Summary
of the thesis Training in Urology_from virtual to reality

Urology training today is still largely based on the apprenticeship model. This means that frequently procedural skill training is patient based. Currently, there are more and more calls to introduce new training methods and models to supplement the traditional educational approach,[1-4] partly owing to legal and ethical objections raised against patient-based training. A report entitled 'Unintentional Injuries in Dutch Hospitals' (Onbedoelde Schade in Nederlandse Ziekenhuizen) states that such injuries occur in 5.7% of all admissions and that the percentage is higher in surgical than in non-surgical departments [5]. Possibly, new, simulator-based training programmes that enable trainees to first practise procedures on simulators outside the operating theatre can go some way towards preventing morbidity and discomfort to patients. In November 2007, the Netherlands Health Care Inspectorate (Inspectie voor de Gezondheidszorg) published a report on adverse events in laparoscopic procedures,[1] in which all disciplines using endoscopic techniques were invited to develop and implement uniform national multi-disciplinary training programmes to reduce the risks that these techniques may pose to patients.

The value of simulations and models for urology training and the transfer from simulation-based skills to performance in practice should be established in validation studies (**Chapter 1**). Simulation models can be validated by subjective and objective studies [6-9]. Subjective studies focus on experts' opinions (content validity) and novices' opinions (face validity) of training models. Objective studies involve experiments aimed at establishing whether 1) simulators discriminate between different levels of expertise and whether a learning curve is observable in novices who use the simulator for training (construct validity) and 2) performance in clinical practice is better following initial simulator training (criterion validity).

Before validation studies are performed, it is advisable to first determine which models have been described and researched. It is also important to seek the opinions of programme directors, involved in the development and implementation of postgraduate surgical training programmes with regard to how simulator-based training in skills centres fits into training programmes.

This dissertation focuses on the general research question:
"*To what extent and in what manner are training models currently integrated in endoscopic urological educational programmes; what is the status of their validity; and how can skills be transferred from the training model to performance in patients?*

Because this question was too broad to be answered by one study, we formulated several specific sub-questions and designed studies to examine these. The study presented in **Chapter 2** was designed to answer sub-question 1: "*What types of endourological training models are described in the literature and to what extent have they been validated?*" The study involved a qualitative review of all papers published between 1980 and 2008 that 1) described one or more endourological training models and/or 2) examined the validity of such training models. Articles dealing with undergraduate medical education or physical examination were not included.

The literature search yielded descriptions of thirty types of training models and 54 validation studies but only three randomised studies with a 1b Oxford Centre for Evidence-Based Medicine (OCEBM) classification score. Of the 36 experimental studies (construct and criterion validity), 33 had small study populations varying from seven to 36 participants. The paucity of published validation studies and the low quality of the available evidence led us to recommend that more randomised studies with larger study populations should be conducted. What is needed are detailed descriptions of models that make it possible for potential users to benefit from prior developments and experiences of others on one hand and, on the other hand, studies of criterion validity in particular which examine the transfer of simulator-acquired skills to performance in real patients.

Chapters 3 and 4 are devoted to sub-question 2. "*How do directors of urological specialist training programmes view the possible shortcomings of endoscopic training and the curriculum design of urology training in the Netherlands?*" All directors of specialist training programmes in general surgery, gynaecology, orthopaedics and urology answered 25 items of a written questionnaire consisting of three parts: 1 demographic data; 2 existing facilities and training programmes; 3 their views on the existing and future training programmes. The response was 73.6% among general surgeons, 68.2% among gynaecologists, 75.0% among orthopaedic surgeons, and 90.9% among urologists.

In 55% of the respondents' hospitals, opportunities for endoscopic skill training were offered. These training modalities varied from training with video instruction to simulator-based training. Of the urologists, only 45% provided the requested estimate of the number of hours postgraduate trainees in their hospital devoted to endoscopic skill training. The estimates varied from 0-400 hours. The majority of the urologists and gynaecologists expressed dissatisfaction with the number of hours of endoscopy-related education provided for trainees. By contrast, the majority of the directors of programmes in general surgery and orthopaedics were satisfied with the number of hours of education. In 77% of the hospitals offering endoscopic skill training, attendance was obligatory (urology 62%). Opinions varied widely with regard to the most effective training models for endoscopy; virtual reality models for procedural skills training were considered most valuable. The majority of the programme directors agreed that endoscopic skill training should be directed on a national level and based on national

guidelines. With the exception of two orthopaedic surgeons, the respondents thought that endoscopy training should be an integral part of surgical training programmes.

In conclusion, the programme directors agreed that endoscopic skill training by means of different types of training models should be an integral part of training programmes. However, such models should first be developed and validated.

Matters concerning the development of training models are considered in **chapter 5**, which examines sub-question 3: *"What is involved in the development of a low-cost, easy to use training model for diagnostic and therapeutic procedures in bladder pathology?*" This chapter focuses on a pig bladder model for various procedures, including flexible and rigid cystoscopy, transurethral resection of bladder tumours of various sizes, biopsies and cystolitholapaxy. Detailed descriptions are given of the materials used, costs are estimated and it is described how a variety of bladder pathologies can be created to maximise the authenticity of tumour resections performed on the model. This model has several advantages: 1) low development and running costs; 2) different urological and gynaecological procedures; 3) realistic tactile feedback and 4) an environment in which complications and errors are acceptable. The drawbacks of this model are that, unlike virtual or augmented reality models, it does not measure objective parameters, bleeding cannot be simulated (yet) and the capacity of pig bladders is approximately 100-300 cc less than that of human bladders.

Chapter 6 describes one of the initial steps of validation, namely the determination of face and content validity of URO Mentor, a virtual reality simulator. This chapter addresses sub-question 4: "*To what degree does UCS (Urethrocystoscopy) and URS (Ureterorenoscopy) simulation on URO Mentor resemble real-time UCS and URS procedures according to urologists and postgraduate trainees in urology?"* Eighty-nine urologists and urology trainees performed a UCS or URS procedure on URO Mentor and evaluated the simulator by completing a questionnaire.

The general impression of URO Mentor was good as is reflected in the score of 7.3 on a ten-point scale from 1=poor to 10=excellent. Regression analysis showed that these results were not associated with the amount of experience in performing endoscopic procedures or the score on URO Mentor. The level of realism of URO Mentor was rated 3.5 or higher, on a five-point scale from 1=poor to 5=excellent, by 25% of respondents. The simulator's usefulness for training programmes was rated 3.5 or higher by 82% of respondents. The results showed no significant differences between specialists and trainees and between UCS and URS procedures. Apparently, URO Mentor is judged to be a realistic and useful simulator. This justifies undertaking further steps of the validation process.

The construct validity of URO Mentor was examined in a study set up to answer sub-questions 5: *"Do novices significantly improve their cystoscopy performances with respect to time, trauma, inspected areas and GRS (global rating scale) scores by training on the URO Mentor VR simulator?"* and 6: *"Is it possible to discriminate between experts' and novices' levels of performance by using the URO Mentor simulator?"* (**Chapter 7**). Fifty novices

(undergraduate medical interns) and thirty experts (urologists, postgraduate trainees and a nurse practitioner who all had performed >50 UCS procedures) performed seven procedures on URO Mentor. The first, fourth, and seventh procedure were 'test procedures,' in which the participants' performance was assessed. Procedure time and number of trauma were measured by the simulator and the percentage of areas inspected and a score on a global rating scale were given by a supervisor.

The novices showed significant improvement in task performance after training on URO Mentor. Effect sizes were large for procedure time and the mean global rating scale score; effect sizes were moderate for trauma and small for improvement in percentage of areas inspected. The fact that urologists had higher ratings on all variables shows that URO Mentor discriminates between experts and novices.

Apart from urologists' and trainees' judgments of the simulator and the ability of the simulator to support learning, one of the key questions in validation studies is whether simulator training leads to improved performance of procedures in patients. **Chapter 8** reports the results of a study conducted to answer sub-question 7: *"Do trainees' urethrocystoscopy performances in real patients improve by training on the URO Mentor VR simulator?"* One hundred medical interns were randomised to performing UCS in a patient either 1) after training the procedure on URO Mentor (UM trained) or 2) without prior training (control group). All participants watched an introductory video to ensure they had the same knowledge of UCS at the start of the study. All participants received an explanation of how to handle the scope in a session in which the bladder was represented by a glass ball. The rationale of this session was that it was considered ethically unacceptable to have an intern perform UCS in a patient without any prior experience in handling a scope. The study was approved by the Medical Ethics Review Boards of the participating hospitals (UMC Groningen and Catherine Hospital Eindhoven). A blinded supervisor rated performance on a global rating scale [10].

Multiple linear regression analysis showed that the UM trained group scored significantly better than the control group. Effect sizes were moderate to large. The group that received training performed significantly better than the controls. No effect of training was observed for participants with a preference for a surgical specialty in two out of five GRS scores. The general introductory session may have been sufficient preparation for the technical aspects of the procedures for these interns. Interns from Catherine Hospital Eindhoven scored more highly than interns from the University Medical Centre Groningen on the handling of the endoscope. This may be associated with the fact that intern performance was rated by different supervisors in the two hospitals. Despite the fact that the global rating scale was evaluated and validated, scoring appears to remain susceptible to individual supervisors' differing opinions. The Groningen interns in the control group experienced more stress than did the Eindhoven interns in that group.

Chapter 9 deals with sub-question 8: *"Does Uro Trainer have sufficient face and content validity to justify conducting experimental studies to investigate its construct and*

criterion validity"? It describes the validation of Uro Trainer, an endourological virtual reality simulator of transurethral prostate (TURP) and bladder (TURBT) resection. A questionnaire on their experiences with TURP and TURBT procedures on the simulator was completed by 97 participants (21% experts and 79% novices) and 64 participants (30% experts and 70% novices), respectively. The questionnaire contained eleven questions with ten-point scales about usefulness, realism and general impression (1=not at all useful, realistic, very poor, 10=extremely useful, realistic, excellent). The participants also indicated whether they would consider purchasing Uro Trainer (yes, no, maybe).

The mean ratings varied between 5.6 and 8.2 with large standard deviations (1.4-2.5). When we searched the literature for criteria to interpret the results, we found widely differing scales and cut-off points in studies that used quantitative measures to determine face and content validity [6;11-16]. Because the literature search provided no clear guidelines, we compared our data with those of various other studies. This led to the conclusion that Uro Trainer did not have satisfactory face and content validity and that improvements to the simulator were needed before further validity studies could be undertaken.

It is imperative that the requirements to be met by a simulator are specified before the simulator is developed or improved and also to ensure that simulator-acquired skills are transferable to performance in real patients [17-19]. A needs analysis should define important components of the training of a certain procedure [17-19]. From our experiences in conducting various validity studies we learned that this type of analysis has not been performed for several endourological procedures. Such an analysis could be conducted by asking experts to list the requirements of simulators (experts are frequently involved in the development of simulators and have been asked this question) and by examining the pitfalls encountered by trainees in performing a procedure in real patients. This is important because in developing simulators we are primarily interested in aspects of a certain procedure that pose difficulties to trainees during real-time performance, because it is these aspects for which simulator-based training is most appropriate.

We were interested in identifying the pitfalls related to performance of TURBT, TURP and URS procedures in patients in order to determine which aspects need to be integrated into simulators and which existing simulators already offer these training modalities. **Chapter 10** examines sub-question 9: *"What pitfalls do novices encounter when learning TURBT, TURP and URS procedures on patients and what lessons can we learn from that for the development of simulator-based training programmes?"* In this study, 37 TURBT, 22 TURP and 21 URS procedures in the operating theatre were observed. The observer was not a member of the operating team and noted all procedure-related interactions between the urologist and trainees. These interactions were labelled as pitfalls or non-pitfalls by the same observer and by a rater who had not been present in the operating theatre (interrater agreement was 0.7). The pitfalls were

classified based on a list derived from urologists' and trainees' answers to a questionnaire with open-ended questions, asking about pitfalls encountered by trainees and a pilot study in which 30 transurethral procedures (which were not included in the final study) were observed.

The most common pitfall in all three procedures turned out to be 'planning/anticipation on new situations', while the second most frequent pitfall was the handling of instruments. This suggests that training programmes and developers of simulators should ensure that these aspects are included in training of TURBT, TURP and URS procedures outside the operating theatre.

When conducting the studies aimed at developing, evaluating and validating simulators and skill laboratory-based training, we were confronted with aspects of these studies that should be addressed or modified in the future. **Chapter 11** presents a review that critically examines different aspects of the validation of surgical simulators. A striking phenomenon is the huge difference between research groups in the use of terminology. The terms face, content and expert are used for subjective validation studies, while referent, discriminative, construct, concurrent, criterion and predictive validity are used in experimental studies.

A literature search aimed at identifying uniform guidelines for subjective validation studies yielded some descriptions of validity but no guidelines regarding the methods to be used in examining these concepts. A comparison of various studies showed that questionnaires vary considerably in the types of answer scale, with studies using four-point, five-point, seven-point and ten-point Likert scales or yes/no questions. Decisions to qualify a model as good or poor were also based on arbitrary cut points even when the same scales were used.

We compared the methods and parameters used to establish the construct validity of endourological simulators in the studies performed between 1980 and 2008 included in the study in Chapter 2. All studies included the parameter 'time'. Although this is undeniably an important parameter, speed of performance is not automatically an indicator of a good procedure outcome.

We found only three endourological studies of criterion validity, although this is often regarded as the most important type of validity to be studied, because it is concerned with the transfer of simulator-acquired skills to real patient performance. The number of studies examining this type of validity for general laparoscopic and endoscopic procedures was larger, but most of these studies had small study populations, were not randomised, had no blinded observers or had participants that differed in basic knowledge and skills.

There was also a large diversity in the way experts and novices were defined. Most studies used number of procedures performed as the criterion to determine level of expertise. It is doubtful, however, whether this is the best approach, considering that earlier studies have shown that specialists' are notoriously unreliable in self assessing their levels of skill and experience [20].

The review resulted in the following main recommendations with regard to the validation and implementation of surgical simulators: 1) needs analysis and programme design should be integral components of the development of training facilities; 2) non-technical factors affecting the performance of procedural skills must also be considered; 3) cooperation of specialists, trainees, educational experts and industrial designers is needed; 4) assessment methods should be developed and evaluated.

Finally, **Chapter 12** presents the general conclusions with regard to the main research question of this dissertation. Additionally, there is a general discussion of the description and validation of simulators, their use in training programmes and the transfer of skills from simulator to patient. This chapter ends with some perspectives for the future and suggestions for further research.

Acknowledgement statement

Thanks are due to Mereke Gorsira for her editorial assistance.