Hospital Doctors' Communication Skills

A randomized controlled trial investigating the effect of a short course and the usefulness of a patient questionnaire

Bård Fossli Jensen

Doctoral Thesis

HØKH, Research Centre, Akershus University Hospital, Norway

Department of Health Services Research, Institute of Clinical Medicine, Campus Ahus, University of Oslo

Department of Behavioural Sciences in Medicine, Institute of Basal Medical Sciences, Faculty of Medicine, University of Oslo

2011

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1198

ISBN 978-82-8264-043-5

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

We are what we repeatedly do.
Excellence, then, is not an act, but a habit.
Aristotle, 384–322 BC

Table of contents:

1 Preface

Looking back, I had my first experience with communication as an easily trained behaviour when I was 16. My mother and I moved to New Zealand. I started college, without knowing anyone. A teacher gave me friendly advice before we moved there; he said, even if you don't feel like it, look at everyone you meet in the corridor and say "hi." According to my own feelings at the time, this seemed more like a recipe for being seen as a weirdo, rather than as a smart move to blend in. But I gave it a shot. It proved a wise thing to do.

Years later, when in medical school, I seemed to have forgotten this advice. My attitude towards communication skills training was coloured by the view that this training was more an academisation of truisms than real, medical training.

Six years after my medical graduation and a few years into specialising in paediatrics, I started to feel confident with regards to my clinical skills. I like to do things. *Better to go and see than to wait and see*. This modus operandi has some advantages when working in a hospital. However, self-reflection and academic approaches to challenges do not benefit from this attitude. When it came to communication with patients, any last doubts with regards to my own capabilities had vaporized. Most of my patients seemed to like me, I was confidant that I communicated well – and I thoroughly enjoyed being at work. My thoughts about discussions regarding communication were that you either have those skills, or you don't. I even expressed this thought out loud.

In the spring of 2006, an advertisement for a communication skills training course provoked my curiosity. Not that I didn't stick to my opinion; instead, I thought more along the lines of, can they still be teaching this? But I was curious, and three days away from the hospital, with free lunch, did not sound bad.

The training was an eye-opener. I felt like I went from being a happy amateur to a curious professional. I was satisfied with my job before. Now I enjoyed it twice as much. Just like that.

I felt I did a better job in all my meetings with patients, not just in situations considered challenging or difficult. Let me, however, share a conversation I had with an upset mom: Her 6-year-old daughter had been diagnosed with a mild attack of a not-so-uncommon diagnosis: Henoch-Schönleins purpura – an inflammation in the small blood vessels, creating bruises on the legs. Apart from the bruises, the child was in excellent health and in no danger. I was called for, as the mother was angry and refused to leave the hospital. I probably would have dismissed her as a *hysterical mom* before my training.

About a minute into our conversation, it turned out that the mother was being influenced by another idea that had been making her very anxious; the child's grandmother had similar bruises on her legs when she passed away in her nursing home a few weeks before. My reply was simple: this was something different. They left shortly after, satisfied. The point is, I would never have been able to reassure her unless I had asked for her opinion. And I would never have asked for her opinion had I not been trained to do so.

The more I thought about my new skills, the more enthralled I became with the discipline. In all other areas of medicine we are up to par with the recommendations that the research has given us. If the literature recommends an antibiotic, then this is normally the antibiotic given. If an operation technique is recommended, then this is the way a patient is operated on. And if there is a development in our knowledge, we make sure to update our colleagues and ourselves. The only area in which we are not up to with what science recommends is when it comes to communication skills.

When I was offered a part in the project that this thesis is about, to go and see if communication training really works, it appealed to me. If it works, fine. Then we would train the doctors. If it doesn't work, that's also fine. We would just know that we shouldn't waste time and money on this kind of training.

The project was already set to go: protocol, financing, statistical model – it was all there. The next step was training doctors and filming more than 500 medical consultations that took place all over the hospital. I saw that I, as a part of this project, could also further develop my own communication skills, and I would also learn research methodology. I was confident that I would be excused for entering the process late by making sure the data set I was about to collect would be solid. My decision to participate was easy.

Throughout the research period I have taken part in teaching communication skills to students and doctors. I also take shifts in the paediatric ward on a monthly basis. This has helped me keep my focus and my motivation. It has enforced my belief that communication skills are like any other clinical skill; they can be trained for and they have to be maintained.

I have enjoyed every aspect of this project. I believe it takes us one step closer to bringing doctors' clinical communication skills to the level at which they should be – in line with the most recent scientific research.

2 Acknowledgements

I would like to thank the Faculty of Medicine at the University of Oslo for including me in the doctoral program. I would also like to thank the Institute of Basic Medical Sciences, the Department of Behavioural Sciences in Medicine, all participants at the Tuesday's meetings, and Helse Sør-Øst Health Services Research for providing a solid learning environment, and Helse Sør-Øst for funding this project.

To my supervisors, Arnstein Finset and Pål Gulbrandsen, you have been outstanding, and I will forever remember the positive, empathic and respectful ways you guided me through the ups and downs of a long-term academic project.

It would not have been possible to conduct the study without the support of Erik Kreyberg Normann, the former managing director at Akershus University Hospital.

Raters Wenche Moastuen, Tonje L. Stensrud, Evelyn Andersson, and Anneli Mellblom did a tremendous job in rating the videos. Your thoroughness has been one of the pillars in this work. Thanks to Haldor Husby for scanning all questionnaires and to Erik Holt for digitalizing and safely storing all the video-recordings.

I would like to thank my co-authors Andrew M. Garratt, Fredrik A. Dahl, and Jurate Saltyte Benth for our collaboration, and a special thanks to my American co-authors Dana Gelb Safran, Richard Frankel, and Ed Krupat for also being inspiring teachers in the communication training course I attended and which sparked my original interest in this field. I also thank the group teachers Øivind Ekeberg, Christofer Lundqvist, and Petr Ricanek. Ellen Kristvik, thank you for your valuable observations during the training, and Reidun Skårerhøgda, thank you for organizing the courses. Tormod Fladby, Harald Hrubos-Strøm and Jan Ole Røvik, your extensive feedback along the way was very helpful, thank you. I would also like to express my appreciation to Barne- og Ungdomsklinikken, AHUS, for finding a place for me during

these years – making it possible for me to see patients and stay in contact with the paediatric ward.

Trond A Mjaaland: every day I have gone to work, I have looked forward to our conversations and your supreme Italian coffee – thank you for your support and good humour.

A special thank to my family for their support. And to Lise - thank you for being supportive far beyond any reason. You are the best.

My last thanks go to each and every doctor and patient who participated in this study – your willingness to contribute by coming to training and permitting us to videotape your appointments throughout the year for the purposes of scientific advancement was generous and highly admirable.

3 List of frequently used abbreviations

4HCS	Four Habits Coding Scheme
4HPQ	Four Habits Patient Questionnaire
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CI	Confidence Interval
ICC	Intra Class Correlation
ISRCTN	International Standard Randomised Controlled Trial Number
OPEQ	OutPatient Experiences Questionnaire
OSCE	Objective Structured Clinical Examination
PDF	Portable Document Format
PG	Pål Gulbrandsen, researcher
SD	Standard Deviation

4 List of Papers

Paper I

Post-recruitment confirmation of informed consent by SMS.

Gulbrandsen, P. & Jensen, B. F. 2010. Journal of medical ethics, 36, 126-8.

Paper II

Interrater reliability for the Four Habits Coding Scheme as part of a randomized controlled trial.

Fossli Jensen, B., Gulbrandsen, P., Benth, J. S., Dahl, F. A., Krupat, E. & Finset, A. 2010. Patient Education and Counseling, 80, 405-9.

Paper III

Effectiveness of a short course in clinical communication skills for hospital physicians: Results of a crossover randomized controlled trial (ISRCTN22153332).

Fossli Jensen, B., Gulbrandsen, P., Dahl, F. A., Krupat, E., Frankel, R. M. & Finset, A. 2010. Patient Education and Counseling, 84, 163-9.

Paper IV

The ability of a behaviour-specific patient questionnaire to identify poorly performing doctors.

Fossli Jensen, B., Dahl, F. A., Safran, D. G., Garratt, A. M., Krupat, E., Finset, A. & Gulbrandsen, P. 2011. BMJ quality & safety.

5 Executive summary

Background

The principles of good communication between doctors and patients are well known. We also know that learner centred training improves doctors' communication skills. Most studies have, however, been done outside of hospitals or exclusively with oncologists. We have found no large-scale studies that have investigated the effect of communication skills training using hospital doctors, from all clinical settings, trained by the same course.

Methods

In a cross-over randomized controlled trial in a large Norwegian hospital we trained doctors for two days following the didactic model of the Four Habits approach to effective clinical communication. We assessed the doctors' communication skills using videotaped encounters with real patients from different clinical settings throughout the hospital, rating the communication skills using the Four Habits Coding Scheme. Patients were recruited using a model developed by us, using SMS after 24 hours to confirm a preliminary consent. We investigated the patients' experiences regarding the doctors' communication skills using the Four Habits Patient Questionnaire - a questionnaire whose content was virtually the same as the coding scheme used to score the videos.

Results

We included 71 of 103 (69%) doctors asked and 497 of 574 (87%) patients. Doctors' baseline communication skills were below scale midpoint. Their communication skills improved significantly when assessed with the total score of the coding scheme, and the doctors who had had prior training improved the most. However, the doctors reported it was hard to maintain the change in communication behaviour over time.

87.8% of the between-doctor variance was not detectable by the patient questionnaire.

Conclusion

Two days of training improved hospital doctors' communication skills. A communication specific patient questionnaire was found to be unsuitable for the purpose of identifying doctors who had been assessed, by expert observers, as performing poorly. Following my findings I suggest that, instead of trying to identify poorly performing doctors in order to train them, all doctors should attend communication skills training sessions on a regular basis.

6 Background

Communication scholars agree upon no singular definition of communication. One pragmatic approach is to define communication as "the relational process of creating and interpreting messages that elicit a response" (Griffin, 2009). As the word clinical is derived from "*klinikē"* - the Greek expression for "bedside," (Oxford, 2010) the phrase "clinical communication" refers to communication between health care professionals and their patients.

This thesis focuses on ways to improve doctors' communication skills and on identifying doctors who perform poorly in this area.

6.1 Short history of research on clinical communication

The milestones in communication research have been well described by leading experts within the field. The Swedish book "Patient-läkarrelationen" (Ottosson, 1999) and the American "The Medical Interview" (Lazare et al., 1995a) both offer well-written overviews, and the following summary is mainly informed by these two books.

After World War II, American doctors treating military personnel became significantly more attentive to the psychological aspects of health and disease (Stoeckle and Billings, 1987). This caused a surge in the development of psychiatric principles. Medical professionals became increasingly aware of the emotional components of disease and health and began incorporating techniques from psychiatry, psychology, and social work in teaching medical interviewing. The number of published studies on communication between doctors and patients started to grow exponentially in the mid-1960s (Anderson and Sharpe, 1991). In 1977, Engel proposed the biopsychosocial model in the journal Science (Engel, 1977), and his work led even more medical practitioners and educators to realize that a more comprehensive approach to interviewing patients was required. Rutter et al. analyzed psychiatric

outpatient interviews to determine the behaviours that affected the quality and quantity of data elicited. They demonstrated the advantages and disadvantages of different interviewing styles for eliciting factual data and emotions (Cox et al., 1981a, Cox et al., 1981b, Cox et al., 1981c, Hopkinson et al., 1981, Rutter and Cox, 1981, Rutter et al., 1981). Shortly after, Mumford et al. consolidated the growing evidence that interviewing and related skills had a significant impact on a wide range of clinical outcomes (Mumford et al., 1982).

Over the decades, there was a great shift in the considerations around the basic principles of how doctors should look at their patients and how clinical communication should be performed. As an example, we can look at the expressed ideal of detachment. In 1958, Aring wrote in JAMA that doctors must remain apart from "the enervating morass of the patient's problems, viewing them detachedly yet interestedly" (Aring, 1958). Lief and Fox stated in 1963 that "The same detachment that enables medical students to dissect a cadaver without fear or disgust seemingly enables them to listen to patients without becoming emotionally involved" (Lief and Fox, 1963). This ideal is no longer supported as the aim of the medical community. We now aim to communicate interest, respect, support, and empathy to our patients as best clinical practice (Lazare et al., 1995b). There is consensus around the principles of good communication, which include building a relationship, using open-ended questions initially, exploring the patient's perspective, displaying empathy, checking for understanding, reaching agreements on problems and plans, and providing closure (Makoul, 2001, Rao et al., 2007, Stewart et al., 1999, Simpson et al., 1991).

6.2 Teaching clinical communication skills

Inui found in 1976 that by teaching doctors to discuss compliance problems with patients, solutions could be found and positive effects on health outcomes (e.g., blood pressure control) were accomplished (Inui et al., 1976). Along with this

increased focus on the importance of clinical communication, there was also a shift in the direction of the research. Up until the 1990s, reviews had their focus on the relationship between communicative behaviour and patient outcome. After this, the focus of reviews shifted more in the direction of investigating the effects of interventions on communication behaviour (Anderson and Sharpe, 1991).

In a doctor-patient setting, training can be aimed at patients, medical students, or doctors. Much of the groundbreaking work within communication training research has been done using the first two groups. Less intervention research has focused on investigating doctors; when this is the case, it has occurred most often within psychiatry, general practice or single medical specialties like oncology. For the groups studied, however, there are no doubts about the positive effect of communication skills training (Aspegren, 1999, Rao et al., 2007).

Considerable research has also been done around the principles of how this teaching and training should be done (Ottosson, 1999, Lazare et al., 1995a). We know learner-centred teaching to be superior to a more traditional teacher-centred approach (Levinson and Roter, 1993); in short, learning by doing is more effective than learning by instruction. However, training requires a minimum of intensity, and according to Aspegren are courses that offer one day's training or less not effective (Aspegren, 1999).

Despite the increased focus on the effect of communication training, none of the randomized controlled studies reported in the reviewed articles included training programs that had been tried out on doctors from all medical specialties. We found it important to investigate whether one course could prove effective for all disciplines. As we began to collect data, we were also motivated by the report from Rao et al. published in the same month. Their main conclusion was that the biggest challenge in this field was now for investigators to design effective patient and doctor

communication interventions that can be integrated into routine practice (Rao et al., 2007).

6.2.1 The Four Habits approach to effective clinical communication

Frankel and Stein structured in 1996 the principles of good, clinical communication into a teaching model for didactic purposes: "The Four Habits model – an approach to effective clinical communication" (Frankel and Stein, 2001). The habits are: invest in the beginning of the encounter to create rapport and set an agenda (Habit I), elicit the patient's perspective (Habit II), demonstrate empathy to provide opportunity for patients to express emotional concerns (Habit III), and invest in the end to provide information and closure (Habit IV). They implemented the model as the basis for teaching programs covering a wide variety of settings and specialties, with duration from 3-4 hours to a five-day course, in Kaiser Permanente, one of the largest health-care organizations in the US. The training was well received by the doctors and the training had positive effect on patient satisfaction surveys in observational studies (Stein et al., 2005). In 2006 the model was well known to my supervisors. They found it appealing due to its clear didactic strategy and also because it was already in use in ordinary practice. However, although the teaching model was widely in use, highly appreciated by the doctors in Kaiser Permanente, and had positive effect on patient satisfaction in observational studies, no experimental effect study had been conducted. It was also not necessarily so that the model could be translated to the context of a Norwegian hospital. My supervisors organized a pilot study to prepare for a randomized controlled trial. They tested a three-day version of the training method used in Kaiser Permanente on 16 Norwegian hospital doctors. I was one of the attendees. The course proved successful according to interviews with the doctors in focus groups (Gulbrandsen et al., 2008) and self-reports investigating the doctors' self-efficacy (Gulbrandsen et al., 2009). However, the participants were all highly motivated doctors interested in clinical communication. To investigate both feasibility

and the effect on a representative population of doctors in a general teaching hospital, my supervisors then organized a randomized controlled trial and assigned me as a PhD student. One of the main purposes if the trail was to investigate research question 1 on page 26. The intervention, following the Four Habits Model, is described in more detail under 8.7.

6.2.2 Possible implications of studying the effect of a communication skills training course

Studying the feasibility and effects of such a course will add important information to help evaluate whether or not training according to the Four Habits Model should be prioritised. If the training shows no effect, resources can be saved by not spending time on training doctors following this model. If the training has an effect it can encourage both the individual doctor and the hospital administration to prioritise communication skills training using the Four Habits Model for communication skills training. As doctors from all clinical specialties can attend the same course, it makes it easier to fit the training into the daily routines of the hospital.

6.3 Evaluating clinical communication skills

Closely related to the research investigating clinical communication skills and the effect of communication skills training is the development of reliable tools to evaluate those skills. In the 1970s, building on work from scientists working with group interventions, investigators developed schemes for coding interactions between patients and doctors, and studied these interactions, one of the most well known being Korsch's work examining visits to a paediatric emergency ward (Francis et al., 1969, Korsch et al., 1971, Korsch and Negrete, 1972). During the 1980s, important studies showed that quantitative methods were reliable and valid when coding interviews (Inui and Carter, 1985), and in the same decade the methods for evaluation of the communication shifted. Earlier research had mainly focused on

ethnography and participant observations (Becker et al., 1961, Fox, 1959, Glaser and Strauss, 1967, Merton et al., 1957, Mumford, 1970), while in the 1980s an increasing number of investigators use audiovisual records as primary source of data (Baron, 1985, Cicourel, 1980, Frankel and Beckman, 1982, Erickson, 1982, Mishler, 1984, West, 1984), and it was demonstrated that quantitative and qualitative methods could be used to complement each other in understanding the complexities of communication between patients and doctors (Roter and Frankel, 1992).

Aspegren identified 10 different methods for measuring communication skills in a review of the literature in 1999. These were 1) training course evaluation, 2) written report by the student of contents of an interview, 3) cognitive testing of knowledge of interviewing, 4) self-rating scales, 5) psychometric tests, 6) direct observation, 7) video- or audio taped interviews, 8) OSCE examination, 9) patients' ratings, and 10) patient health outcome (Aspegren, 1999). We decided to use videotaped interviews with real patients, rating the videos using the Four Habits Coding Scheme (4HCS) described in 8.8.

When a method requiring coding was decided upon, the raters had to be trained to prove valid scoring values. Agreement among raters has been discussed for decades. One well known article was published in 1960 by Jacob Cohen (Cohen, 1960) and complex statistical models have been developed for different types of scales and different types of data. A common term used to describe agreement among raters is interrater reliability (Wikipedia, 2010). When it comes to establishing interrater reliability among raters coding videos, we found only sparse descriptions in the literature on how to ensure satisfactory interrater reliability when coding communication skills. This is supported by reports stating that for research related to coding with multiple raters, the method used to calculate interrater reliability is often unsatisfactorily described. Many papers are criticised for how interrater reliability is achieved and reported when coding videos (Stemler, 2004).

As interrater reliability is a property of the testing situation, not of the tool itself, we needed to establish satisfactory interrater reliability among our raters for this study (Stemler, 2004). The method of how to calculate interrater reliability was decided after we considered what would likely be the method best-suited when using two coders. We soon realised we would need more coders than first anticipated, which led us to also consider other methods for calculating interrater reliability.

Validating the coding tool was essential to be able to report reliable results. However, if successful, a report including both the theoretical background for choice of methods and a thorough description of how we established interrater reliability should be welcomed by other researchers looking to use the same or similar coding tools. We therefore reported our experiences in a methodological paper (Paper II).

6.4 Comparing patient experiences to expert coding

Large resources are required to videotape doctors with real patients, and hence another question we wanted to investigate was whether video filming could be replaced by patient experience questionnaires. Patient questionnaires have the advantages of being simple to administer and are far less costly than video recordings. However, many questionnaires designed to measure doctor communication lack external validation and caution should be used when interpreting patients' ratings of their doctors (Epstein et al., 2005). Furthermore the patient questionnaires that have been validated have often been validated against other questionnaires, seldom against tools that rely on assessing the communication with audio- or videotapes. We therefore thought it important to compare patient-experienced observations with the coding of objective raters in order to investigate whether a patient questionnaire would be a reliable tool when it comes to identifying poorly performing doctors.

6.4.1 Possible implications of studying a communication-specific patient questionnaire's ability to detect variance

If a communication-specific patient questionnaire can be proven to be equally good at identifying low-performing doctors, as compared to video-recordings, this would be beneficial considering the reduced need for resources, and could possibly also reduce stress on patients and doctors who would not need to be videotaped in order to evaluate the doctors' communication skills.

6.5 The need for a new approach to informed consent

Early on in the period of preparation and planning, we recognized the challenge presented by working within the routines of a fully operational hospital across all medical disciplines. It is important that patients gain relevant information before they consent to a trial, and they also need a cooling-off period before making a final decision. The recommended time for patients to reflect before they give consent to attend a clinical trial is a minimum 24 hours (Wager et al., 1995). This was also a strict requirement of the regional ethics committee. It is fully possible to carry out this kind of study in many clinical settings, but it has mostly been tried out in clinical settings in one or very few locations. We aimed at studying 512 patients in a minimum of 192 different locations, and filming was meant to take place at bedsides during rounds, during practical procedures, in the outpatient clinic, and in the emergency room. We considered it impossible to plan at least 24 hours ahead of one specific patient visiting the emergency room, due to the fact most patients would not know they would be ill this long in advance. But the more planned encounters would also be impossible to schedule when it comes to the individual doctor and patient. As an example – if the doctor is scheduled for bedside rounds at a certain date, the patients she will meet are not scheduled until the same morning. And even if lists of patients could be ready earlier, doctors' work schedules are often changed just a few days or hours ahead. We decided the only way would be to approach the doctors

and film them with the patient they were attending “there and then.” This left us with a very large number of potential patients. If we were to follow the traditional approach of sending out information material to all these patients in advance, we would end up sending out material to a six-digit number of patients and only approach a few thousandths of those giving consent. This would, in turn, severely compromise inclusion rates and lead to a large inclusion bias. We were in need of a new and more robust approach. The initial response from the ethical committee was that the 24 hour rule had to be followed, but having heard our arguments, our suggested approach was accepted.

6.5.1 Possible implications of a new approach for informed consent

Researchers need methods where the ethical aspects above are carefully taken care of and where, at the same time, the practical procedure is both realistic and does not severely compromise inclusion rates. A report on how to achieve a good, valid patient consent rate in a large clinical communication trial in a hospital can help to achieve high patient inclusion rates in similar studies in the future. This will also help to ensure research on encounters that would not be studied when following the traditional procedure.

7 Research Questions

The overall aim of my thesis was to investigate whether a communication training course, put in practice in a running hospital, proved feasible and effective, and to investigate the usefulness of a patient questionnaire in identifying poorly performing doctors.

Related to this background we asked the following research questions:

1: Does a 20-hour communication training course, addressing general communication skills taught by the principles of the Four Habits approach to effective clinical communication, change communication behaviour of hospital doctors across clinical settings and specialties? The answer is found under 9.3.1 on page 42.

2: When compared to videotapes, can a patient experience questionnaire, whose content is derived from the same conceptual model and whose content is virtually identical to the scoring tool used by video coders, differentiate highly performing doctors from poorly performing doctors when it comes to communication skills? The answer is found under 9.3.2 on page 44.

During the process, challenges emerged with regards to informed consent and interrater reliability. This led to the investigations reported below, as I believe it will be of benefit for a future researcher within this field:

a) The feasibility of a new approach to obtain informed patient consent based on re-confirmation of a preliminary consent. Result found under 9.3.3 on page 46

b) Reliability of the Four Habits Coding Scheme used for rating communications skills of hospital doctors across clinical settings and disciplines within a hospital. Result found under 9.3.4 on page 47

8 Material and Methods

As preparation for this thesis, my supervisors conducted a pilot study with two main purposes. One was to explore how the Four Habits Approach was experienced by Norwegian Hospital doctors. A second was to develop and test the feasibility of the Four Habits Patient Questionnaire (4HPQ) - a patient questionnaire aimed to be concordant with the observational items in the Four Habits Coding Scheme. The US model for training in the Four Habits Approach proved applicable with only small adjustments and none of the participants doubted the effectiveness of the elements of the Four Habits. For description of the development of Four Habits Patient Questionnaire see 8.10.1 and Table 1.

8.1 The hospital

The study took place in Akershus University Hospital, whose main tasks are research, medical education, and medical treatment. The hospital, located in the capital area of Norway, with 762 somatic beds and 6,300 employees, is one of the largest hospitals in Norway - providing specialized health services to 470,000 people (AHUS, 2011).

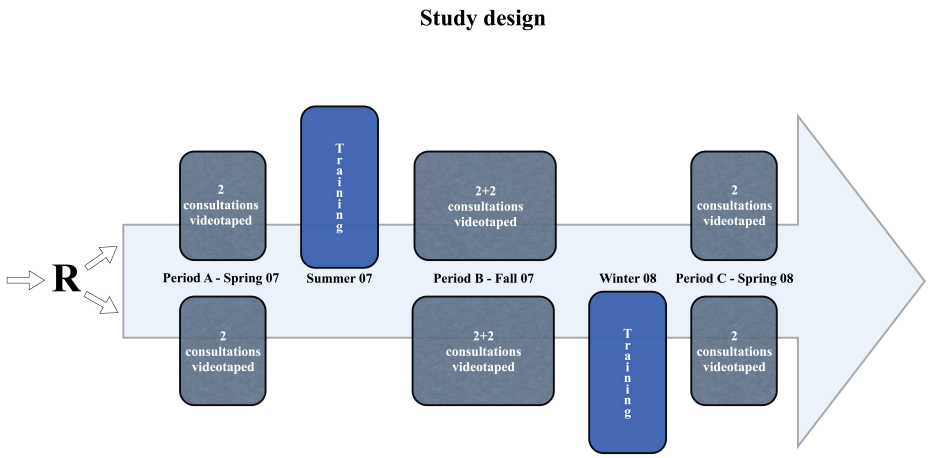
8.2 The design

We wanted to test the effect of the communication training course, and for effect studies a randomized controlled trial is the preferred method. We used a crossover design with doctors as their own controls in a randomized controlled trial, with the intervention at different time-points in the two arms (Figure 1 below).

All doctors included had two encounters videotaped before the first course (period A – baseline). After the first course, all doctors had four videotaped encounters (period B). Then, the doctors who had not participated in the first course received the

intervention, followed by the videotaping of another two encounters for all doctors (period C).

Figure 1



8.3 ISRCTN registration

We applied for registration in the International Standard Randomised Controlled Trial Number Register on 23. April 2007. The registration number is ISRCTN22153332 (ISRCTN).

8.4 Sample size and statistical power

A statistician computed sample size estimation under the assumption that the effect size was 0.4 SD, which is considered a small to medium effect size. We applied a multilevel analysis, with doctors at the upper level and patients at the lower level, and assumed ICC = 0.1, alpha = 0.05, beta = 0.80, and used two-tailed tests. This showed that the sample size would have to be 32 doctors in each group – a total of

64. Altogether 512 consultations would be needed. Anticipating a 10% loss in follow-up, we needed 72 doctors to secure the data for this sample size.

8.5 The doctors

All authorized staff doctors below 60 years of age working in clinical departments (anaesthesiology, paediatrics, surgery, internal medicine, gynaecology/obstetrics, neurology, orthopaedics, ear-nose-throat (ENT)) by February 2007 were made available, a total body of 249 doctors. Psychiatrists and radiologists were not included due to the particular clinical settings of these two specialities; the doctors in psychiatry were not included, as their work requires more than basic communication skills; radiologists were not included because the meetings with patients are most often merely technical. Characteristics of the doctors are given in Table 1 in Paper III.

8.5.1 Random selection and recruitment of doctors

Using a random, stratified draw from the total body of doctors we produced a list of eighty doctors to serve as primary participants. In anticipation of 20% of doctors not being willing to participate, we added another 40 doctors to serve as substitutes to the list according to the same principles. Two statisticians, Jurate Saltyte Benth and Fredrik A. Dahl, performed randomization on February 15, 2007. They used the excel file “Legeliste til random 15 feb 2007” (which listed doctors, their department and position), and extracted the preset number of doctors of senior and junior positions in each department. Technically, adding a column with random decimal numbers between 0 and 1 made this possible. Thereafter, they sorted all rows, with department name as the first key, position as second key, and the random number as last key. This sorting procedure listed all doctors in the same department and in the same position consecutive, but randomly within these groups. After that, the statisticians manually extracted doctors for inclusion, up to the preset number for each department and position. The courses in June were named a and b, and the

courses in December/January c and d. The doctors were assigned to the groups a-b-c-d, consecutively in the order that they were extracted. Finally, the statisticians made a PDF file of the list of 80 (primary participants) + 40 (substitutes) and mailed it to PG Feb 15, 2007 at 1:33 PM.

Before inviting any doctors, a meeting with the managing director of the hospital was arranged. After the study was explained to him, he then endorsed it and recommended all department heads allow doctors a paid leave of absence to attend the communication-training course. We arranged separate meetings with all heads of departments and explained the study and answered possible questions. We made it clear that no more than two doctors per department would be absent simultaneously. After meetings with the heads of departments, an email was sent out to all doctors informing them about the study and preparing them for the possibility that they might be drawn upon to participate. We sent hard copy personalized invitations to the selected doctors, sent by the internal mail system in the hospital. In addition, an email invitation was sent out through the hospital's email system. Doctors who did not respond to any of these two invitations were contacted by telephone. If a doctor declined to participate, the next doctor from the stratified sample was asked and this procedure was repeated until either the required number of doctors had agreed to participate or there were no more doctors to include from that particular stratum.

8.6 The patients

We included real patients being treated at the hospital by the doctors in the study. Exclusions were made either because of patient refusal or because the researchers or the doctors found them ineligible. Patients were found ineligible who did not have the language skills to understand what was being said verbally or in the written material or who were too ill to talk or fill out papers. Participating doctors were also allowed to exclude patients if they considered the patient to be in a particularly vulnerable situation.

The diagnoses of the patients were not collected. However, the main areas of medical complaints are indicated by the clinical speciality of the doctor. Characteristics of the patients are given in Table 3 in Paper IV.

8.6.1 Recruitment of patients

When a patient arrived at the hospital and was waiting to see the doctor, I, or sometimes my supervisor PG, approached him or her and explained the study. We also handed the patient written information. If the patient consented initially, his or her consultation was videotaped with a video camera on a tripod and an external microphone. No researchers were present in the room during the consultation. Immediately after the consultation, the patients filled out patient questionnaires and signed the consent form. After 24 hours, all patients were contacted by SMS (or in person, or by phone or regular mail if they did not use SMS) and asked to answer yes or no to whether they stood by their consent. All SMSs were transferred to a computer and saved for future documentation of the consent. For patients who could not use SMS, I noted the date and time of the verbal reconfirmation on the consent form.

8.7 The intervention

The intervention was a communication skills training course for the doctors based on the principles of the Four Habits Model. The training took place at a different location than the hospital and the participants wore their private clothing. The rationale for this was mainly to make sure they could focus on the course without distractions like beepers or colleagues needing help, but it was also important to have a safe setting that helped them feel open to sharing emotions, their own insecurity or work-related frustrations.

The doctors participated in the 20-hour (in sets of 45 minutes) course over two consecutive days. The decision to spend two days in training was mainly based on

what was possible to achieve considering the practical implications for taking doctors out of their daily work. The course consisted of a 50/50 mix of theory and 45 minute group sessions (3-7 participants and two teachers per group) including role-plays, with plenary debriefs after each group. The theory-based plenary sessions were focused on the core issues of good communication.

The course was based on the same content as the 5-day Communication Skills Intensive course offered by Kaiser Permanente (Stein et al., 2005). The main differences between our two-day course and this 5-day course were that our plenary sessions were more compressed and that the group sessions were able to incorporate less focus on the individual doctor's development. In the role-plays, doctors played both doctor and patient roles. We did not use actors or real patients and we did not videotape any of the doctors as part of the training. Clinical scenarios suited for the training of each habit were available, and adjusted to the specialty of the doctor playing him/herself. Some instructions were given separately to the two role players. The patient instructions included imagining the patient's family situation, beliefs, expectations, and emotions, as well as basic symptom descriptions. After feedback, role-plays were rerun by the same players or by a new pair, depending on what would likely be most instructive. Most participants acted at least once both as the doctor and as the patient during the course.

There were six group sessions; one for each habit, one for specific training based on participants' interests, and one dedicated to further post-course training.

At the conclusion of the course, all participants received a one-sheet overview of the Four Habits to carry in their pockets as reminder during their everyday work.

About three months after each intervention, all participants were invited to join a two-hour group session. The object was to discuss any thoughts they might have had

after the training and ask about their experiences with implementing what they had learned into their clinical routines.

8.8 The coding tool and variables expected to change

To evaluate communication in the videotaped encounters, an obvious tool for us to consider was the Four Habit Coding Scheme - Figure 2 on page 34. It was already developed and had been validated, although in general practice and not in a hospital setting like ours. We needed a coding tool that was quick to use. Krupat et al. reported that when using the Four Habits Coding Scheme, it took about 5 minutes more than the actual time of the encounter to rate the communication behaviour (Krupat et al., 2006). Based on the above information, we decided to use the Four Habits Coding Scheme in our study. In Four Habits Coding Scheme, the videos are rated using a 23-item scale with the items organized according to the Four Habits model. Item categories were 1, 3, and 5, from 1 representing “not very efficient behaviour” to 5 “highly efficient behaviour.” The original codebook was translated to Norwegian, giving examples of communication behaviour that qualifies as good practice in each of the categories. Coders were instructed to use these categories and only use categories 2 and 4 if they felt strongly that the behaviour being coded was directly between two of these categories, as stated on the Four Habits Coding Scheme. The habit scores consist of six items for Habit I, three for Habit II, four for Habit III, and ten for Habit IV.

Figure 2

Four Habits Coding Scheme

Code each of the items below using the categories 1, 3 or 5. If you feel strongly that the behaviour being coded is directly between categories, you may use the categories 2 or 4.

	Not very Effective				Highly Effective
1. Invest in the Beginning					
A. Shows familiarity with patient	1	(2)	3	(4)	5
B. Greets patient warmly	1	(2)	3	(4)	5
C. Makes small talk	1	(2)	3	(4)	5
D. Uses primarily open-ended questions	1	(2)	3	(4)	5
E. Encourages expansion of patient's concerns	1	(2)	3	(4)	5
F. Elicits the full agenda of concerns	1	(2)	3	(4)	5
2. Elicit the patient's Perspectice					
A. Interested in patient's understanding of problem	1	(2)	3	(4)	5
B. Asks about patient's goal for visit	1	(2)	3	(4)	5
C. Shows interest in impact on patient's life	1	(2)	3	(4)	5
3. Demonstrate Empathy					
A. Encourages expression of emotions	1	(2)	3	(4)	5
B. Accepts / validates patient's feelings	1	(2)	3	(4)	5
C. Helps to identify / label feelings	1	(2)	3	(4)	5
D. Displays effective nonverbal behaviour	1	(2)	3	(4)	5
4. Invest in the End					
A. Frames information using patient's perspective	1	(2)	3	(4)	5
B. Allows time for information to be absorbed	1	(2)	3	(4)	5
C. Explains clearly / uses little jargon	1	(2)	3	(4)	5
D. Explains rationale for tests and treatments	1	(2)	3	(4)	5
E. Effectively tests for comprehension	1	(2)	3	(4)	5
F. Encourages involvement in decision-making	1	(2)	3	(4)	5
G. Explores acceptability of treatment plan	1	(2)	3	(4)	5
H. Explores barriers of implementation	1	(2)	3	(4)	5
I. Encourages additional questions	1	(2)	3	(4)	5
J. Makes clear plans for follow-up	1	(2)	3	(4)	5

For each doctor we first computed the average total score from the Four Habits Coding Scheme for the periods A, B and C (see Figure 1 on page 28), denoted a, b, and c, respectively. For a doctor randomized to the summer course, (b-a) was the estimated improvement over the intervention period, while (c-b) was the estimated

improvement over the control period. Using this approach, the estimated treatment effect was Δ=(b-a)-(c-b)=-a+2b-c. Note that if we split B into two parts, B1 and B2, with average scores b1 and b2, and define the intervention effect estimate as (b1-a)-(c-b2), this equals -a+2b-c. For doctors randomized to the winter course, the treatment effect estimate was Δ=a-2b+c. The null hypothesis H0 was that the treatment had no effect, which means that the expected treatment effect estimate would be zero: E(Δ)=0. The H1 hypothesis was that the treatment had a positive effect: E(Δ)>0. We estimated E(Δ) as a weighted average d of the individual Δ values. For robustness, we used a standard two-tailed t-test. Note that the observation unit is the doctor.

If the intervention had an effect we expected improvement in the score in the Four Habits Coding Scheme. We also investigated whether there was a change in the duration of the encounters using the same statistical method.

8.9 Methods to reach acceptable interrater reliability between raters in our material

When a video was rated using the Four Habits Coding Scheme, a total score was calculated. The score has a range from 23 to 115 points, with higher numbers indicating better communication. Raters were trained for 18 hours and then interrater reliability was calculated using this total score and ICC over the first 20 videos they rated. The number of 20 videos was chosen due to what senior researchers found reasonable. We found no consensus or guidelines addressing how many videos that need to be included in order to calculate interrater reliability.

8.10 Methods in studying the ability of patient questionnaires to identify poorly performing doctors

As far as we could see, no large-scale study had investigated the correspondence of a patient questionnaire and objective coding using instruments whose contents were derived from the same conceptual model and whose contents were virtually identical. During and after the pilot course, the Four Habits Patient Questionnaire, a patient questionnaire with highly specific items directly related to the skills being taught, was tested and developed for this purpose (Gulbrandsen et al., 2008) – see 8.10.1 below. We also included more established tools. As instruments widely used to assess patient experiences and satisfaction with health care include dimensions of communication as well (Epstein et al., 2005, Garratt et al., 2005, Roland et al., 2009, Garratt et al., 2008, Sitzia, 1999, CAHPS®), patients completed the Four Habits Patient Questionnaire along with communication and information-specific items of the OutPatient Experiences Questionnaire (OPEQ) (Garratt et al., 2005) – see 8.10.2 – and the global satisfaction item of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) – see 8.10.3 (CAHPS®).

8.10.1 The Four Habits Patient Questionnaire (4HPQ)

The Four Habits Patient Questionnaire is related directly to the Four Habits, and maps the specific items of the Four Habits Coding Scheme. The Four Habits Patient Questionnaire-pilot had 23 items – one item corresponding to each item in the Four Habits Coding Scheme. It was validated in the pilot study, as 210 questionnaires were analyzed, leaving 10 questions suitable for a questionnaire following the formal criteria for inclusion (Gulbrandsen et al., 2008). When designing the final version of the Four Habits Patient Questionnaire used in our main study, all items were evaluated again. Five items were retained, as they considered crucial elements of doctor-patient communication and were hence important for content validity, leaving the Four Habits Patient Questionnaire with a total of 15 items. The items had a four-

point scale, between ”definitely yes,” ”somewhat yes,” ”somewhat no,” and ”definitely no.” The 15 items covered Habit I (four items), Habit II (two items), Habit III (three items), and Habit IV (six items) – see Table 1 below.

Table 1

Questions in the Four Habits Patient Questionnaire (4HPQ)

I	1	Did the doctor seem to know the important information about your medical history?[b]
I	2	At the beginning of the visit, did the doctor meet you in a way that helped put you at ease?[a]
I	3	In exploring your health concerns, did the doctor give you a good chance to express yourself in your own words?[a]
I	4	Did the doctor encourage you to fully describe your health concerns?[a]
II	5	Did the doctor seem interested in finding out how you thought about the health concerns?[a]
II	6	Did the doctor seem interested in finding out how your current health problems are affecting your daily life?[a]
III	7	Did you get good eye contact with the doctor?[a]
III	8	Did the doctor seem sensitive to your feelings?[a]
III	9	Did you feel that the doctor was interested in you as a person?[a]
IV	10	Did the doctor give you information that directly addressed the concerns you had expressed?[a]
IV	11	When the doctor gave you information, did s/he give you as much time as you needed to understand it and absorb it?a
IV	12	When the doctor gave you information, was it clear and in words you could easily understand?b
IV	13	After the doctor gave you information, did s/he make sure to find out how well you understood the information?b
IV	14	Did the doctor encourage you to be as much involved as you would like in the decisions about your health care?b
IV	15	Toward the end of the visit, did the doctor make clear and specific plans about what you should do as a follow-up?b

[a] Questions obtained after a principal component analysis of the questions tried out in the pilot study.
[b] In addition to the questions in [a], five more were considered to also address crucial elements of doctor-patient communication and hence included in 4HPQ.

8.10.2 The OutPatient Experiences Questionnaire (OPEQ)

The OPEQ has been used as measure of care quality, with evidence for reliability and validity following a Norwegian national survey of patients (Garratt et al., 2005). Six of the 24 items in OPEQ are related to doctor’s communication, and we included these questions in our study. The specific questions and scale can be seen in Paper IV.

8.10.3 The Consumer Assessment of Healthcare Providers and Systems (CAHPS)

The CAHPS has been used as measure of care quality, with evidence for reliability and validity from the US (CAHPS®). We included one of the 39 items in CAHPS as a measure of global satisfaction, asking the patients “using any number from 0 to 10, where 0 is the worst doctor possible and 10 is the best doctor possible, what number would you use to rate this doctor?”

8.10.4 Variables investigated when studying variance in doctor communication detected by the 4HPQ

Directly after the encounters, patients filled in the three questionnaires above. To investigate to what degree the Four Habits Patient Questionnaire could identify between-doctor variance in communication, a two-level analysis was performed. Each doctor had up to eight filmed encounters, each encounter was treated as the lower level and each doctor as the upper level. The number of films for each doctor was taken into account. We calculated how much of the between-doctor variance could be detected by the use of the Four Habits Patient Questionnaire and used this to determine to what extent the patient questionnaires could be used to identify poorly performing doctors. The full calculation can be found in the addendum of Paper IV.

9 Results

The answers to the research questions from page 26 will follow in the synopsis of the articles below, after providing some general results.

The regional ethics committee approved our suggested approach in 8.6.1 when it came to including patients, and we conducted the inclusion of patients as described.

One of 16 department heads was reluctant to agree that the doctors in his department should be invited to the study. He did, however, agree with the explanation that he did so because everyone else had.

To reach the intended number of 72 participants, we had to invite 103 doctors (see Figure 3 below). Collected characteristics of the doctors who participated, and the doctors who refused participation, as given in Table 2 below, did not reveal any significant differences between the two groups.

Figure 3

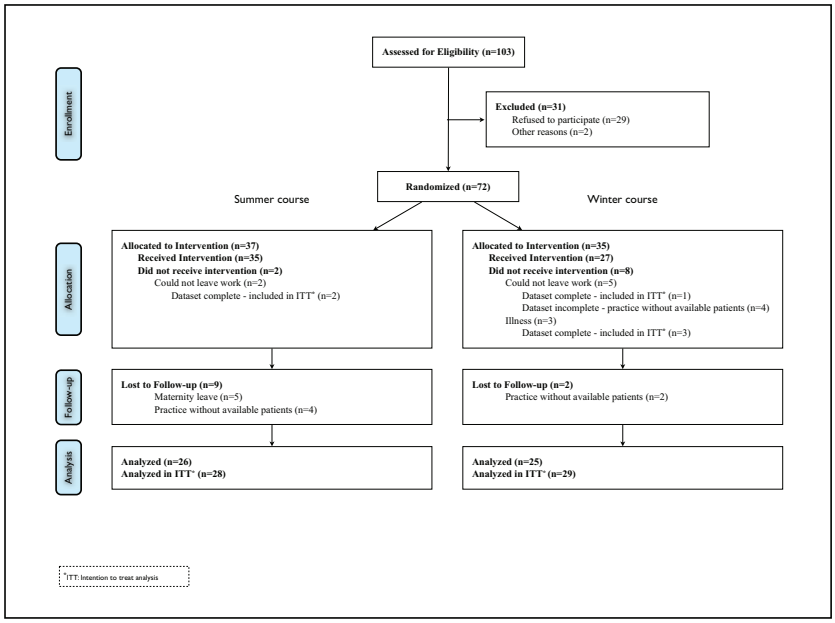


Table 2

Representativeness of doctors[a]			
	Participated (N=71)	Refused participation (N=32)	p-value
Age doctor mean years (SD)	40.3 (8.6)	42.7 (9.2)	0.378
Sex doctor			0.877
Females	30 (42)	13 (41)	
Males	41 (58)	19 (59)	
Work status of doctor			0.580
Residents	33 (47)	13 (41)	
Consultants	38 (54)	19 (59)	
Doctors medical specialty			0.120[e]
Medical[b]	25 (35)	9 (28)	
Surgical	23 (32)	17 (53)	
General[c], Orthopaedics, ENT[d]	18 (25)	10 (31)	
Anaesthesiology	5 (7)	7 (22)	
Other	23 (32)	6 (19)	
Neurology	8 (11)	2 (6)	
Paediatrics	8 (11)	1 (3)	
Gynaecology	7 (10)	3 (9)	

[a] Data presented as No (%) except as noted. Some groups do not add up to 100% due to rounding.

[b] Cardiology, Respiratory diseases, General Internal Medicine common, Nephrology, Endocrinology, Infectious diseases, Haematology, Gastroenterology

[c] Gastro surgery, Urology, Vascular Surgery

[d] Ear-Nose-Throat

[e] Pearson chi-square between the specialty sub classifications medical, surgical, and other.

9.1 Loss to analysis

574 encounters were available. Researchers excluded 18 patients due to the described exclusion criteria; one patient was excluded because the patient was also a doctor included in the present study. One patient was excluded because the interpreter refused, and one patient was excluded because the doctor became ill after the patient was asked. Of the 553 patients found eligible, 530 (96%) gave initial positive response to consent. Due to the technical failures for three encounters discovered immediately, 527 were eligible for confirmation of their consent. 519 (99%) of these confirmed their consent, giving a rate of consent of 519 of 553 (94%). 20 video films had to be excluded from the analysis due to technical errors discovered at a later stage, and two video films had to be excluded because one doctor withdrew from the project, leaving a total of 497 (87%) video films to study.

After inclusion, there could be three main reasons for exclusion of doctors from the main study; 1) not attending the communication training course (the intervention), 2) not having enough filmed encounters, and 3) withdrawal. Of the 72 included, nine did not receive intervention, 11 did not obtain enough filmed consultations for different reasons, and one doctor withdrew from the project, leaving 51 doctors for the main analysis. We compared these 51 doctors to the 52 doctors not included, and found that doctors in surgical disciplines were underrepresented (Table 1 in Paper III). Anaesthesiologists were particularly reluctant to participate – five of 12 consented and of these five, only one was available to be filmed enough times to be included in the main study. Of the four not available to be filmed enough times, we have no registration of reasons, but according to my experience, these doctors were simply more reluctant to participate than the other doctors in the study.

9.2 Baseline values assessed by video and by patient questionnaires

9.2.1 Doctors' baseline communication skills level assessed by video observation

Using the Four Habits Coding Scheme to evaluate the encounters filmed before any of the doctors in the main study had received intervention, we created a baseline score. The mean baseline score (SD) from these 102 encounters was 60.3 (9.9). Transformed to a scale of 0-100%, where higher numbers indicate better communication skills, the score was 41%. The scores were below the scale midpoint for all habits and lowest for Habit II, eliciting the patient's perspective, with a mean score of 5.4 (1.8) on the scale 3-15 for the three items in this Habit, equivalent to 20% of highest achievable score.

9.2.2 Baseline duration of the encounters

Duration (mm:ss) (SD) of the encounters at baseline had a mean of 21:34 (12:13) and a range of 04:14 – 57:43.

9.2.3 Doctors' baseline communication skills level assessed by patient questionnaires

We measured the patients' experiences of the doctors' communication skills with three instruments, 4HPQ, OPEQ, and CAHPS. The baseline levels, when transformed to a 0-100% scale, were 85%, 81%, and 86% respectively. The correlations between the tools were .79 for 4HPQ vs. OPEQ, .80 for 4HPQ vs. CAHPS, and .71 for OPEQ vs. CAHPS, all $p < .01$.

9.3 Answers to research questions - Synopsis of articles

9.3.1 Answer to research question 1 – Synopsis Paper III

Paper III: Effectiveness of a short course in clinical communication skills for hospital doctors: results of a crossover randomized controlled trial.

The objective was to test the hypothesis that a two-day course in clinical communication, based on the Four Habits model for effective clinical communication, can improve the communication skills of doctors across clinical disciplines working in a hospital.

Communication skills were evaluated using video films of real encounters over a period of 14 months. The mean (SD) number of weeks from the end of training to the video filming of the encounters was 22 (13), with the range 2-50.

In the crossover design, we found an increase in the Four Habits Coding Scheme score of 7.5 points (p=0.01, 95% CI 1.6 to 13.3), fairly evenly distributed across subgroups. For the 15 doctors who reported to have had prior communication skill training the improvement was 14.4 points (p=0.04, 95% CI 0.7-28.1). Patient satisfaction with the doctors' communication, measured with one question that rates the doctors on a scale from 0 to 10 according to CAHPS (CAHPS®), did not change significantly (0.3 (p=0.38, 95% CI -0.3; 0.8)). However, we experienced a ceiling effect well known from other patient experience reports – in our study, indicated at the relative high mean baseline score of 8.6 (1.4). The duration of the encounters (min:sec) were also sustained without significant change (-1:03 (p=0.69, 95% CI -6:13; 4:07)).

The paper concludes that the Four Habits model for effective clinical communication, originally developed in the US for an outpatient-clinic setting, can be used in a hospital setting in Norway. Hospital doctors' communication skills may be improved by a 20-hour course according to the model tested here, with the most improvement enjoyed by doctors who have had communication training before, and without influencing the average duration of the encounters.

1/3 of the doctors attended the meetings about 3 months after their intervention. These meetings revealed information in addition to what was reported in Paper III. In

a substudy published in the Journal of the Norwegian Medical Association, we documented that the doctors' self-efficacy regarding communication skills improved after the intervention (Gulbrandsen et al., 2009). Despite this, doctors told us they find it hard to focus on their communication and that it was too easy to fall back into their old habits. Many of the doctors told us there was no workplace culture that encouraged focus on improving communication. The meetings were not compulsory, and not part of the study in which the doctors agreed to participate – the meetings were voluntary. We did not obtain information regarding the experiences of those who did not attend. We could not, according to the doctors' age, sex, experience, speciality, self-efficacy, and communication skills, observe any difference between the doctors who attended follow-up meetings and those who did not.

9.3.2 Answer to research question 2 – Synopsis Paper IV

Paper IV: The ability of a behaviour-specific patient questionnaire to identify poorly performing doctors.

Assessing doctors' communication skills using observations and ratings, for example, with the help of video recordings, is expensive and time consuming. Patients' questionnaires addressing patient experiences and satisfaction with regards to doctors' communication skills are often used as proxies.

The objective of this paper was to investigate whether a patient questionnaire, the Four Habits Patient Questionnaire, designed to evaluate the specific recommended communication skills of the doctor, can replace video recordings when it comes to distinguishing between doctors with different levels of skills, especially when identifying poorly performing doctors.

In all of the 497 encounters representing 71 doctors, patients completed post-visit questionnaires. We calculated how much of the between-doctor variance observed

when coding the video recordings with Four Habits Coding Scheme could be predicted by the Four Habits Patient Questionnaire.

Although the Four Habits Patient Questionnaire correlated significantly with the video observed behaviour at doctor level (Pearson's $r=0.42$, $p<0.01$), the portion of the between-doctor variance not detectable by the Four Habits Patient Questionnaire was 87.8%. We have also illustrated the findings in a scatter plot in Paper IV. The scatter plot shows that a majority of the doctors with a low score on Four Habits Coding Scheme get good ratings on the Four Habits Patient Questionnaire, and therefore the latter has low sensitivity regardless of where one would choose to set a cut-off score. When trying to identify doctors below the mean score on the Four Habits Coding Scheme, even when using the best possible cut-off score of 82% on the Four Habits Patient Questionnaire, the sensitivity and specificity are 32% and 97% respectively, with positive and negative predicative values of 92% and 61%.

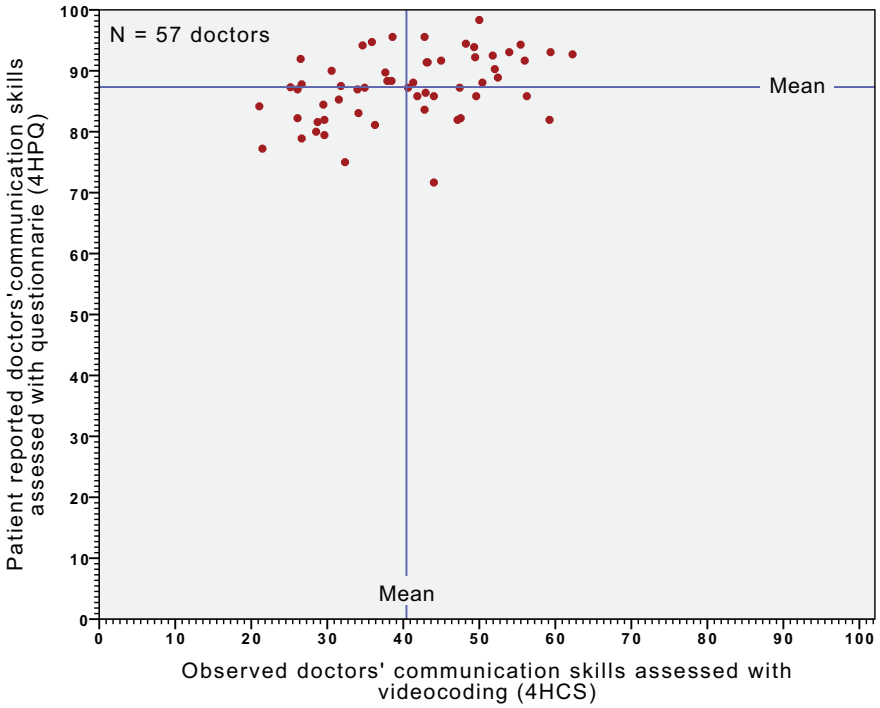
The 71 doctors are represented by 1 to 8 films, and they are displayed in the scatter plot as dots of equal size and give the impression that all dots are based on the same number of observations. Nevertheless, when producing a scatter plot including only the 57 doctors of whom 7 or more films were made (Figure 4 on page 46), the main picture is the same, except we no longer see an outlier, represented by one film only, on the bottom left.

The appearance of the scatter-plots supports the high positive predictive value of a very low mean score on the Four Habits Patient Questionnaire (below 80 % of maximum score). The standard error of the means for these doctors was, however, large.

The paper concludes that there was a significant positive correlation between patient reports of doctor communication skills and behaviour observed on video. In particular, doctors assessed to be poor communicators by their patient panel were

similarly viewed poorly by expert observers. However, since a large portion of doctors assessed by patients as having good communication skills were assessed unfavourably by expert observers, the paper suggests that patient reports alone may not be sufficient to identify all doctors who would benefit from communication skill improvement training.

Figure 4



9.3.3 Report addressing investigation of 7 a) – Synopsis Paper I

Paper I: Post-recruitment confirmation of informed consent by SMS

Of the 497 encounters, 375 (76%) were outpatient clinic visits, 81 (16%) bedside on rounds, and 41(8%) encounters in the emergency room. The paper concludes that a preliminary consent given at the time of information, followed by mandatory confirmation of the consent, sent by SMS, is a possible simplification of the procedure of securing a valid, informed consent in studies like this.

9.3.4 Report addressing investigation of 7 b)– Synopsis Paper II

Paper II: Interrater reliability for the Four Habits Coding Scheme as part of a randomized controlled trial.

After having rated the video recorded encounters, we presented our experiences with establishing interrater reliability among the coders as a poster during the International Conference on Communication in Health Care in Miami 2009. We were invited to write an article for the special issue of Patient Education and Counseling, to be published in relation to the following conference in 2010. The invitation fits with findings in the literature stating that the process of establishing interrater reliability and the rationale for choice of method on how it is calculated is often not sufficiently described (Stemler, 2004). The object of this paper was therefore to describe the process for developing interrater reliability for the Four Habits Coding Scheme in detail, including our considerations when choosing method(s) to calculate interrater reliability.

Interrater reliability was tested on four occasions, including checking for drifting, and the ICC was satisfactory with values >.74.

The paper concludes that three days of training in using the Four Habits Coding Scheme produced a satisfactory interrater reliability for the total score in our heterogeneous material. Extremely low variance on some items, either due to lack of relevance or to poor performance by most doctors, led to a reduced reliability on a more detailed level. Attempts to improve interrater reliability, by searching through habits and single items in order to identify areas to improve and focus on during further training, were not helpful. The practical implication of this is that the Four Habits Coding Scheme total score can now be recommended for intervention studies to evaluate clinical communication in several clinical settings in hospitals.

10 Discussion

I will discuss my research and findings by approaching two different areas that need discussion: a discussion of *methodological issues* and discussion of *findings.*

10.1 Methodology and random errors

One of the purposes of using the randomized controlled trial's design is to minimize random errors. We included enough doctors to prevent random skewing in distribution in the different groups. But even if a random error in this distribution could have occurred, since doctors were used as their own controls in the cross-over design, the validity of the results would not have been influenced by this. On this basis, we claim that random errors have not influenced the conclusions deducted from any of the calculations.

10.2 Methodology and systematic errors

10.2.1 Selection bias

Selection bias can be defined as a distortion of evidence or data that arises from the way that the data are collected (Webster's, 2010b).

10.2.1.1 Selection bias in recruitment of patients

Paper I is solely dedicated to the process of recruiting patients, and the selection process and possible inclusion bias of patients are also discussed in 10.3.

The most obvious place where a possible selection bias of patients may occur in our design is when the researcher is about to approach a patient waiting to see his doctor. Exclusion was to be decided if the patient had lack of Norwegian language skills, suffered from psychiatric disorders, or were too ill. This was open to interpretation by the researcher. I did most of the inclusions, and even if I did my utmost to not let myself be influenced by characteristics like patient name (which

could indicate a lack of language skills, if it is an obviously foreign name), or an appearance that could, for example, indicate that a patient might require more time or might have a psychiatric disorder, I cannot rule out the possibility that such factors have influenced the inclusion.

Another source for selection bias is the type of medical encounter that occurred. We did not collect data regarding this factor, but it is my experience that filming one doctor when in her office in the out-patient clinic is less challenging and requires less work when it comes to setting up the technical equipment than filming the same doctor bedside on rounds. This led to doctors that were available for filming both on rounds and in the out-patient clinic being filmed more often in the outpatient clinic than in other settings – and hence filming relatively more encounters in out-patient settings than in other settings.

10.2.1.1.1 Strengths and weaknesses of the sample of patients

The 87% rate of approached encounters included in the study is conservative regarding the positive response rates of patients. Included in this number are 21 films where patients were not asked, as they were not found eligible, and 25 films where films were excluded due to reasons independent of the patients. 522 (94%) of the 553 patients approached ended up giving a valid consent after confirmation of their initial consent.

One can argue that inclusion rates should not only take into account the number of patient refusals, leaving out the 25 films excluded due to reasons independent of the patients, as the question of inclusion is irrevocably linked to the total concept of what a patient is being asked to take part in. As an example, technical errors may likely happen more often in studies where a technician is not present during the encounter. But a technician being present might influence the number of patients who say yes in

the first place. Nevertheless, an inclusion rate of 87% of patients is sufficient to claim the generalizability of findings to the population in our hospital. I have not found other similar studies across medical specialties in hospitals to compare these rates with. However, in primary care, we can see that Roter et al. had a patient consent rate of 72% when studying 69 doctors' communication skills as assessed with audiotapes (Roter et al., 1995).

Patients who met exclusion criteria are not represented in this study, and hence the results cannot be considered representative for these groups. As the evaluation of language skills and clinical condition had to be evaluated by the researcher or the doctor to be filmed, there is also a chance of inclusion bias here. However, the number of patients who were excluded by doctors or researchers suggests that the selection bias introduced by exclusion procedures is low.

10.2.1.2 Selection bias in recruitment of doctors

We experienced that the procedure of recruiting doctors by first engaging the administration of the hospital, and then the heads of the different departments, was helpful when it came to encouraging doctors to accept invitations to join the study. However, the method of informing the heads of the departments included the potential risk of leaving us in a situation where the attitude of the department head influenced the doctors' decisions regarding whether or not they would agree to participate.

There are three main ways a training session can be implemented by a hospital administration.

1. The administration can have the full initiative and order someone from the outside, like our research team, to come to the different departments and implement the training.

2. Someone from the outside can suggest the training for the administration and the administration can order the departments to cooperate.

3. Someone from the outside can suggest the training for the administration and the administration recommends the departments take part, but leaves the final decision to each department head.

We chose the last method, as we believe this gives the overall best attitude towards the project and in the end provides better inclusion rates. Regardless of which of these approaches is chosen, there is a possible disadvantage of anchoring the project at the administrative level that occurred to me while attending a national conference for anaesthesiologists. Here it became obvious that, as the Norwegian health system undergoes profound changes, many of the doctors experience the changes that are implemented by the administration as based on agendas other than the direct improvement of the working conditions for the doctors and the best possible treatment for each individual patient. Many had with the impression that changes often have political or economic motives. This can cause scepticism towards interventions implemented by the hospital administration. As an example, doctors have asked me if the aim of *effective* clinical communication is meant to save time and resources. This tells me that my interpretation of the term “effective communication,” namely how to spend the time available for communication in the most effective way to *improve patients’ health*, is not obvious and intuitively shared by everyone. Nevertheless, I am under the opinion that anchoring the project within the administration of the hospital at an early stage is of vital importance when it comes to recruiting doctors to communication training. The pitfalls mentioned above will most likely be less damaging for recruitment than the practical difficulties one would meet when trying to find time off for the doctors without commitment from the administrative level.

10.2.1.2.1 Strengths and weaknesses of the sample of doctors

I have not found similar studies with which to compare inclusion rates, but we see it as a strength that as many as 71 out of 103 were willing to participate. However, 20 of the 71 could not be included in the main analysis, either because they did not receive intervention or because they were lost in the follow up process. This limits the generalizability when it comes to the surgical disciplines, as they were underrepresented in the 51 doctors included in the main analysis. I believe the main reasons for this is a combination of the scepticism mentioned above, together with a workload that does not make it easy to prioritize two days of training, while colleagues will have to make sure the clinical daily work and routines are taken care of.

We do not have any information about the communication skills of the 32 doctors who declined participation, but we do see that the six doctors who agreed to participate and were filmed throughout the whole period, but for some reason did not receive the intervention, had a lower baseline score (SD) compared to those who received intervention (49.7 (10.3) vs. 60.3 (9.9)). This can indicate that doctors who for some reason do not end up undergoing intervention are less skilled than the doctors who receive intervention, hence indicating that less skilled doctors were also underrepresented.

10.2.2 Information bias

An information bias can be described as a bias due to the object of a study influencing the nature, precision and completeness of the study data collected about that object (Webster's, 2010a).

In our study there is a risk of information bias, especially when it comes to raters' coding of videotapes. A common argument against rating communication is that much of the quality of the communication depends on the personality of the doctor,

which some call the x-factor, and a more subtle relationship between the doctor and patient, which some refer to as human chemistry. We did not ask our raters to describe either their thoughts nor their emotions around this human chemistry between the patient and the doctor. Neither did we systematically ask the raters their personal opinion around whether they liked the doctor or not, or to what extent they felt affected by this when rating. We have, however, anecdotal stories in which the raters said they were surprised by the low scores for doctors whom they liked - and vice versa. It is not unlikely that raters would be affected by this x-factor or human chemistry and thus skew the doctors' communication score in the direction of their impression. It is impossible to rule out this effect completely, but in future research I suggest including a question for the raters about to what degree they like the doctor, and maybe also in what way – e.g., do they perceive the doctor to be charming, funny, good-looking, trust-worthy etc. I also suggest studying the same consultations with regards to more objective measures on attractiveness, like sex, face symmetry, height, pitch of voice, body stature, etc, as there is evidence that factors like these influence the scoring (McConnell et al., 2008).

When it comes to patients filling out different scoring sheets, the risk of information bias occurs for many different questions. For example, asking the patient if he got good eye contact with the doctor (question 7 in 4HPQ) can lead to answers skewed in a positive direction; if the patient did not get eye contact it could be because the patient might be shy, ashamed or have other reasons for avoiding eye-contact. The patient can then think that a lack of eye contact is due to a quality within him, and hence, as it is obvious we are rating the doctor, still give the doctor a good score, as the patient will think the doctor is not to blame for lack of eye contact. We focus on this kind of information bias, and other consequences of limited ability to meta-observe the communication that you yourself are a part of, in paper IV, and it is further discussed in section 10.3.5. I believe that the amount of information bias in

our study is on par with similar research within the field – and our study actually is given strength by the fact that we investigate communication with different tools and from different angles making our conclusions less vulnerable to any information bias specific to one single tool.

10.3 Findings

10.3.1 Measuring clinical communication

We have shown that a 20-hour communication-training course for hospital doctors had a positive effect on the doctors' communication behaviour.

One major limitation when it comes to measuring the effect of an intervention using a quantitative approach, which is necessary to conduct a randomized controlled trial, is the transformation of real interactions between two people into a number. This way of transforming a complex interaction into a quantitative measure will necessarily give a simplified picture of what is actually going on in the encounter, and hence information will be lost in the process. In our study, where we do not register simply whether a behaviour occurs or not, but also how effective it might be according to the context, there is a risk of information not only being lost, but also misinterpreted. The same goes for variables that try to describe patients' observations of the doctors' communication skills.

Although the Four Habits Coding Scheme had been validated against RIAS and an acceptable interrater reliability was reported in Krupat's original study, (Krupat et al., 2006) that study used a homogenous sample of primary care consultations. This may speak against the validity of the same tool being used in a study with a heterogeneous sample from a variety of hospital settings. We still decided to use this tool as we considered it to be the best suited for our rather large body of material. The Four Habits Coding Scheme has been used in several more recent studies as well (Rouf et al., 2009, Blanch-Hartigan et al., 2010, Gulbrandsen et al., 2010). The

reported satisfactory interrater reliability in the three studies above was reproduced in our study, supporting the assumption of validity of the Four Habits Coding Scheme when used in our material.

10.3.2 Level of baseline skills

We have reported a mean baseline score well below the scale midpoint, and especially low for Habit II – eliciting the patient's perspective. One can argue that the top of the scale on Four Habits Coding Scheme represents an ideal, theoretical level of communication skills based on 23 different items that are all important in clinical communication, but not necessarily relevant in all encounters in this broad range of clinical settings. The high evaluations from patients may suggest that the ratings given by our coders using the Four Habits Coding Scheme are too severe. This is further discussed in 10.3.6.

10.3.3 Design of main study

We considered a classic randomized controlled trial with parallel groups, but had two main concerns. The first was regarding possible selective attrition when it came to participation of doctors. We were afraid that doctors would be more likely to drop out of a study with a classic design, some because they would find it acceptable to be filmed, but not if they had to attend two days of training, and some because they would feel that being filmed throughout a year without receiving any training would not be worth it. The second concern had to do with the possibility that the working conditions might change during the year of observation. For example – if the workload in the hospital is generally lower in the period after the intervention occurs, one might expect the doctors to perform better because they have more time, not because of the intervention. Or, what is more likely, the stresses of a busy hospital in winter might camouflage an improvement. Due to these concerns we decided to make it clear to all doctors that everybody would receive the intervention, and we decided we would use the doctors as their own controls. This crossover design is

robust and meets many of the challenges mentioned earlier. But it has one statistical disadvantage: if the doctors who receive the intervention first continue to improve throughout the observation period, and hence become better in the last period, C, than in the middle period, B, even though they have no training in-between, this will decrease the calculated effect of the intervention. The opposite will happen if they forget their training over time, which will inflate the result.

10.3.4 Findings in the main study

The main finding was an improvement in communication, as assessed by the Four Habits Coding Scheme, of 7.5 points. This shows that the training did change the doctors' behaviour in a positive direction, but it also raises two obvious questions. The first question relates to the magnitude of this increase as compared to other interventions, and the second relates to which communication skills that have improved the most or the least. The main reasons why it is difficult to compare the increase in score to other intervention studies is that there as of today are multiple coding tools and no agreed method on how to evaluate communication skills. When we look at the eight studies from Rao et al.'s review from 2007 which had practice of practicing physicians as part of the intervention, we se that they all used different methods. One counted different behaviours based on the training manual in the study (Roter et al., 1995), one used the Roter coding system sorting utterances into 40 categories, and also counted the frequency with which physicians elicited all of a patient's concerns (Joos et al., 1996), one investigated patients' ratings using the Art of Medicine survey (Brown et al., 1999), one used the GATHARES-CP 13 item dichotomised coding tool (Moral et al., 2001), one counted utterances following the medical interaction process system (MIPS) (Fallowfield et al., 2002), one utilized the scale OPTION (observing patient involving), a standardized score from 12 observed behaviours (Elwyn et al., 2004), one looked at time spent and number of tests, referrals, and prescribed medications (Margalit et al., 2004), and one investigated

care seeking behaviour of mothers of sick children (Mohan et al., 2004). This illustrates why comparing effect to other studies will not be very fruitful at this stage, and it also illustrates the need for agreement around one or a few coding tools in the near future. If we, nevertheless, should evaluate the effect of the intervention, we note that it was 75% of the standard deviation at baseline, which is close to Cohen's criterion for a large effect (80%, d = 0.8) (Cohen, 1988).

As to the second question that arose, there are two main reasons why we do not try to conclude what habits or single items that improved. First, the study was not designed to differentiate between the different habits or single items. Second, we encouraged the doctors not to try to improve everything at once. Instead they were encouraged to focus on a few skills of their own choice and practice these. Their decision on what to focus on was based on individual preferences – but it is likely they first focused on elements where they expected the most improvement and highest payoff when related to how much effort they could put into it. Caution should therefore be shown if looking for improvements on specific items. The course offers training in all the central elements in communication skills. The improvement we found, regardless of what items might have contributed more, I see as an expression of where improvement was the easiest to achieve for these doctors, rather than as an indication of what behaviour this training improves. In another group with different references and baseline skills, the group might adopt different skills. Another way of looking at it is that even if this training model caused improvements in certain skills, next time the doctors would receive training after the same model they might acquire other skills, as now there would be other areas more easily improved. I would like to emphasise the significant and large improvement for doctors who had had prior communication skills training (Table 2 in Paper III) and argue that the more often the doctors are allowed to train, and the more focus is given to communication skills, the better the effect and larger the return on the investment will be.

We found neither improvement nor deterioration over time - the calculated average was slightly lower in period C than period B for doctors who received training between period A and B, but not significantly. We did not find any evidence to support that doctors would continue to improve by themselves after being given a training course. Supported both by the interviews with the doctors after three months and the score in period C, it is more likely that, unless there is also a focus on communication skills in the daily practice from others besides the doctors who have taken a course, the doctors will fall back to old habits over time.

10.3.5 4HPQ used to identify poorly performing doctors

When describing a level of communication skills, there is an intuitive ring to asking the patients about the effectiveness of the doctors' communication, rather than having a third party interpreting this aspect. In our study, as frequently reported elsewhere, the patient questionnaires suffer from low variability and the ceiling effect.

To meet the challenge of highly skewed answers in patient questionnaires, some argue that only reporting the proportion of the top ratings given by patients is more useful than summarizing scores via means (Makoul et al., 2007). We chose not to do this as that was not how the questionnaire was designed and we would lose information in the process.

10.3.6 Interrater reliability and training of raters

There are different ways to look at agreement between the raters when using the Four Habits Coding Scheme. In addition to the method of looking at agreement on the total score as described in 8.9, across multiple videos, one can also look at agreement within one coding sheet, representing only one video. This would show agreement across the 23 items with the five-item scale from 1 (not very effective) to 5 (highly effective). It is also possible to calculate interrater reliability across many videos item by item – also with the five-item scale.

We decided to use the Four Habits Coding Scheme total score for several reasons. First, the main object of the study was to study the overall effect of communication training; hence, looking at the overall score is intuitive and natural also when it comes to looking at agreement between raters. Second, the distribution of the scores on single items were skewed, not giving a normal distribution. Third, the variable when looking at single items is a discrete ordinal variable, not well suited for calculations of consistency estimates of interrater reliability like Pearson's r and ICC. When adding the scores for each of the 23 items, we treat this sum score as a ratio-scaled variable with a higher likelihood of normal distribution and more suited for mathematical calculations.

Although the raters agree satisfactorily, according to the calculations of the interrater reliability, one must consider the uncertainty represented if the raters are observers and are not in fact experiencing the encounters themselves. It is a risk that scoring can contribute to a distorted image of what is going on, for example, if things go unsaid because the patient is well known to the doctor, or due to other mechanisms that cause communication to not be picked up by the raters when watching video recordings. The fact that all our raters were experienced psychology students might also influence their ratings, compared to how ratings would have been done with raters with a different academic background. It is difficult to predict what direction raters with a different background might take in terms of severity of rating. However, in our experiences during the training courses for the raters, both PG and myself – both educated as medical doctors – rated doctors more severely than the raters who coded the videotapes. This speaks against any fear of our raters being too severe.

We concluded ICC to be our preferred method to calculate interrater reliability in similar studies in the future. One advantage of ICC is that it gives one value for the interrater reliability of multiple raters. One disadvantage can be that it incorporates both rater correlation and rater severity. Pearson's r is solely a measure for

correlation between raters and indicates correlation regardless of raters' severity. This means that a low Pearson's r means low correlation, but a low ICC means a difference in rater severity or low correlation, or both. This makes it more difficult to determine the source if interrater reliability is low. ICC was suitable in our study, despite the comments above, and we feel confident when recommending ICC for investigating interrater reliability in similar studies.

We spent more time on rating than we expected. Part of this was due to the complexity of the material, but much of the time spent was related to our raters being very thorough when it came to documenting *why* they would give the score they did. In future research, time can be saved by better instructing the raters about how much they are expected to document. About halfway through our rating period, when the raters were informed that it was not necessary to document their ratings to the same extent they had done while discussing ratings and establishing interrater reliability, rating sped up quite noticeably – without any effect on interrater reliability. One can argue that raters were then experienced enough to reduce their note-taking, but when asked, the coders stated that a lot of the notes were made to document what they did, not to ensure a more accurate score. We should have emphasised that documentation was only required for training and for internal discussion – not while coding independently.

10.3.7 Patient informed consent

When it comes to valid, informed consent from patients, the advantage of our approach was that after the encounter, when final consent was given, the patient knew what was shown on the videotape and could relate the decision whether or not to participate. Arguments against this would be that the recruiting researchers might adapt a more superficial style of oral information, referring to the possibility of regretting that the encounter had been filmed, and also perhaps downplay the cons of participating. Persuading recruiters, however, is a well-known problem in clinical

research. One can argue that patients would have a more difficult time withdrawing after they have initially said yes to having the encounter filmed. On the other hand, one can also argue it will be easier for patients to withdraw after they know they are unlikely to see the researcher again and just can send an answer by SMS. For most patients, and at least for all outpatient clinic visits, that means they can give their answer after they have left the hospital.

I am confident that the amount of persuasion in our study was not greater than in other studies where patients are met with information, then given time to consider the implications, and asked for consent after 24 hours. We did not put more pressure on patients than that which was necessarily exerted by the mere fact that we appeared and asked them. Some patients did initially decline, and some also declined after their initial, preliminary positive response. This indicates that the pressure was acceptable and within ethical guidelines. What struck me as the most important impression in the process of recruiting patients was, however, not that patients needed to be persuaded. It was rather the contrary. Despite what I would describe as a natural reluctance to be videotaped during their encounter with a doctor, most patients were very positive towards our research and were more than willing to contribute. Expressions stating how important they found it that someone focused on doctors' communication were quite overwhelming, and I will state here my belief that the high percentage of patients who consented is not an expression of persuasion, but rather a signal of how important patients found it that we were focusing on doctors' communication skills.

10.4 Suggested research in the future

One major incitement for investigating whether patient questionnaires can replace video-observations of encounters is the large resources it takes to code a full-length encounter. Recently the use of "thin slices" when coding with RIAS – where only segments of the video are coded and the results are considered representative for

what is happening in the rest of the encounter - have received increased attention (Roter et al., 2011). This method is evolving. Hopefully “slicing it thin” will overcome the challenge of the time-consuming process of coding with RIAS. This can save resources and make it easier to code encounters using both RIAS and Four Habits Coding Scheme, and thus provide an even more complete description of doctors’ communication skills in the future.

Despite our sparse findings when investigating the Four Habits Patient Questionnaire and its ability to identify poorly performing doctors, exploring patients’ experiences is important in order to register as much information around the interaction between the patient and the doctor as possible. One patient questionnaire that has found momentum is the communication assessment tool (CAT) (Makoul et al., 2007). 20,000 questionnaires were recently collected over the time period of one month and analyses are being done to identify areas within doctors’ communication that should be prioritized in communication training (Makoul, 2010). This, however, only identifies sub-items within clinical communication that are given the lowest patient scores. It does not identify which doctors to train. Further studies need to be done to clarify the degree to which new or improved survey measures can be used to identify poorly performing doctors. A suggestion for future research would be to systematically validate patient questionnaires according to Epstein et al. by using related scales with the same patient population, ideally randomizing the order of items and utilizing a similar response format. At the same time, researchers should validate the tools against observations in a manner similar to paper IV in our study – at least for national indicators.

Another interesting area of research would be to have patients rate the video recording with the patient questionnaire, rather than filling out the questionnaire by memorizing what went on in the encounter. This might help to answer whether the high ratings given by patients represent a correct measure of the communication in

the encounter or to what extent it is difficult for patients to meta-observe themselves while taking part in the encounter themselves.

This thesis fits in with the research suggested by Rao in 2007 (Rao et al., 2007), showing that a general communication course, organized within the frames of the daily routine of a hospital, improved doctor communication behaviour. Soon, the next step within this area should be to move the level of focus and intervention up from each individual doctor to the level of the department – comparing one department in one hospital that is given communication skills training to a similar department in another hospital. This will make it possible to reliably investigate the effect of communication training on variables involving health outcomes.

11 Conclusions

The conclusions below are deducted from a randomized controlled trial with crossover design utilizing 497 videotaped encounters with a patient participation rate of 87%. The high rate can mainly be credited to our new approach for obtaining valid, informed consent using confirmation of a preliminary consent using SMS.

The Four Habits Coding Scheme is suitable when coding a diverse material from a hospital setting. More than two raters are needed for coding materials in larger studies.

A 20-hour general communication skills training course for hospital doctors across medical specialties showed an improving effect on the doctors' communication skills when assessed by coding of videotaped encounters with real patients.

The Four Habits Patient Questionnaire, addressing patient experiences with doctors' communication skills, could not identify all the poorly performing doctors that were identified by video coding. Following my findings I suggest that, instead of trying to identify poorly performing doctors in order to train them, all doctors should receive communication skills training on a regular basis.

12 References

AHUS. 8. Mar 2011. *RE: AHUS key figures.* Type to FOSSLI JENSEN B.

ANDERSON, L. A. & SHARPE, P. A. 1991. Improving patient and provider communication: A synthesis and review of communication interventions. *Patient Education and Counseling,* 17**,** 99-134.

ARING, C. D. 1958. Sympathy and empathy. *J Am Med Assoc,* 167**,** 448-52.

ASPEGREN, K. 1999. BEME Guide No. 2: Teaching and learning communication skills in medicine-a review with quality grading of articles. *Medical Teacher,* 21**,** 563 - 570.

BARON, R. J. 1985. An introduction to medical phenomenology: I can't hear you while I'm listening. *Ann Intern Med,* 103**,** 606-11.

BECKER, H. S., GEER, B., HUGHES, E. C. & STRAUSS, A. L. 1961. *Boys in white: student culture in medical school,* Chicago, Ill, University of Chicago Press.

BLANCH-HARTIGAN, D., HALL, J. A., ROTER, D. L. & FRANKEL, R. M. 2010. Gender bias in patients' perceptions of patient-centered behaviors. *Patient Educ Couns,* 80**,** 315-20.

BROWN, J. B., BOLES, M., MULLOOLY, J. P. & LEVINSON, W. 1999. Effect of clinician communication skills training on patient satisfaction. A randomized, controlled trial. *Ann Intern Med,* 131**,** 822-9.

CAHPS®. Consumer Assessment of Healthcare Providers and Systems. Available: https://www.cahps.ahrq.gov/content/products/PDF/PocketGuide.htm [Accessed Version: 9/21 2005].

CICOUREL, A. V. 1980. Three models of discourse analysis: The role of social structure. *Discourse Process,* 3**,** 101-132.

COHEN, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* 20**,** 37-46.

COHEN, J. 1988. *Statistical Power Analysis for the Behavioral Sciences (Second Edition)*, Lawrence Erlbaum Associates.

COX, A., HOLBROOK, D. & RUTTER, M. 1981a. Psychiatric interviewing techniques. VI. Experimental study: eliciting feelings. *Br J Psychiatry,* 139**,** 144-52.

COX, A., HOPKINSON, K. & RUTTER, M. 1981b. Psychiatric interviewing techniques II. Naturalistic study: eliciting factual information. *Br J Psychiatry,* 138**,** 283-91.

COX, A., RUTTER, M. & HOLBROOK, D. 1981c. Psychiatric interviewing techniques V. Experimental study: eliciting factual information. *Br J Psychiatry,* 139**,** 29-37.

ELWYN, G., EDWARDS, A., HOOD, K., ROBLING, M., ATWELL, C., RUSSELL, I., WENSING, M. & GROL, R. 2004. Achieving involvement: process outcomes from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. *Fam Pract,* 21**,** 337-46.

ENGEL, G. L. 1977. The need for a new medical model: a challenge for biomedicine. *Science,* 196**,** 129-36.

EPSTEIN, R. M., FRANKS, P., FISCELLA, K., SHIELDS, C. G., MELDRUM, S. C., KRAVITZ, R. L. & DUBERSTEIN, P. R. 2005. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. *Soc Sci Med,* 61**,** 1516-28.

ERICKSON, F. 1982. Audiovisual Records as a Primary Data Source. *Sociological Methods & Research,* 11**,** 213-232.

FALLOWFIELD, L., JENKINS, V., FAREWELL, V., SAUL, J., DUFFY, A. & EVES, R. 2002. Efficacy of a Cancer Research UK communication skills training model for oncologists: a randomised controlled trial. *Lancet,* 359**,** 650-6.
FOX, R. 1959. *Experiment Perilous,* New York, NY, Free Press.
FRANCIS, V., KORSCH, B. M. & MORRIS, M. J. 1969. Gaps in doctor-patient communication. Patients' response to medical advice. *N Engl J Med,* 280**,** 535-40.
FRANKEL, R. M. & BECKMAN, H. B. 1982. *Impact: An Interaction-Based Method for Preserving and Analyzing Clinical Transactions,* Nashville, Tenn, Humana Inc.
FRANKEL, R. M. & STEIN, T. 2001. Getting the most out of the clinical encounter: the four habits model. *J Med Pract Manage,* 16**,** 184-91.
GARRATT, A. M., BJAERTNES, O. A., KROGSTAD, U. & GULBRANDSEN, P. 2005. The OutPatient Experiences Questionnaire (OPEQ): data quality, reliability, and validity in patients attending 52 Norwegian hospitals. *Qual Saf Health Care,* 14**,** 433-7.
GARRATT, A. M., SOLHEIM, E. & DANIELSEN, K. 2008. National and cross-national surveys of patient experiences: a structured review. Report No 7-2008. *Oslo: Norwegian Knowledge Centre for the Health Services*.
GLASER, B. & STRAUSS, A. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research* Chicago, Ill, Aldine.
GRIFFIN, E. A. 2009. *A first look at communication theory,* New York, McGraw-Hill Higher Education.
GULBRANDSEN, P., JENSEN, B. F. & FINSET, A. 2009. [Self-efficacy among doctors in hospitals after a course in clinical communication]. *Tidsskr Nor Laegeforen,* 129**,** 2343-6.
GULBRANDSEN, P., KRUPAT, E., BENTH, J. S., GARRATT, A., SAFRAN, D. G., FINSET, A. & FRANKEL, R. 2008. "Four Habits" goes abroad: report from a pilot study in Norway. *Patient Educ Couns,* 72**,** 388-93.
GULBRANDSEN, P., MADSEN, H. B., BENTH, J. S. & LAERUM, E. 2010. Health care providers communicate less well with patients with chronic low back pain--a study of encounters at a back pain clinic in Denmark. *Pain,* 150**,** 458-61.
HOPKINSON, K., COX, A. & RUTTER, M. 1981. Psychiatric interviewing techniques. III. Naturalistic study: eliciting feelings. *Br J Psychiatry,* 138**,** 406-15.
INUI, T. S. & CARTER, W. B. 1985. Problems and prospects for health services research on provider-patient communication. *Med Care,* 23**,** 521-38.
INUI, T. S., YOURTEE, E. L. & WILLIAMSON, J. W. 1976. Improved outcomes in hypertension after physician tutorials. A controlled trial. *Ann Intern Med,* 84**,** 646-51.
ISRCTN. Available: http://www.controlled-trials.com/ISRCTN22153332/ISRCTN22153332 [Accessed 8. Mar 2011].
JOOS, S. K., HICKAM, D. H., GORDON, G. H. & BAKER, L. H. 1996. Effects of a physician communication intervention on patient care outcomes. *J Gen Intern Med,* 11**,** 147-55.
KORSCH, B. M., FREEMON, B. & NEGRETE, V. F. 1971. Practical implications of doctor-patient interaction analysis for pediatric practice. *Am J Dis Child,* 121**,** 110-4.
KORSCH, B. M. & NEGRETE, V. F. 1972. Doctor-patient communication. *Sci Am,* 227**,** 66-74.
KRUPAT, E., FRANKEL, R., STEIN, T. & IRISH, J. 2006. The Four Habits Coding Scheme: validation of an instrument to assess clinicians' communication behavior. *Patient Educ Couns,* 62**,** 38-45.
LAZARE, A., PUTNAM, S. M. & LIPKIN JR, M. 1995a. *The medical interview : clinical care, education, and research / Mack Lipkin, Jr., Samuel M. Putnam, Aaron*

Lazare, editors ; J. Gregory Carroll, Richard M. Frankel, associate editors ; Allen Keller, Terri Klein, P. Kay Williams, assistant editors, New York, Springer-Verlag.
LAZARE, A., PUTNAM, S. M. & LIPKIN JR, M. 1995b. Three functions of the medical interview. *In:* LIPKIN JR, M., PUTNAM, S. M. & LAZARE, A. (eds.) *The Medical Interview : clinical care, education, and research.* New York: Springer-Verlag.
LEVINSON, W. & ROTER, D. 1993. The effects of two continuing medical education programs on communication skills of practicing primary care physicians. *J Gen Intern Med,* 8**,** 318-24.
LIEF, H. & FOX, R. C. 1963. *Training for `detached concern' in medical students. In: Lief HI, Lief VF, Lief NR, eds. The psychological basis of medical practice,* New York, Harper & Row.
MAKOUL, G. 2001. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Academic Medicine,* 76**,** 390-3.
MAKOUL, G. Year. Priorities for effective communication with patients: patterns emerging from CAT data. *In:* INTERNATIONAL CONFERENCE ON COMMUNICATION IN HEALTHCARE 2010 09/07 2010 Verona, Italy.
MAKOUL, G., KRUPAT, E. & CHANG, C. H. 2007. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. *Patient Educ Couns,* 67**,** 333-42.
MARGALIT, A. P., GLICK, S. M., BENBASSAT, J. & COHEN, A. 2004. Effect of a biopsychosocial approach on patient satisfaction and patterns of care. *J Gen Intern Med,* 19**,** 485-91.
MCCONNELL, A. R., RYDELL, R. J., STRAIN, L. M. & MACKIE, D. M. 2008. Forming implicit and explicit attitudes toward individuals: social group association cues. *J Pers Soc Psychol,* 94**,** 792-807.
MERTON, R., READER, C. G. & KENDALL, P. L. 1957. *The Student Physician,* Cambridge, Mass Harvard University Press.
MISHLER, E. G. 1984. *The Discourse of Medicine: Dialectics of Medical Interviews,* Norwood, NJ, Ablex Publishing Corporation.
MOHAN, P., IYENGAR, S. D., MARTINES, J., COUSENS, S. & SEN, K. 2004. Impact of counselling on careseeking behaviour in families with sick children: cluster randomised trial in rural India. *BMJ,* 329**,** 266.
MORAL, R. R., ALAMO, M. M., JURADO, M. A. & DE TORRES, L. P. 2001. Effectiveness of a learner-centred training programme for primary care physicians in using a patient-centred consultation style. *Fam Pract,* 18**,** 60-3.
MUMFORD, E. 1970. *Interns: From students to physicians,* Cambridge, Mass, Harvard University Press
MUMFORD, E., SCHLESINGER, H. J. & GLASS, G. V. 1982. The effect of psychological intervention on recovery from surgery and heart attacks: an analysis of the literature. *Am J Public Health,* 72**,** 141-51.
OTTOSSON, J.-O. 1999. Patient-läkarrelationen. Läkekonst på vetenskaplig grund. *In:* OTTOSSON, J.-O. (ed.). Stockholm: Natur ock Kultur.
OXFORD, D. 2010. *Clinical - From Oxford Dictionaries* [Online]. Available: http://oxforddictionaries.com/view/entry/m_en_gb0155110#m_en_gb0155110 [Accessed 30. Sep 2010].
RAO, J. K., ANDERSON, L. A., INUI, T. S. & FRANKEL, R. M. 2007. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. *Med Care,* 45**,** 340-9.
ROLAND, M., ELLIOTT, M., LYRATZOPOULOS, G., BARBIERE, J., PARKER, R. A., SMITH, P., BOWER, P. & CAMPBELL, J. 2009. Reliability of patient responses in pay for performance schemes: analysis of national General Practitioner Patient Survey data in England. *BMJ,* 339**,** b3851.

ROTER, D. & FRANKEL, R. 1992. Quantitative and qualitative approaches to the evaluation of the medical dialogue. *Soc Sci Med,* 34**,** 1097-103.
ROTER, D. L., HALL, J. A., BLANCH-HARTIGAN, D., LARSON, S. & FRANKEL, R. M. 2011. Slicing it thin: New methods for brief sampling analysis using RIAS-coded medical dialogue. *Patient Educ Couns*.
ROTER, D. L., HALL, J. A., KERN, D. E., BARKER, L. R., COLE, K. A. & ROCA, R. P. 1995. Improving physicians' interviewing skills and reducing patients' emotional distress. A randomized clinical trial. *Arch Intern Med,* 155**,** 1877-84.
ROUF, E., CHUMLEY, H. & DOBBIE, A. 2009. Patient-centered interviewing and student performance in a comprehensive clinical skills examination: is there an association? *Patient Educ Couns,* 75**,** 11-5.
RUTTER, M. & COX, A. 1981. Psychiatric interviewing techniques: I. Methods and measures. *Br J Psychiatry,* 138**,** 273-82.
RUTTER, M., COX, A., EGERT, S., HOLBROOK, D. & EVERITT, B. 1981. Psychiatric interviewing techniques. IV. Experimental study: Four contrasting styles. *Br J Psychiatry,* 138**,** 456-65.
SIMPSON, M., BUCKMAN, R., STEWART, M., MAGUIRE, P., LIPKIN, M., NOVACK, D. & TILL, J. 1991. Doctor-patient communication: the Toronto consensus statement. *BMJ,* 303**,** 1385-7.
SITZIA, J. 1999. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care,* 11**,** 319-28.
STEIN, T., FRANKEL, R. M. & KRUPAT, E. 2005. Enhancing clinician communication skills in a large healthcare organization: a longitudinal case study. *Patient Educ Couns,* 58**,** 4-12.
STEMLER, S. E. 2004. A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. Practical Assessment, Research & Evaluation, Retrieved February 28, 2010 from http://PAREonline.net/getvn.asp?v=9&n=4 Available: http://pareonline.net/getvn.asp?v=9&n=4 [Accessed February 28, 2010].
STEWART, M., BROWN, J. B., BOON, H., GALAJDA, J., MEREDITH, L. & SANGSTER, M. 1999. Evidence on patient-doctor communication. *Cancer Prev Control,* 3**,** 25-30.
STOECKLE, J. D. & BILLINGS, J. A. 1987. A history of history-taking: the medical interview. *J Gen Intern Med,* 2**,** 119-27.
WAGER, E., TOOLEY, P. J., EMANUEL, M. B. & WOOD, S. F. 1995. How to do it. Get patients' consent to enter clinical trials. *BMJ,* 311**,** 734-7.
WEBSTER'S, D. 2010a. *Information Bias - From Webster's Online Dictionary* [Online]. Webster's Online Dictionary. Available: http://www.websters-online-dictionary.org/definitions/information+bias?cx=partner-pub-0939450753529744%3Av0qd01-tdlq&cof=FORID%3A9&ie=UTF-8&q=information+bias&sa=Search#962 [Accessed December 6. 2010].
WEBSTER'S, D. 2010b. *Selection Bias - From Webster's Online Dictionary* [Online]. Webster's Online Dictionary. Available: http://www.websters-online-dictionary.org/definitions/selection+bias?cx=partner-pub-0939450753529744%3Av0qd01-tdlq&cof=FORID%3A9&ie=UTF-8&q=selection+bias&sa=Search#914 [Accessed December 6. 2010].
WEST, C. 1984. *Routine Complications: Troubles with Talk Between Doctors and Patients,* Bloomington, Ind, Indiana University Press.
WIKIPEDIA. 2010. *Interrater reliability - From Wikipedia, the free encyclopedia* [Online]. Available: http://en.wikipedia.org/wiki/Inter-rater_reliability [Accessed September 30. 2010].

I

II

III

IV