Abstract

The objective of this thesis was to build a vascular simulation device for assessment based on previously acquired knowledge and to design a validation process for said assessment.
Although simulation is becoming more significant in surgical education, the validation process seldom adheres to the current framework.
We incorporated the contemporary unitary framework of validity into the development and the evaluation of the simulation tool. Domain knowledge on endovascular procedures was acquired utilizing Cognitive Task Analysis and the Think-Aloud technique. A blueprint of endovascular procedures, including data on the sub steps and complications, was created and used as a base for the simulation design. The outline for an assessment process with a discussion on the inclusion of the five sources of validity evidence was created. The prototype of the simulator and the assessment process should be revised after the evaluation to incorporate any elaborated enhancements.

Das Ziel dieser Arbeit war es, basierend auf davor erarbeitetem Wissen, einen vaskulären Simulator zur Beurteilung von Chirurgen zu bauen und einen entsprechenden Validationsprozess zu entwerfen.
Obwohl Simulation in der chirurgischen Ausbildung und Bewertung einen immer wichtigeren Stellenwert einnimmt, entspricht der Validationsprozess nur selten den gegenwärtigen Vorgaben.
Wir haben den zeitgemäßen einheitlichen Ansatz von Validität in die Entwicklung und die Evaluation des Simulators als Beurteilungsmethode miteinbezogen. In dieser Arbeit wurde eine Zusammenfassung von vaskulären Eingriffen entworfen, die Daten über die Zwischenschritte und Komplikationen enthält. Zudem wurde ein Entwurf für einen Validierungsprozess erstellt und die Angabe von Beweisen aus allen fünf Quellen der modernen Theorie diskutiert. Der Prototyp des Simulators sollte nach dem Bewertungsprozess überarbeitet werden, um alle erarbeiteten Verbesserungen einzubringen.

Contents

Part I.

Introduction and Theory

1. Introduction

Surgical procedures are traditionally taught through the apprenticeship model in which the trainee's education is guided by a practicing surgeon proficient in his field. The mentee learns procedures by initially observing them and discussing the diagnosis, procedure and possible complications with his or her mentor. Thereupon they move on to assisting the performing surgeon and gradually advance to taking over parts of the surgery themselves under supervision. Only after satisfactorily passing these steps do the resident surgeons move on to performing the operation themselves. However this model has some significant drawbacks.

1.1. Disadvantages of the Apprenticeship Model

The education through a mentor-mentee relationship is exceedingly time consuming and therefore occupies a massive amount of work hours for both the expert surgeon and the trainee. However, as the European Working Time Directive (EWTH) ensures to reduce working hours throughout Europe, this time might not be sufficiently available in the future[22]. In fact, the reductions have let to a considerable decrease in hours that can be spent on practicing for surgeons in training[14]. Consequently the question arises, whether the quality of training will suffer from this development.
Furthermore, the current teaching process is often criticized for not being standardized. The proficiency level trainees can achieve is strongly dependent on the quality of interaction with their mentor and the availability of cases in their hospital. Nevertheless, although the apprenticeship model has been criticized as being inefficient and unpredictable, there is little regulation on the assessment of performance. To ensure an adequate surgical training, the uniformity of training and the evaluation of post-training skills has to be equalized and controlled[13][1].
Another predicament of this training method is the inevitable increase of risk of patient harm or discomfort, as the junior surgeon has to perform the procedure on a real patient ab initio. Although there is a general consensus in the area of skill acquisition that proficiency cannot be achieved exclusively by visual observation of a task, there are little to no options of practicing the procedure outside of actual surgery. The lack of other opportunities to train jeopardizes patient safety and creates a stressful and possibly detrimental work environment[14] [22].

Some of these deficiencies can be compensated through the integration of simulation devices into medical education.

1.2. Medical Simulation

The origin of simulation in complex work environments with high safety standards lies in aviation. Similar to the surgical field, pilots are required to handle a variety of complications and intricate operational sequences while minimizing the risk potential. However, simulation technology is widely accepted and adopted in aviation and therefore represents an essential part of the teaching process, whereas the medical field has not fully incorporated these devices yet[23]. Simulators in aviation feature a more life-like surrounding as well as an abundance of typical and exceptional scenarios, which still have to be cultivated in medicine [20]. Like aviation, medicine can benefit from exploiting the advantages of simulators.
For instance, simulators offer the opportunity to create training modules that cover a variety of different procedures, cases and even complications. This way, the trainees can learn from a predefined assortment instead of the random cases dependent on the patients in their hospital. The versatile offer allows students to try several approaches, learn from their mistakes without harming a patient and to experience uncommon cases in a low-stress and predetermined environment. Through focused training sessions, direct feedback and progress tracking the learning progress can be made more effective and predictable. This benefits the trainees and makes it possible to assess their improvement in a standardized manner[7][17].

1.3. Proposed Project

While simulation is beneficial for the education and assessment of all surgical procedures, minimally invasive procedures, such as interventional radiology or cardiology are particularly well suited to be simulated, because the performance in these operations is eminently influenced by the surgeon's ability to interpret a two-dimensional image, e.g. fluoroscopy or ultrasound.[13][11] These procedures can be modeled through a combination of 2D visual representation and haptic feedback, unlike open surgery, which is more challenging due to its 3D work environment[23]. Moreover, many movements in these processes are hard to entirely comprehend, as they are only perceptible through the imaging and the outside manipulation of the instruments by the surgeon. The main clue to reenact the movements, the haptic feedback, is only accessible to the trainees in practice. To minimize the potential harm for patients, simulation can offer the experience of these tactile responses without risking harm[14].

Endovascular surgery is particularly well fit to be simulated, as it is a potentially life-saving procedure that is mainly led by angiography and fluoroscopy.
The objective of this project was to simulate these procedures without the use of actual radiation, as this also poses a threat to the performing surgeon. If the operations, such as percutaneous transluminal angioplasties, can be practiced beforehand both the surgeon and the patient can profit from the gain in efficiency as it may lead to a decrease in radiation dose.
We will discuss the design and the validation of the simulator in the following.

2. Related Work

Because of the high risk potential in the medical field the requirements for simulators are particularly high. To assure that the simulation can actually benefit or even partly replace the traditional surgical education and skill assessment every part of the development and testing process has to be thoroughly documented and analyzed. Therefore there are several fields of research to consider when aspiring to create a simulation device which is able to fulfill these tasks. We will discuss some of these concepts specifically in the following.

2.1. Cognitive Task Analysis

The challenge to consider, especially in medicine, is teaching intricate and lengthy procedures as a sequence of steps while acknowledging the underlying thought process. While different approaches and instruments can be memorized and taught verbally, the actual performance of the procedure, including the cognitive processes, is more challenging to transfer. This "procedural knowledge how" is acquired through extensive practice. According to Lanzer et al.[20] the transfer of this knowledge is currently highly dependent on the quality of interaction between the student and his or her mentor. This dependency can be neutralized by accessing the underlying cognitive processes and implementing the relationship of tactile cues and appropriate behavior into the workflow of a simulator[2].

2.1.1. Method

Cognitive Task Analysis (CTA) describes a collection of techniques designed to elicit knowledge from experts. It is typically divided into a knowledge elicitation and an analysis and representation phase. While the progress in the second phase is dependent on the preceding work, there are several methods of knowledge extraction to consider in the first stage[5][2].
Commonly used approaches include semi-structured or structured interviews. These can be accompanied by data elicitation from literature to collect substantial domain knowledge of the procedure in question. Special emphasis has to be put on the integrity of steps including exhaustive descriptive data[5]. Important details include required

information, tactile clues and possible mistakes or complications. To identify the risks in particular, one can also have the experts describe past critical incidents and the skills and patterns they used to salvage the situation[2].
Some researchers also recommend the identification of automated steps. These actions are described as requiring little cognitive involvement for experienced surgeons and are often overlooked in traditional education [26].
This kind of data acquisition can be supplemented by observing and videotaping the procedure[14].

2.1.2. Think-Aloud

As a final step in surgical simulator development, research ordinarily tries to prove that the developed tool can actually perform the desired task. For training simulators this task is to have a beneficial effect on the improvement of a surgeon's skill. This assumption can generally be established through expert opinion or transferability studies. As long as the simulation does not teach wrong techniques or habits, which can be prevented through an expert evaluation, the use of said tool will not have a negative impact on the surgeon's education [16].
In contrast, the use of simulators for assessment could potentially have significant consequences on both surgeons and patients. Wrongful passes or failures in such a test due to mistakes in the simulation development and evaluation could lead to the certification of unfit surgeons, bad patient care and various other unforeseen repercussions [16].
Therefore researchers should refer to scientific frameworks to provide evidence to support the accuracy of a test. The theory behind this process, which is referred to as validation, should be carefully examined and the resulting inferences should be included in both the development and the assessment process of the simulation tool.
Research generally aims to show that the developed simulation tool is reliably able to distinguish the participants' skill levels. For this purpose several participants with a background in the specific field are asked to perform a number of tasks on the simulator and evaluated by analysis of their performance regarding the predefined metrics. Afterwards the dependency of scores and actual skill level is assessed to determine whether the simulator is a sufficient assessment tool.
However, this process seldom adheres to predefined requirements to prove its significance. Although the *Standards for Educational and Psychological Testing (The Standards)* include definite information on currently accepted validity concepts, this data is insufficiently represented in literature[16].
We will discuss the theory and implementation of the contemporary framework in the following.

2.2. Validity

Understanding the concepts and goals behind validation is a crucial part of creating a sensible and conclusive assessment. Without validation, the simulator's scores are meaningless. Therefore research dictates certain characteristics to consider during the validation process. Validity itself is defined as

> "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests"[24].

This means, it describes how dependable test scores concerning a specific intended goal truly are.
Although it seems most instinctive to validate an instrument, such as a simulator, as a whole through a predefined standardized method, the actual process is more fluid.
Firstly, it is not the simulator itself which gets validated but its propriety to assess a certain value, such as the surgeon's psycho-motor or cognitive skills concerning a specific task. This abstract concept represents the hypothesis the validation is based on. Analogous to other fields of science, this hypothesis is tested and the accumulated evidence is then used to either support or refute the underlying theory. Therefore validity is not a dichotomy, but a series of tests and evidence leading to revision and refinement of the hypothesis [3] [6].
Validity has traditionally been divided into content, criterion and construct validity. However the current theory has combined these arbitrary distinctions to the overarching framework of construct validity. This approach accentuates the dependence of hypothesis and instrument scores[24][3].

2.2.1. Sources of Evidence

Within the current unitary concept of construct validity five sources of evidence are identified to support the construct: content, response process, internal structure, relation to other variables and consequences. Evidence from these categories should be acquired individually and combined to create an extensive representation. Each category is associated with a variety of exemplary types of evidence (see Table 2.1) [3].

The five sources of evidence are explained in the following:

- **Content.** This, most essential, type of evidence should show that the content of the test is truthfully depicting the intended construct. Therefore, the steps taken to ensure that the test scores represent the construct have to be thoroughly documented and evaluated.

- **Response Process.** Is has to be shown that all possible sources of errors within the test administration were minimized to assure data integrity. This can be done by analyzing the thought processes of the participants or observers of the test to guarantee the appropriateness of the test methods.
- **Internal Structure.** The internal structure of the assessment process relates to the reliability and generalizability of the test. This means, the test should still yield the same results when reproduced under different circumstances, such as with other participants.
- **Relationship to Other Variables.** The test can be validated by comparing it to an existing measure with ascertained accuracy.
- **Consequences.** The possible consequences of the intended or unintended outcomes of the test results on the students and/or society have to be carefully evaluated [6][3].

To fully comprehend the implementation of the framework into the correct design and evaluation of a simulator, researchers should familiarize themselves with positive examples from literature [16].

2.2.2. Current Situation

While the definition of validity through *the Standards* seems unambiguous, the actual realization of this framework in research aiming to validate simulation tools greatly varies. A paper by Korndorffer Jr et al. [16] from 2010 revealed that only 23 percent of the examined studies on validation in the laparoscopic simulator education partly or fully used the contemporary framework. This image is supported by Cook et al. [4], who analyzed papers which 'evaluated the validity of simulation-based assessment scores using two or more evidence sources'. They found that only 3 percent of these papers referenced the five sources of validity evidence and 24 percent made no reference to any validity framework at all.
These statistics show that the assessment of simulation-based tests lacks the execution of a scientific validation process. Surgical research seems to have developed its own definition of validation in isolation of scientific advancements, which greatly differs from *the Standards*. Instead, they often rely on 'face validity', that means on whether the test scores seem valid according to experts or based on the analysis of the results [9].
However, the use of this type of validity source is discouraged in the contemporary framework, as it has no scientific foundation. For instance, validity is often claimed after the test scores show a significant difference between novices and experts. As these categories of participants frequently differ from having none or little experience to

being proficient in the task, these scores have no real implication. To conclude that the test can accurately depict a surgeon's skill level would be like concluding a math test is valid after comparing the scores of fifth graders and college professors [16].
As the contemporary framework is seldom applied in surgical education literature, there have been several papers published to contribute to the proper implementation into research. The consensus is that researchers should use positive examples as a starting point to base their validity efforts on and abandon outdated validation criteria in favor of the contemporary unitary framework [25] [16].

Table 2.1.: Description and examples for the five sources of evidence for validity [3][6].

Content	Response Process	Internal Structure	Relationship to other variables	Consequences
• Examination blueprint • Test specifications • Quality of test questions • Item writer qualifications • Sensitivity review	• Quality control of electronic scanning/ scoring • Student format familiarity • Key validation of preliminary scores	• Item analysis data: - Item difficulty/ discrimination - Item-total correlations • Standart errors of measurement: - Generalizability - Dimensionality	• Correlation with other relevant variables • Convergent correlations - similar tests • Test-criterion correlations	• False positives/ negatives • Reasonableness of method of establishing pass-fail score

Part II.

Knowledge Acquisition

3. Knowledge Elicitation

Our first goal in this project was to obtain comprehensive domain knowledge on endovascular interventions through Cognitive Task Analysis. This information was both necessary to create a realistic simulation device as well as evaluate the specific demands and idiosyncrasies of the field.
To obtain the sufficient knowledge we completed several semi-structured interviews with an expert in the field of vascular surgery. These sessions ranged from 30 to 90 minutes and were consistently audiotaped and retrospectively transcribed. The interviews were supplemented by observation and videotaping of assorted surgeries.

3.1. Initial Semi-Structured Interview

The first session's main objective was to get a comprehensive overview of endovascular procedures in general, as well as establish necessary requirements for the simulator. Therefore the interview was divided into two parts.
The first part was a constructive conversation on the prerequisites and possible areas of application of vascular simulation. The main focus was put on the exchange of expectations and possibilities to weigh them and set a mutually agreed upon goal.
The second part was a CTA closely adhering to the approach described by Tjam et al.[26]. Therefore, the expert was asked to go over a sequence of procedural steps derived from literature and correct any mistakes or ambiguities. After agreement on the main steps of the procedure was reached, these steps were divided into sub steps and separately analyzed, adhering to the following questions:

- Can this step be divided into sub-steps?
- Is this step automatic?
- What materials are needed?
- Which complications are likely to occur in this step? How can you prevent them?

Additionally, the expert was asked to suggest metrics the skill assessment with the simulator could be based on.

3.2. Surgery Observation

Thereafter several surgical procedures were observed to acquire more knowledge about the surgical working environment and workflow. One procedure, a percutaneous transluminal angioplasty (PTA) was singled out in particular for videotaping and further analysis. The PTA was well fit to exemplify the general procedure during endovascular interventions, as the steps are generally comparable.
Although the classic Think-Aloud method would likely have been most efficient to elicit knowledge during the procedure, the circumstances demanded a less distracting approach. Although the theory behind this method dictates an immediate verbalization of thought processes, some researchers suggest that the verbalization can also be recorded in retrospect instead of concurrently. Therefore we decided not to force immediate and constant 'thinking-aloud' to avoid compromising patient safety.
As suggested by Lundgrén-Laine and Salanterä[18], we instructed the performing surgeon preparatory to the surgery. We asked him to describe the steps and underlying decisions as instantly as possible and gave him an example of what we meant by 'thinking-aloud': "Right now I am thinking about what catheter I should use. To make this decision, I have to know how severe the stenosis is." However, the surgeon was not consistently urged to keep on speaking whilst being mid step, but rather to continue afterwards, as to not disrupt his concentration. The surgery was videotaped with two differently faced cameras. The first one focused on the performing surgeon(Figure 3.1), so his actions could later be reconstructed and transcribed. The second camera videotaped the displays in the operating room(Figure 3.2). These displays show the fluoroscopy, angiography and digital subtraction angiography(DSA) images the surgeon uses to navigate instruments inside the patient's body. As these images are one of the main clues for the surgeon's decisions and actions, they are particularly important for the simulation.

3.3. Conclusive Semi-Structured Interview

After the analysis of the first interview and the surgery observation, we were able to create a representation of the usual steps during an endovascular intervention. The next step was to validate this representation by having it corrected by our expert. Additionally, further details about the steps were elicited by CTA. This time, the approach from the first semi-structured interview was altered and several questions were added. After the steps of the procedure were agreed upon, the following questions where answered for each step:

- What are the sub tasks for this step?

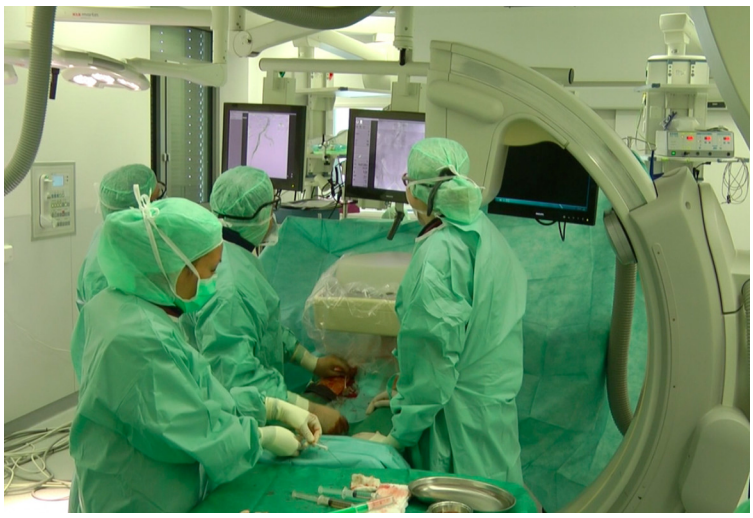


Figure 3.1.: Example frame from the video focused on the performing surgeon.

- Regarding overall understanding:
  - What actions and behavior occur in this step?
  - Does any kind of imaging occur in this step? If yes, what kind of imaging?
  - What techniques are applied in this step and what materials and instruments are used?
  - What is the purpose of this step?
- Regarding the evaluation of cognitive processes:
  - How much cognitive involvement (i.e., attention, concentration, mental alertness) does this step require for you?
  - What decisions are made in this step?
  - What cues and information are relevant for these decisions?
- Additional Information:
  - Is there specific knowledge you need to fulfill this step?
  - What are potential communication demands with the anesthetist, nursing staff or the patient in this step?
  - What are potential and typical errors and complications in this step?
- Ratings:

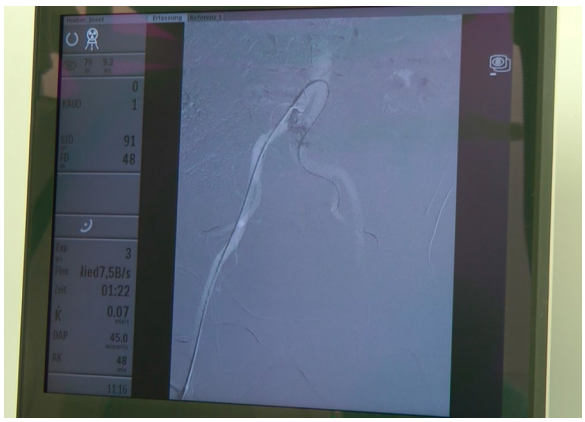


Figure 3.2.: Example frame from the video focusing on the displays.

- This step is of critical importance to a successful procedure.
  Strongly agree - Agree - Neutral - Disagree - Strongly Disagree
- This step is associated with increased safety risks.
  Strongly agree - Agree - Neutral - Disagree - Strongly Disagree

The questions were partly derived from literature and partly developed as a direct result of questions arising during our work. The 'Ratings' were added to determine Critical Performance Steps (CPS) of the procedure. As these steps are rated as being of critical importance, special emphasis should be put on the training and assessment of their performance with the simulator[13].
We concluded the interview with questions on the assessment process for surgeons in the hospital and on possible metrics to distinguish the skill level of a surgeon. Finally, we asked the expert to complete a questionnaire (Figure 3.3) designed to determine the viability of some metrics frequently used in literature. The questionnaire contained several statements relating to the performance in these areas.

Questionnaire

The following statements all refer to some aspects of endovascular procedures. Please rate how much you agree with them on the following scale: strongly agree - agree - undecided - disagree - strongly disagree

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
Short surgery times increase the likeability of a good outcome for the surgery.	O	O	O	O	O
Surgeons with better mental-mapping abilities will be able to get to the stenosis quicker.	O	O	O	O	O
Surgeons with better mental-mapping abilities will be able to get to the stenosis with using less fluoroscopy.	O	O	O	O	O
Even slight differences in the placing of the stent or balloon can make a difference.	O	O	O	O	O
The exact placing of the balloon or stent makes a difference in the outcome of the surgery.	O	O	O	O	O
The use of fluoroscopy should be kept as short as possible.	O	O	O	O	O
Using more contrast leads to a better understanding of the situation.	O	O	O	O	O
There are usually many different approaches in endovascular surgery that lead to an ideal outcome.	O	O	O	O	O
Using more contrast increases the risks for patient safety.	O	O	O	O	O
Different instruments, such as guidewires and catheters, are often interchangeable.	O	O	O	O	O
When not placed exactly over the stenosis, the balloon or stent might not work sufficiently.	O	O	O	O	O
There is usually one ideal approach in an endovascular surgery.	O	O	O	O	O
A well trained surgeon will be able to shorten the fluoroscopy time during surgery.	O	O	O	O	O
A skilled surgeon will reach the stenosis significantly faster than a less skilled surgeon.	O	O	O	O	O
Patients are usually not affected when the surgery takes longer.	O	O	O	O	O
When using more contrast, a surgeon ultimately avoids mistakes.	O	O	O	O	O

Figure 3.3.: Questionnaire to determine the viability of the frequently used metrics.

4. Data Analysis and Knowledge Representation

In the next step the recorded interviews and surgeries were transcribed and analyzed to summarize and present the acquired knowledge. Statements considered irrelevant were omitted from the transcripts. The conclusions derived from this data were discussed and evaluated by the research team to guarantee reflexivity[5].

4.1. Main Procedural Steps

After extracting the main procedural steps from literature and the CTA transcripts, we illustrated them in a Flowchart (Figure 4.1), which was later corrected by an expert surgeon to assure accuracy.

The Flowchart shows the procedure in an endovascular intervention to resolve a stenosis in the groin or legs through balloon inflation. However, the main steps for most angioplasties in general or stent placements are comparable. Specifically the steps for the insertion and placement of catheters, balloons and stents are largely alike.
As the Flowchart shows, the surgery starts with the positioning and preparation of the patient. Thereafter the surgeon punctures the arteria femoralis using the Seldinger technique and performs the first angiography. Ideally the surgeon would only use one guidewire and balloon to resolve the stenosis afterwards. However, in practice, guidewires and catheters often have to be exchanged for instruments with other properties. These circumstances will be further explained later.
Finally, the balloon has to be inflated using the right pressure. Hence the surgeon might choose to increase the pressure step-by-step, as to not damage the vessel and not leave a residual stenosis.
All steps can be guided by medical imaging, such as fluoroscopy, angiography and digital subtraction angiography (DSA). Fluoroscopy is usually used to observe the movement of the instruments while advancing them through the arteries (see Figure 4.2) and the behavior of the balloon when it is being inflated. However, blood vessels themselves are not visible with fluoroscopy.

The stenosis and its influence on the blood stream can only be seen in the angiography or DSA images after inserting contrast into the bloodstream (see Figure 4.3). Therefore

the quality of these images also depends on the appropriate administration of the contrast. For instance, if the contrast is injected through a catheter which is too far apart from the stenosis or too little contrast is injected the contrast of the resulting image might be unsatisfactory.

The kind of imaging used in every step in specific is not listed in the Flowchart, as it is also dependent on the surgeon's preferences and the different types of imaging are often combined. For instance, a surgeon might choose to perform an angiography to visualize the blood vessels first. This image can then be used additionally to the fluoroscopy to use as a roadmap whilst guiding the instruments through the vessels.

4.2. Sub Step Analysis

After agreeing on the main and sub steps of the procedure, we compiled additional knowledge regarding each sub step to gain deeper insight into the processes during surgery. These details were predominantly derived from the final semi-structured interview (see Section 3.3). The insertion of guidewires and catheters was summarized into one step in the analysis, as the steps are strongly alike.
When analyzing the data, the expert's answers and pre gained knowledge were divided into the following categories: Main steps, sub steps, actions, imaging, techniques and materials, purpose of step, cognitive involvement, decisions, specific knowledge, and typical errors complications.
The ratings to distinguish Critical Performance Steps was not significant, as the expert rated all steps after the patient positioning as both being critically important for the success of the procedure and being associated with increased patient safety risk. This means, due to the nature of the procedure, all steps have to be meticulously practiced and assessed.
The results are shown in the following Tables (Table 4.1 and 4.2) and discussed in the following. All sub steps with no significant characteristics have been omitted from the respective representation.

4.2.1. Purpose and Actions

To gain insight into the details of the procedure, particularly the surgeon's motoric performance, we analyzed the specific actions executed in each step and the desired outcome of the step. Both aspects are critical to fully comprehend the sub step. For instance, whilst the performance of an angiography seems straightforward, the details include repositioning the C-Arm and positioning the sheath or catheter appropriately to receive a conclusive image. While the diagnostic angiography at the beginning of the procedure is supposed to be focused on the stenosis, the final angiography should

show a larger section to secure proper blood flow in the whole system.
Hence Table 4.1 shows the purpose of each sub step during the procedure and the according actions performed by the surgeon.

Sub Step	Purpose	Actions
Location of Artery		The surgeon tries to feel the vessel in the groin area by locating the blood pressure with his fingers.
Puncture with Trocar	Access to the vessel without damaging it.	The surgeon punctures the skin where he feels the vessel and pushes the needle into the vessel.
Control Angiography	Check proper insertion of needle.	The surgeon and operative staff move the C-Arm over the incision. The surgeon administers the contrast through the already entered trocar. The surgeon can review the images taken with the C-Arm on one of the screens
Insertion Guidewire	Leading the sheath into the vessel.	The surgeon inserts the guidewire into the vessel through the lumen of the trocar.
Removal Trocar		The surgeon removes the trocar from the incision while leaving the guidewire inside the patient.
Insertion Sheath	Insertion of all following instruments through the sheath.	The surgeon inserts the sheath into the vessel over the guidewire and might fixate it with adhesive tape.
Removal Guidewire		The surgeon removes the guidewire through the sheath.
Angiography	View of artery in perspective to surrounding anatomy. View of blood flow within the vessel.	After administering the contrast and taking an image the surgeon can replay and stop the images to assess the blood flow and the vessel.

First Analysis	Assessment of situation and choice of approach.	
Choice of Wire	Appropriate choice for balloon catheter.	The surgeon chooses a guidewire and communicates his choice to the staff.
Bending the Tip	Crossing corners or curves within the vessel without damaging it.	The surgeon or the nurse bends the tip of the guidewire.
Insertion Guidewire		The surgeon enters the guidewire through the sheath. Depending on the type of wire, it has to be wetted beforehand.
Positioning	Crossing the stenosis.	The surgeon moves the guidewire through the vessels and positions it over the lesion by rotating and pushing it. If needed, the C-Arm is repositioned to get a better view of the wire inside the patient.
Choice of Balloon	Sufficient coverage of stenosis and appropriate catheter.	
Insertion Catheter		The surgeon inserts the balloon catheter through the sheath over the guidewire.
Positioning	Coverage of the whole stenosis.	The surgeon positions the balloon catheter over the stenosis, over the guidewire. If needed, the C-Arm is repositioned.
Removal Guidewire		The surgeon removes the guidewire from within the catheter.

Inflation of Balloon	Recanalization of the vessel.	The surgeon inflates the balloon for a certain time, e.g. one minute, starting with a lower pressure and then raising it if necessary.
Final Angiography	View of the artery and the surrounding area after the intervention.	see Control Angiography
Removal of Instruments		The surgeon removes the catheter from the sheath and then removes the sheath from the body. He or she puts pressure on the wound.

Table 4.1.: Purpose and Actions.

4.2.2. Typical Errors and Complications

As the simulation should detect any possible errors of the participant and include difficult uncommon cases it is important to recognize all possible complications during the procedure (Table 4.2). For endovascular surgery, the most typical error made by the surgeon is wrong choice of instrument. This does not directly harm the patient, however, it does increase surgery and consequently fluoroscopy time, as the instruments have to be exchanged. Additionally to being time consuming, this mistake is also expensive and can frequently be avoided by planning the procedure ahead. The knowledge of what kind of instrument should be used for a specific procedure and vessel diameter should be internalized. Other, directly harmful, complications include the damaging of the vessel wall. This can lead to occlusion or dissection of the vessel wall, which might have to be resolved by placing a stent at the affected area.

4.3. Additional Information

The interviews in combination with literature yielded the following possible performance metrics: procedure time, fluoroscopy time, contrast fluid used, stent or balloon placement accuracy, residual stenosis, lesion coverage, stent-vessel ratio, maximum stent deployment pressure, time to diagnostic aortogram, time to cross stenosis, time to inflation of angioplasty balloon, mistakes made and choice of approach[19][15][28]. However the questionnaire (Figure 3.3) to distinguish the viability of metrics indicated that the choice of approach is only partially suitable to distinguish surgeons' skill levels,

as the metric can be ambiguous. Table 4.3 shows the outcome of the questionnaire, with the Likert Scale being transferred into a grading scale ranging from 1 (the ratings strongly supported statements validating the metric) to 5 (the ratings strongly opposed the statements validating the metric).

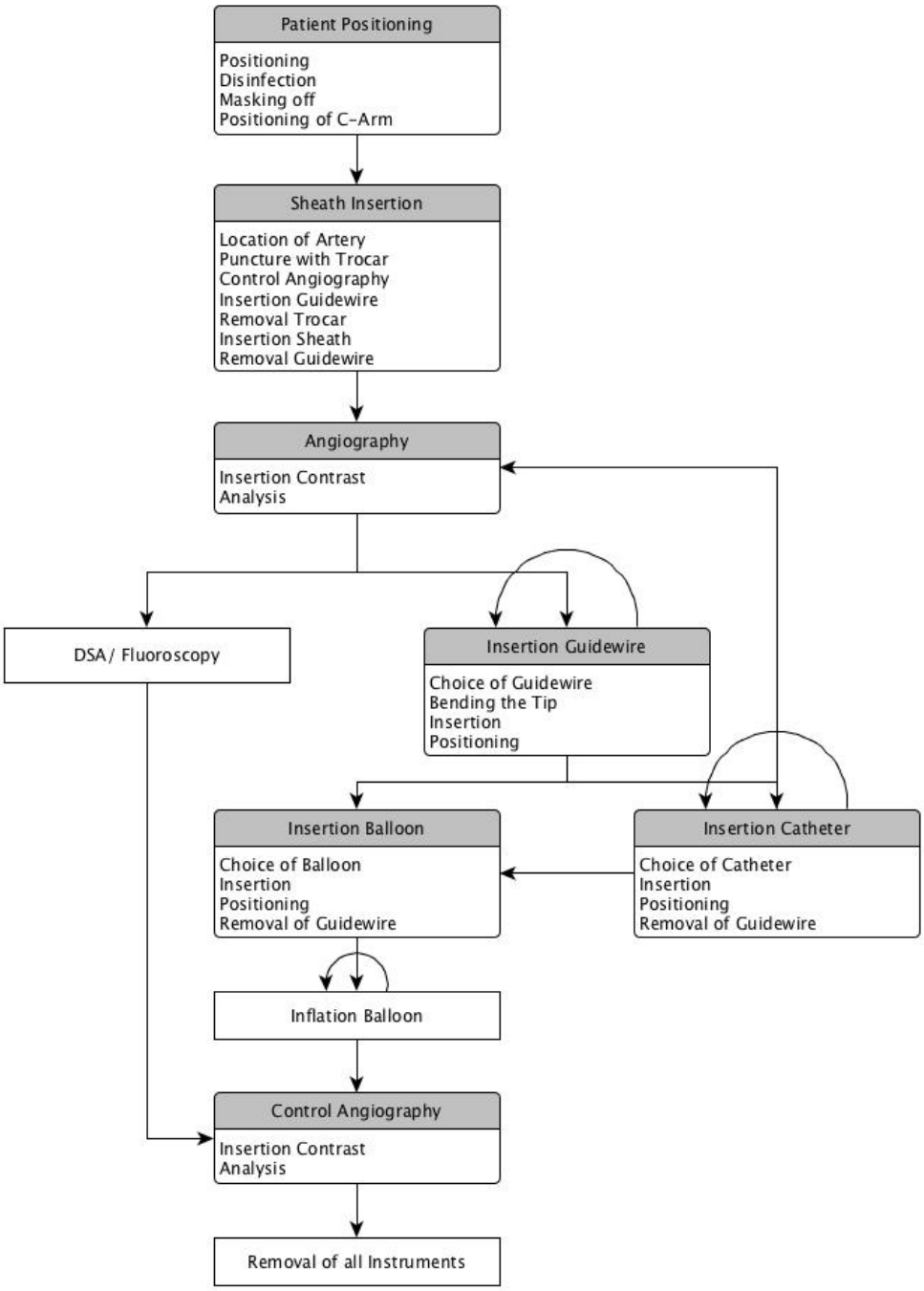


Figure 4.1.: Flowchart depicting the main steps in an endovascular procedure.

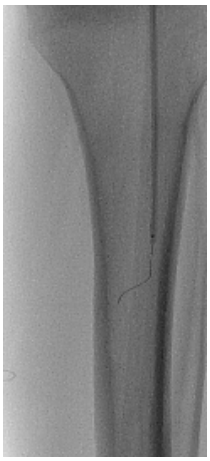


Figure 4.2.: Intraoperative fluoroscopy image of a percutaneous transluminal angioplasty showing the navigation of a guidewire.

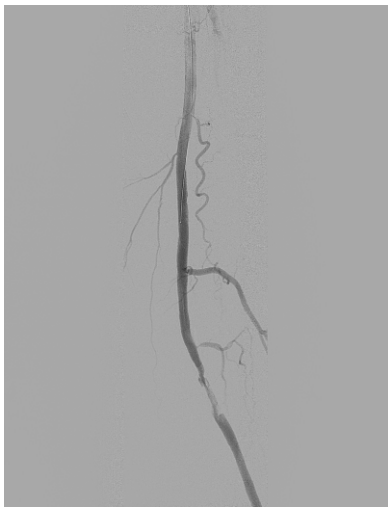


Figure 4.3.: Intraoperative digital subtraction angiography image of a percutaneous transluminal angioplasty showing the stenosis of the vessel.

Sub Step	Typical errors or Complications
Location of Artery	Vessel cannot be located ("felt"), e.g. because the patient is obese, there is limited blood flow due to a stenosis or calcification in the groin.
Puncture with Trocar	Destruction or damage of the vessel wall (lower risk with a plastic sheet around the needle). Not finding the vessel. Puncturing the wrong area, causing a lesion at a nerve, vein, bone or muscle. Risk of bleeding for high blood pressure.
Insertion Sheath	Movement of sheath during surgery.
First Analysis	Misdiagnosis of the situation and therefore wrong choice of approach.
Choice of Wire	Wrong choice of wire (a new wire will have to be inserted using a support catheter).
Positioning Wire	Dissection of the vessel wall. Failure in placing the wire.
Choice of Balloon	Wrong choice of balloon (e.g. too short to cover stenosis).
Positioning Balloon	Dissection of the vessel wall. Failure in placing the wire.
Inflation balloon	Using too much pressure and thereby damaging the vessel or causing an occlusion.
Final Angiography	Taking a too narrow image and therefore not being able to see the whole vessel system and blood flow.

Table 4.2.: Sub steps and their respective typical errors and complications.

Table 4.3.: Evaluation of questionnaire to determine the viability of frequently used metrics. Rated from 1(strongly supported) to 5(strongly opposed).

Proposed Metric	Score
Surgery Time	2,25
Fluoroscopy Time	1,33
Contrast Used	2,67
Choice of Approach	3,67
Balloon Placement	1,67

Part III.

Design and Validation

5. Simulator Design

Before deciding on the metrics and the assessment process used in the simulator, we had to design and build the physical components and the software used for the simulation device. As the preliminary aim of the project was to create a simulator without the use of actual radiation, we had to find an alternative imaging process. Furthermore, it had become clear in our prior studies that haptic feedback is an essential cue for making decisions and assessing the condition of the vessels. Therefore our final design had to incorporate these requirements.

5.1. Workstations

When dealing with complicated procedures it is often advisable not to simulate the operation as a whole but to partition it into several work stations. For endovascular procedures the steps can be divided into preoperative planning, visible work and work inside the arteries not visible without additional imaging techniques.

5.1.1. Preoperative Planning

The first stage consists of the inspection of the preoperative imaging and the planning of the procedure. These plans can include whether to use a balloon, stent or laser and which C-Arm positions will likely be necessary. The surgeon usually tries to identify the blocked arteries and estimate the severity of the calcification by measuring the outer and inner diameter of the vessel in the image. This step is not performed for all endovascular procedures. For instance, cardiovascular interventions frequently do not use preoperative CTA or MRA imaging.
We chose to combine this stage with current research on the ideal positioning of C-Arms at our university. In their paper Fallavollita et al.[8] present a system to shift the intraoperative imaging process to a 'desired-view' control. This means, the surgeon would already be able to choose his or her optimal views based on the preoperative CT or CTA images using an intuitive user interface before the actual procedure. During the surgery, the surgeon would not have to position the C-Arm but could simply communicate one of the predefined views and the system itself would compute the corresponding position dependent on the patient anatomy and positioning [8].

The metrics used in the paper, time to define viewpoints in seconds and number of optimal viewpoints defined, would be recorded to reveal differences according to skill level of the participant and correlations to the following performance. Furthermore, an expert should rate the quality of the chosen viewpoints as an additional representation of the participant's skill.

5.1.2. Visible Work

The second category almost exclusively comprises the Seldinger technique to get access to the artery before entering any additional instruments. This technique has many specific requirements on simulation, such as recreating the properties of skin and blood flow. However, as this step is not specific to endovascular procedures, we chose to exclude it from the prototype. If needed, an additional workstation could be incorporated into the simulation.
To fully meet the requirements of this task one can refer to 'Development and Validation of a Virtual Reality Simulator: Human Factors Input to Interventional Radiology Training' by Johnson et al. The paper describes the layout and the determination of several metrics specific to the simulation of the Seldinger Procedure [14].

5.1.3. Main Procedure

Consequently, the main focus of our initial simulation was the work inside of the patient. The main workstation requires the participant to choose adequate instruments, insert them into the model, navigate them through the model with the help of imaging and ultimately resolve the simulated stenosis. The details of the simulation device finally agreed upon and implemented will be illustrated in the following.

5.2. 3D model

To represent vascular procedures we chose to combine an imaging software with an actual 3D model of the arteries.
To assure a realistic depiction of the vessel structures, we chose to use 3D printing for the fabrication of this model. This type of production also has the advantage of being inexpensive, which will allow us to experiment with different anatomies and materials. We extracted the mesh for our first model from a preoperative CTA image provided to us by our expert advisor and printed it using an almost see-through plastic. We later refined the material to smoothen the surface and make the material more transparent. The last step in particular was necessary to be able to track the instruments inside the model.

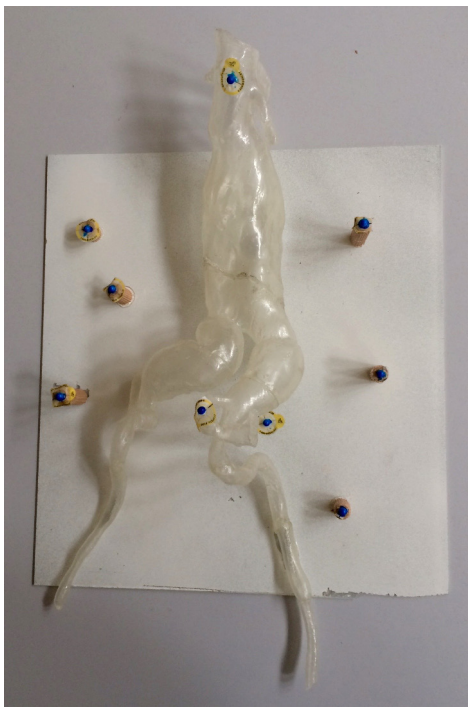


Figure 5.1.: 3D printed model of a vessel system derived from a CTA image.

The finished model, as shown in Figure 5.1, served as a prototype to use for the refinement of the imaging software and to develop the base design of the simulation. However, as the model depicts one specific case, further models will have to be extracted and printed for each varying case.

As the mesh was extracted from actual preoperative data, it was also possible to reconstruct the according surgery based on the intraoperative angiograms. This means, we could use the actual procedure as a prototype and try to adjust the simulation to it.

5.3. Imaging Software

An essential part of the project was to create a software to adequately reproduce the intraoperative imaging, including repositioning of the imaging device. The model itself should not be visible during the simulation, as the mental mapping of the 2D image into the 3D work environment of the body is one of the main cognitive tasks during surgery.
Furthermore it was important to recreate the movement of the C-Arm as its ideal positioning has a strong influence on the resulting image quality.

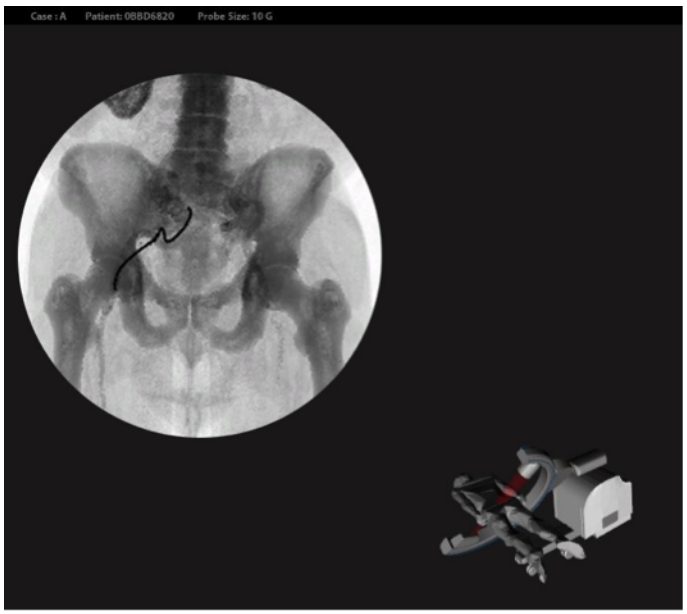


Figure 5.2.: The User Interface used in the simulation. The image can be seen on the left, the position of the C-Arm on the right.

The final software displays the artery model as well as the surrounding anatomy. As seen in Figure 5.2, the user interface includes the actual imaging of the arteries on the left and an additional representation of the position of the C-Arm in relation to the patient on the right. This way the participant can comprehend the relation between C-Arm placement and resulting image, as he or she would in real surgery.

Figure 5.3 shows the controller used during the simulation to move both the representation of the C-Arm and the resulting image.
The main image also shows the movement of the inserted instruments, as seen in Figure 5.4. The tracking of these instruments is done with the help of markers, which poses some requirements on the properties of the model. For instance, as the marker has to be well visible throughout the simulation, the model has to be as translucent as possible. Most work regarding the imaging software and the visual tracking of the instruments and the model was done by Loïse Ulrich.

5.4. Metrics

After the main layout and functionality of the endovascular simulator were specified, the next step was to find expressive and sensible metrics to build the assessment process

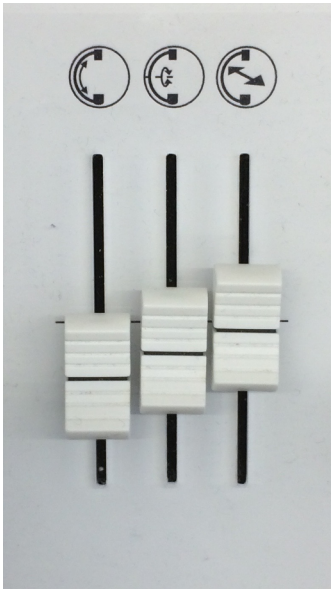


Figure 5.3.: The controller used during the simulation to move the C - Arm.

on. This means the metrics have to accurately indicate a participant's skill level and be reliably implemented in the simulation.
Based on the knowledge of validity and performance metrics in general we first decided on a goal the simulation tool should achieve. Therefore we decided, that:

"The test on the simulator should reflect a surgeon's ability to perform an endovascular procedure. It should include a combined rating of the participants cognitive and psychomotor performance as well as his or her overall knowledge of the procedure and instruments. When passing the test, a surgeon is approved to perform an endovascular procedure on an actual patient without supervision."

To achieve this goal, we assessed all metrics derived from the literature and interviews as seen in Section 4.3 on their significance. This, the practicability and the expert's rating on the viability lead to the following metrics:

- **Number of guidewires used:** The ideal number of guidewires used would usually be one. The more guidewires a surgeon has to use to complete the surgery, the longer and more expensive the surgery becomes.
- **Time to pass the stenosis:** Time it takes the surgeon to navigate the guidewire past the stenosis in seconds. This metric relates to the surgeon's mental mapping skills as well as his knowledge of guidewire properties.

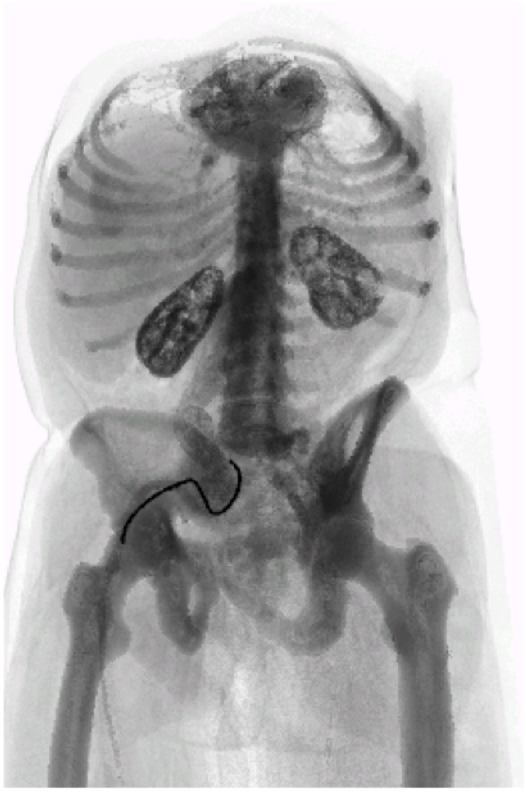


Figure 5.4.: Overlay of a CT image and the reconstruction of the catheter inside the vessel model.

- **Time the instrument is moving against the direction:** This metric records the time in which the instruments have to be pulled back in seconds. This number should be as low as possible, as having to pull back the instrument is generally a consequence of wrong navigation.
- **C-Arm movement:** Number of times the C-Arm is repositioned. A skilled surgeon will be more likely to find an ideal position and not have to move the C-Arm as often.
- **Coverage:** The percentage of the stenosis covered with the balloon or stent. If the percentage is too little, the stenosis might not be fully resolved. If it is too high, the vessel wall can get damaged.
- **Placement acurracy:** Distance from the middle of the stent/balloon to the middle of the stenosis in millimeters.

- **Maximum stent deployment pressure:** Difference between maximum used deployment pressure and ideal deployment pressure. Can also be used for balloon deployment.
- **Imaging time:** The sum of time a surgeon uses the imaging software during the simulation in seconds.

So far, fluoroscopy time and contrast used have been excluded from this list, although the initial expert rating (see Table 4.3) suggested they are suitable to distinguish a surgeon's skill level. This is because the imaging software creates an image which shows both the vessels and the instruments, therefore combining properties of both angiography and fluoroscopy. As the two imaging modalities cannot be differentiated in the simulation we have combined them into the metric 'imaging time'.
The ideal scores achieved in each metric and the resulting final score have to determined with the help of an expert.

6. Assessment Process

To prove that the simulation tool is actually able to produce scores which depict the determined construct, the development of the simulation tool has to be accompanied by a simultaneous validation process. This process should evaluate the current state of the test validity and yield improvements on the simulation and the used metrics if needed. The planned process and the inclusion of the five sources of validity will be discussed in this chapter.

6.1. Outline

The following assessment process was constructed incorporating instructions derived from *the Standards* and from literature on proper validation. The evidence to support validity based on the outcome of the assessment process will be presented after the outline of the method.

6.1.1. Participants

The participants of the assessment process should ideally be endovascular surgeons in their second or third year of training, as this is generally the phase of education when students start to perform surgeries on their own. As the scores derived from the simulation should reveal whether a surgeon has the abilities to conquer this step, participants from this stage will provide the most significant scores.
Participants from other training stages can be included, as a lack of correlation between the differing level of experience and the resulting scores would be strong evidence against validity. At least two additional experts should be included in the assessment process, so they can evaluate their experiences with the simulator.

6.1.2. Method

Before beginning the actual simulation process, the researchers briefly introduce the simulator and all important information to the participants. This includes instructions on the proper execution of the 'Think-Aloud' method. They are also told to perform the procedure as they would in a real surgical environment.

After a short break after the instructions, the participants are asked what they expect the simulation tool to be like. At this point, they should not have seen the simulator yet, to assure unbiasedness. This way, the researchers can identify the participants associations and demands and derive possible requirements on the simulation tool.
Afterwards they are asked to perform the first case. This stage of the process should give the participants a chance to get accustomed with the simulator. The participants are asked to "think-aloud" during their performance. This helps the researchers to assess the design of the simulator and give evidence for validity. To assure an ideal outcome, the participants are given examples of 'think-aloud' ahead of the simulation and a training scenario in which they are asked to solve an arithmetic problem while verbalizing their thought process. During the performance, the observing researcher can remind them to 'keep on talking' if they are silent for longer than 30 seconds [18]. The metrics should be recorded and analysed, although they do not present the official score.
A second, different test case should be performed to actually assess the participants' skill levels. The recorded scores of the participants in each category should be combined to give an overall rating of their performance.
During this case at least one expert observer should rate the participants' performance with help of the 'Objective structured assessment of technical skill' (OSATS) to give a comparable measure of the performance. The observers will be blinded to the level of proficiency of the students to assure unbiasedness.

6.2. Validation

As previously explained, validity is not a definite state but should rather be treated as a hypothesis which can be supported or disputed using different sources of evidence. In the following we will discuss the five sources of evidence in correlation to our predefined construct. Although some of the evidence is reliant on the outcome of the assessment process, we can already evaluate the significance of the possible results.

6.2.1. Content

The content forms the base of the assessment and oftentimes contradictory findings in other categories can be traced back to content underrepresentation. To provide content evidence one has to show that there is a logical correlation between the test content and the construct. This usually entails explaining the development of the test scores or the simulation tool [4].
The most frequently used method of development is using an expert panel to develop

a blueprint of the assessment construct. Other common approaches include the modification of an existing well-trusted instrument or revision of a prototype to incorporate expert opinions [4].
Within our project, we used a combination of these techniques to provide sufficient content evidence. Although the test blueprint was not created by an expert educator, it was developed using expert knowledge acquired through several scientific knowledge elicitation methods. This assures that the acquired data is both extensive and trustworthy. The knowledge acquisition and its implementation into the simulator design is illustrated in the previous chapters.
However, one could argue that the content evidence is weakened by the fact that we only consulted one expert. We tried to confirm the acquired data by interviewing additional surgeons from other hospitals and visiting their surgeries. Unfortunately, due to time constraints, it was not possible to perform another extensive CTA with other experts.
To find sufficient content evidence nonetheless we chose to incorporate revision of the simulator prototype into our validation process. This means, the simulation device should be tested by experts to make sure the simulation process represents the construct. This is why we chose to include at least two additional experts into the assessment process. While the Think-Aloud technique should reveal possible errors of test administration with less experienced participants (see Subsection 6.2.2), the experts should be told to put special focus on mistakes or insufficiencies of the simulation. This way, mistakes in the test development can be identified and revised.

6.2.2. Response Process

To provide evidence for the response process category one has to eliminate all possible errors in the test administration. This includes unbiasedness of raters, measuring errors and the final combination of the test results into a definite score [4] [9].
For our assessment process, we chose to include several different sources of evidence. Firstly, the participants get to interact with the simulation tool in an introductory test case, to assure that all participants are familiar with the device. This way, the scores will be more likely to be a representation of the construct, instead of describing the participants skill to adjust to a new work environment.
Secondly, we chose to utilize the 'Think-Aloud' technique, as recommended by Downing et al.[6]. In this case, the method should reveal all possible errors in the test administration. For instance, if a participant did not fully understand the instructions at the beginning, we should be able to identify these problems.
Furthermore, we should eliminate all possible errors of measurement due to faulty electronics or software. Therefore several trials should be executed before the actual

assessment process to refine the measuring process if necessary.
Another evidence belonging to this category is the sensible combination of specific measures or items into one final representative score. For the vascular simulation device, this means combining the performance in all metrics into one score. While the construction of the score is classified as response process evidence, the determination of a pass/fail cutscore belongs to the consequences category. However, as these two tasks strongly correspond, they should be approached collectively (see Subsection 6.2.5).

6.2.3. Internal Structure

This source of validity evidence generally refers to the reliability of the assessment process. In other words, when reproduced with different participants, observers, at another time or other differing parameters, would the assessment still yield the same results? To prove the reproducibility of the scores, research often analyzes interrater reliability, which is not applicable to our study, as the final scores do not rely on a rater. As this type of evidence does not apply to our research, we deduced that the most effective evidence would be to perform a generalizability study on our assessment process. Generalizability (G) theory aims to find all factors within an assessment that are potential sources of errors. These factors are referred to as facets and can include persons, time, settings, raters and other elements. After identifying all facets, the researcher has to quantify the amount of errors caused by each facet in specific and by the interaction of the facets. A G study results in a final score which is a direct representation of the reliability. However, G theory is an extensive field of study and outcomes are reliant on the proper execution of the study itself [27].

6.2.4. Relation to Other Variables

To find evidence belonging to this category, one has to compare the new test to an older, trusted measure. The simulator scores would ideally be compared to actual performance during surgery [16]. However, this measure cannot be accessed within the scope of a validation process.
Therefore we had to find an alternative and more definitive method of assessment.
As previously explained, most assessment in surgical education is done by observation by an expert. Due to the fact, that this method is highly subjective, the surgical research community has developed various tools to measure performance in the operation room in a more standardized manner. These measures include:

- **Operative Performance Rating System (OPRS):** A set of procedure-specific rating instruments and recommended methods for observing and judging single performances of nine operative procedures.

- **Objective Structured Assessment of Technical Skill (OSATS):** A rating scheme (see Table 6.1) including seven different categories and explanations of the scores.
- **Ottawa Surgical Competency Operating Room Evaluation (O-SCORE)**: An 11-item tool including one item on the assessment of procedural competence, two feedback items and eight items to be rated on a 5-point competency scale [10].

All of the above methods of assessment have been frequently used in research and their validity has been supported by several papers [9]. We chose the OSATS to use as a comparative measure as they are not surgery specific and result in a definite point rating.
This rating should be compared with the scores the participants achieved on their performance with the simulator. A strong correlation between the scores suggests strong evidence, whereas no correlation or divergent scored would weaken the hypothesis.
Additionally, participants who have already performed an endovascular procedure on their own should have significantly higher scores than those who have not. This is because these surgeons have been deemed skilled enough to achieve this goal through an internal assessment process which is usually accompanied by observation and mentoring. Therefore one can assume that participants who have performed an endovascular procedure by themselves have shown more advanced intraoperative skills beforehand.

6.2.5. Consequences

This source of validity evidence is the one least studied and therefore also least discussed in literature. It represents the positive, negative, intended and unintended consequences of test scores on the learners.
An obvious result of the scores in our defined construct is that a surgeon will be either be approved to do an endovascular procedure by himself or herself or will have to undergo further training. To properly and comprehensively defend these consequences, the reasons behind the determination of the pass/fail cut point have to be clear. There are several different methods from standard-setting studies to establish the minimum passing score (MPS).
To find supportive consequences evidence, we would have to identify an applicable method of determining the MPS and evaluate the possible repercussions. This is usually done with the help of experts to distinguish the difficulty of certain items or the estimated amount of students who will be able to pass a test.
Undoubtedly, these considerations lie far in the future, as they are only useful after the rest of the simulation tool is sufficiently elaborated.

Please circle the number corresponding to the candidate's performance in each category, irrespective of training level				
Respect for Tissue:				
1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue but occasionally caused inadvertent tissue damage	4	5 Consistently handled tissue appropriately with minimal damage
Time and Motion:				
1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of movement and maximum efficiency
Instrument Handling:				
1 Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments	2	3 Competent use of instruments, but occasionally appeared stiff or awkward	4	5 Fluid moves with instruments and no awkwardness
Knowledge of Instruments:				
1 Frequently asked for wrong instrument or used inappropriate instruments	2	3 Knew names of most instruments and used appropriate instrument	4	5 Obviously familiar with the instruments and their names
Flow of Operation:				
1 Frequently stopped operating and seemed unsure of next move	2	3 Demonstrated some forward planning with reasonable progression of procedure	4	5 Obviously planned course of operation with effortless flow from one move to the next
Use of Assistants:				
1 Consistently placed assistants poorly and failed to use assistants	2	3 Appropriate use of assistants most of the time	4	5 Strategically used assistants to the best advantage at all time
Knowledge of Specific Procedure:				
1 Deficient knowledge. Needed specific instructions at most steps	2	3 Knew all important steps of operation	4	5 Demonstrated familiarity with all aspects of operation

Table 6.1.: Objective Structured Assessment of Technical Skill(OSATS) rating scheme [21]

Part IV.

Discussion and Future Work

7. Future Work

The following chapter will discuss the further plans for the development of the vascular simulator, especially the proposed assessment process. To guarantee validity the simulation process will probably have to be repeatedly tested, to incorporate improvements on any exposed insufficiencies. Some inadequacies and possibilities concerning the simulation tool have already been identified and will be discussed in the following including potential implementations.

7.1. Completion of the Prototype

So far, some of the envisaged features still need to be implemented into the simulation. Firstly the metrics we decided on have to be integrated into the simulation. This entails their exact recording and possibly a representation visible to the participant. As the entities of the metrics are quite detailed the recording has to be particularly precise.
As research suggests that simulation tools are more effective when the participants get a direct feedback of their performance, some form of immediate response should also be considered [26].
Finally, the real-time tracking of the instruments inside the model and its representation with the imaging software has to be expanded and refined.

7.2. Further Development

There are several improvements to the simulation tool to consider, once the main setup is done.
To begin with, the model itself should be further refined. While being a sufficient prototype, the model has some shortcomings that should be revised.
Firstly, the model, as it is right now, is printed from a rather stiff plastic, however most vessels are flexible, especially in the abdomen. Unlike the simulation model, the arteries adjust to the form of the entered instruments. Therefore the surgeon will likely feel no resistance in real life when navigating the instruments unless there is a calcification. There are two possible solutions to this challenge. One can either experiment with alternate materials for the fabrication of the model, such as silicone, or choose a case

with comparably stiff arteries for the prototype, for instance in the legs.
Additionally, the tactile reproduction of the surgery does not feature a bloodstream yet. However, some instruments react to the contact with blood by becoming more flexible or easier to navigate. To fully emulate this behavior one has to find a substitution for the bloodstream, such as a glycerin solution [12].
Furthermore, the two final workstations actually related to surgery could be combined into one simulation device to give the impression of a patient during surgery. This way the participants could find it easier to associate the simulation with their usual work environment.
Also suggested by the experts was to include several scenarios on how to resolve complications into the simulation. These cases should be studied, however, they are obviously avoided in practice. The simulator could provide a platform to train and assess the performance in these stressful scenarios.

7.3. Validation Process

The validation process described in Chapter 6 has to be performed and analyzed. Additional validity evidence can be derived from the results. If the results weaken the construct, these aspects of the simulation have to be revisited and enhanced.
As discussed before, this could include seeking the advice of additional experts to assure strong content evidence.
After the completion of the first assessment process and the incorporation of all detected improvements to the simulation tool further assessments can be arranged. As validation is a process and not a definitive state, one should keep refining the design and evaluation of the simulator assessment as long as there is no strong evidence to support the construct found from several sources.

8. Conclusion

Due to work hour restrictions and a call for standardization in the medical field, experts are suggesting the supplementation of the traditional apprenticeship model by the use of simulation for training and assessment. In the scope of this project we created a simulation tool for the assessment of vascular surgeons and outlined a validation process. Stefanidis et al. [25] identified the use of outdated validation criteria as a significant gap in surgical simulation and proposed the use of the contemporary unitary framework as a necessary improvement.
We included this demand into the development and the evaluation of the simulation device.
After consulting an expert, we were able to create a blueprint of the surgical procedure during an endovascular intervention including potential complication and errors. We utilized Cognitive Task Analysis and the Think-Aloud technique to assure an efficient process and a thorough and extensive knowledge acquisition.
The collected data was used to design and build a vascular simulation device for surgical assessment. The main workstation includes a physical reproduction of a patient-specific blood vessel system and a software emulating the intraoperative imaging process. The finished simulator will score the participants performance based on several metrics derived from literature and expert interviews.
We put special emphasis on the proper implementation of the current validation framework into the outline of the assessment process and tried to find evidence for all five sources named by *the Standards*. This mindset influenced both the development of the simulation tool itself and the design of the validation process.
After the prototype is completed and the assessment has been conducted, the results should be reviewed and used to revise the simulation tool.

List of Figures

List of Tables

Bibliography

[1] J. Bismuth, M. A. Donovan, M. K. O'Malley, H. F. El Sayed, J. J. Naoum, E. K. Peden, M. G. Davies, and A. B. Lumsden. "Incorporating simulation in vascular surgery education." In: *Journal of Vascular Surgery* 52.4 (Oct. 2010), pp. 1072–1080. ISSN: 1097-6809. DOI: 10.1016/j.jvs.2010.05.093.

[2] J. Cannon-Bowers, C. Bowers, R. Stout, K. Ricci, and A. Hildabrand. "Using cognitive task analysis to develop simulation-based training for medical tasks." In: *Military Medicine* 178.10 (Oct. 2013), pp. 15–21. ISSN: 1930-613X. DOI: 10.7205/MILMED-D-13-00211.

[3] D. A. Cook and T. J. Beckman. "Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application." In: *The American Journal of Medicine* 119.2 (Feb. 2006), 166.e7–166.e16. ISSN: 0002-9343. DOI: 10.1016/j.amjmed.2005.10.036.

[4] D. A. Cook, B. Zendejas, S. J. Hamstra, R. Hatala, and R. Brydges. "What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment." In: *Advances in Health Sciences Education: Theory and Practice* 19.2 (May 2014), pp. 233–250. ISSN: 1573-1677. DOI: 10.1007/s10459-013-9458-4.

[5] C. Craig, M. I. Klein, J. Griswold, K. Gaitonde, T. McGill, and A. Halldorsson. "Using cognitive task analysis to identify critical decisions in the laparoscopic environment." In: *Human Factors* 54.6 (Dec. 2012), pp. 1025–1039. ISSN: 0018-7208.

[6] S. M. Downing. "Validity: on the meaningful interpretation of assessment data." In: *Medical Education* 37.9 (Sept. 1, 2003), pp. 830–837. ISSN: 1365-2923. DOI: 10.1046/j.1365-2923.2003.01594.x.

[7] M. P. Eason. "Simulation devices in cardiothoracic and vascular anesthesia." In: *Seminars in Cardiothoracic and Vascular Anesthesia* 9.4 (Dec. 2005), pp. 309–323. ISSN: 1089-2532.

[8] P. Fallavollita, A. Winkler, S. Habert, P. Wucherer, P. Stefan, R. Mansour, R. Ghotbi, and N. Navab. "'Desired-View' controlled positioning of angiographic C-arms." In: *Proceedings of the 17th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI), Boston* (2014).

[9] I. Ghaderi, F. Manji, Y. S. Park, D. Juul, M. Ott, I. Harris, and T. M. Farrell. "Technical skills assessment toolbox: a review using the unitary framework of validity." In: *Annals of Surgery* 261.2 (Feb. 2015), pp. 251–262. ISSN: 1528-1140. DOI: 10.1097/SLA.0000000000000520.

[10] W. T. Gofton, N. L. Dudek, T. J. Wood, F. Balaa, and S. J. Hamstra. "The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence." In: *Academic Medicine: Journal of the Association of American Medical Colleges* 87.10 (Oct. 2012), pp. 1401–1407. ISSN: 1938-808X. DOI: 10.1097/ACM.0b013e3182677805.

[11] S. M. Green, A. J. Klein, S. Pancholy, S. V. Rao, D. Steinberg, R. Lipner, J. Marshall, and J. C. Messenger. "The current state of medical simulation in interventional cardiology: a clinical document from the Society for Cardiovascular Angiography and Intervention's (SCAI) Simulation Committee." In: *Catheterization and Cardiovascular Interventions: Official Journal of the Society for Cardiac Angiography & Interventions* 83.1 (Jan. 1, 2014), pp. 37–46. ISSN: 1522-726X. DOI: 10.1002/ccd.25048.

[12] S. Ikeda, C. T. Villagran, T. Fukuda, Y. Okada, F. Arai, M. Negoro, M. Hayakawa, and I. Takahashi. "Patient-Specific IVR Endovascular Simulator with Augmented Reality for Medical Training and Robot Evaluation." In: *Journal of Robotics and Mechatronics* 20.3 (2008), pp. 441–448.

[13] S. J. Johnson, C. M. Hunt, H. M. Woolnough, M. Crawshaw, C. Kilkenny, D. A. Gould, A. England, A. Sinha, and P. F. Villard. "Virtual reality, ultrasound-guided liver biopsy simulator: development and performance discrimination." In: *The British Journal of Radiology* 85.1013 (May 2012), pp. 555–561. ISSN: 1748-880X. DOI: 10.1259/bjr/47436030.

[14] S. J. Johnson, S. M. Guediri, C. Kilkenny, and P. J. Clough. "Development and validation of a virtual reality simulator: human factors input to interventional radiology training." In: *Human Factors* 53.6 (Dec. 2011), pp. 612–625. ISSN: 0018-7208.

[15] D. Klass, M. D. B. S. Tam, J. Cockburn, S. Williams, and A. P. Toms. "Training on a vascular interventional simulator: an observational study." In: *European Radiology* 18.12 (Dec. 2008), pp. 2874–2878. ISSN: 1432-1084. DOI: 10.1007/s00330-008-1090-y.

[16] J. R. Korndorffer, S. J. Kasten, and S. M. Downing. "A call for the utilization of consensus standards in the surgical education literature." In: *American Journal of Surgery* 199.1 (Jan. 2010), pp. 99–104. ISSN: 1879-1883. DOI: 10.1016/j.amjsurg.2009.08.018.

[17] C. L. Lake. "Simulation in cardiothoracic and vascular anesthesia education: tool or toy?" In: *Seminars in Cardiothoracic and Vascular Anesthesia* 9.4 (Dec. 2005), pp. 265–273. ISSN: 1089-2532.

[18] H. Lundgrén-Laine and S. Salanterä. "Think-aloud technique and protocol analysis in clinical decision-making research." In: *Qualitative Health Research* 20.4 (Apr. 2010), pp. 565–575. ISSN: 1049-7323. DOI: 10.1177/1049732309354278.

[19] S. K. Neequaye, R. Aggarwal, R. Brightwell, I. Van Herzeele, A. Darzi, and N. J. W. Cheshire. "Identification of skills common to renal and iliac endovascular procedures performed on a virtual reality simulator." In: *European Journal of Vascular and Endovascular Surgery: The Official Journal of the European Society for Vascular Surgery* 33.5 (May 2007), pp. 525–532. ISSN: 1078-5884. DOI: 10.1016/j.ejvs.2006.12.022.

[20] P. L. Niels Taatgen. "Procedural Knowledge in Percutaneous Coronary Interventions." In: *Journal of Clinical & Experimental Cardiology* (2013). ISSN: 21559880. DOI: 10.4172/2155-9880.S6-005.

[21] H. Niitsu, N. Hirabayashi, M. Yoshimitsu, T. Mimura, J. Taomoto, Y. Sugiyama, S. Murakami, S. Saeki, H. Mukaida, and W. Takiyama. "Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room." In: *Surgery Today* 43.3 (Mar. 2013), pp. 271–275. ISSN: 0941-1291. DOI: 10.1007/s00595-012-0313-7.

[22] V. A. Pandey, S. A. Black, A. M. Lazaris, J. R. Allenberg, H. H. Eckstein, G. W. Hagmüller, J. Largiader, J. H. N. Wolfe, and European Vascular Workshop, Pontresina. "Do workshops improved the technical skill of vascular surgical trainees?" In: *European Journal of Vascular and Endovascular Surgery: The Official Journal of the European Society for Vascular Surgery* 30.4 (Oct. 2005), pp. 441–447. ISSN: 1078-5884.

[23] V. A. Pandey and J. H. N. Wolfe. "Expanding the use of simulation in open vascular surgical training." In: *Journal of Vascular Surgery* 56.3 (Sept. 2012), pp. 847–852. ISSN: 1097-6809. DOI: 10.1016/j.jvs.2012.04.015.

[24] N. Ratanawongsa, P. A. Thomas, S. S. Marinopoulos, T. Dorman, L. M. Wilson, B. H. Ashar, J. L. Magaziner, R. G. Miller, G. P. Prokopowicz, R. Qayyum, and E. B. Bass. "The reported validity and reliability of methods for evaluating continuing medical education: a systematic review." In: *Academic Medicine: Journal of the Association of American Medical Colleges* 83.3 (Mar. 2008), pp. 274–283. ISSN: 1040-2446. DOI: 10.1097/ACM.0b013e3181637925.

[25] D. Stefanidis, N. Sevdalis, J. Paige, B. Zevin, R. Aggarwal, T. Grantcharov, and D. B. Jones. "Simulation in Surgery." In: *Annals of Surgery* SEPTEMBER (2014), p. 1. ISSN: 0003-4932. DOI: 10.1097/SLA.0000000000000826.

[26] I. M. Tjiam, B. M. A. Schout, A. J. M. Hendrikx, A. J. J. M. Scherpbier, J. A. Witjes, and J. J. G. van Merriënboer. "Designing simulator-based training: an approach integrating cognitive task analysis and four-component instructional design." In: *Medical Teacher* 34.10 (2012), e698–707. ISSN: 1466-187X. DOI: 10.3109/0142159X.2012.687480.

[27] N. Webb and R. Shavelson. "Generalizability theory: Overview." In: *Encyclopedia of statistics in behavioral science* 2 (2005), pp. 717–19.

[28] M. C. M. Willems, J. A. van der Vliet, V. Williams, L. J. S. Kool, D. Bergqvist, and J. D. Blankensteijn. "Assessing endovascular skills using the Simulator for Testing and Rating Endovascular Skills (STRESS) machine." In: *European Journal of Vascular and Endovascular Surgery: The Official Journal of the European Society for Vascular Surgery* 37.4 (Apr. 2009), pp. 431–436. ISSN: 1532-2165. DOI: 10.1016/j.ejvs.2008.12.021.